

# Untersuchung der Auswirkungen des G 8 auf die Leistungsentwicklung am Beispiel des Hamburger KESS-Jahrgangs

Stanislav Ivanov, Roumiana Nikolova, Ulrich Vieluf

Oktober 2014

In Hamburg startete die Umstellung vom neunstufigen Gymnasium (G 9) auf das achtstufige Gymnasium (G 8) im Schuljahr 2002/03 mit der Jahrgangsstufe 5, sodass der erste G 8-Jahrgang im Sommer 2010 das Abitur ablegte, gemeinsam mit dem letzten G 9-Jahrgang. Anders als in anderen Bundesländern (vgl. hierzu Klemm, 2013; Kühn, 2013) erfolgte diese Schulzeitverkürzung ohne Kürzung der Unterrichtszeit – das bisherige Unterrichtsvolumen von insgesamt 265 Wochenstunden in den Sekundarstufen I und II blieb unverändert: Die Wochenstundenzahl in der Sekundarstufe I (Jahrgangsstufen 5 bis 10) wurde um 19 auf 197 Wochenstunden erhöht, die Wochenstundenzahl in der Sekundarstufe II wurde um 19 auf 68 Wochenstunden reduziert. Die bisherige Vorstufe entfiel (bzw. wurde in die Jahrgangsstufe 10 vorgezogen), sodass die gymnasiale Oberstufe nur mehr die zweijährige Studienstufe umfasst. Parallel zu dieser strukturellen Veränderung wurde ein Bündel von Maßnahmen zur Qualitätssteigerung im Schulwesen eingeführt. Dazu gehören die Implementierung neuer Bildungspläne, die regelmäßige Durchführung von Lernstandserhebungen und die Einführung zentraler Abschlussprüfungen in ausgewählten Fächern. Es wurde die Profiloberstufe eingeführt, die Belegverpflichtungen in der gymnasialen Oberstufe wurden erhöht, indem die Kernfächer in vierstündigen Kursen auf grundlegendem bzw. erhöhtem Anforderungsniveau durchgehend zu belegen sind, die Wahlmöglichkeiten der Prüfungsfächer im Abitur wurden eingeschränkt.<sup>1</sup>

Der KESS-Jahrgang ist der zweite Hamburger G 8-Jahrgang. Da bei allen fünf Lernstandserhebungen, die im Rahmen der KESS-Studie durchgeführt wurden, u. a. Tests in ausgewählten Kompetenzbereichen eingesetzt wurden, die bereits in der sieben Jahre zuvor durchgeführten Studie „Aspekte der Lernausgangslage und der

---

<sup>1</sup> Anders als in anderen Bundesländern erfolgte diese Reform in Hamburg nicht vor dem Hintergrund sinkender Schülerzahlen, sondern wurde vom damaligen Senat im Wesentlichen damit begründet, dass deutsche Schülerinnen und Schüler im internationalen Vergleich überdurchschnittlich alt seien, wenn sie mit Studium und Ausbildung beginnen und in das Berufsleben eintreten (vgl. Bürgerschaft der Freien und Hansestadt Hamburg, 2003).

Lernentwicklung“ (LAU) verwendet worden waren (vgl. Lehmann et al., 1997, 1999, 2002, 2004, 2006; Trautwein et al., 2006), besteht die Möglichkeit, im direkten Vergleich beider Schülerjahrgänge mögliche Effekte der Schulzeitverkürzung zu überprüfen.

An beiden Studien nahm der komplette Schülerjahrgang teil, wenngleich infolge der Rotationsdesigns nicht alle Schülerinnen und Schüler dieselben Tests bearbeiteten. (vgl. hierzu Bos & Pietsch, 2006; Bos, Bonsen & Gröhlich, 2009; Bos & Gröhlich, 2010; Vieluf, Ivanov & Nikolova, 2011).

Im Folgenden beschränken wir uns auf die vergleichende Analyse der Lernausgangslagen der Gymnasiastinnen und Gymnasiasten beider Schülerjahrgänge zu Beginn der Jahrgangsstufe 7 und der von ihnen am Ende der gymnasialen Oberstufe in ausgewählten Kompetenzbereichen erreichten Lernstände. Dem liegt die Überlegung zugrunde, dass in beiden Schülerjahrgängen bei der Anmeldung für die weiterführende Schule ein unbeschränktes Elternwahlrecht galt, Eltern also auch gegen die Empfehlung der Zeugniskonferenz der abgebenden Grundschule ihr Kind ohne Prüfungsverfahren auf einem Gymnasium anmelden konnten. Erst am Ende der Jahrgangsstufe 6 entschieden die Zeugniskonferenzen der weiterführenden Schulen auf der Grundlage der Lernentwicklungen und der Leistungsstände in den Unterrichtsfächern über die Berechtigung zum Eintritt in die Jahrgangsstufe 7 des Gymnasiums. So hatten Abschlüssen, aber auch die Neuaufnahme von Schülerinnen und Schülern aus den anderen Schulformen in beiden Jahrgängen im Verlauf der Jahrgangsstufen 5 und 6 zu nicht unerheblichen Veränderungen der Schülerzusammensetzung geführt (Lehmann et al., 1999, Bos et al., 2009, Hillebrand, 2014).

Grundlage der nachfolgenden vergleichenden Analysen sind neben den erhobenen Leistungsdaten die aus Schülerakten sowie im Rahmen von Schüler- und Elternbefragungen gewonnenen Angaben zu einer Vielzahl von Hintergrundmerkmalen der Schülerschaften beider Jahrgänge. Dieser umfangreiche Datensatz macht es möglich, nicht allein die Lernstände eines vollständigen Schülerjahrgangs, der das Abitur unter den Rahmenbedingungen des G 8 erworben hat, mit den Lernständen eines vollständigen Schülerjahrgangs, der das Abitur unter den Rahmenbedingungen des G 9 erworben hat, in ausgewählten Kompetenzbereichen zu vergleichen, sondern darüber hinaus möglichen Zusammenhängen zwischen Leistungsunterschieden und Veränderungen hinsichtlich der soziodemografischen Zusammensetzung der Schülerschaften beider Jahrgänge nachzugehen. Hieraus leitet sich

folgendes Vorgehen ab: Zunächst werden die zentralen Fragestellungen, denen hier nachgegangen wird. Anschließend werden die Datengrundlage und das methodische Vorgehen beschrieben. Im darauffolgenden Abschnitt werden die Ergebnisse der Datenanalysen vorgestellt, die im letzten Abschnitt diskutiert werden

## **Fragestellungen**

Anhand des Kohortenvergleichs sollen die folgenden Fragestellungen beantwortet werden:

1. Inwieweit unterscheiden sich die in den Kompetenzbereichen Englisch, Mathematik und Naturwissenschaften am Ende der gymnasialen Oberstufe erreichten Lernstände zwischen dem KESS- und dem LAU-Jahrgang?

2. Falls ja: Inwieweit können diese Unterschiede auf die Schulzeitverkürzung (einschließlich aller Begleitmaßnahmen) zurückgeführt werden?

3. Treffen die über alle Schülerinnen und Schüler ermittelten Befunde hinsichtlich des Zusammenhangs zwischen Schulzeitverkürzung und erreichten Lernständen gleichermaßen für Schülerinnen und Schüler mit hohen bzw. niedrigen Lernausgangslagen zu Beginn der Jahrgangsstufe 7 zu?

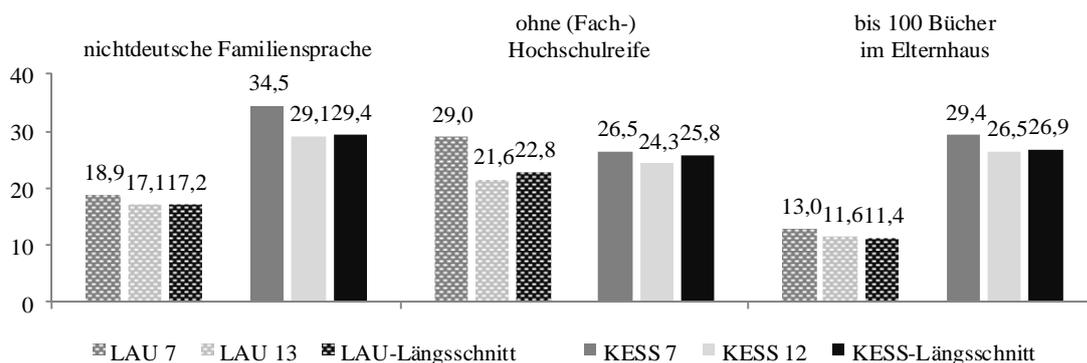
Während sich die erste Fragestellung auf deskriptiver Ebene beantworten lässt, setzt die Beantwortung der beiden anderen Fragestellungen eine quasi-experimentelle Versuchsanordnung voraus. Dabei wird der KESS-Jahrgang als Versuchsgruppe und der LAU-Jahrgang als Kontrollgruppe hinsichtlich möglicher Auswirkungen der Schulzeitverkürzung betrachtet. Da die beiden Vergleichsgruppen nicht randomisierte, sondern natürliche, bereits existierende Schülergruppen sind, können Leistungsunterschiede am Ende der gymnasialen Oberstufe nicht ohne Weiteres auf die Schulzeitverkürzung zurückgeführt werden. Vielmehr ist es notwendig, beide Gruppen bezüglich personenbedingter leistungsrelevanter Hintergrundvariablen, wie die individuellen Lernausgangslagen zu Beginn der Jahrgangsstufe 7 (Vorwissen) und der soziokulturelle und ethnische Hintergrund, ausreichend zu parallelisieren. Die längsschnittlich angelegten LAU- und KESS-Studien stellen dafür eine Fülle schülerbezogener Daten zur Verfügung.

## Datengrundlage und methodisches Vorgehen

### Datengrundlage

4.953 von insgesamt 13.219 Schülerinnen und Schülern des LAU-Jahrgangs<sup>2</sup> besuchten zu Beginn der Jahrgangsstufe 7 ein grundständiges Gymnasium, das entspricht 37,5 Prozent. Im KESS-Jahrgang waren es 5.815 von 14.200 Schülerinnen und Schülern, das entspricht 41,0 Prozent. Die Zahl der Abiturientinnen und Abiturienten an den grundständigen Gymnasien ist von 3.517 im LAU-Jahrgang auf 4.675 im KESS-Jahrgang bzw. um 33 Prozent gestiegen.<sup>3</sup> Dieser Anstieg hat zu einer veränderten soziodemografischen bzw. -kulturellen Zusammensetzung der Schülerschaft geführt, wie die in Abbildung 1 für ausgewählte Hintergrundmerkmale wiedergegebenen prozentualen Anteile erkennen lassen.

Abb. 1: LAU – KESS: Prozentuale Unterschiede des soziokulturellen Hintergrundes beider Kohorten (grundständige Gymnasien)



So ist der Anteil der Schülerinnen und Schüler mit nichtdeutscher Familiensprache zwischen LAU 7 und KESS 7 von 18,9 auf 34,5 Prozent gestiegen. Diese Differenz hat sich zwar infolge des überproportionalen Ausscheidens dieser Schülergruppe im KESS-Jahrgang von 15,6 auf 12,0 Prozent (alle Abiturientinnen und Abiturienten des Jahrgangs) bzw. 12,2 Prozent (nur die Schülerinnen und Schüler, die an beiden Erhebungen teilgenommen haben) bis zum Ende der gymnasialen Oberstufe verringert, ist aber immer noch beträchtlich. Deutlich geringer sind die Unterschiede

<sup>2</sup> Ohne Sonderschulen.

<sup>3</sup> Im LAU-Jahrgang betrug der Rückgang der Schülerzahl an den grundständigen Gymnasien zwischen den Jahrgangsstufen 7 und 13 1.436, das entspricht 29,0 Prozent, im KESS-Jahrgang waren es 1.140 bzw. 19,6 Prozent. Dieser Rückgang ist nur teilweise auf ein vorzeitiges Ausscheiden bzw. einen Abbruch des Bildungsweges zum Abitur zurückzuführen; eine nennenswerte Teilmenge der Schülerinnen und Schüler hat im Verlauf oder am Ende der Mittelstufe die Schulform gewechselt und den Bildungsgang auf einer Gesamtschule, einem Aufbaugymnasium oder einem Beruflichen Gymnasium fortgesetzt. Exakte Angaben zu den Bildungsverläufen liegen allerdings nicht vor.

zwischen beiden Kohorten hinsichtlich des Anteils an Schülerinnen und Schülern, deren beide Eltern höchstens den Mittleren Schulabschluss erworben haben. Er betrug im LAU-Jahrgang zu Beginn der Jahrgangsstufe 7 29,0 Prozent, im KESS-Jahrgang 26,5 Prozent. Infolge des deutlich überproportionalen Ausscheidens dieser Schülergruppe im LAU-Jahrgang liegt ihr Anteil am Ende der gymnasialen Oberstufe im KESS-Jahrgang um 2,7 Prozent (alle Abiturientinnen und Abiturienten des Jahrgangs) bzw. 3,0 Prozent (nur die Längsschnittschülerinnen und -schüler) über deren Anteil im LAU-Jahrgang. Schließlich ist der Anteil an Schülerinnen und Schülern mit einem Buchbestand im Elternhaus von maximal 100 Büchern zwischen LAU 7 und KESS 7 um 16,4 Prozent gestiegen, am Ende der gymnasialen Oberstufe betrug die Differenz zwischen beiden Jahrgängen infolge des leicht überproportionalen Ausscheidens dieser Schülergruppe im KESS-Jahrgang 14,9 Prozent (alle Abiturientinnen und Abiturienten des Jahrgangs) bzw. 15,5 Prozent (nur die Längsschnittschülerinnen und -schüler).

Betrachtet man die leistungsbezogenen Lernvoraussetzungen der Schülerinnen und Schüler beider Jahrgänge zu Beginn der Mittelstufe, so ergibt sich für den KESS-Jahrgang – ungeachtet seiner ungünstigeren soziodemografischen und -kulturellen Schülerzusammensetzung – mit einer Effektstärke von  $d = 0,12$  ein leichter Vorsprung im Kompetenzbereich *Mathematische Grundbildung*, in den Kompetenzbereichen *Deutsch-Leseverständnis* und *Englisch-Sprachverständnis* liegen beide Kohorten (annähernd) gleichauf ( $d \leq 0,01$ ).

In beiden Schülerjahrgängen wurden die Tests *Deutsch – Leseverständnis* und *Mathematische Grundbildung* von allen Schülerinnen und Schülern der Jahrgangsstufe 7 bearbeitet, während der Test *Englisch – Allgemeines Sprachverständnis* nur im LAU-Jahrgang von allen Schülerinnen und Schülern, im KESS-Jahrgang hingegen von einer Teilstichprobe (2.863 der insgesamt 5.815 Gymnasiastinnen und Gymnasiasten, das entspricht 49,2 Prozent) bearbeitet wurde. Die andere Hälfte der Schülerschaft des KESS-Jahrgangs bearbeitete den Test *Naturwissenschaftliche Grundbildung*, der im LAU-Jahrgang nicht in der Jahrgangsstufe 7 eingesetzt worden war. Der Test *Deutsch – Leseverständnis* wurde nur zu Beginn der Mittelstufe eingesetzt. Die Tests *Englisch – Allgemeines Sprachverständnis* und *Mathematische Grundbildung* wurden am Ende der gymnasialen Oberstufe von allen Schülerinnen und Schülern beider Schülerjahrgänge bearbeitet. Außerdem wurden in beiden Schülerjahrgängen am Ende der gymnasialen Oberstufe der *Test of English as a*

*Foreign Language* (TOEFL) sowie die Tests *Voruniversitäre Mathematik* und *Naturwissenschaftliche Grundbildung* in Teilstichproben durchgeführt (vgl. Tabelle 1).

Tab. 1: Übersicht über die in LAU und in KESS zu Beginn der Jahrgangsstufe 7 und am Ende der gymnasialen Oberstufe eingesetzten Tests

Test	Jahrgangsstufe 7		Jahrgangsstufe 12/13	
	LAU	KESS	LAU	KESS
Deutsch – Leseverständnis	Vollerhebung	Vollerhebung	---	---
Englisch – Allgemeines Sprachverständnis	Vollerhebung	Stichprobe	Vollerhebung	Vollerhebung
TOEFL – Gesamttest	---	---	Stichprobe	Stichprobe
TOEFL – Structure & Written Expression	---	---	Vollerhebung	Vollerhebung
TOEFL – Vocabulary & Reading Comprehension	---	---	Stichprobe	Stichprobe
TOEFL – Listening Comprehension	---	---	Stichprobe	Stichprobe
Mathematische Grundbildung	Vollerhebung	Vollerhebung	Vollerhebung	Vollerhebung
Voruniversitäre Mathematik	---	---	Stichprobe	Stichprobe
Naturwissenschaftliche Grundbildung	---	---	Stichprobe	Stichprobe

Von den 3.517 Abiturientinnen und Abiturienten des LAU-Jahrgangs haben 3.347 bzw. 95 Prozent an der Erhebung teilgenommen, im KESS-Jahrgang waren es 3.822 von 4.675 Abiturientinnen und Abiturienten, das entspricht einer Teilnahmequote von 82 Prozent. Die erheblich geringere Teilnahmequote im KESS-Jahrgang geht in erster Linie auf organisatorische Engpässe zurück, die durch eine veränderte Gestaltung des Zeitraums zwischen schriftlichem und mündlichem Abitur bedingt waren, hat aber nicht zu einer systematischen Verzerrung der Datenbasis geführt. So betrug die Abiturdurchschnittsnote für die Teilnehmerinnen und Teilnehmer 2,35, für die Nichtteilnehmerinnen und Nichtteilnehmer 2,27.

Um weiteren möglichen Verzerrungen durch Datenausfall aufgrund von Testrotationen, unvollständiger Testbearbeitung oder Nichtteilnahme an einzelnen Tests zu begegnen, wurden fehlende Werte in beiden Studien und zu beiden Messzeitpunkten nach dem Verfahren der *Multiplen Imputation* (Rubin, 1987; Schafer, 1997) mit der Software *Mplus* (Muthén, L. & Muthén, B., 2010) anhand fünf vollständiger (imputierter) Datensätze geschätzt und ergänzt. Dabei wurden auf der Individual-ebene leistungsrelevante soziodemografische und -kulturelle Hintergrundinformationen wie *Familiensprache*, *Geburtsland der Eltern*, *höchster Schulabschluss der Eltern* sowie *Buchbestand im Elternhaus*, auf der Schulebene die Schulmittelwerte in den einzelnen Tests berücksichtigt.

Mit Blick auf die übergeordnete Fragestellung werden in den nachfolgenden Analysen ausschließlich die Testergebnisse derjenigen Schülerinnen und Schüler berücksichtigt, die sowohl zu Beginn der Mittelstufe (Jahrgangsstufe 7) als auch am Ende der gymnasialen Oberstufe (Jahrgangsstufe 13 bzw. 12) an der LAU- bzw. an der KESS-Untersuchung teilgenommen haben, von denen also mindestens aus einem der Leistungstests am Ende der gymnasialen Oberstufe Daten vorliegen. Die empirische Basis der nachfolgenden Analysen bilden die Ergebnisse von insgesamt 2.518 Abiturientinnen und Abiturienten des LAU-Jahrgangs und 2.974 Abiturientinnen und Abiturienten des KESS-Jahrgangs<sup>4</sup> in den folgenden Kompetenztests: dem Test *Englisch – Allgemeines Sprachverständnis*, dem *Test of English as a Foreign Language* (TOEFL-Gesamtskala) mit den Untertests *Structure & Written Expression*, *Vocabulary & Reading Comprehension* und *Listening Comprehension*, den beiden Mathematiktests *Mathematische Grundbildung* und *Voruniversitäre Mathematik* sowie dem Test *Naturwissenschaftliche Grundbildung*. Für dieselben Schülerinnen und Schüler standen darüber hinaus die Ergebnisse in den Tests *Deutsch – Leseverständnis*, *Englisch – Allgemeines Sprachverständnis* und *Mathematische Grundbildung* zur Verfügung, die zu Beginn der Mittelstufe (Jahrgangsstufe 7) in LAU 7 bzw. in KESS 7 erhoben worden waren.<sup>5</sup>

---

<sup>4</sup> Die LAU 13-Erhebung fand im Frühjahr 2005, die KESS 12-Erhebung im Frühjahr 2011 statt, jeweils zwischen der schriftlichen und der mündlichen Abiturprüfung.

<sup>5</sup> Detaillierte Beschreibungen der eingesetzten Tests finden sich in Bos et al. (2009), Trautwein et al. (2010) sowie in Lehmann et. al. (2006, 2011).

## Methodisches Vorgehen

Zur Beantwortung der ersten Fragestellung wurden die Testergebnisse der Abiturientinnen und Abiturienten beider Jahrgänge unter Anwendung des einparametrischen Rasch-Modells<sup>6</sup> einer simultanen Skalierung mit der Software *ConQuest*<sup>7</sup> (Wu et al., 1998) unterzogen und auf acht gemeinsamen Skalen verortet. Anschließend wurden alle Personenparameter auf eine Skalenmetrik mit dem Mittelwert  $M = 500$  und der Standardabweichung  $SD = 100$  linear transformiert. Da die KESS 12-Untersuchung als Replikation der LAU 13-Untersuchung konzipiert und mit nahezu unveränderten Testheften durchgeführt worden war, basierten die Skalierung und der Leistungsvergleich beider Jahrgänge am Ende der gymnasialen Oberstufe ausschließlich auf identischen Items, die in derselben Reihenfolge dargeboten wurden.<sup>8</sup>

Zur Beantwortung der zweiten Fragestellung wurden in einem ersten Schritt die Ergebnisse der Abiturientinnen und Abiturienten des KESS-Jahrgangs in den Tests *Deutsch – Leseverständnis*, *Englisch – Allgemeines Sprachverständnis* und *Mathematische Grundbildung* zu Beginn der Mittelstufe (KESS 7) anhand von 16, 48 bzw. 20 Items, die auch in der LAU 7- Untersuchung eingesetzt worden waren, in einem Verankerungsverfahren auf die Skalenmetrik der jeweiligen LAU-Skalen transformiert.<sup>9</sup>

Die drei für die beiden Jahrgänge nun gemeinsamen Skalen wurden im nächsten Schritt zu einem latenten Faktor *Allgemeiner Fachleistungsindex zu Beginn der Jahrgangsstufe 7 (AFI 7)* zusammengefasst mit dem Ziel, die individuellen leistungsbezogenen Lernvoraussetzungen der Gymnasiastinnen und Gymnasiasten beider Schülerjahrgänge zu Beginn der Mittelstufe abzubilden. Anschließend wurden der *AFI 7*, die Hintergrundmerkmale *Geschlecht*, *Familiensprache*, *Geburtsland der Eltern*, *höchster Schulabschluss der Eltern* und *Buchbestand im Elternhaus* sowie die Kohortenzugehörigkeitsvariable *LAU/KESS* als Prädiktoren (unabhängige Variablen) für die am Ende der gymnasialen Oberstufe erreichten Lernstände (abhängige

---

<sup>6</sup> Die psychometrischen Vorzüge des Rasch-Modells und die Methode der Rasch-Skalierung sind u. a. bei Baumert, Bos & Lehmann (2000, S. 60 ff.) anschaulich erläutert.

<sup>7</sup> In diesem Programm werden die Item- und Personenparameter nach der Marginal-Maximum-Likelihood-(MML)-Methode geschätzt.

<sup>8</sup> Booklet- sowie Itempositionseffekte konnten auf diese Weise weitgehend ausgeschlossen werden.

<sup>9</sup> Die technischen Schritte bei diesem Verfahren bestanden u. a. darin, die Schwierigkeitsparameter der identischen Testaufgaben (Ankeritems) auf ihre Werte in der LAU-Studie zu fixieren. Das Verfahren des Fixierens von Ankeritems ist u.a. bei Davier & Davier (2007) beschrieben.

Variablen) in acht testbezogenen Regressionsmodellen untersucht. Zur Beantwortung der dritten Fragestellung wurden die gleichen Analysen für die 500 Testbesten (obere Leistungsgruppe – oLG) sowie für die 500 Testschwächsten (untere Leistungsgruppe – uLG) beider Jahrgänge zu Beginn der Mittelstufe in den beiden längsschnittlichen Tests *Englisch – Allgemeines Sprachverständnis* und *Mathematische Grundbildung* durchgeführt.<sup>10</sup>

Von besonderem Interesse sowohl für die zweite als auch für die dritte Fragestellung ist der *LAU/KESS-Prädiktor*, der im Kontext der übrigen Prädiktoren als Effekt der Schulzeitverkürzung (einschließlich der begleitenden pädagogischen Maßnahmen) betrachtet werden kann. Ein signifikanter eigenständiger Beitrag dieses Prädiktors (signifikantes standardisiertes  $\beta$ -Gewicht) kann als Hinweis darauf interpretiert werden, dass Unterschiede in den am Ende der gymnasialen Oberstufe erreichten Lernständen zwischen dem LAU- und dem KESS-Jahrgang nicht allein durch (Kohorten-)Unterschiede in den leistungsbezogenen sowie soziodemografischen und -kulturellen Lernausgangslagen zu Beginn der Mittelstufe erklärt, sondern darüber hinaus in einen Zusammenhang mit der Schulzeitverkürzung (einschließlich aller Begleitmaßnahmen) gebracht werden können.

## **Ergebnisse**

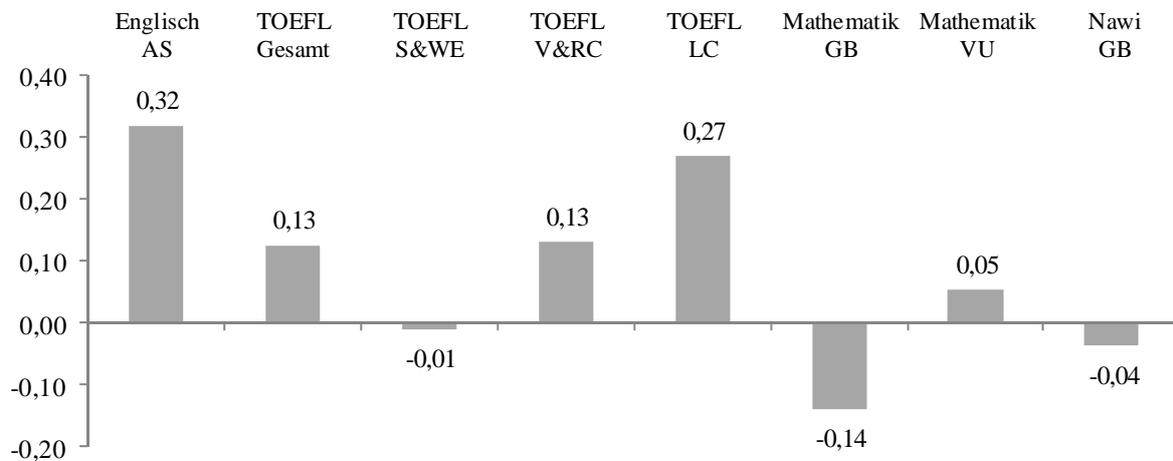
### **Deskriptive Befunde**

Für sechs der acht Kompetenzskalen ergeben sich signifikante Unterschiede zwischen den Lernständen der Abiturientinnen und Abiturienten des LAU- und des KESS-Jahrgangs am Ende der gymnasialen Oberstufe. Lediglich im TOEFL-Untertest *Structure & Written Expression* und im Test *Naturwissenschaftliche Grundbildung* sind die Mittelwertunterschiede nicht signifikant (vgl. Abb. 2).

---

<sup>10</sup> Aufgrund der deutlich unterschiedlichen Verteilung der Schülerinnen und Schüler auf die Kursniveaus wurden die 500 Testbesten (obere Leistungsgruppe – oLG) und die 500 schwächsten Testteilnehmerinnen und -teilnehmer (untere Leistungsgruppe – uLG) beider Jahrgänge betrachtet. So hatten im LAU-Jahrgang im Fach Englisch 60 Prozent einen Grundkurs und 33 Prozent einen Leistungskurs belegt, 7 Prozent hatten eine andere Fremdsprache gewählt. Im Fach Mathematik waren es 88 Prozent, die einen Grundkurs belegt hatten, 12 Prozent hatten sich für einen Leistungskurs entschieden. Demgegenüber waren es im KESS-Jahrgang im Fach Englisch 21 Prozent, die einen Kurs auf grundlegendem Anforderungsniveau belegt hatten, 77 Prozent hatten sich für einen Kurs auf erhöhtem Anforderungsniveau entschieden, 2 Prozent hatten eine andere Fremdsprache gewählt. Im Fach Mathematik waren es 58 Prozent, die einen Kurs auf grundlegendem Anforderungsniveau belegt hatten, 42 Prozent hatten sich für einen Kurs auf erhöhtem Anforderungsniveau entschieden.

Abb. 2: Lernstände am Ende der gymnasialen Oberstufe im Vergleich: Mittelwertdifferenzen in Effektstärken (Referenz: LAU 13)



Englisch AS = *Allgemeines Sprachverständnis*, TOEFL Gesamt = *Test of English as a Foreign Language – Gesamtttest*, TOEFL S&WE = *Untertest Structure & Written Expression*, TOEFL V&RC = *Untertest Vocabulary & Reading Comprehension*, TOEFL LC = *Untertest Listening Comprehension*, Mathematik GB = *Mathematische Grundbildung*, Mathematik VU = *Voruniversitäre Mathematik*, Nawi GB = *Naturwissenschaftliche Grundbildung*

Die Abiturientinnen und Abiturienten des KESS-Jahrgangs verzeichnen in dem Test *Englisch – Allgemeines Sprachverständnis* mit  $d = 0,32$  und in dem *TOEFL-Untertest Listening Comprehension* mit  $d = 0,27$  jeweils einen deutlichen Leistungsvorsprung vor den Abiturientinnen und Abiturienten des LAU-Jahrgangs. Im TOEFL-Gesamtttest sowie im TOEFL-Untertest *Vocabulary & Reading Comprehension* ergibt sich mit  $d = 0,13$  jeweils ein etwas höherer mittlerer Lernstand für den KESS-Jahrgang, auch im Test *Voruniversitäre Mathematik* verzeichnet der KESS-Jahrgang mit  $d = 0,05$  einen leichten Vorsprung. Im TOEFL-Untertest *Structure & Written Expression* sowie im Test *Naturwissenschaftliche Grundbildung* besteht kein signifikanter Unterschied zwischen den Abiturientinnen und Abiturienten beider Jahrgänge. Demgegenüber bleiben die Abiturientinnen und Abiturienten des KESS-Jahrgangs im Test *Mathematische Grundbildung* mit  $d = -0,14$  signifikant unter dem mittleren Lernstand der Abiturientinnen und Abiturienten des LAU-Jahrgangs. Diese Befunde belegen, dass die Schulzeitverkürzung einschließlich ihrer Begleitmaßnahmen unter den spezifischen Hamburger Rahmenbedingungen zumindest in den beiden betrachteten Jahrgängen nicht mit generellen Leistungseinbußen einhergegangen ist, sondern – wie im Fach Englisch – sogar eine deutliche Leistungsverbesserung beobachtet werden kann.

## **Befunde der Regressionsanalysen**

Um zu überprüfen, inwieweit die festgestellten Leistungsunterschiede auf die Schulzeitdauer (LAU-Jahrgang: dreizehnjährig, KESS-Jahrgang: zwölfjährig) zurückgeführt werden können, wurden in einem weiteren Schritt testbezogene Regressionsanalysen durchgeführt. Im Ergebnis zeigt sich, dass der *LAU/KESS*-Prädiktor für die sechs der acht Kompetenzskalen, für die sich bereits signifikante Unterschiede hinsichtlich der mittleren Lernstände beider Kohorten ergeben hatten, auch nach Kontrolle der soziodemografischen und -kulturellen sowie der leistungsbezogenen Lernvoraussetzungen ein signifikantes  $\beta$ -Gewicht aufweist (vgl. Tabelle 2). Dieser Befund weist darauf hin, dass der Indikator für die Schulzeitverkürzung (einschließlich ihrer Begleitmaßnahmen) eigenständig zur Erklärung der unterschiedlichen Lernstände beider Kohorten beiträgt, und zwar in dem Test *Englisch – Allgemeines Sprachverständnis* (Englisch-AS), im TOEFL-Gesamttest (TOEFL-ges) und den beiden Untertests *Vocabulary & Reading Comprehension* (TOEFL-V&RC) und *Listening Comprehension* (TOEFL-LC) sowie in beiden Mathematiktests.

Tab. 2: Prädiktoren der Lernstände in Englisch, Mathematik und Naturwissenschaften am Ende der Sekundarstufe II: standardisierte  $\beta$ -Gewichte und Standardfehler (in Klammern)<sup>11</sup>

	AFI 7	Jahrgang LAU = 0	Geschlecht w = 0	Familien- sprache nicht D = 0	Geburts- land Eltern ABI = 0	Schulbildung g Eltern ohne ABI = 0	Bücher im Elternhaus <100 = 0
$\beta$ (SE)							
Englisch- AS	<b>0,46**</b> (0,02)	<b>0,17**</b> (0,01)	-0,01 (0,01)	<b>0,04**</b> (0,02)	0,03 (0,02)	<b>0,06**</b> (0,01)	<b>0,05**</b> (0,01)
Englisch- AS oLG	<b>0,43**</b> (0,04)	<b>0,10**</b> (0,03)	0,00 (0,03)	0,03 (0,04)	0,01 (0,04)	<b>0,07*</b> (0,03)	0,05 (0,03)
Englisch- AS uLG	<b>0,11**</b> (0,04)	<b>0,31**</b> (0,03)	<b>0,07**</b> (0,03)	<b>0,15*</b> (0,05)	-0,02 (0,05)	<b>0,07*</b> (0,03)	<b>0,07*</b> (0,03)
TOEFL- S&WE	<b>0,46**</b> (0,02)	0,01 (0,01)	<b>0,07**</b> (0,01)	-0,01 (0,02)	-0,03 (0,02)	<b>0,11**</b> (0,01)	<b>0,09**</b> (0,01)
TOEFL- V&RC	<b>0,34**</b> (0,02)	<b>0,09**</b> (0,01)	<b>0,11**</b> (0,01)	<b>0,07**</b> (0,02)	<b>-0,07**</b> (0,02)	<b>0,07**</b> (0,01)	<b>0,11**</b> (0,01)
TOEFL- LC	<b>0,30**</b> (0,02)	<b>0,17**</b> (0,01)	<b>0,05**</b> (0,01)	0,02 (0,02)	<b>-0,06**</b> (0,02)	<b>0,15**</b> (0,01)	<b>0,16**</b> (0,01)
TOEFL- ges	<b>0,34**</b> (0,02)	<b>0,10**</b> (0,01)	<b>0,08**</b> (0,01)	<b>0,07**</b> (0,02)	<b>-0,05**</b> (0,02)	<b>0,13**</b> (0,01)	<b>0,16**</b> (0,01)
Mathe-GB	<b>0,46**</b> (0,01)	<b>-0,07**</b> (0,01)	<b>0,17**</b> (0,01)	0,02 (0,02)	0,00 (0,02)	<b>0,05**</b> (0,01)	<b>0,04**</b> (0,01)
Mathe-GB oLG	<b>0,35**</b> (0,05)	0,04 (0,03)	<b>0,14**</b> (0,03)	0,07 (0,04)	-0,04 (0,04)	<b>0,09**</b> (0,03)	0,01 (0,03)
Mathe-GB uLG	-0,06 (0,04)	<b>-0,10**</b> (0,03)	<b>0,14**</b> (0,03)	0,08 (0,05)	0,00 (0,05)	<b>0,07*</b> (0,03)	0,03 (0,04)
Mathe- VU	<b>0,25**</b> (0,02)	<b>0,05**</b> (0,01)	<b>0,20**</b> (0,01)	0,03 (0,02)	0,03 (0,02)	<b>0,04**</b> (0,01)	<b>0,07**</b> (0,01)
Nawi- GB	<b>0,11**</b> (0,02)	0,02 (0,01)	<b>0,21**</b> (0,01)	<b>0,09**</b> (0,02)	0,02 (0,02)	<b>0,04**</b> (0,01)	<b>0,12**</b> (0,01)

\*  $p \leq 0,05$ ; \*\*  $p \leq 0,01$ ; RMSEA = 0,09; SRMR = 0,06;

R<sup>2</sup>: Englisch-AS = 0,24, TOEFL-S&WE = 0,24, TOEFL-V&RC = 0,11, TOEFL-LC = 0,16, TOEFL-ges = 0,18, Mathe-GB = 0,25, Nawi-GB = 0,09

Um diesen Effekt gegen eine Konfundierung mit der nur für den KESS-Jahrgang relevanten Oberstufenreform (u.a. Einführung der Profiloberstufe) abzusichern, wurden für die beiden längsschnittlich erfassten Kompetenzdomänen *Englisch* –

<sup>11</sup> Die Merkmale *LAU-Jahrgang*, *weibliches Geschlecht*, *Familiensprache nicht Deutsch*, *mindestens ein Elternteil im Ausland geboren*, *beide Eltern ohne (Fach-)Hochschulreife* und *Buchbestand im Elternhaus maximal 100 Bücher* dienen hier als Referenzgrößen und wurden mit 0 kodiert. Ein positives  $\beta$ -Gewicht für den LAU/KESS-Prädiktor deutet auf einen Leistungsvorsprung der Abiturientinnen und Abiturienten des KESS-Jahrgangs, ein negatives  $\beta$ -Gewicht auf einen Leistungsvorsprung der Abiturientinnen und Abiturienten des LAU-Jahrgangs hin.

*Allgemeines Sprachverständnis* und *Mathematische Grundbildung* die oben beschriebenen Regressionsanalysen wiederholt, diesmal wurden aber nicht die Lernstände am Ende der gymnasialen Oberstufe (Jahrgangsstufe 13 bzw. 12), sondern die Lernstände zu Beginn der gymnasialen Oberstufe (Jahrgangsstufe 11) als abhängige Variablen eingesetzt. Die Berechnungen haben ergeben, dass unter Berücksichtigung der soziodemografischen und -kulturellen sowie der leistungsbezogenen Lernvoraussetzungen zu Beginn der Jahrgangsstufe 7 der Indikator für die Schulzeitverkürzung bereits zu Beginn der gymnasialen Oberstufe für den Test *Englisch – Allgemeines Sprachverständnis* einen hochsignifikanten eigenständigen Beitrag zugunsten des KESS-Jahrgangs aufweist ( $\beta=0,32$ ). Einen gleichfalls hochsignifikanten, mit  $\beta=0,07$  jedoch deutlich schwächer ausgeprägten eigenständigen Beitrag verzeichnet der LAU/KESS-Prädiktor für den Test *Mathematische Grundbildung*, wiederum zugunsten des KESS-Jahrgangs.

Der Tabelle 2 ist des Weiteren zu entnehmen, dass in nahezu allen untersuchten Testbereichen die leistungsbezogenen Lernausgangslagen der Schülerinnen und Schüler zu Beginn der Mittelstufe (*AFI 7*) erwartungsgemäß den größten eigenständigen Beitrag leisten. Dieser fällt im *TOEFL-Gesamttest (TOEFL-ges)* wie auch in den drei Untertests *Structure & Written Expression (TOEFL-S&WE)*, *Vocabulary & Reading Comprehension (TOEFL-V&RC)* und *Listening Comprehension (TOEFL-LC)*, im Test *Englisch – Allgemeines Sprachverständnis (Englisch-AS)* und im Test *Mathematische Grundbildung (Mathe-GB)* stärker aus als im Test *Voruniversitäre Mathematik (Mathe-VU)*. Das geringste  $\beta$ -Gewicht hat der *AFI 7* für den Test *Naturwissenschaftliche Grundbildung (Nawi-GB)*.<sup>12</sup>

Für alle Testbereiche sind die zu Beginn der Mittelstufe erfassten Merkmale *höchster Schulabschluss der Eltern* und *Buchbestand im Elternhaus* signifikante Prädiktoren der am Ende der gymnasialen Oberstufe erreichten Lernstände. Diese fallen für die Abiturientinnen und Abiturienten aus eher bildungsnahen Familien im Durchschnitt höher aus als für die Abiturientinnen und Abiturienten aus eher bildungsfernen Elternhäusern. Besonders hoch ist das relative Gewicht der beiden Prädiktoren im *TOEFL-Gesamttest* und im *TOEFL-Untertest Listening Comprehension*.

---

<sup>12</sup> Dies ist insofern modellkonform, als die naturwissenschaftlichen Vorkenntnisse der Siebtklässlerinnen und Siebtklässler nicht im *AFI 7* berücksichtigt werden konnten, da dieser Kompetenzbereich in *LAU 7* nicht erfasst worden war.

Auch der Prädiktor *Geschlecht* liefert mit Ausnahme des Kompetenzbereichs *Englisch – Allgemeines Sprachverständnis* einen eigenständigen Beitrag zur Erklärung der Leistungsunterschiede am Ende der gymnasialen Oberstufe. Dieser ist in den mathematisch-naturwissenschaftlichen Tests annähernd doppelt so hoch (zugunsten der Abiturienten) wie im TOEFL (ebenfalls zugunsten der Abiturienten).

Die Familiensprache weist unter Kontrolle aller anderen Prädiktoren einen engen Zusammenhang mit den Ergebnissen im Test *Englisch – Allgemeines Sprachverständnis*, im TOEFL-Gesamttest, im TOEFL-Untertest *Listening Comprehension* und im Test *Naturwissenschaftliche Grundbildung* auf. Dabei verzeichnen die Abiturientinnen und Abiturienten mit der Familiensprache Deutsch im Durchschnitt höhere Lernstände als die Abiturientinnen und Abiturienten mit einer nichtdeutschen Familiensprache. Demgegenüber lässt sich für die beiden TOEFL-Untertests *Structure & Written Expression* und *Vocabulary & Reading Comprehension* sowie für beide Mathematiktests ein solcher Zusammenhang nicht nachweisen.

Nach Berücksichtigung aller anderen Prädiktoren hat das *Geburtsland der Eltern* nur im TOEFL-Gesamttest sowie in den beiden TOEFL-Untertests *Vocabulary & Reading Comprehension* und *Listening Comprehension* eine eigenständige Erklärungskraft, wobei die Lernstände der Schülerinnen und Schüler mit mindestens einem im Ausland geborenen Elternteil im Durchschnitt etwas höher liegen als die der Abiturientinnen und Abiturienten, deren beide Elternteile in Deutschland geboren sind.

Differenziert nach Leistungsgruppen ergibt sich, dass der *LAU/KESS*-Prädiktor sowohl für die Schülerinnen und Schüler aus der oberen Leistungsgruppe (oLG) wie auch für die Schülerinnen und Schüler aus der unteren Leistungsgruppen (uLG) im Test *Englisch – Allgemeines Sprachverständnis* ein signifikantes positives  $\beta$ -Gewicht aufweist. Dabei ist dieser Effekt für die Schülerinnen und Schüler aus der unteren Leistungsgruppe dreimal so hoch wie für die Schülerinnen und Schüler aus der oberen Leistungsgruppe. Demgegenüber ist im Test *Mathematische Grundbildung* ein signifikanter negativer Zusammenhang zwischen den am Ende der gymnasialen Oberstufe erreichten Lernständen und dem *LAU/KESS*-Prädiktor für die Schülerinnen und Schüler der unteren Leistungsgruppe nachweisbar, nicht aber für die Schülerinnen und Schüler der oberen Leistungsgruppe.

## Diskussion

Verglichen wurden die Lernstände zweier Schülerjahrgänge am Ende der gymnasialen Oberstufe anhand identischer Leistungstests in den Kompetenzbereichen Englisch, Mathematik sowie Naturwissenschaften und unter Berücksichtigung der soziodemografischen und -kulturellen sowie leistungsbezogenen Lernvoraussetzungen zu Beginn der Jahrgangsstufe 7. Beide Schülerjahrgänge hatten zwar dasselbe Unterrichtsvolumen (265 Wochenstunden in den Sekundarstufen I und II), jedoch hatten die Schülerinnen und Schüler des LAU-Jahrgangs das Abitur nach dreizehn Jahren (G 9) abgelegt, die Schülerinnen und Schüler des KESS-Jahrgangs hingegen nach zwölf Jahren (G 8). Daher richtete sich der Fokus der empirischen Analysen auf die Frage, inwieweit ein Zusammenhang zwischen der Dauer der Schulzeit und den erreichten Lernständen besteht. Im Ergebnis hat sich gezeigt, dass im Test *Englisch – Allgemeines Sprachverständnis* und im TOEFL-Untertest *Listening Comprehension* deutliche Unterschiede ( $d > 0,20$ ) zwischen beiden Jahrgängen bestehen, und zwar zugunsten des KESS-Jahrgangs (G 8). Auch im TOEFL-Gesamttest, dem TOEFL-Untertest *Vocabulary & Reading Comprehension* und dem Test *Voruniversitäre Mathematik* ergab sich ein Leistungsvorsprung für den KESS-Jahrgang, allerdings schwächer ausgeprägt ( $0,05 \leq d \leq 0,20$ ). Im TOEFL-Untertest *Structure & Written Expression* wie auch im Test *Naturwissenschaftliche Grundbildung* liegen die mittleren Lernstände beider Jahrgänge annähernd gleichauf ( $d \leq 0,05$ ). Lediglich im Test *Mathematische Grundbildung* ergab sich ein Leistungsvorsprung ( $d = 0,14$ ) für den LAU-Jahrgang (G 9).

Die festgestellten Unterschiede gehen mit einer stark veränderten Schülerzusammensetzung einher. So ist der Anteil der Schülerinnen und Schüler mit Migrationshintergrund sowie aus eher bildungsfernen Elternhäusern im KESS-Jahrgang erheblich gestiegen. Die regressionsanalytischen Modelle haben gezeigt, dass auch nach Kontrolle der soziodemografischen und -kulturellen sowie der leistungsbezogenen Lernvoraussetzungen der Schülerinnen und Schüler beider Jahrgänge zu Beginn der Jahrgangsstufe 7 die bezogen auf die am Ende der gymnasialen Oberstufe erreichten Lernstände festgestellten Unterschiede statistisch hochsignifikant ( $p_{\text{zweiseitig}} \leq .001$ ) bleiben. Unter der Annahme, dass mit den leistungsbezogenen Lernvoraussetzungen zu Beginn der Jahrgangsstufe 7 (*AFI 7*), dem *Geschlecht*, der *Familiensprache*, dem *Geburtsland der Eltern*, dem *höchsten Schulabschluss der Eltern* und dem *Buchbestand im Elternhaus* wesentliche leistungs-

relevante Faktoren berücksichtigt und somit Kohorteneffekte (weitestgehend) neutralisiert sind, lässt sich damit ein Zusammenhang zwischen der Schulzeitverkürzung (einschließlich ihrer Begleitmaßnahmen) und den am Ende der gymnasialen Oberstufe in den untersuchten Kompetenzbereichen erreichten Lernständen nachweisen, allerdings – entgegen den allgemeinen Erwartungen – weitgehend zugunsten des KESS-Jahrgangs (G 8).

Vertiefende Analysen ergaben darüber hinaus, dass die für die Gesamtkohorten nachgewiesenen Zusammenhänge nicht gleichermaßen für Schülerinnen und Schüler mit unterschiedlichen domänenspezifischen Lernausgangslagen zu Beginn der Jahrgangsstufe 7 zutreffen. So sind die für die Gesamtkohorten ermittelten Effekte vor allem auf die Lernergebnisse der Schülerinnen und Schüler aus der unteren Leistungsgruppe zurückzuführen, das gilt sowohl für den im Test *Englisch – Allgemeines Sprachverständnis* festgestellten Leistungsvorsprung als auch für den im Test *Mathematische Grundbildung* festgestellten Leistungsrückstand des KESS-Jahrgangs.

Diese Befunde belegen, dass die Einführung des achtjährigen Bildungsgangs zum Abitur in Hamburg nicht zu einer generellen Absenkung des Leistungsniveaus in den untersuchten Kompetenzbereichen geführt hat. Dies ist umso bemerkenswerter, als der gymnasiale Anteil am Gesamtjahrgang sowie der Anteil an Schülerinnen und Schüler aus bildungsferneren Elternhäusern im Vergleich zum LAU-Jahrgang gestiegen ist. Ergänzende regressionsanalytische Berechnungen haben ergeben, dass bei Berücksichtigung der soziodemografischen und -kulturellen sowie der leistungsbezogenen Lernvoraussetzungen zu Beginn der Jahrgangsstufe 7 die Schülerinnen und Schüler des KESS-Jahrgangs bereits zu Beginn der gymnasialen Oberstufe in den Tests *Englisch – Allgemeines Sprachverständnis* und *Mathematische Grundbildung* einen hochsignifikanten Leistungsvorsprung gegenüber den Schülerinnen und Schülern des LAU-Jahrgangs aufweisen. Dabei zeigen sich sowohl zu Beginn als auch am Ende der gymnasialen Oberstufe domänenspezifische wie auch differenzielle Effekte, auf die bereits Baumert und Watermann (2000) hingewiesen haben.

Diese Befundlage gibt Anlass, die fachbezogenen Veränderungen, die mit dem KESS-Jahrgang eingeführt worden sind, genauer zu untersuchen. So könnten die verbesserten Englischleistungen des KESS-Jahrgangs beispielsweise mit der Einführung von Englisch in der Grundschule, der Implementierung kompetenz-

orientierter Rahmenpläne in Verbindung mit einem erhöhten Anforderungsniveau, einer Ausweitung des bilingualen Englischunterrichts sowie gezielten Fortbildungsmaßnahmen für die Fachlehrkräfte im Zusammenhang stehen. Das ungünstigere Abschneiden der Schülerinnen und Schüler des KESS-Jahrgangs in der mathematischen Grundbildung könnte demgegenüber auf die Reduzierung der Unterrichtszeit in der Oberstufe (Jahrgangsstufen 11 und 12) zurückzuführen sein. Dem ließe sich beispielsweise durch eine Erhöhung der Unterrichtszeit im Fach Mathematik von drei auf vier Wochenstunden in den Jahrgangsstufen 9 und 10 (bei Beibehaltung des Unterrichtsvolumens insgesamt) oder durch die Bereitstellung zusätzlicher Förderzeiten begegnen.

Insgesamt bestätigt der leistungsbezogene KESS-LAU-Kohortenvergleich bereits vorliegende Befunde, denen zufolge das G 8 nicht per se zu einem Niveauverlust führen muss. Im Gegenteil: Wie am Beispiel der Englischleistungen des KESS-Jahrgangs ausgewiesen, kann eine Schulzeitverkürzung durchaus auch mit Leistungsvorteilen einhergehen. Entscheidende Parameter für die am Ende der Schulzeit erreichten Lernstände resp. Kompetenzen wären demnach weit eher der Umfang und die Qualität des erteilten Fachunterrichts, nicht aber die Dauer der Schulzeit. Vor diesem Hintergrund erscheint es wünschenswert, in Folgestudien Merkmale guten Fachunterrichts verstärkt in den Blick zu nehmen und auf dieser Basis weitere Maßnahmen für die Entwicklung der Unterrichtsqualität zu ergreifen. Dabei ist zu bedenken, dass in Hamburg im Unterschied zu anderen Bundesländern die Einführung des G 8 nicht mit einer Verminderung des Unterrichtsvolumens verbunden gewesen ist, weshalb die Ergebnisse der Studie nicht generalisierbar sind.