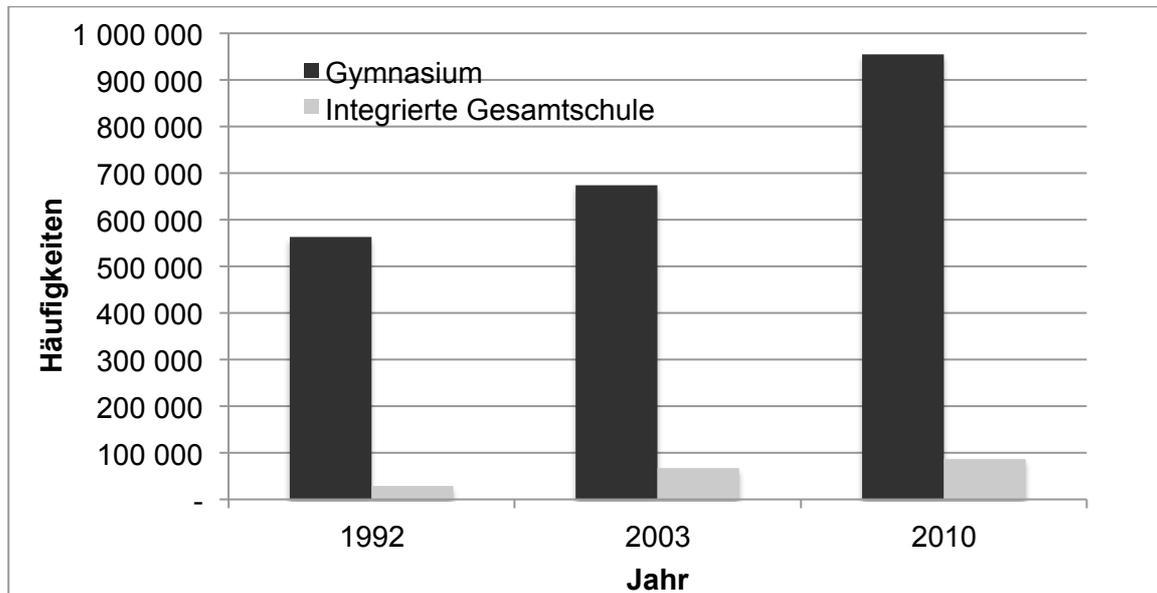


Wege zur Hochschulreife und Leistungen von Abiturienten

1 Einleitung

Als Folge der Bildungsexpansion, die in den 1950er Jahren in den alten Bundesländern einsetzte und zu einem massiven Ausbau des Gymnasiums führte, ist die Nachfrage nach höheren Bildungsabschlüssen erheblich angestiegen. So konnten im Jahre 2010 über 40 Prozent der 25 bis 30-jährigen eine fachgebundene oder allgemeine Hochschulreife nachweisen. In der Gruppe der 60 bis 65-jährigen betrug dieser Anteil lediglich etwas über 20 Prozent. Umgekehrt wiesen in dieser Altersgruppe noch rund 50 Prozent einen Hauptschulabschluss auf, in der Gruppe der 25-30-jährigen waren es lediglich rund 20 Prozent. Die Zunahme in der Nachfrage nach höheren Bildungszertifikaten scheint ungebrochen zu sein; dies bestätigt die Abbildung 1, in der die Schülerzahlen in der Sekundarstufe II an allgemeinbildenden Gymnasien und integrierten Gesamtschulen zwischen 1992 und 2010 abgetragen sind.

Abbildung 1: Schülerinnen und Schüler (absolute Häufigkeiten) in der Sekundarstufe II nach Jahr und Schulform

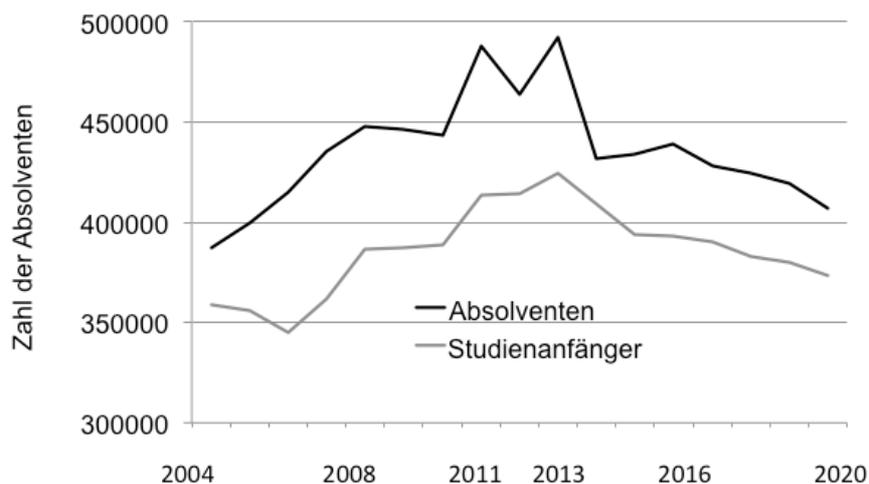


Quelle: Statistisches Bundesamt, Fachserie 11, Bildung und Kultur. Allgemeinbildende Schulen, 2011.

Befanden sich im Jahr 1992 noch unter 600.000 Schülerinnen und Schüler in einer Sekundarstufe II an allgemeinbildenden Gymnasien, so ist die Zahl im Jahr 2010 auf über 950.000 angestiegen. Dieser Trend zeigt sich auch auf niedrigerem Niveau in integrierten Gesamtschulen, er wird sich allerdings in den kommenden Jahren aufgrund des starken Rückgangs der Geburtenzahlen abschwächen bzw. längerfristig sogar umkehren. Dies belegt beispielsweise die Prognose der Studienanfängerzahlen durch die Kultusministerkonferenz (KMK, 2009), die in Abbildung 2 dargestellt ist. Danach wird der Höhe-

punkt mit Doppelabiturjahrgängen in einigen Bundesländern im Jahre 2013 erreicht. Anschließend werden die Zahlen absinken, allerdings über dem Niveau von 2004 bleiben.

Abbildung 2: Absolventinnen/Absolventen mit allgemeiner bzw. fachgebundener Hochschulreife und Studienanfängerzahlen in Deutschland nach Jahr

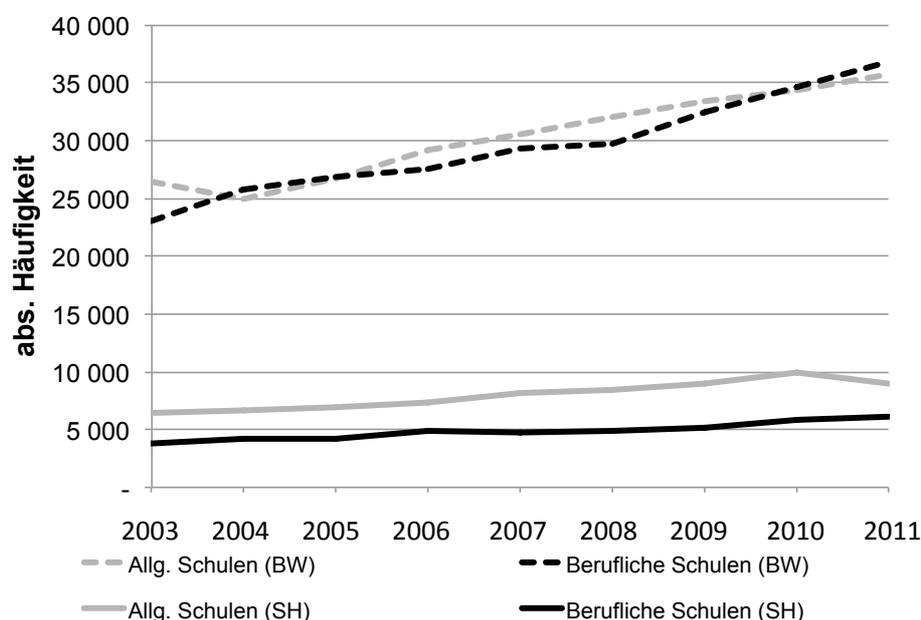


Quelle: Prognose KMK, 2009.

Der Anstieg der Absolventen mit Hochschulzugangsberechtigung ist nicht allein der Expansion des allgemeinbildenden Gymnasiums geschuldet. Vielmehr sind die Wege zur Hochschulreife insgesamt flexibler geworden (vgl. z. B. Köller et al., 2004). Im Schuljahr 2009/2010 erhielten rund 25 Prozent aller Abiturienten ihre Hochschulzugangsberechtigung nicht mehr an einem allgemeinbildenden Gymnasium. Die quantitativ wichtigsten alternativen Zugänge zur Hochschulreife sind dabei das berufliche Gymnasium und die integrierte Gesamtschule. Beispielsweise erwarben in Schleswig-Holstein im Jahr 2010 rund 37 Prozent der Absolventen ihre fachgebundene oder allgemeine Hochschulreife in beruflichen Schulen. Die Abbildung 3 zeigt diese Zahlen im Trend und ergänzt sie um die Statistik aus Baden-Württemberg, wo mittlerweile mehr Hochschulzugangsberechtigungen im berufsbildenden System erworben werden.

Ebenfalls im Schuljahr 2009/2010 stammten 18 Prozent der Absolventen mit Hochschulreife in Nordrhein-Westfalen von integrierten Gesamtschulen. Darüber hinaus hat sich mit den Strukturreformen in allen Bundesländern nach PISA 2000 (Deutsches PISA-Konsortium, 2001) im allgemeinbildenden Sekundarschulsystem die Tendenz zu einem zweigliedrigen Schulsystem herauskristallisiert, in dem auch die nichtgymnasiale Säule teilweise in der Sekundarstufe II ausgebaut ist, sodass sie einen alternativen Weg zur Hochschulreife bietet. In Schleswig-Holstein stehen die Gemeinschaftsschulen für diese Entwicklung. Somit sind vielfältige Wege entstanden, welche die Hochschulzugangsberechtigung vergeben.

Abbildung 3: Absolventinnen/Absolventen mit allgemeiner bzw. fachgebundener Hochschulreife nach Bildungsgang, Land und Jahr



Quelle: Statistisches Bundesamt, Fachserie 11, Bildung und Kultur. Allgemeinbildende Schulen, 2011; SH: Schleswig-Holstein; BW: Baden-Württemberg

Mit der Expansion des Gymnasiums, den alternativen Bildungsgängen und den in Folge deutlich angewachsenen Anteilen von jungen Erwachsenen mit fachgebundener und allgemeiner Hochschulreife stellt sich die Frage, ob bei einer so weit fortgeschrittenen Öffnung überhaupt noch fachliche Standards, welche die Vergabe der Hochschulreife legitimieren, von breiten Teilen der Schülerschaft erreicht werden. Letztendlich verbindet sich damit auch die Frage nach der Studierfähigkeit der Absolventen. Im Rahmen dieses Beitrags werden diese Gedanken aufgenommen und Befunde vorgestellt, die folgende Fragen beantworten sollen:

- Wie wichtig sind schulisch erworbene Kompetenzen für die Studierfähigkeit?
- Wie sichert die Sekundarstufe II die Erreichung von fachlichen Standards?
- Was wissen wir über die Leistungen von Abiturienten in Deutschland?

2 Zur Relevanz voruniversitärer Leistungen für den Studienerfolg

Es existieren mittlerweile viele empirische Studien, die zeigen können, dass voruniversitär erworbene schulische Kompetenzen gute Indikatoren für die allgemeine Studierfähigkeit sind und Studienleistungen bzw. Studienerfolg vorhersagen können (im Überblick Formazin, Schroeders, Köller, Wilhelm & Westmeyer, 2011). Studierfähigkeit umfasst das, was den Schülerinnen und Schülern inhaltlich und methodisch aus der gymnasialen Oberstufe mitgegeben wird (vgl. KMK, 1977/1978) und sie auf das wissenschaftliche Arbeiten/Studieren an der Universität vorbereitet. Aus einer bildungstheoretischen Perspektive konstituiert sich Studierfähigkeit aus vertiefter Allgemeinbildung in den Kernfä-

chern Deutsch, Mathematik und fortgeführter Fremdsprache und der Wissenschaftspropädeutik, wobei letzterer eine Schlüsselrolle zukommt (vgl. Trautwein & Lüdtke, 2004): „Die Wissenschaftspropädeutik hat die Funktion der Enkulturation: Die Gymnasiastinnen und Gymnasiasten wachsen hierdurch in die Kultur der Wissenschaft hinein. Wissenschaftspropädeutik meint Bildung zur Wissenschaft und zielt auf eine Haltung, die dem Einzelnen Wissenschaft öffnet und erschließt und den Blick dafür schärft“ (Schmidt, 1991, S. 200). Weitere Kriterien einer hinreichenden Studieneignung sind nach Einschätzung der KMK-Expertenkommission von 1995 sogenannte „Schlüsselqualifikationen“ in den Bereichen der (1) sozialen Kompetenz, (2) Selbststeuerung des Lernens und (3) Eigenverantwortlichkeit.

Interessanterweise wird diese bildungstheoretische Diskussion von den Abnehmern an Universitäten kaum explizit aufgenommen. Dort wird das Konzept der Studierfähigkeit pragmatischer gehandhabt und bezieht sich auf kognitive und motivationale Merkmale der Studienanfänger. Die Tabelle 1 zeigt die Ergebnisse einer Befragung von Professorinnen und Professoren (Konegen-Grenier, 2002).

Tabelle 1: Ergebnisse einer Befragung von Professorinnen und Professoren zu Faktoren der Studierfähigkeit

Kognitive Merkmale	Persönlichkeits- und Motivationsmerkmale
<ul style="list-style-type: none"> ▪ Analytische Fähigkeiten ▪ Abstraktionsfähigkeit ▪ Differenzierungsvermögen ▪ Synthesefähigkeit ▪ Transferfähigkeit ▪ Kreativität ▪ Sprachliche Ausdruckfähigkeit 	<ul style="list-style-type: none"> ▪ Inhaltliches Interesse ▪ Leistungsmotivation ▪ Genauigkeit ▪ Zielstrebigkeit ▪ Beharrlichkeit ▪ Eigeninitiative ▪ Selbstorganisation ▪ Hohes Anspruchsniveau ▪ Kompetenter Umgang mit Unsicherheit ▪ Fähigkeit zur Selbstreflexion

Quelle: Konegen-Grenier, 2002.

Auf Seiten der kognitiven Fähigkeiten werden im Wesentlichen drei Faktoren genannt, Intelligenz, Kreativität und sprachliche Kompetenzen, auf Seiten der nicht-kognitiven Personmerkmale sind es zusammengefasst Motivation und Selbstregulation, welche die Studierfähigkeit ausmachen. Passend hierzu berichtet Heldmann (1984) eine Befragung, in der Hochschullehrerinnen und Hochschullehrer von ihren Studentinnen und Studenten (1) eine starke Lern- und Leistungsbereitschaft, (2) eine hohe Selbständigkeit und Motivation, (3) besondere kommunikative Fähigkeiten, (4) eine hohe Ausdauer und Belastbarkeit sowie (5) intellektuelle Neugier erwarten. Fachliche Kompetenzen spielen hier eine untergeordnete Reihe. Auf der andere Seite hat der Deidesheimer Kreis (1997) in seinen Ausführungen zur „Feststellung der Studieneignung im Rahmen der Hochschulzulassung“ auf die besondere Rolle der Fachleistungen hingewiesen. In ihren Ratschlägen zur Verwendung von Instrumenten zur Diagnose der Studieneignung argumentieren die Mitglieder dieses Kreises, dass Studienfächer je nach Inhalt gänzlich unterschiedliche

Anforderungen stellen und fachspezifische gegenüber globalen Verfahren zur Feststellung der Studierfähigkeit vorzuziehen sind. In diesem Sinne können auch Befunde interpretiert werden, die sich in Validierungsuntersuchungen ergaben, die ergänzend zur TIMS-Studie durchgeführt wurden (vgl. Baumert et al., 2000).

2.1 Abiturnote und Studienerfolg

Die Abiturnote kann als genereller Indikator voruniversitärer Fachleistungen verstanden werden, wobei unbestritten ist, dass neben kognitiven auch motivationale und andere lernförderliche Einflüsse in die Abiturnote eingehen (vgl. Köller & Baumert, 2002). Bei aller Kritik an der Vergleichbarkeit von Abiturnoten ist es verblüffend, welche substantielle Prädiktionskraft sie für Studienerfolge haben. In nationalen Metaanalysen zeigte sich, dass sie die spätere Studienleistung mit $r = .31$ (Trapmann, Hell, Weigand & Schuler, 2007) bzw. $r = .35$ (Baron-Boldt, Schuler & Funke, 1988; Schuler, Funke & Baron-Boldt, 1990) gut vorhersagen konnten. Nach Korrekturen für die Unreliabilität des Kriteriums und die Varianzeinschränkungen stiegen diese Werte auf $r = .52$ (Trapmann et al., 2007) bzw. $r = .46$ (Schuler et al., 1990). Internationale Befunde ergeben ein vergleichbares Bild. Für Varianzeinschränkung korrigierte Korrelationen aus Validitätsuntersuchungen lagen bei $r = .54$ (Bridgeman, McCamley-Jenkins & Ervin, 2000; Kobrin et al., 2008) bzw. $r = .59$ (Ramist, Lewis & McCamley-Jenkins, 2001) zwischen der Schulabschlussnote und der durchschnittlichen Studiennote nach dem ersten Studienjahr. Einzelnoten sind generell weniger prädiktiv als die Abiturdurchschnittsnote (Baron-Boldt et al., 1988; Formazin et al., 2011).

2.2 Leistungstests und Studienerfolg

Standardisierte Leistungstests wurden bislang im deutschen Sprachraum eher selten eingesetzt. Belastbare Befunde liegen für den „Test für medizinische Studiengänge TMS“ vor, der von 1986 bis 1997 bei der Zulassung zu medizinischen Studiengängen deutschlandweit eingesetzt wurde. Mit dem TMS wurden in neun Untertests Fähigkeiten erfasst, die für die erfolgreiche Bewältigung des Medizinstudiums als notwendig erachtet wurden. Die Tests umfassten allerdings keine Inhalte, die typischerweise in der gymnasialen Oberstufe unterrichtet werden. Validitätsstudien fanden mittlere Korrelationen von $r = .45$ zwischen der TMS-Leistung und der Note in der ärztlichen Vorprüfung. Die Testleistung band über die Abiturdurchschnittsnote hinaus inkrementelle Varianz im Kriterium Studienleistung ($\Delta R^2 = .07$; Trost et al., 1998; Trost, Klieme & Nauels, 1997).

Hell, Trapmann und Schuler (2007) führten einer Metaanalyse zur Validität von Leistungstests im Rahmen der Studierendenauswahl im deutschen Sprachraum durch. Die Mehrheit der berücksichtigten Studien bezog sich ebenfalls auf den TMS und ergab einen korrigierten Korrelationskoeffizienten von $r = .51$ für den Studiengang Humanmedizin.

International werden Leistungstests sehr viel häufiger eingesetzt. Insbesondere in den USA ist ihr Einsatz breit untersucht worden und es existieren viele Untersuchungen zur prädiktiven Validität dieser Tests. Bei der Auswahl für Bachelor-Studiengänge kommen neben der Abschlussnote (Grade Point Average GPA) an mehr als 3800 Colleges Zulassungstests (SAT I und SAT II) zum Einsatz (The College Board, 2010). Für die Zulas-

sung zu Masterstudiengängen werden an vielen Institutionen neben der Abschlussnote des 1. Studienabschnittes (Undergraduate GPA) ebenfalls Tests eingesetzt (GRE General und GRE Subject Tests; ETS, 2010).

Sowohl der SAT I als auch der GRE General Test erfassen schlussfolgerndes Denken mit numerischem und verbalem Material (Bridgeman et al., 2000; ETS, 2009), d.h. die Aufgaben können eher als Indikatoren der generellen Studierfähigkeit verstanden werden. Für die Vorhersage der Noten nach dem 1. Studienjahr wurden für den SAT I Validitätswerte von $r = .52$ (Bridgeman et al., 2000) bzw. $r = .53$ (Kobrin et al., 2008) berichtet. Für den GRE General Test wurde in einer Metaanalyse die Abschlussnote des Graduiertenstudiums vorhergesagt. Die Korrelationen lagen bei $r = .32$ für den GRE General Test mit quantitativem Material und $r = .36$ für den GRE General Test mit verbalem Material (Kuncel, Hezlett & Ones, 2001) bzw. bei $r = .37$ für den GRE General Test Gesamtscore (Kuncel & Hezlett, 2007).

Im Gegensatz zu den allgemeinen Leistungstests SAT I und GRE General Test erfassen die SAT II: Subject Tests und die GRE Subject Tests domänenspezifisches voruniversitäres Wissen, bspw. in Geschichte oder in einer Fremdsprache (SAT II; Ramist et al., 2001) bzw. in einem spezifischen Fach aus dem 1. Studienabschnitt, z.B. Chemie oder Psychologie (GRE Subject Tests; Kuncel et al., 2001). Die Vorhersagegüte dieser Verfahren ist ebenfalls hoch. Sie lag für die SAT II bei $r = .58$ (Ramist et al., 2001) und für die GRE Subject Tests bei $r = .41$ (Kuncel & Hezlett, 2007; Kuncel et al., 2001).

Dass solche Befunde auf Deutschland übertragbar sind, zeigt die Arbeit von Formazin et al. (2011). In dieser Studie wurden zur Auswahl von Studierenden ($N = 1.049$) im Fach Psychologie Abiturdurchschnittsnoten, globale Indikatoren der Studierfähigkeit (schlussfolgerndes Denken) und Indikatoren voruniversitären Wissens (Textverständnis, Mathematik, Englisch, Biologie) eingesetzt. Von den zugelassenen Studierenden ($N = 79$) lagen später Studienleistungen (Noten in Modulabschlussprüfungen) vor. Die Analyse von Strukturgleichungsmodellen ergab zunächst, dass schlussfolgerndes Denken (standardisierter Regressionskoeffizient $\beta = .32$) und voruniversitäres Wissen ($\beta = .59$) immerhin 44 Prozent der Varianz der Studienleistungen erklärten. Wurde zusätzlich die Abiturdurchschnittsnote berücksichtigt, so stieg die erklärte Varianz auf 52 Prozent. Die Abiturnote allein erklärte 15 Prozent der Varianz. Man kann die Ergebnisse dahingehend zusammenfassen, dass voruniversitäres Wissen bzw. voruniversitäre Kompetenzen, die in der gymnasialen Oberstufe erworben werden, in der Tat eine erhebliche Vorhersagekraft für Studienleistungen haben. Ein hohes Vorwissen steigert die Chance im Studium erfolgreich zu sein, ein geringes Vorwissen senkt diese Chance. Damit verbunden ist natürlich die Forderung der Sicherung von Leistungsstandards in der Sekundarstufe II.

3 Sicherung fachlicher Standards in der gymnasialen Oberstufe

Herkömmliche Steuerungsstrategien in der gymnasialen Oberstufe folgen einer Input-Orientierung, bei der davon ausgegangen wird, dass Vorgaben für Stundenzahlen, Vertiefung von Fächern und die Regularien für die Abiturprüfung hinreichende Leistungs-niveaus auf Seiten der Schülerinnen und Schüler am Ende der Sekundarstufe II sicherstellen. Inwieweit die kürzlich von der Kultusministerkonferenz verabschiedeten Bildungs-

standards für das Abitur in den Fächern Deutsch, Mathematik und fortgeführte Fremdsprachen (Englisch/Französisch) eine Abkehr von dieser Philosophie darstellen werden, ist offen.

In ihren Beschlüssen von 2002 hat die Kultusministerkonferenz die Einheitlichen Prüfungsanforderungen für das Abitur (EPA) revidiert, um auch zukünftig die Qualität der allgemeinen Hochschulreife als schulische Abschlussqualifikation zu sichern und die allgemeine Studierfähigkeit sowie der Übergang in eine berufliche Ausbildung zu gewährleisten. Die EPA sollen für den gesamten Fächerkanon der gymnasialen Oberstufe ein zwischen den Ländern vergleichbares Orientierungswissen, fachliche Standards und Grundkompetenzen sichern, die über den Auftrag des jeweiligen Faches hinausgehen. Die EPA stellen dabei explizit länderübergreifende Festlegungen für die Abiturprüfung dar, indem sie konkrete Lern- und Prüfungsbereiche beschreiben und auch anforderungsspezifische Vorgaben für die Bewertung von Abiturleistungen machen. „Als für die Studierfähigkeit grundlegende Kompetenzen werden die sprachliche Ausdrucksfähigkeit, insbesondere die schriftliche Darlegung eines konzisen Gedankengangs, das verständige Lesen komplexer fremdsprachlicher Sachtexte sowie der sichere Umgang mit mathematischen Symbolen und Modellen hervorgehoben.“ (KMK, 2002, S. 1). Darüber hinaus wird immer wieder die wissenschaftspropädeutische Funktion der gymnasialen Oberstufe betont, mit der Konsequenz, dass selbstständiges Lernen, wissenschaftsorientiertes Arbeiten sowie die Entwicklung der Kommunikationsfähigkeit und der Kooperationsbereitschaft als Kernziele genannt werden. Beibehalten wird die Forderung, dass die gymnasiale Oberstufe den Zugang zu vertiefter Allgemeinbildung ermöglichen muss.

Während die EPA also länderübergreifend einen Orientierungsrahmen für die Abiturprüfung darstellen, bieten die zentralen Abiturprüfungen einen Rahmen, der innerhalb des einzelnen Landes qualitätssichernd sein soll. Identische Aufgaben, die in den Abiturprüfungen eines Landes eingesetzt werden, lassen Unterschiede zwischen Schulen deutlich werden und können der Schulaufsicht Hinweise geben, wo möglicherweise Mindestanforderungen an die Leistungsfähigkeit von Abiturienten unterschritten werden.

4 Studien zu den voruniversitären Leistungen von deutschen Schülerinnen und Schülern

Mit der empirischen Wende in der Erziehungswissenschaft sind Schulleistungsstudien in der Primarstufe und Sekundarstufe I etabliert worden, die detailliert über die Kompetenzstände deutscher Schülerinnen und Schüler in den Kerndomänen schulischer Arbeit (Mathematik, Deutsch, 1. Fremdsprache, Naturwissenschaften) Auskunft geben (vgl. Köller, 2010). Bislang ist es nicht gelungen, ein vergleichbares System in der gymnasialen Oberstufe zu etablieren. Trotz dieser Defizite in der Datenlage liegen seit Mitte der 1990er Jahre auch Ergebnisse für voruniversitäre Leistungen deutscher Abiturientinnen und Abiturienten vor, die auf die Fächer Mathematik und Englisch beziehen und damit zwei der drei Fächer abbilden, für die die KMK insbesondere vertiefte Kenntnisse am Ende der gymnasialen Oberstufe fordert. Die vorliegenden Daten stammen aus der Studie „Bildungsverläufe und psychosoziale Entwicklung im Jugendalter“ (BIJU; vgl. Köller, Baumert & Schnabel, 1999), der Dritten Internationalen Mathematik- und Naturwissen-

schaftsstudie (TIMSS, vgl. Baumert, Bos & Lehmann, 2000), der Studie Transformation des Sekundarschulsystems und akademische Karrieren (TOSCA, vgl. Köller et al., 2004), der Lernausgangslagenuntersuchung in Klasse 13 (LAU 13; vgl. Trautwein et al., 2007) und der TOSCA-Repeat Studie (TOSCA-R; vgl. Trautwein et al., 2010). Die Tabelle 2 gibt einen Überblick über die berücksichtigten Fächer und Kohorten. Die Studien sollen im Folgenden mit einigen Kernbefunden vorgestellt werden. Da sich die Studien TOSCA und TOSCA-R hinsichtlich Anlage und Fragestellungen erheblich ähneln, wird nur die TOSCA-Studie beschrieben.

Tabelle 2: Empirische Studien, in denen voruniversitäre Leistungen von Schülerinnen und Schülern der Sekundarstufe II untersucht wurden

Studie	Erhebungsjahr	Voruniversitäre Leistungsbereiche	Kohorten
BIJU	Frühjahr 1997	Mathematik, Englisch	Schülerinnen und Schüler der 12. Jahrgangsstufe in integrierten Gesamtschulen und allgemeinbildenden Gymnasien in Nordrhein-Westfalen
TIMSS	Frühjahr 1996	Mathematik, Physik	National repräsentative Stichprobe von Abiturienten an Oberstufen unterschiedlicher Schulformen
TOSCA	Frühjahr 2002	Mathematik, Englisch	Repräsentative Stichprobe von Abiturienten an allgemeinbildenden und beruflichen Gymnasien in Baden-Württemberg
LAU 13	Frühjahr 2005	Mathematik, Englisch,	Vollerhebung der Abiturienten aller gymnasialer Oberstufen in der Stadt Hamburg
TOSCA-R	Frühjahr 2006	Mathematik, Englisch	Repräsentative Stichprobe von Abiturienten an allgemeinbildenden und beruflichen Gymnasien in Baden-Württemberg

4.1 Leistungen an Oberstufen integrierter Gesamtschulen und allgemeinbildender Gymnasien in Nordrhein-Westfalen (BIJU)

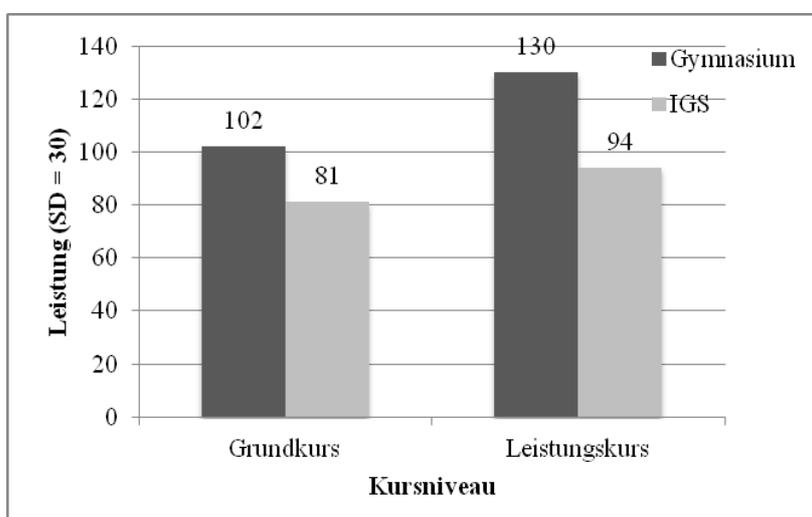
Im Rahmen der BIJU-Studie wurden Bildungskarrieren im Jugend- und frühen Erwachsenenalter in einem Mehrkohorten-Längsschnitt, der eine vergleichende Analyse von Entwicklungsverläufen unter differenziellen institutionellen Rahmenbedingungen erlaubt, untersucht. Eine ausführliche Darstellung der Anlage der Untersuchung findet sich bei Köller et al., 2010), Details zur Oberstufenuntersuchung sind bei Köller et al. (1999) beschrieben. Dort fand bei der Berichterstattung eine Beschränkung auf die Mathematikleistungen statt, die im Folgenden übernommen wird.¹

Die untersuchte nordrhein-westfälische Stichprobe umfasste insgesamt $N = 1.573$ junge Erwachsene (56,5 Prozent Frauen) der 12. Jahrgangsstufe, die im Frühjahr 1997 an

¹ Die Englischergebnisse haben Köller, Baumert, Cortina, Trautwein und Watermann (2004) publiziert. Sie entsprachen weitgehend den berichteten Befunden für das Fach Mathematik.

den Erhebungen teilnahmen. Aus Oberstufen an allgemeinbildenden Gymnasien lagen Daten von $N = 991$ Schülerinnen und Schülern vor; insgesamt $N = 582$ Schülerinnen und Schüler kamen aus 12. Jahrgangsstufen von integrierten Gesamtschulen. Für rund ein Viertel der untersuchten Schülerinnen und Schüler lagen bereits Testleistungen am Ende der Sekundarstufe I vor. Der in der 12. Jahrgangsstufe verwendete Test enthielt teilweise Aufgaben, die curriculare Inhalte der Sekundarstufe I abbildeten, die übrigen Items stammten aus der TIMSS-Oberstufenstudie (Baumert et al., 2000) und thematisierten Inhalte der Sekundarstufe II. Die Leistungen wurden auf einer Skala mit einem Mittelwert von $M = 100$ und einer Streuung von $SD = 30$ abgetragen.

Abbildung 4: Mathematikleistungen in der 12. Jahrgangsstufe in Nordrhein-Westfalen nach Schulform (Gymnasium vs. integrierte Gesamtschule) und Kursniveau



Quelle: Köller et al., 1999.

Ein Kernbefund dieser Untersuchung ist in Abbildung 4 aufgeführt. Es ergaben sich große Leistungsunterschiede zwischen den Schulformen. So erreichten Schülerinnen und Schüler der Mathematik-Leistungskurse an Gesamtschuloberstufen im Mittel nicht das Leistungsniveau der Grundkurse an Gymnasien. Die Leistungen eines erheblichen Anteils der Schülerschaft an Gesamtschulen lagen auf einem Niveau, bei dem auch Mindeststandards, wie man sie in der gymnasialen Oberstufe erwartet, verfehlt wurden. Mindeststandards wurden hier definiert als mittleres Leistungsniveau aller Schülerinnen und Schüler, die im Mathematik-Grundkurs des letzten Zeugnisses die Note „ausreichend“ erhalten hatten. Vor allem die geringen Kompetenzstände in den Leistungskursen der integrierten Gesamtschulen warfen die Frage auf, ob dort die Ziele einer vertieften Allgemeinbildung noch erreicht werden konnten. Ergänzende Analysen wiesen aber auch darauf hin, dass die Schülerinnen und Schüler an integrierten Gesamtschulen schon mit deutlich geringeren Kompetenzen in die Sekundarstufe II übergetreten waren. Weitere Analysen zeigten, dass sich trotz grundlegend unterschiedlicher Leistungsniveaus die Zeugnisnoten zwischen den Schülerinnen und Schülern beider Schulformen kaum unterschieden. Hinter identischen Noten standen also nicht mehr vergleichbare Kompetenzstände. Die Autoren diskutieren dies vor dem Hintergrund einer möglichen Verteilungsungerechtigkeit beim Zugang zu zulassungsbeschränkten Studiengängen.

4.2 Leistungen deutscher Schülerinnen und Schüler in voruniversitärer Mathematik am Ende der gymnasialen Oberstufe – Befunde der TIMS-Studie

Im Rahmen der TIMS-Studie (Baumert et al., 2000) wurde eine repräsentative Stichprobe von Schülerinnen und Schülern am Ende der gymnasialen Oberstufe getestet. Erfasst wurden die Leistungen in voruniversitärer Mathematik und Physik. Die Daten erlauben einen internationalen Vergleich der am Ende der gymnasialen Oberstufe erreichten fachspezifischen Kompetenzen. Da in Deutschland nur solche Schülerinnen und Schüler die Tests bearbeiteten, die auch einen Grund- oder Leistungskurs im jeweiligen Fach besucht hatten, fiel die Stichprobe in Physik relativ klein aus. Es soll daher im Folgenden wiederum eine Beschränkung auf das Fach Mathematik stattfinden.

Der Fachleistungstest für den voruniversitären Mathematikunterricht beinhaltete Aufgaben, die entlang des Lehrstoffs voruniversitärer Fachkurse am Ende der Sekundarstufe II entwickelt wurden (vgl. Klieme, 2000). Zielführend für die Testkonstruktion waren die fachliche Breite und Tiefe sowie die Passung zu den Lerninhalten und Lernzielen in der Sekundarstufe II. Es standen die fachimmanenten Konzepte, Operationen und Gesetzmäßigkeiten selbst im Vordergrund (zu den Details der Testentwicklung vgl. Baumert et al., 2000; Klieme, 2000). Der größte Teil der Aufgaben bezog sich auf Standardstoffe der gymnasialen Oberstufe. Außerdem wurde auf Kenntnisse und Fähigkeiten zurückgegriffen, die bereits in der Mittelstufe vermittelt worden waren, aber in der gymnasialen Oberstufe wieder aufgenommen oder aber in schwierigere Problemstellungen eingebettet werden mussten. Die Leistungen wurden auf einem internationalen Maßstab mit einem Mittelwert von $M = 500$ und einer Streuung von $SD = 100$ abgetragen. Zur Erleichterung der Interpretation der Testwerte wurden vier Kompetenzstufen definiert, die in Tabelle 3 beschrieben sind. Während Schülerinnen und Schüler auf der Stufe II im Wesentlichen den Stoff der Sekundarstufe I beherrschen, wird erst ein Niveau voruniversitären Kompetenzen auf den Stufen III und IV erreicht. Schülerinnen und Schüler auf Stufe I weisen Kompetenzen auf, wie man sie typischerweise nach der 6. Jahrgangsstufe erwartet.

Tabelle 3: Kompetenzniveaus in der TIMSS-Oberstufenuntersuchung und ihre inhaltliche Beschreibung

Schwellenwert	Kompetenzniveau	Zur Lösung der Aufgaben notwendige Operationen
≤ 400	Stufe I	Elementares Schlussfolgern
401 – 500	Stufe II	Anwendung mathematischer Begriffe und Regeln
501 – 600	Stufe III	Anwendung von Lerninhalten der gymnasialen Oberstufe
> 600	Stufe IV	Selbstständiges Lösen mathematischer Probleme auf Oberstufenniveau

Quelle: Watermann et al., 2004.

In TIMSS zeigte sich, dass 22 Prozent der Abiturientinnen und Abiturienten Leistungen auf Stufe I erreichten, 48 Prozent auf Stufe II, 23 Prozent auf Stufe III und 7 Prozent auf Stufe IV. Mit anderen Worten verfehlten rund 70 Prozent der getesteten Schülerinnen und Schüler am Ende der gymnasialen Oberstufe die Lernziele des voruniversitären Mathematikunterrichts.

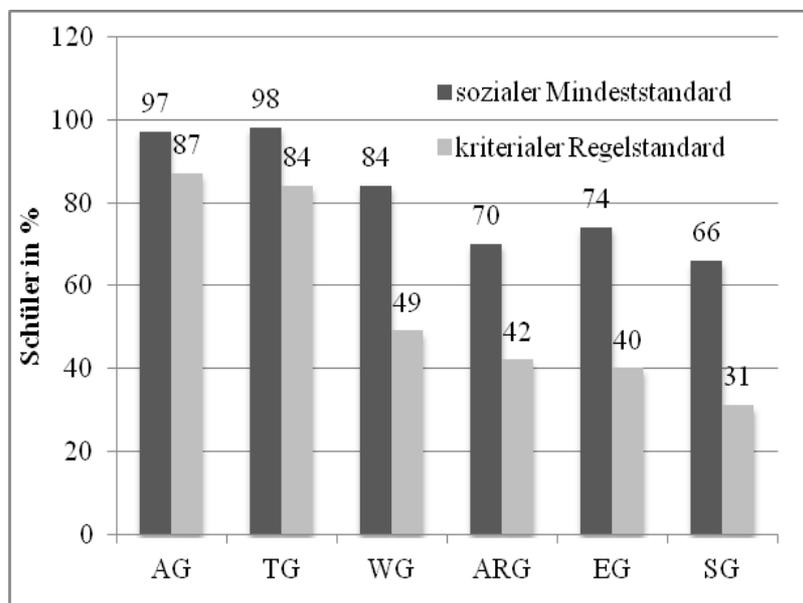
4.3 Befunde zu Leistungsständen an allgemeinbildenden und beruflichen Gymnasien in Baden-Württemberg

Im Land Baden-Württemberg wurden die beruflichen Gymnasien seit den 1970er Jahren als alternative Wege zum Abitur systematisch ausgebaut. Technische, wirtschaftswissenschaftliche, agrarwissenschaftliche, sozialpädagogische, ernährungswissenschaftliche und biotechnologische Zweige vergeben dort die allgemeine Hochschulreife. Berufliche Gymnasien rekrutieren typischerweise eine Schülerschaft, die in der großen Mehrzahl in der Sekundarstufe I einen nicht-gymnasialen Bildungsgang (in der Regel einen, der zu Mittleren Schulabschluss führt) besucht hat. Mittlerweile erwirbt rund ein Drittel aller Abiturienten in Baden-Württemberg die Hochschulzugangsberechtigung an einem beruflichen Gymnasium. Inwieweit die im Abitur erreichten Kompetenzniveaus an beruflichen und allgemeinbildenden Gymnasien vergleichbar sind, haben Köller et al. (2004) in der Studie Transformation des Sekundarschulsystems und akademische Karrieren (TOSCA) für die Fächer Mathematik und Englisch systematisch untersucht. Berücksichtigt wurden am Ende des Schuljahres 2001/2002 die Leistungen von $N = 2.854$ Schülerinnen und Schülern allgemeinbildender und $N = 1.876$ Schülerinnen und Schülern beruflicher Gymnasien in der 13. Jahrgangsstufe. Mit Ausnahme biotechnologischer Gymnasien wurden alle Zweige beruflicher Gymnasien berücksichtigt. Mit der TOSCA-Repeat Studie (TOSCA-R; Trautwein et al., 2010) ist der Versuch einer Replikationsstudie unter den Bedingungen einer veränderten Oberstufenstruktur in Baden-Württemberg vorgenommen worden.

Die Mathematikleistungen in TOSCA wurden mit dem in TIMSS (vgl. Baumert et al., 2000) eingesetzten Test erfasst, die Englischleistungen mit einer Paper-Pencil-Version des Test of English as a Foreign Language (TOEFL; zu Details Köller et al., 2004). Hinsichtlich der erreichten Testleistungen wurden soziale und kriteriale Lernziele definiert. Bei der Definition von Mindeststandards im Fach Mathematik wurde analog zur BIJU-Studie vorgegangen. Ausgehend von der Annahme, dass Mathematiklehrkräfte in der gymnasialen Oberstufe ein geteiltes professionelles Verständnis davon besitzen, welche Kompetenzen hinter einer ausreichenden Leistung (5 Punkte) in einem Mathematik-Grundkurs stehen, wurde die mittlere Testleistung aller Grundkurschüler, die im letzte Zeugnis 5 Punkte erreicht hatten, als (sozialer) Mindeststandard definiert. Kriteriale Standards im Sinne von Regelstandards wurden auf der Basis des in TIMSS verwendeten Kompetenzstufenmodells definiert (s.o.). In diesem Modell beschreibt die zweite Kompetenzstufe ein Leistungsniveau, bei dem Kompetenzen, die typischerweise in der Sekundarstufe I erworben werden, sicher beherrscht werden. Die Kompetenzstufe III beschreibt die sichere Anwendung von in der Oberstufe erworbenen Lerninhalten. Letztere wurde im Sinne eines Regelstandards für Leistungskurschülerinnen und -schüler definiert, die Kompetenzstufe II als Regelstandard für die Grundkurschülerinnen und -schüler. Die Abbildung 5 zeigt den Prozentsatz der Schülerinnen und Schüler an den verschiedenen Zweigen, welche den sozialen Mindeststandard bzw. kriterialen Regelstandard erreicht haben. Erkennbar ist, dass die Erreichung der unterschiedlichen Standards für die allgemeinbildenden und technischen Gymnasien unproblematisch ist, ein anderes Bild zeigt sich an den übrigen Zweigen beruflicher Gymnasien, an denen erhebliche Schüleranteile die definierten Standards nicht erreichen. Die Leistungsunterschiede

zwischen den Schulzweigen bildeten sich allerdings nicht in den entsprechenden Mathematiknoten ab, die sich kaum unterschieden (zu den Details Watermann et al., 2004).

Abbildung 5: Prozentsatz der Schülerinnen und Schüler der TOSCA-Studie nach Schulform und Erreichen der sozialen und kriterialen Standards im Fach Mathematik am Ende der gymnasialen Oberstufe



Anmerkungen: AG: allgemeinbildendes Gymnasium; TG: technisches Gymnasium; WG: Wirtschaftsgymnasium; ARG: agrarwissenschaftliches Gymnasium; EG: ernährungswissenschaftliches Gymnasium; SG: sozialpädagogisches Gymnasium.

Quelle: Watermann et al., 2004.

Im Fach Englisch wurde, wie oben beschrieben, der TOEFL in einer Paper-Pencil-Version eingesetzt. Dieser Test wird seit den 1960er Jahren verwendet, um die fremdsprachlichen Kompetenzen von internationalen Studienplatz-Bewerbern in den USA und Kanada zu überprüfen. Amerikanische Universitäten und Colleges definieren üblicherweise Mindestleistungen im TOEFL, die internationale Bewerberinnen und Bewerber erreichen müssen, um dort studieren zu können. Die Werte im TOEFL lassen sich auch auf den Niveaus des Gemeinsamen Europäischen Referenzrahmens für Sprachen (GER; Europarat, 2001) abtragen. Der GER beschreibt ausführlich, was Lernende leisten müssen, „um eine Sprache für kommunikative Zwecke zu benutzen, und welche Kenntnisse und Fertigkeiten sie entwickeln müssen, um in der Lage zu sein, kommunikativ erfolgreich zu handeln“ (Europarat, 2001, S. 14). Kommunikative Aktivitäten können auf unterschiedlichem Niveau ausgeführt werden. Diesem Tatbestand wird im GER durch Niveaustufen Rechnung getragen. Es werden drei Grundniveaus – A, B und C – unterschieden, die in je zwei Teilniveaus aufgespalten sind. Insgesamt ergeben sich also sechs Niveaustufen (A1, A2, B1, B2, C1, C2). Am Ende der Sekundarstufe I geben die Bildungsstandards der KMK vor, dass Schülerinnen und Schüler, die den Mittleren Abschluss anstreben, das Niveau B1 erreichen sollten, am Ende der gymnasialen Oberstufe sollte auf vertieftem Niveau im Leistungskurs C1 erreicht werden, ansonsten B2. Hier-

über geben die Einheitlichen Prüfungsanforderung an die Abiturprüfung (EPA) Auskunft. Die Analysen der TOSCA-Daten ergaben das in Tabelle 4 dargestellte Bild.

Tabelle 4: Verteilung der Schülerinnen und Schüler auf die Niveaustufen des GER nach Schulform und Kursniveau

	Allgemeinbildendes Gymnasium			Berufliches Gymnasium		
	Abgewählt	GK	LK	abgewählt.	GK	LK
A1/A2	16%	7%	2%	77%	44%	16%
B1/B2	74%	76%	59%	22%	53%	72%
C1/C2	10%	17%	40%	1%	3%	12%

Anmerkung: GK: Grundkurs; LK: Leistungskurs.

Quelle: Köller et al., 2004.

Setzt man B1/B2 als Standard für die Grundkurse und C1/C2 als Standard für die Leistungskurse, so wird deutlich, dass nennenswerte Anteile der Schülerinnen und Schüler diese Erwartungen nicht erfüllen. Dies gilt insbesondere für die beruflichen Gymnasien, in deren Grundkursen über 40 Prozent der Schülerinnen und Schüler hinter den Erwartungen der Sekundarstufe I zurückbleiben.

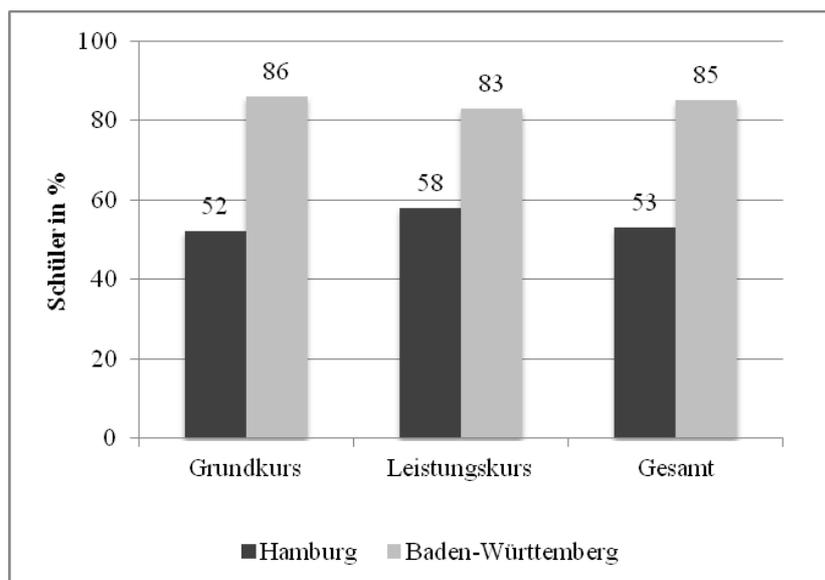
Zweifellos rückt mit diesen Befunden erneut die Frage nach der Vergleichbarkeit der Abschlüsse und nach der Verteilungsgerechtigkeit beim Zugang zu zulassungsbeschränkten Studienfächern in den Vordergrund, implizieren doch die Ergebnisse, dass identische Noten an beruflichen Gymnasien für andere Kompetenzniveaus stehen dürften als an allgemeinbildenden Gymnasien.

4.4 Mathematik- und Englischleistungen in allgemeinbildenden Gymnasien Hamburgs und Baden-Württembergs: Ergebnisse der LAU- und TOSCA-Untersuchung

Die im vorherigen Abschnitt beschriebenen Tests zur voruniversitären Mathematik und zu den Englischkenntnissen wurden auch in der Lernausgangslagenuntersuchung LAU 13 eingesetzt. In LAU 13 wurde der gesamte Hamburger Abiturjahrgang am Ende der 13. Jahrgangsstufe im Schuljahr 2004/2005 in verschiedenen Fächern hinsichtlich der erreichten Fachleistungen untersucht (s. hierzu ausführlich Trautwein et al., 2007). Die leitende Fragestellung der Analysen betraf die Vergleichbarkeit der Abschlüsse zwischen unterschiedlichen Ländern. Betrachtet man das Abschneiden der Gymnasiasten beider Länder in PISA 2003 (PISA-Konsortium Deutschland, 2005) oder PISA 2006 (PISA-Konsortium Deutschland, 2008), so weist Baden-Württemberg in den mathematischen Kompetenzen einen Vorsprung von 21 Punkten auf (599 vs. 570 Punkten). Die Ergebnisse des Ländervergleichs von 2009 (Köller et al., 2010) geben Hinweise auf die Englischleistungen. Hier schnitten Gymnasiastinnen und Gymnasiasten Hamburgs im Hörverstehen etwas besser ab (598 vs. 592 Punkte in Baden-Württemberg), im Lesen kehrte sich der Vorteil zugunsten Baden-Württembergs um (587 vs. 576 Punkte). Ausgehend von der Annahme, dass Vorwissen ein wichtiger Prädiktor für die am Ende der Sekundarstufe II erreichten Leistungen ist, konnte dementsprechend vorhergesagt werden, dass die

untersuchten Schülerinnen und Schüler aus TOSCA im Fach Mathematik denen aus LAU überlegen sein sollten. In Englisch sollten sich dagegen keine systematischen Unterschiede zeigen. Die Bestätigung dieser Vorhersagen für Mathematik findet sich in der Abbildung 6 wieder. Dort ist im Sinne der oben definierten kriterialen Standards für die voruniversitäre Mathematik abgetragen, wie hoch die Anteile in Grund- und Leistungskursen beider Länder sind, welche die Standards erreichen.

Abbildung 6: Prozentsatz der Schülerinnen und Schüler an allgemeinbildenden Gymnasien Hamburgs und Baden-Württembergs nach Kursniveau und Erreichen der kriterialen Standards im Fach Mathematik am Ende der gymnasialen Oberstufe



Quelle: Trautwein et al., 2007.

In Baden-Württemberg liegen die Anteile über 80 Prozent, in Hamburg erreicht dagegen nur gut die Hälfte der Schülerinnen und Schüler die kriterialen Standards.

Für die Leistungen im TOEFL zeigt die Tabelle 5 die Werte an allgemeinbildenden/grundständigen Gymnasien beider Länder. Hier ergeben sich in den Leistungskursen Vorteile zugunsten der Schülerinnen und Schüler in Hamburg (50% vs. 40%). In den Grundkursen und bei den Schülerinnen und Schülern, welche Englisch in der gymnasialen Oberstufe ausgewählt haben, zeigen sich Vorteile zugunsten Baden-Württembergs.

Wie oben beim Schulformenvergleich (allgemeinbildende vs. berufliche Gymnasien) ergibt sich auch hier (zumindest im Fach Mathematik) das Problem der regionalen Vergleichbarkeit des Abiturs und damit verbunden die Frage der Verteilungsgerechtigkeit beim Zugang zu zulassungsbeschränkten Fächern, in denen nach der Föderalismusreform und der Novelle der Hochschulrahmengesetzgebung vom August 2004 Absolventen beider Länder um die verfügbaren Plätze konkurrieren, ohne dass die früheren länderspezifischen Quoten beim Zugang zu den Numerus clausus-Fächern noch ihre Gültigkeit besäßen.

Tabelle 5: Anteile der Abiturientinnen und Abiturienten auf den GER-Stufen nach Land und Kursniveau in der gymnasialen Oberstufe

	Baden-Württemberg			Hamburg		
	abgewählt	GK	LK	abgewählt	GK	LK
A1/A2	16%	7%	2%	32%	16%	2%
B1/B2	74%	76%	59%	60%	68%	49%
C1/C2	10%	17%	40%	8%	17%	50%

Anmerkungen: GK: Grundkurs; LK: Leistungskurs.

Quelle: Trautwein et al., 2007

Zusammenfassung

Obwohl es in der Gemeinschaft der 16 Länder kein abgestimmtes Bildungsmonitoring-Programm für die Sekundarstufe II gibt, sind in den letzten fast 20 Jahren einige Untersuchungen durchgeführt worden, die Auskunft über Leistungsstände von Schülerinnen und Schülern am Ende der gymnasialen Oberstufe geben. Aus den vorliegenden Ergebnissen für die Fächer Mathematik und Englisch lässt sich schlussfolgern, dass in unterschiedlichen Ländern und unterschiedlichen Schulformen mit substantiellen Leistungsunterschieden zu rechnen ist. Vor allem gilt dies für Einrichtungen, die alternative Wege zur Hochschulreife eröffnen. Hier stellt sich in der Tat die Frage, wie die wünschenswerte Öffnung der Wege zum Abitur durch Maßnahmen der Qualitätsentwicklung und Qualitätssicherung flankiert werden kann. Dass die EPA dies nicht erschöpfend leisten können, sollten die hier präsentierten Leistungsvergleiche eindrucksvoll untermauert haben. Dasselbe dürfte für die Bildungsstandards für das Abitur gelten, die 2013 veröffentlicht werden. Da nicht geplant ist, sie mit Maßnahmen der Überprüfung zu verbinden, dürfte ihre steuernde Kraft über die Länder hinweg bescheiden bleiben.

5 Schlussfolgerungen

Die Bildungsexpansion und die Eröffnung alternativer Wege zur Hochschulreife haben die Zahl der Abiturientinnen und Abiturienten in den letzten Jahren dramatisch ansteigen lassen. Aus den ehemaligen Eliteanstalten Gymnasium und Universität sind Massenveranstaltungen geworden, in denen breite Anteile eines Jahrgangs Bildungsangebote erhalten und für akademische Berufe qualifiziert werden. Dieser Modernisierungsprozess im Bildungswesen ist infolge der gestiegenen Anforderungen an Absolventen wünschenswert und trägt der großen Nachfrage nach Akademikerinnen und Akademikern Rechnung. Gleichzeitig ist mit nicht-intendierten „Nebenwirkungen“ solcher Veränderungen im Bildungssystem zu rechnen, deren systematische Untersuchung Aufgabe der Bildungsforschung ist. Die bereitgestellten Befunde belegen, im unteren Leistungsbereich nur geringe Kompetenzstände erreicht werden. Dies gilt umso stärker in beruflichen Gymnasien und integrierten Gesamtschulen, in denen auch Ende der Sekundarstufe II teilweise Leistungsniveaus erreicht werden, die einen erfolgreichen Übergang in Studium fraglich erscheinen lassen. Die Studienabbruchquoten, die beispielsweise im Jahr 2010 in

den Ingenieurwissenschaften bei 48 Prozent lagen (vgl. Heublein et al., 2012), mögen ein Hinweis darauf sein, dass bei vielen Studienanfängerinnen und Studienanfängern eine erhebliche Lücke zwischen Vorkenntnissen und Anforderungen im Studium besteht. Man kann solchen Diskrepanzen auf zwei Wegen begegnen: zum einen durch intensivere Förderung in der gymnasialen Oberstufe, zum anderen durch Eingangskurse an den Universitäten, die die Studierenden zunächst auf die Anforderungen der universitären Fächer vorbereiten. Letzteres findet momentan zunehmend an deutschen Universitäten Eingang und wird hoffentlich dazu beitragen die Abbrecherquoten zu senken.

Literatur

- Baron-Boldt, J., Schuler, H. & Funke, U. (1988). *Prädiktive Validität von Schulabschlußnoten: Eine Metaanalyse*. Zeitschrift für Pädagogische Psychologie, 2 (2), 79–90.
- Baumert, J., Bos, W. & Lehmann, R. (Hg.) (2000). *Dritte Internationale Mathematik- und Naturwissenschaftsstudie: Mathematische und naturwissenschaftliche Bildung am Ende der Schullaufbahn. Bd. 2: Mathematische und physikalische Kompetenzen am Ende der gymnasialen Oberstufe*. Opladen: Leske + Budrich.
- Baumert, J., Roeder, P. M. & Watermann, R. (2003). *Das Gymnasium – Kontinuität im Wandel*. In K. S. Cortina, J. Baumert, A. Leschinsky, K. U. Mayer & L. Trommer (Hrsg.), *Das Bildungswesen in der Bundesrepublik Deutschland* (S. 487–524). Reinbek: Rowohlt.
- Bridgeman, B., McCamley-Jenkins, L. & Ervin, N. (2000). *Predictions of freshman grade-point average from the revised and recentered SAT I: Reasoning Test* (College Board Research Report No. 2001-01). New York: College Entrance Examination Board.
- Deidesheimer Kreis. (1997). *Hochschulzulassung und Studieneignungstests: Studienfeldbezogene Verfahren zur Feststellung der Eignung für Numerus-clausus- und andere Studiengänge*. Göttingen: Vandenhoeck & Ruprecht.
- ETS - Educational Testing Service (2010). *Sign up for the Free GRE® Search Service*. Retrieved March 9, 2010, from http://www.ets.org/gre/general/about/tools/search_service/index.html.
- Europarat. (2001). *Gemeinsamer Europäischer Referenzrahmen für Sprachen: lernen, lehren, beurteilen*. Berlin: Langenscheidt.
- Formazin, M., Schroeders, U., Köller, O., Wilhelm, O. & Westmeyer, H. (2011): *Studierendenauswahl im Fach Psychologie: Testentwicklung und Validitätsbefunde*. *Psychologische Rundschau*, 62 (4), 221–236.
- Hell, B., Trapmann, S. & Schuler, H. (2007). *Eine Metaanalyse der Validität von fachspezifischen Studierfähigkeitstests im deutschsprachigen Raum*. *Empirische Pädagogik*, 21 (3), 251–270.
- Helmke, A. & Weinert, F. E. (1997). *Bedingungsfaktoren schulischer Leistungen*. In F. E. Weinert (Hrsg.), *Psychologie des Unterrichts und der Schule. Enzyklopädie der Psychologie, Serie Pädagogische Psychologie* (S. 71–176). Göttingen: Hogrefe.
- Heublein, U., Richter, J., Schmelzer, R. & Sommer, D. (2012) Klieme, E. (2000). *Die Entwicklung der Schwund- und Studienabbruchquoten an den deutschen Hochschulen. Statistische Berechnungen auf der Basis des Absolventenjahrgangs 2010*. Hannover: HIS GmbH.

- Klieme, E. (2000). *Fachleistungen im voruniversitären Mathematik- und Physikunterricht: Theoretische Grundlagen, Kompetenzstufen und Unterrichtsschwerpunkte*. In J. Baumert, W. Bos & R. Lehmann (Hrsg.), *TIMSS/III. Dritte Internationale Mathematik- und Naturwissenschaftsstudie – Mathematische und naturwissenschaftliche Bildung am Ende der Schullaufbahn: Bd. 2. Mathematische und physikalische Kompetenzen am Ende der gymnasialen Oberstufe* (S. 57–128). Opladen: Leske + Budrich.
- KMK – Kultusministerkonferenz (1977/1978). *Empfehlungen zur Arbeit in der gymnasialen Oberstufe*. *Bildung und Erziehung*, 31, 561–574.
- KMK – Kultusministerkonferenz (1995). *Weiterentwicklung der Prinzipien der gymnasialen Oberstufe und des Abiturs. Abschlussbericht der von der Kultusministerkonferenz eingesetzten Expertenkommission*. Bonn: Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland.
- KMK – Kultusministerkonferenz (2002). *Vereinbarung über Einheitliche Prüfungsanforderungen in der Abiturprüfung*. (Beschluss der Kultusministerkonferenz vom 01.06.1979 i.d.F. vom 24.05.2002). Bonn: Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland.
- KMK – Kultusministerkonferenz (2009). *Vorausberechnung der Schüler- und Absolventenzahlen 2010 bis 2025*. Bonn: Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland.
- Kobrin, J. L., Patterson, B. F., Shaw, E. J., Mattern, K. D. & Barbuti, S. M. (2008). *Validity of the SAT for predicting first-year college grade point average* (College Board Research Report No. 2008-5). New York: The College Board.
- Köller, O. (2010). *Standardsetzung im Bildungssystem*. In H. Reinders., H. Ditton, C. Gräsel & B. Gniewosz (Hrsg.), *Empirische Bildungsforschung. Strukturen und Methoden* (S. 179–192). Wiesbaden: VS Verlag.
- Köller, O. & Baumert, J. (2001). *Leistungsgruppierungen in der Sekundarstufe I und ihre Konsequenzen für die Mathematikleistung und das mathematische Selbstkonzept der Begabung*. *Zeitschrift für Pädagogische Psychologie*, 15 (2), 99–110.
- Köller, O. & Baumert, J. (2002). *Das Abitur – Immer noch ein gültiger Indikator für die Studierfähigkeit?* *Aus Politik und Zeitgeschichte*, B26/2002, 12–19.
- Köller, O. & Baumert, J. (2012). *Schulische Leistungen und ihre Messung*. In W. Schneider & U. Lindenberger (Hrsg.), *Entwicklungspsychologie* (7. Auflage, S. 639–655). Weinheim: Beltz/PVU.
- Köller, O., Baumert, J., Cortina, K. S. & Trautwein, U. (2010). *Bildungsverläufe und psychosoziale Entwicklung im Jugendalter und jungen Erwachsenenalter*. In C. Spiel, B. Schober, P. Wagner & R. Reimann (Hrsg.), *Bildungspsychologie* (S. 245–252). Göttingen: Hogrefe.
- Köller, O., Baumert, J., Cortina, K. S., Trautwein, U. & Watermann, R. (2004). *Öffnung von Bildungswegen in der Sekundarstufe II und die Wahrung von Standards: Analysen am Beispiel der Englischleistungen von Oberstufenschülern an integrierten Gesamtschulen, beruflichen und allgemein bildenden Gymnasien*. *Zeitschrift für Pädagogik*, 50 (5), 679–700.
- Köller, O., Baumert, J. & Schnabel, K. (1999). *Wege zur Hochschulreife: Offenheit des Systems und Sicherung vergleichbarer Standards. Analysen am Beispiel der Mathematikleistungen von Oberstufenschülern an integrierten Gesamtschulen und Gymnasien in Nordrhein-Westfalen*. *Zeitschrift für Erziehungswissenschaft*, 2, 370–405.
- Köller, O., Knigge, M. & Tesch, B. (Hrsg.) (2010). *Sprachliche Kompetenzen im Ländervergleich*. Münster: Waxmann.

- Köller, O., Trautwein, U., Cortina, K. S. & Baumert, J. (2006). *Rezeptive Kompetenzen in Englisch am Ende der gymnasialen Oberstufe*. Unterrichtswissenschaft, 34 (3), 239–255.
- Köller, O., Watermann, R., Trautwein, U. & Lüdtkke, O. (Hrsg.) (2004). *Wege zur Hochschulreife in Baden-Württemberg. TOSCA – Eine Untersuchung an allgemeinbildenden und beruflichen Gymnasien*. Opladen: Leske+Budrich.
- Konegen-Grenier, C. (2002). *Studierfähigkeit und Hochschulzugang*. Forschung & Lehre, 9, 481–483.
- Kuncel, N. R. & Hezlett, S. A. (2007). *Standardized tests predict graduate students' success*. Science, 315, 1080–1081.
- Kuncel, N. R., Hezlett, S. A. & Ones, D. S. (2001). *A comprehensive meta-analysis of the predictive validity of the Graduate Record Examinations: Implications for graduate student selection and performance*. Psychological Bulletin, 127 (1), 162–181.
- PISA-Konsortium Deutschland (2005). *PISA 2003: Der zweite Vergleich der Länder in Deutschland – Was wissen und können Jugendliche?* Waxmann, Münster.
- PISA-Konsortium Deutschland (2008). *PISA 2006 in Deutschland: die Kompetenzen der Jugendlichen im dritten Ländervergleich*. Waxmann, Münster.
- Ramist, L., Lewis, C. & McCamley-Jenkins, L. (2001). *Using achievement tests/SAT II: Subject Tests to demonstrate achievement and predict college grades: Sex, language, ethnic, and parental education groups* (College Board Research Report No. 2001-5). New York: College Entrance Examination Board.
- Roeder, P. M., Baumert, J., Sang, F. & Schmitz, B. (1986). *Expansion des Gymnasiums und Leistungsentwicklung*. Zeitschrift für Soziologie, 15 (3), 210–220.
- Schmidt, A. (1991). *Das Gymnasium im Aufwind. Entwicklung, Struktur, Probleme seiner Oberstufe*. Aachen: Hahner Verlagsgesellschaft.
- Schuler, H., Funke, U. & Baron-Boldt, J. (1990). *Predictive validity of high-school grades: A meta-analysis*. Applied Psychology: An International Review, 39 (1), 89–103.
- The College Board (2010). *Find a college*. Retrieved March 9, 2010, from <http://www.collegeboard.com/student/csearch/index.html>.
- Trapmann, S., Hell, B., Weigand, S. & Schuler, H. (2007). *Die Validität von Schulnoten zur Vorhersage des Studienerfolgs – eine Metaanalyse*. Zeitschrift für Pädagogische Psychologie, 21 (1), 11–27.
- Trautwein, U., Köller, O., Lehmann, R. H. & Lüdtkke, O. (Hrsg.) (2007). *Schulleistungen von Abiturienten. Regionale, schulformbezogene und soziale Disparitäten*. Münster: Waxmann.
- Trautwein, U. & Lüdtkke, O. (2004). *Aspekte von Wissenschaftspropädeutik und Studierfähigkeit*. In O. Köller, R. Watermann, U. Trautwein & O. Lüdtkke (Hrsg.), *Wege zur Hochschulreife in Baden-Württemberg. TOSCA – Eine Untersuchung an allgemein bildenden und beruflichen Gymnasien* (S. 327–367). Opladen: Leske + Budrich.
- Trautwein, U., Neumann, M., Nagy, G., Lüdtkke, O. & Maaz, K. (Hrsg.) (2010). *Schulleistungen von Abiturienten: Die neu geordnete gymnasiale Oberstufe auf dem Prüfstand* (S. 327–366). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Watermann, R., Nagy, G. & Köller, O. (2004). *Mathematikleistungen in allgemeinbildenden und beruflichen Gymnasien*. In O. Köller, R. Watermann, U. Trautwein & O. Lüdtkke (Hrsg.), *Wege zur Hochschulreife in Baden-Württemberg. TOSCA – Eine Untersuchung an allgemein bildenden und beruflichen Gymnasien* (S. 205–283). Opladen: Leske + Budrich.