

Originalartikel

Effekte der Gymnasialprofilzugehörigkeit auf Leistungsentwicklungen im Fach Englisch

Michael Leucht^{1,2}, Jan Retelsdorf¹, Hans Anand Pant³, Jens Möller⁴ und Olaf Köller¹

¹Leibniz-Institut für die Pädagogik der Naturwissenschaften und Mathematik (IPN), Kiel

²ISBS, Universität Vechta

³Institut zur Qualitätsentwicklung im Bildungswesen (IQB) an der Humboldt-Universität zu Berlin

⁴Institut für Psychologie, Christian-Albrechts-Universität zu Kiel

Zusammenfassung: Im Kontext der zunehmenden Aufgabe des Kurssystems zugunsten neigungsbezogener Profile in den Oberstufen allgemeinbildender Gymnasien prüfen wir, ob die Wahl eines sprachlichen Profils zu günstigeren Leistungsentwicklungen im Fach Englisch führt als die Wahl eines nicht-sprachlichen Profils. Insbesondere aufgrund der profilbezogen unterschiedlichen Belegungen von zweiter, dritter und ggf. vierter Fremdsprache wäre eine günstigere Entwicklung des sprachlichen Profils durch positiven Transfer zwischen schulisch erlernten Fremdsprachen zu rechtfertigen. Wir nutzen längsschnittliche Daten von $N = 1.171$ Schülerinnen und Schülern in 11. bzw. 13. Jahrgangsstufen aus Schleswig-Holstein. Wir beschreiben zunächst die an den fünf Profilen in der Qualifikationsphase belegten Fremdsprachen. Bezugnehmend auf den zurückliegenden Ländervergleich im Fach Englisch zeichnen wir Leistungsentwicklungen in zwei Teildomänen, dem Lese- und Hörverstehen, von der 9. über die 11. bis hin zur 13. Jahrgangsstufe nach. Mehrebenenanalysen sprechen sodann für eine schwache Bestätigung profilbezogener Transfereffekte, die sich gegenüber Effekten der leistungsbezogenen Komposition auf Klassenebene abgrenzen lassen.

Schlüsselwörter: Gymnasialprofile, rezepive Englischleistungen, Sekundarstufe II, Mehrebenenanalysen, differenzielle Lernmilieus

Study Profile Effects on English as a First Foreign Language Development

Abstract: In the academic track of German upper secondary education, the course level system is increasingly replaced by thematic study profiles. We investigate whether opting for a language profile (in which up to four foreign languages are taught) leads to a more favorable development of first foreign language (English) achievement than opting for a non-language profile (in which instruction in a second, third and fourth language can be avoided). We rely on longitudinal data of $N = 1.171$ 11th and 13th graders from one federal state, Schleswig-Holstein. We report to what extent foreign languages are taught in the five profiles. By relating to a recent national assessment we trace the development of students' achievement in two skill domains, reading and listening comprehension, over the grades 9, 11, and 13. Multilevel analyses finally provide a weak confirmation of transfer-related study profiles' effects that are robust against effects of achievement-related student body composition.

Keywords: thematic study profiles, receptive English skills, upper secondary school, multilevel analyses, differential learning environments

1 Einleitung

Zu den großen Gymnasialreformen der letzten Jahre zählt neben der Verkürzung der Beschulungsdauer (sogenanntes G8) und der Einführung von Bildungsstandards für die Allgemeine Hochschulreife die weitgehende Aufgabe des

bisherigen Systems aus Grund- und Leistungskursen in der Oberstufe. Die konkrete Umsetzung dieser letztgenannten Neuerung variiert von Bundesland zu Bundesland. Den meisten Ländern gemeinsam sind eine Aufwertung sogenannter Kernfächer und die Möglichkeit zur neigungsbezogenen Ausgestaltung von Beschulung durch ein indivi-

duell frei wählbares Profilmfach sowie eines oder mehrere Profilergänzungsfächer. Welche Effekte diese Neuerung auf Leistungsentwicklungen im Verlauf der Sekundarstufe II hat, ist bisher noch nicht untersucht worden. Die vorliegende Studie schließt diese Forschungslücke, indem sie die Leistungsentwicklung im Fach Englisch von der 11. bis zur 13. Jahrgangsstufe in fünf Profilen des allgemeinbildenden Gymnasiums in Schleswig-Holstein nachzeichnet. Konkret soll die Frage untersucht werden, ob die Wahl eines sprachlichen Profils, in dem Unterrichtsangebote in bis zu vier Fremdsprachen denkbar sind, zu günstigeren Leistungsentwicklungen führt als die Wahl eines nicht-sprachlichen Profils, in dem über die erste bzw. fortgeführte Fremdsprache hinausgehende fremdsprachliche Unterrichtsangebote auch durch andere Fächerkombinationen umgangen werden können. Damit beabsichtigen wir gleichermaßen eine Annäherung an ggf. differenzielle Beschulungseffekte in Gymnasialprofilen und einen Beitrag zum Diskurs zu Transfereffekten zwischen verschiedenen sprachlichen Fächern.

1.1 Leistungsentwicklung im Fach Englisch in der Sekundarstufe II

Anders als in den verkehrssprachlichen Leistungsdomänen (verkehrssprachliches Englisch: s. z. B. Hill, Bloom, Black & Lipsey, 2008; verkehrssprachliche Leseverstehensleistungen im Fach Deutsch: z. B. Retelsdorf, Becker, Köller & Möller, 2012; Klieme, 2006) ist beim schulischen Fremdsprachenlernen (s. Möller, Zaunbauer & Leucht, 2010) im Fach Englisch auch noch in der Sekundarstufe II von substantiellen jährlichen Leistungszuwächsen auszugehen. Die detailliertesten Aussagen für Deutschland und dort insbesondere das allgemeinbildende Gymnasium erlauben die Studien LAU (*Aspekte der Lernausgangslage und der Lernentwicklung*; z. B. Trautwein, Köller, Lehmann & Lüdtke, 2007) und KESS (*Kompetenzen und Einstellungen von Schülerinnen und Schülern*; z. B. Vieluf, Ivanov & Nikolova, 2011). Beide nahmen das Bundesland Hamburg in den Blick und verwendeten mit dem C-Test¹ ein Maß zur globalen Sprachkompetenz (Sigott, 2004). In LAU zeigte sich eine insgesamt akzelerierte Leistungsentwicklung: Betrag der Lernzuwachs zwischen dem jeweiligen Ende der Jahrgänge 9 und 11 in der späteren Sekundarstufen-II-Kohorte des Gymnasiums $d = 0.29$, stieg er zwischen Jahrgang 11 und 13 (Testung: Juni 2005) weiter auf $d = 0.58$. Sechs Kalenderjahre später belegte die KESS-Studie im dann achtjährigen gymnasialen Bildungsgang (G8) einen dezelerierten Lernverlauf: Der Leistungsunterschied zwischen Jahrgang 8 und 10 (jeweiliges Schuljahresende) entsprach $d = 0.66$, zwischen dem Beginn der 11. und dem Ende der 12. Jahrgangsstufe (Testung: Juni 2011) hingegen wurde nur ein Leistungsunterschied von $d = 0.38$ fest-

gestellt. Weder LAU noch KESS erlauben u.E. Aussagen zur eingangs genannten Neuordnung des Kurssystems in der gymnasialen Oberstufe.

1.2 Neugeordnete gymnasiale Oberstufe in Schleswig-Holstein (Profiloberstufe)

Die im Land Schleswig-Holstein im Schuljahr 2008/2009 erfolgte Neuordnung des Kurssystems der gymnasialen Oberstufe war mit folgenden Änderungen verbunden: Je nach Größe und Angebot der Gymnasien werden bis zu fünf Profile angeboten: (1) naturwissenschaftlich, (2) sprachlich, (3) gesellschaftswissenschaftlich, (4) ästhetisch und (5) sportlich. Unabhängig vom Profil müssen die Kernfächer Deutsch, Mathematik sowie erste bzw. fortgeführte Fremdsprache in der 12. und 13. Jahrgangsstufe («Qualifikationsphase») vierstündig belegt werden. Der Unterricht erfolgt auf erhöhtem Anforderungsniveau. Neben den Kernfächern wird ein profilgebendes Fach vierstündig auf erhöhtem Niveau unterrichtet. Das Profil wird arrondiert durch zwei profilergänzende Fächer, die i. d. R. dreistündig unterrichtet werden. Kernfächer, profilgebende und -ergänzende Fächer werden, so dies an der Einzelschule leistbar ist, im Klassenverband unterrichtet.

Die neigungsbezogene Schwerpunktsetzung führt dazu, dass sich das Ausmaß der Beschulung in sprachlichen Fächern insgesamt zwischen den Profilen unterscheidet: So werden im naturwissenschaftlichen Profil i. d. R. eine Naturwissenschaft (z. B. Biologie) als profilgebendes Fach und zwei weitere Naturwissenschaften (z. B. Physik, Chemie) als profilergänzende Fächer in der Qualifikationsphase unterrichtet. Neben Deutsch und der ersten bzw. fortgeführten Fremdsprache (zwei der drei Kernfächer; acht Wochenstunden) muss jedoch keine weitere Fremdsprache belegt werden. Dies gilt ebenfalls für die anderen drei nicht-sprachlichen Profile. Im sprachlichen Profil hingegen werden dieselben beiden Kernfächer (Deutsch, Fremdsprache) i. d. R. durch ein fremdsprachliches profilgebendes Fach (vier Wochenstunden) sowie maximal zwei fremdsprachliche profilergänzende Fächer (zusammen weitere sechs Wochenstunden) ergänzt. Diese modellhaften Überlegungen besitzen allerdings keine Verbindlichkeit, da die konkrete Ausgestaltung von Profilen letztlich ein Ergebnis von Entscheidungsprozessen auf Einzelschulebene ist.

1.3 Transfer zwischen sprachlichen Fächern

Aus spracherwerbtheoretischer Perspektive wird eine positive wechselseitige Beeinflussung der in verschiedenen sprachlichen Fächern erworbenen Kompetenzen unterstellt. Derartige Transferprozesse sind umfassend für

¹ Wir beschränken die Darstellung von LAU und KESS auf die dort im Längsschnitt untersuchten Testkonstrukte im Fach Englisch.

das Leseverstehen untersucht, wobei zwischen *Bottom-Up*- und *Top-Down*-Verarbeitungsprozessen (z.B. Carr & Levy, 1990; Perfetti, 1999; Koda, 2005) unterschieden wird. *Bottom-Up*-Verarbeitung besteht im Dekodieren von Textinhalten und dem Erstellen der Textoberflächen-Repräsentation sowie deren Reduktion zu Propositionen. *Top-Down*-Verarbeitung kennzeichnet die Anreicherung von Propositionen um Vorwissensinhalte, die abhängig ist von der automatisierten Anwendung metakognitiver Wissensbestände (d.h. Wissen um bestimmte Textgenres und diesbezügliche Erwartungen; z.B. Alexander, Schallert & Hare, 1991; Schoonen, Hulstijn & Bossers, 1998) sowie dem willentlichen Einsatz von Lesestrategien, die z.B. den Leseprozess supervidieren sowie kompensatorische Strategien für den Fall unzureichender *Bottom-Up*-Verarbeitung bereithalten (z.B. Anderson, 1991). Das Kernargument der ursprünglich in der Bilingualitätsforschung beheimateten Arbeiten zum Transfer (z.B. Cummins, 2000) besagt, dass insbesondere *Top-Down*-Verarbeitungsprozesse über Sprachsysteme hinweg geteilte Kompetenzen sind. Es folgt, dass beim Erwerb einer Fremdsprache auf Kompetenzen aus der Verkehrssprache sowie einer ggf. zuvor erlernten Fremdsprache rekurriert werden kann. Weiterhin ist gemäß van Gelderen, Schoonen, De Stool, De Gloppe und Hulstijn (2007) davon auszugehen, dass der Erwerb einer (ggf. weiteren) Fremdsprache eine Steigerung der transferierbaren Teilkompetenzen bewirkt, so dass hiervon auch eine zeitlich später stattfindende Verarbeitung in der Verkehrssprache sowie in der ggf. zuvor erlernten Fremdsprache profitieren wird.

Van Gelderen et al. (2007) belegten die Gültigkeit entsprechender Annahmen für das Leseverstehen am Ende der Sekundarstufe I. *Top-Down*-Verarbeitung als Träger von Transferprozessen ist dabei *kein* lesespezifischer Prozess, sondern kommt auf ähnliche Weise auch bei akustischer Sprachverarbeitung zum Einsatz (s. Perfetti, 1999; Kürschner & Schnotz, 2008). Erfolgreicher Transfer ist jedoch erst möglich, wenn in der neu gelernten Sprache eine hinreichende «Schwelle» sprachbezogener Wissensbestände (z.B. Wortschatz) sowie basaler Sprachverarbeitung überwunden wurde (vgl. die *Linguistic Threshold Hypothesis* nach Alderson, 1984). Eine zeitliche Präzisierung der Schwelle steht bislang aus. Die Arbeiten von Schoonen et al. (1998) sowie van Gelderen et al. (2007) legen nahe, dass sie im Mittel nach drei Jahren Beschulung überwunden wird. Darüber hinaus scheint erfolgreicher Transfer gebunden an die schriftsprachliche (zugrunde liegendes Alphabet, dessen Eigenschaften) und grammatikalische Ähnlichkeit beider Sprachen (z.B. Birch, 2007; Georgiou, Parrila & Papadopoulos, 2008) sowie kulturübergreifende Variationen einzelner Wortbedeutungen (z.B. Dewaele & Pavlenko, 2002; Schwanenflugel, Blout & Lin, 1991). Beide letztgenannten Limitationen verweisen darauf, dass für die eng verschränkt ablaufende *Bottom-Up*- und *Top-Down*-Verarbeitung nur ein gemeinsames, begrenztes Maß an Arbeitsgedächtniskapazität zur Verfügung steht.

1.4 Alternative Zugänge zur Erklärung eventueller Profileffekte

Jenseits von Transferprozessen sind mehrere andere Ansätze zur Erklärung ggf. differierender Zuwachsraten in Gymnasialprofilen denkbar: Etwa erscheint plausibel, dass Klassenverbände im sprachlichen Profil bereits zu Beginn der Qualifikationsphase über höhere mittlere Ausgangsleistungen im Fach Englisch verfügen. Zahlreiche Arbeiten haben auf die Bedeutung der leistungsbezogenen Komposition auf Klassenebene für nachfolgendes Lernen hingewiesen (für zweitsprachliche Französischleistungen s. Neumann et al., 2007). Derartige Effekte (vgl. im Detail Harris, 2010, sowie Dumont, Neumann, Maaz & Trautwein, 2013) werden erstens auf einen *Ansteckungseffekt* in lernrelevanten Verhaltensnormen sowie konkretem Lernverhalten in der Peergruppe «Klasse» rückgeführt. Zweitens wird argumentiert, dass Lehrkräfte auf das Leistungsniveau der Klasse *reagieren*, indem sie ihre Instruktion spezifisch auf den jeweiligen Lernstand der Klasse ausrichten, so dass Lehrkräfte in von ihnen als «fähiger» eingeschätzten Klassen i. d. R. auch anspruchsvolleren Unterricht realisieren werden.

Weiter ist denkbar, dass zwischen Englischlehrkräften Unterschiede bestehen und dass diese bei der Zuweisung von Lehrkräften zu Profilen Beachtung finden (z.B. besonders gut ausgebildete Englischlehrkräfte werden eher dem sprachlichen Profil zugeordnet). Dass Unterschiede in der Lehramtsausbildung Effekte insbesondere auf das Fach- und fachdidaktische Wissen von Lehrkräften haben, wurde für das Fach Mathematik durch COACTIV belegt (z.B. Baumert et al., 2010). Zusätzlich ist zu vermuten, dass die Wahl eines sprachlichen Profils auf Schülerebene motivational mit einem höheren Fachinteresse sowie einer höheren Wertschätzung von fremdsprachlichen Fächern einhergeht. Der zumeist reziproke Zusammenhang von Fachinteresse und Fachleistungen wurde vielfach belegt, auch für das Fach Englisch (s. z.B. Köller & Baumert, 2012). Auch Unterschiede in der Wahrnehmung außerschulischer Lerngelegenheiten sind denkbar. LAU-13 legte deutliche Effekte längerer Auslandsaufenthalte auf die Englischleistungen am Gymnasium nahe (s. Jonkmann, Köller & Trautwein, 2007). Angesichts der veränderten Verfügbarkeit medialer Angebote (z.B. englischsprachige Fernsehserien und Filme) wäre weiter davon auszugehen, dass Schülerinnen und Schüler im sprachlichen Profil aufgrund von höheren Leistungen, höherem Fachinteresse und höherer Wertschätzung von Englisch in ihrer Freizeit eher entsprechende Angebote nutzen werden als Schülerinnen und Schüler in nicht-sprachlichen Profilen.

2 Fragestellungen

Unsere Studie zielt *erstens* auf eine Deskription der in der Qualifikationsphase an den fünf Profilen konkret belegten Fremdsprachen. *Zweitens* beabsichtigen wir eine Abschät-

Tabelle 1

Verteilung der untersuchten Schülerinnen und Schüler auf die fünf Profile sowie Anteile der Schülerinnen

| Profil | Prozentuale Verteilung | | Anteil weiblich in «gewichtet» |
|-------------------------------|--------------------------|----------------|--------------------------------|
| | ungewichtet ^a | gewichtet | |
| Naturwissenschaftlich | 386 (33.0 %) | 2.012 (31.9 %) | 766 (38.1 %) |
| Sprachlich | 228 (19.5 %) | 1.196 (19.0 %) | 966 (80.7 %) |
| Gesellschaftswissenschaftlich | 369 (31.5 %) | 1.998 (31.7 %) | 725 (36.3 %) |
| Ästhetisch | 115 (9.8 %) | 699 (11.1 %) | 599 (85.6 %) |
| Sportlich | 73 (6.2 %) | 400 (6.3 %) | 136 (34.1 %) |
| Gesamt | 1171 (100.0 %) | 6305 (100.0 %) | 3192 (50.6 %) |

Anmerkungen: ^a Anzahl der Klassenverbände: Profil «naturwissenschaftlich»: 22 Klassen, Profil «sprachlich»: 13 Klassen, Profil «gesellschaftswissenschaftlich»: 22 Klassen, Profil «ästhetisch»: 7 Klassen, Profil «sportlich»: 4 Klassen. $M_{\text{Klassengröße}} = 17.2$, $SD_{\text{Klassengröße}} = 4.7$ ($Min = 6$, $Max = 25$).

zung der in den Profilen erwartbaren Leistungszuwächse zwischen der 11. und 13. Jahrgangsstufe. Unsere Stichprobe war 2009 bereits Bestandteil der Population des von Köller, Knigge und Tesch (2010) berichteten Ländervergleichs. Da unserer Studie dieselbe Metrik zugrunde liegt, ist für die Gesamtgruppe am allgemeinbildenden Gymnasium eine Abschätzung der durchschnittlichen Leistungsentwicklung von der 9. über die 11. bis hin zur 13. Jahrgangsstufe möglich. Über LAU- und KESS-Ergebnisse «zusammenfassend» betrachtet könnte die Annahme einer gleichförmig ansteigenden Lernentwicklung in diesem Zeitraum gerechtfertigt werden. Andererseits untersuchen wir wie die LAU-Studie den neunjährigen Bildungsgang am Gymnasium, was einen akzelerierten Lernverlauf (d. h. ein Ansteigen der Zuwachsraten) nahelegt. Über LAU und KESS hinausgehend ermöglicht unsere Arbeit erstmalig eine Unterscheidung ggf. differenzieller Entwicklungen in den zwei Teildomänen des Englischen, dem Lese- und Hörverstehen.

Nachfolgend rücken wir *drittens* das Zusammenspiel von Ausgangsleistungen in der 11. Jahrgangsstufe, gewähltem Profil sowie nachfolgender Leistungsentwicklung bis zum Ende der 13. Jahrgangsstufe in den Fokus. Nicht alle der genannten Erklärungsansätze lassen sich hier empirisch prüfen. Um Störvariablen auszuschließen, werden nur Schülerinnen und Schüler betrachtet, für die Englisch die erste schulisch erlernte Fremdsprache ist und die vierstündigen Unterricht im Fach Englisch erhalten. Sodann beabsichtigen wir eine Annäherung an eine empirische Überprüfung der Transferhypothese in drei Schritten. Im Einzelnen sollen folgende Fragen geprüft werden:

- 3a) Resultieren nach Kontrolle der individuellen Ausgangsleistungen des 11. Jahrgangs am sprachlichen Profil bis zum Ende des 13. Jahrgangs höhere Leistungszuwächse als an nicht-sprachlichen Profilen?
- 3b) Sind diese eventuellen Profileffekte robust gegenüber Effekten der leistungsbezogenen Komposition auf Klassenebene?

3c) Führt die zusätzliche Berücksichtigung individueller Fremdsprachenbelegungen zu einer Veränderung eventueller Profileffekte?

3 Methode

3.1 Stichprobe

Die Untersuchung erfolgte im Rahmen des fünften (LISA 5) und sechsten (LISA 6) Messzeitpunkts der LISA-Studie². Die zum ersten Messzeitpunkt (Beginn des Schuljahres 2004/2005) gezogene Schulstichprobe war repräsentativ für Schleswig-Holstein. Ab LISA 5 wurden an den Gymnasien komplette Jahrgangsstufen untersucht. Die Testungen für LISA 5 erfolgten im Januar/Februar 2011 (11. Jahrgangsstufe). LISA 6 folgte im April/Mai 2013 (13. Jahrgangsstufe), d. h. zwischen den Erhebungen lagen rund 27 Monate. Die Stichprobenziehung erfolgte am IEA *Data Processing and Research Center* (DPC) in Hamburg, das auch die Datenerhebungen durch trainierte Testleiter verantwortete.

Nachfolgend betrachten wir den Überschneidungsbereich der am allgemeinbildenden Gymnasium anvisierten Stichproben beider Messzeitpunkte. Wir verwenden eine Gewichtung der Daten, die einen Rückschluss auf alle im Schuljahr 2012/2013 in Schleswig-Holstein am allgemeinbildenden Gymnasium unterrichteten Schülerinnen und Schüler zulässt. Hierzu approximiert die am DPC erstellte Gewichtung die in der Fachserie 11, Reihe 1, für das Schuljahr 2012/13 (Statistisches Bundesamt, 2014; s. dort Tabelle 3.4) angegebene Schülerzahl von 7695 Schülerinnen und Schülern. Tabelle 1 zeigt die prozentuale Verteilung der betrachteten Teilstichprobe von LISA 5 und LISA 6 ($N = 1.171$) auf die fünf Profile für ungewichtete und gewichtete Daten sowie zusätzlich profilbezogenen Anteile der Schülerinnen für unsere gewichteten Daten.

2 Bis einschließlich LISA 5 stand das Akronym «LISA» für «Lesen in der Sekundarstufe». Zu LISA 6 wurde das Studiendesign erweitert. Dem wurde auch durch einen veränderten Namen der Studie Rechnung getragen: «LISA 6: Eine empirische Studie zu Lernergebnissen an beruflichen und allgemeinbildenden Gymnasien in Schleswig-Holstein».

Die Abweichung der gewichteten Gesamtschülerzahl ($N = 6.305$) von dem in Fachserie 11 angegebenen Vergleichswert resultiert zunächst aus Schulwechslern: Genau 10.0 % der zu LISA 6 untersuchten Schülerinnen und Schüler waren zwischen LISA 5 und LISA 6 von anderen allgemeinbildenden Gymnasien auf eine der LISA-Schulen gewechselt. Für diese Schulwechsler liegen keine Daten aus LISA 5 vor, daher werden sie ausgeschlossen. Andererseits ergibt sich die Abweichung aus einem Ausschluss der Schülerinnen und Schüler mit einer anderen ersten Fremdsprache als Englisch (1.6 %) sowie weniger als vier Wochenstunden im Fach Englisch (5.2 %).

Zum Teil rekurren wir nachfolgend auf Ergebnisse des Ländervergleichs von Köller et al. (2010). Dies erscheint zulässig, da die in LISA untersuchte Gymnasialkohorte der Population des allgemeinbildenden Gymnasiums entstammt, die im Schuljahr 2008/2009 am Ländervergleich teilgenommen hatte. Der Ländervergleich fand im April bis Juni 2009 statt, zwischen ihm und LISA 5 lagen rund 21 Monate.

3.2 Instrumente

3.2.1 Lernstände in der ersten Fremdsprache Englisch

Zur Erfassung des Lese- und Hörverstehens wurde eine Teilmenge der Aufgaben des Ländervergleichs von Köller et al. (2010) eingesetzt. Sie erfassen die in den Bildungsstandards für die erste Fremdsprache Englisch (KMK [Ständige Konferenz der Kultusminister der Länder der Bundesrepublik Deutschland], 2004) beschriebenen kommunikativen Fertigkeiten des Lese- und Hörverstehens. Diese werden – unter anderem durch Bezugnahme auf den Gemeinsamen Europäischen Referenzrahmen für Sprachen (Europarat, 2001) – übergreifend als Fähigkeiten zum Verstehen geschriebener sowie auditiv vermittelter Fremdsprache aufgefasst. Die hierbei zu erbringenden rezeptiven Teilleistungen reichen vom Verstehen spezifischer Details bis hin zum globalen, schlussfolgernden Verstehen (s. auch Leucht, Retelsdorf, Möller & Köller, 2010). Zur Erfassung des englischsprachigen Lese- und Hörverstehens werden dabei ausschließlich authentische, auf reale Sprachverwendungssituationen bezogene Texte verwendet. Die Operationalisierung in konkreten Testaufgaben erfolgte durch eine geschulte Gruppe von Aufgabenentwicklern unter Anleitung durch eine internationale Expertin im Bereich des *Language Assessment* (s. im Detail Rupp, Vock, Harsch & Köller, 2008). Vor dem Ländervergleich erfolgte zudem eine systematische Pilotierung und Normierung der Aufgabenmengen (s. Köller et al., 2010). In LISA 5 wurden 133 Lese- sowie 118 Hörverstehensaufgaben eingesetzt (*Multi-Matrix-Design*). Die Schätzung von Personenfähigkeiten erfolgte auf Basis einparametrischer *Item-Response-Modelle* in *ConQuest* (Version 3.0; Wu, Adams, Wilson & Haldane, 2007). Als Schätzer für die Personenfähigkeiten wurden *Plausible*

Values (PVs) verwendet. Beide Tests sind zufriedenstellend reliabel (Leseverstehen: .77; Hörverstehen: .75). Zu LISA 6 wurden 42 Lese- sowie 32 Hörverstehensaufgaben aus LISA 5 erneut eingesetzt (*PV-Reliabilität* jeweils .82). Die Interkorrelation der Lese- und Hörverstehens-*PVs* beträgt .56 zu LISA 5 und .77 zu LISA 6.

3.2.2 Zusätzliche Variablen

Zur Erfassung der Intelligenz wurde in LISA 5 die Skala N2 (Figurenanalogien) des *Kognitiven Fähigkeits-test* (KFT) von Heller und Perleth (2000) verwendet. Im Rasch-Modell resultierte eine *PV-Reliabilität* von .73. Im Vorfeld von LISA 5 und 6 wurden an den Schulen so genannte Schülerteilnahmelisten (STL) ausgefüllt, die jeweils alle im 11. bzw. 13. Jahrgang unterrichteten Schülerinnen und Schüler listeten. Hiermit wurden Angaben zur Profizugehörigkeit, zum Geschlecht, zum Alter sowie zur Frage erhoben, ob Englisch die erste schulisch erworbene Fremdsprache ist. Spezifisch in LISA 5 berücksichtigte die STL Deutsch-, Mathematik- und Englischnoten des Abschlusszeugnisses der 10. Jahrgangsstufe, spezifisch für LISA 6 die im Fach Englisch belegten Wochenstunden (Schuljahr 13/1). STL-Daten liegen vollständig vor. Ebenfalls vollständig sind die Angaben zu Zeugnisnoten des Schulhalbjahres 13/2 (Deutsch, Mathematik, Englisch) sowie zu Abiturgesamtnoten, die im Juni 2013 per Fax von den Schulen erfragt wurden.

Die in der Qualifikationsphase (d.h. in der 12. bzw. 13. Jahrgangsstufe) belegten Fremdsprachen wurden in LISA 6 mit Hilfe eines Schülerfragebogens erfragt. Obschon die obigen Ausführungen die Möglichkeit von vier belegten Fremdsprachen am sprachlichen Profil nahelegen, wurden im Fragebogen nur die ersten drei schulisch erlernten Fremdsprachen berücksichtigt («ja/nein» zur Belegung in Jahrgang 12 und 13 sowie offene Angaben zur Nennung konkret schulisch erlernter erster, zweiter und dritter Fremdsprachen). Die Teilnahme am Fragebogen war freiwillig, 56.0 % unserer Stichprobe haben ihn bearbeitet. Zwischen den Profilen bestehen Unterschiede in der Teilnahmequote (*Min* = 44.7 %, *Max* = 72.8 %), in der Gesamtstichprobe gibt es jedoch kaum Hinweise auf einen systematischen Stichprobenausfall, wie in weiterführenden Selektivitätsanalysen zum Geschlecht ($d = 0.08$), zur Abiturgesamtnote ($d = -0.05$) sowie zur Intelligenz ($d = 0.09$) gezeigt werden konnte.

3.3 Analytisches Vorgehen

3.3.1 Umgang mit fehlenden Werten

Die Teilnahme an den Leistungstests in LISA 5 und 6 war per Ministeriumserlass verpflichtend. Dennoch war ein substanzieller Anteil der auf den STL jeweils geführten Schülerschaften am Tag der Testung nicht anwesend: In LISA 5 betrug diese Ausschöpfungsquote 88.2 %, für LISA 6

82.2 %. Zur Schätzung der *PVs* für die am Testtag nicht anwesenden Personen wurde auf Angaben der STL und Faxabfrage Bezug genommen. Die Hintergrundmodelle (d. h. *ConQuest-Option Regression*) für LISA 5 berücksichtigten die Profiltugehörigkeit, das Geschlecht, das Alter der Schülerinnen und Schüler sowie die Deutsch-, Mathematik- und Englischnoten (10. Jahrgang). Zusätzlich wurden Klassenmittelwerte der Punktschätzer aus den *PV*-Verteilungen (*Expected A-posteriori, EAPs*) aus initialen Skalierungen aller am Testtag anwesender Schülerinnen und Schüler (Lese-, Hörverstehen, KFT) berücksichtigt. Das Hintergrundmodell für LISA 6 Englisch berücksichtigte wiederum Profiltugehörigkeit³, Geschlecht, Alter, gleichlautende Zeugnisnoten (13/2), die Abiturgesamtdurchschnittsnote sowie erneut Klassenmittelwerte der initialen Skalierungen. Alle Hintergrundmodelle wurden faktorisiert (orthogonale Hauptkomponenten, die 90 % der Varianz extrahieren).

Auf eine Imputation der Fragebogendaten wurde aus folgenden Gründen verzichtet: a) Der Anteil fehlender Werte erschien mit 44.0 % vergleichsweise hoch, b) belegte Fremdsprachen und Fremdsprachenfolgen bilden profilspezifisch nicht-normalverteilte Daten und c) jenseits besuchter Profile und der Englischnoten (13/2) lagen keine vielversprechenden Prädiktoren für ein Imputationsmodell vor. Daraus folgt, dass die Fragestellungen 1 und 3c nur für eine reduzierte Teilstichprobe von 607 (gewichtet: 3528) Schülerinnen und Schülern beantwortet werden, die sich zu 31.2, 15.2, 36.0, 9.4 und 8.2 % (gewichtete Daten) auf die fünf genannten Profile verteilen.

3.3.2 Skalen-Equating auf die Metriken des Ländervergleichs

Für die eingesetzten Englischtests erfolgte ein *Equating* (z. B. Kolen & Brennan, 2004) der Skalen auf die Metriken des Lese- bzw. Hörverstehens im Ländervergleich. Die Itemschwierigkeiten wurden auf ihre damaligen Werte verankert und die resultierenden *PVs* wurden unter Verwendung der gewichteten Mittelwerte und Standardabweichungen des Ländervergleichs auf dessen Metrik transformiert ($M = 500$ und $SD = 100$ im 9. Jahrgang des allgemeinbildenden Schulsystems in Deutschland im Schuljahr 2008/09). Alle Analysen berücksichtigen diese fünf transformierten *PVs*, ihre Ergebnisse wurden entsprechend den Vorgaben von Rubin (1987) gemittelt. Wie im Ländervergleich beruhen alle Analysen auf gewichteten Daten.

3.3.3 Mehrebenenanalytisches Vorgehen

Die hierarchische Datenstruktur hat zur Folge, dass die effektive Stichprobengröße aufgrund der Intra-Klas-

senkorrelation reduziert ist. Dies führt typischerweise zu einer Unterschätzung der Standard- und damit zu einer Inflation der Alpha-Fehler (z. B. Hox, 2010). Zur Untersuchung der Fragestellung 3 erfolgte daher eine Trennung von Effekten auf Individual- und Klassenebene. Dies realisierten wir in einem mehrebenenanalytischen Zugang in *Mplus* (Version 7.0; Muthén & Muthén, 1998–2012; Analyseoption *Type = Twolevel*). Hierfür war eine Zuordnung von Schülerinnen und Schülern zu Klassenverbänden erforderlich, die eindeutig genau einem der fünf Profile zuzuordnen sind. Dies wurde durch eine Kombination von Schul-Identifikationsnummer, Klassenbezeichnung in LISA 6 (z. B. «13A») und dem individuell besuchten Profil realisiert. So resultierten 68 Klassenverbände ($M_{\text{Klassengröße}} = 17.2$, $SD_{\text{Klassengröße}} = 4.7$). Die Klassenanzahlen je Profil sind der Anmerkung zu Tabelle 1 zu entnehmen. In der reduzierten Teilstichprobe mit vollständigen Fragebogendaten resultierten 46 Klassen ($M_{\text{Klassengröße}} = 12.0$, $SD_{\text{Klassengröße}} = 4.3$).

4 Ergebnisse

4.1 In der Qualifikationsphase belegte Fremdsprachen

Tabelle 2 zeigt deskriptive Ergebnisse zur Fragestellung 1. Wir fokussieren hier Fragebogenangaben zur 12. Jahrgangsstufe (im 13. Jahrgang ergaben sich nur marginale Abweichungen; jeweils < 2 %). Die Zahlen weisen tatsächlich auf einen möglichen Vorteil des sprachlichen Profils hin, da hier die Mehrheit der Schülerinnen und Schüler drei Fremdsprachen belegt hatte. Die Zahlen verdeutlichen aber auch Unterschiede in den Lerngelegenheiten zwischen den nicht-sprachlichen Profilen: Während am naturwissenschaftlichen und sportlichen Profil tatsächlich so gut wie ausschließlich nur eine Fremdsprache besucht wurde, belegten substantielle Anteile der Schülerinnen und Schüler des gesellschaftswissenschaftlichen und ästhetischen Profils zwei Fremdsprachen.

Wie sich weiterhin zeigen lässt, erfolgte die Beschulung in der zweiten und dritten Fremdsprache *nicht* innerhalb der gebildeten Klassenverbände. Der jeweils klassenbezogene Anteil «Belegung einer zweiten Fremdsprache» variierte z. B. am naturwissenschaftlichen Profil zwischen 0.0 und 82.0 % ($SD = 0.24$). Ähnliche Streuungen ($SD = 0.14$ bis 0.26) zeigten sich auch an den anderen nicht-sprachlichen Profilen (Belegung zweiter Fremdsprache) sowie am sprachlichen Profil für die Belegung einer zweiten (0.08) bzw. dritten (0.15) Fremdsprache.

3 Genau 4.2 % der 1.171 untersuchten Schülerinnen und Schüler haben zwischen LISA 5 und 6 das Profil gewechselt. Für die unten beschriebene Klassenverbandsvariable sowie für alle Gruppenbildungen im Ergebnisteil wurden Profiltugehörigkeiten aus LISA 6 zugrunde gelegt.

Tabelle 2

Prozentuale Anteile von Schülerinnen und Schülern mit einer, zwei oder drei belegten Fremdsprachen, aufgebrochen nach Profil

| Profil | Prozentuale Anteile von Schülerinnen und Schülern mit ... | | |
|-------------------------------|---|--------------------------------|--------------------------------|
| | ...einer belegten Fremdsprache | ...zwei belegten Fremdsprachen | ...drei belegten Fremdsprachen |
| Naturwissenschaftlich | 82.6% | 17.4% | - |
| Sprachlich | 15.8% | 8.1% | 76.1% |
| Gesellschaftswissenschaftlich | 59.5% | 40.5% | - |
| Ästhetisch | 50.8% | 49.2% | - |
| Sportlich | 83.2% | 16.2% | - |
| Gesamt | 61.3% | 27.2% | 11.6% |

Anmerkungen: Fremdsprachenfolgen (i. S. v. erster, zweiter und dritter Fremdsprache): «Englisch, Französisch, Latein» (34.9%), «Englisch, Latein, Französisch» (26.2%), «Englisch, Französisch, Spanisch» (16.8%), «Englisch, Latein, Spanisch» (15.0%) sowie vier weitere Kombinationen, auf die zusammengenommen 7.1% entfallen.

4.2 Leistungsentwicklung von der 9. bis hin zur 13. Jahrgangsstufe

Abbildung 1 zeigt deskriptiv die Leistungsentwicklung im Lese- und Hörverstehen von der 9. (Vergleichswerte für Schleswig-Holstein aus Tabelle 5.4 bei Köller et al., 2010) über die 11. (LISA 5) bis zur 13. Jahrgangsstufe (LISA 6). Im Leseverstehen kam es insgesamt zu einer Verlangsamung der Leistungsentwicklung. Betrug der Leistungsunterschied zwischen der 9. und 11. Jahrgangsstufe 0.81 *SD* (bzw. 2.8 Punkte pro Monat), lautete der Vergleichswert zwischen Jahrgang 11 und 13 nur noch 0.25 *SD* (0.6 Punkte pro Monat). Die profilweisen Mittelwertsunterschiede korrespondieren mit Effektstärken zwischen $d = 0.15$ (sportliches) und $d = 0.38$

(sprachliches Profil). Ein im Ausmaß anderes Bild zeigt sich im Hörverstehen: Der Leistungsunterschied zwischen Jahrgang 9 und 11 setzte mit 0.94 *SD* (bzw. 3.0 Punkte pro Monat) etwas höher an und blieb zwischen Jahrgang 11 und 13 annähernd stabil (1.18 *SD* bzw. 2.7 Punkte pro Monat). Die profilbezogenen Mittelwertsunterschiede zwischen Jahrgang 11 und 13 weisen auch hier für das sprachliche Profil den höchsten ($d = 1.35$) und für das sportliche Profil den niedrigsten ($d = 1.04$) Zuwachs aus.

Zudem bestanden erhebliche Unterschiede im absoluten Niveau: So erreichte der 11. Jahrgang des sprachlichen Profils im Leseverstehen bereits mittlere Leistungen, die z. T. weit jenseits dessen liegen, was die anderen vier Profile erst am Ende des 13. Jahrgangs leisteten.

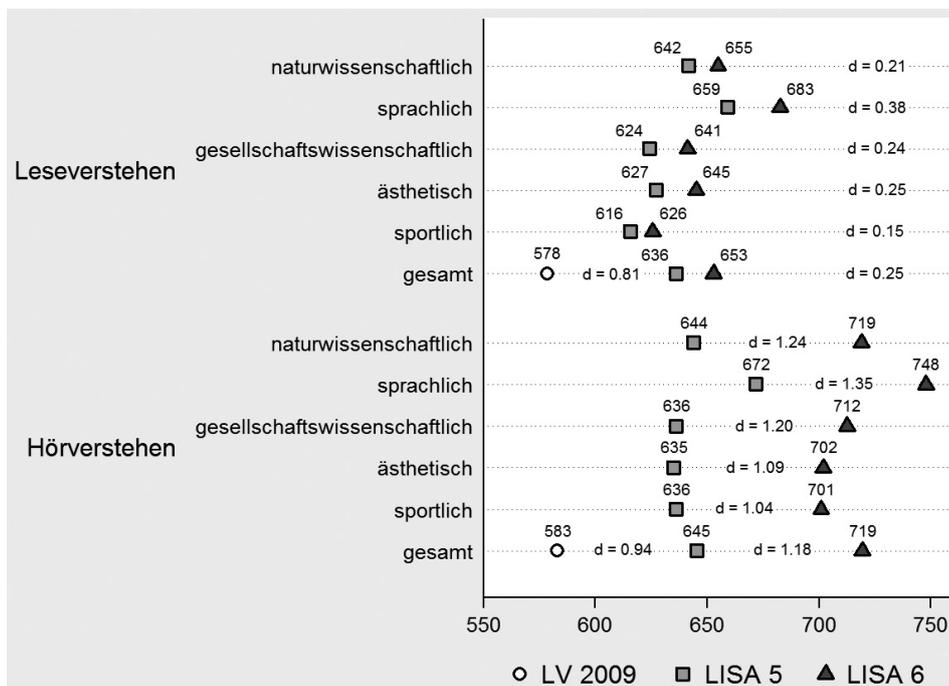


Abbildung 1. Leistungsentwicklung im englischsprachigen Lese- (oben) und Hörverstehen vom Ende der 9. über die Mitte der 11. bis hin zum Ende der 13. Jahrgangsstufe.

Tabelle 3

Vorhersage der Lernstände im Fach Englisch am Ende der 13. Jahrgangsstufe. Unstandardisierte Regressionskoeffizienten aus Mplus.

| | Modell 1 ^a | | Modell 2 ^b | | Modell 3 ^b | | Modell 4 ^{b,c} | |
|---|----------------------------|----------------------------|----------------------------|----------------------------|----------------------------|----------------------------|----------------------------|----------------------------|
| | b_{Lesen} (SE) | $b_{\text{Hören}}$ (SE) | b_{Lesen} (SE) | $b_{\text{Hören}}$ (SE) | b_{Lesen} (SE) | $b_{\text{Hören}}$ (SE) | b_{Lesen} (SE) | $b_{\text{Hören}}$ (SE) |
| <i>Individuum</i> | | | | | | | | |
| Englischleistung 11. Jg ^d | 21.9* (2.2) | 18.7* (2.5) | 22.1* (2.2) | 18.7* (2.5) | 22.0* (2.2) | 18.6* (2.5) | 25.7* (2.5) | 20.3* (2.7) |
| Geschlecht (1 = w.) | -3.6 (4.3) | 0.5 (3.7) | -1.2 (4.3) | 2.1 (3.6) | -3.7 (4.2) | 0.5 (3.7) | -7.9 (5.7) | -8.0 (5.4) |
| Intelligenz 11. Jg ^d | 7.4* (2.0) | 9.0* (2.0) | 7.2* (2.0) | 8.6* (2.0) | 7.4* (2.0) | 9.0* (2.0) | 7.6* (2.7) | 9.3* (2.5) |
| Belegung von zwei Fremdsprachen | - | - | - | - | - | - | 9.8 (6.7) | 9.3 (6.5) |
| Belegung von drei Fremdsprachen | - | - | - | - | - | - | 9.7 (14.3) | 13.8 (18.5) |
| R^2 | .18 | .16 | .16 | .14 | .17 | .14 | .23 | .19 |
| <i>Klasse</i> | | | | | | | | |
| Englischleistung 11. Jg ^e | - | - | -9.2* (3.4) | -5.6 (3.6) | -15.0* (4.0) | -10.4* (3.6) | -17.8* (5.1) | -14.0* (3.7) |
| Profil «naturwiss.» | -24.4* (6.7) | -23.1* (7.0) | - | - | -25.2* (7.4) | -22.7* (7.8) | -25.2 (14.4) | -17.6 (16.2) |
| Profil «gesellschafts- wiss.» | -31.3* (7.9) | -23.5* (8.3) | - | - | -32.7* (10.9) | -23.0* (9.3) | -31.5* (14.8) | -17.6 (16.5) |
| Profil «ästhetisch» | -25.1* (9.7) | -33.3* (12.0) | - | - | -26.6* (11.1) | -32.9* (12.7) | -18.2 (16.4) | -23.4 (20.9) |
| Profil «sportlich» | -50.0* (10.3) | -40.2* (10.9) | - | - | -51.8* (12.0) | -39.6* (11.7) | -47.0* (17.2) | -32.1 (19.7) |
| R^2 | .36 | .30 | .30 | .30 | .52 | .48 | .59 | .42 |

Anmerkungen: ^a 125 Replikationen ($5 \times 5 \times 5$ PVs). ^b 625 Replikationen ($5 \times 5 \times 5 \times 5$ PVs), Gruppenmittelwert-Zentrierung des Individualmerkmals «Englischleistung 11. Jg». ^c Befunde für reduzierte Teilstichprobe (s. Abschnitt 3.3.1). ^d z-Standardisierung kontinuierlicher Prädiktoren auf Ebene 1. ^e Kompositionseffekt = Effekt_{Klasse} - Effekt_{Individuum}. * $p < .05$.

4.3 Differenzielle Leistungsentwicklungen an den fünf Profilen

Tabelle 3 präsentiert die Ergebnisse der Mehrebenenanalysen. Zunächst sprechen die Intraklassenkorrelationen von jeweils 0.13 dafür, dass Klassenmerkmale für die Vorhersage der Lernstände in der 13. Jahrgangsstufe insgesamt eher nachgeordnet bedeutsam waren. Vor diesem Hintergrund erlaubt Modell 1 die Beantwortung der Fragestellung 3a, wobei wir zusätzlich interindividuelle Unterschiede in der Intelligenz (11. Jahrgang) sowie im Geschlecht kontrollieren. Auf Klassenebene resultierten an den nicht-sprachlichen Profilen niedrigere Zuwachsraten als am sprachlichen Profil: Etwa erreichten Klassen des sportlichen Profils eine mittlere Leistung, die auf der Bildungsstandard-Metrik 50 bzw. 40 Punkte unter dem von mit Blick auf Ausgangsleistungen und Intelligenz vergleichbaren Klassen des sprachlichen Profils lag. Die Lernrückstände für die anderen nicht-sprachlichen Profile fielen mit 23 bis 33 Punkten weniger deutlich aus. Trotz numerisch und statistisch bedeutsamer Gruppenunter-

schiede lag der auf die Profilvariable zurückgehende Anteil an der Gesamtvarianz – als Folge der niedrigen Intra-klassenkorrelation – in Modell 1 bei lediglich 4.7 (Lese-) bzw. 3.9 % (Hörverstehen).

Fragestellung 3b kann mit Hilfe der Modelle 2 und 3 beantwortet werden. Zur Abschätzung der Kompositionseffekte wurde der von Marsh, Lüdtke, Robitzsch et al. (2009) vorgeschlagene *Doubly-Manifest Approach* verwendet. Modell 2 belegt allein im Leseverstehen einen gering ausgeprägten Kompositionseffekt. Anders als in der Literatur beschrieben ist das Vorzeichen des Effektes negativ, d.h. höhere mittlere Ausgangsleistungen in Jahrgang 11 korrespondierten offenkundig mit geringeren Leistungszuwächsen bis zum Ende der Jahrgangsstufe 13. Modell 3 illustriert, dass die profilbezogen differenziellen Zuwachsraten im Lese- und Hörverstehen unabhängig sind von der leistungsbezogenen Komposition (und spricht im Ansatz für eine Interaktion von Komposition und Profiltugehörigkeit im Hörverstehen). Da sich die Koeffizienten der Profiltugehörigkeiten in Modell 3 gegenüber Modell 1 nur marginal veränderten (0.4 bis 1.8 Punkte), verzichteten wir hier auf

einen statistischen Test. Durch simultane Berücksichtigung von Profileffekten und Komposition erhöhte sich der erklärte Anteil der Gesamtvarianz auf 6.8 (Lese-) bzw. 6.2 % (Hörverstehen).

4.4 Untersuchung möglicher Transfereffekte

Eine Überprüfung von Fragestellung 3c erlaubt Modell 4. Auf Individualebene resultierten für die individuellen Belegungen zweiter und dritter Fremdsprachen keine belastbaren Effekte. Allerdings bedingte deren Berücksichtigung eine z.T. deutliche Erhöhung der Standardfehler profilbezogener Effekte auf Klassenebene. Diese verfehlten nun insbesondere im Hörverstehen das Signifikanzniveau ($p < 0.05$). Dennoch wurden in *Mplus* Wald-Tests berechnet, um exakt prüfen zu können, inwieweit Modell 4 eine statistisch bedeutsame Veränderung der Profileffekte gegenüber Modell 3 kennzeichnete. Referenzwerte der Wald-Tests waren die profilbezogenen Regressionskoeffizienten aus Modell 3 in der reduzierten Teilstichprobe mit vollständigen Fragebogendaten (s. Abschnitt 3.3.1)⁴. Beide Wald-Tests verfehlten statistische Signifikanz (Leseverstehen: $\text{Wald-Test}_{[df=4]} = 0.136$, $p = 0.998$; Hörverstehen: $\text{Wald-Test}_{[df=4]} = 0.140$, $p = 0.998$), was also gegen eine statistisch belastbare Veränderung der profilbezogenen Koeffizienten spricht.

5 Diskussion

Im Kontext der Reform des Systems von Grund- und Leistungskursen zugunsten der Profileroberstufe wurde die Frage untersucht, inwieweit die Beschulung an den fünf Profilen des allgemeinbildenden Gymnasiums in Schleswig-Holstein unterschiedliche Leistungsentwicklungen in der ersten Fremdsprache Englisch bedingt. Mehrere Erklärungen für eventuelle Profileffekte wurden herausgestellt. In drei Schritten wurde eine Überprüfung der Transferhypothese approximiert, die insbesondere einen Ausschluss der konkurrierenden Hypothese leistungsbezogener Kompositionseffekte vorsah. Zudem erfolgte eine Untersuchung übergreifender Leistungsentwicklungen, d.h. am allgemeinbildenden Gymnasium insgesamt, von der 9. bis hin zur 13. Jahrgangsstufe. Über vergleichbare Arbeiten für das Bundesland Hamburg (LAU, KESS) hinausgehend erfolgte diesbezüglich erstmalig eine Unterscheidung der zwei Teildomänen des Lese- und Hörverstehens.

5.1 Zusammenfassung und Diskussion zentraler Befunde

Die deskriptiven Analysen zu den an den fünf Profilen in der Qualifikationsphase besuchten Fremdsprachen (*Fra-*

gestellung 1) unterstrichen – im dritten Jahrgang der im Schuljahr 2008/2009 in Schleswig-Holstein eingeführten Profileroberstufe – eine Heterogenität von fremdsprachenbezogenen Lerngelegenheiten. Tatsächlich besuchten Schülerinnen und Schüler des sprachlichen Profils in der Mehrzahl drei, des naturwissenschaftlichen und sportlichen Profils in der Mehrzahl nur eine (hier: die erste schulisch erlernte) Fremdsprache. Am gesellschaftswissenschaftlichen und ästhetischen Profil wurden zu etwa gleichen Teilen eine bzw. zwei Fremdsprachen besucht. Die Fremdsprachenbelegung korrespondierte nur bedingt mit den deskriptiven Befunden zu Leistungsentwicklungen von der 11. bis hin zur 13. Jahrgangsstufe. Zwar wiesen das naturwissenschaftliche und das sportliche Profil im Leseverstehen die numerisch niedrigsten Zuwächse aus, im Hörverstehen war dieser Trend so jedoch nicht auszumachen. In beiden Domänen zeigten sich am sprachlichen Profil die deskriptiv höchsten Zuwachsraten. In der Gesamtgruppe des allgemeinbildenden Gymnasiums ergab sich über den gesamten Zeitraum von der 9. bis hin zur 13. Jahrgangsstufe (*Fragestellung 2*) im Leseverstehen eine merkliche Dezeleration der Leistungsentwicklung. Im Hörverstehen blieben Leistungszuwächse im selben Zeitraum weitgehend stabil. Ferner konnten deutliche Unterschiede im absoluten Leistungsniveau ausgemacht werden, die im Leseverstehen insofern als umso gravierender zu bewerten sind, als dass dort im Verlauf der Sekundarstufe II nur noch sehr geringe Leistungszuwächse zu verzeichnen waren.

Mit Blick auf die Annäherung an eine Überprüfung der Transferhypothese ist festzuhalten, dass bei Kontrolle individueller Ausgangsleistungen (sowie der Intelligenz und des Geschlechts) an den fünf Profilen in der Tat unterschiedliche Leistungsentwicklungen in der ersten Fremdsprache Englisch ermöglicht wurden (*Fragestellung 3a*). Obwohl hierdurch nur ein vergleichsweise kleiner Anteil der Gesamtvarianz erklärt werden konnte, erscheinen die Gruppenunterschiede praktisch durchaus bedeutsam, legte Abbildung 1 doch nahe, dass in den oberen beiden Jahrgangsstufen in der Gesamtgruppe mit jährlichen Leistungszuwächsen von neun (Lese-) bzw. 37 Punkten (Hörverstehen) zu rechnen ist. Die Profileffekte waren robust gegenüber Effekten der leistungsbezogenen Komposition (*Fragestellung 3b*), die – anders als in der Literatur üblicherweise beschrieben – negativ ausfielen. Wir schlagen vor, die negativen Effekte der leistungsbezogenen Komposition als «Sättigungseffekte» zu interpretieren. Demnach würden Englischlehrkräfte in Klassen mit bereits hohen Lernständen in Jahrgangsstufe 11 in den nachfolgenden zwei Schuljahren weniger Zeit auf eine fortwährende Schulung der von der Kultusministerkonferenz beschriebenen rezeptiven kommunikativen Kompetenzen (KMK, 2012) verwenden.

Nachfolgend wurde ein *direkter Test* der Transferhypothese beabsichtigt (*Fragestellung 3c*). Ein solch direkter

4 Vollständige Befunde auf Anfrage beim Erstautor erhältlich.

Nachweis wäre u.E. dann erfolgt, wenn die zusätzliche Berücksichtigung der Fremdsprachbelegungen zu einem vollständigen Verschwinden von Profilizugehörigkeitseffekten auf Leistungszuwächse geführt hätte. Dieser direkte Nachweis ist nicht gelungen, d.h. auf Klassenebene verblieben numerisch unveränderte Effekte der Profilizugehörigkeit. Dass die Profileffekte aufgrund erhöhter Standardfehler nun die statistische Signifikanz verfehlten, illustriert allerdings eine sehr enge Verknüpfung der Fremdsprachbelegungen und profilizugehörigen Zuwachsraten. Diese Interpretation wird unterstrichen durch den Umstand, dass sich die Genauigkeit der ebenfalls auf Klassenebene erfolgten Kompositionseffekte-Schätzungen durch den Einbezug der Fremdsprachbelegungen nicht veränderte.

5.2 Limitationen der Studie

Einschränkend ist zunächst festzuhalten, dass die Deskription belegter Fremdsprachen sowie die Untersuchung von Fragestellung 3c nur für eine Teilstichprobe der untersuchten Schülerinnen und Schüler erfolgen konnte (zur Begründung s. Abschnitt 3.1.1). Weiterhin war es nicht möglich, profilizugehörige Zuwachsraten mit tatsächlichen Fremdsprachenfolgen (s. die *Anmerkung* zur Tabelle 2) in Beziehung zu setzen. Entsprechende Angaben involvierten offene Angaben im Fragebogen, für die die Anzahl gültiger Werte weiter absank, wobei die dann resultierende Teilstichprobe von $n_{\text{ungewichtet}} = 377$ als zu klein für Mehrgruppenanalysen erschien. Aus dem gewählten Umgang mit fehlenden Werten in Leistungstest und Fragebogen resultiert als weitere Einschränkung, dass auf Individual-ebene zusätzlich denkbare Kontrollvariablen (z.B. sozio-ökonomischer Status, Fachinteresse) nicht berücksichtigt werden konnten.

Durch die unter 3.1 beschriebene Strategie zur Stichprobenauswahl wird die Belastbarkeit unserer Aussagen zusätzlich limitiert. Einschränkungen resultieren ebenfalls aus einer nicht erfolgten Berücksichtigung von Testwiederholungseffekten auf Leistungszuwächse (z.B. Freund & Holling, 2011), die zumindest für *einen Teil* der Schülerinnen und Schüler in Rechnung zu stellen sind, da eine Teilmenge der bereits zu LISA 5 im *Multi-Matrix-Design* (d.h. jede Schülerin und jeder Schüler erhält einen Teil aller Aufgaben) eingesetzten Aufgaben in LISA 6 als für alle Schülerinnen und Schüler gleichförmiger Test erneut verwendet wurde. Ferner resultiert eine Einschränkung unserer Aussagen dahingehend, dass die auf Klassenebene lokalisierten 13 % Varianz konfundierte Varianz von Klassenverbänden geschachtelt in Schulen darstellen. Da die von uns untersuchten Fragestellungen keine sinnvoll auf Schulebene modellierbaren Prädiktoren bereithielten, verzichteten wir auf ein Drei-Ebenen-Modell. Mit Blick auf

weiterführende Studien soll jedoch nicht ausgeschlossen werden, dass etwa der oben beschriebene Mechanismus der Lehrkraftzuweisung tatsächlich ein schulspezifischer sein könnte (d.h. Englischlehrkräfte *zwischen Schulen* unterscheiden sich systematisch z.B. in ihrer Ausbildung⁵). Abschließend sei herausgestellt, dass der gewählte Analysezugang Leistungsuntergruppen innerhalb von Profilen (z.B. besonders leistungsschwache Schülerinnen und Schüler profitieren in besonderer Weise etwa von den Angeboten am sprachlichen Profil) unberücksichtigt ließ.

5.3 Zur Frage differenzieller Leistungsentwicklungen im Lese- und Hörverstehen

Ausgehend von den Befunden zu Fragestellung 2 lag die Vermutung nahe, die Lehrplanvorgaben für den Sekundarbereich II Schleswig-Holsteins räumten der Beschulung beider Teildomänen einen unterschiedlichen Stellenwert ein. Eine nachgeordnete Analyse des – aus der Zeit vor der Einführung der Profileroberstufe stammenden, aber nach wie vor gültigen – Lehrplans für das Fach Englisch spricht gegen diese Annahme. So beschreibt der Lehrplan (s. MBWFK [Ministerium für Bildung, Wissenschaft, Forschung und Kultur des Landes Schleswig-Holstein], 2002) übergreifend, der Kompetenzzuwachs von Schülerinnen und Schülern zeige sich unter anderem in deren größer werdenden Fähigkeit, «komplexe Lese- und Hörtexte [...] zu verstehen und zu bewerten» (MBWFK, 2002, S. 32). Er benennt unter «Lesen/Leseverständnis» (MBWFK, 2002, S. 34) eindeutig diejenigen Teilfähigkeiten, die auch bei der Operationalisierung des hier verwendeten Tests (vgl. Rupp et al., 2008) berücksichtigt wurden. Für die Schulung der «Text- und Medienkompetenz» seien ferner alle literarischen Gattungen zu berücksichtigen, wobei mindestens drei Ganzschriften zu lesen seien. Auch müssten explizit nicht-literarische, authentische Texte Berücksichtigung finden. Der zugrunde gelegte Textbegriff rekurriert demnach bewusst auf die jenseits schulischer Ausbildung liegende Relevanz der Fremdsprache Englisch u. a. für den internationalisierten Arbeitsmarkt, die sich auch in der Anlage der Bildungsstandards findet. Folglich seien *alle* Vermittlungsformen von Text (d.h. Druck, Bild, Ton, elektronische Medien) sowie «mehrfach-kodierte» Texte (d.h. z.B. Filme) in den Unterricht mit einzuschließen. Dies schlägt die Brücke zum Hörverstehen, zu dem sich ähnlich umfangreiche Angaben finden, ohne dass allerdings die Angabe erfolgen würde, dass das eine gegenüber dem anderen stärker zu gewichten sei.

Eine gemäß Lehrplanvorgaben unterschiedliche gewichtete Beschulung beider Teildomänen ist somit unwahrscheinlich. Alternativ scheint denkbar, dass eine Beschulung im englischsprachigen Leseverstehen auch

5 Entsprechende Befunde für den Sekundarbereich I (d.h. für einen Vergleich von Englischlehrkräften an Haupt-, Real-, Gesamtschulen und Gymnasien) wurden von Helmke et al. (2008) vorgelegt.

in gymnasialen Bildungsbiographien zeitlich früher ansetzt als die im Hörverstehen. Ausgehend hiervon wäre plausibel, wenn Leseleistungen auch zeitlich früher ein «Decken»-Niveau erreichen. Retelsdorf et al. (2012) belegten eine solche Dezeleration für das muttersprachliche Leseverstehen bereits in der Sekundarstufe I. Ergänzend könnte abschließend argumentiert werden, die im Hörverstehen höheren Leistungszuwächse seien gar nicht zwingend *allein* ein Ergebnis innerschulischer Lernprozesse, sondern der oben beschriebenen *privaten* Nutzung medialer Angebote (v.a. Fernsehserien, Filme) geschuldet. Inwieweit sich diese Annahme empirisch bestätigen lässt, müssen nachfolgende Studien zeigen.

Literatur

- Alderson, J. C. (1984). Reading in a foreign language: A reading or a language problem? In J.C. Alderson & A.H. Urquhart (Eds.), *Reading in a foreign language* (pp. 1–24). London: Longman.
- Alexander, P. A., Schallert, D. L. & Hare, V. C. (1991). Coming to terms: How researchers in learning and literacy talk about knowledge. *Review of Educational Research*, 61, 315–343.
- Anderson, N. J. (1991). Individual differences in strategy use in second language reading and testing. *Modern Language Journal*, 75, 460–472.
- Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A. et al. (2010). Teachers mathematical knowledge, cognitive activation in the classroom, and student progress. *American Educational Research Journal*, 47, 133–180.
- Birch, B. (2007). *English L2 reading: Getting to the bottom*. Mahwah, N.J.: Lawrence Erlbaum Associates.
- Carr, T. H. & Levy, B. (Eds.). (1990). *Reading and its development: Component skills approaches*. San Diego: Academic Press.
- Cummins, J. (2000). *Language, power, and pedagogy: Bilingual children in the crossfire*. Clevedon, UK: Multilingual Matters.
- Dewaele, J.-M. & Pavlenko, A. (2002). Emotion vocabulary in interlanguage. *Language Learning*, 52, 263–322.
- Dumont, H., Neumann, M., Maaz, K. & Trautwein, U. (2013). Die Zusammensetzung der Schülerschaft als Einflussfaktor für Schülerleistungen. *Psychologie in Erziehung und Unterricht*, 60, 198–213.
- Europarat. (2001). *Gemeinsamer europäischer Referenzrahmen für Sprachen: lernen, lehren, beurteilen*. Berlin: Langenscheidt.
- Freund, P. A. & Holling, H. (2011). Retest effects in matrix test performance: Differential impact of predictors at different hierarchy levels in an educational setting. *Learning and Individual Differences*, 21, 597–601.
- Georgiou, G. K., Parrila, R. & Papadopoulos, T. C. (2008). Predictors of word decoding and reading fluency across languages varying in orthographic consistency. *Journal of Educational Psychology*, 100, 566–580.
- Harris, D. N. (2010). How do school peers influence student educational outcomes? Theory and evidence from economics and other social sciences. *Teachers College Record*, 112, 1163–1197.
- Heller, K. A. & Perleth, C. (2000). *Kognitiver Fähigkeits-Test (Revision) für 5.–12. Klassen (KFT 5–12+R)*. Göttingen: Beltz-Testgesellschaft.
- Helmke, A., Helmke, T., Schrader, F.-W., Wagner, W., Nold, G. & Schröder, K. (2008). Soziodemografische und fachdidaktisch relevante Merkmale von Englischlehrpersonen. In DESI-Konsortium (Hrsg.), *Unterricht und Kompetenzerwerb in Deutsch und Englisch. Ergebnisse der DESI-Studie*. Weinheim: Beltz.
- Hill, C. J., Bloom, H. S., Black, A. R. & Lipsey, M. W. (2008). Empirical benchmarks for interpreting effect sizes in research. *Child Development Perspectives*, 2, 172–177.
- Hox, J. J. (2010). *Multilevel analysis. Techniques and applications* (2nd ed.). New York: Routledge.
- Jonkmann, K., Köller, O. & Trautwein, U. (2007). Englischleistungen am Ende der Sekundarstufe II. In U. Trautwein, O. Köller, R. Lehmann & O. Lüdtke (Hrsg.), *Schulleistungen von Abiturienten: Regionale, schulformbezogene und soziale Disparitäten* (S. 113–142). Münster: Waxmann.
- Klieme, E. (2006). Zusammenfassung zentraler Ergebnisse der DESI-Studie. Zugriff am 05. Januar 2015: www.kmk.org/.../2006/2006_03_01-DESI-Ausgewahlte-Ergebnisse.pdf
- KMK (Ständige Konferenz der Kultusminister der Länder der Bundesrepublik Deutschland). (2004). *Bildungsstandards für die erste Fremdsprache (Englisch / Französisch) für den Mittleren Schulabschluss*. Beschluss vom 04.12.2003. München: Luchterhand.
- KMK. (2012). *Bildungsstandards für die fortgeführte Fremdsprache (Englisch/Französisch) für die Allgemeine Hochschulreife*. Beschluss der Kultusministerkonferenz vom 18.10.2012. München: Luchterhand.
- Koda, K. (2005). *Insights into second language reading: A cross-linguistic approach*. Cambridge: Cambridge University Press.
- Kolen, M. J. & Brennan, R. L. (2004). *Test equating, scaling, and linking: Methods and practices*. Berlin, Germany: Springer.
- Köller, O. & Baumert, J. (2012). Schulische Leistungen und ihre Messung. In W. Schneider & U. Lindenberger (Hrsg.), *Entwicklungspsychologie* (7. Aufl., S. 645–661). Weinheim: Beltz/PVU.
- Köller, O., Knigge, M. & Tesch, B. (Hrsg.). (2010). *Sprachliche Kompetenzen im Ländervergleich. Überprüfung der Erreichung der Bildungsstandards für den Mittleren Schulabschluss für Deutsch und die erste Fremdsprache in der neunten Jahrgangsstufe*. Münster: Waxmann.
- Kürschner, C. & Schnotz, W. (2008). Verhältnis gesprochener und geschriebener Sprache bei der Konstruktion mentaler Repräsentationen. *Psychologische Rundschau*, 59, 139–149.
- Leucht, M., Retelsdorf, J., Möller, J. & Köller, O. (2010). Zur Dimensionalität rezeptiver Kompetenzen im Fach Englisch. *Zeitschrift für Pädagogische Psychologie*, 24, 123–138.
- Marsh, H. W., Lüdtke, O., Robitzsch, A., Trautwein, U., Asparouhov, T., Muthén, B. et al. (2009). Doubly-latent models of school contextual effects: Integrating multilevel and structural equation approaches to control measurement and sampling error. *Multivariate Behavioral Research*, 44, 764–802.
- MBWFK (Ministerium für Bildung, Wissenschaft, Forschung und Kultur des Landes Schleswig-Holstein). (2002). *Lehrplan für die Sekundarstufe II (Gymnasium, Gesamtschule, Fachgymnasium): Englisch*. Kiel: Autor.
- Möller, J., Zaunbauer, A. & Leucht, M. (2010). Fremdsprachenlernen. In D. H. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (4. Aufl., S. 219–225). Weinheim: Beltz.
- Muthén, L. K. & Muthén, B. O. (1998–2012). *Mplus User's Guide. Seventh Edition*. Los Angeles, CA: Muthén & Muthén.
- Neumann, M., Schnyder, I., Trautwein, U., Niggli, A., Lüdtke, O., & Cathomas, R. (2007). Schulformen als differenzielle Lernmilieus: Institutionelle und kompositionelle Effekte auf

- die Leistungsentwicklung im Fach Französisch. *Zeitschrift für Erziehungswissenschaft*, 10, 399–420.
- Perfetti, C. A. (1999). Comprehending written language: A blueprint of the reader. In C. Brown & P. Hagoort (Eds.), *The neurocognition of language* (pp. 167–208). Oxford, N.Y.: Oxford University Press.
- Retelsdorf, J., Becker, M., Köller, O. & Möller, J. (2012). Reading development in a tracked school system: A longitudinal study over 3 years using propensity score matching. *British Journal of Educational Psychology*, 82, 647–671.
- Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. New York: Wiley.
- Rupp, A. A., Vock, M., Harsch, C. & Köller, O. (2008). *Developing standards-based assessment tasks for English as a first foreign language – Context, processes, and outcomes in Germany*. Münster: Waxmann.
- Schoonen, R., Hulstijn, J., & Bossers, B. (1998). Language-dependent and language-independent knowledge in native and foreign language reading comprehension: An empirical study among Dutch students in Grades 6, 8 and 10. *Language Learning*, 48, 71–106.
- Schwanenflugel, P. J., Blout, P. J. & Lin, P.-J. (1991). Cross-cultural aspects of word meaning. In P. J. Schwanenflugel (Ed.), *The psychology of word meanings* (pp. 74–90). Hillsdale, N.J.: Erlbaum.
- Sigott, G. (2004). *Towards identifying the C-Test construct*. Frankfurt am Main: Peter Lang.
- Statistisches Bundesamt. (2014). *Fachserie 11, Reihe 1. Bildung und Kultur – Allgemeinbildende Schulen. Schuljahr 2012/13* (korrigierte Fassung vom 14. Februar 2014). Wiesbaden: Autor.
- Trautwein, U., Köller, O., Lehmann, R., & Lüdtke, O. (Hrsg.). (2007). *Schulleistungen von Abiturienten: Regionale, schulformbezogene und soziale Disparitäten*. Münster: Waxmann.
- van Gelderen, A., Schoonen, R., De Stool, R. D., De Gloppe, K. & Hulstijn, J. (2007). Development of adolescent reading comprehension in language 1 and language 2: A longitudinal analysis of constituent components. *Journal of Educational Psychology*, 99, 477–491.
- Vieluf, U., Ivanov, S. & Nikolova, R. (2011). *Kompetenzen und Einstellungen von Schülerinnen und Schülern an Hamburger Schulen am Ende der Sekundarstufe I und zu Beginn der gymnasialen Oberstufe*. Hamburg: Behörde für Schule und Berufsbildung.
- Wu, M. L., Adams, R. J., Wilson, M. & Haldane, S. A. (2007). *ACER ConQuest. Version 2.0. Generalised Item Response Modeling Software* [Computer software]. Camberwell: ACER Press.

Dr. Michael Leucht

Leibniz-Institut für die Pädagogik der Naturwissenschaften und der Mathematik (IPN)
Christian-Albrechts Universität zu Kiel
Olshausenstraße 62
24118 Kiel
Deutschland

leucht@ipn.uni-kiel.de