

Bildungspotenziale erkennen und fördern.

Eine Machbarkeitsstudie

Prof. Dr. Knut Schwippert, Dr. Daniel Kasper, Dr. Katrin Schulz-Heidorf & Jana Steen, M.A.

Universität Hamburg

8. Januar 2020

Prof. Dr. Knut Schwippert  
Universität Hamburg  
Fakultät für Erziehungswissenschaft  
Postadresse: Von-Melle-Park 8  
20146 Hamburg  
Besucheradresse: Alsterterrasse 1  
20354 Hamburg

Dr. Daniel Kasper  
Universität Hamburg  
Fakultät für Erziehungswissenschaft  
Postadresse: Von-Melle-Park 8  
20146 Hamburg  
Besucheradresse: Alsterterrasse 1  
20354 Hamburg

Dr. Katrin Schulz-Heidorf  
Universität Hamburg  
Fakultät für Erziehungswissenschaft  
Postadresse: Von-Melle-Park 8  
20146 Hamburg  
Besucheradresse: Alsterterrasse 1  
20354 Hamburg

Jana Steen, M.A. (bis 30.06.2019)  
Universität Hamburg  
Fakultät für Erziehungswissenschaft  
Postadresse: Von-Melle-Park 8  
20146 Hamburg  
Besucheradresse: Alsterterrasse 1  
20354 Hamburg

# Inhalt

1	Zusammenfassung.....	1
2	Darstellung der Ausgangslage .....	3
2.1	Vergleichbare Zugangschancen im Bildungssystem.....	4
2.2	Kompetenzstände von Schülerinnen und Schülern .....	9
2.3	Bildungsdisparitäten von Schülerinnen und Schülern.....	11
3	Perspektiven und Ableitungen .....	12
3.1	Perspektive: Vergleichbarkeit von Querschnittanalysen ermöglichen .....	12
3.2	Perspektive: Soziale Bildungsdisparitäten.....	13
3.3	Zwischenfazit und Ableitung einer allgemeinen Fragestellung.....	13
4	Empirische Modellierung .....	14
4.1	Methodisches Vorgehen .....	14
4.2	Theoretisches Rahmenkonzept .....	15
4.3	Datenmaterial.....	18
4.4	Equating-Verfahren .....	18
4.5	Exemplarische Analysen .....	19
4.5.1	Daten und Variablen.....	19
4.5.2	Equating.....	20
4.5.3	Identifizierung von Schülerinnen und Schülern mit erwartungswidrigen Leistungen ..	20
5	Befunde der exemplarischen Analysen .....	21
6	Zusammenfassende inhaltliche Interpretation der Modellierung .....	23
7	Fazit und Ausblick .....	24
	<b>Literatur</b> .....	<b>26</b>
	<b>Anhang A: Für Equating nutzbare Items</b> .....	<b>30</b>
	<b>Anhang B: Equating-/Linking-Verfahren</b> .....	<b>32</b>
	<b>Anhang C: Rechnerische Umsetzung des Equatings</b> .....	<b>34</b>

# 1 Zusammenfassung

Zentrales Bemühen bei der Bildung und Ausbildung von Kindern und Jugendlichen ist es, ihnen – im Sinne von Klafki (vgl. 2007) – die Möglichkeit zu eröffnen, ein selbstbestimmtes Leben zu führen und gleichzeitig an der Weiterentwicklung der Gesellschaft mitzuwirken. Hinter diesem Anspruch steht ein Bildungsverständnis, das kognitive, emotionale, ästhetische, soziale und praktisch-technische Dimensionen umfasst und zugleich auf konkreten Kulturtechniken basiert, die im Bildungssystem vermittelt werden müssen. Dazu gehören neben Lese- und Schreibkompetenzen auch Kompetenzen in mathematischen und naturwissenschaftlichen Fächern sowie inzwischen auch der kompetente Umgang mit digitalen Medien.

Eine oder mehrere dieser Kompetenzen nicht zu besitzen und bestimmte Lernfortschritte (z.B. Übergang in ein Gymnasium) sowie Abschlusserfolge (z.B. die allgemeine Hochschulreife) nicht zu erreichen, hat weitreichende Folgen für den Einzelnen. Es kann bedeuten, dass man an bestimmten Teilen des gesellschaftlichen Lebens nicht teilnehmen kann. Verschiedene Studien (TIMSS, PIRLS und PISA)<sup>1</sup> zeigen seit Jahren, dass Lernerfolge, das Erreichen von Bildungsabschlüssen sowie der Erwerb von Kompetenzen nicht nur von den kognitiven Voraussetzungen oder dem Vorwissen der Schülerinnen und Schüler abhängen, sondern dass diese vielmehr auch maßgeblich durch ihren sozialen Status bedingt werden (vgl. Reiss, Sälzer, Schiepe-Tiska, Klieme & Köller, 2016; Wendt, Bos et al., 2016; Hußmann, Wendt et al., 2017). Zugleich lassen sich Lern- und Abschlusserfolge nicht nur durch die erreichten Kompetenzen erklären. Die sich hier zeigenden Diskrepanzen sind als Bildungsdisparitäten anzusehen.

Bildungsdisparitäten lassen sich dabei zum einen entlang des sozialen Status beobachten, d.h. Schülerinnen und Schüler mit höherem sozialen Status verzeichnen in der Regel mehr Lernerfolge, erreichen höherwertige Bildungsabschlüsse und erwerben mehr Kompetenzen als Schülerinnen und Schüler mit niedrigem sozialen Status. Zum anderen lassen sich Bildungsdisparitäten jedoch auch innerhalb von Gruppen mit gleichem sozialen Status beobachten, d.h. von Schülerinnen und Schülern mit vergleichbarem sozialen Status weisen einige erwartungswidrig hohe Bildungserfolge andere erwartungswidrig niedrige Bildungserfolge auf. In dieser Untersuchung werden diese beiden Gruppen (Schülerinnen und Schüler mit erwartungswidrig hohen Leistungen und Schülerinnen und Schüler mit erwartungswidrig niedrigen Leistungen) in den Blick genommen, um herauszufinden in welchen Merkmalen, die sich in Bezug auf Lernprozesse kurz bzw. mittelfristig verändern lassen, sich diese beiden Gruppen unterscheiden. Durch die Betrachtung der Extremgruppen lassen sich die Merkmale besser identifizieren, die den Unterschied bei der Ausschöpfung von Bildungspotenzialen erläutern. Dabei wird der Frage nachgegangen, ob sich Merkmale finden, die bei systematischer Variation zur Ausschöpfung von Bildungspotenzialen beitragen können.

Die vorliegende Machbarkeitsstudie<sup>2</sup> stellt dabei einen möglichen Ausgangspunkt für einen perspektivisch und sequenziell angelegten Projektplan dar. Zentrales Ziel des sequenziell angelegten Projektplans ist es, kurz bzw. mittelfristig veränderbare Merkmale der Schulsteuerung zu identifizieren. Zentrales Ziel der Machbarkeitsstudie ist es, einen Forschungsrahmen zu skizzieren, mit dem dieses Ziel erreicht werden kann und der damit eine mögliche Anschlussperspektiven bietet. Ausgehend von

---

<sup>1</sup> Trends in International Mathematics and Science Study (TIMSS), Progress in International Reading Literacy Study (PIRLS), Programme for International Student Assessment (PISA).

<sup>2</sup> Wir danken der Bertelsmann Stiftung, die aufgrund ihrer Zuwendung von 2018 bis 2019 diese Machbarkeitsstudie ermöglicht hat.

der Darstellung der Ausgangslage und dem Aufzeigen relevanter Veränderungen in der Schullandschaft wird cursorisch der Forschungsstand zu Bildungsdisparitäten zusammengefasst (Kapitel 2). Im Anschluss werden zwei Perspektiven der Studie skizziert: Zum einen die Perspektive, Querschnittsdaten verschiedener Studien miteinander vergleichbar zu machen, zum anderen die Perspektive, den sozialen Hintergrund der Schülerinnen und Schüler in den Vordergrund zu rücken. Basierend auf diesen Vorstellungen werden die allgemeinen Fragestellungen der Studie vorgestellt (Kapitel 3). Im Kern der Machbarkeitsstudie stehen die Darstellung des theoretischen Rahmens und die Darstellungen von Überlegungen über mögliche Analysestrategien (Kapitel 4). In Kapitel 5 werden die Ergebnisse der Machbarkeitsstudie zusammengefasst; die darauf basierende inhaltliche Interpretation wird in Kapitel 6 dargestellt. Ein Fazit und Ausblick schließen diesen Bericht ab (Kapitel 7).

Für diesen Bericht wurde geprüft, ob und inwiefern die folgenden Fragen in einer ausführlichen Studie bearbeitet werden können und sollen:

- 1.) Wie sieht der Ist-Stand in Bezug auf Bildungsdisparitäten im Bildungssystem aus?
- 2.) Wie fallen die Leistungsunterschiede von Schülerinnen und Schülern je nach sozialem Hintergrund, Geschlecht und Migrationsstatus aus?
- 3.) Wie sehen die Bildungsdisparitäten aus, wenn man den sozialen Hintergrund, das Geschlecht und den Migrationsstatus kontrolliert?
- 4.) Welche Merkmale, die sich in Bezug auf Lernprozesse verändern lassen, tragen dazu bei, diese Unterschiede zu erklären?
- 5.) Welche weiteren Merkmale lassen sich identifizieren, die zum Abbau der Bildungsdisparitäten beitragen können?
- 6.) Wie kann einem Entstehen bzw. einer Vergrößerung von Bildungsdisparitäten präventiv begegnet werden?

Die Erkenntnisse der Machbarkeitsstudie belegen, dass die Bearbeitung der zuvor genannten Fragestellungen im Rahmen einer ausführlichen Anschlussstudie zielführend erscheint. Dabei erscheinen uns insbesondere die folgenden Perspektiven vielversprechend.

- Der gewählte Fokus auf zwei Extremgruppen von Schülerinnen und Schülern, nämlich denen, die deutlich hinter den von ihnen erwartbaren Leistungen zurückbleiben versus denen, die die erwarteten Leistungen übertreffen, sollte weiterverfolgt werden.
- Zunächst sollte die weitere, über die hier gewählten exemplarischen Daten- und Themenzugänge hinausgehende Analyse der vorliegenden Primarstufendaten erfolgen und erst in einem zweiten Zyklus sollten die sich dabei bewährenden Verfahren auf die Sekundarstufe ausgeweitet werden.
- Die Möglichkeit einer sequenziell angelegten, wiederholt durchgeführten Analyse der beiden Bildungsbereiche im Sinne eines kontinuierlichen Monitorings sollte geprüft werden.
- Die Daten der amtlichen Statistik als beschreibendes Merkmal einer sich verändernden Schullandschaft und Schülerschaft sollten mit einbezogen werden.
- Das Verfahren des Equatings sollte mit einbezogen werden, um die verschiedenen zur Verfügung stehenden Studien vergleichbar zu machen.
- Bei den explorativen Analysen sollte der Fokus auf kurz- bis mittelfristig veränderbare Faktoren gerichtet werden.

In an diese Machbarkeitsstudie anschließenden Untersuchungen (z.B. anhand von Trend-Daten (IGLU,<sup>3</sup> TIMSS oder PISA) oder anhand von Panel-Daten (NEPS)) sollte weiter nach Möglichkeiten des Abbaus von Bildungsdisparitäten durch kurzzeitig veränderbare Bedingungsfaktoren gesucht werden. Weiterhin sollte analysiert werden, wie auf Schul- und Unterrichtsebene systematisch Hürden ab- und Unterstützungsfaktoren ausgebaut werden können, um alle, vor allem aber auch jene Kinder und später auch Jugendliche besser auf eine erfolgreiche Teilhabe an der Gesellschaft vorzubereiten, die ihre diesbezüglichen Bildungspotenziale bislang nicht ausschöpfen können.

## 2 Darstellung der Ausgangslage

Bildung sollte das Ziel haben, Schülerinnen und Schüler zu unterstützen, ein selbstbestimmtes Leben führen zu können und ihnen damit die Teilhabe an gesellschaftlichen Entwicklungen mit zu ermöglichen. Der Anspruch auf ein selbstbestimmtes Leben und auf die Teilhabe an gesellschaftlichen Entwicklungen sollte unabhängig von persönlichen Charakteristika wie etwa Geschlecht oder Herkunft bestehen. Das Bildungssystem muss demnach in seiner Kompetenzvermittlung so ausgelegt sein, dass zum einen der Erwerb von Kompetenzen möglich ist und zum anderen diese Möglichkeit nicht von eben solchen persönlichen Charakteristika abhängt. Oder anders formuliert: Im Bildungssystem muss das Prinzip der Chancengleichheit gelten (vgl. Schwippert, 2009).

Die Erfüllung der genannten Prämissen, und damit von Chancengleichheit im Bildungssystem, würde sich empirisch etwa in den erreichten Kompetenzständen der Schülerinnen und Schüler in verschiedenen Bereichen zeigen. Diese Kompetenzstände müssten zum einen hinreichend ausgeprägt sein, zum anderen dürften erreichbare Kompetenzstände im Bildungssystem bei sonst gleichen akademischen Voraussetzungen nicht von Merkmalen wie etwa dem sozialen Hintergrund der Schülerinnen und Schüler abhängen. Des Weiteren würde man annehmen, dass Schülerinnen und Schüler unter gleichen persönlichen (wie z.B. Vorwissen, Motivation, Interesse) und institutionellen (z.B. Verfügbarkeit von Lernmaterialien in den Schulen, Gestaltung des Unterrichts) Voraussetzungen auch dieselben Kompetenzstände erreichen. Lässt sich einer oder mehrere dieser Sachverhalte empirisch nicht beobachten, würde man schlussfolgern, dass das Prinzip der Chancengleichheit nicht erfüllt ist.

Im Fokus der Machbarkeitsstudie steht die Untersuchung des letztgenannten Sachverhaltes: Erreichen Schülerinnen und Schüler unter vergleichbaren Voraussetzungen dieselben Kompetenzstände? Wir werden in dieser Machbarkeitsstudie zeigen, dass dies nicht der Fall ist. Unter vergleichbaren schulischen Lernbedingungen lassen sich Schülerinnen und Schüler mit erwartungswidrig hohen Bildungserfolgen von Schülerinnen und Schülern mit erwartungswidrig niedrigen Bildungserfolgen unterscheiden. Um das Prinzip der Chancengleichheit erfüllen zu können, ist zu untersuchen, welche Faktoren zu dieser Situation beitragen. Deswegen soll der Frage nachgegangen werden, welche Merkmale sich in Bezug auf Lernprozesse kurz bzw. mittelfristig verändern lassen, die zur Erklärung dieser Bildungsdisparität beitragen.

Im Folgenden wird ein Überblick darüber gegeben, inwieweit sich ausgewählte Merkmale, die als Indikatoren über die Veränderungen von Zugangschancen im Bildungssystem Auskunft geben, gewandelt haben, und es wird in die Diskussion über (Bildungs-)Disparitäten eingeführt.

---

<sup>3</sup> Internationale Grundschul-Lese-Untersuchung.

## 2.1 Vergleichbare Zugangschancen im Bildungssystem

Im Bildungsbereich lassen sich Chancenunterschiede einerseits *normativ* beschreiben, nämlich z.B. als Konsens über eine privilegierte bzw. minimale materielle bzw. ideelle Ausstattung von Personen mit Bildungsgütern und andererseits *empirisch* beschreiben über relative Positionen bzw. relative Möglichkeiten im Vergleich innerhalb einer Gruppe bzw. Gesellschaft. Hier setzt formale Chancengleichheit im Bildungssystem insbesondere auch den Zugang zu angemessenen Bildungspotenzialen für alle Kinder voraus. Jedoch sind die Voraussetzungen dafür, dass jedes Kind gleichermaßen am Bildungssystem teilnehmen und sein Bildungspotenzial entfalten kann, nicht gegeben. Zudem variieren die Rahmenbedingungen im Bildungssystem über die Zeit hinweg, wie uns der Blick in die beschreibenden amtlichen Schulstatistiken zeigen wird. Hier ist es wichtig, kontinuierlich im Blick zu behalten, unter welchen unterschiedlichen Voraussetzungen Schülerinnen und Schüler in Deutschland lernen, um einen möglichen Einfluss von systemischen Bedingungsfaktoren auf (erwartungswidrige) Lernergebnisse nicht aus dem Blick zu verlieren.

Um einen ersten Einblick in sich wandelnde Situationen im Bildungsbereich zu geben, werden nachfolgend exemplarisch Indikatoren aus den amtlichen Statistiken bzw. aus zyklisch angelegten Studien vorgestellt, die auf mögliche Einschränkungen beim Vergleich von Untersuchungen aus unterschiedlichen Erhebungszeitpunkten verweisen. Exemplarisch sind hier Informationen zu den Themen Inklusion (von Schülerinnen und Schülern mit sonderpädagogischem Förderbedarf), Ausbau von Ganztagschulen und Verfügbarkeit digitaler Medien in der Schule skizziert. Der sich hierbei abzeichnende Wandel im Bildungssystem wird bei der Interpretation der Analysen als Hintergrundinformation zu berücksichtigen sein.

### *Stichwort: Inklusion*

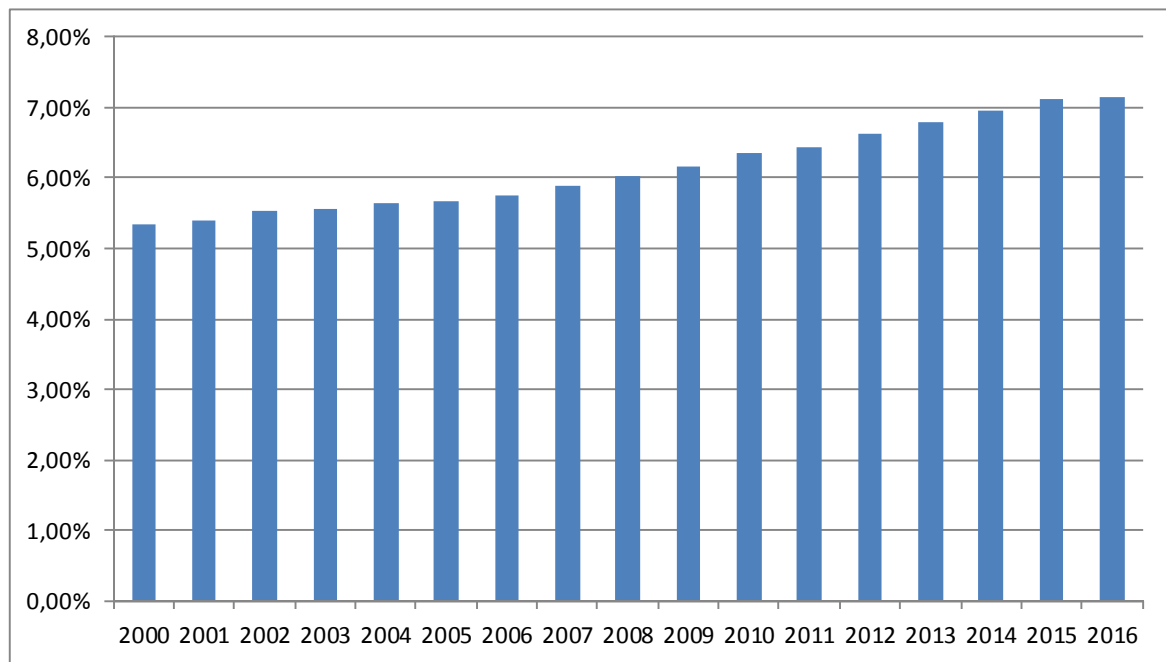
An Schulen erhobene Leistungsdaten könnten in einer Anschlussstudie als Output betrachtet und ihre Entwicklung im Trend seit Beginn der 2000er Jahre nachvollzogen werden. In diesem Zeitraum hat sich die Zusammensetzung der Schülerschaft stark verändert, insbesondere durch die voranschreitende Inklusion: Während zu Beginn der 2000er Jahre die meisten Schülerinnen und Schüler mit sonderpädagogischem Förderbedarf in Förderschulen unterrichtet wurden, sind heute viele an allgemeinen Schulen inkludiert. Zurückgeführt werden kann dies auf die 2009 auch in Deutschland in Kraft getretene UN-Behindertenrechtskonvention (BRK). Für die vorliegende Arbeit ist die Entwicklung inklusiver Beschulung von 2000 bis 2016 relevant, da Daten aus diesem Zeitraum verwendet werden. Grundlage für die nachstehenden Berechnungen sind die Daten der Kultusministerkonferenz (vgl. KMK, n.d.). Hierbei rücken insbesondere die Förderquote von Schülerinnen und Schülern, die Exklusions- und Inklusionsquoten in den Blickpunkt:

1. Förderquote:<sup>4</sup> Der Anteil der Schülerinnen und Schüler mit Förderbedarf an allen Schülerinnen und Schülern im Alter der Vollzeitschulpflicht (Klasse 1 bis 9/10 und Förderschulen) steigt im betrachteten Zeitraum kontinuierlich. Insbesondere ab 2009 wird der Anteil von Schülerinnen und Schülern, für die ein Förderbedarf diagnostiziert wird immer größer, obwohl die Gesamtzahl der Schülerinnen und Schüler in Vollzeitschulpflicht im gleichen Zeitraum abnimmt (vgl. Abb. 1).

---

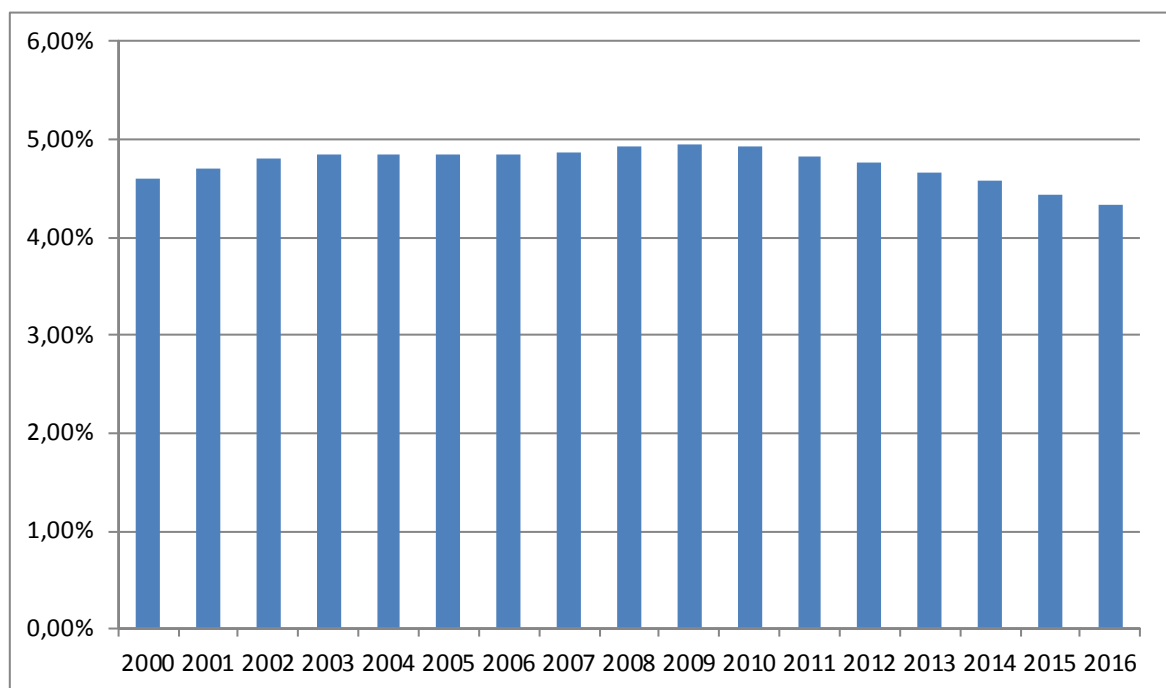
<sup>4</sup> Die Förderquote beschreibt den Anteil der Schülerinnen und Schüler, die sonderpädagogisch gefördert werden, unabhängig davon, ob sie eine Förder- oder allgemeine Schule besuchen. „Sie wird gebildet als Quotient aus der Zahl der Schüler/innen mit sonderpädagogischer Förderung und der Gesamtzahl der Schüler/innen im Primar- und Sekundarbereich I“ (KMK, 2018, S. 39).

Abbildung 1: Förderquote: Anteil der Schülerinnen und Schüler mit Förderbedarf an allen Schülerinnen und Schülern im Alter der Vollzeitschulpflicht in Prozent



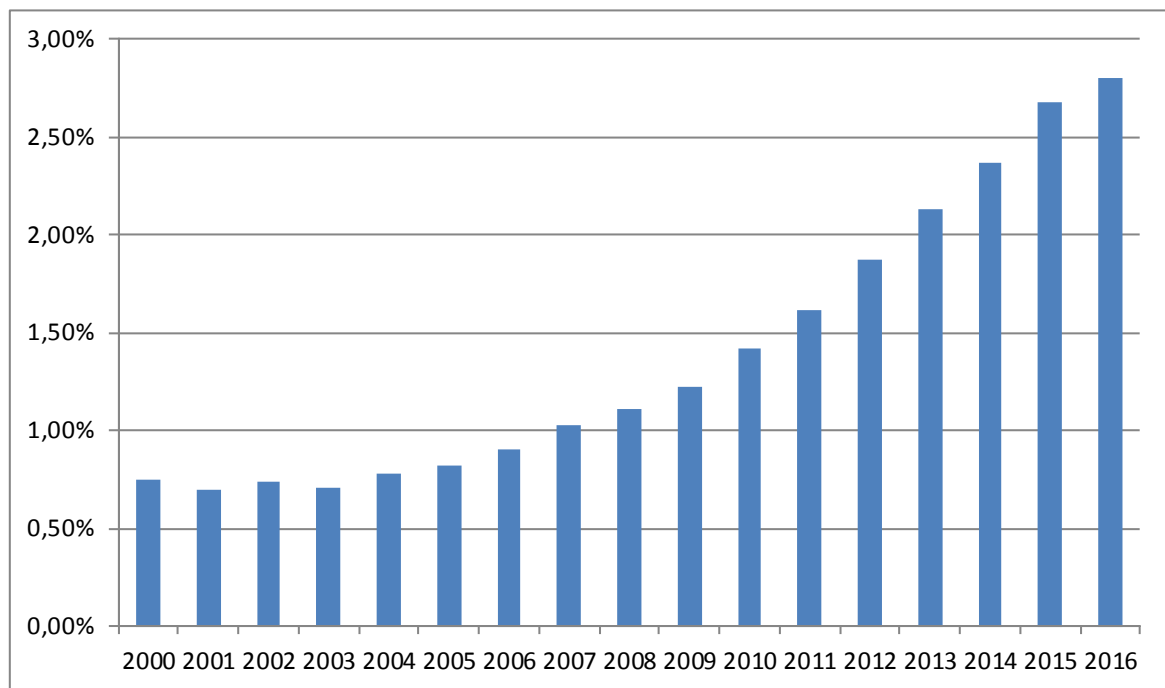
2. Exklusionsquote: Der Anteil der Schülerinnen und Schüler, die eine Förderschule besuchen, steigt von 2000 (4,60 %) bis 2003 (4,84 %). Sie stagniert nahezu bis 2006 und steigt anschließend bis 2009 (4,95 %) wieder an. Dort hat sie ihren Höhepunkt erreicht und sinkt dann kontinuierlich bis 2016 (4,34 %) (vgl. Abb. 2).

Abbildung 2: Exklusionsquote: Anteil der Schülerinnen und Schüler, die eine Förderschule besuchen in Prozent



3. Inklusionsquote: Im Jahr 2000 liegt die Inklusionsquote bei 0,75 %. Sie steigt bis 2009 um durchschnittlich 0,05 % auf 1,22 %. Mit Inkrafttreten der BRK beginnt sie deutlich schneller zu steigen – um durchschnittlich 0,23 % pro Jahr. Sie erreicht 2016 mit 2,81 % ihren höchsten Stand. Der Anteil der Schülerinnen und Schüler mit Förderbedarf, die eine allgemeine Schule besuchen, hat sich im untersuchten Zeitraum nahezu verdreifacht (vgl. Abb. 3).

Abbildung 3: Inklusionsquote: Anteil der Schülerinnen und Schüler mit Förderbedarf, die eine allgemeine Schule besuchen in Prozent



Die Inklusion ist seit 2000 in Deutschland insgesamt (bei starken Unterschieden zwischen den Bundesländern) (Klemm, 2018) stark vorangeschritten. Die Anzahl an inklusiv beschulten Schülerinnen und Schülern hat sich in den 17 untersuchten Jahren nahezu verdreifacht, obwohl die Schülerzahlen insgesamt rückläufig sind. Für einen Teil dieses Anstiegs ist sicher die steigende Förderquote verantwortlich. Diese steigt jedoch im Vergleich zur Inklusionsquote deutlich langsamer.

Die Exklusionsquote steigt bis 2009 noch leicht an und sinkt seit dem Inkrafttreten der BRK kontinuierlich. Im Gegensatz dazu hat die Inklusionsquote bereits 2004 ihren Extrempunkt erreicht und steigt seitdem an – ab 2009 jedoch deutlich schneller als zuvor. 2016 findet man damit wesentlich mehr Schülerinnen und Schüler mit Förderbedarf an allgemeinen Schulen. Die Zusammensetzung der Schülerschaft hat sich im relevanten Zeitraum durch die Inklusion verändert. Dadurch sind die Daten aus verschiedenen Zyklen der Schulleistungsstudien nur bedingt miteinander vergleichbar.

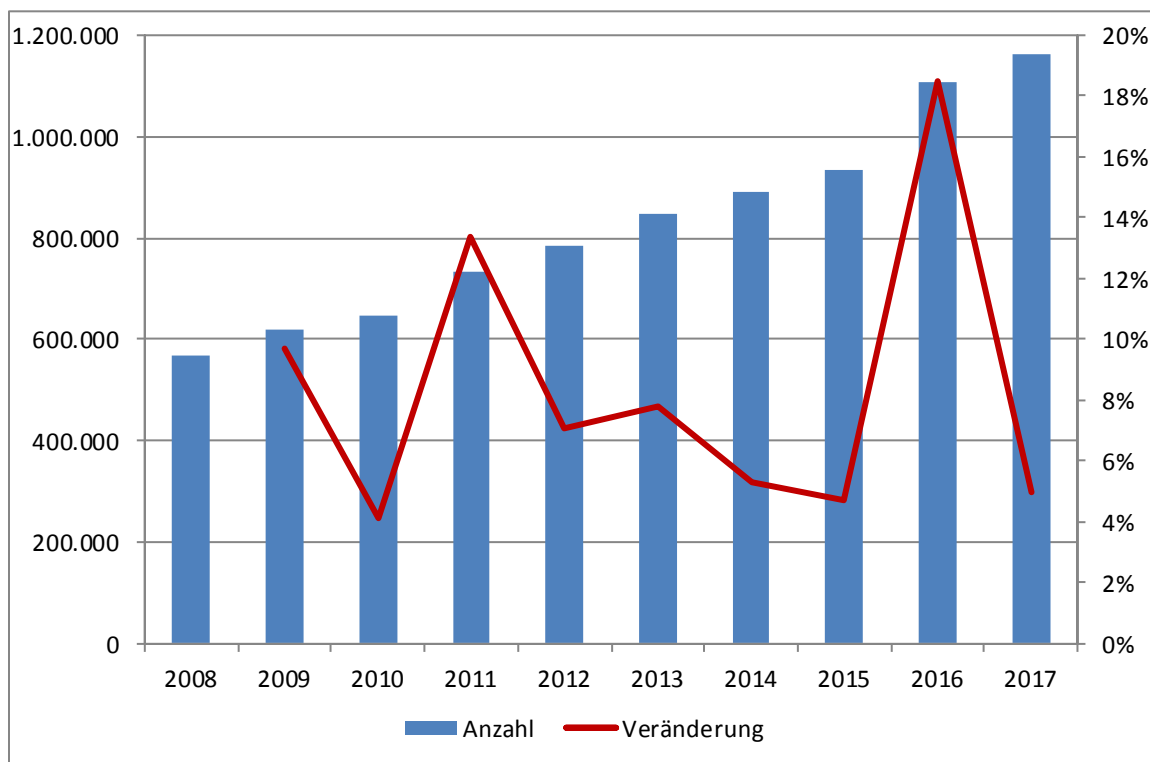
Ähnlich wie im Zusammenhang mit dem Thema Inklusion angedeutet, ist davon auszugehen, dass auch andere Entwicklungen im Schulsystem – wie z.B. der Ausbau des Ganztagsangebots, das Angebot und die Nutzung digitaler Medien in Schule und Unterricht wie auch im außerschulischen Kontext die Lernbedingungen der Schülerinnen und Schüler derart verändert haben, dass ein direkter Vergleich von Leistungsmerkmalen über solch einen langen Zeitraum nur unter Berücksichtigung dieser Entwicklungen möglich und sinnvoll ist.



### Stichwort: Ausbau der Ganztagschule

Die gesellschaftliche Bedeutung der schulischen Ganztagsangebote steigt kontinuierlich. Als Ursachen werden die Vereinbarkeit von Familie und Beruf sowie die Verbesserung der Rahmenbedingungen des schulischen Lernens gesehen (vgl. KMK, n.d.). Der Ausbau des Ganztagsschulangebots ist ein gesellschaftlich bedeutendes Thema, und die amtlichen Daten, die über die KMK zur Verfügung gestellt werden, zeigen in den letzten zehn Jahren einen durchgehenden Anstieg an Schülerinnen und Schülern, die am Ganztagsschulbetrieb teilnehmen (vgl. KMK, n.d.).

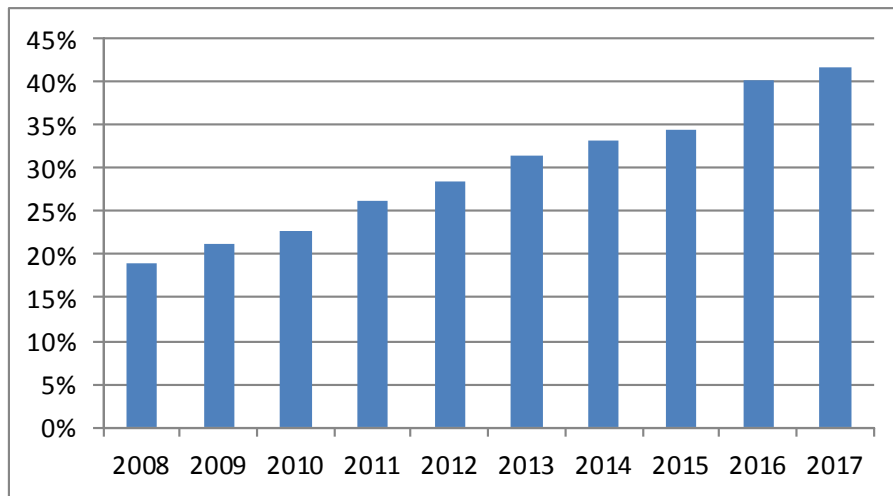
Abbildung 4: Anzahl der am Ganztagsschulbetrieb teilnehmenden Grundschülerinnen und -schüler sowie Veränderungen zum Vorjahr in Prozent



Die Anzahl der Schülerinnen und Schüler verdoppelt sich von 2008 (566.103) bis 2017 (1.162.269). Dies könnte durch das von der Bundesregierung geförderte Investitionsprogramm „Zukunft Bildung und Betreuung“ (2003–2009) bedingt sein, in dessen Zuge, durch finanzielle Unterstützung, Konzepte und Maßnahmen ein Ausbau des Ganztagsschulbetriebs stattfand (vgl. BMBF, 2009). Prozentual verlief der Anstieg sehr unterschiedlich über die Jahre. Der Graph (rote Linie) verdeutlicht einen starken Anstieg in den Jahren 2008, 2011 und besonders im Jahr 2016 (vgl. Abb. 4).

Bei der Betrachtung des prozentualen Anteils von Grundschülerinnen und Grundschülern im Ganztag wird der generelle Wandel deutlich. Innerhalb von zehn Jahren ist die prozentuale Teilnahme am Ganztag stark gestiegen (vgl. KMK, n.d.). 2008 haben 19 % der Grundschülerinnen und Grundschüler das Ganztagsangebot genutzt. 2017 liegt der Anteil bei 42 %. Damit liegt eine Steigerung von 23 Prozentpunkten innerhalb von zehn Jahren vor (vgl. Abb. 5).

Abbildung 5: Anteil der Schülerinnen und Schüler an Grundschulen, die am Ganztag teilnehmen in Prozent

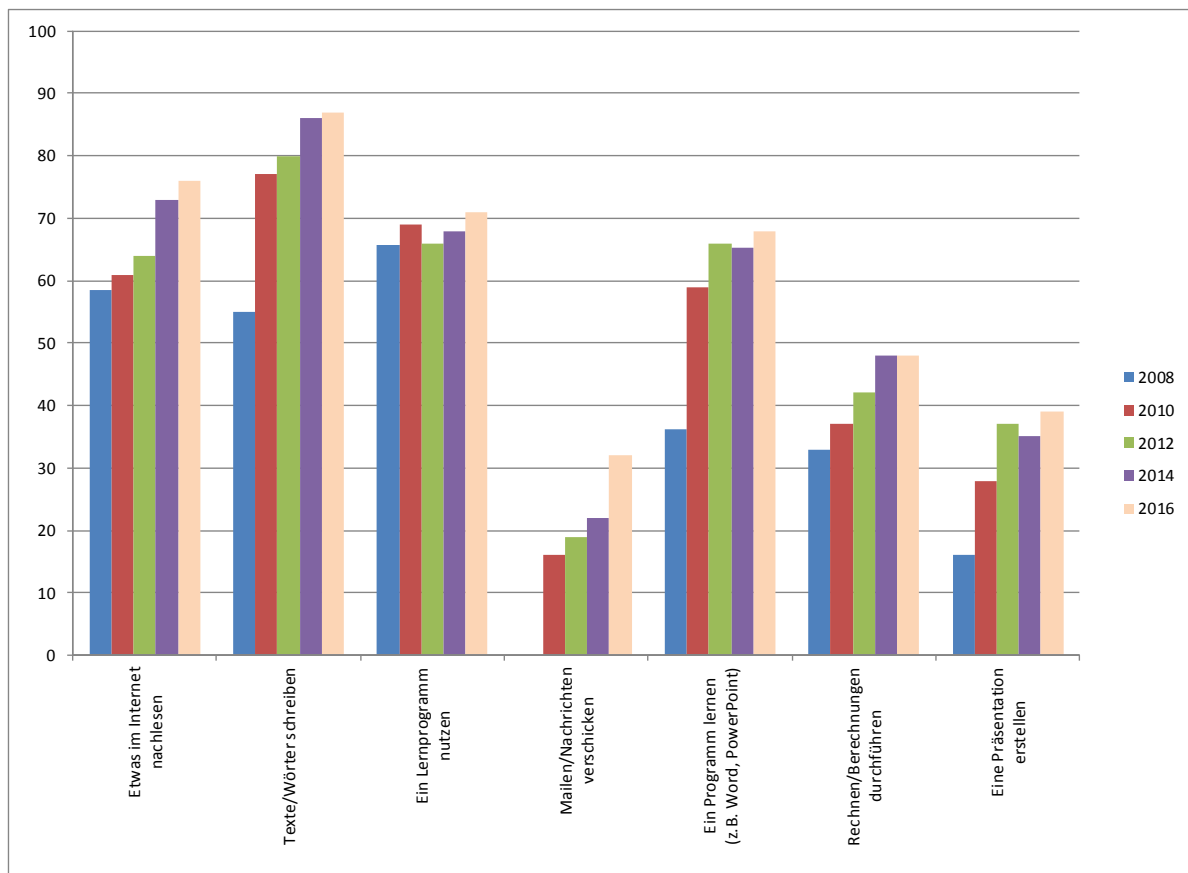


### *Stichwort: Digitalisierung in der Schule*

Die Relevanz der Nutzung von digitalen Medien für die Teilhabe am gesellschaftlichen Leben steigt kontinuierlich. Die Medienkompetenz bildet mittlerweile „eine Schlüsselqualifikation für den schulischen Alltag und für die Aneignung von Bildung generell“ (Mpfs, 2017, S. 3). Die Nutzung digitaler Medien als Mittel zur Wissensvermittlung in der Schule ist derzeit noch eher sporadisch. Mit zunehmendem Alter und dem Wechsel in die Sekundarstufe I steigt die Nutzung jedoch. Grund hierfür könnten die noch spärliche digitale Ausstattung von Schulen und das Alter der Geräte sein. Die Bedeutung von Auskünften über schulische und außerschulische Nutzung digitaler Medien steigt in den letzten Jahren zunehmend. Mit der Betrachtung der Nutzung von Computern in der Schule können aufschlussreiche Informationen von Grundschulkindern, Lehrkräften und über schulischen Rahmenbedingungen erschlossen werden (vgl. Eickelmann, Lorenz, Vennemann, Gerick & Bos, 2014). Bei der Nutzung von Computern in der Schule kann angenommen werden, dass es sich um einen pädagogisch begleiteten oder zu mindestens angeleiteten Umgang mit Medien handelt, da Lehrkräfte, Pädagoginnen und Pädagogen zu dieser Zeit vor Ort und verantwortlich sind (vgl. Mpfs, 2017). Die hier zusammengeführten Daten stammen aus den KIM-Studien (Studie „Kindheit, Internet, Medien“ des Medienpädagogischen Forschungsverbunds Südwest mpfs). In den Studien werden Schülerinnen und Schüler im Alter zwischen 6 und 13 Jahren wiederholt zur ihrer Nutzung von Medien befragt.

Um Schülerinnen und Schülern die Teilhabe am gesellschaftlichen Leben zu ermöglichen, müssen sie über Kompetenzen im Umgang mit Computern verfügen. Bei der Nutzung des Computers in der Schule kann generell festgehalten werden, dass seit 2008 in allen Kategorien der Nutzung ein Anstieg vorliegt. Innerhalb der letzten neun Jahre ist bei den Kategorien „etwas im Internet nachlesen“, „Texte/Wörter schreiben“ und „ein Lernprogramm nutzen“ ein leichter kontinuierlicher Anstieg zu verzeichnen. Bei der Kategorie „Mailen/Nachrichten verschicken“ ist von 2014 bis 2016 ein sehr großer Anstieg von knapp 50% zu verzeichnen. In den Kategorien „ein Programm lernen“, „Rechnen/Berechnungen durchführen“ und „eine Präsentation erstellen“ kann über die ersten Jahre ein Anstieg verzeichnet werden und hin zu 2016 liegt ein leichter Rückgang bei der Nutzung vor (vgl. Abb. 6).

Abbildung 6: Computernutzung in der Schule: mindestens 1 Mal pro Woche. Angaben in Prozent von Kindern, die den Computer in der Schule nutzen



Interessant ist die Betrachtung der Veränderung nach der Umsetzung des von der Bundesregierung auf den Weg gebrachten „DigitalPakt Schule“, mit dem die digitale Ausstattung der Schulen verbessert werden soll, geeignete pädagogische Konzepte entwickelt und Lehrkräfte besser ausgebildet werden sollen. Gerick, Vennemann, Lorenz und Eickelmann (2014) betonen dabei, dass die Qualität der Lernprozesse eine große Rolle spielt, dabei folgen sie der Frage, wie digitale Medien in den Fachunterricht eingebunden werden können. Die Wirkung, welche durch die Nutzung digitaler Medien beim Kompetenzerwerb von Grundschülerinnen und Grundschulern erreicht werden kann, wird durch diese Daten nur angedeutet. Mit weiteren Investitionen in die digitale Bildung kann die Schule im „digitalen Zeitalter ihren Bildungs- und Erziehungsauftrag erfüllen und Schülerinnen und Schüler auf das Leben gut vorbereiten“ (BMBF, 2017).

## 2.2 Kompetenzstände von Schülerinnen und Schülern

Über eine allgemeine Beschreibung der Schullandschaft hinaus rücken in der vorgelegten Studie insbesondere zentrale Kulturtechniken in den Fokus, die maßgeblich in der Schule vermittelt werden und die als notwendige Ressource vorhanden sein sollten, um aktiv am gesellschaftlichen Leben teilzuhaben. Zu diesen im Bildungssystem zu vermittelnden Kulturtechniken gehören neben Lese- und Schreibkompetenzen auch Kompetenzen in mathematischen und naturwissenschaftlichen Fächern. Inzwischen ist auch der kompetente Umgang mit digitalen Medien nicht mehr aus unserem Alltag wegzudenken. Wie wir mittlerweile aus zahlreichen Studien der empirischen Bildungsforschung wissen, bleibt ein Teil von Schülerinnen und Schülern regelmäßig auf einem Kompetenzniveau, welches

nicht ausreicht, um aktiv am gesellschaftlichen Leben teilzunehmen (vgl. OECD, 2016; Martin, Mullis, Foy & Hooper, 2016; Mullis, Martin, Foy & Hooper, 2016; Mullis, Martin, Foy & Hooper, 2017). Darüber hinaus zeigen sich systematische Chancenungleichheiten bei Schülerinnen und Schülern z.B. mit Migrationshintergrund oder aus bildungsferneren Familien, wenn es um ihre Kompetenzstände geht (vgl. Reiss et al., 2016; Wendt et al., 2016; Hußmann, Wendt et al., 2017). Aufgrund der Tatsache, dass Kinder wichtige Kompetenzen durch ihr Elternhaus erwerben, ist schließlich u.a. bekannt, dass die Leseleistungen mit dem sozioökonomischen Status der Familie zusammenhängen (vgl. Schulz-Heidorf, 2016). Auch diese Bildungsdisparitäten gilt es systematisch im Blick zu behalten und die zentralen Bedingungsfaktoren zu identifizieren, die zum Abbau der Unterschiede beitragen können.

#### Exkurs: Kompetenzstände am Ende der Grundschule

Im Bereich der empirischen Bildungsforschung sprechen Hartig und Klieme (2006) von Kompetenzen als *kontextspezifische kognitive Leistungsdispositionen* unter Ausschluss motivationaler und affektiver Faktoren. Die Lesekompetenz sowie die mathematische und naturwissenschaftliche Kompetenz werden auch als Schlüsselkompetenzen beschrieben (vgl. ebd.; Rychen, 2008). Die Kompetenzstände von Schülerinnen und Schülern werden u.a. im Rahmen internationaler Schulleistungsstudien (TIMSS, PIRLS, PISA) erfasst. In diesen groß angelegten Studien liegt der Testentwicklung ein Kompetenzmodell zugrunde, das getrennt für einzelne Domänen Inhaltsbereiche und kognitive Anforderungsbereiche spezifiziert.

Die Erfassung unterschiedlicher Kompetenzniveaus wird bei TIMSS und PIRLS durch die Einteilung in die Kompetenzstufen I-V (Mathematik, Naturwissenschaften, Lesen) erreicht. Einzelne Kompetenzstufen stehen hierbei für verschiedene Leistungsniveaus. Untere Kompetenzstände weisen darauf hin, dass Schülerinnen und Schüler noch keine ausreichende Basis für den weiteren selbständigen Lernprozess aufweisen. Bei Erreichen der Kompetenzstufe V in der Domäne der Naturwissenschaften weisen Schülerinnen und Schüler bspw. auch elementare Fähigkeiten im Bereich der naturwissenschaftlichen Denk- und Arbeitsweisen auf. Im Folgenden werden exemplarisch aktuelle Ergebnisse aus TIMSS 2015 und PIRLS 2016 bezüglich der Kompetenzstände von Grundschülerinnen und Grundschulern in Deutschland vorgestellt.

*Mathematische Kompetenzen* – Schülerinnen und Schüler in Deutschland erreichen einen Leistungsmittelwert von 522 Punkten im internationalen Vergleich. Hierbei erzielten fast ein Viertel der Schülerinnen und Schüler (23 %) Leistungen, die unterhalb der Kompetenzstufe III liegen, d.h. Leistungen, die kein ausreichendes Ausgangsniveau für die weiterführende Schule darstellen. 4 % sind der untersten Kompetenzstufe (I) und 5 % der obersten Kompetenzstufe V zuzuordnen (vgl. Wendt, Bos et al., 2016).

*Naturwissenschaftliche Kompetenzen* – Das Kompetenzniveau liegt hier bei 528 Punkten, womit sich Schülerinnen und Schüler in Deutschland in der oberen Hälfte der Teilnehmerstaaten befinden. 22 % der Schülerinnen und Schüler erreichen Leistungen auf den beiden unteren Kompetenzstufen I und II, was bedeutet, dass diese Schülerinnen und Schüler kein Kompetenzniveau erreicht haben, auf dem in der Sekundarstufe aufgebaut werden könnte. Der Kompetenzstufe V sind 8 % der Schülerinnen und Schüler zuzuordnen (ebd., S. 18).

Beim *Leseverständnis* erreichen 2016 rund 11 % der deutschen Grundschülerinnen und -schüler die höchste Kompetenzstufe, während rund 19 % ein Leseverständnis aufweisen, das für die sich anschließenden Klassenstufen als nicht ausreichend zu charakterisieren ist. Schülerinnen und Schüler auf diesem Niveau sind noch dabei, Lesen zu lernen, ohne jedoch – wie in den weiterführenden Schulen erwartet – in der Lage zu sein, lesend zu lernen (Bremerich-Vos, Wendt & Bos 2017).

## 2.3 Bildungsdisparitäten von Schülerinnen und Schülern

Im Nachkriegsdeutschland erfährt bei der Diskussion über Bildungsbeteiligung, Bildungsbenachteiligung bzw. Bildungsdisparität insbesondere die soziale Herkunft eine besondere Betonung: Seinerzeit war der Anteil von Schülerinnen und Schülern aus Arbeiterfamilien, die das Gymnasium besuchten bzw. ein Studium aufgenommen haben, gering. Gleichzeitig wurde angenommen, dass es einen beträchtlichen Anteil von Schülerinnen und Schülern geben müsste, der die notwendigen Voraussetzungen für einen höheren Bildungsabschluss hat (vgl. Maaz, Baumert & Cortina, 2008). Die gleichzeitig geprägte Kunstfigur des „katholischen Arbeitermädchens vom Lande“ (Peisert, 1967) weist auf multiple benachteiligte Schülergruppen hin und wird bis heute zur Veranschaulichung von Bildungsungleichheiten genutzt, auch wenn sich die Gruppenzusammensetzung inzwischen deutlich von Peiserts Modell unterscheidet (vgl. u.a.: Kasper, Wendt, Bos & Köller 2016, Hußmann, Stubbe & Kasper 2017).

Der Anstieg der relativen Anteile von Kindern in höheren Bildungsgängen zeigt sich seit den 1960er Jahren durch einen prozentualen Anstieg von Schülerinnen und Schülern in Schulen, die zu mittleren bzw. höheren Schulabschlüssen führen (vgl. z.B. Baumert, Cortina & Leschinsky, 2008). Ein Trend, der sich bis heute fortgesetzt hat und in verschiedenen Bundesländern zu Gymnasialquoten<sup>5</sup> von rund 50 % geführt hat (vgl. KMK, 2019). Trotz dieses strukturell deutlichen Wandels unterscheidet sich der schulische Erfolg nach wie vor, wenn Herkunftsmerkmale in den Blick genommen werden. Dabei ist es unerheblich, ob die Kompetenzen von Schülerinnen und Schülern getestet oder kumulierte Information wie Übergangs- oder Abschlussquoten in den Blick genommen werden.

Bei der Betrachtung von Bildungsdisparitäten werden unterschiedliche Wirkmechanismen der Herkunft in den Blick genommen. So unterscheidet Boudon (1974) zwischen primären und sekundären Benachteiligungseffekten: Primäre Benachteiligungseffekte werden hierbei auf Hintergrundmerkmale und vorliegenden Voraussetzungen zurückgeführt, während sekundäre Effekte auf schichtspezifische Entscheidungen im Bildungsverlauf (z.B. bei Übergängen) zurückgeführt werden (vgl. Maaz, Baumert & Trautwein, 2009). Eine etwas andere Charakterisierung legt Roemer (1998) vor, der „circumstances“ und „efforts“ unterscheidet und somit einerseits auf nicht oder nur schwer variierbare Hintergrundmerkmale verweist und andererseits auf volitional steuerbare Anstrengungen aufmerksam macht. Gerade letztere zeigen in ihren Auswirkungen, was bei Boudon als (mögliche) sekundäre Benachteiligungseffekte beschrieben wird.

In einem Bildungssystem wie Deutschland gilt die Idee des meritokratischen Systems, in dem Anstrengung und die daraus resultierende Leistung anerkannt wird, positiv konnotiert, während herkunftsbedingte Privilegien eher kritisch bewertet werden. Bei Analysen zum Kompetenzerwerb zeigt sich jedoch, dass beobachtete Kompetenzunterschiede zwischen Schülerinnen und Schülern einerseits durch herkunftsbedingte Merkmale, wie Bildungsniveau der Eltern, Verfügbarkeit von bildungsrelevanten Ressourcen oder auch der Migrationshintergrund erklärt werden können, andererseits aber auch in einem Zusammenhang mit Persönlichkeitsmerkmalen wie Interesse oder Motivation stehen. In der Tradition des deutschen Bildungssystems ist es ein Leitgedanke, herkunftsbedingte Bildungsdisparitäten nach Möglichkeit in der Schule auszugleichen (oder zumindest abzuschwächen), sodass Anstrengungen der Schülerinnen und Schüler im Lernprozess zum Abbau der Bildungsdispari-

---

<sup>5</sup> Die Gymnasialquote beschreibt den Anteil von Schülerinnen und Schülern, die die gymnasiale Oberstufe besuchen. Mit dem mittleren Abschluss sind Schülerinnen und Schüler befähigt, die gymnasiale Oberstufe zu besuchen.

täten beitragen. In rezenten Untersuchungen wie PIRLS, TIMSS und PISA wurde in den letzten Jahren wiederholt dokumentiert, dass in Deutschland herkunftsbedingte Merkmale besonders ausgeprägt im Zusammenhang mit den erfassten Kompetenzständen der Schülerinnen und Schüler stehen (vgl. Reiss et al., 2016; Wendt, Bos et al., 2016; Hußmann, Wendt et al., 2017).

Soziale Bildungsdisparitäten im Bildungsbereich wurden bisher anhand verschiedener Indikatoren beschrieben (Döbert & Seeber, 2009). Den Zusammenhang dieser Indikatoren mit etwa der mathematischen und naturwissenschaftlichen Kompetenz konnten etwa Stubbe, Schwippert & Wendt (2016) nachweisen. Als Standardvariablen zur Beschreibung bzw. Kontrolle von sozialen Hintergrundmerkmalen werden in aktuellen Untersuchungen insbesondere Informationen über den heimischen Buchbesitz, dem Besitz bildungsrelevanter Güter im Haushalt, dem Migrationshintergrund, dem Bildungsstand der Eltern bzw. des Status des von ihnen ausgeübten Berufs – oder aber auch Kompositionen aus verschiedenen vorliegenden Merkmalen – genutzt. In Ergänzung zu diesen Indikatoren können auch Merkmale wie etwa das Leseverhalten der Eltern in den Blick genommen werden (vgl. Goy, Valtin & Hußmann, 2017).

Was meint Bildungsdisparität heute?

Wenn über Bildungsdisparitäten gesprochen wird, ist heute insbesondere die Bedeutung der sozialen Herkunft von Interesse. Seinerzeit war der Anteil von Schülerinnen und Schülern aus Arbeiterfamilien, die das Gymnasium besuchten oder ein Studium aufnahmen gering. Seit den 1960er Jahren ist ein prozentualer Anstieg von Schülerinnen und Schülern in Schulen, die zu höheren Schulabschlüssen führen zu verzeichnen. Trotz des strukturellen Wandels zeigt sich bei der Betrachtung von Herkunftsmerkmalen, dass sich der schulische Erfolg noch immer unterscheidet.

Bei der Betrachtung von Bildungsdisparitäten werden heute verschiedene Wirkungsmechanismen sozialer Herkunft unterschieden, Roemer (1998) unterscheidet zwischen „circumstances“ und „efforts“. Erstere verweisen auf nicht oder nur schwer veränderbare Hintergrundmerkmale und Letztere auf willentlich steuerbare Anstrengungen.

## 3 Perspektiven und Ableitungen

### 3.1 Perspektive: Vergleichbarkeit von Querschnittanalysen ermöglichen

Wie oben gezeigt, werden zur empirischen Aufzeichnung von Bildungsdisparitäten Merkmale des sozialen Hintergrunds, des Migrationsstatus und (verstärkt in den letzten Jahren) der Inklusion verwendet (vgl. Hußmann, Stubbe et al., 2017; Kasper et al., 2016; Schulz-Heidorf, 2016; Stubbe et al., 2016; Wendt & Schwippert 2017; Wendt, Schwippert & Stubbe, 2016). Die konkret verwendeten Merkmale unterscheiden sich dabei hinsichtlich ihrer Komplexität und auch in Bezug auf ihre Messeigenschaften. Darüber hinaus variieren die verwendeten Merkmale zum Teil sowohl zwischen unterschiedlichen empirischen Bildungsstudien als auch (bei Studien im Trend) über die Zeit. Damit unterscheiden sich aber auch die Aussagen hinsichtlich der Stärke und der Art der vorhandenen Bildungsdisparitäten teilweise erheblich zwischen den Studien, und ein Vergleich dieser Ergebnisse ist bisher nur begrenzt möglich.

Eine methodische Lösung dieser Herausforderung stellt die Möglichkeit des *Equating* dar. Ein solches *Equating* wurde von uns bereits erfolgreich im Rahmen von TIMSS 2015 eingesetzt (vgl. Wendt, Kasper, Bos, Vennemann & Goy, 2017). Es trägt dazu bei, die Ergebnisse von inhaltsgleichen aber verschieden erfassten Merkmalen direkt mit einander vergleichen zu können.

Exkurs: Wie wird die Vergleichbarkeit gewährleistet?

Merkmale werden in verschiedenen Studien unterschiedlich gemessen, d.h. beispielsweise mit verschiedenen Indizes dargestellt. Diese Indizes werden herangezogen, um die Leistung der Lernenden zu erklären. Um die Indizes und ihre Auswirkungen zwischen den Studien vergleichbar zu machen, wird das Verfahren des Equating herangezogen.

Als Beispiel könnten hier die Ergebnisse zur Untersuchung der sozialen Bildungsdisparität herangezogen werden. Sowohl in TIMSS als auch in PISA werden Fragen zum sozialen Hintergrund der Schülerinnen und Schüler (z.B. welchen Beruf üben deine Eltern aus) gestellt und anschließend im Sinne von Indikatoren zur Berechnung eines Gesamtscores der sozialen Lage, in TIMSS als *Home Resources for Learning Scale* (HRL; vgl. Martin & Mullis, 2012) und in PISA als *Index of Economic, Social and Cultural Status* (ESCS; vgl. OECD, 2017) bezeichnet, verwendet. Allerdings basieren der HRL Index und der ESCS Index zum Teil auf unterschiedlichen Indikatoren und zum Teil auf unterschiedlichen Berechnungsroutinen. Damit sind die Werte auf diesen beiden Indizes nicht direkt miteinander vergleichbar und somit ist auch ein direkter Vergleich beispielsweise der Zusammenhänge dieser beiden Indizes mit der Schülerleistung nicht möglich. Durch das Verfahren des Equating können die beiden Skalen auf eine Metrik gebracht werden und diese metrisch äquivalenten Skalen dann weiterverwendet werden, um etwa soziale Bildungsdisparitäten gemäß TIMSS mit sozialen Bildungsdisparitäten gemäß PISA zu vergleichen. Wie am Beispiel von TIMSS 2015 gezeigt (Kasper et al., 2016) können somit gesellschaftlich bzw. gesetzlich induzierte leistungsrelevante Veränderungen in der Schülerschaft aufgezeigt werden.

### 3.2 Perspektive: Soziale Bildungsdisparitäten

Bildungsdisparitäten wurden bisher vor allem als Zusammenhang der vorgenannten Indikatoren mit verschiedenen Maßen für Lern- und Abschlusserfolge sowie Kompetenzen aufgefasst. Die zuvor genannten Merkmale können zum Teil auch als Merkmale der betrachteten Gruppen (wie Schulklassen oder Schulen) zusammengefasst werden, um so nicht nur die Individuen selber, sondern auch deren unterschiedliche Lernumfelder zu berücksichtigen. Solche zusammengefassten Merkmale können als sogenannte Kompositionseffekte von Gruppen genutzt werden. Eine starke Konzentration bestimmter Kompositionsmerkmale in unterschiedlichen Klassen kann auf mögliche Zuweisungs- bzw. Aufnahmequotierungen aufmerksam machen, die dann als Selektionseffekte beschrieben werden können. Ähnliche Kompositionsmerkmale werden inzwischen auch im Rahmen der Steuerung von Bildungssystemen eingesetzt. Ein Beispiel hierfür ist der in Hamburg entwickelte Sozialindex (vgl. Bos, Gröhlich & Bosen, 2009).

Ein Ziel dieser Machbarkeitsstudie ist es, empirische Modelle zu entwickeln und zu testen, die es erlauben – unter Kontrolle herkunftsbedingter Merkmale – Merkmale zu identifizieren, die entsprechend ihres Ursprungs auf unterschiedlichen Ebenen kompensatorisch zwischen den Schülerinnen und Schülern wirken. Hierbei werden zum einen Persönlichkeitsmerkmale der Lernenden aber auch Merkmale der Lernumgebung (Ausstattung, pädagogisches oder didaktisches Handeln) in den Blick genommen.

### 3.3 Zwischenfazit und Ableitung einer allgemeinen Fragestellung

Um die Frage nach dem Ausmaß und der Entwicklung von Bildungsdisparitäten beantworten zu können, sind zum einen Indikatoren für Bildungsdisparitäten mit Leistungserfolgen in Zusammenhang zu bringen, um u.a. Profile multipler Benachteiligungen zu analysieren. Zum anderen ist eine besondere

Aufmerksamkeit auf jene veränderbaren Bedingungsmerkmale auf Leistung zu richten, die erklären können, warum auch nach Kontrolle relevanter Hintergrundmerkmale Schülerinnen und Schüler hinter den statistisch von ihnen erwarteten Leistungen zurückbleiben, während andere Schülerinnen und Schüler über ihrem erwarteten Leistungspotenzial liegen. Um dabei der Heterogenität der Indikatoren zwischen den verwendeten Studien ein Stück weit Rechnung zu tragen, sind dabei Verfahren des Equating anzuwenden.

Zusammenfassend lassen sich im Wesentlichen zwei – aufeinander aufbauende – Forschungsdesiderate beschreiben:

- Wie lassen sich Bildungsdisparitäten insbesondere in der Kontrastierung von Schülerinnen und Schülern mit unterschiedlichen Ausgangslagen und Schulerfolgen beschreiben? Welche kurzzeitig durch ideelle bzw. materielle Förderung veränderbaren Bedingungsmerkmale für den Leistungserfolg erweisen sich als trennscharf in Hinblick auf die Unterscheidung verschiedener Schülergruppen?
- Welche gesellschaftlich bzw. gesetzlich induzierten leistungsrelevanten Veränderungen in der Schülerschaft können Trendentwicklungen erklären? Lassen sich Aussagen bezüglich Bildungsdisparitäten, unterschiedlich gemessen in verschiedenen Studien und über die Zeit, durch Verfahren des Equating direkter aufeinander beziehen?

In der vorliegenden Untersuchung wird unter Berücksichtigung sozialer Hintergrundmerkmale untersucht, wie kurzzeitig durch ideelle bzw. materielle Förderung veränderbare Bedingungsmerkmale mit dem Schulerfolg im Zusammenhang stehen.

## 4 Empirische Modellierung

### 4.1 Methodisches Vorgehen

In der hier vorgestellten Machbarkeitsstudie wird der Entwicklung von Chancen bzw. Benachteiligungen (die hier nicht aktiv im Sinne von Diskriminierung, sondern passiv im Sinne von fehlenden Möglichkeiten verstanden werden) im deutschen Bildungssystem nachgegangen. Im Rahmen der Untersuchung werden somit die relativen Positionen bzw. relativen Möglichkeiten von Individuen im deutschen Bildungssystem in den Blick genommen. Zu diesem Zweck fokussieren wir auf den Lernerfolg von Schülerinnen und Schülern, der anhand von Leistungstests in einer Reihe von für Sekundäranalysen zur Verfügung stehenden Daten ermittelt wurde. Hierbei stehen in der Regel zentrale Merkmale wie Leseverständnis, mathematische und naturwissenschaftliche Kompetenz im Fokus.

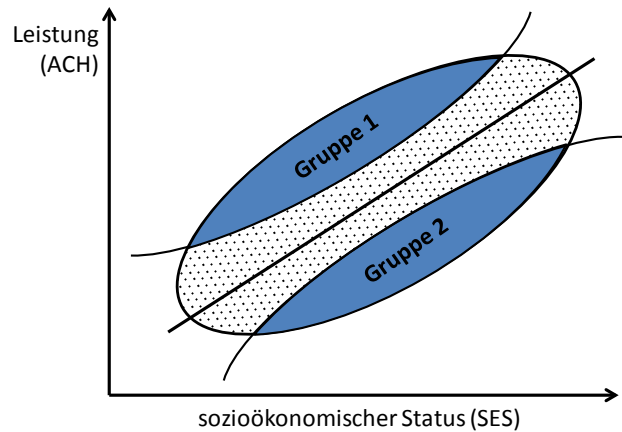
Anhand von Regressionsmodellen kann der generelle Effekt des sozialen Hintergrunds auf den Kompetenzstand untersucht werden. Darüber hinaus ermöglicht diese Analyse, den sozialen Hintergrund zu „kontrollieren“ und unabhängig von dessen Effekt weitere erklärende Merkmale in das Modell aufzunehmen. Auf diese Weise können Aussagen über die Wirkung von Merkmalen unabhängig vom sozialen Hintergrund gemacht werden. Abbildung 7 ist die schematische Darstellung eines solchen Modells zu entnehmen. Die durch die Ellipse gehende Gerade kennzeichnet den ermittelten Effekt des sozialen Hintergrunds (SES) auf die Leistung (ACH) der Schülerinnen und Schüler, während die farblich hinterlegten Bereiche die Abweichungen von der Vorhersage der Leistung durch SES kennzeichnen, die durch weitere Variablen beschrieben werden können.

Etwas weniger technisch formuliert: Je privilegierter der Sozialstatus der Familie, desto besser fallen im Regelfall auch die Leistungen von Schülerinnen und Schülern aus. Gleichwohl zeigen sich inner-



halb der Schülergruppe Teilgruppen, für die dieser Zusammenhang nicht gilt: Diese Gruppen zeigen erwartungswidrig hohe (Gruppe 1) oder erwartungswidrig niedrige (Gruppe 2) Leistungen.

Abbildung 7: Schematische Darstellung einer Regression mit Bereichen unerklärter Abweichungen



In dieser Studie liegt der Fokus auf der Analyse der Gruppe 1 und Gruppe 2. Dabei sollen zwei Fragen genauer beantwortete werden:

- Wie lassen sich als erwartungswidrig identifizierte Gruppen charakterisieren?
- Welche Merkmale, die sich in Bezug auf Lernprozesse kurz bzw. mittelfristig verändern lassen, helfen diese beiden Gruppen zu unterscheiden?

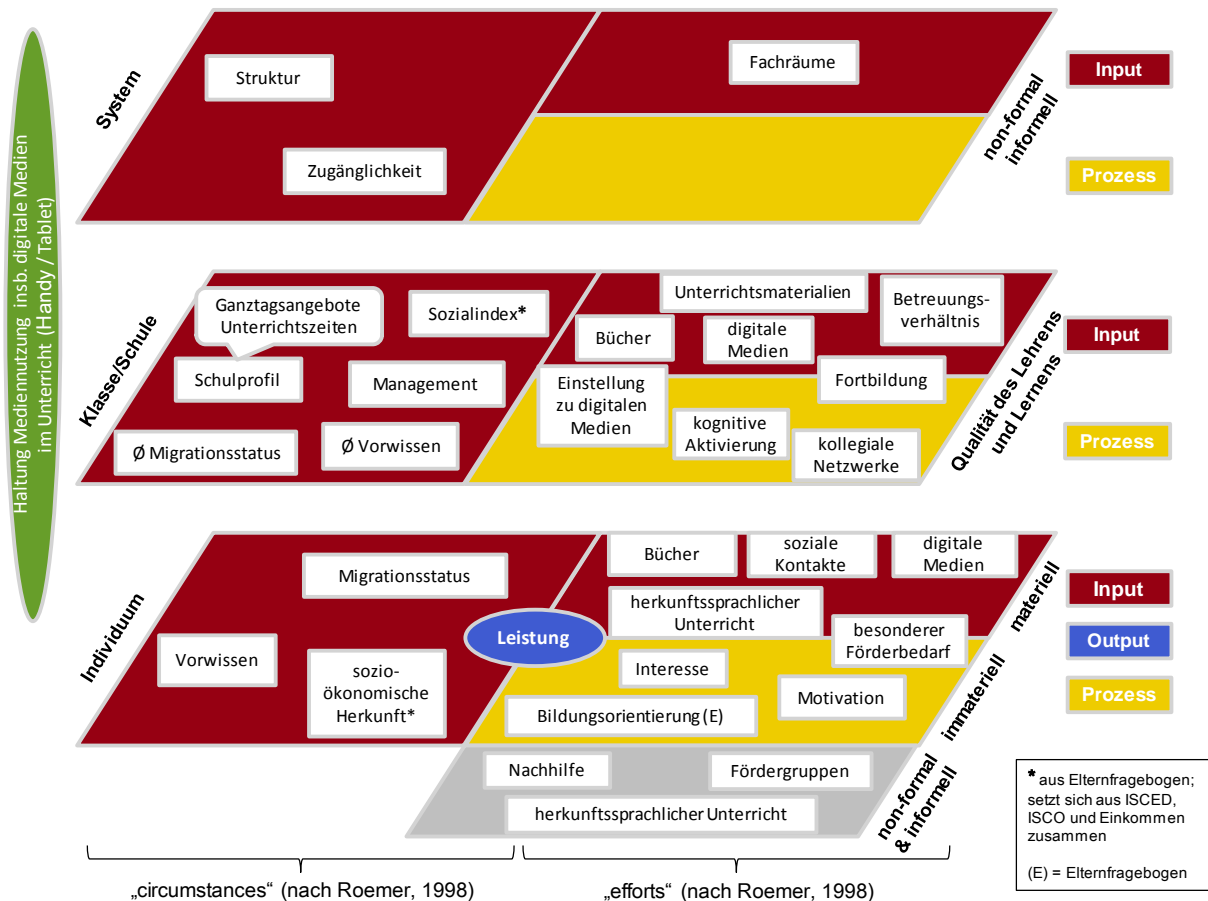
Wie der Abbildung 7 zu entnehmen ist, werden zwei Gruppen von Schülerinnen und Schülern miteinander kontrastiert. Diese Kontrastierung wurde gewählt, um möglichst Trennscharfe Indikatoren für erwartungswidrige Leistungen zu identifizieren. Wenn klar ist, in welchen kurz- bzw. mittelfristig veränderbaren Lernbedingungen sich die stärksten Schülerinnen und Schüler (Gruppe 1) von den schwächsten Schülerinnen und Schülern (Gruppe 2) unterscheiden, dann können diese Merkmale genutzt werden, um die Schülerinnen und Schüler der Gruppe 2 zu fördern, oder anders formuliert: Das hier vorgeschlagene Modell verfolgt den Ansatz des Lernens von den Stärksten. Würde man anstelle von zwei Gruppen mehrere Gruppen kontrastieren, dann würde man zwar mehr Leistungsvarianz erklären können, aber die Orientierung an den Stärksten würde verloren gehen.

## 4.2 Theoretisches Rahmenkonzept

Untersuchungen des Unterrichts in Bezug auf eine gelingende Förderung der Kompetenzen von Schülerinnen und Schülern sind in der Regel anders konzipiert als Studien, die Informationen über das Bildungssystem bzw. über Strukturen im Bildungssystem Auskunft geben sollen. Dennoch werden oft auch Fragen z.B. der Schul- und Unterrichtsforschung in Systemmonitoringstudien aufgenommen, ohne dass deren Studiendesign hierfür optimiert ist. Um mit den jeweils speziell ausgerichteten Schul- und Unterrichtsstudien dennoch anschlussfähig an andere Studien zu sein, werden beim Untersuchungsdesign und der Konzeption der Befragungs- und Testinstrumente in der Regel sogenannte Ankerpunkte hergestellt. Da es in der vorliegenden Untersuchung darum geht, ein breites Repertoire an Schul- und Unterrichtsstudien, Untersuchungen zum Bildungsmonitoring und von amtlichen Statistiken (mit einer ganz eigenen Intention zur Systemsteuerung) zu verbinden, haben wir verschiedene Modelle zur empirischen Erfassung von Informationen aus dem Bildungssektor zusam-

mengefasst (vgl. Abb. 8). Ziel dieser Zusammenstellung ist, für die Sichtung der vorliegenden Daten eine Systematik zu entwickeln, die zum einen hilft, relevante Variablen (ggf. vergleichend zwischen den Studien) zu identifizieren und die zum anderen in der Literatur als erklärungskräftig identifizierte Modelle bzw. Forschungstraditionen zusammenführt.

Abbildung 8: Kombiniertes Mehrebenen Input-Prozess-Output-Modell<sup>6</sup>



Dies ist noch weiter auszuführen: Das theoretische Rahmenmodell zu dieser Studie basiert zum einen auf dem *Dynamischen Modell zur Beschreibung von Schuleffektivität* (vgl. Creemers & Kyriakides, 2008) und zum anderen auf dem *Input-Prozess-Output-Modell zur Beschreibung der Qualität des Lehrens und Lernens* (Ditton, 2002; Helmke, 2004). Zur Systematisierung der relevanten Variablen werden in dem Modell außerdem *immaterielle und materielle Faktoren* sowie *Circumstances und Efforts* (vgl. Roemer, 1998) unterschieden.

Generell wird in dem Modell widerspiegelt, dass alle Input- bzw. Prozessmerkmale als relevant für das Outputmerkmal Leistung (im Sinne eines Lernerfolgs in den Schulfächern) angesehen werden. So ist etwa davon auszugehen, dass das individuelle Circumstances-Inputmerkmal *Migrationsstatus* die

<sup>6</sup> Um veränderte Förderbemühungen, aber auch sich wandelnde Anforderungen im Bildungssystem nachzuzeichnen, müssen die Analysen im zeitlichen Verlauf betrachtet werden, d.h. es ist zu überprüfen, inwiefern etwa eine Veränderung in der Bereitstellung von digitalen Medien zum Zeitpunkt  $t_1$  die Zugehörigkeit zu der Gruppe der Schülerinnen und Schüler mit erwartungswidrig hohen Leistungen zum Zeitpunkt  $t_2$  vorhersagt. In Bezug auf das kombinierte Mehrebenen Input-Prozess-Output-Modell würde das bedeuten, dass eine zeitliche Achse in dieses Modell einzufügen wäre. Aus Gründen der Übersichtlichkeit haben wir auf die Darstellung dieser Achse in Abbildung 8 jedoch verzichtet.

Schülerleistung beeinflusst. Allerdings lässt sich der Migrationsstatus von Schülerinnen und Schülern nicht ändern. Dieses Merkmal (und alle anderen Circumstances-Merkmale in dem Modell) kann in statistischen Analysen zwar als Kontrollvariable genutzt werden, um zum einen Bildungsdisparitäten aufzuzeigen und zum anderen die Gruppen der Schülerinnen und Schüler mit erwartungswidrig niedrigen bzw. erwartungswidrig hohen Leistungen zu identifizieren, es lässt sich aber nicht kurzfristig bzw. mittelfristig verändern. Die Efforts-Merkmale wiederum lassen sich kurzfristig bzw. mittelfristig verändern und stehen damit im Zentrum dieser Studie.

Im Sinne des unter Abschnitt 4.1 aufgeführten methodischen Vorgehens und basierend auf dem kombinierten Mehrebenen Input-Prozess-Output-Modell geht es in der Studie zunächst darum, anhand von Circumstances-Merkmalen die Gruppen der Schülerinnen und Schüler mit erwartungswidrig niedrigen bzw. erwartungswidrig hohen Leistungen zu identifizieren. In einem weiteren Schritt werden die Efforts-Merkmale genutzt, um die Unterschiede in der Leistung dieser beiden Gruppen zu erklären. Flankierend zu diesen Analysen werden die beiden Gruppen anhand deskriptiver Daten (z.B. Alter oder Geschlecht) beschrieben.

Um der Frage nach den Ursachen von Bildungsdisparitäten angemessen empirisch nachgehen zu können, wird die Untersuchung insbesondere auf vorliegende Daten aus Schulvergleichsuntersuchungen (z.B. TIMSS) gestützt werden, die im Rahmen von Re-Analysen erschlossen und ggf. perspektivisch auch um entsprechende Fragen erweitert werden sollen. Auf dieser Basis kann empirisches Handlungswissen über Möglichkeiten des Abbaus von Bildungsdisparitäten durch die Beeinflussung kurzzeitig veränderbarer Bedingungsfaktoren (siehe Kasten 1) gewonnen werden. Diese Perspektive in den Blick zu nehmen, bedeutet Schülerinnen und Schüler zu identifizieren, die hinter den für sie erwarteten Bildungspotenzialen bleiben – unabhängig davon, an welcher Stelle des Leistungskontinuums diese Reserven verortet sind.

Kasten 1: Auswahl von Prozessmerkmalen kurzzeitig veränderbarer Lernbedingungen

Folgende exemplarische Merkmale stellen kurzzeitig veränderbare Lernbedingungen dar, welche möglicherweise zwischen Schülerinnen und Schülern differenzieren, die ihre Bildungspotenziale ausgenutzt haben und solchen, für die dies (noch) nicht der Fall ist.

- Auf Schulebene: Bereitstellung von Ressourcen (z.B. differenzierte Lernmaterialien oder Bibliotheken), Ausstattung mit unterschiedlichen Medien, Bereitstellung von Räumlichkeiten für non-formale Bildungsangebote, Weiterentwicklung des individuellen Beratungsangebots für Lehrerinnen und Lehrer, Schülerinnen und Schüler sowie Eltern.
- Auf Unterrichtsebene: Neu-Fokussierung von Feedbackstrukturen, Anpassung von Lehr- und Lernmaterialien beispielsweise hinsichtlich des Anspruchsniveaus, Anpassung des Umgangs mit Lehr-/Lernmaterialien.
- Auf Individualebene: Bereitstellung von persönlichen Lernmaterialien (z.B. Bücher, neue Medien), Initiierung von individuellen Förderangeboten im non-formalen Bildungskontext (z.B. Nachhilfe, Interessengruppen, Projektgruppen), Motivationstrainings.

Um Aussagen über sich wandelnde Bildungschancen vornehmen zu können, sind die sich verändernden Rahmenbedingungen in den Blick zu nehmen, die zum Beispiel durch den demographischen Wandel, aber auch durch gesetzliche Veränderungen bedingt sein können (s. Abschnitt 2.1). So konnten etwa Kasper et al. (2016) in TIMSS für Deutschland nachweisen, dass sich die Leistung der Viert-

klässlerinnen und Viertklässler in Naturwissenschaften von TIMSS 2007 nach TIMSS 2015 vermutlich um 20 Punkte verbessert hätte, wenn die u.a. durch Bildungsreformen in dieser Zeit bedingte zunehmende Heterogenisierung der Schülerschaft nicht stattgefunden hätte. Anders formuliert: Auch wenn individuell betrachtete Leistungsindikatoren über die Zeit keine Veränderung dokumentieren, kann sich dieses unter Berücksichtigung sich verändernder Umstände dennoch als Erfolg zum Abbau von Bildungsdisparitäten herausstellen.

### 4.3 Datenmaterial

In der Machbarkeitsstudie wurden zunächst die Daten von drei Studien gesichtet: 1.) TIMSS erfasst mathematische und naturwissenschaftliche Grundverständnisse von Grundschülerinnen und Grundschülern, 2.) IGLU, englischsprachig unter PIRLS geführt, erfasst das Leseverständnis und 3.) das *Nationale Bildungspanel* (NEPS) erfasst die gesamte Lebensspanne, dabei werden in unterschiedlichen Erhebungen verschiedene thematische Schwerpunkte gesetzt. TIMSS und IGLU sind internationale Schulleistungsstudien, die 4. Jahrgangsstufen untersuchen. Auf internationaler Ebene ist die *International Association for the Evaluation of Educational Achievement* (IEA) für die Organisation und für das zur Verfügung stellen der Daten verantwortlich. Die nationalen Daten der beiden Studien sind über das Forschungsdatenzentrum des IQB Berlin zu beantragen. NEPS ist eine Längsschnittstudie, die Bildungsverläufe von Personen erfasst, um Entwicklungen in den Bildungsverläufen der Menschen und auch bei den Übergängen, wie zum Beispiel von der Grundschule in eine weiterführende Schule, nachzuzeichnen. NEPS wird unter der Leitung des Leibniz-Institut für Bildungsverläufe (LifBi) durchgeführt. Die Daten der Studie sind über das LifBi Forschungsdatenzentrum zu beantragen.

Alle drei herangezogenen Studien erheben Daten von Viertklässlerinnen und Viertklässlern, deren Eltern, Lehrkräften und Schulleiterinnen und Schulleitern. In allen drei Studien werden die Kompetenzen über Leistungstests erfasst. Für ein ganzheitliches Bild werden parallel Merkmale der Schülerinnen und Schüler, ihrer Familien und der schulischen Umgebung erfasst. Mit dem Datenmaterial liegen Informationen zu verschiedenen Kompetenzbereichen und Informationen von verschiedenen Personengruppen zu den Viertklässlerinnen und Viertklässlern vor. 2011 ist der Erhebungszeitpunkt von TIMSS und IGLU auf das gleiche Jahr gefallen, damals entschied Deutschland sich, die beiden Studien mit derselben Stichprobe zu erheben. Mit der Nutzung von TIMSS und IGLU 2011, TIMSS 2015, IGLU 2016 und der zweiten Erhebungswelle von NEPS aus dem Jahr 2015/2016 kann für Schülerinnen und Schüler ein Trend über sechs Jahren am Ende der Grundschulzeit betrachtet werden.

Mit dem Datenmaterial aus diesen drei Studien können über das Equating-Verfahren die Daten verglichen werden, um am Ende gemeinsam Aussagen über die drei Studien zu formulieren. Dafür wurden aus den Studien Variablen zu verschiedenen Merkmalen auf Individual-, Klassen-/ Schul- und Systemebene aus den Skalenhandbüchern herausgesucht. Die Merkmale wurden anhand des kombinierten Mehrebenen Input-Prozess-Output-Modells (vgl. Abb. 8) ausgewählt. Im Anhang A ist eine Übersicht nutzbarer Items für das Equating der beschriebenen Studien zusammengestellt.

### 4.4 Equating-Verfahren

In der empirischen Bildungsforschung werden häufig alternative Testformen desselben Tests angewendet. So werden in TIMSS 2015 etwa im Rahmen des multiplen Matrix Designs 14 Testhefte eingesetzt, die alle dieselben Kompetenzen (Mathematik und Naturwissenschaften) messen sollen. Durch den Einsatz unterschiedlicher Testformen desselben Tests kann etwa verhindert werden, dass die Schülerinnen und Schüler voneinander abschreiben. Der Einsatz unterschiedlicher Testformen des-

selben Tests wirft allerdings die Frage auf, inwiefern die Testwerte der alternativen Testformen miteinander vergleichbar sind. Um die Testwerte dieser alternativen Testformen miteinander vergleichen zu können, müssen diese auf derselben Skala abgebildet werden. Die Prozedur, mit der die Werte auf dieselbe Skala gebracht werden wird als Equating oder auch Linking bezeichnet (vgl. Kolen & Brennan, 2004).

Vereinfacht kann man sich das Equating durch folgendes Beispiel vorstellen. Angenommen man misst bei zwei Personen A und B die Größe. Über Person A erfährt man, dass sie 2 Meter groß ist, über Person B, dass sie 180 Zentimeter groß ist. Rein nach dem Zahlenwert ist Person B also größer. Erst nach *Umrechnen* in dieselbe Metrik (z.B. cm) lassen sich die Zahlen direkt miteinander vergleichen und es wird klar, dass Person A größer ist.

Ein Equating ist in der empirischen Forschung also immer dann angezeigt, wenn Skalen, Index, Summenwerte oder dergleichen verschiedener (inhaltsgleicher) Messinstrumente miteinander verglichen werden sollen. Im Rahmen der Machbarkeitsstudie sind dies etwa die Skalen zum sozialen Hintergrund. Zur technischen Erläuterung der Equating- bzw. Linking-Verfahren sei auf Anhang B verwiesen.

## 4.5 Exemplarische Analysen

Die exemplarischen Analysen sollen dazu dienen, das methodische Vorgehen zur empirischen Analyse sozialer Ungleichheiten zu erläutern. Dazu wird in einem ersten Schritt der soziale Status der Schülerinnen und Schüler genutzt, um deren Leistung vorherzusagen und die Gruppen der Schülerinnen und Schüler mit erwartungswidrig niedrigen und hohen Leistungen zu identifizieren (im Folgenden auch *Gruppe der Ausreißer* im Sinne der linearen Regression genannt). Anschließend werden diese beiden Gruppen fokussiert, um herauszufinden, in welchen Merkmalen des kombinierten Mehrebenen Input-Prozess-Output-Modells sich diese beiden Gruppen unterscheiden.

### 4.5.1 Daten und Variablen

Für die exemplarischen Analysen werden die nationalen Datensätze für Deutschland der Schülerinnen und Schüler, der Eltern und der Lehrkräfte von TIMSS 2015 und PIRLS 2016 genutzt (vgl. IEA, 2019). Diese Datensätze können perspektivisch über das FDZ am IQB bezogen werden.

Als Leistungsvariable für den ersten Schritt der Analysen zur Identifizierung von Ausreißern dienten im Falle des TIMSS-Datensatzes die Kompetenzen in den Naturwissenschaften und in Mathematik. Im Falle von PIRLS diente die Kompetenz in Lesen als abhängige Variable.

Der soziale Status der Schülerinnen und Schüler wurde durch einen Sozial-Index abgebildet. Die Bildung des sozialen Index basiert auf den Variablen zum höchsten sozioökonomischen Status der Eltern (ISEI), dem Besitz eines eigenen Zimmers, der Anzahl der Bücher im Haushalt, dem verfügbaren Geld im Haushalt, der Anzahl der Besuche im Theater im letzten Jahr, der Anzahl der Besuche in einem Museum oder einer Kunstaustellung im letzten Jahr, der Anzahl der Besuche in der Oper, im Ballett oder in einem klassischen Konzert im letzten Jahr und dem Migrationsstatus der Eltern. Diese Auswahl der Items zur Bildung des sozialen Index hat sich in verschiedenen Studien bereits bewährt (vgl. Smith, Wendt & Kasper, 2017). Details zu diesen Variablen finden sich im Anhang A zu diesem Bericht.

Aus dem kombinierten Mehrebenen Input-Prozess-Output-Modell wurden folgende Variablen verwendet, um die Gruppenzugehörigkeit der Schülerinnen und Schüler mit erwartungswidrig niedrigen

und hohen Leistungen zu erklären: Die Motivation der Schülerinnen und Schüler, die Bildungsorientierung der Eltern, das Interesse der Schülerinnen und Schüler am Lesen bzw. an Naturwissenschaften und Mathematik, die Verfügbarkeit von Computern (einschließlich Tablets) im Leseunterricht, die Häufigkeit der Computernutzung (einschließlich Tablets) zu Hause, um Aufgaben für die Schule zu erledigen, die Häufigkeit der Computernutzung (einschließlich Tablets) in der Schule, um Aufgaben für die Schule zu erledigen und die Verwendung von Computerprogrammen zur Leistungsdifferenzierung. Details zu diesen Variablen finden sich im Anhang A zu diesem Bericht.

#### 4.5.2 Equating

Zur Vorhersage der Kompetenzen der Schülerinnen und Schüler in Mathematik, Naturwissenschaften und Lesen wurde der soziale Index der Schülerinnen und Schüler genutzt. Der soziale Index gibt als aggregiertes Merkmal Information darüber, inwiefern die Schülerinnen und Schüler mit sozialen Potenzialen ausgestattet sind. Ein hoher Wert auf dem Index steht für eine relativ hohe Verfügbarkeit von sozialen Potenzialen, ein niedriger Index verweist auf eine relativ geringe Verfügbarkeit von sozialen Potenzialen. Zur Bildung des sozialen Index wurde das *partial credit model* (PCM) verwendet; die einbezogenen Indikatoren sind in Abschnitt 4.5.1 aufgeführt. Die rechnerische Umsetzung des Equatings ist im Anhang C beschrieben.

#### 4.5.3 Identifizierung von Schülerinnen und Schülern mit erwartungswidrigen Leistungen

Zur Ermittlung der Gruppen von Schülerinnen und Schülern mit erwartungswidrig niedrigen und hohen Leistungen wurden die Leistungen der Schülerinnen und Schüler auf den sozialen Index regressiert. Tabelle 1 zeigt die Ergebnisse dieser Regression für die Domänen Mathematik, Naturwissenschaften und Lesen. Der soziale Index erklärt in allen drei Kompetenzbereichen die Leistung in dem Sinne, dass Schülerinnen und Schüler mit einer relativ hohen Verfügbarkeit von sozialen Potenzialen auch höhere Leistungen in den Domänen zeigen. Die erklärten Varianzanteile für die drei Modelle liegen bei  $R^2 = 0.14$  (Mathematik),  $R^2 = 0.18$  (Naturwissenschaften) und  $R^2 = 0.19$  (Lesen) und verweisen auf eine mittlere Effektstärke.

Tabelle 1: Regression der Kompetenz in Mathematik, Naturwissenschaften und Lesen auf den sozialen Index

	Mathematik		Naturwissenschaften		Lesen	
	<i>b</i>	(SE)	<i>b</i>	(SE)	<i>b</i>	(SE)
Konstante	221.4**	(16.6)	156.7**	(17.6)	121.0**	(28.0)
Sozialer Index	3.1**	(0.2)	3.8**	(0.2)	4.2**	(0.3)

b = Regressionsgewicht (unstandardisiert).

Signifikanzniveau: \*\* = signifikant ( $p < .01$ ).

Bei jeder der drei Regressionen lassen sich Abweichungen der Schülerleistung von der Regressionsgeraden nach oben und nach unten beobachten. Dabei handelt es sich um Schülerinnen und Schüler, die Leistungen zeigen, die nicht auf deren sozialen Status zurückgeführt werden können. Im Anschluss an die Regressionen wurden pro Quartil des sozialen Status die größten fünf Prozent und die kleinsten fünf Prozent dieser Datenpunkte als Schülerinnen und Schüler mit erwartungswidrig hohen

bzw. erwartungswidrig niedrigen Leistungen definiert.<sup>7</sup> Basierend auf dieser Definition wurde eine neue abhängige Variable konstruiert. Auf dieser abhängigen Variablen haben Schülerinnen und Schüler mit erwartungswidrig hohen Leistungen eine eins, Schülerinnen und Schüler mit erwartungswidrig niedrigen Leistungen eine null, und alle anderen Schülerinnen und Schüler einen fehlenden Wert. Diese neue abhängige Variable wurde im nächsten Schritt der Analyse genutzt, um zu identifizieren, in welchen Merkmalen des kombinierten Mehrebenen Input-Prozess-Output-Modells sich die beiden Gruppen (Schülerinnen und Schüler mit erwartungswidrig hohen Leistungen vs. Schülerinnen und Schüler mit erwartungswidrig niedrigen Leistungen) unterscheiden.

## 5 Befunde der exemplarischen Analysen

Um zu identifizieren, in welchen Merkmalen des kombinierten Mehrebenen Input-Prozess-Output-Modells sich Schülerinnen und Schüler mit erwartungswidrig hohen von Schülerinnen und Schülern mit erwartungswidrig niedrigen Leistungen unterscheiden, wurden logistische Regressionen berechnet. In diesen logistischen Regressionen ging die Gruppenzugehörigkeit als abhängige Variable ein, und als Prädiktoren wurden ausgewählte Merkmale des kombinierten Mehrebenen Input-Prozess-Output-Modells verwendet (s. Abschnitt 4.2). Tabelle 2 gibt die Ergebnisse der logistischen Regression wieder.

Wie zu erkennen, steigt die Wahrscheinlichkeit zur Gruppe der Schülerinnen und Schüler mit erwartungswidrig hohen Leistungen in Mathematik zu gehören mit dem Interesse an Mathematik ( $b = 0.4$ ). Die Häufigkeit der Computernutzung (einschließlich Tablets) in der Schule, um Aufgaben für die Schule zu erledigen hängt allerdings negativ mit der Gruppenzugehörigkeit zusammen: Schülerinnen und Schüler, die angeben den Computer jeden Tag ( $b = -1.9$ ) oder 1- bis 2-mal pro Woche ( $b = -1.5$ ) in der Schule zu nutzen, um Aufgaben für die Schule zu erledigen haben eine niedrigere Wahrscheinlichkeit zur Gruppe der Schülerinnen und Schüler mit erwartungswidrig hohen Leistungen in Mathematik zu gehören als Schülerinnen und Schüler die angeben, den Computer nie oder fast nie zu nutzen.

Im Falle von Naturwissenschaften sagt die Häufigkeit der Computernutzung (einschließlich Tablets) zu Hause, um Aufgaben für die Schule zu erledigen die Gruppenzugehörigkeit vorher. Schülerinnen und Schüler, die den PC zu Hause jeden Tag ( $b = -1.4$ ) nutzen, um Aufgaben für die Schule zu erledigen haben eine signifikant geringere Wahrscheinlichkeit zur Gruppe der Schülerinnen und Schüler mit erwartungswidrig hohen Leistungen in Naturwissenschaften zu gehören als Schülerinnen und Schüler, die den PC nie oder fast nie zur Erledigung dieser Aufgaben zu Hause nutzen.

In Bezug auf die Lesekompetenz sagt die Verfügbarkeit von Computern im Leseunterricht die Gruppenzugehörigkeit vorher. Offensichtlich nehmen Schülerinnen und Schüler mit erwartungswidrig niedrigen Leistungen häufiger an Leseunterricht teil, in dem Computer zur Verfügung stehen als Schülerinnen und Schüler mit erwartungswidrig hohen Leistungen ( $b = -0.3$ ). Das Interesse, die Moti-

---

<sup>7</sup> Die Setzung der fünf Prozent Marke erfolgte aus inhaltlich-methodischen Überlegungen: Zielmarken unterhalb von 5 Prozent führen bei den vorliegenden Datensätzen zu einer Fallzahl pro Gruppe (weniger als 200 Schülerinnen und Schüler pro Gruppe), mit der selbst mittlere Effektstärken von kurzfristig veränderbaren Lernbedingungen kaum mehr entdeckt werden könnten; bei Zielmarken oberhalb von 5 Prozent besteht hingegen zum einen die Gefahr, praktisch irrelevante Effekte von kurzfristig veränderbaren Lernbedingungen als statistisch signifikant auszuweisen, zum anderen wäre es mitunter auch nicht mehr möglich, trennscharfe kurzfristig veränderbare Lernbedingungen zu identifizieren, d.h. das Prinzip vom Lernen von dem Stärksten (s. Abschnitt 4.1) wäre ausgehebelt.

vation und auch die Bildungsorientierung der Eltern haben indes keinen Einfluss auf die Gruppenzugehörigkeit.

Tabelle 2: Logistische Regression der Gruppenzugehörigkeit (Schülerinnen und Schüler mit erwartungswidrig hohen vs. Schülerinnen und Schüler mit erwartungswidrig niedrigen Leistungen) auf ausgewählte Merkmale des kombinierten Mehrebenen Input-Prozess-Output-Modells

	Mathematik		Naturwissenschaften		Lesen	
	<i>b</i>	(SE)	<i>b</i>	(SE)	<i>b</i>	(SE)
Konstante	-2.7 <sup>ns</sup>	(2.4)	0.1 <sup>ns</sup>	(2.1)	-4.3 <sup>*</sup>	(2.3)
Motivation	-0.2 <sup>ns</sup>	(0.2)	0.0 <sup>ns</sup>	(0.2)	0.1 <sup>ns</sup>	(0.2)
Interesse	0.4 <sup>*</sup>	(0.2)	0.0 <sup>ns</sup>	(0.2)	0.3 <sup>ns</sup>	(0.2)
Bildungsorientierung	0.1 <sup>ns</sup>	(0.1)	0.0 <sup>ns</sup>	(0.1)	0.1 <sup>ns</sup>	(0.1)
Verfügbarkeit von Computern (einschließlich Tablets) im Leseunterricht					-0.3 <sup>*</sup>	(0.2)
Nutzung des PC in der Schule (jeden Tag) <sup>A</sup>	-1.9 <sup>*</sup>	(1.0)				
Nutzung des PC in der Schule (1- bis 2-mal pro Woche) <sup>B</sup>	-1.5 <sup>*</sup>	(0.8)				
Nutzung des PC in der Schule (1- bis 2-mal pro Monat) <sup>C</sup>	-0.4 <sup>ns</sup>	(6.6)				
Nutzung des PC zu Hause (jeden Tag) <sup>D</sup>			-1.4 <sup>**</sup>	(0.5)		
Nutzung des PC zu Hause (1- bis 2-mal pro Woche) <sup>E</sup>			0.0 <sup>ns</sup>	(0.6)		
Nutzung des PC zu Hause (1- bis 2-mal pro Monat) <sup>F</sup>			1.0 <sup>ns</sup>	(0.7)		
Computerprogrammen zur Leistungsdifferenzierung	21.7 <sup>ns</sup>	(13.8)	13.4 <sup>ns</sup>	(15.8)		

*b* = Regressionsgewicht (unstandardisiert).

Signifikanzniveau: ns = nicht signifikant; \* = signifikant ( $p < .05$ ); \*\* = signifikant ( $p < .01$ ).

A = Häufigkeit der Computernutzung (einschließlich Tablets) in der Schule, um Aufgaben für die Schule zu erledigen (0 = nie oder fast nie; 1 = jeden Tag oder fast jeden Tag).

B = Häufigkeit der Computernutzung (einschließlich Tablets) in der Schule, um Aufgaben für die Schule zu erledigen (0 = nie oder fast nie; 1 = 1- bis 2-mal pro Woche).

C = Häufigkeit der Computernutzung (einschließlich Tablets) in der Schule, um Aufgaben für die Schule zu erledigen (0 = nie oder fast nie; 1 = 1- bis 2-mal pro Monat).

D = Häufigkeit der Computernutzung (einschließlich Tablets) zu Hause, um Aufgaben für die Schule zu erledigen (0 = nie oder fast nie; 1 = jeden Tag oder fast jeden Tag).

E = Häufigkeit der Computernutzung (einschließlich Tablets) zu Hause, um Aufgaben für die Schule zu erledigen (0 = nie oder fast nie; 1 = 1- bis 2-mal pro Woche).

F = Häufigkeit der Computernutzung (einschließlich Tablets) zu Hause, um Aufgaben für die Schule zu erledigen (0 = nie oder fast nie; 1 = 1- bis 2-mal pro Monat).



## 6 Zusammenfassende inhaltliche Interpretation der Modellierung

Mit den bis hier vorgestellten exemplarischen Analysen wurde gezeigt, inwiefern es möglich ist, in Datensätzen der empirischen Bildungsforschung die Gruppe der Schülerinnen und Schüler mit erwartungswidrig hohen Leistungen von der Gruppe der Schülerinnen und Schüler mit erwartungswidrig niedrigen Leistungen zu unterscheiden. Wir wollten auch herausfinden, inwiefern Verfahren des Equating dabei zu berücksichtigen sind. Letztlich sollte auch überprüft werden, ob sich Merkmale des kombinierten Mehrebenen Input-Prozess-Output-Modells finden lassen, in denen sich die Gruppe der Schülerinnen und Schüler mit erwartungswidrig hohen Leistungen von der Gruppe der Schülerinnen und Schüler mit erwartungswidrig niedrigen Leistungen deutlich unterscheiden. Wir haben dabei den Ansatz der linearen Regression mit der Kontrollvariable des sozialen Index gewählt. Im Weiteren haben wir dann pro Quartil des sozialen Status die größten fünf Prozent und die kleinsten fünf Prozent der Datenpunkte als Schülerinnen und Schüler mit erwartungswidrig hohen bzw. erwartungswidrig niedrigen Leistungen definiert. Die resultierende Gruppenzugehörigkeit wurde als abhängige Variable in den weiteren Analysen genutzt.

Die Ergebnisse zum Equating zeigen deutlich, dass Equating-Verfahren einen Einfluss auf die Ergebnisse und damit die Schlussfolgerungen der vorgeschlagenen Analysen haben können. Ohne Equating unterscheiden sich die Schülerinnen und Schüler in TIMSS 2015 nicht von den Schülerinnen und Schülern in PIRLS 2016 in Bezug auf den sozialen Status. Nach erfolgtem Equating zeigte sich, dass die Schülerinnen und Schüler in TIMSS im Durchschnitt einen niedrigen sozialen Status aufwiesen gegenüber den PIRLS 2016 Schülerinnen und Schülern.

Die Analysen zur Erklärung von erwartungswidrigen Leistungen legen nahe, dass es möglich ist, mit Merkmalen des kombinierten Mehrebenen Input-Prozess-Output-Modells diese erwartungswidrigen Leistungen zu erklären. In Bezug auf die Kompetenz in Mathematik können das Interesse und die Häufigkeit der Computernutzung (einschließlich Tablets) in der Schule, um Aufgaben für die Schule zu erledigen erwartungswidrige Leistungen erklären. Es ist also zu vermuten, dass Chancenungleichheit kompensiert werden kann, wenn das Interesse an Mathematik bei Schülerinnen und Schülern mit erwartungswidrig niedrigen Leistungen in Mathematik gefördert wird. In Bezug auf die Häufigkeit der Computernutzung (einschließlich Tablets) in der Schule, um Aufgaben für die Schule zu erledigen fällt der Befund etwas komplexer aus: Schülerinnen und Schüler mit erwartungswidrig niedrigen Leistungen in Mathematik nutzen den Computer bereits häufiger in der Schule, um Aufgaben für die Schule zu erledigen als Schülerinnen und Schüler mit erwartungswidrig hohen Leistungen.

Eine einfache Erhöhung der Ressourcen scheint hier nicht die Lösung zu sein. Vielmehr wäre zu ergründen, was genau sich hinter diesem Befund verbirgt. Haben die Lehrerinnen und Lehrer diese Schülerinnen und Schüler vielleicht bereits animiert, mehr mit dem Computer zu arbeiten; dann wäre in Paneldesigns zu überprüfen, ob dieses im Zeitverlauf einen positiven Effekt auf die Leistung der Schülerinnen und Schüler hat. Oder nutzen die Schülerinnen und Schüler mit erwartungswidrig niedrigen Leistungen den Computer zwar häufiger zu Lösung von Aufgaben für die Schule, sind aber mit der Bedienung der jeweiligen Softwareprogramme nicht vertraut bzw. nutzen die falschen Programme; dann wären geeignetere Programme bzw. Schulungen für diese Schülerinnen und Schüler zu etablieren. In jedem Fall bedarf dieser Befund weiterer vertiefender Analysen, die mit dem hier genutzten Datenmaterial allerdings nicht möglich sind.

Dieselbe Schlussfolgerung lässt sich auch für die naturwissenschaftliche Kompetenz und die Kompetenz in Lesen ableiten. In Bezug auf die Naturwissenschaften ist es die Häufigkeit der Computernutzung (einschließlich Tablets) zu Hause, um Aufgaben für die Schule zu erledigen, die negativ mit der Gruppenzugehörigkeit korreliert. In Bezug auf die Lesekompetenz korreliert die Verfügbarkeit von Computern im Leseunterricht negativ mit der Gruppenzugehörigkeit. Auch hier wäre in vertiefenden Analysen zu untersuchen, wie genau diese Computernutzung aussieht bzw. was die Verfügbarkeit von Computern im Leseunterricht impliziert. Generell lässt sich über alle drei Domänen allerdings festhalten, dass der digitale Faktor sehr wahrscheinlich erfolgsversprechend genutzt werden kann, um Chancenungleichheiten zu kompensieren.

## 7 Fazit und Ausblick

Die vorliegenden Befunde und Arbeitsstände aus der Machbarkeitsstudie zeigen, dass eine weitergehende und ausgeweitete Untersuchung bezüglich der Konfiguration von sozialen Hintergrundmerkmalen der Schülerinnen und Schüler und durch das Bildungssystem bedingten Merkmalen des Lehrens und Lernens auch für ausführliche und vertiefende Untersuchungen im Primar- und Sekundarbereich zielführend erscheint. Darüber könnte auf relevante Ausstattungsmerkmale aufmerksam gemacht werden, die einen Faktor bei der Beschreibung von (kontrollierten) Leistungsdisparitäten darstellen. Die Analysen verweisen darauf, dass Chancenungleichheit kompensiert werden könnte, wenn das Interesse z.B. an Mathematik bei Schülerinnen und Schülern mit erwartungswidrig niedrigen Leistungen in Mathematik gefördert würde. Des Weiteren legt diese Machbarkeitsstudie nahe, die angedachten Analysen nicht einmalig in einer Anschlussstudie zu realisieren, sondern im Sinne eines Monitorings wiederholt – auch mit Blick auf andere Etappen des Bildungssystems – durchzuführen. Diese Folgerung leitet sich aus verschiedenen in der Machbarkeitsstudie dargelegten Gründen ab: Zum einen sind kontinuierliche Veränderungen im Bildungssystem zu beobachten, die einen Einfluss auf die Ergebnisse einer Anschlussstudie und damit auch auf darauf basierenden Empfehlungen zur Ausschöpfung von Bildungspotenzialen haben können. Zum anderen handelt es sich bei den meisten zur Verfügung stehenden Datensätzen der empirischen Bildungsforschung um Trenddaten, die kausale Aussagen nur bei Betrachtungen über die Zeit hinreichend plausibilisieren.

Aus methodischer Perspektive zeigt sich, dass sowohl der technisch gewählte Ansatz des Equatings als auch die analytische Identifizierung von erwartungswidrig hohen bzw. niedrigen Leistungen zielführend erscheint. Die vorgelegte Betrachtung der Lage in den Grundschulen bildet nur eine erste Etappe der schulisch begleiteten Lernprozesse ab. Um festzustellen ob die identifizierten Effekte von Interesse und Motivation und der Computernutzung auch im weiteren Bildungsverlauf eine relevante Rolle spielen und in welcher Konfiguration sie sich zeigen, ist der Blick auf die Sekundarstufe zu erweitern. Basierend auf den Ergebnissen und Erfahrungen der Machbarkeitsstudie lassen sich folgende Empfehlungen für mögliche Expertisen im Rahmen einer Anschlussstudie ableiten.

1. Die vorliegende Machbarkeitsstudie beschränkt sich auf Analysen von Daten aus der Grundschule, da sich hier die Schulstruktur homogener (auch zwischen den Bundesländern) zeigt als in der anschließenden Sekundarstufe. Die Frage der Kompensation von Bildungsdisparitäten stellt sich dennoch auch für die älteren Schülerinnen und Schüler – gerade auch vor dem Hintergrund, dass sich Benachteiligungseffekte über die Zeit kumulieren (können). Bei der Betrachtung von Schülerinnen und Schülern in der Sekundarstufe (z.B. anhand von Daten aus den vorliegenden PISA bzw. NEPS Erhebungen) wäre zu prüfen, ob eine Analyse der Entwicklung (und damit differenzieller Entwicklungsverläufe insbesondere unter Berücksichtigung unterschiedlicher schulischer bzw.

unterrichtlicher Rahmenbedingungen) oder eine Fokussierung auf reale Abschlüsse für die Frage des Abbaus von Bildungsdisparitäten zielführend erscheint. Auch wenn testtheoretisch die in groß angelegten Schulvergleichsuntersuchungen festgestellten Kompetenzstände bessere empirische Eigenschaften aufweisen, so stellen doch die Abschlüsse (Abschlussnoten bzw. Zusammensetzung von Abschlussnoten) die Datenrealität dar, an denen sich aufnehmende Institutionen wie Schule, aber auch Arbeitgeber orientieren. Der Übergang in einen (bestenfalls) weiteren Ausbildungsschritt dokumentiert die zu diesem Zeitpunkt tatsächlich realisierte Chance.

2. Bei dem Design der genutzten IEA-Studien zeigen sich Grenzen der Analysen. Bei der Identifizierung der erwartungswidrig hoch bzw. niedrig abscheidenden Schülerinnen und Schüler ergeben sich (aufgrund der für Populationsschätzungen optimierten Bestimmung der Kompetenzen von Schülerinnen und Schülern mit Plausible Values) immer verschiedene Gruppeneinstellungen (in Abhängigkeit von den Plausible Values), sodass eine direkte Zuordnung von Klassen bzw. Schulmerkmalen nicht möglich ist, d.h. Merkmale des kombinierten Mehrebenen Input-Prozess-Output-Modells auf der Klassen-/Schulebene bzw. Systemebene lassen sich nicht ohne weiteres in die statistischen Modelle mit aufnehmen. Für anschließende Untersuchungen wäre ein Lösungsweg, dem in der NEPS-Studie eingeschlagenen Weg zu folgen und die Kompetenzen der Schülerinnen und Schüler optimiert zur Feststellung der Individualleistungen anhand von WLE Schätzern zu bestimmen. Den damit in Kauf genommenen Nachteil der eingeschränkten optimierten Beschreibung des Bildungssystems würde dafür eine Verknüpfbarkeit der Individualdaten mit Klassen- bzw. Schuldaten entgegenstehen.
3. Die Machbarkeitsstudie verfolgte den Ansatz des Lernens vom Stärksten. Darauf basierend wurden die Gruppen der erwartungswidrigen Leistungen mit der fünf Prozentmarke definiert. Dem gegenüber stünde ein Ansatz der Beschreibung der Gesamtpopulation (differentielles Lernen von allen Schülerinnen und Schülern). In diesem Ansatz würden nicht nur zwei Extremgruppen gebildet, sondern beispielsweise fünf Gruppen (z.B. 5 %; 25 %; 40 %; 25 %; 5 % – der Verteilung der Abweichungswerte). Diese Gruppen könnten dann beispielsweise mit Hilfe von multinominallogistischen Regressionen verglichen werden. Der eingeschränkten Klarheit der Gegenüberstellung von nur zwei Extremgruppen (wie in dieser Machbarkeitsstudie gewählt) stünde eine umfassende Beschreibung der Gesamtstichprobe gegenüber.
4. In einigen Bundesländern wurden bzw. werden die Möglichkeiten geschaffen, amtliche Schulstatistiken mit den in den jeweiligen Bundesländern erhobenen Leistungsdaten (z.B. aus Vergleichsarbeiten, Bildungstrend) zu verknüpfen und für wissenschaftliche Reanalysen zur Verfügung zu stellen. Diese Daten weisen aktuell jedoch wesentliche Einschränkungen auf: Die Daten sind teilweise regional stark begrenzt und damit nur auf ein spezifisches Bundesland beziehbar. Auch wenn sich diese Untersuchungen an hohen methodischen Standards orientieren, weisen sie dennoch inhaltlich eine sehr heterogene Bandbreite in der Realisierung der Erhebungsinstrumente und Routinen auf, was bundeslandübergreifende Analysen sehr erschwert bzw. unmöglich macht. Neben dieser strukturellen Einschränkung erweist sich auch die Dokumentation der Erhebungsinstrumente als herausfordernd. Die wissenschaftlichen Standards der Übermittlung von Daten und deren Dokumentation von durch Fördermittel finanzierten Untersuchungen (wie z.B. PIRLS, TIMSS oder PISA) an Forschungsdatenzentren ist weder bei amtlichen Schulstatistiken noch bei länderspezifischen Schul- bzw. Unterrichtsbezogenen Untersuchungen durch die jeweiligen Landesinstitutionen Standard. Sollte die Nutzung von amtlichen Schulstatistiken (und den in den Bundesländern erhobenen Kompetenzständen der Schülerinnen und Schüler) in den Blick genommen werden, wäre zunächst zu prüfen, ob die Bundesländer bzw. Schulstatistiken ausrei-

chend differenzierte Informationen über beispielsweise die soziokulturelle bzw. sozioökonomische Situation der Schülerinnen und Schüler und deren Familien enthalten.

5. Wenn Skalen, Index oder Summenwerte verschiedener (inhaltsgleicher) Messinstrumente miteinander verglichen werden sollen, dann sollte auf Verfahren des Equating zurückgegriffen werden. Nur so kann gewährleistet werden, dass die Werte auf diesen Skalen nicht nur inhaltlich, sondern auch numerisch vergleichbar sind.
6. In der Machbarkeitsstudie wurden Trenddaten verwendet. Es wurde allerdings auch aufgezeigt, dass sich die Bildungsbedingungen über die Zeit stark verändern. Diese Veränderungen sind bei der Interpretation der Befunde zu berücksichtigen. Interessant wäre in diesem Sinne etwa der Vergleich von Trenddaten mit Analyseergebnissen aus Paneldaten (wie etwa aus der Studie Kompetenzen und Einstellungen von Schülerinnen und Schülern; KESS 4; Bos & Pietsch, 2006).

## Literatur

- Baumert, J., Cortina, K. S. & Leschinsky, A. (2008). Grundlegende Entwicklungen und Strukturprobleme im allgemeinbildenden Schulwesen. In K. S. Cortina, J. Baumert, A. Leschinsky, K. U. Mayer & L. Trommer (Hrsg.), *Das Bildungswesen in der Bundesrepublik Deutschland. Strukturen und Entwicklungen im Überblick* (S. 53–130). Reinbek: Rowohlt.
- BMBF (Bundesministerium für Bildung und Forschung). (2009). *Gut angelegt. Das Investitionsprogramm Zukunft Bildung und Betreuung*. Berlin: BMBF.
- BMBF (Bundesministerium für Bildung und Forschung). (2017). *Mit dem DigitalPakt Schulen zukunftsfähig machen*. Verfügbar unter: <https://www.bmbf.de/de/mit-dem-digitalpakt-schulen-zukunftsfahig-machen-4272.html> [25.04.2019].
- Bos, W., Gröhlich, C. & Bensen, M. (2009). Der Belastungsindex für die Schulen der Sekundarstufe I in Hamburg. In W. Bos, M. Bensen & C. Gröhlich (Hrsg.), *KESS 7. Kompetenzen und Einstellungen von Schülerinnen und Schülern an Hamburger Schulen zu Beginn der Jahrgangsstufe 7* (S. 123–132). Münster: Waxmann.
- Bos, W. & Pietsch, M. (2006). *KESS 4. Kompetenzen und Einstellungen von Schülerinnen und Schülern am Ende der Jahrgangsstufe 4 in Hamburger Grundschulen*. Münster: Waxmann.
- Boudon, R. (1974). *Education, opportunity, and social inequality. Changing prospects in Western societies*. New York: Wiley.
- Bremerich-Vos, A., Wendt, H. & Bos, W. (2017). Lesekompetenzen im internationalen Vergleich: Testkonzeption und Ergebnisse. In A. Hußmann, H. Wendt, W. Bos, A. Bremerich-Vos, D. Kasper, E.-M. Lankes, N. McElvany, T. C. Stubbe & R. Valtin (Hrsg.), *IGLU 2016. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 79–142). Münster: Waxmann.
- Creemers, B. P. M. & Kyriakides, L. (2008). *The dynamics of educational effectiveness: A contribution to policy, practice and theory in contemporary school*. London: Routledge.
- Davies, M., Holland, P. W. & Thayer, D. T. (2004). *The kernel method of test equating*. New York: Springer.
- Davies, M., Xu, X. & Carstensen, C. H. (2011). Measuring growth in longitudinal large-scale assessment with general latent variable model. *Psychometrika*, 76 (2), 318–336.
- Ditton, H. (2002). Unterrichtsqualität – Konzeptionen, methodische Überlegungen und Perspektiven. *Unterrichtswissenschaft*, 20 (3), 197–212.

- Döbert, H. & Seeber, S. (2009). *Gutachten zur Machbarkeit eines ‚Chancenspiegels‘ im Projekt ‚Heterogenität und Bildung‘ der Bertelsmann Stiftung – Machbarkeitsstudie*. Gütersloh: Bertelsmann Stiftung.
- Eickelmann, B., Lorenz, R., Vennemann, M., Gerick, J. & Bos, W. (Hrsg.). (2014). *Grundschule in der digitalen Gesellschaft. Befunde aus den Schulleistungsstudien IGLU und TIMSS 2011*. Münster: Waxmann.
- Fischer, G. H. (1996). IRT-Modelle als Forschungsinstrumente der Differentiellen Psychologie. In K. Pawlik (Hrsg.), *Grundlagen und Methoden der Differentiellen Psychologie* (S. 673–729). Göttingen: Hogrefe.
- Gerick, J., Vennemann, M., Lorenz, R. & Eickelmann, B. (2014). Schulische Ausstattung mit digitalen Medien in der Grundschule. In B. Eickelmann, R. Lorenz, M. Vennemann, J. Gerick & W. Bos (Hrsg.), *Grundschule in der digitalen Gesellschaft. Befunde aus den Schulleistungsstudien IGLU und TIMSS 2011* (S. 19–34). Münster: Waxmann.
- Goy, M., Valtin, R. & Hußmann, A. (2017). Leseselbstkonzept, Lesemotivation, Leseverhalten und Lesekompetenz. In A. Hußmann, H. Wendt, W. Bos, A. Bremerich-Vos, D. Kasper, E.-M. Lankes, N. McElvany, T. C. Stubbe & R. Valtin (Hrsg.), *IGLU 2016. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 143–176). Münster: Waxmann.
- Hartig, J. & Klieme E. (2006). Kompetenz und Kompetenzdiagnostik. In K. Schweizer (Hrsg.), *Leistung und Leistungsdiagnostik* (S. 127–143). Heidelberg: Springer.
- Helmke, A. (2004). *Unterrichtsqualität erfassen, bewerten, verbessern* (3. Aufl.). Seelze: Kallmeyer.
- Hußmann, A., Stubbe, T. C. & Kasper, D. (2017). Soziale Herkunft und Lesekompetenzen von Schülerinnen und Schülern. In A. Hußmann, H. Wendt, W. Bos, A. Bremerich-Vos, D. Kasper, E.-M. Lankes, N. McElvany, T. C. Stubbe & R. Valtin (Hrsg.), *IGLU 2016. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 195–218). Münster: Waxmann.
- Hußmann, A., Wendt, H., Bos, W., Bremerich-Vos, A., Kasper, D., Lankes, E.-M., McElvany, N., Stubbe, T. C. & Valtin, R. (2017). *IGLU 2016. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich*. Münster: Waxmann.
- IEA. (2019). *IEA study data repository*. Verfügbar unter: <https://www.iea.nl/repository/studies> [11.02.2018].
- Kasper, D., Wendt, H., Bos, W. & Köller, O. (2016). Trends in mathematischen und naturwissenschaftlichen Kompetenzen am Ende der Grundschulzeit in Deutschland. In H. Wendt, W. Bos, C. Selter, I. Köller, K. Schwippert & D. Kasper (Hrsg.), *TIMSS 2015. Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 367–382). Münster: Waxmann.
- Klafki, W. (2007). *Neue Studien zur Bildungstheorie und Didaktik. Zeitgemäße Allgemeinbildung und kritisch-konstruktive Didaktik* (6., neu ausgestattete Aufl.). Weinheim: Beltz.
- Klemm, K. (2018). *Unterwegs zur inklusiven Schule. Lagebericht 2018 aus bildungsstatistischer Perspektive*. Gütersloh: Bertelsmann Stiftung. Verfügbar unter: [www.bertelsmann-stiftung.de/fileadmin/files/BSt/Publikationen/GrauePublikationen/Studie\\_IB\\_Unterwegs-zur-inkluisiven-Schule\\_2018.pdf](http://www.bertelsmann-stiftung.de/fileadmin/files/BSt/Publikationen/GrauePublikationen/Studie_IB_Unterwegs-zur-inkluisiven-Schule_2018.pdf) [06.11.2019].
- KMK (Sekretariat der ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland). (n.d.). *Allgemeinbildende Schulen in Ganztagsform in den Ländern in der Bundesrepublik*. Verfügbar unter: <https://www.kmk.org/dokumentation-statistik/statistik/schulstatistik/allgemeinbildende-schulen-in-ganztagsform.html> [25.04.2019].

- KMK (Sekretariat der ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland – Kommission für Statistik). (2018). *Definitionenkatalog zur Schulstatistik 2018*. Verfügbar unter: <https://www.kmk.org/fileadmin/Dateien/pdf/Statistik/Defkat2018.pdf> [24.04.2019].
- KMK (Sekretariat der ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland). (2019). *Statistische Veröffentlichung der Kultusministerkonferenz. Dokumentation Nr. 217. Schüler, Klassen, Lehrer und Absolventen der Schule 2008 bis 2017*. Verfügbar unter [https://www.kmk.org/fileadmin/Dateien/pdf/Statistik/Dokumentationen/SKL\\_2017\\_Dok\\_217.pdf](https://www.kmk.org/fileadmin/Dateien/pdf/Statistik/Dokumentationen/SKL_2017_Dok_217.pdf) [30.04.2019].
- Kolen, M. J. & Brennan, R. L. (2004). *Test equating, scaling, and linking. Methods and practices*. New York: Springer.
- Maaz, K., Baumert, J. & Cortina, K. S. (2008). Soziale und regionale Ungleichheit im deutschen Bildungssystem. In K. S. Cortina, J. Baumert, A. Leschinsky, K. U. Mayer & L. Trommer (Hrsg.), *Das Bildungswesen in der Bundesrepublik Deutschland: Strukturen und Entwicklungen im Überblick* (S. 205–243). Reinbek: Rowohlt.
- Maaz, K., Baumert, J. & Trautwein, U. (2009). Genese sozialer Ungleichheit im institutionellen Kontext der Schule: Wo entsteht und vergrößert sich soziale Ungleichheit? *Zeitschrift für Erziehungswissenschaft, Sonderheft 12*, 11–46.
- Martin, M. O. & Mullis, I. V. S. (2012). *Methods and procedures in TIMSS and PIRLS 2011*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Martin, M. O., Mullis, I. V. S., Foy, P. & Hooper, M. (2016). *TIMSS 2015 international results in science*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Martin, M. O., Mullis, I. V. S. & Hooper, M. (2016). *Methods and procedures in TIMSS 2015*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Martin, M. O., Mullis, I. V. S. & Hooper, M. (2017). *Methods and procedures in PIRLS 2016*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Mpfs (Medienpädagogischer Forschungsverbund Südwest). (2017). *KIM-Studie 2016 Kindheit, Internet, Medien. Basisstudie zum Medienumgang 6- bis 13-Jähriger in Deutschland*. Stuttgart: mpfs.
- Mullis, I. V. S., Martin, M. O., Foy, P. & Hooper, M. (2016). *TIMSS 2015 international results in mathematics*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Mullis, I. V. S., Martin, M. O., Foy, P. & Hooper, M. (2017). *PIRLS 2016 International Results in Reading*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- OECD. (2016). *PISA 2015 results (volume I): Excellence and equity in education*. Paris: OECD Publishing. <http://dx.doi.org/10.1787/9789264266490-en>.
- OECD. (2017). *PISA 2015. Technical report*. Paris: OECD Publishing.
- Peisert, H. (1967). *Soziale Lage und Bildungschancen in Deutschland*. München: Piper.
- Reiss, K., Sälzer, C., Schiepe-Tiska, A., Klieme, E. & Köller, O. (2016). *PISA 2015. Eine Studie zwischen Kontinuität und Innovation*. Münster: Waxmann.
- Roemer, J. (1998). *Equality of opportunity*. Cambridge, MA: Harvard University Press.
- Rost, J. (2006). Item-Response-Theorie. In F. Petermann & M. Eid (Hrsg.), *Handbuch der psychologischen Diagnostik*. Göttingen: Hogrefe.
- Rychen, D. S. (2008). OECD Referenzrahmen für Schlüsselkompetenzen – ein Überblick. In I. Bormann & G. de Haan (Hrsg.), *Kompetenzen der Bildung für nachhaltige Entwicklung. Operationalisierung, Messung, Rahmenbedingungen, Befunde* (S. 15–22). Wiesbaden: VS Verlag für Sozialwissenschaften.

- Schulz-Heidorf, K. (2016). *Individuelle Förderung im Unterricht: Eine Möglichkeit, soziale Herkunft und Schulerfolg zu entkoppeln? Eine Re-Analyse aus IGLU-E 2011*. Berlin: epubli.
- Schwippert, K. (2009). Gibt es Chancengleichheit in einer heterogenen Gesellschaft? In I. Sylvester, I. Sieh, M. Menz, H.-W. Fuchs & J. Behrendt (Hrsg.), *Bildung – Recht – Chancen. Rahmenbedingungen, empirische Analysen und internationale Perspektiven zum Recht auf chancengleiche Bildung* (S. 81–94). Münster: Waxmann.
- Smith, D. S., Wendt, H. & Kasper, D. (2017). Social reproduction and sex in German primary schools. *Compare. A journal of Comparativ and International Education*, 47 (2), 240–256. doi:10.1080/03057925.2016.1158643
- Stubbe, T. C., Schwippert, K. & Wendt, H. (2016). Soziale Disparitäten der Schülerleistungen in Mathematik und Naturwissenschaften. In H. Wendt, W. Bos, C. Selter, O. Köller, K. Schwippert & D. Kasper (Hrsg.), *TIMSS 2015. Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 229–316). Münster: Waxmann.
- Tian, F. (2011). *A comparison of equating/linking using the Stocking-Lord method and concurrent calibration with mixed-format tests in the non-equivalent groups common-item design under IRT*. Chestnut Hill, MA: Boston College, Lynch School of Education. Verfügbar unter: <http://hdl.handle.net/2345/2370> [15.06.2019].
- Wendt, H., Bos, W., Selter, C., Köller, O., Schwippert, K. & Kasper, D. (2016). *TIMSS 2015. Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich*. Münster: Waxmann.
- Wendt, H., Kasper, D., Bos, W., Vennemann, M. & Goy, M. (2017). Wie viele Punkte auf der TIMSS-Metrik entsprechen einem Lernjahr? Leistungszuwächse in Mathematik und Naturwissenschaften am Ende der Grundschulzeit. In T. Eckert & B. Gniewosz (Hrsg.), *Bildungsgerechtigkeit* (S. 121–153). Wiesbaden: Springer VS.
- Wendt, H. & Schwippert, K. (2017). Lesekompetenzen von Schülerinnen und Schülern mit und ohne Migrationshintergrund. In A. Hußmann, H. Wendt, W. Bos, A. Bremerich-Vos, D. Kasper, E.-M. Lankes, N. McElvany, T. C. Stubbe & R. Valtin (Hrsg.), *IGLU 2016. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 219–234). Münster: Waxmann.
- Wendt, H., Schwippert, K. & Stubbe, T. C. (2016). Mathematische und naturwissenschaftliche Kompetenzen von Schülerinnen und Schülern mit Migrationshintergrund. In H. Wendt, W. Bos, C. Selter, I. Köller, K. Schwippert & D. Kasper (Hrsg.), *TIMSS 2015. Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 317–332). Münster: Waxmann.
- Wendt, H., Stubbe, T. C. & Schwippert, K. (2012). Soziale Herkunft und Lesekompetenzen von Schülerinnen und Schülern. In W. Bos, I. Tarelli, A. Bremerich-Vos & K. Schwippert (Hrsg.), *IGLU 2011. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 175–190). Münster: Waxmann.

## Anhang A: Für Equating nutzbare Items

Tabelle 3 gibt einen Überblick über die Merkmale, nach denen die Skalenhandbücher der ausgewählten Studien durchsucht wurden.

Tabelle 3: Ausgewählte Merkmale, nach denen die Skalenhandbücher durchsucht wurden

Individualebene	Schul-/Klassenebene
Sozioökonomische Herkunft	Unterrichtsmaterialien
Migrationsstatus	Digitale Medien
Individuelle Merkmale	Unterrichtsqualität
Motivation	Schulprofil
Interesse	Fördergruppen
Außerschulischer Kontext	Fortbildung
Vorwissen	Kollegiale Netzwerke
Elterliches Unterstützungsverhalten	Schulform

Tabelle 4 (Equating-Design) gibt einen Überblick über die Anzahl an Items pro Merkmal, die zwischen den Studien identisch sind.

Tabelle 4: Für das Equating nutzbare Items für die individuellen und Schul-/Klassenmerkmale

	TIMSS & NEPS	TIMSS & IGLU	NEPS & IGLU	TIMSS, IGLU & NEPS
<b>INDIVIDUUM</b>				
Sozioökonomischer Hintergrund	–	11	–	3
Migrationsstatus	–	10	–	1
Individuelle Merkmale	–	18	–	2
Motivation	–	–	–	–
Interesse	–	–	–	–
Außerschulischer Kontext	–	2	–	–
Vorwissen	2	2	–	–
Elterliche Unterstützung	–	3	–	–
<b>KLASSE/SCHULE</b>				
Unterrichtsmaterialien	–	1	–	–
Digitale Medien	–	2	–	–
Unterrichtsqualität	–	1	–	–
Schulprofil	–	1	–	1
Fördergruppen	–	1	1	–
Fortbildung	–	1	–	–
Kollegiale Netzwerke	–	2	–	–
Schulform	–	–	–	–

Da es sich um verschiedene Studien handelt, weisen die Merkmale zum Teil unterschiedliche Antwortkategorien auf. Um eine höhere Anzahl an gleichen Items für das Equating-Verfahren zu erhalten



ten, können wo möglich, diese unterschiedlichen Kategorien angepasst werden. Ein Beispiel für diese Zusammenfassung liefert Tabelle 5. Das Merkmal *Bücherbesitz zu Hause* wurde in allen drei Studien erfasst. In TIMSS und IGLU wurden dieselben Antwortkategorien vorgegeben, in NEPS wurden im oberen Bereich differenziertere Antwortkategorien vorgegeben (201 bis 500 Bücher und mehr als 500 Bücher). Vor dem Equating können diese beiden Kategorien in eine Kategorie (mehr als 200 Bücher) zusammengefasst werden.

Tabelle 5: Beispiel für das neukodieren der Variable „Bücherbesitz zu Hause“ des Elternfragebogens

<b>Alte Kodierung NEPS</b>	<b>Neue Kodierung NEPS</b>	<b>Variable TIMSS &amp; IGLU</b>
1 = 0 bis 10 Bücher	1 = 0 bis 10 Bücher	1 = 0 bis 10 Bücher
2 = 11 bis 25 Bücher	2 = 11 bis 25 Bücher	2 = 11 bis 25 Bücher
3 = 26 bis 100 Bücher	3 = 26 bis 100 Bücher	3 = 26 bis 100 Bücher
4 = 101 bis 200 Bücher	4 = 101 bis 200 Bücher	4 = 101 bis 200 Bücher
5 = 201 bis 500 Bücher 6 = mehr als 500 Bücher	5 = mehr als 200 Bücher	5 = mehr als 200 Bücher

## Anhang B: Equating-/Linking-Verfahren

Für die Auswahl eines geeigneten Linking-Verfahrens ist die Art der Datenerhebung relevant. Im Rahmen des Linking werden drei Formen der Datenerhebung unterschieden: 1.) single-group Design, 2.) equivalent-group Design und 3.) non-equivalent group Design mit gemeinsamen Items bzw. non-equivalent group Design mit Ankeritems (vgl. Davier, Holland & Thayer, 2004). Im single-group Design werden die alternativen Testformen derselben Gruppe von Personen vorgelegt, d.h. Unterschiede in der Schwierigkeit der Tests können nicht auf unterschiedlichen Personencharakteristika basieren. Im equivalent-group Design werden die parallelen Testformen verschiedenen Gruppen zugeteilt, wobei die Gruppenzugehörigkeit häufig per Zufall bestimmt wird, sodass in diesem Fall signifikante Unterschiede in der Schwierigkeit der Tests wahrscheinlich nicht auf unterschiedlichen Personencharakteristika basieren. Im non-equivalent group Design werden die parallelen Testformen von unterschiedlichen Gruppen bearbeitet, allerdings haben die parallelen Testformen zum Teil identische Items. Dieses Design wird beispielsweise in TIMSS und PIRLS verwendet und es würde auch gelten, wenn etwa TIMSS-Instrumente mit NEPS-Instrumenten verknüpft werden sollen. Deswegen wird im Folgenden auf Linking-Verfahren für dieses Design eingegangen.

Eine Möglichkeit, verschiedene Instrumente zu verlinken besteht durch die Verwendung von Modellen der Item Response Theorie (IRT) (vgl. Rost, 2006; Fischer, 1996). Mit diesen Modellen können generell Itemparameter und Personenparameter geschätzt werden. Die Itemparameter umfassen beispielsweise die Schwierigkeit eines Items oder dessen Diskriminationsfähigkeit, die Personenparameter sagen etwas über die Fähigkeit bzw. Kompetenz einer Person aus. Beim Linking werden zunächst die Itemparameter der Testformen geschätzt und diese auf einer gemeinsamen Skala abgebildet, danach werden die Personenparameter für die einzelnen Testformen bei gegebenen Itemparametern geschätzt. Da Itemparameter und Personenparameter in IRT-Modellen auf einer Skala abgebildet werden, sind die so ermittelten Personenparameter der verschiedenen Testformen auch auf einer Skala abgebildet.

Für die Schätzung der Itemparameter und die Abbildung dieser auf einer gemeinsamen Skala haben sich zwei Vorgehensweisen etabliert: Entweder die Itemparameter aller Testformen werden zusammen geschätzt (*concurrent calibration*) und befinden sich dadurch schon auf derselben Metrik, oder die Itemparameter der Testformen werden separat geschätzt und müssen nachträglich noch auf dieselbe Skala transformiert werden (vgl. Davier, Xu & Carstensen, 2011). Da sich *die concurrent calibration* als effizienter gegenüber der separaten Skalierung erwiesen hat (vgl. Tian, 2011), wurde sie in dieser Machbarkeitsstudie angewendet.

In der praktischen Anwendung der *concurrent calibration* werden die Datensätze der parallelen Testformen in einen Datensatz geschrieben, wobei die Daten der gemeinsamen Items untereinander und die Daten verschiedener Items nebeneinander geschrieben werden. Für die Items dieses kombinierten Datensatzes werden dann die Itemparameter geschätzt, wodurch die Itemparameter für die gemeinsamen Items zwischen den Testformen identisch und auf derselben Skala sind, und die übrigen Itemparameter auf derselben Skala liegen wie die Itemparameter der gemeinsamen Items, d.h. alle Itemparameter aller Testformen sind auf derselben Skala. Nach der Schätzung der Itemparameter werden dann die Personenparameter geschätzt.

Ergänzend muss noch gesagt werden, dass das hier beschriebene Verfahren zwar vor allem im Kontext von Leistungsdaten, d.h. Testformen, entwickelt wurde. Aus mathematisch-statistischer Sicht spricht aber nichts dagegen, dieses Verfahren auch zum verlinken von Fragebogenitems zu verwenden.

den (siehe etwa die Analysen im Rahmen von TIMSS und PIRLS; vgl. Martin, Mullis & Hooper, 2017; Martin, Mullis & Hooper, 2016; Martin & Mullis, 2012). So wurden in den nachfolgend berichteten Analysen der Machbarkeitsstudie etwa die Items zum sozialen Status der Schülerinnen und Schüler von TIMSS und PIRLS verlinkt, um einen Index für den sozialen Status der Schülerinnen und Schüler zu ermitteln.

### Anhang C: Rechnerische Umsetzung des Equatings

Insgesamt wurden drei Skalierungen mit dem PCM umgesetzt. In der ersten Skalierung wurde der soziale Index nur für den TIMSS 2015-Datensatz bestimmt, in der zweiten Skalierung wurde der soziale Index nur für den PIRLS 2016-Datensatz berechnet und in der dritten Skalierung wurde der soziale Index für den kombinierten TIMSS- und PIRLS-Datensatz bestimmt, d.h. in der dritten Skalierung wurde das Verfahren der *concurrent calibration* angewendet. Diese dritte Skalierung ist notwendig, um die nachfolgenden Ergebnisse basierend auf dem TIMSS-Datensatz mit den Ergebnissen basierend auf dem PIRLS-Datensatz vergleichbar zu machen (s. Abschnitt 4.5.3 und Abschnitt 5). Nach der Skalierung wurde der soziale Index auf eine Metrik mit einem Mittelwert von 100 und eine Standardabweichung von 10 transformiert.

Die deskriptiven Statistiken für den sozialen Index für TIMSS und PIRLS aller drei Skalierungsläufe finden sich in Tabelle 6. In TIMSS liegt der mittlere soziale Status bei 100 Punkten bei separater Skalierung und bei 99.4 Punkten bei gemeinsamer Skalierung mit dem PIRLS-Datensatz. Für PIRLS beträgt der mittlere soziale Status ebenfalls 100 Punkte (dieser Wert von 100 wie auch der Wert von 100 für TIMSS ist durch die vorgenommene Transformierung des sozialen Index nach Skalierung zu erwarten) und 100.4 Punkte bei gemeinsamer Skalierung mit dem TIMSS-Datensatz. Wie ein Vergleich des mittleren sozialen Status der Schülerinnen und Schüler in TIMSS und PIRLS zeigt, wiesen die Schülerinnen und Schüler in TIMSS mit 99.4 Punkten einen etwas geringeren mittleren sozialen Status auf als die Schülerinnen und Schüler in PIRLS mit 100.4 Punkten. Wie die dazugehörigen Konfidenzintervalle zeigen ist dieser Unterschied statistisch signifikant. Da nur der soziale Index, der mithilfe der *concurrent calibration* gewonnen wurde eine Vergleichbarkeit der nachfolgenden Ergebnisse zwischen TIMSS und PIRLS erlaubt, wurde für die Analysen nur noch dieser Index verwendet.

Tabelle 6: Deskriptive Statistiken für den sozialen Index für TIMSS 2015 und PIRLS 2016 Daten

Studie	Sozialer Index	Mittelwert	Standardfehler	95-% Konfidenzintervall	
				Untere Schranke	Obere Schranke
TIMSS 2015	Separate Skalierung	100.0	0.22	99.45	100.33
	Concurrent Calibration	99.4	0.22	98.92	99.80
PIRLS 2016	Separate Skalierung	100.0	0.28	99.31	100.44
	Concurrent Calibration	100.4	0.28	99.83	100.94