

**Rainer H. Lehmann, Susanne Hunger,  
Stanislav Ivanov und Rüdiger Gänsfuß  
unter Mitarbeit von Ellen Hoffmann**

# **LAU 11**

**Aspekte der Lernausgangslage  
und der Lernentwicklung  
Klassenstufe 11**

Ergebnisse einer längsschnittlichen  
Untersuchung in Hamburg

## **Danksagung**

Dank gilt allen, die seit nunmehr gut acht Jahren die Arbeitsgruppe an der Humboldt-Universität zu Berlin bei der wissenschaftlichen Auswertung des Hamburger Projekts „Aspekte der Lernausgangslage und der Lernentwicklung“ unterstützen und denen, die an der erfolgreichen Durchführung der LAU-11-Erhebung mitgewirkt haben.

Im Amt für Bildung hat zunächst Frau Dr. Steinbach als Projektleiterin mit Frau Brose und Frau Nielsen wichtige Impulse für die konzeptionelle Entwicklung gegeben. Als nachfolgender Projektleiter war Herr Dr. Poerschke ein kritischer und konstruktiver Partner für die Berliner Forschungsgruppe, der mit Präzisierungen und Anregungen die Studie aufmerksam begleitet hat. Herr Vieluf als Leiter der Abteilung Qualitätsentwicklung und Standardsicherung am Hamburger Landesinstitut für Lehrerbildung und Schulentwicklung hat mit seinem Team wichtige Beiträge geleistet. Ihnen allen verdankt der jetzt vorgelegte Bericht wesentliche Impulse.

Die Hamburger Schülerinnen und Schüler haben fast ohne Ausnahme mit Eifer und Ernsthaftigkeit die Mühe auf sich genommen, in bis zu sechs Unterrichtsstunden die Tests und einen Fragebogen zu bearbeiten. Schulleitungen und Lehrkräfte haben durch ihre aktive Unterstützung die Studie gefördert.

Astrid Neumann hat auf der Grundlage umfangreicher einschlägiger Erfahrungen in Berlin für die Deutsch-Testkomponente „Textproduktion“ das Beurteilungsschema optimiert sowie den Auswertungsprozess koordiniert, kontrolliert und dokumentiert.

Ellen Hoffmann, Anne Ladstätter und Steffi Schulz haben mit großem Engagement und Gewissenhaftigkeit dafür gesorgt, dass die Vorbereitungsaufgaben für die elektronische Datenerfassung, die Codierung des Englisch-Tests sowie der Code-Abgleich zu vorherigen Untersuchungen erfolgreich vollbracht werden konnten.

Ohne diese vielfältige Zusammenarbeit hätte die Studie nicht durchgeführt und erfolgreich abgeschlossen werden können.

Rainer H. Lehmann,  
wissenschaftlicher Leiter

# Inhaltsverzeichnis

<b>1</b>	<b>Ziele der Untersuchung</b>	<b>5</b>
<b>2</b>	<b>Anlage und Durchführung der Untersuchung</b>	<b>7</b>
2.1	Organisation der Erhebung	7
2.2	Instrumentierung	7
2.3	Datenstruktur im Längsschnitt	19
<b>3</b>	<b>Fachleistungen zu Beginn der Klassenstufe 11</b>	<b>24</b>
3.1	Methodische Vorüberlegungen	24
3.2	Analyse der fachbezogenen Leistungstests	28
3.2.1	Fachleistung Deutsch: Leseverständnis	30
3.2.2	Fachleistung Deutsch: Rechtschreiben	37
3.2.3	Fachleistung Mathematik I	41
3.2.4	Fachleistung Mathematik II	47
3.2.5	Fachleistung Englisch: Wortergänzungstest (C-Test)	53
3.2.6	Fachleistung Englisch: Leseverständnis	60
3.2.7	Zusammenfassung der allgemeinen Fachleistung zu Beginn der Klassenstufe 11	66
3.3	Zusammenfassung	71
<b>4</b>	<b>Entwicklung der Fachleistungen im Längsschnitt</b>	<b>74</b>
4.1	Methodische Vorüberlegungen	74
4.2	Entwicklung der Fachleistung Deutsch: Leseverständnis	79
4.3	Entwicklung der Fachleistung Deutsch: Rechtschreiben	90
4.4	Entwicklung der Fachleistung Mathematik	96
4.5	Entwicklung der Fachleistung Englisch	107

4.6	Zusammenfassung	116
<b>5</b>	<b>Individueller und sozialer Kontext der Schulleistungen</b>	<b>121</b>
5.1	Kognitive Lernvoraussetzungen	121
5.2	Alter und Geschlecht	125
5.3	Migrationshintergrund	135
<b>6</b>	<b>Schlussbetrachtung</b>	<b>144</b>
	<b>Literaturverzeichnis</b>	<b>147</b>
	<b>Glossar</b>	<b>149</b>

# 1 Ziele der Untersuchung

Die Untersuchung *Aspekte der Lernausgangslage und der Lernentwicklung – Klassenstufe 11 (LAU 11)* ist Teil einer Längsschnittstudie, die nahezu alle Hamburger Schülerinnen und Schüler der damaligen fünften Klassen seit 1996 im Abstand von zwei Schuljahren erfasst hat. Im Schuljahr 2002/03 wurden nun diese Schülerinnen und Schüler erneut getestet. Da sich ein Teil der Jugendlichen in der Klassenstufe 11 der gymnasialen Oberstufe und ein anderer Teil in einer beruflichen Ausbildung befindet, wurde zusätzlich zur *LAU 11* das Projekt *ULME (Untersuchung der Leistung, Motivation und Einstellungen zu Beginn der beruflichen Ausbildung)* konzipiert. *LAU* dokumentiert in zweijährigen Abständen die jeweils erreichten Lernstände, Lernentwicklungen und schulbezogenen Einstellungen vom Ende der Grundschulzeit bis in die Klassenstufe 11 der gymnasialen Oberstufe bzw. in die Einführungsstufe des Aufbaugymnasiums, *ULME* analysiert die Kompetenzen von Schülerinnen und Schülern, die sich im Jahre 2002 am Anfang einer Ausbildung an einer Berufsschule oder an einer Berufsfachschule befanden.

Im September 1996 fand die erste Erhebung der Längsschnittstudie statt. Sie betraf den Übergang von der Grundschule in die Sekundarstufe I und konzentrierte sich auf ausgewählte Aspekte der Lernausgangslage, auf die der Unterricht an den weiterführenden Schulen aufbaut (vgl. LEHMANN & PEEK 1997). Die zweite Erhebung im September 1998 richtete sich auf die Ausprägung und Entwicklung von Fachleistungen, Problemlösekompetenz und schulbezogenen Einstellungen desselben Schülerjahrgangs in den Beobachtungsstufen der Haupt- und Realschulen bzw. der Gymnasien sowie in den Jahrgangsstufen 5 und 6 der Gesamtschulen – und damit zugleich auf die zu Beginn des Unterrichts in der Jahrgangsstufe 7 erreichten Lernstände (vgl. LEHMANN, GÄNSFUß & PEEK 1999). Im Jahr 2000 folgte die dritte Erhebung, die die weitere Lernentwicklung der Hamburger Schülerinnen und Schüler zum Gegenstand hatte (vgl. LEHMANN, PEEK, GÄNSFUß & HUSFELDT 2002). Mit der LAU-11-Erhebung schließt sich nun die Beschreibung und Analyse der weiteren Lernentwicklung derjenigen Jugendlichen an, die im Herbst 2002 in die gymnasiale Oberstufe an Gymnasien, Gesamtschulen, Aufbau-, Wirtschafts- und technischen Gymnasien oder in die Einführungsstufe des Aufbaugymnasiums eingetreten sind.

Für den hier vorgelegten Bericht steht die Aufklärung von Fragen zu Fachleistungen und Lernentwicklungen in den Klassenstufen 9 und 10 sowie zu dem Einfluss kognitiver Lernvoraussetzungen, des Geschlechts und eines Migrationshintergrundes im Vordergrund:

- Mit welchen Lernständen in ausgewählten Fachleistungsbereichen (Deutsch: Leseverständnis, Rechtschreiben; Mathematik; Englisch: Leseverständnis, Lücken-Test) treten die Schülerinnen und Schüler in die verschiedenen Formen der gymnasialen Oberstufe bzw. in die Einführungsstufe des Aufbaugymnasiums ein?
- Wie haben sich die Lernstände der Schülerinnen und Schüler vom Eintritt in die Klassenstufe 5 bis zum Übergang in die gymnasiale Oberstufe bzw. in die Einführungsstufe des Aufbaugymnasiums entwickelt bzw. verändert?
- Welche Zusammenhänge bestehen zwischen der Lernentwicklung und den kognitiven Lernvoraussetzungen der Schülerinnen und Schüler? Inwiefern zeigen sich Besonderheiten in der Lernentwicklung von Mädchen und Jungen und von Schülerinnen und Schülern aus Migrationsfamilien?

Diesen Zielsetzungen entsprechend ist der vorliegende Bericht aufgebaut: Nach Ausführungen zur Anlage und zur Durchführung der empirischen Untersuchung in Kapitel 2 (Organisation der Erhebung, Instrumentarium, Datenstruktur im Längsschnitt) werden in den Kapiteln 3 und 4 Ergebnisse zu den zu Beginn der Klassenstufe 11 bzw. zu Beginn der Einführungsstufe des Aufbaugymnasiums erreichten Lernständen sowie zur bisherigen Lernentwicklung der Schülerinnen und Schüler vorgestellt. In Kapitel 5 werden die Lernstände und Lernentwicklungen unter Berücksichtigung der kognitiven Grundfähigkeiten der Schülerinnen und Schüler, des Geschlechts und des Migrationshintergrundes thematisiert. Der Bericht endet mit einer Zusammenfassung der Ergebnisse in Kapitel 6.

Weitere Aspekte der Lernausgangslage und der Lernentwicklung, die im Rahmen der LAU 11 erhoben wurden und die im folgenden Kapitel mit aufgeführt und erläutert werden, werden nach Abschluss der aufwändigen Analysen zu einem späteren Zeitpunkt veröffentlicht.

## **2 Anlage und Durchführung der Untersuchung**

In diesem Kapitel werden Informationen zur Organisation der Erhebung gegeben (2.1), es werden die grundlegenden Konstruktionsprinzipien und der Aufbau der Erhebungsinstrumente vorgestellt (2.2), und es werden die erzielten Teilnahme- und Bearbeitungsquoten im Längsschnitt mitgeteilt (2.3).

### **2.1 Organisation der Erhebung**

Bei der Organisation der Erhebung wurde auf erprobte und bewährte Verfahren aus der LAU-5-, der LAU-7- und der LAU-9-Erhebung zurückgegriffen. Die allgemeine Steuerung der Untersuchung lag wie bisher in der Verantwortung einer Begleitkommission (BeKo), die eine beratende Funktion gegenüber den Entscheidungsträgern im Amt für Bildung und gegenüber der wissenschaftlichen Forschungsgruppe hat. Die Information der Schulen über die Ziele und Verfahren der Studie und die Organisation der Datenerhebung (Versorgung der Schulen mit Test- und Befragungsmaterialien; Beratung der Testleiterinnen und Testleiter in den Schulen) erfolgte über ein Planungsbüro im Amt für Bildung.

Für die schulinterne Organisation war jeweils eine von der Schulleitung benannte Ansprechperson aus dem Kollegium verantwortlich. Die Tests und Befragungen wurden von (Klassen-) Lehrkräften, darunter auch Referendarinnen und Referendare, durchgeführt.

Die Datenerfassung, -aufbereitung und -auswertung sowie die Abfassung des Berichts über die Untersuchung lag in der Verantwortung der wissenschaftlichen Forschungsgruppe an der Humboldt-Universität zu Berlin.

### **2.2 Instrumentierung**

Im Rahmen der Hauptuntersuchung im Herbst 2002 wurden standardisierte Fachleistungstests (SL-HAM 10/11), ein spezielles Verfahren zur Ermittlung von Textverarbeitungswissen, Aufgaben zur Textproduktion – sowohl für Deutsch als auch für Englisch – sowie ein Schülerfragebogen eingesetzt. Zusätzlich wurde ein zweiter Mathematiktest eingesetzt, der zum einen TIMSS-Aufgaben enthält, zum anderen Aufgaben, die vom Amt für Bildung in Hamburg entwickelt wurden. Diese Instrumente wurden bei einem Teil der untersuchten Jugendlichen durch einen nonverbalen Test zum schlussfolgernden Denken ergänzt. Dieser Test (CFT 20) wurde denjenigen Schülergruppen vorgelegt, die bisher noch nicht an der Längsschnittuntersuchung teilgenommen hatten: Jugendliche, die ein Aufbau-, Wirtschafts- oder technisches Gymnasium be-

suchen. Die Aufgaben Deutsch-Textproduktion und Englisch-Textproduktion wurden jeweils nur in einer Stichprobe eingesetzt. In den hier einbezogenen Klassen erhielten die Schülerinnen und Schüler der einen Klassenhälfte die Textproduktionsaufgabe Deutsch, die der anderen Klassenhälfte eine Textproduktionsaufgabe für die Fremdsprache Englisch. Darüber hinaus wurden Angaben aus den amtlichen Schülerakten entnommen.

Insgesamt wurden in der LAU 11 15 Erhebungsinstrumente an bis zu drei Testtagen eingesetzt. In der Tabelle 1 wird unterschieden zwischen vier Gruppen von Erhebungsinstrumenten:

- Gruppe I: Instrumente, die für die LAU konzipiert und speziell für die LAU 11 im Herbst 2001 in Hamburg pilotiert wurden
- Gruppe II: Instrumente aus bereits vorliegenden empirischen Untersuchungen, die im Herbst 2001 in Hamburg pilotiert wurden
- Gruppe III: Instrumente, die für die LAU 11 konzipiert wurden, aber *nicht* in Hamburg pilotiert wurden
- Gruppe IV: Instrumente aus bereits vorliegenden empirischen Untersuchungen, die nicht für die LAU 11 pilotiert wurden

Tabelle 1 Erhebungsinstrumente, beauftragte Institutionen

Erhebungsinstrumente	entwickelt von:
<b>Gruppe I:</b>	
1) Deutsch – Leseverständnis (kontinuierliche Texte)	Universität Koblenz-Landau (Prof. Dr. Jäger),
2) Mathematik I	Humboldt-Universität zu Berlin (HU), Behörde für Bildung und Sport,
3) Englisch – Leseverständnis	Hamburg (BBS)
4) Englisch – C-Test	Hamburg (BBS)
5) Fragebogen zu Lernstrategien	Universität Bremen (Prof. Dr. Straka), BBS und HU
<b>Gruppe II:</b>	
6) Deutsch – Lesegeschwindigkeit	Universität Würzburg,
7) metakognitives Wissen über Text- verarbeitung	(Prof. Dr. Schneider)



**Gruppe III:**

8)	diskontinuierliche Texte (Mathe IIa)	BBS, modifiziert übernommen aus LAU 9
9)	LAU-11-Schülerfragebogen	
10)	Erhebungen aus den Schülerakten	BBS

**Gruppe IV:**

11)	Rechtschreiben – Fehlersuche	übernommen aus LAU 9
12)	Deutsch – Textproduktion	Humboldt-Universität zu Berlin
13)	Englisch – Textproduktion	Universität Augsburg, Prof. Dr. Schröder
14)	Mathematik IIb	TIMSS III
15)	CFT 20 (Kurzform)	Cattell / Weiß; Hogrefe-Verlag

Der Bestimmung der erreichten Leistungsstände diene in erster Linie ein standardisierter Schulleistungstest: der *Hamburger Schulleistungstest für zehnte und elfte Klassen – SL-HAM 10/11*<sup>1</sup>. Dieses Instrument als der eigentliche Leistungstest umfasst Untertests zu den fachbezogenen Lernbereichen *Mathematik* (I und II), *Deutsch* (Leseverständnis und passives Rechtschreibwissen) und *Englisch* (Leseverständnis und C-Test). Die endgültige Auswahl der Aufgaben wurde von der Forschungsgruppe der Humboldt-Universität zu Berlin nach einer breit angelegten Voruntersuchung für die Haupterhebung unter inhaltlichen, teststatistischen und auswertungsökonomischen Gesichtspunkten optimiert.

Der Test wurde in der „Testversion II“ bei all denjenigen Schülerinnen und Schülern eingesetzt, die einen Mittleren Schulabschluss (Realschulabschluss bzw. einen dem Realschulabschluss gleichwertigen Schulabschluss) erworben haben und entweder in die Einführungsstufe des Aufbaugymnasiums oder in die Vorstufe des Aufbau-, Wirtschafts- oder technischen Gymnasiums eingetreten sind; die „Testversion I“ kam bei denjenigen Jugendlichen zum Einsatz, die unmittelbar in die gymnasiale Oberstufe versetzt worden sind.

---

<sup>1</sup> Herausgeber: Behörde für Bildung und Sport, Amt für Bildung, Hamburg 2002; Testentwicklung: Zentrum für empirische pädagogische Forschung (ZEPF) an der Universität Landau, die Abteilung Empirische Bildungsforschung und Methodenlehre an der Humboldt-Universität zu Berlin sowie die Behörde für Bildung und Sport in Hamburg.

Vor dem Hintergrund unterschiedlicher Bildungsgänge und den mit ihnen verbundenen Zugangsberechtigungen erschien es inhaltlich sinnvoll, folgende Gruppen zu unterscheiden:

- Schülerinnen und Schüler in der Regelform der gymnasialen Oberstufe an Gymnasien (*Testversion I*)
- Schülerinnen und Schüler in der Regelform der gymnasialen Oberstufe an Gesamtschulen (*Testversion I*)
- Schülerinnen und Schüler an Wirtschafts- und technischen Gymnasien mit Versetzung in die gymnasiale Oberstufe (*Testversion I*)
- Schülerinnen und Schüler an Wirtschafts- und technischen Gymnasien mit einem Mittleren Schulabschluss (*Testversion II*)
- Schülerinnen und Schüler in der Vorstufe eines Aufbaugymnasiums (je nach Abschlusszeugnis entweder *Testversion I* oder *Testversion II*)
- Schülerinnen und Schüler in der Einführungsstufe eines Aufbaugymnasiums (*Testversion II*)

### **Deutsch – Leseverständnis (kontinuierliche Texte)**

Der Deutsch-Leseverständnistest enthielt in der LAU 11 insgesamt 32 im Multiple-Choice-Format gestellte Aufgaben. Den Aufgaben zum Leseverständnis lag ein Testkonzept zugrunde, das bereits in den vorausgegangenen Erhebungen LAU 5, LAU 7 und LAU 9 leitend war: Es wurden unterschiedlich schwierige Aufgaben verwendet, um über die beiden Testversionen hinweg den gesamten Anforderungsbereich von einfacher Informationsentnahme bis hin zum eigenständigen schlussfolgernden bzw. interpretatorischen Umgang mit Texten zu erfassen.

Für jede der beiden Testversionen lagen vier Texte mit jeweils acht Aufgaben vor. Von diesen vier Texten waren jeweils zwei im längsschnittlichen „Ankerbereich“, ein Text im querschnittlichen testversionübergreifenden „Kernbereich“ und ein Text im testversionspezifischen Bereich situiert. Die beiden Ankertexte mit ihren je acht Ankeraufgaben als Berechnungsgrundlage für Längsschnittanalysen stammten aus dem im Jahr 2000 eingesetzten *Hamburger Schulleistungstest für achte und neunte Klassen – SL-HAM 8/9*. Die kern- und testversionspezifischen Texte wurden nach einer Pilotierung mit 1.510 Schülerinnen und Schülern im Herbst 2001 für die Hauptuntersuchung ausgewählt.

Die Tests wurden jeweils in einer A-Form und in einer Parallelförmigen B ausgegeben; die Formen unterscheiden sich nur in der Anordnung der Texte („Pseu-

do-Parallelformen“). Die Bearbeitungszeit betrug einschließlich einer organisatorischen Einführung in den Test für alle Schülerinnen und Schüler 60 Minuten.

## Mathematik I

Um die curricularen Anforderungen mit den beiden Testversionen des Mathematik-I-Tests möglichst breit abzudecken, erhielt jede Schülerin bzw. jeder Schüler über einen gemeinsamen Kernbestand von 20 Aufgaben (Aufgaben der Schwierigkeitsstufe 2 in der folgenden Tabelle 2) hinaus weitere 21 Aufgaben, die nach Testversion differenziert waren. In beiden Testversionen wurden jeweils die drei Stoffgebiete Algebra, Arithmetik und Geometrie berücksichtigt. Die Testversion I enthielt zusätzlich zwei Stochastik-Aufgaben. Hinsichtlich der Anforderungen handelte es sich nicht nur um typische Mathematikaufgaben vom Beginn der gymnasialen Oberstufe. Der Kernbestand von 20 Aufgaben führte 14 so genannte „Ankeraufgaben“ aus der LAU 9 mit. Sie ermöglichten eine Verkopplung der Ergebnisse mit den Befunden aus den Untersuchungen des Längsschnitts. Bei einer Testlänge von geplanten 41 Items wurden 21 realschulspezifische Aufgaben in den Test der Hauptuntersuchung übernommen, wobei 15 aus der Pilotierung stammten und – zur weiteren Stützung der curricularen Validität – sechs von Hamburger Experten neu konstruiert wurden. Für die Hauptuntersuchung enthielten beide Testversionen je 41 Items. Von den 21 spezifischen Items der Testversion I wurden fünf von einer Hamburger Expertengruppe nach der Pilotierung neu konstruiert.

Tabelle 2 verdeutlicht das Design des Mathematik-I-Tests. Durch die Überschneidungen zwischen den Testversionen wurde es möglich, allen Schülerinnen und Schülern unabhängig von der jeweiligen Testversion einen Lernstandswert auf einer einheitlichen Skala („Metrik“) zuzuweisen.

Tabelle 2 Design des Mathematiktests im *SL-HAM 10/11*

<i>Schwierigkeitsstufe 1</i> 21 Aufgaben	<i>Schwierigkeitsstufe 2</i> 14 Aufgaben   6 Aufgaben		<i>Schwierigkeitsstufe 3</i> 21 Aufgaben
21 Aufgaben	14 Aufgaben	6 Aufgaben	
<i>Testversion für Schülerinnen und Schüler mit einem Mittleren Schulabschluss (Testversion II)</i>			
	14 Aufgaben	6 Aufgaben	21 Aufgaben
	<i>Testversion für Schülerinnen und Schüler mit Vertretung in die gymnasiale Oberstufe (Testversion I)</i>		

In dieser Kombination der beiden unterschiedlichen Versionen enthielt der Mathematik-I-Test 21 Algebra-, 16 Arithmetik-, 23 Geometrie- und 2 Stochastik-Aufgaben. Die 62 Aufgaben wurden in Form von Mehrfachwahlantworten (Multiple-Choice-Items mit vier Antwortalternativen) gestellt. Die Bearbeitungszeit betrug einschließlich einer organisatorischen Einführung in den Test 50 Minuten.

### **Englisch – Leseverständnis**

In der LAU 11 / ULME wurde erstmals ein Leseverständnistest in der Fremdsprache Englisch eingesetzt. Die Verankerung mit einer bereits früher definierten Skala war daher nicht möglich. Wegen der zu erwartenden großen Bandbreite der Fachleistungen war jedoch eine querschnittliche Verankerung über einen Kerntext erforderlich.

Nach der Pilotierung von sechs Textpassagen bei insgesamt 1.260 Schülerinnen und Schülern (Hauptschule: 350; Realschule: 347; Gymnasium: 563) wurden ein Kerntext sowie eine realschul- und eine gymnasialspezifische Textpassage für die Hauptuntersuchung ausgewählt.

Jede schulformspezifische Version des Englisch-Leseverständnistests umfasste somit zwei Textpassagen, die in einem Zeitraum von 25 Minuten zu bearbeiten waren.

### **Englisch – C-Test**

Nach dem Leseverständnistest wurde den Schülerinnen und Schülern der Englisch-C-Test vorgelegt, dessen Format zu Beginn des gesamten Englisch-Testteils zehn Minuten lang eingeübt wurde.

Das Bearbeitungsformat dieses Instruments gilt als äußerst robust; es liefert in aller Regel ein gutes Maß für die generelle fremdsprachliche Kompetenz. Die Testkonstruktion des so genannten C-Tests beruht auf einer Variante des Cloze-Prinzips. Dabei werden, im Unterschied zum Cloze-Test, nicht ganze Wörter, sondern nur die Buchstaben der zweiten Hälfte jedes vierten Wortes im Text gelöscht. Die Schülerinnen und Schüler haben dann die Aufgabe, die entsprechenden Wörter zu vervollständigen. Den in der Literatur formulierten Anforderungen (vgl. RAATZ & KLEIN-BRALEY 1983; GROTHJAHN 1995), dass ein C-Test – um die Reliabilität zu sichern – mindestens 100 Tilgungen aufweisen, aus verschiedenen kurzen Texten bestehen und exakt auswertbar sein sollte, kam der eingesetzte Englisch-C-Test folgendermaßen nach: Er enthielt vier zusammenhängende Texte mit jeweils 23 bis 35 Wortergänzungen, von denen zwei von

allen Schülerinnen und Schülern bearbeitet, die übrigen beiden je nach Testversion differenziert eingesetzt wurden.

Zur Verfügung standen die Erprobungsdaten von 1.201 Schülerinnen und Schülern (Hauptschule: 335; Realschule: 314; Gymnasium: 552). Erprobt wurden zehn Texte mit jeweils bis zu 35 Wortergänzungsaufgaben. Aufgrund der Pilotergebnisse wurden vier Texte in die Hauptuntersuchung aufgenommen, davon gehörten zwei Texte zum Anker- bzw. Kernbereich, weitere zwei Texte wurden testversionspezifisch eingesetzt. Auf Vorschlag der Behörde für Bildung und Sport wurde je ein nicht erprobter Text für die Testversion I sowie für die Testversion II hinzugefügt:

Tabelle 3 Design des Englisch-C-Tests im *SL-HAM 10/11*

Text	Testversion I	Testversion II
Crossing the Channel (CC)	X	X
A Look at the Future 20 Years Ago	X	X
The Famous Spanish Painter Picasso (FSP)		X
Spring Break in America (SBA)		X
Come to Yellowstone (CY)	X	
Life in Silicon Valley (LSV).	X	

Von den insgesamt 174 Wortergänzungen waren 167 eindeutig, in sieben Fällen waren jedoch zwei Ergänzungen möglich und demnach als richtig anzuerkennen. Auch für diese Leistungskomponente wurde das bereits in der LAU 9 gewählte Prinzip der doppelten (längs- und querschnittlichen) Verankerung gewählt. Einer der beiden Texte aus dem Kernbereich war bereits Bestandteil des LAU-9-Instrumentariums. Die hierauf bezogenen 23 Aufgaben bildeten die Ankeritems, die für die längsschnittliche Skalenerweiterung genutzt werden konnten.

### Fragebogen zu Lernstrategien

Um der Frage nach allgemeinen Lerndispositionen nachzugehen, die in der gymnasialen Oberstufe von hoher Bedeutung sind, aber auch das lebenslange Lernen betreffen, wurde ein an der Universität Bremen (Prof. Straka) entwickeltes und dort bereits präpilotiertes Instrument erprobt, das fünf Hauptdimensionen von lernstrategischem Verhalten (1. Aneignen; 2. Planen; 3. Organisieren; 4. metakognitive Kontrolle; 5. kognitive Kontrolle) mit je zwei Subdimensionen

erfasst. Für die Festlegung der endgültig verwendeten Version spielte indessen auch der pragmatische Gesichtspunkt der benötigten Befragungszeit eine Rolle, so dass angesichts der ansprechenden Präpilotierungs-Kennwerte eine Kürzung vorgenommen werden konnte und aus genannten Gründen auch vorgenommen wurde.

In der Pilotierungs-Befragung von ca. 1.250 Schülerinnen und Schülern wurden 59 Items erprobt. Zur Anwendung gelangten exploratorische Faktorenanalysen und klassische Reliabilitätsprüfungen. Trotz der analytischen Unterscheidung von fünf Hauptdimensionen erwies sich das Instrument wegen der hohen Interkorrelationen als kürzbar. Eine starke erste Hauptkomponente war nach Kürzung des Fragebogens auf 24 Items immer noch durch eine sehr gute interne Konsistenz gekennzeichnet. Um jedoch die entwicklungsbestimmende Systematik weitgehend zu bewahren, wurde entschieden, drei Subskalen zu bilden:

- „Lernprozesse planen und organisieren“,
- „Lernprozesse strukturieren“ und
- „Lernprozesse kognitiv und metakognitiv reflektieren“.

Dieses Instrument ist in den LAU-11-Schülerfragebogen integriert, für dessen Beantwortung die Schülerinnen und Schüler insgesamt 20 Minuten Zeit erhielten.

## **Deutsch – Lesegeschwindigkeit**

Anregungen aus dem PISA-Konsortium folgend wurde die Einbeziehung eines Lesegeschwindigkeitstests in die Studie erwogen, der es erlaubt hätte, ein differenzierteres Bild der am Lesen beteiligten kognitiven Prozesse zu zeichnen. Zur Verfügung stand ein an der Universität Würzburg (Prof. Dr. Schneider) entwickeltes Instrument. Die Schülerinnen und Schüler hatten darin einen Text zu lesen und zur Kontrolle kursorisch ca. fünf Aufgaben pro Seite zu bearbeiten. Am Ende der Bearbeitungszeit sollten sie markieren, wie weit sie im Text gekommen sind. Leider konnte dieses Instrument nicht in der vorliegenden Analyse berücksichtigt werden, da sich das elektronisch gesteuerte Einlesen des Tests nachträglich als derart fehleranfällig erwies, dass die ermittelten Daten vorläufig nicht verwendet werden können.

## **Metakognitives Wissen über Textverarbeitung**

Bei diesem Fragebogen ging es in erster Linie darum herauszufinden, welche Strategien Schülerinnen und Schüler verwenden, um Texte optimal zu verstehen und aus ihnen zu lernen. Dieses von Prof. Schneider und Mitarbeitern (Universi-

tät Würzburg) entwickelte Instrument hat sich im Kontext von PISA als sehr erklärungskräftig erwiesen. Bei den Aufgaben 1 bis 7 wurden den Schülerinnen und Schülern jeweils fünf Antwortvorschläge, bei Aufgabe 8 acht Vorschläge vorgegeben, die jeweils durch eine Pseudo-Notenvergabe bewertet werden sollten. Die Durchführungszeit für die acht Aufgaben betrug 10 Minuten.

An der Pilotierung nahmen 1.237 Schülerinnen und Schüler teil (Hauptschule: 279; Realschule: 382; Gymnasium: 576). Die Ergebnisse der Erprobung waren insgesamt hoch zufriedenstellend.

### **Diskontinuierliche Texte (Teil des Mathematik-II-Tests)**

Die diskontinuierlichen Texte bildeten zusammen mit 17 in TIMSS III bewährten Aufgaben zur mathematischen Grundbildung den Mathematik-II-Test, für dessen Bearbeitung insgesamt 40 Minuten vorgesehen waren.

Es handelte sich bei diesem Instrument um zwei Texte, die unter Aufnahme von Aspekten des Testkonzepts von PISA in die Untersuchung einbezogen wurden. Im LAU-Testkonzept wurden ähnliche Aufgaben zur Informationsentnahme bereits zu Beginn der Klassenstufe 5 eingesetzt. Ausgewählt wurden Aufgaben, die sich an der deutschen Fassung des *International Adult Literacy Survey* von 1994/95 (IALS; vgl. OECD / STATISTICS CANADA 1995) orientieren, sich auf aktuelles Material beziehen und als geschlossene Aufgaben konzipiert sind. Eine Pilotierung dieser Aufgaben konnte aus Zeitgründen nicht erfolgen. Angesichts der Ähnlichkeit der Aufgaben mit international bewährten und robusten Skalen schien die Übernahme in die Hauptuntersuchung nur mit geringen Risiken behaftet.

Aufgenommen wurde ein Block von 11 Aufgaben (5 Aufgaben zum Test „Fahrpläne“; 6 Aufgaben zum Test „Arbeitslosenstatistik“) ohne Differenzierung nach Anforderungsniveaus.

### **Schülerfragebogen**

Mit Hilfe eines Schülerfragebogens wurden schul- und unterrichtsbezogene Einstellungen der Schülerinnen und Schüler erhoben. Der Schülerfragebogen setzte sich aus Items bzw. Skalen zusammen, die bereits in ähnlicher Form in anderen Studien zur Messung schul- und unterrichtsbezogener Einstellungen eingesetzt wurden. Er enthielt insgesamt 28 Aussagen zu den Themenbereichen *Schulzufriedenheit*, *leistungsbezogenes Selbstkonzept* und *Anstrengungsbereitschaft*, die jeweils auf einer vierstufigen Skala eingeschätzt werden sollten. Darüber hinaus umfasste der Schülerfragebogen zentrale Aspekte der außerschulischen Lern-

und Lebenswelt der Schülerinnen und Schüler, vor allem Informationen zu ihrer Wohnsituation und zu Lernangeboten zu Hause, zu ihrer Muttersprache, zu ihrer außerschulischen Alltagsgestaltung und zu ihren Hausaufgabenzeiten. Zusammen mit den 24 Fragen zum lernstrategischen Verhalten (s. o.) hatte der LAU-Schülerfragebogen insgesamt 70 Items, für deren Bearbeitung die Schülerinnen und Schüler 20 Minuten Zeit erhielten. Die Bearbeitung des Schülerfragebogens erfolgte freiwillig.

### **Erhebungen aus den Schülerakten**

Wie schon 1996, 1998 und 2000 wurden Informationen (u. a. Geburtsjahr, Geschlecht, Staatsangehörigkeit, Postleitzahl, Noten usw.) den in den Schulen vorhandenen Schülerakten entnommen.

Neben den persönlichen Daten der Schülerinnen und Schüler ging es bei diesem Instrument in erster Linie um Aspekte des Lernerfolgs und der Schullaufbahn in den letzten beiden Jahren vor der Datenerhebung in LAU 11: Wiederholen bzw. Überspringen einer Klasse, Schul(form)wechsel sowie die Noten in den letzten Zeugnissen. Hinzu kamen Organisationsdaten zur Rekombination der ehemaligen zehnten Klassen.

Für die Schülerinnen und Schüler an den Wirtschafts- und technischen Gymnasien in der LAU 11 wurden die entsprechenden Angaben aus der Schülerakte in den jeweiligen Schulen in elektronischer Form erzeugt und als Diskette zur Auswertung verfügbar gemacht.

### **Rechtschreiben – Fehlersuche**

Neben Kompetenzen im Bereich Leseverständnis prüft der *SL-HAM 10/11* eine zweite wesentliche Dimension des Umgangs mit der deutschen Sprache: Kenntnisse der Rechtschreibung. Das *passive Rechtschreibwissen* wird hier in Form der Fehlersuche in einem vorgegebenen Text, der bereits in der LAU 9 eingesetzt wurde, erhoben. Für die Feststellung von Lernfortschritten in den Klassenstufen 9 und 10 gab es keine Alternative zum erneuten Einsatz dieses Instruments, zumal der Text „Jugend und Freizeit an der Jahrtausendwende“ auch für die Klassenstufe 11 als altersgemäß erschien. Leistungsrelevante Erinnerungseffekte waren nahezu auszuschließen. Mithin war eine Pilotierung nicht notwendig.

Der Text zum Freizeitverhalten von Jugendlichen mit insgesamt 285 Wörtern enthielt 30 Wörter mit Rechtschreibfehlern. Die Schülerinnen und Schüler hatten die Aufgabe, sämtliche Fehlschreibungen zu unterstreichen. In der Auswer-



tung wurden dann zusätzlich alle fälschlich unterstrichenen Richtigschreibungen berücksichtigt. Für den Untertest zum passiven Rechtschreibwissen gab es folglich nur eine Testversion; die Bearbeitungszeit betrug 15 Minuten.

### **Deutsch – Textproduktion**

Im Rahmen der LAU 9 war ein aus überregionalen Leistungsmessungen (IEA 1985) bereits bekanntes Verfahren zur Bewertung von Schülertexten weiterentwickelt worden, das sich auf

- die Auszählung von Mikroelementen,
- die analytische Unterscheidung von Bewertungsdimensionen,
- die Mehrfachbewertung durch unabhängig arbeitende Bewerter und
- die Verwendung von so genannten „*Benchmarks*“

stützte. Das Verfahren hat sich insgesamt bewährt. Es ist auch für die Studie *Deutsch-Englisch-Schülerleistungen International* (DESI) der Kultusministerkonferenz (Koordination: Deutsches Institut für Internationale Pädagogische Forschung (DIPF); Frankfurt am Main) übernommen worden. Eine eigene Pilotierung der für LAU 11 bzw. ULME vorgesehenen Aufgabe („Jugendclub“) war daher nicht erforderlich und wurde auch nicht vorgenommen.

Die Aufgabe Textproduktion, für die 45 Minuten Bearbeitungszeit vorgesehen waren, bestand darin, einen Beschwerdebrief an den Gemeinderat zu verfassen, der auf die amtliche Mitteilung über die Schließung eines Jugendclubs Bezug nimmt. Um auch typisch gymnasiale Anforderungen zu erfassen, wurde die Testversion I um eine Zusatzaufgabe ergänzt.

### **Englisch – Textproduktion**

Wie bei dem Untertest Deutsch: Textproduktion wurde auch für die Fremdsprache mit einer Teilstichprobe die Fähigkeit der Schülerinnen und Schüler zum Verfassen von Texten untersucht. Damit wurde an Vorerprobungen aus dem Kontext der DESI angeknüpft. Von Prof. K. Schröder (Universität Augsburg) wurden zwei Aufgabenformen vorgeschlagen; in einem Fall wurde ein verbaler Stimulus vorgegeben, im anderen Fall ging es um die Produktion einer freien Geschichte nach einem von vier dargebotenen fotografischen Stimuli. Die Aufgaben waren in einer kleinen Stichprobe vorerprobt worden. In der Hauptuntersuchung wurden für die Testversion II zwei verschiedene Texte eingesetzt, wobei die Schülerinnen und Schüler frei zwischen beiden Texten wählen konnten. Um auch die typisch gymnasialen Anforderungen zu erfassen, wurde für die Schülerinnen und Schüler, die die Testversion I bearbeiten sollten, die Schreib-

aufgabe zusätzlich um die oben genannte freie Geschichte über eine von vier dargebotenen Fotografien ergänzt.

Für die Aufgabe Englisch: Textproduktion waren 45 Minuten Bearbeitungszeit vorgesehen. Das Bewertungsschema wurde in wesentlichen Teilen analog dem in der Komponente Deutsch: Textproduktion verwendeten aufgebaut.

## **Mathematik II**

Um die LAU 11 zusätzlich mit anderen Studien zu vernetzen und insgesamt den ‚authentischen‘, weniger curriculumgebundenen Charakter der Untersuchung zu stärken, wurde entschieden, bewährte Items aus der Komponente „Mathematische Grundbildung“ der TIMSS/III zu übernehmen. Angesichts der belegten Robustheit dieses Itempools erschien eine eigene Pilotierung nicht angezeigt. Zusammen mit den beiden diskontinuierlichen Texten bildeten die 17 TIMSS-III-Aufgaben zur mathematischen Grundbildung den Mathematik-II-Test, für dessen Bearbeitung insgesamt 40 Minuten vorgesehen waren.

## **CFT 20 (Kurzform)**

Die Schulleistungstests wurden durch den nonverbalen *Culture Fair Intelligence Test – CFT 20* in seiner Kurzform ergänzt. Dieser Test zielt auf die Messung von Aspekten des schlussfolgernden Denkens, also von allgemeinen kognitiven Lernvoraussetzungen. Es handelt sich bei diesem Verfahren um einen bundesweit geeichten und im Übrigen auch schon im Rahmen der LAU 5 bzw. LAU 9 eingesetzten Test, was den Vergleich mit Altersnorm- bzw. Klassenstandardwerten ermöglicht (vgl. WEIß 1997). Die Aufgaben des *CFT 20* sind in vier Subtests (Reihen fortsetzen; Klassifikationen; Matrizen; topologische Schlussfolgerungen) untergliedert. Die einzelnen Subtests bestehen aus sprachfreien, in zeichnerischer Form dargestellten und nach Schwierigkeiten geordneten Einzelaufgaben im Multiple-Choice-Format. Seine anschaulich-figurale Gestaltung macht diesen standardisierten Test vom Grad der Beherrschung der deutschen Sprache sowie von fachgebundenem Wissen und Können weitgehend unabhängig und gestattet es, Übereinstimmungen und Diskrepanzen zwischen dem kognitiven Potenzial einer Schülerin bzw. eines Schülers und ihren bzw. seinen fachgebundenen Fähigkeiten zu erkennen. Für die Bearbeitung des *CFT 20* in seiner Kurzform mit insgesamt 46 Aufgaben wurden 40 Minuten vorgesehen. Angewendet wurde der Test in den Wirtschafts- und technischen Gymnasien sowie in den Aufbaugymnasien. Die Teilnahme war freiwillig.

## 2.3 Datenstruktur im Längsschnitt

Diese vierte Datenerhebung im Rahmen der LAU fand an drei Tagen zwischen dem 16. und dem 20. September 2002 statt. Es wurde angestrebt, mit der Untersuchung im September 2002 unmittelbar nach dem Übergang in die Klassenstufe 11 oder in die Einführungsstufe des Aufbaugymnasiums Informationen über möglichst all diejenigen Schülerinnen und Schüler zu erhalten, die bereits an den Erhebungen in den Klassenstufen 5, 7 und 9 teilgenommen hatten. Zusätzlich wurden auch diejenigen Jugendlichen der Klassenstufe 11 und der Einführungsstufe des Aufbaugymnasiums in die Studie einbezogen, die aus verschiedenen Gründen an den Erhebungen 1996, 1998 bzw. 2000 noch nicht teilgenommen hatten. Neben zugezogenen Schülerinnen und Schülern sowie Schülerinnen und Schülern, die die Klassenstufe 11 wiederholten (teilweise im Anschluss an einen Auslandsschulbesuch), stammte eine beträchtliche Zahl der „Neuzugänge“ aus der vorjährigen Einführungsstufe des Aufbaugymnasiums und aus beruflichen Bildungsgängen (insbesondere Berufsfachschulen).

Die folgende Tabelle 4 gibt an, wie viele Schulen, Klassen sowie Schülerinnen und Schüler in die Untersuchung einbezogen wurden, für wie viele Schülerinnen und Schüler aus Klassen bzw. Schulen also mindestens Angaben aus den Schülerbögen vorliegen.

Tabelle 4 Datenstruktur der Erhebungen: Anzahl der Schulen, Klassen sowie Schülerinnen und Schüler nach Bildungsgang

	Schulen	Klassen	Schülerinnen und Schüler
<b>Gymnasium</b>	65	160	3.604
<b>Gesamtschule</b>	15	36	891
<b>Wirtschafts- und technisches Gymnasium (WTG-G)</b>			160
	11	45	
<b>Wirtschafts- und technisches Gymnasium (WTG-R)</b>			865
		22	323
<b>Aufbaugymnasium VS*</b>	9		
<b>Aufbaugymnasium ES</b>		23	568
<b><i>insgesamt</i></b>	<i>100</i>	<i>286</i>	<i>6.411</i>

\* Ohne diejenigen Klassen, die aus der vorjährigen Einführungsstufe hervorgegangen sind und bereits an der Pilotierung der Erhebungsinstrumente teilgenommen hatten.

Die nachfolgende Tabelle 5 zeigt die Teilnahmequoten für die einzelnen Bestandteile der Untersuchung. Die Bearbeitung der Schulleistungstests war verbindlich, die Bearbeitung des Schülerfragebogens und des *CFT 20* waren freiwillig, die Textproduktionsaufgaben in Deutsch und in Englisch wurden, wie oben beschrieben, nur von einem Teil der Jugendlichen bearbeitet.

Tabelle 5 Teilnahmequoten (Schülerinnen und Schüler) nach Bildungsganggruppen (in Prozent)

<i>SL HAM 10/11</i>					
	<b>Mathe- matik I</b>	<b>Mathe- matik II</b>	<b>Deutsch – Lesever- ständnis</b>	<b>Deutsch – passives Recht- schreib- wissen</b>	<b>Englisch – C-Test</b>
<b>Gymnasium</b>	96,3 %	96,3 %	96,9 %	94,5 %	95,5 %
<b>Gesamtschule</b>	96,6 %	96,4 %	95,8 %	94,7 %	95,2 %
<b>WTG-G</b>	97,5 %	97,5 %	98,1 %	92,5 %	93,8 %
<b>WTG-R</b>	97,2 %	96,9 %	97,6 %	95,6 %	94,5 %
<b>Aufbaugymnasium VS*</b>	94,4 %	95,7 %	96,6 %	94,1 %	93,5 %
<b>Aufbaugymnasium ES</b>	97,7 %	97,4 %	96,1 %	90,3 %	91,0 %
<b><i>insgesamt</i></b>	<i>96,5 %</i>	<i>96,5 %</i>	<i>96,8 %</i>	<i>94,3 %</i>	<i>94,8 %</i>
	<b>Englisch – Lesever- ständnis</b>	<b>Text- erschlie- ßung</b>	<b>CFT 20</b>	<b>Schüler- frage- bogen</b>	<b>lernstrate- gisches Verhalten</b>
<b>Gymnasium</b>	95,4 %	94,2 %	(2,8 %)	91,4 %	91,3 %
<b>Gesamtschule</b>	95,2 %	93,9 %	(0,0 %)	85,1 %	85,1 %
<b>WTG-G</b>	93,8 %	89,4 %	95,6 %	93,8 %	93,8 %
<b>WTG -R</b>	94,3 %	91,6 %	91,8 %	92,8 %	92,7 %
<b>Aufbaugymnasium VS*</b>	93,5 %	90,7 %	65,3 %	88,9 %	88,5 %
<b>Aufbaugymnasium ES</b>	90,8 %	87,1 %	76,9 %	83,6 %	83,8 %
<b><i>insgesamt</i></b>	<i>94,7 %</i>	<i>92,9 %</i>	–	<i>89,9 %</i>	<i>89,9 %</i>

\* Ohne diejenigen Klassen, die aus der vorjährigen Einführungsstufe hervorgegangen sind und bereits an der Pilotierung der Erhebungsinstrumente teilgenommen haben.

Angesichts dieser hohen Teilnahmequoten kann von einer soliden Datengrundlage gesprochen werden, auf die sich die weiteren Analysen stützen. Dabei ist zu beachten, dass an den Aufbaugymnasien diejenigen Klassen der Vorstufe,

die aus der Einführungsstufe hervorgegangen sind und bereits im Jahr zuvor an der Pilotierung der Erhebungsinstrumente teilgenommen hatten, nicht erneut getestet worden sind.

Die Untersuchung wurde als Vollerhebung konzipiert. Die sich ergebenden Veränderungen in der Zusammensetzung von Klassen bringen nur geringfügige Auswertungsprobleme im Längsschnitt mit sich, weil in den meisten Fällen die Organisationsdaten aus den Schülerbögen den Abgleich zwischen Klassen vom Beginn der Klassenstufe 5 bis zum Beginn der Klassenstufe 11 bzw. zum Eintritt in die Einführungsstufe des Aufbaugymnasiums erlauben. Beim Abgleich der Daten im Längsschnitt ist dennoch mit einigen Schwankungen zu rechnen. Ein nicht unbeträchtlicher Teil der Schülerinnen und Schüler hat zwischen den Klassenstufen 5 und 11 bzw. dem Eintritt in die Einführungsstufe des Aufbaugymnasiums entweder eine Klasse wiederholt oder – selten – übersprungen oder ist aus Hamburg weggezogen, während dafür andere in der Population neu hinzugekommen sind (z. B. Jugendliche aus zugezogenen Familien, Wiederholer aus höheren bzw. Springer aus unteren Klassen, Schülerinnen und Schüler, die aus der vorjährigen Einführungsstufe in die Vorstufe des Aufbaugymnasiums übergetreten sind, Schülerinnen und Schüler aus beruflichen Bildungsgängen). Zusätzlich ergeben sich Schwankungen vor dem Hintergrund, dass sich Sonderschulen, Privatschulen und siebenstufige Gymnasien 1998 oder 2000 einmalig oder erstmalig an der Untersuchung beteiligt haben. Mit dem Übergang in die gymnasiale Oberstufe bzw. in die Einführungsstufe des Aufbaugymnasiums wurden in der LAU 11 die Haupt- und der überwiegende Teil der Realschülerinnen und Realschüler sowie der Gesamtschülerinnen und Gesamtschüler nicht mehr erfasst, dafür aber zum Teil in der Parallelstudie ULME.

Mit den Erhebungen zu Beginn der Klassenstufen 5, 7, 9 und 11 wurden insgesamt 18.087 Schülerinnen und Schüler erfasst. Darüber, wie viele von den 6.411 Jugendlichen, die in der LAU 11 getestet wurden, bereits an früheren Erhebungen der LAU teilgenommen haben, gibt die folgende Tabelle 6 Auskunft.

Tabelle 6 Schülerinnen und Schüler der LAU 11 nach Erhebungszeitraum

<b>Schüler insgesamt in LAU 11</b>	<b>N = 6.411</b>	<b>100 %</b>
<b>erstmals in LAU 5</b>	N = 3.868	60,3 %
<b>erstmals in LAU 7</b>	N = 342	5,3 %
<b>erstmals in LAU 9</b>	N = 476	7,4 %
<b>neu in LAU 11</b>	N = 1.725	26,9 %

Tabelle 7 enthält detaillierte Angaben darüber, wie viele Schülerinnen und Schüler aus der LAU 11 in LAU 9 und/oder in LAU 7 bzw. in LAU 5 getestet worden sind.

Tabelle 7 Einbezogene Schülerinnen und Schüler zu den vier Erhebungszeitpunkten 2002 (LAU 11), 2000 (LAU 9), 1998 (LAU 7) und 1996 (LAU 5), differenziert nach Bildungsgang

	Gym- nasium	Gesamt- schule	WTG-G	WTG-R	Aufbau- gymna- sium VS	Aufbau- gymna- sium ES	N
<b>Klassenstufe 11</b>	514	157	91	570	228	165	1.725 (26,9 %)
<b>Klassenstufen 9 und 11</b>	350	27	9	32	19	39	476 (7,4 %)
<b>Klassenstufen 7 und 11</b>	–	1	2	7	1	4	15 (0,2 %)
<b>Klassenstufen 7, 9 und 11</b>	198	40	4	28	12	45	327 (5,1 %)
<b>Klassenstufen 5, 7 und 11</b>	11	10	–	18	1	17	57 (0,9 %)
<b>Klassenstufen 5, 7, 9 und 11</b>	2.531	656	54	210	62	298	3.811 (59,4 %)
<b>insgesamt</b>	<i>3.604</i>	<i>891</i>	<i>160</i>	<i>865</i>	<i>323</i>	<i>568</i>	<b>6.411</b> (100 %)

Tabelle 8 gibt Aufschluss über die Herkunft der Schülerinnen und Schüler (die in Klassenstufe 9 besuchte Schulform) und informiert über Schülerwanderungen beim Übergang von der Klassenstufe 10 in die Klassenstufe 11 der gymnasialen Oberstufen bzw. in die Einführungsstufe des Aufbaugymnasiums. Die Prozentangaben unter den Schülerzahlen beziehen sich jeweils auf die Gesamtschülerschaft des betreffenden Bildungsgangs in der LAU 11. Insgesamt haben 4.614 Schülerinnen und Schüler aus LAU 11 bereits an der LAU 9 teilgenommen. 3.204 von den insgesamt 6.411 Schülerinnen und Schülern, die an der LAU 11 teilgenommen haben, stammen aus einer Gymnasialklasse, 919 Schülerinnen und Schüler aus einer Gesamtschulklasse, 395 Schülerinnen und Schüler hatten in der Klassenstufe 9 eine Realschulklasse besucht, 3 Schülerinnen und Schüler eine Hauptschulklasse und 93 Jugendliche kamen aus einer Integrierten Haupt- und Realschule.

Tabelle 8 Herkunft der Schülerinnen und Schüler nach Bildungsgang zu Beginn der Klassenstufe 11 und nach Schulform zu Beginn der Klassenstufe 9

an <i>LAU 9</i> bereits teilgenommen	<i>LAU 11</i> Gym- nasium N=3.604	Gesamt- schule N=891	WTG-G N=160	WTG-R N=865	Aufbau- gymna- sium VS N=323	Aufbau- gymna- sium ES N=568	ins- gesamt N=6.411
<b>Gymnasial- klasse</b>	3.059 (84,9 %)	65 (7,3 %)	36 (22,5 %)	20 (2,3 %)	10 (3,1 %)	14 (2,5 %)	3.204 (50 %)
<b>Gesamtschul- klasse</b>	20 (0,6 %)	654 (73,4 %)	27 (16,9 %)	77 (8,9 %)	2 (0,6 %)	139 (24,5 %)	919 (14,3 %)
<b>Realschul- klasse</b>	–	4 (0,4 %)	3 (1,9 %)	138 (16,0 %)	71 (22,0 %)	179 (31,5 %)	395 (6,2 %)
<b>Hauptschul- klasse</b>	–	–	–	1 (0,1 %)	1 (0,3 %)	1 (0,2 %)	3 (0,05 %)
<b>IHR-Klasse</b>	–	–	1 (0,6 %)	34 (3,9 %)	9 (2,8 %)	49 (8,6 %)	93 (1,5 %)
<b>insgesamt wiedergefun- dene Schüler aus LAU 9</b>	<b>3.079 (85,4 % von 3.604)</b>	<b>723 (81,1 % von 891)</b>	<b>67 (41,9 % von 160)</b>	<b>270 (31,2 % von 865)</b>	<b>93 (28,8 % von 323)</b>	<b>382 (67,3 % von 568)</b>	<b>4.614 (72,0 % von 6.411)</b>

Betrachtet man die Schülerwanderungen von der Klassenstufe 9 aus, ist festzustellen, dass von den insgesamt 4.660 Gymnasiastinnen und Gymnasiasten, die an der LAU 9 teilgenommen hatten, im Jahr 2002 noch 3.059 (65,6 %) in der Regeloberstufe eines Gymnasiums und 145 (3,1 %) in den anderen Bildungsgängen angetroffen wurden. An der LAU 9 hatten 3.703 Schülerinnen und Schüler in Gesamtschulklassen teilgenommen, von ihnen befanden sich im Herbst 2002 654 (17,7 %) in der Regeloberstufe derselben Schulform und 265 (7,2 %) in einer der anderen Schulformen. Von den 2.070 Realschülerinnen und Realschülern der LAU 9 haben 216 (10,4 %) der Jugendlichen den direkten Übergang in die gymnasiale Oberstufe vollzogen, weitere 179 Schülerinnen und Schüler (8,6 %) sind in die Einführungsstufe des Aufbaugymnasiums eingetreten. Von den 709 Schülerinnen und Schülern, die in der Klassenstufe 9 eine Integrierten Haupt- und Realschule (IHR-Schule) besuchten, sind 44 (6,2 %) direkt in die gymnasiale Oberstufe übergetreten, weitere 49 Schülerinnen und Schüler (6,9 %) besuchten im Jahr 2002 die Einführungsstufe des Aufbaugymnasiums.

### **3 Fachleistungen zu Beginn der Klassenstufe 11**

#### **3.1 Methodische Vorüberlegungen**

Die Erhebungen zu den Fachleistungen, dem Wissen zur Texterschließung sowie zum lernstrategischen Verhalten fanden zu Beginn der Klassenstufe 11 bzw. der Einführungsstufe des Aufbaugymnasiums statt, zu einem Zeitpunkt also, zu dem seit Ende der Klassenstufe 10 vermutlich keine nennenswerten Lernfortschritte stattgefunden haben.

Die Schülerschaft in der Regelform der gymnasialen Oberstufe an den Gymnasien und Gesamtschulen besteht aus Schülerinnen und Schülern, die in die gymnasiale Oberstufe versetzt worden sind.

Gymnasien und Gesamtschulen mit Aufbauklassen hingegen führen auch Jugendliche mit einem qualifizierten Mittleren Schulabschluss zum Abitur; Voraussetzung für den Eintritt in die Einführungsstufe des Aufbaugymnasiums ist eine Durchschnittsnote von mindestens 3,0 über alle Fächer (ohne Sport), während für den unmittelbaren Eintritt in die Vorstufe des Aufbaugymnasiums darüber hinaus eine Durchschnittsnote von mindestens 3,0 in den Fächern Deutsch, Mathematik und erste Fremdsprache erforderlich ist.

Die Schülerschaft an den Wirtschafts- und technischen Gymnasien setzt sich zu einem Großteil aus Jugendlichen mit einem qualifizierten Mittleren Schulabschluss (Durchschnittsnote 3,0 über alle Fächer, ohne Sport; 865 Jugendliche) und zu einem deutlich geringeren Anteil (160 Jugendliche) aus Gymnasiastinnen und Gymnasiasten sowie Gesamtschülerinnen und Gesamtschülern, die die Voraussetzungen für den direkten Übergang in die gymnasiale Oberstufe erfüllen, zusammen.

Um den unterschiedlichen Voraussetzungen für den Eintritt in die verschiedenen Formen der gymnasialen Oberstufe bzw. die Einführungsstufe des Aufbaugymnasiums Rechnung zu tragen, wurden zwei Testversionen entwickelt, die sich hinsichtlich der curricularen Anforderungen unterscheiden; mit Testversion I wurden alle Schülerinnen und Schüler mit Versetzung in die gymnasiale Oberstufe getestet, mit Testversion II alle Schülerinnen und Schüler mit einem Mittleren Schulabschluss (Realschulabschluss bzw. eine von der zuständigen Behörde als gleichwertig anerkannte Vorbildung).

Wo nicht ausdrücklich anderes vermerkt ist, beziehen sich die Analysen konsequent auf die folgende Zuordnung (Tabelle 9):



Tabelle 9 Bildungsgänge in der LAU 11

<b>Bildungsganggruppen und Testversionen</b>	<b>Kurzform</b>
Schülerinnen und Schüler in der Regelform der gymnasialen Oberstufe an Gymnasien ( <i>Testversion I</i> )	Gymnasium
Schülerinnen und Schüler in der Regelform der gymnasialen Oberstufe an Gesamtschulen ( <i>Testversion I</i> )	Gesamtschule
Schülerinnen und Schüler an Wirtschafts- und technischen Gymnasien mit Versetzung in die gymnasiale Oberstufe ( <i>Testversion I</i> )	WTG-G
Schülerinnen und Schüler an Wirtschafts- und technischen Gymnasien mit qualifiziertem Mittlerem Schulabschluss ( <i>Testversion II</i> )	WTG -R
Schülerinnen und Schüler in der Vorstufe des Aufbaugymnasiums (je nach Vorbildung – Mittlerer Schulabschluss oder Versetzung in die Vorstufe des Aufbaugymnasiums – entweder <i>Testversion II</i> oder <i>Testversion I</i> )	Aufbaugymnasium VS
Schülerinnen und Schüler in der Einführungsstufe des Aufbaugymnasiums mit qualifiziertem Mittlerem Schulabschluss ( <i>Testversion II</i> )	Aufbaugymnasium ES

Schulformwechsel, aber auch Klassenwiederholung, das Überspringen einer Klassenstufe und regionale Mobilität der Familien, in geringerem Umfang auch Datenausfälle durch Nichtteilnahme einer Schule, einer Klasse oder eines Kurses an den LAU-Erhebungen haben unvermeidlich Konsequenzen für den Umfang der Untersuchungsgruppe. Von den 12.620 Jugendlichen, die im Rahmen der LAU 9 getestet wurden, nahmen 4.614 (36,6 %) an der LAU 11 teil, weitere 2.346 Schülerinnen und Schüler (18,6 %) beteiligten sich an der ULME-Erhebung, die parallel zur LAU 11 in den beruflichen Schulen durchgeführt wurde.

Eine Unterscheidung nach den Gründen für das Ausscheiden eines Schülers oder einer Schülerin aus der Untersuchungsgruppe ist bei der gegebenen Datenlage nicht möglich. Ob Schulabgang ohne Abschluss, Übergang in die Berufsausbildung nach dem Hauptschulabschluss, Klassenwiederholung, Überspringen einer Klassenstufe, Wegzug, Auslandsschulbesuch, Berufsausbildung außerhalb von Hamburg, freiwilliges soziales oder ökologisches Jahr oder Schulabgang ohne Abschluss maßgeblich waren, ließ sich im Rahmen dieser Untersuchung nicht ermitteln. Eine Aussage zu den schulformbezogenen Quoten der Klassenwiederholung ist demnach nicht möglich.

Da keine Leistungsdaten für Klassenwiederholer der Klassenstufen 9 und 10 vorliegen und vor allem diejenigen nicht erfasst sind, die nicht in die gymnasiale

Oberstufe bzw. in die Einführungsstufe des Aufbaugymnasiums übergetreten sind, werden die erreichten Lernstände *aller* ehemaligen Zehntklässlerinnen und Zehntklässler insgesamt erheblich niedriger sein als die in diesem Kapitel berichteten. Aus diesem Grunde beziehen sich alle nachfolgenden Analysen *nicht* auf Lernstandswerte am Ende der Klassenstufe 10, sondern nur auf Lernstandswerte von Schülerinnen und Schülern zu Beginn der Klassenstufe 11.

Die „Neuzugänge“ (Jugendliche, die an keiner der vorangegangenen Untersuchungen teilgenommen hatten) weisen in den Bereichen Deutsch-Leseverständnis, Mathematik I und Englisch (C-Test) im Vergleich zu den Schülerinnen und Schülern des Längsschnitts zu Beginn der Klassenstufe 11 im Mittel einen Leistungsrückstand auf, der im Leseverständnis mit einer Effektstärke von  $-0,26$  etwas geringer ist als in Mathematik und Englisch ( $-0,42$  bzw.  $-0,43$ ).

Vergleicht man die Mittelwertunterschiede beider Gruppen in den verschiedenen Formen der gymnasialen Oberstufe, so ergibt sich ein uneinheitliches Bild: In den Wirtschafts- und technischen Gymnasien verzeichnen die Neuzugänge unter den Schülerinnen und Schülern mit Versetzung in die gymnasiale Oberstufe vor allem im Leseverständnis und in Englisch einen nennenswerten Leistungsvorsprung gegenüber den Schülerinnen und Schülern des Längsschnitts, während die Neuzugänge in der gymnasialen Oberstufe an den Gymnasien in Mathematik, in der gymnasialen Oberstufe an den Gesamtschulen dagegen im Leseverständnis und in Englisch einen Leistungsrückstand gegenüber den Schülerinnen und Schülern des Längsschnitts zeigen.

Vergleichsweise homogen sind die Lernausgangslagen der Schülerinnen und Schüler beider Gruppen in der Einführungsstufe des Aufbaugymnasiums. Erwartungsgemäß liegen die mittleren Fachleistungen dieser Schülerschaft deutlich unterhalb der Lernstände in den Regelformen der gymnasialen Oberstufe, allerdings sind die Leistungsrückstände insbesondere im Vergleich zu den Schülerinnen und Schülern mit einem Mittleren Schulabschluss, die in die Vorstufe der Wirtschafts- und technischen Gymnasien eingetreten sind, eher gering, in Englisch liegt der Mittelwert sogar oberhalb des Mittelwertes dieser Vergleichsgruppe, was angesichts der Tatsache, dass die Schülerinnen und Schüler die einjährige Einführungsstufe zunächst erfolgreich durchlaufen müssen, bevor sie in die Vorstufe des Aufbaugymnasiums eintreten können, überrascht.

In Tabelle 10 sind die entsprechenden Mittelwerte, Standardabweichungen und Effektstärken wiedergegeben.

Tabelle 10 Fachleistungen der Neuzugänge in der Klassenstufe 11 im Vergleich mit den Schülerinnen und Schülern des Längsschnitts: Mittelwerte, Standardabweichungen und Effektstärken

		<b>Deutsch – Leseverständnis</b> Skalenpunkte und Gesamtstandard- abweichung (GSA)	<b>Mathematik</b> Skalenpunkte und Gesamtstandard- abweichung (GSA)	<b>Englisch – C-Test</b> Skalenpunkte und Gesamtstandard- abweichung (GSA)
<b>Gymnasium</b>	Längsschnitt	147,0 (N=2.998)	146,0 (N=2.986)	158,1 (N=2.966)
	Zugänge	145,7 (N= 493)	142,3 (N= 485)	158,0 (N= 475)
	GSA	15,1	15,7	14,5
	Effektstärke	-0,09	-0,24	-0,01
<b>Gesamt- schule</b>	Längsschnitt	143,7 (N= 707)	137,2 (N= 713)	146,5 (N= 704)
	Zugänge	140,9 (N= 147)	136,8 (N= 148)	143,9 (N= 144)
	GSA	14,0	14,6	14,5
	Effektstärke	-0,20	-0,03	-0,18
<b>WTG-G</b>	Längsschnitt	136,0 (N= 68)	137,2 (N= 69)	144,0 (N= 66)
	Zugänge	142,2 (N= 89)	138,5 (N= 87)	147,9 (N= 84)
	GSA	19,2	15,1	19,6
	Effektstärke	0,32	0,09	0,20
<b>WTG-R</b>	Längsschnitt	135,8 (N= 291)	133,2 (N= 293)	139,4 (N= 288)
	Zugänge	137,5 (N= 553)	131,0 (N= 548)	138,5 (N= 529)
	GSA	17,8	16,3	16,3
	Effektstärke	0,10	-0,13	-0,06
<b>Aufbau- gymnasium VS</b>	Längsschnitt	138,3 (N= 90)	135,5 (N= 89)	142,9 (N= 89)
	Zugänge	140,7 (N= 222)	135,0 (N= 216)	145,4 (N= 213)
	GSA	16,0	14,2	15,0
	Effektstärke	0,15	-0,04	0,17
<b>insgesamt</b>	<b>Längsschnitt</b>	145,3 (N=4.154)	143,2 (N=4.150)	154,3 (N=4.113)
	<b>Neuzugänge</b>	141,2 (N=1.504)	136,3 (N=1.484)	147,0 (N=1.445)
	<b>zusammen</b>	144,2 (N=5.658)	141,4 (N=5.634)	152,4 (N=5.558)
	<b>GSA</b>	16,0	16,5	16,8
	<b>Effektstärke</b>	-0,26	-0,42	-0,43
<b>Aufbau- gymnasium ES</b>	Längsschnitt	136,0 (N= 394)	130,3 (N= 395)	141,4 (N= 377)
	Neuzugänge	135,9 (N= 152)	129,0 (N= 160)	139,9 (N= 140)
	GSA	17,1	15,6	13,8
	Effektstärke	-0,01	-0,08	-0,11

### 3.2 Analyse der fachbezogenen Leistungstests

Auf der Grundlage jeweils *aller* zur Verfügung stehenden Aufgabenlösungen war es das Ziel, die erreichten Lernstände für alle getesteten Schülerinnen und Schüler insgesamt zu bestimmen. Im Rahmen der probabilistischen Testtheorie (Item Response Theory; vgl. FISCHER & MOLENAAR 1995; HAMBLETON, SWAMINATHAN & ROGERS 1991) wurden die Aufgaben der fachbezogenen Untertests *Deutsch: Leseverständnis, Mathematik I, Englisch: C-Test* und *Englisch: Leseverständnis* im *SL-HAM 10/11* unter Nutzung des einparametrischen Rasch-Modells skaliert. Im Unterschied zu den so genannten klassischen Methoden der Testauswertung wird bei diesem Verfahren nicht vorausgesetzt, dass alle Aufgaben gleich schwierig sind. Vielmehr wird das Leistungsniveau (die ‚Fähigkeit‘) einer Schülerin bzw. eines Schülers unter Berücksichtigung der populationsunabhängigen Schwierigkeit der gelösten Aufgaben ermittelt<sup>2</sup>. Das Verfahren erlaubt es,

- die Einschlägigkeit von Testaufgaben für die theoretisch bestimmte Fähigkeitsdimension empirisch zu prüfen und ungeeignete Items auszusondern,
- den Schwierigkeitsgrad von Testaufgaben und die Fähigkeit von Personen auf demselben Maßstab abzubilden und
- die Fähigkeit einer Person zuverlässig zu schätzen, auch wenn nur eine Teilmenge der Aufgaben bearbeitet wurde.

Die Fähigkeit einer Schülerin bzw. eines Schülers wird bei diesem Skalierungsverfahren über Wahrscheinlichkeiten definiert. Die geschätzte Fähigkeit ist dabei umso höher, je schwieriger die richtig gelösten Aufgaben sind. Die Festsetzung der Aufgabenschwierigkeiten wiederum hängt von der Anzahl richtiger Lösungen ab. Wenn sich nun die Antwortmuster der Schülerinnen und Schüler in guter Näherung so darstellen lassen, dass die Aufgaben stets nur bis zu einer für die einzelne Schülerin bzw. für den einzelnen Schüler bestimmbaren Schwierigkeit gelöst werden, können für alle Aufgaben und Fähigkeitsgruppen die Lösungswahrscheinlichkeiten berechnet werden (LEHMANN, PEEK, GÄNSFUß, HUSFELDT: 2002).

Nachdem die Aufgabenschwierigkeiten und die Schülerfähigkeiten innerhalb des Rasch-Modells auf dieselbe Skala projiziert sind, lassen sich die Aufgaben eines bestimmten Schwierigkeitsgrades als diejenige Leistung interpretieren, die von einer Schülerin bzw. einem Schüler des entsprechenden Fähigkeitsniveaus

---

<sup>2</sup> Eine gut nachvollziehbare Erklärung der Rasch-Skalierung findet sich bei BAUMERT, BOS & LEHMANN (2000, S. 60 ff.).

mit einer Wahrscheinlichkeit von  $p = 0,5$  erbracht wird<sup>3</sup>. Übersteigt die auf solcher Grundlage geschätzte Fähigkeit einer Schülerin oder eines Schülers die Schwierigkeit einer Aufgabe, so wird sie bzw. er diese mit umso größerer Wahrscheinlichkeit lösen, je größer die Differenz zwischen Fähigkeits- und Schwierigkeitsparameter ist. Unterschreitet hingegen die geschätzte Fähigkeit des Schülers oder der Schülerin die Aufgabenschwierigkeit, so wird ihm oder ihr, abhängig von der Differenz, mit umso höherer Wahrscheinlichkeit die Aufgabe misslingen.

Insgesamt haben 4.614 Schülerinnen und Schüler, die im Rahmen der LAU 11 getestet wurden, auch an der LAU 9 teilgenommen. Von ihnen wurden 3.811 Schülerinnen und Schüler bereits 1996 erstmals mit der LAU 5 erfasst (vgl. S. 22). Um Aussagen über die Lernentwicklung in allen Phasen der Untersuchung treffen zu können, wurden in den Untertests *Deutsch: Leseverständnis*, *Mathematik I* und *Englisch: C-Test* des *SL-HAM-10/11* die jeweilige Schwierigkeit der Ankeraufgaben nach Maßgabe der LAU 9 festgelegt. Dadurch konnte gewährleistet werden, dass die Metrik der LAU-Reihe „verlängert“ wird, und zwar so, dass nun die Ergebnisse für alle vier Erhebungszeitpunkte 1996 (LAU 5), 1998 (LAU 7), 2000 (LAU 9) und 2002 (LAU 11) auf einer einzigen Skala verortet sind. Bei der Fachleistung *Englisch*, die zu Beginn der Klassenstufe 5 noch nicht gemessen werden konnte, beschränkt sich die Verankerung auf die Klassenstufen 7, 9 und 11<sup>4</sup>.

Die Schwierigkeitsparameter der Ankeritems für die Leistungen der Neuntklässlerinnen und Neuntklässler in den Bereichen *Mathematik I*, *Deutsch: Leseverständnis* und *Englisch: C-Test* lagen bei der Skalenbildung 2003 bereits vor, deshalb wurde die nunmehr erweiterte Metrik so geeicht, dass diese Werte gleich geblieben sind, während sie für neue, in der elften Klassenstufe erstmals eingesetzte Items diesem Maßstab entsprechend berechnet werden konnten.

---

<sup>3</sup> Die Festlegung dieser Lösungswahrscheinlichkeit für die Zuordnung von Fähigkeits- und Schwierigkeitsparametern beruht auf einer pragmatisch begründeten Entscheidung. Mit der Wahl eines strengeren Maßstabs (vgl. z. B. die in der TIMS-Studie festgelegte Lösungswahrscheinlichkeit von 0,65) verringert sich schülerseitig das Fähigkeitsniveau nur formal. Die „Beherrschung“ eines bestimmten Schwierigkeitsniveaus ist dann an niedrigere Fehlerquoten gebunden. Die Unterschiede innerhalb der Personengruppen und zwischen ihnen ändern sich dabei nicht.

<sup>4</sup> Das Verfahren des Fixierens von Ankeritems in einer Längsschnittuntersuchung ist bei NIKOLOVA & LEHMANN (2003) erläutert.

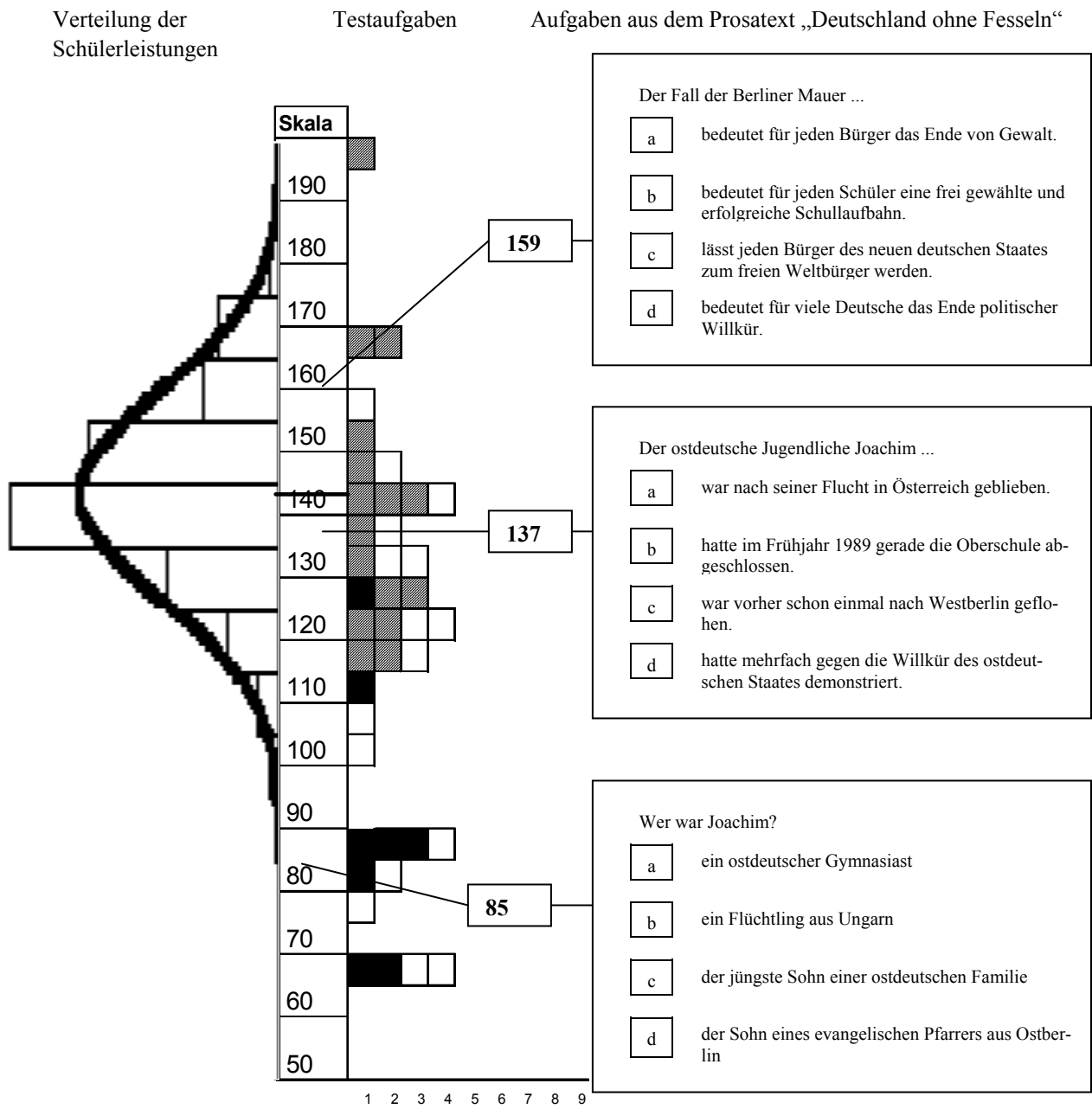
### 3.2.1 Fachleistung Deutsch: Leseverständnis

Die für den Deutschunterricht relevanten Fachleistungen werden zum einen mit einem Leseverständnistest und zum anderen mit einem Rechtschreibtest zur Ermittlung des passiven Rechtschreibwissens im *SL-HAM 10/11* erfasst. Die hierauf bezogenen Befunde für die Schülerinnen und Schüler zu Beginn der Klassenstufe 11 sollen im Folgenden vorgestellt werden.

Die Fähigkeiten der Schülerinnen und Schüler im Leseverständnis sowie die Eigenschaften der Aufgaben des *Leseverständnistests* sind, wie die in den Untertests *Mathematik I* und *Englisch-C-Test* gemessenen Fähigkeiten, über die Rasch-Skalierung zugleich miteinander verschränkt und aufeinander projizierbar. Die Gesamtskala für das Leseverständnis erfasst auf der Testseite unterschiedliche Textformen (Prosa-, Sach- und Gebrauchstexte) und Schwierigkeitsgrade, auf der Verständnissseite unterschiedliche Fähigkeits- und Leistungsniveaus. Dabei ist es möglich und sinnvoll, die Verteilung der Aufgabenschwierigkeiten derjenigen des gemessenen Leseverständnisses – wie bereits in der LAU 9 – grafisch gegenüberzustellen. Abbildung 1 zeigt das Ergebnis für die Gesamtgruppe der getesteten Schülerinnen und Schüler, wobei die einzelnen Aufgaben nach den unterschiedlichen Textformen ausgewiesen sind, auf die sie sich beziehen.

Bei einem Mittelwert von 119,1 auf der Seite der Aufgabenschwierigkeiten hat sich eine breite Verteilung von 67 bis 197 Skalenpunkten ergeben. Im Vergleich zu den Prosa- und Sachtextaufgaben konzentrieren sich die Fragestellungen, die sich auf Gebrauchstexte beziehen, deutlicher auf die untere Hälfte der Skala. Offensichtlich ist die Ursache in der Textform selbst zu finden: Gebrauchstexte werden verfasst, um elementare Informationen möglichst unmissverständlich mitzuteilen. Es war – nicht zuletzt aufgrund der Erfahrungen aus der LAU 9 – zu erwarten, dass die Jugendlichen mit solchen Texten verhältnismäßig wenig Schwierigkeiten haben würden und dass deshalb auch die Aufgabenschwierigkeiten in diesem Bereich nur eine geringe Streuung zeigten. Ein geringeres Schwierigkeitsgefälle besteht zwischen den Aufgaben zu literarischen Texten und solchen zu Sachtexten. Die Aufgaben zu den Sachtexten sind im Mittel die anspruchsvollsten und sind eher dem oberen Bereich der Skala (115 bis 190) zuzuordnen. Die prosatextspezifischen Aufgabenstellungen sind in ihren Anforderungen am weitesten gestreut (65 bis 160) und – bis auf den obersten Bereich – auf die gesamte Skala verteilt. Diese Aufgaben erfordern, so lässt sich aufgrund der breiten Verteilung der Aufgabenschwierigkeiten schlussfolgern, sehr spezielle Verständnisleistungen.

Abbildung 1 Verteilung der Schülerleistungen in Deutsch-Leseverständnis im Vergleich mit den Schwierigkeiten der Testaufgaben im *SL-HAM-10/11-Leseverständnistest* (nach Textformen)



Gesamtstichprobe (N = 6.204)

Textform:

■ Gebrauchstext    ▨ Sachtext    □ Prosatext

Leistungsmittelwert aller Schülerinnen und Schüler (einschließlich der Schülerinnen und Schüler aus der Einführungsstufe des Aufbaugymnasiums): 143,5.

Die Kurve zeigt die Aufteilung der Fachleistungen, wenn sie normalverteilt wären.

Im Gegensatz zur LAU 9, bei der die Verteilungskurve der Leistungen in Deutsch-Leseverständnis rechtssteil war – einige Schülerinnen und Schüler hatten extrem schwache Verständnisseleistungen, während verhältnismäßig viele gute und sehr gute Werte erzielt werden konnten –, ist in der LAU 11 das Leseverständnis annähernd normalverteilt: Nur wenige Jugendliche haben extrem schwache Verständnisseleistungen gezeigt, doch es konnten auch nur wenige Schülerinnen und Schüler sehr gute Werte erzielen. Der Mittelwert für das Leseverständnis liegt mit einem Raschwert von 143,5 (Standardabweichung: 16,3) deutlich höher als die mittlere Aufgabenschwierigkeit (119,1). Dabei ist zu berücksichtigen, dass dieses scheinbar günstige Leistungsbild vor allem durch Selektionsprozesse verursacht wurde, die an der Gelenkstelle zwischen Sekundarstufe I und II stattfanden.

Um möglichst den gesamten Anforderungsbereich von einfachen Decodierungsaufgaben bis hin zum eigenständigen schlussfolgernden Umgang mit Texten zu erfassen, wurden bei der Testkonzeption wiederum unterschiedlich schwierige Items entwickelt. Wenn man die jeweils geforderten Leistungen genauer untersucht, wird dieser Sachverhalt auch empirisch bestätigt. Die Kompetenzstufen, die von einfachen Aufgaben der Informationsentnahme zu komplexen mentalen Operationen reichen, entsprechen den neueren Theorien des Leseverständnisses. Immer erscheinen solche Aufgaben am leichtesten, die mit dem Auffinden eines Wortes, einer kürzeren Passage oder auch eines Symbols lediglich einen direkten Abgleich zwischen Text und Frage verlangen. Die Schwierigkeit steigt in dem Maße, wie komplexe Randbedingungen präsent gehalten werden müssen und wie der Gehalt umfangreicherer Passagen zunächst verdichtet und gegebenenfalls weiterverarbeitet werden muss (LEHMANN, PEEK, GÄNSFUß UND HUSFELDT: 2002).

In der Abbildung 1 sind drei Beispielaufgaben aus dem Prosatext „*Deutschland ohne Fesseln*“ ausgewiesen, die das Prinzip der Hierarchie zunehmend schwierigerer Anforderungen verdeutlichen. Die Frage: „*Wer war Joachim?*“ (Beispielaufgabe 1) erfordert auf einer relativ niedrigen Anforderungsstufe mit dem Raschwert von 85 die Fähigkeit, eine im gleichen Wortlaut gegebene Information aufzufinden, die direkt im zweiten Absatz des Textes steht (richtige Antwort: *der Sohn eines evangelischen Pfarrers aus Ostberlin*). 99,7 Prozent aller getesteten Schülerinnen und Schüler haben diese Fähigkeitsstufe der einfachen Informationsentnahme erreicht bzw. übertroffen. Beispiel 2 ist mit der Aufgabenformulierung „*Der ostdeutsche Jugendliche Joachim ...*“ deutlich anspruchsvoller. Die richtige Antwort „*... hatte mehrfach gegen die Willkür des ostdeutschen Staates demonstriert.*“ mit einem Raschwert von 137 verlangt die Fähigkeit, aus dem Text einen Sachverhalt zu rekonstruieren, wobei die Formu-



lierungen in Aufgabe und Text sprachlich unterschiedlich sind und aus vorgegebenen Einzelsituationen eine Verallgemeinerung generiert werden muss. Oberhalb dieses Skalenwertes lagen 65,6 Prozent der getesteten Jugendlichen. Die dritte Beispielaufgabe: „*Der Fall der Berliner Mauer ...*“ (Raschwert 159) steht für ein vergleichsweise anspruchsvolles Fähigkeitsniveau, das mit knapp 20 Prozent nur noch von einer Minderheit der Schülerinnen und Schüler erreicht bzw. übertroffen wurde; von 81,3 Prozent der Getesteten wurde dieses Niveau nicht erreicht. Die richtige Antwort: „*bedeutet für viele Deutsche das Ende politischer Willkür.*“ erfordert auf einem hohen Niveau – über die vorgelagerten Leseverständnisstufen hinaus – die Fähigkeit, Gelesenes umzusetzen, Verknüpfungen zu leisten und Schlussfolgerungen zu ziehen.

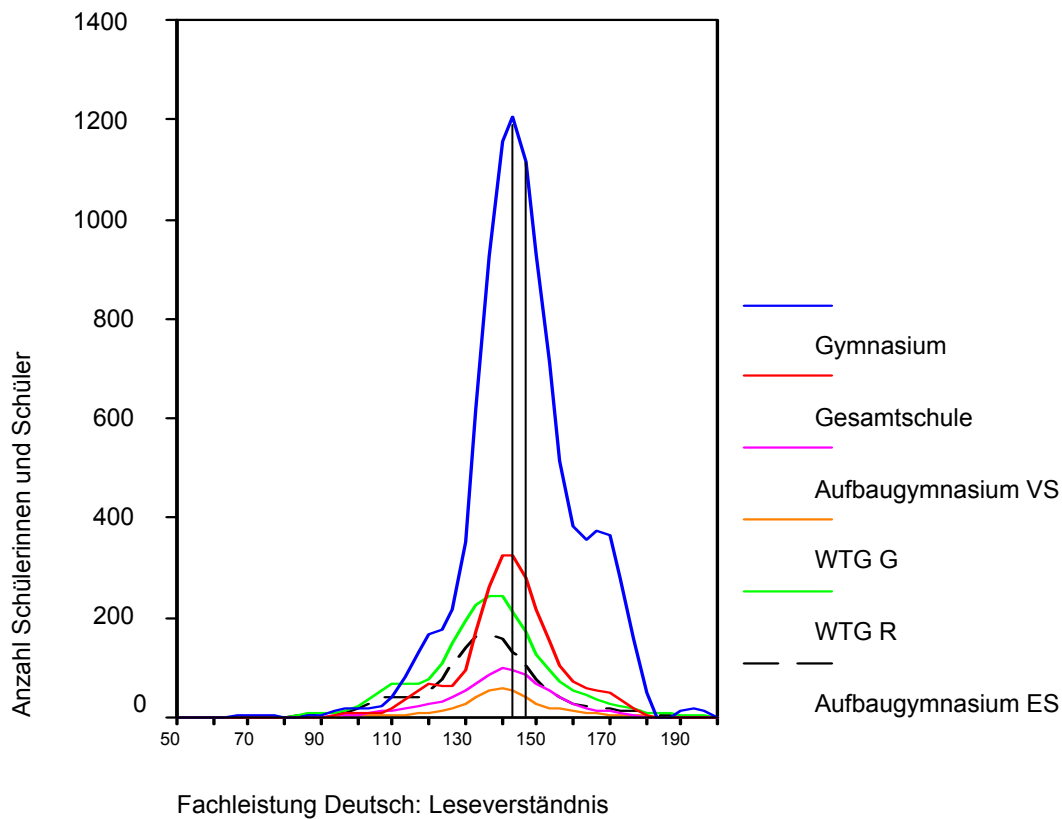
### **Leseverständnis, differenziert nach Bildungsganggruppen**

Im Folgenden soll – differenziert für die Schülerinnen und Schüler aus den unterschiedlichen Bildungsganggruppen – die Verteilung der erreichten Leistungen im Leseverständnistest zu Beginn der Klassenstufe 11 bzw. zu Beginn der Einführungsstufe des Aufbaugymnasiums dargestellt und diskutiert werden.

In der Abbildung 2 sind für die insgesamt 6.204 berücksichtigten Jugendlichen die gruppenbezogenen Verteilungen eingezeichnet. Dabei markiert die rechte Senkrechte den Mittelwert der grundständigen Gymnasiastinnen und Gymnasiasten; die linke Senkrechte stellt den allgemeinen Mittelwert aller Schülerinnen und Schüler der Klassenstufe 11 dar (ohne die Schülerinnen und Schüler aus der Einführungsstufe des Aufbaugymnasiums).

Bei der Interpretation dieser Grafik ist die unterschiedliche Verteilung der Jugendlichen in den verschiedenen Bildungsganggruppen zu berücksichtigen (61,7 Prozent der Schülerschaft befindet sich in der Gruppe „Gymnasium“ im Unterschied beispielsweise zu den 2,8 Prozent, die der Gruppe „WTG-G“ angehören). Die angenäherte Normalverteilung, die schon in Abbildung 1 für die Gesamtgruppe der Schülerinnen und Schüler zu erkennen war, zeigt sich auch bei der grafischen Darstellung der einzelnen Bildungsganggruppen. Zu erkennen ist aber auch, dass eine Mehrheit der Jugendlichen aus der Gruppe „Gymnasium“ über dem allgemeinen Mittelwert liegt, der sich auf die Gesamtheit aller Schülerinnen und Schüler der Klassenstufe 11 bezieht. Obwohl viele Schülerinnen und Schüler dieser Gruppe die Mehrzahl der Aufgaben gelöst haben, kann man nicht von einem Deckeneffekt sprechen, da „perfekte“ Ergebnisse selten sind.

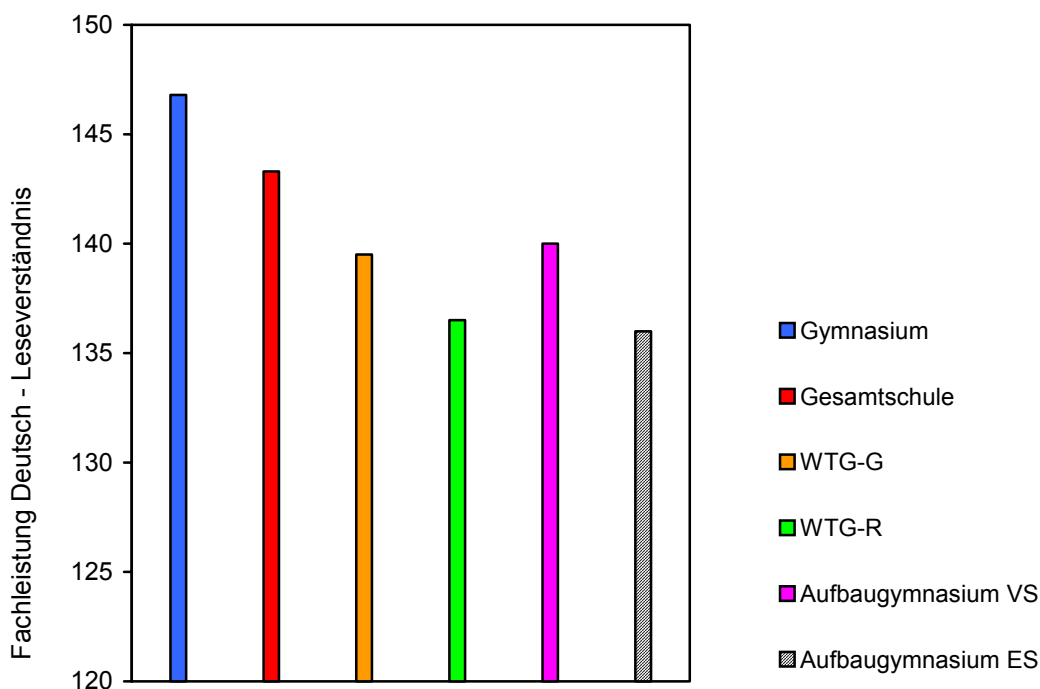
Abbildung 2 Deutsch-Leseverständnis: Verteilung der Lernstände zu Beginn der Klassenstufe 11 nach Bildungsgang (absolut)



Die Abbildung zeigt, dass der Anteil derjenigen Jugendlichen, die den Mittelwert der grundständigen Gymnasiastinnen und Gymnasiasten erreichen bzw. übertreffen (rechte Senkrechte), zwischen den verschiedenen Bildungsganggruppen deutlich variiert. So haben 39,9 Prozent der grundständigen Gesamtschülerinnen und Gesamtschüler den Leistungsdurchschnitt der grundständigen Gymnasiastinnen und Gymnasiasten von 146,8 Skalenpunkten erreicht oder übertroffen, 31,2 Prozent der Schülerinnen und Schüler aus der Gruppe WTG-G, 35,9 Prozent aus der Gruppe Aufbaugymnasium VS sowie 27,1 Prozent aus der Gruppe WTG-R; unter den Schülerinnen und Schülern aus der Einführungsstufe des Aufbaugymnasiums sind es 24,2 Prozent.

In Abbildung 3 sind die mittleren Lernstände, die von den Schülerinnen und Schülern der verschiedenen Bildungsganggruppen erreicht worden sind, in einem Säulendiagramm dargestellt.

Abbildung 3 Deutsch-Leseverständnis: Verteilung der Lernstände zu Beginn der Klassenstufe 11 nach Bildungsgang



Die in der Tabelle 11 dargestellten statistischen Kennwerte für das erreichte Leseverständnis zeigen die Wertigkeit der Mittelwertunterschiede zu Beginn der Klassenstufe 11 bzw. zu Beginn der Einführungsstufe des Aufbaugymnasiums. Die durchschnittlichen Leistungen der Gymnasiastinnen und Gymnasiasten (Mittelwert: 146,8) liegen – bei der insgesamt zweitgeringsten Leistungsstreuung ( $s = 15,1$ ) – mit einer Effektstärke von  $d = 0,16$  oberhalb des Mittelwertes aller Getesteten, gefolgt von den Gesamtschülerinnen und Gesamtschülern. Die Schülerinnen und Schüler der Gruppen WTG-G und Aufbaugymnasium VS liegen mit Effektstärken von  $d = -0,29$  bzw.  $d = -0,26$  unterhalb des allgemeinen Leistungsdurchschnitts und damit  $-0,45$  bzw.  $-0,42$  Standardabweichungen unter der mittleren Leistung der grundständigen Gymnasiastinnen und Gymnasiasten.

Auffällig ist der geringe Leistungsunterschied zwischen den Schülerinnen und Schülern der Gruppe WTG-R und den Schülerinnen und Schülern, die die Einführungsstufe des Aufbaugymnasiums besuchen. Beide Gruppen liegen mit  $d = -0,46$  bzw.  $d = -0,51$  deutlich unterhalb des Leistungsdurchschnitts aller Vorstufenschülerinnen und -schüler und etwa zwei Drittel einer Standardabweichung unterhalb der mittleren Leistung der grundständigen Gymnasiastinnen und Gymnasiasten. Dies entspricht einerseits der vergleichbaren Vorbildung der Schülerinnen und Schüler (Mittlerer Schulabschluss mit dem Notendurchschnitt 3,0 über alle Fächer), andererseits ist zu bedenken, dass die Schülerinnen und Schüler aus der Einführungsstufe erst nach erfolgreichem Durchlaufen dieser

Vorbereitungsphase in die gymnasiale Oberstufe des Aufbaugymnasiums versetzt werden. Nimmt man die Mittelwertdifferenz zwischen den Schülerinnen und Schülern aus der Vorstufe und aus der Einführungsstufe des Aufbaugymnasiums als Maßstab, so sollte man eigentlich bei denjenigen, die unmittelbar aus einer Realschule in die Vorstufe eines Fachgymnasiums eingetreten sind, einen um etwa eine viertel Standardabweichung höheren Lernstand erwarten.

Tabelle 11 Deutsch-Leseverständnis zu Beginn der Klassenstufe 11: Mittelwerte, Standardabweichungen und Effektstärken, insgesamt und getrennt nach Bildungsgang

	Mittelwert	Standardabweichung	Effektstärke $d^5$	<i>N</i>
<b>Gymnasium</b>	146,8	15,1	0,16	3.491 (61,7 %)
<b>Gesamtschule</b>	143,3	14,0	-0,06	854 (15,1 %)
<b>WTG-G</b>	139,5	19,2	-0,29	157 (2,8 %)
<b>WTG-R</b>	136,9	17,8	-0,46	844 (14,9 %)
<b>Aufbaugymnasium VS</b>	140,0	16,0	-0,26	312 (5,5 %)
<b>insgesamt</b>	144,2	16,0	—	5.658 (100 %)
<b>Aufbaugymnasium ES</b>	136,0	17,1	-0,51	546

<sup>5</sup> Die *Effektstärke d* ist hier so berechnet worden, dass die Differenz zwischen dem Mittelwert einer Bildungsganggruppe und dem Gesamtmittelwert durch die Gesamtstandardabweichung geteilt wurde (ohne die Leistungen der Schülerinnen und Schüler aus der Gruppe Aufbaugymnasium ES). Dies hat den Vorzug, dass Unterschiede zwischen den Bildungsganggruppen unmittelbar aus den Differenzen zwischen den einzelnen Effektstärken ermittelt werden können (vgl. LEHMANN, GÄNSFUß, PEEK 1999, S. 50). Z. B. beträgt die Differenz zwischen den Schülerschaften der Gruppen Gymnasium und Gesamtschule 0,22 (Gesamt-) Standardabweichungen, während die Differenz zwischen der Schülerschaft der Gruppen Gesamtschule und Aufbaugymnasium VS 0,20 Standardabweichungen beträgt. Wählte man die Standardabweichung innerhalb der Bildungsganggruppe zum Maßstab, würden die *d*-Koeffizienten etwas größere Werte annehmen, ohne zu anderen Verhältnissen zu kommen. Mit der gewählten Darstellungsweise liegt ein standardisiertes Maß für die Differenzen zwischen den Bildungsganggruppen vor, das sich auch für Vergleiche zwischen den verschiedenen Aspekten der Fachleistung heranziehen lässt.

### 3.2.2 Fachleistung Deutsch: Rechtschreiben

Die Rechtschreibkenntnisse der Schülerinnen und Schüler wurden in LAU 11 als *passives Rechtschreibwissen* erhoben. Das passive Rechtschreibwissen wurde im *SL-HAM 10/11* gruppenübergreifend in der anspruchsvollen Form der Markierung von Fehlschreibungen in einem vorgegebenen Text geprüft. Dabei handelte es sich um einen Sachtext zum Thema „*Jugend und Freizeit an der Jahrtausendwende*“, der bereits in der LAU 9 eingesetzt worden war und insgesamt 285 Wörter umfasste. In diesem Text wird das Freizeitverhalten von Jugendlichen unter soziologischen Aspekten thematisiert. Insgesamt enthielten 30 der 285 Wörter einen Rechtschreibfehler. Die Testteilnehmerinnen und Testteilnehmer hatten die Aufgabe, die Fehlschreibungen zu unterstreichen.

Im Folgenden ist – um einen Einblick in die Anforderungsstruktur des Rechtschreibtests zu geben – eine Auswahl von sechs der falsch geschriebenen Wörter mit den dazugehörigen Anteilen derjenigen Schülerinnen und Schüler aufgeführt, die sie als Fehlschreibungen erkannt haben<sup>6</sup>.

Von den insgesamt 6.043 Jugendlichen, die den Test bearbeiteten, haben annähernd 90 Prozent die folgenden beiden fett und kursiv gesetzten Wörter als Fehlschreibungen markiert. In einem Fall handelt es sich um die fehlende Kennzeichnung eines kurzen Vokals durch eine Konsonantenkombination:

- ... und vor allem ohne **Leistungsdruk** ... (89,8 %),

im anderen Fall um die Schreibung des gedehnten *i* im Suffix *-ieren* bei Fremdwörtern:

- ... stärker an der Arbeit und der Familie **orientirt**, ... (93,3 %).

Ca. 75 Prozent der Getesteten war geläufig, dass das Partizip I bei „*anstrengend*“ mit dem Suffix *-d* endet; ebenfalls etwa drei Viertel der Schülerinnen und Schüler haben die Zusammenschreibung zweier getrennt zu schreibender Wörter als Fehler erkannt:

- *Freizeit ist für die heutige Jugend ganz schön **anstrengent***. (77,5 %);
- **Dazugehört** dann die entsprechend modische Sportkleidung, ... (74,1 %).

---

<sup>6</sup> Da der Rechtschreibtest für alle Bildungsganggruppen einheitlich war, konnte auf eine Rasch-Skalierung verzichtet werden; sie hätte bei der großen Zahl der Items auch zu Problemen geführt. Stattdessen wurden bei der Auswertung fälschlich unterstrichene Richtigschreibungen so berücksichtigt, dass jeder erkannte Rechtschreibfehler mit dem Faktor „Anzahl richtiger Markierungen geteilt durch Anzahl aller Markierungen“ gewichtet wurde (LEHMANN, PEEK, GÄNSFUß UND HUSFELDT: 2002, S. 45).

Schließlich gibt es Fehlschreibungen, die von ungefähr der Hälfte der getesteten Jugendlichen erkannt wurden. In dem folgenden Fall:

- ... eine starke **Enttäuschung** am Staat ... (50,9 %)

gilt es zu wissen, dass bei der Zusammensetzung von Wörtern die Morpheme an der Fugengrenze beibehalten werden, und bei einem der schwierigsten Wörter muss das Stammprinzip beherrscht werden:

- „...“, **ergenz**t Horst Opaschowski. (47,8 %).

52,2 Prozent der Schülerinnen und Schüler haben nicht erkannt, dass in der Ableitung der Stammvokal *a* der Wurzel „ganz“ in *ä* umgelautet wird.

Auch bei dem Untertest Deutsch-Rechtschreiben wird deutlich, dass es erhebliche Leistungsunterschiede zwischen den verschiedenen Gruppen gibt (vgl. Tabelle 12).

Tabelle 12 Passives Rechtschreibwissen zu Beginn der Klassenstufe 11: Mittelwerte, Standardabweichungen und Effektstärken, insgesamt und getrennt nach Bildungsgang

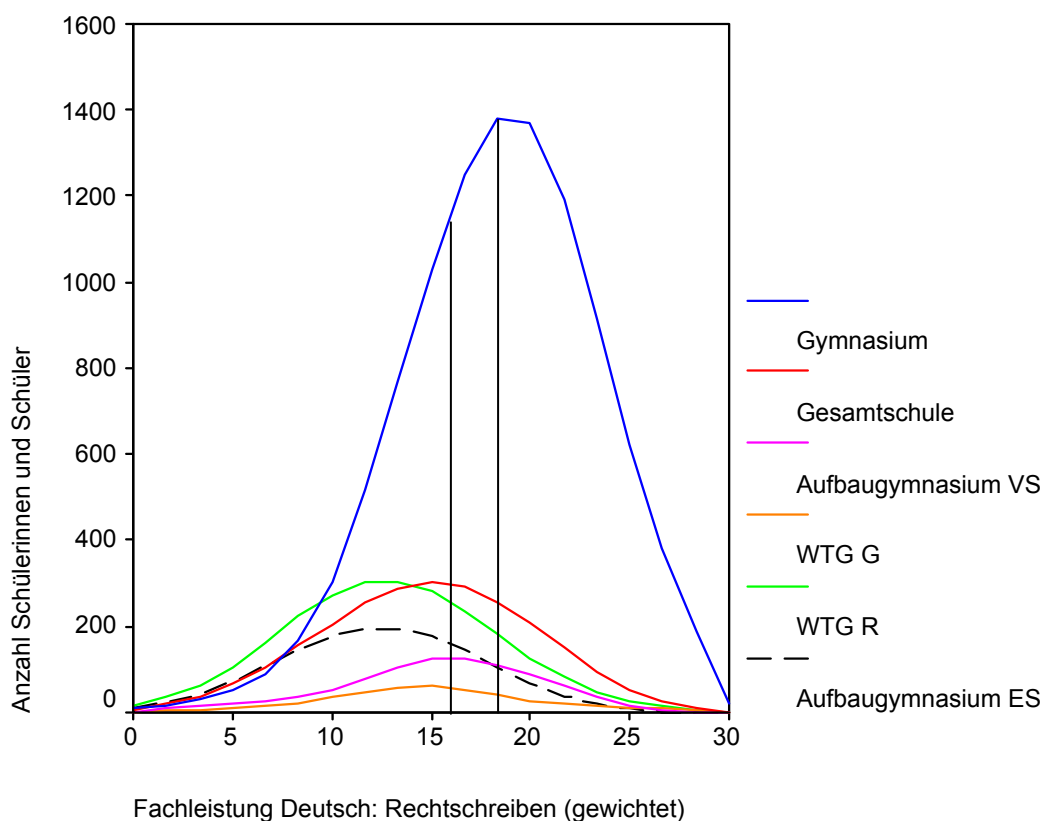
	Mittelwert	Standardabweichung	Effektstärke <i>d</i>	<i>N</i>
<b>Gymnasium</b>	18,3	4,6	0,28	3.407 (61,6 %)
<b>Gesamtschule</b>	14,9	5,1	-0,36	844 (15,3 %)
<b>WTG-G</b>	15,0	5,4	-0,34	148 (2,7 %)
<b>WTG-R</b>	12,9	5,0	-0,74	827 (14,9 %)
<b>Aufbaugymnasium VS</b>	15,5	5,0	-0,25	304 (5,5 %)
<b>Insgesamt</b>	16,8	5,3	–	5.530 (100 %)
<b>Aufbaugymnasium ES</b>	12,5	4,9	-0,81	513

Bei anteiliger Berücksichtigung fälschlich unterstrichener Richtigschreibungen ergibt sich – bei insgesamt 30 vorgegebenen Fehlern – für alle getesteten Vorstufenschülerinnen und -schüler insgesamt ein Durchschnitt von 16,8 richtig erkannten Fehlschreibungen. Die grundständigen Gymnasiastinnen und Gymnasiasten

haben als leistungsstärkste Gruppe etwa 60 Prozent der vorgegebenen Fehlschreibungen erkannt. Mit einem Mittelwert von 18,3 und einer Effektstärke von  $d = 0,28$  liegen sie über dem Gesamtdurchschnitt und mehr als eine halbe Standardabweichung vor der nächststärksten Gruppe, den Vorstufenschülern des Aufbaugymnasiums. Die Schülerinnen und Schüler der Gruppen Gesamtschule und WTG-G konnten annähernd die Hälfte aller Fehlschreibungen identifizieren. Die leistungsschwächste Gruppe unter den Vorstufenschülerinnen und -schülern sind die Testteilnehmerinnen und Testteilnehmer aus der Gruppe WTG-R, die drei Viertel einer Standardabweichung unter dem Gesamtmittelwert und sogar mehr als eine Standardabweichung unter dem Mittelwert der grundständigen Gymnasiastinnen und Gymnasiasten liegen. Wie im Leseverständnis, so ist auch im passiven Rechtschreibwissen die Mittelwertdifferenz zwischen den Gruppen WTG-R und Aufbaugymnasium ES mit 0,07 Standardabweichung äußerst gering.

Wie bereits für den Leseverständnistest, so lässt sich auch für den Rechtschreibtest kein Deckeneffekt nachweisen, da es keine Schülerinnen und Schüler mit perfekten Antworten gab. Abbildung 4 zeigt für alle Bildungsganggruppen eine annähernde Normalverteilung. Die linke Senkrechte zeigt den Mittelwert aller Vorstufenschülerinnen und -schüler.

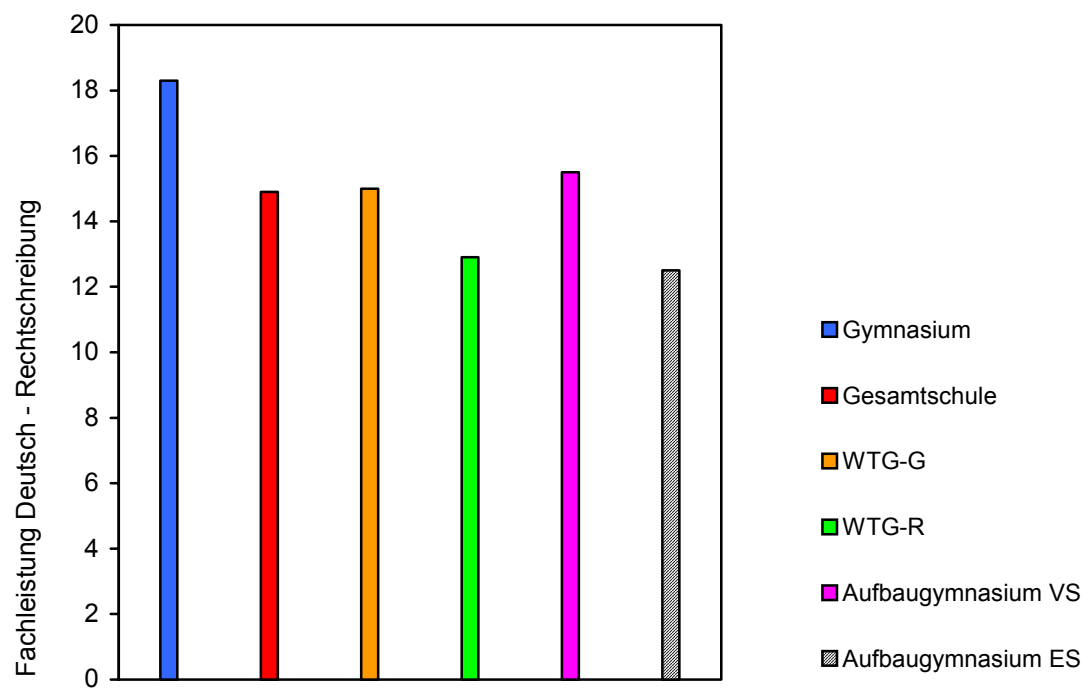
Abbildung 4 Passives Rechtschreibwissen: Verteilung der Lernstände zu Beginn der Klassenstufe 11 nach Bildungsgang (absolut)



Deutlich zu erkennen ist, dass nur ein geringer Teil der Jugendlichen aus den anderen Bildungsganggruppen den Mittelwert der grundständigen Gymnasias-tinnen und Gymnasiasten erreicht bzw. übertrifft (rechte Senkrechte). Den Leis-tungsdurchschnitt der grundständigen Gymnasiastinnen und Gymnasiasten von 18,3 Punkten haben 23,8 Prozent der grundständigen Gesamtschülerinnen und Gesamtschüler, 23,6 Prozent der Schülerinnen und Schüler aus der Gruppe WTG-G, 26,6 Prozent aus der Gruppe Aufbaugymnasium VS und 14,6 Prozent aus der Gruppe WTG-R erreicht oder übertroffen, unter den Schülerinnen und Schülern aus der Einführungsstufe des Aufbaugymnasiums sind es 11,1 Prozent.

In der Abbildung 5 ist die Verteilung der Schülerleistungen für die einzelnen Bildungsganggruppen noch einmal in einem Säulendiagramm dargestellt.

Abbildung 5 Passives Rechtschreibwissen: Verteilung der Lernstände zu Be-ginn der Klassenstufe 11 nach Bildungsgang





### 3.2.3 Fachleistung Mathematik I

Nach dem oben dargestellten Auswertungsverfahren lassen sich auch die Struktur des Mathematik-I-Tests auf der einen Seite und die Fähigkeitsstruktur der Schülerinnen und Schüler auf der anderen Seite anschaulich darstellen. Unter der Voraussetzung der Eindimensionalität (vgl. BAUMERT, BOS & LEHMANN 2000, S. 94 ff.) wurden auch hier die Schwierigkeiten der einzelnen Aufgaben der Verteilung der gemessenen mathematischen Testleistungen grafisch gegenübergestellt und dabei gleichzeitig die verschiedenen Themenbereiche kenntlich gemacht. Die Ergebnisse im Mathematik-I-Test, ergänzt um Beispielaufgaben aus dem Bereich Algebra, sind der Abbildung 6 zu entnehmen.

Auf der rechten Seite der Skala sind die 59 Testaufgaben<sup>7</sup> nach ihrem testanalytisch bestimmten Schwierigkeitsgrad geordnet eingetragen. So liegen z. B. sechs Aufgaben aus den Bereichen *Arithmetik*, *Algebra* und *Geometrie* mit geringerer Schwierigkeit im Skalenbereich 80 bis 100, während sich die Aufgaben zur *Stochastik* mit relativ hoher Schwierigkeit im Skalenbereich 150 bis 170 befinden. Die schwersten Aufgaben (*Algebra* und *Geometrie*) liegen im Skalenbereich 170 bis 190.

Die Komplexität der Aufgaben, die Plausibilität der angegebenen Alternativantworten und der Einbezug nicht im engeren Sinne mathematischer Fähigkeiten wie z. B. des Leseverständnisses gehören zu einer Vielzahl von Faktoren, die letztlich die Aufgabenschwierigkeit bestimmen. Wenn über den gesamten Leistungsbereich hinweg mit gleicher Präzision gemessen werden soll, wäre eine Gleichverteilung der Aufgabenschwierigkeiten ideal. Hingegen hat sich das dominante Motiv realistischer Lösungserwartungen bei der Auswahl von geeigneten Aufgaben dahingehend ausgewirkt, dass sich eine Verteilung der Aufgabenschwierigkeiten mit relativ vielen Aufgaben im mittleren Bereich eingestellt hat.

Die Aufgaben des Mathematik-I-Tests sind weitgehend über die gesamte Skala verteilt. Allerdings sind die Aufgaben aus dem Bereich *Arithmetik* eher in der unteren Hälfte (leichtere Aufgaben) zu finden, während die Aufgaben aus dem Bereich *Geometrie* und die wenigen (vor allem für die leistungsstärkeren Schülerinnen und Schüler bestimmten) Aufgaben aus der *Stochastik* sich in der oberen Hälfte der Skala konzentrieren.

Die Verteilung der mathematischen Leistungen der Testteilnehmerinnen und Testteilnehmer ist auf der linken Seite der Skala abgebildet. Wie Abbildung 6 zeigt, sind die Ergebnisse annähernd normalverteilt: Nur wenige Schülerinnen und Schüler haben extrem schwache bzw. (nahezu) perfekte Leistungen gezeigt,

---

<sup>7</sup> Eine Aufgabe konnte aus teststatistischen Gründen nicht in die Analyse einfließen.

während viele Leistungen im Bereich um den Skalenmittelwert 140,3 (Standardabweichung  $s = 16,7$ ) liegen.

Die Anforderungen an die Schülerinnen und Schüler im Aufgabenbereich *Algebra* verlangen bereits in der Mittelstufe sichere Kenntnisse beim Lösen von einfachen Gleichungen und bei Grundrechenoperationen mit Dezimalbrüchen. Schülerinnen und Schüler in der gymnasialen Oberstufe sollten Rechengesetze im Umgang mit Klammern und Variablen sicher beherrschen sowie Teilbarkeitsregeln anwenden und Geradengleichungen lösen können.

Um einen Eindruck von den Aufgabenstellungen des Tests und ihren unterschiedlichen Anforderungsniveaus zu geben, sind in die Abbildung 6 Beispielaufgaben aus dem Bereich *Algebra* eingetragen. Die Darstellung ist so gewählt, dass die Aufgaben bestimmten Fähigkeitsniveaus zugeordnet werden, die erreicht sein müssen, um sie mit einiger Sicherheit – d. h. mit einer Wahrscheinlichkeit von 50 Prozent – lösen zu können.

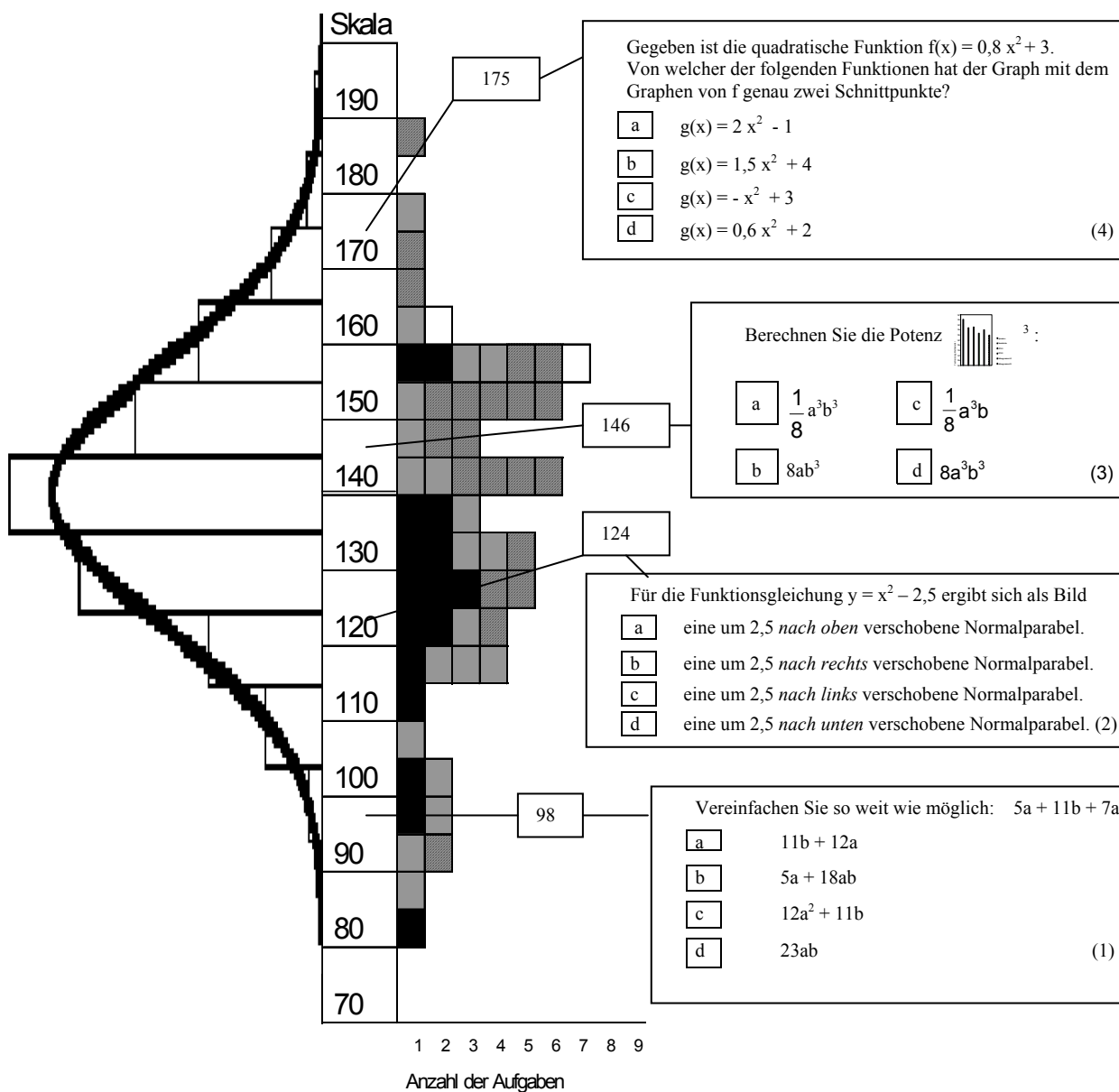
Die Datenauswertung hat ergeben, dass z. B. das Leistungsniveau mit dem Skalenwert 98 nahezu von allen Schülerinnen und Schülern (99,3 Prozent) erreicht oder überschritten wurde. Die Beispielaufgabe 1 mit der Anforderung, eine Additionsaufgabe mit zwei Unbekannten zu vereinfachen, ist eine Aufgabe dieser Art und Schwierigkeit.<sup>8</sup> Die Aufgabe 2, bei der von den Schülerinnen und Schülern eine gegebene Funktionsgleichung analysiert werden sollte, um die dazugehörige Grafik zu finden, ist ein Beispiel für ein Fähigkeitsniveau von 124, das 84,9 Prozent der Getesteten erreicht bzw. übertroffen haben. Die Lösung einer Aufgabe, die sowohl den sicheren Umgang mit Potenzen als auch die Beherrschung von Grundrechenarten bei Berücksichtigung der Klammerregeln erfordert, gelang mit einiger Sicherheit ab einem Skalenwert von 146; die Beispielaufgabe 3 spiegelt das Leistungsniveau wider, das 34,4 Prozent der Getesteten erreicht oder überschritten haben. Dem Skalenwert von 175 entspricht die Beispielaufgabe 4, bei der diejenige quadratische Funktion ermittelt werden sollte, die mit der quadratischen Funktion in der Aufgabenstellung genau zwei Schnittpunkte bildet. Hier liegt eine typische Aufgabe für ein mathematisches Leistungsniveau vor, das 2,5 Prozent der Schülerinnen und Schüler erzielt oder überschritten haben.

---

<sup>8</sup> Die hier genannten Prozentangaben beziehen sich nicht auf die einzelnen Aufgaben, sondern auf das erreichte Schwierigkeitsniveau, für das die Beispielaufgaben stellvertretend stehen.

Abbildung 6 Verteilung der Schülerleistungen im Fach Mathematik im Vergleich mit den Schwierigkeiten der Testaufgaben im *SL-HAM-10/11-Mathematik-I-Test* (nach Themenbereichen)

Verteilung der Schülerleistungen      Testaufgaben      Aufgaben aus dem Bereich Algebra



Gesamtstichprobe (6.189)

Themenbereiche:

Arithmetik       Algebra       Geometrie       Stochastik

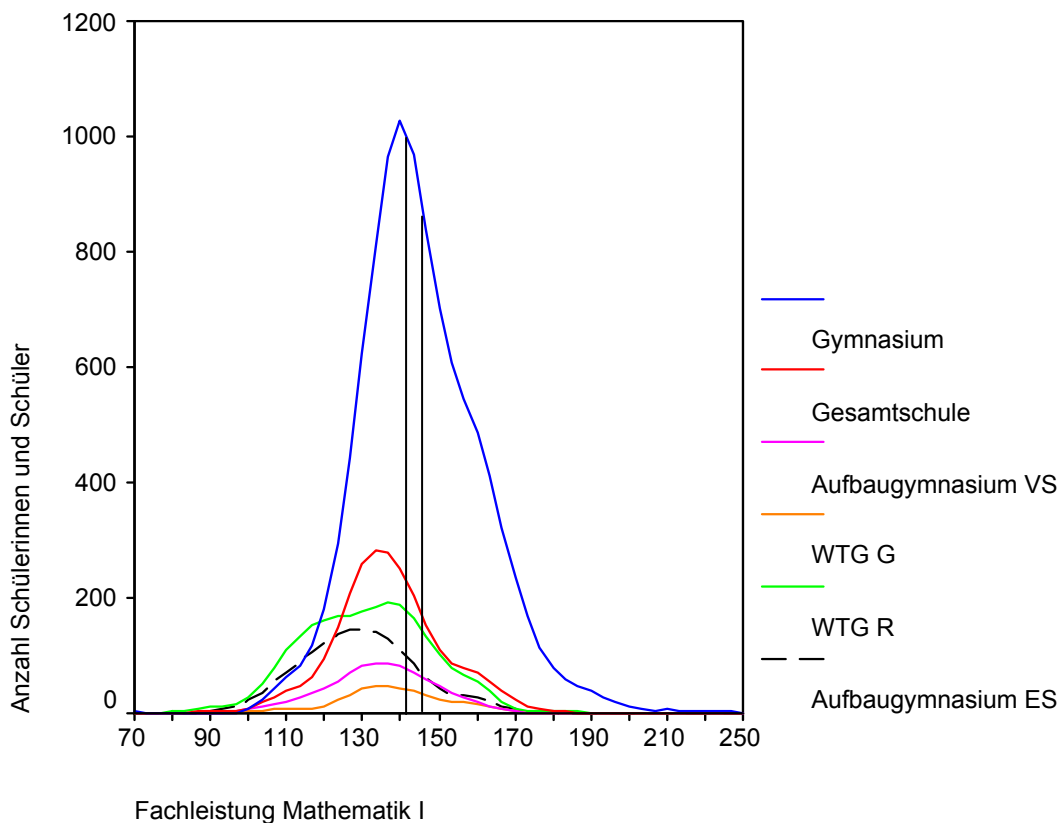
Leistungsmittelwert aller Schülerinnen und Schüler (einschließlich der Schülerinnen und Schüler aus der Einführungsstufe des Aufbaugymnasiums): 140,3.

Die Kurve zeigt die Aufteilung der Fachleistungen, wenn sie normalverteilt wären.

## Fachleistung Mathematik I, differenziert nach Bildungsgang

Abbildung 7 zeigt die Verteilung der mathematischen Fachleistungen in absoluten Häufigkeiten zu Beginn der Klassenstufe 11, getrennt nach Bildungsganggruppen. Bei der Interpretation dieser Grafik sind wiederum die zwischen den einzelnen Gruppen stark differierenden Schüleranteile (vgl. unten Tabelle 13) zu berücksichtigen. Diese Form der Darstellung wurde wiederum deshalb gewählt, um die Überschneidungen der einzelnen Gruppen und die typischen bildungsgangspezifischen Bereiche sichtbar zu machen. Bei der Interpretation dieser Grafik ist also zu beachten, dass der hohe Anteil der grundständigen Gymnasias-tinnen und Gymnasiasten (61,6 Prozent der insgesamt getesteten Schülerinnen und Schüler) und die erheblich geringere Anzahl von Testpersonen in den anderen Gruppen die Differenz zwischen dem Mittelwert der Gesamtgruppe (ohne die Schülerinnen und Schüler aus der Einführungsstufe des Aufbaugymnasiums; linke Senkrechte) und dem Mittelwert der grundständigen Gymnasiastinnen und Gymnasiasten (rechte Senkrechte) auf einen niedrigen Wert festlegen.

Abbildung 7 Fachleistung Mathematik I: Verteilung der Lernstände zu Beginn der Klassenstufe 11 nach Bildungsgang (absolut)



Aufgrund der positiven Auslese von Schülerinnen und Schülern, die an LAU 11 teilgenommen haben, und der neuen Form der äußeren Differenzierung sind die Überschneidungen zwischen den Bildungsgängen im Fach Mathematik im

Unterschied zu den Ergebnissen der LAU 9 erwartungsgemäß größer. Die Überschneidungen der sechs Gruppen deuten darauf hin, dass die Standards in allen Fällen ähnlich sind. Es gibt keinen nach den Vorgaben früherer LAU-Erhebungen sinnvoll definierbaren „typisch gymnasialen Leistungsbereich“ bzw. keinen typischen Leistungsbereich für die Gesamtschülerinnen und Gesamtschüler oder für die Schülerinnen und Schüler der Wirtschafts- und technischen Gymnasien. Gleichwohl unterscheiden sich die Leistungsverteilungen in den verschiedenen Bildungsganggruppen, wie Abbildung 7 veranschaulicht.

Für die grundständigen Gymnasiastinnen und Gymnasiasten entspricht der Leistungsdurchschnitt einem Lernstandswert von 145,4 Skalenpunkten. Dieser Wert wurde von 24,0 Prozent der grundständigen Gesamtschülerinnen und Gesamtschüler, 28,2 Prozent der Schülerinnen und Schüler aus der Gruppe WTG-G und 23,0 Prozent der Vorstufenschülerinnen und -schüler an den Aufbaugymnasien erreicht oder übertroffen, während es in der Gruppe WTG-R 19,9 Prozent sind. In der Gruppe der Schülerinnen und Schüler aus der Einführungsstufe des Aufbaugymnasiums sind es 14,2 Prozent.

In Abbildung 8 sind die Mittelwertunterschiede zwischen den einzelnen Bildungsganggruppen in einem Säulendiagramm veranschaulicht.

Abbildung 8 Fachleistung Mathematik I: Verteilung der Lernstände zu Beginn der Klassenstufe 11 nach Bildungsgang

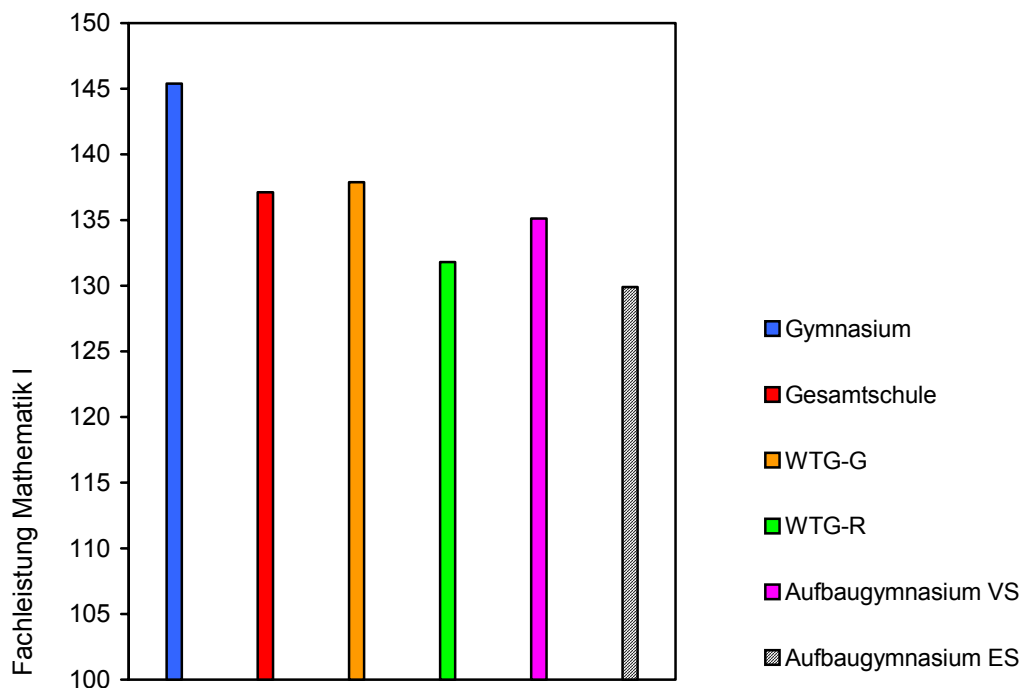


Tabelle 13 enthält die numerischen Kennwerte, die die Unterschiede zwischen den sechs Gruppen in den erreichten Mathematikleistungen zu Beginn der

Klassenstufe 11 bzw. zu Beginn der Einführungsstufe des Aufbaugymnasiums ausweisen. Wie diesen Werten zu entnehmen ist, bestanden zu Beginn der Klassenstufe 11 beträchtliche Unterschiede zwischen den Bildungsganggruppen: Die grundständigen Gymnasiastinnen und Gymnasiasten haben Durchschnittsleistungen erzielt, die trotz des mit 61,6 Prozent hohen Anteils<sup>9</sup> an der hier berücksichtigten Schülerschaft um 0,24 Standardabweichungen über dem allgemeinen Mittelwert liegen. Mit Effektstärken von -0,21 (WTG-G) bis -0,58 (WTG-R) liegen die Mathematikleistungen aller anderen Gruppen deutlich unter dem allgemeinen Durchschnitt.

Tabelle 13 Fachleistungen in Mathematik I zu Beginn der Klassenstufe 11: Mittelwerte, Standardabweichungen und Effektstärken, insgesamt und getrennt nach Bildungsgang

	Mittelwert	Standard- abweichung	Effektstärke <i>d</i>	<i>N</i>
<b>Gymnasium</b>	145,4	15,7	0,24	3.471 (61,6 %)
<b>Gesamtschule</b>	137,1	14,6	-0,26	861 (15,3 %)
<b>WTG-G</b>	137,9	15,1	-0,21	156 (2,8 %)
<b>WTG-R</b>	131,8	16,3	-0,58	841 (14,9 %)
<b>Aufbaugymnasium VS</b>	135,1	14,2	-0,38	305 (5,4 %)
<b><i>insgesamt</i></b>	<i>141,4</i>	<i>16,5</i>	–	<i>5.634</i> <i>(100 %)</i>
<b>Aufbaugymnasium ES</b>	129,9	15,6	-0,70	555

Wie bereits beim Textverständnis und dem passiven Rechtschreibwissen fällt auch bei Mathematik der mit der Effektstärke  $d = 0,12$  vergleichsweise geringe Mittelwertunterschied zwischen den Schülerinnen und Schülern aus der Gruppe WTG-R und den Schülerinnen und Schülern aus der Einführungsstufe des Aufbaugymnasiums auf. Die Lernausgangslagen der beiden Gruppen WTG-G (Schülerinnen und Schüler mit Versetzung in die gymnasiale Oberstufe) und WTG-R (Schülerinnen und Schüler mit einem Mittleren Schulabschluss) diffe-

<sup>9</sup> Je größer die Fallzahl einer Gruppe ist, desto repräsentativer ist auch ihr Mittelwert und umso aussagekräftiger ist der Unterschied zu den anderen Gruppen.

rieren mit einer Effektstärke von 0,37 Standardabweichungen sogar noch etwas stärker als die Lernausgangslagen der Schülerinnen und Schüler aus der Einführungsstufe und aus der Vorstufe des Aufbaugymnasiums ( $d = 0,32$ ) – mit dem Unterschied, dass diese nicht in den gleichen Klassen unterrichtet werden, sondern ein Schuljahr zwischen ihnen liegt.

Bei einem Vergleich der Standardabweichungen in den vorangegangenen LAU-Erhebungen war ersichtlich geworden, dass sich die Streuung der Mathematikleistungen im Hamburger Schulsystem vom Beginn der Klassenstufe 7 bis zum Beginn der Klassenstufe 9 insgesamt verringert hatte (LAU 7:  $s = 23,8$ ; LAU 9:  $s = 19,6$ ), und zwar – wie die Vergleiche für einzelne Schulformen zeigten – jeweils in analogen Verhältnisgrößen. Aufgrund der positiven Auslese von Schülerinnen und Schülern beim Übergang von der Klassenstufe 10 in die Klassenstufe 11 der gymnasialen Oberstufe war nun nochmals eine Verringerung der Streuung zu erwarten. Tatsächlich hat sich die Streuung auf  $s = 16,5$  vermindert. Dieser Wert ist jedoch größer als derjenige für Gymnasien am Ende der Klassenstufe 8 (vgl. LEHMANN, PEEK, GÄNSFUß & HUSFELDT 2002, S. 33). Interessanterweise hat auch die Streuung der Leistungen der grundständigen Gymnasiastinnen und Gymnasiasten zugenommen.

### 3.2.4 Fachleistung Mathematik II

Der zweite Untertest des *SL-HAM 10/11*, der für den Mathematikunterricht relevante Leistungen erfasst, ist der Mathematik-II-Test. Im Unterschied zu den Aufgaben des Mathematik-I-Tests wurde hier keine Differenzierung nach abgebender Schulform und Bildungsgang/Oberstufenform vorgenommen. Alle Schülerinnen und Schüler hatten eine einheitliche Testversion zu bearbeiten<sup>10</sup>.

Der Mathematik-II-Test umfasst zum einen Aufgaben zu diskontinuierlichen Texten, zum anderen Aufgaben aus TIMSS III. Sechs der Aufgaben zu den diskontinuierlichen Texten beziehen sich auf eine Grafik der Bundesanstalt für Arbeit, die über den Einfluss der Jahreszeiten auf die Arbeitslosenstatistik über einen Zeitraum von fünfzehn Monaten Aufschluss gibt. Weiterhin gehören zu den diskontinuierlichen Texten fünf Aufgaben zu Abfahrts- und Ankunftszeiten, die sich auf zwei Busfahrpläne beziehen. Aus dem Kontext von TIMSS III

---

<sup>10</sup> Hingegen wurde in der ULME zusätzlich eine hauptschulspezifische Testversion (HS-Version) eingesetzt. Beide Testversionen unterscheiden sich nur in der Anzahl der Aufgaben. Die HS-Version umfasst insgesamt 23, die RS/Gy-Version 28 Aufgaben. Aufgrund des höheren Leistungsstandes der ehemaligen Realschülerinnen und Realschüler sowie der Schülerschaft an Gymnasien enthält die Testversion RS/Gy fünf Aufgaben aus der TIMS-Studie mehr, die sich durch einen höheren Schwierigkeitsgrad von den anderen Aufgaben unterscheiden.

stammen siebzehn Aufgaben zur mathematischen Grundbildung, durchgehend im Multiple-Choice-Format.

Die Leistungen wurden als Rohpunktwerte für die sechs Bildungsganggruppen analysiert. Eine probabilistische Skalierung wurde bei diesem Instrument nicht vorgenommen, da die vorhandenen Daten weder eine Verknüpfung mit der LAU 9 gestatten noch eine Verankerung auf der TIMSS-Metrik, die auf dem Verfahren der „plausible values“ beruht und mit den gegebenen Daten nicht ohne weiteres reproduziert werden kann. Jedoch garantiert die Information über die Anzahl der gelösten Aufgaben eine ausreichende Messqualität.

Um einen Einblick in die Anforderungsstruktur des Mathematik-II-Tests zu geben, sind im Folgenden die Lösungshäufigkeiten von sieben ausgewählten Aufgaben aufgeführt. Die ersten beiden Aufgaben in Tabelle 14 sind dem Teil ‚Arbeitslosenstatistik‘ zugeordnet, die dritte, vierte und fünfte Aufgabe gehören zum TIMSS-Teil und die beiden letzten Aufgaben entstammen dem Testteil ‚Fahrplan‘. Die zweite Spalte der Tabelle zeigt, wie viel Prozent der insgesamt 6.184 Schülerinnen und Schüler (einschließlich der Schülerinnen und Schüler aus der Einführungsstufe des Aufbaugymnasiums), die den Mathematik-II-Test bearbeitet haben, die Beispielaufgaben lösen konnten.

Tabelle 14 Beispielaufgaben des Untertests Mathematik II mit Prozentanteil der Schülerinnen und Schüler, die die Aufgaben gelöst haben

<i>Beispielaufgaben</i>	<i>Prozent aller Schülerinnen und Schüler, die die Aufgabe gelöst haben</i>
<i>Arbeitslosenstatistik</i>	
Wie hoch war die Zahl der Arbeitslosen (Originalwert) im März 2002? ( <b>4 156 000</b> )	31,8 %
Wird der Originalwert in der Arbeitslosenstatistik vom Januar 2001 mit dem vom Januar 2002 verglichen, so zeigt sich:	81,4 %
<ul style="list-style-type: none"> <li>a) Die Arbeitslosenzahl hat abgenommen.</li> <li><b>b) Die Arbeitslosenzahl hat zugenommen.</b></li> <li>c) Die Arbeitslosenzahl ist gleich geblieben.</li> <li>d) Dazu wird keine Aussage gemacht.</li> </ul>	



**TIMSS III**

Ein 45.000 Liter fassender Wassertank wird mit einer Geschwindigkeit von 220 Litern pro Minute gefüllt. Schätzen Sie auf eine halbe Stunde genau, wie lange es dauert, den Tank zu füllen. a) 4 Stunden <b>b) 3 ½ Stunden</b> c) 3 Stunden d) 2 ½ Stunden	48,3 %
In einer Kiste sind 3.000 Glühbirnen. 100 Glühbirnen werden zufällig ausgewählt und getestet. Fünf Glühbirnen dieser Auswahl sind defekt. Wie viele defekte Glühbirnen sind voraussichtlich in der ganzen Kiste? a) 60 <b>b) 150</b> c) 300 d) 600	75,5 %
Ein Warnsystem besteht aus zwei unabhängigen Alarmanlagen, die bei einem Notfall mit den Wahrscheinlichkeiten 0,95 bzw. 0,90 ansprechen. Suchen Sie die Wahrscheinlichkeit, dass in einem Notfall mindestens eine der Alarmanlagen anspricht. a) <b>0,995</b> b) 0,975 c) 0,900 d) 0,855	23,8 %
<b>Fahrpläne</b>	
Frau Haverveld fährt montags und dienstags um 07.38 Uhr mit dem Bus 145 ab S Harburg zur Arbeit. Wann kommt sie am Rabenstein an? <b>(07.51)</b>	66,5 %
Wann fährt sonnabends der erste Bus nach 09.00 Uhr ab Knoopstraße / Bremer Straße? <b>(09.07)</b>	19,0 %

Bei insgesamt 28 Aufgaben und 28 erreichbaren Punkten ergab sich ein Leistungsdurchschnitt von 13,5 Punkten (ohne die Schülerinnen und Schüler aus der Einführungsstufe des Aufbaugymnasiums). Die Mittelwertunterschiede zwischen den verschiedenen Bildungsganggruppen entsprachen weitgehend den bereits für den Mathematik-I-Test berichteten.

Tabelle 15 Leistungen in Mathematik II zu Beginn der Klassenstufe 11: Mittelwerte, Standardabweichungen und Effektstärken, insgesamt und getrennt nach Bildungsgang

	Mittelwert	Standardabweichung	Effektstärke $d$	$N$
<b>Gymnasium</b>	14,5	4,3	0,23	3.469 (61,6 %)
<b>Gesamtschule</b>	12,6	4,1	-0,20	859 (15,2 %)
<b>WTG-G</b>	12,7	4,2	-0,18	156 (2,8 %)
<b>WTG-R</b>	11,3	4,0	-0,50	838 (14,9 %)
<b>Aufbaugymnasium VS</b>	11,7	4,0	-0,41	309 (5,5 %)
<b>insgesamt</b>	13,5	4,4	–	5.631 (100 %)
<b>Aufbaugymnasium ES</b>	10,5	3,5	-0,68	553

Die grundständigen Gymnasiastinnen und Gymnasiasten liegen mit einer Effektstärke von 0,23 über dem allgemeinen Durchschnitt (Mathematik I:  $d = 0,24$ ), während die Schülerinnen und Schüler aus der Gruppe WTG-G mit einer Effektstärke von -0,18 (Mathematik I: -0,21) und die grundständigen Gesamtschülerinnen und Gesamtschüler mit einer Effektstärke von -0,20 (Mathematik I: -0,26) unter dem allgemeinen Durchschnitt liegen. Es folgen die Schülerinnen und Schüler aus der Vorstufe des Aufbaugymnasiums (-0,41; Mathematik I: -0,38), die Schülerinnen und Schüler aus der Gruppe WTG-R liegen eine halbe Standardabweichung unter dem allgemeinen Durchschnittswert (Mathematik I: -0,58) und drei Viertel einer Standardabweichung unter dem Mittelwert der grundständigen Gymnasiastinnen und Gymnasiasten. Ihr Leistungsrückstand gegenüber den Schülerinnen und Schülern aus der Gruppe WTG-G hat eine Effektstärke von -0,32 und ist noch ausgeprägter als der Leistungsunterschied

zwischen den Schülerinnen und Schülern aus der Einführungsstufe und aus der Vorstufe des Aufbaugymnasiums ( $d = -0,27$ ).

Vergleicht man wieder die einzelnen Bildungsganggruppen mit dem Leistungsstand der grundständigen Gymnasiastinnen und Gymnasiasten, so haben 31,3 Prozent der grundständigen Gesamtschülerinnen und Gesamtschüler und 34,0 Prozent der Schülerinnen und Schüler aus der Gruppe WTG-G den Leistungsdurchschnitt (14,5 Punkte, rechte Senkrechte in Abbildung 9) erreicht oder übertroffen, während es unter den Vorstufenschülerinnen und -schülern an den Aufbaugymnasien 23,3 Prozent und in der Gruppe WTG-R 22,4 Prozent sind. In der Gruppe der Schülerinnen und Schüler aus der Einführungsstufe des Aufbaugymnasiums sind es 13,2 Prozent.

Abbildung 9 Mathematik II: Verteilung der Lernstände zu Beginn der Klassenstufe 11 nach Bildungsgang (absolut)

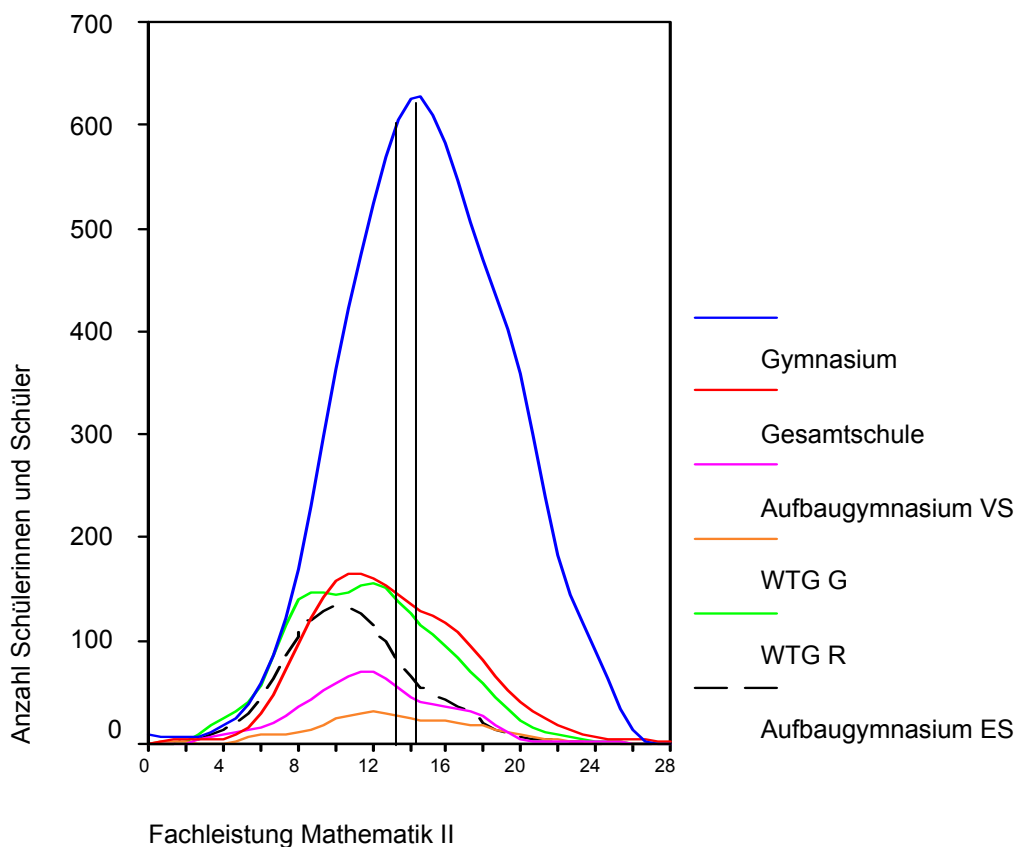
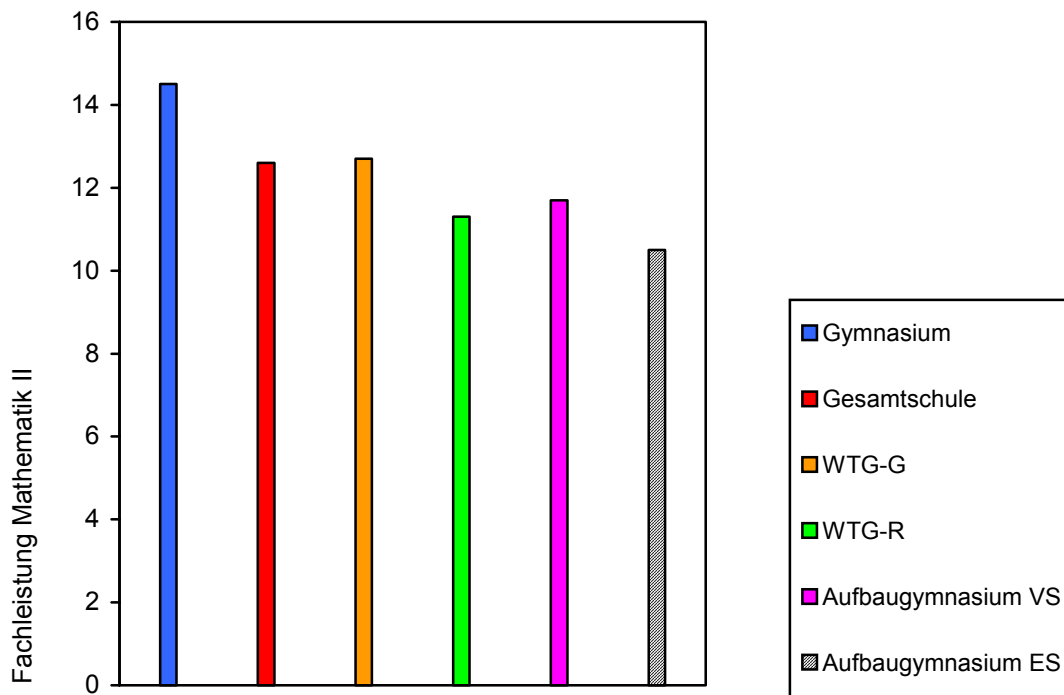


Abbildung 9 zeigt, dass die Leistungen der grundständigen Gymnasiastinnen und Gymnasiasten in guter Näherung normalverteilt sind. Hingegen sind bei den Schülerinnen und Schülern aller anderen fünf Gruppen tendenziell linkssteile Verteilungen zu erkennen. Offenbar waren die Aufgaben dieses an TIMSS angelehnten, aber auch Verständnisaufgaben aus dem Bereich grafisch-nummerischer Darstellungen umfassenden Tests insgesamt für die Schülerschaft außer-

halb des Regelgymnasiums relativ schwer. Die linke Senkrechte kennzeichnet wiederum den allgemeinen Mittelwert aller Vorstufenschülerinnen und -schüler.

In Abbildung 10 sind die Mittelwertunterschiede zwischen den einzelnen Bildungsganggruppen wiederum in einem Säulendiagramm veranschaulicht.

Abbildung 10 Fachleistung Mathematik II: Verteilung der Lernstände zu Beginn der Klassenstufe 11 nach Bildungsgang



### 3.2.5 Fachleistung Englisch: C-Test

Der eine Unterrichtsstunde umfassende *SL-HAM 10/11-Englishtest* erhebt zum einen im Format eines *C-Tests* Wort-, Rechtschreib-, Syntax- und Grammatikkenntnisse der Schülerinnen und Schüler und zum anderen im Format eines Leseverständnistests Verständnisleistungen aus dem Textzusammenhang.

Die Feststellung der globalen Sprachkompetenz im Bereich Englisch als Fremdsprache sowie der Wort- und Grammatikkenntnisse erfolgte im modifizierten Cloze-Format mit insgesamt sechs Lückentexten, wobei jede Testversion zwei Kerntexte und zwei testversionspezifische Texte enthielt. Abgesehen vom ersten und letzten Satz war jedes vierte Wort nur zur Hälfte ausgeschrieben und musste von den Schülerinnen und Schülern vervollständigt werden (bei ungerader Buchstabenzahl  $i$  waren  $(i + 1) / 2$  Buchstaben zu ergänzen). Insgesamt enthielt der Englisch-C-Test 174 Wortergänzungen, von denen 170 bei der Auswertung den teststatistischen Anforderungen der Rasch-Skalierung genügten. In der Abbildung 11 (siehe unten, S. 55) sind auf der rechten Seite der Rasch-Skala die 170 Aufgaben – geordnet nach ihrem Schwierigkeitsgrad – und auf der linken Seite die darauf bezogene Verteilung der Schülerleistungen eingetragen.

Die Testaufgaben verteilen sich bei einer mittleren Schwierigkeit um den Wert 147 über den Bereich von 96 bis 230 Skalenpunkten, wobei allerdings nur zwei Aufgaben einen Schwierigkeitsparameter über 200 haben; der Mittelwert für die Schülerleistungen (einschließlich der Schülerinnen und Schüler aus der Einführungsstufe des Aufbaugymnasiums) liegt mit 151,4 Skalenpunkten ( $s = 16,9$ ) etwas über der mittleren Aufgabenschwierigkeit. Wie man der Abbildung 11 entnehmen kann, sind – bei angenäherter Normalverteilung der Schülerleistungen – die Schwierigkeiten der Testaufgaben und die Fähigkeiten bzw. Leistungen der getesteten Schülerinnen und Schüler insgesamt recht gut aufeinander abgestimmt.

Die 170 Wortergänzungsaufgaben wurden ihrer Schwierigkeit nach in fünf Anforderungsniveaus mit jeweils 34 Items untergliedert. Pro Anforderungsniveau steigt der Schwierigkeitsgrad der jeweils zu ergänzenden Wörter. Anforderungsniveau 1 beinhaltet vor allem Wortergänzungen, die hauptsächlich den unteren Bereich des Grundwortschatzes repräsentieren: einfache Substantive, Pronomina und Präpositionen, die zum Teil mehrfach vertreten sind, einfache Verbformen wie *be* und *will* sowie Artikel und Bindewörter wie *the* und *and*. In der zweiten Itemgruppe sind größtenteils Wörter des Grundwortschatzes vertreten, vor allem Substantive und Präpositionen. Auch auf dem Anforderungsniveau 3 sind noch Vokabeln des Grundwortschatzes gefragt, relativ häufig in kurzen (hinweisarmen) oder mehrdeutigen Formen. Wörter wie *performance*,

*authorities* und orthografisch anspruchsvolle Verben wie *continue* charakterisieren das Anforderungsniveau 4. Wortergänzungen, die dem Niveau 5 zugeordnet wurden, stellen die höchsten Anforderungen an Wortschatz, Rechtschreibung und Grammatik. Tabelle 16 gibt einen Überblick über typische Wortergänzungen der verschiedenen Anforderungsniveaus.

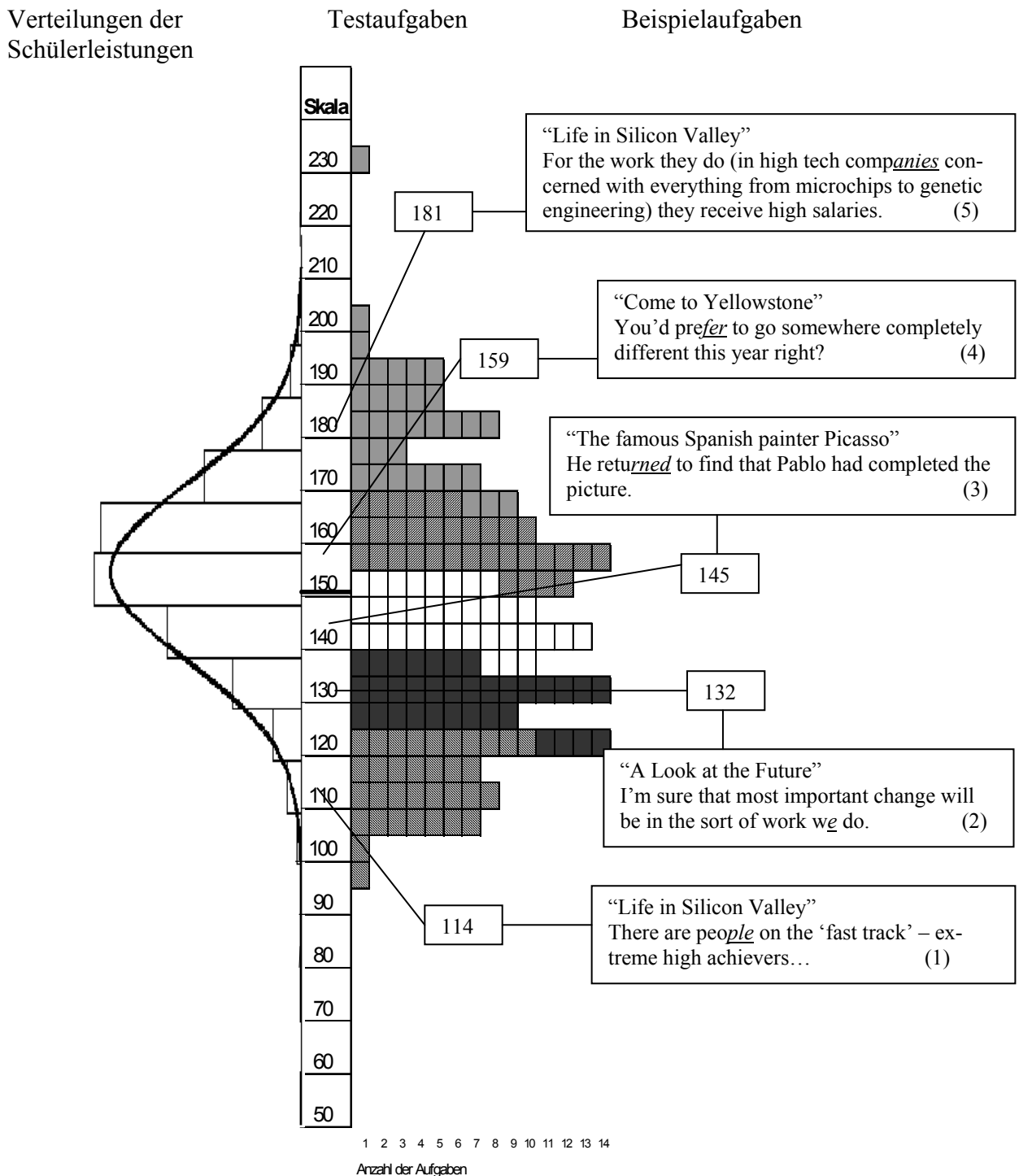
Tabelle 16 Beispiele für Wortergänzungen der unterschiedlichen Anforderungsniveaus im Englisch-C-Test

Wortart	Anforderungsniveau 1	Anforderungsniveau 2	Anforderungsniveau 3	Anforderungsniveau 4	Anforderungsniveau 5
<i>Substantive</i>	4 * people year picture	minutes mountains ferry Europe parents	Change Example Artist Work City	computers performance authorities visitors canyons	crowds engineering companies standard nightmare
<i>Verben</i>	be will do come	is	Returned Went do breaking (will) mean	needn't kept continue leave prefer	growing enjoy sort agree fire
<i>Präpositionen</i>	over of in	to between from about	at on from when	when as if by	as
<i>Adverbien / Adjektive</i>	most almost	much little	lot often every fastest	soft only awful unless	completely extremely lonely exhausting
<i>Artikel, Bindewörter</i>	5 * the 3 * and	the and	the/their this/that		
<i>Pronomen</i>	he his	him we		who	

Um einen Eindruck davon zu vermitteln, welche Leistungsniveaus mit bestimmten Skalenwerten verknüpft sind, sind in Abbildung 11 fünf Beispielaufgaben eingetragen. Beispiel 1 – die korrekte Ergänzung des Wortes „*people*“ im Satz- bzw. Textkontext – steht für ein Leistungsniveau (Skalenwert 114), das mit 50-prozentiger Wahrscheinlichkeit von 97,5 Prozent aller getesteten Schülerinnen und Schüler erreicht bzw. übertroffen wurde. Die weiteren Beispiele – auch hier sind die geforderten Wortergänzungen kursiv geschrieben – stehen für zunehmend anspruchsvolle Leistungsniveaus. Der Skalenwert 132 für die Beispielaufgabe 2 wurde von 87,7 Prozent aller Testteilnehmerinnen und -teilnehmer erreicht bzw. übertroffen; 68,8 Prozent aller Schülerinnen und Schüler

erreichten bzw. übertrafen den Skalenwert 145 (Beispiel 3) und 32,0 Prozent den Skalenwert 159 (Beispiel 4). 2,7 Prozent aller Getesteten haben einen Skalenwert von 181 und höher (höchstes Anforderungsniveau; Beispiel 5) erreicht.

Abbildung 11 Verteilung der Schülerleistungen in *Englisch* im Vergleich mit den Schwierigkeiten der Testaufgaben im *SL-HAM 10/11-Englisch: C-Test*



Gesamtstichprobe: N = 6.075; Leistungsmittelwert aller Schülerinnen und Schüler: 151,4.  
Die Kurve zeigt die Aufteilung der Fachleistungen, wenn sie normalverteilt wären.

Die 14 Aufgaben, die dem Skalenbereich von 155 bis 160 (Anforderungsniveau 3) zugeordnet sind, können als typische Aufgaben für die Gruppe der grundständigen Gymnasiastinnen und Gymnasiasten angesehen werden, deren Mittelwert bei 158,1 liegt. Die 10 Aufgaben, die dem Skalenbereich von 146 bis 150 (Anforderungsniveau 2) zuzuordnen sind, entsprechen den Mittelwerten der Gruppen Gesamtschule und WTG-G. 13 Aufgaben sind dem Skalenbereich von 141 bis 144 (ebenfalls Anforderungsniveau 2) zuzuordnen, die den mittleren Leistungen der Schülerinnen und Schüler aus den Gruppen Aufbaugymnasium VS und ES entsprechen. Weitere 10 Aufgaben, die dem Skalenbereich 136 bis 140 (Anforderungsniveau 1) zuzuordnen sind, entsprechen den mittleren Leistungen der Schülerinnen und Schüler aus der Gruppe WTG-R.

Vergleicht man nun die verschiedenen Aufgabengruppen miteinander, so lassen sich qualitativ folgende Unterschiede feststellen: Die „typisch gymnasialen“ Aufgaben der Testversion I sind in der Regel Wortergänzungen, die höhere Anforderungen an die orthografischen und grammatischen Kompetenzen stellen. So sind häufiger Verben zu ergänzen als Substantive. Zusätzlich erfordern die Verbergänzungen eine höhere Flexibilität beim Zugriff auf das englische Sprachregister sowie einen sicheren Umgang mit Kasus, Tempus und Numerus. Folgende Beispiele verdeutlichen das: *continue, get, watched, prefer* oder *you'll / you'd*. Hingegen sind für den Bereich der anderen Gruppen einfache Verben wie *do* oder häufig gebrauchte Nomina wie *work, city, men* oder *holiday* typisch.

### **Fachleistung Englisch: C-Test, differenziert nach Bildungsganggruppen**

Die folgende Tabelle 17 zeigt die gruppenspezifischen Ergebnisse für den *Englisch-C-Test*. Die grundständigen Gymnasiastinnen und Gymnasiasten liegen auch in diesem *SL-HAM 10/11*-Untertest leistungsmäßig vorn. Mit einem Durchschnittswert von 158,1 Skalenpunkten und einer vergleichsweise geringen Streuung (14,5) stellen sie mit 5,7 Skalenpunkten über dem allgemeinen Mittelwert bei einer Effektstärke von 0,34 die leistungsstärkste Gruppe dar. Die Mittelwerte aller anderen Bildungsganggruppen liegen deutlich unterhalb des allgemeinen Gesamtmittelwerts, wobei insbesondere die mit Version II getestete Gruppe WTG-R mit einer sehr hohen Effektstärke ( $d = -0,81$ ) unter dem allgemeinen Leistungsdurchschnitt und sogar unterhalb der mittleren Leistung der Schülerinnen und Schüler aus der Einführungsstufe des Aufbaugymnasiums liegt. Entsprechend ist der Leistungsunterschied zwischen den beiden Gruppen WTG-G und WTG-R mit 7,4 Skalenpunkten und einer Effektstärke von 0,44 noch stärker ausgeprägt als zwischen den Schülerinnen und Schülern aus der Vorstufe und der Einführungsstufe des Aufbaugymnasiums (3,6 Skalenpunkte;  $d = 0,22$ ).



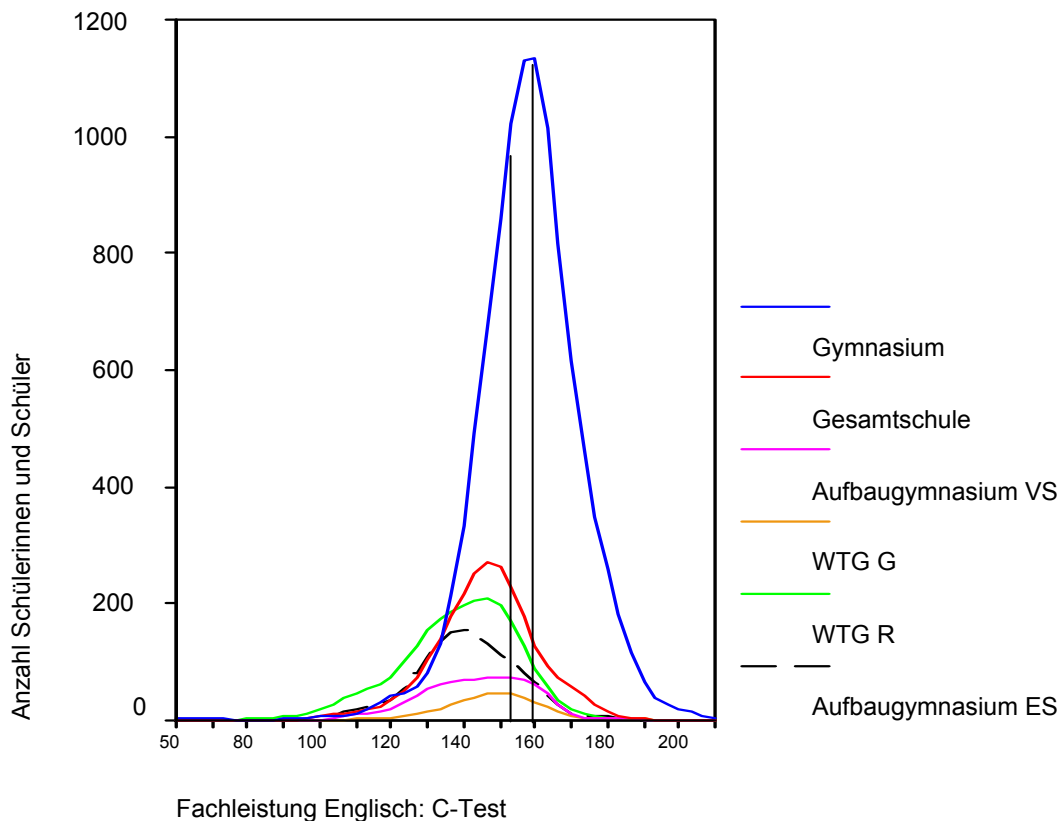
Tabelle 17 Fachleistung Englisch: C-Test zu Beginn der Klassenstufe 11: Mittelwerte, Standardabweichungen und Effektstärken, insgesamt und getrennt nach Bildungsgang

	Mittelwert	Standard- abweichung	Effektstärke <i>d</i>	<i>N</i>
<b>Gymnasium</b>	158,1	14,5	0,34	3.441 (61,9 %)
<b>Gesamtschule</b>	146,0	14,5	-0,38	848 (15,3 %)
<b>WTG-G</b>	146,2	19,6	-0,37	150 (2,7 %)
<b>WTG-R</b>	138,8	16,3	-0,81	817 (14,7 %)
<b>Aufbaugymnasium VS</b>	144,6	15,0	-0,46	302 (5,4 %)
<b><i>insgesamt</i></b>	<i>152,4</i>	<i>16,8</i>	–	<i>5.558</i> <i>(100 %)</i>
<b>Aufbaugymnasium ES</b>	141,0	13,8	-0,68	517

Die Schülerinnen und Schüler an den Wirtschafts- und technischen Gymnasien zeigen die größte Streuung (WTG-G:  $s = 19,6$ , WTG-R:  $s = 16,3$ ), die Schülerinnen und Schüler aus der Einführungsstufe des Aufbaugymnasiums mit  $s = 13,8$  die geringste. Die Abbildung 12 (siehe unten, S. 58) veranschaulicht diese Befunde grafisch.

Die grundständigen Gymnasiastinnen und Gymnasiasten stellen – wie stets – die mit Abstand größte Schülergruppe. Da die Anforderungen, die in der gymnasialen Oberstufe gestellt werden, primär auf diese Gruppe bezogen sind, erscheint es auch an dieser Stelle sinnvoll, sich vergleichend auf die Leistungsverteilung dieser Bildungsganggruppe (rechte Senkrechte) zu beziehen. Auf diese Weise wird erkennbar, wie viele der getesteten Schülerinnen und Schüler aus der jeweiligen Bildungsganggruppe unterhalb bzw. oberhalb des mittleren Leistungsbereichs der grundständigen Gymnasiastinnen und Gymnasiasten liegen. Die linke Senkrechte markiert wieder den Mittelwert aller getesteten Schülerinnen und Schüler (ohne die Schülerinnen und Schüler aus der Einführungsstufe des Aufbaugymnasiums).

Abbildung 12 Englisch-C-Test: Verteilung der Lernstände zu Beginn der Klassenstufe 11 nach Bildungsgang (absolut)

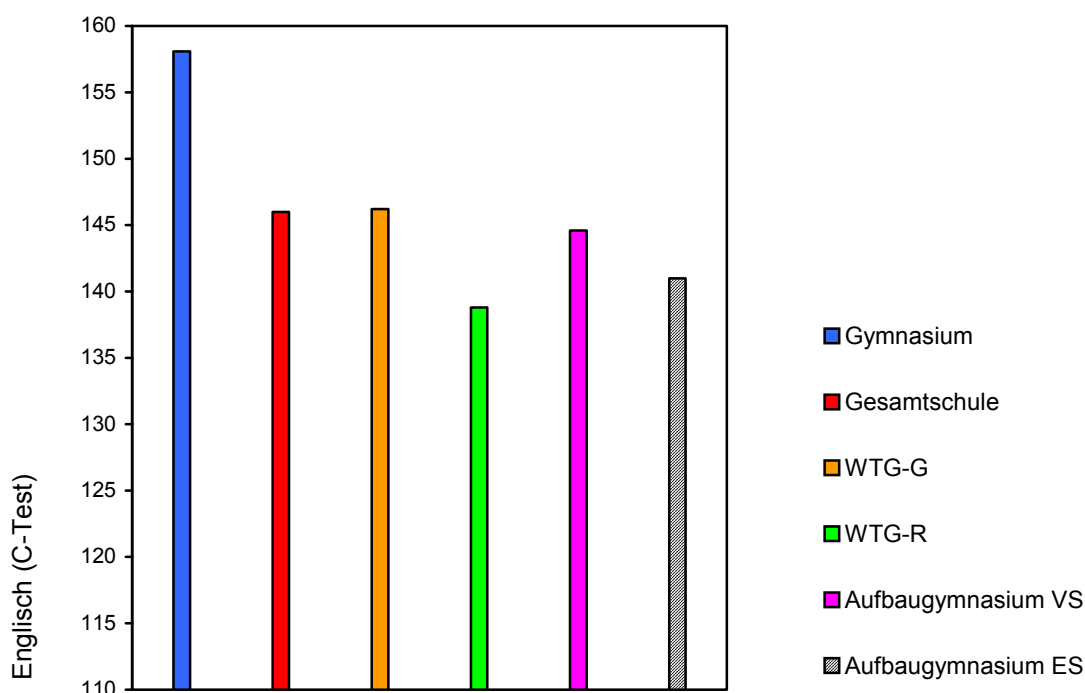


Sichtbar wird, dass nur ein kleiner Teil der anderen Bildungsganggruppen den Mittelwert der grundständigen Gymnasiastinnen und Gymnasiasten erreicht bzw. übertrifft.

Zur Orientierung darüber, wie viel Prozent aller Schülerinnen und Schüler den Kennwert der Gruppe Gymnasium von 158,1 Skalenpunkten überschreiten, sollen hier wieder die entsprechenden Quoten berichtet werden: Es handelt sich um 18,5 Prozent aus der Gruppe Gesamtschule, 24,7 Prozent aus der Gruppe WTG-G und 16,6 Prozent aus der Gruppe Aufbaugymnasium VS; aus der Gruppe Aufbaugymnasium ES sind es 9,7 Prozent und 8,6 Prozent der Schülerinnen und Schüler aus der Gruppe WTG-R.

In Abbildung 13 werden die Mittelwertunterschiede zwischen den Bildungsganggruppen noch einmal mit Hilfe eines Säulendiagramms veranschaulicht. Die erreichten individuellen Lernstände befinden sich im Bereich von 49,5 bis 206,5 Skalenpunkten. Der besseren Übersicht halber wird in der Grafik nur der Bereich von 130 bis 160 Skalenpunkten wiedergegeben.

Abbildung 13 Fachleistung Englisch-C-Test: Verteilung der Lernstände zu Beginn der Klassenstufe 11 nach Bildungsgang



### 3.2.6 Fachleistung Englisch: Leseverständnis

Den Schülerinnen und Schülern wurden im Rahmen des *Englisch-Leseverständnistests* zwei Texte in englischer Sprache vorgelegt. Alle zu testenden Jugendlichen erhielten je einen Kerntext und einen schulformspezifischen Text. Ihre Aufgabe bestand darin, acht (Testversion I) bzw. zehn (Testversion II) in deutscher und neun (Testversion I) bzw. sieben (Testversion II) in englischer Sprache vorgegebene Verständnisfragen im Multiple-Choice-Format zu beantworten<sup>11</sup>. Da der Englisch-Leseverständnistest in der LAU 9 noch nicht eingesetzt wurde, erübrigte sich eine Verzahnung mit den Längsschnittdaten.

Da in der LAU 11 nur diejenigen Schülerinnen und Schüler getestet wurden, die in die gymnasiale Oberstufe bzw. in die Einführungsstufe des Aufbaugymnasiums eingetreten sind, der Test aber in drei Versionen vorlag<sup>12</sup>, wurden zunächst der LAU- und der ULME-Datensatz gemeinsam skaliert. Anschließend wurde speziell für die LAU 11 der Mittelwert der Englisch-Leseverständnisleistungen auf 100 und die Standardabweichung auf 25 festgelegt. Schülerkom-

<sup>11</sup> Ursprünglich hatte die Testversion I insgesamt achtzehn Aufgaben. Aus teststatistischen Gründen konnte jedoch eine Aufgabe in der Analyse nicht berücksichtigt werden.

<sup>12</sup> In der ULME wurden die Hauptschul-, Realschul- und Gymnasialversion eingesetzt, in der LAU 11 nur die Realschul- und die Gymnasialversion

petenzen und Aufgabeneigenschaften des Englisch-Leseverständnistests sind somit auch hier auf einer Skala miteinander verschränkt.

Abbildung 14 zeigt die Verteilung der Aufgabenschwierigkeiten, die den gemessenen Leseverständnisseleistungen der Testteilnehmerinnen und Testteilnehmer gegenübergestellt wurden. Die Grafik zeigt eine leicht rechtssteile Verteilung der Schülerleistungen; ein nennenswerter Anteil aller getesteten Jugendlichen (55,7 %) erbringt Leistungen, die über dem allgemeinen Mittelwert von 100 Skalenpunkten liegen. Einem Leistungsmittelwert von 100 und einer Variationsweite von -9 bis 163 Skalenpunkten<sup>13</sup> steht ein Mittelwert von 75,6 auf der Seite der Aufgabenschwierigkeiten gegenüber. Offenbar sind die Schülerinnen und Schüler der gymnasialen Oberstufe in ihrer großen Mehrheit mit den Anforderungen dieses Untertests besonders gut zurechtgekommen.

Ähnlich wie der Deutsch-Leseverständnistest deckt auch der Englisch-Leseverständnistest den gesamten Anforderungsbereich von einfachen Dekodierungsaufgaben bis hin zu schlussfolgernden interpretatorischen Leistungen ab. So konnten die Aufgaben der drei in der LAU 11 eingesetzten Sachtexte ihrem Anforderungsniveau nach in 3 Niveaugruppen unterteilt werden. Anforderungsniveau 1 (Skalenbereich 38 bis 76) beinhaltet größtenteils Aufgaben, deren Lösung einfache Übersetzungen aus dem englischen Grundwortschatz ins Deutsche erfordert. Aufgaben, die das Auffinden von in der Fragestellung fast wörtlich angegebenen Sachverhalten im englischen Text oder die Kenntnis elementarer englischer Grammatikstrukturen erfordern – wie die Unterscheidung zwischen aktiver und passiver Verbverbindung – werden ebenfalls dieser Niveaustufe zugeordnet. Das zweite Anforderungsniveau, das dem Skalenbereich 59 bis 94 zuzuordnen ist, repräsentiert vor allem Aufgaben, deren Lösung die Verknüpfung mehrerer Textpassagen und deren Verständnis verlangt. Zusätzlich werden Vokabeln vorausgesetzt, die eher selten verwendet werden (z. B. *employee*). Auf diesem Anforderungsniveau wird teilweise das Auffinden der sinnvollsten von mehreren möglichen Antworten verlangt. Für die Lösung von Aufgaben auf dem Anforderungsniveau 3 (Skalenbereich 80 bis 131) war die Übersetzung anspruchsvoller Vokabeln im Kontext (z. B. die Übertragung von „*silence*“ in „*Sprachlosigkeit*“), das Verstehen des gesamten Textes sowie einmalig die Verfügbarkeit von nicht textimplizitem Hintergrundwissen notwendig.

In der Abbildung 14 ist für jedes Anforderungsniveau jeweils eine Beispielaufgabe eingetragen, entsprechend verknüpfte Skalenwerte können abgelesen werden.

---

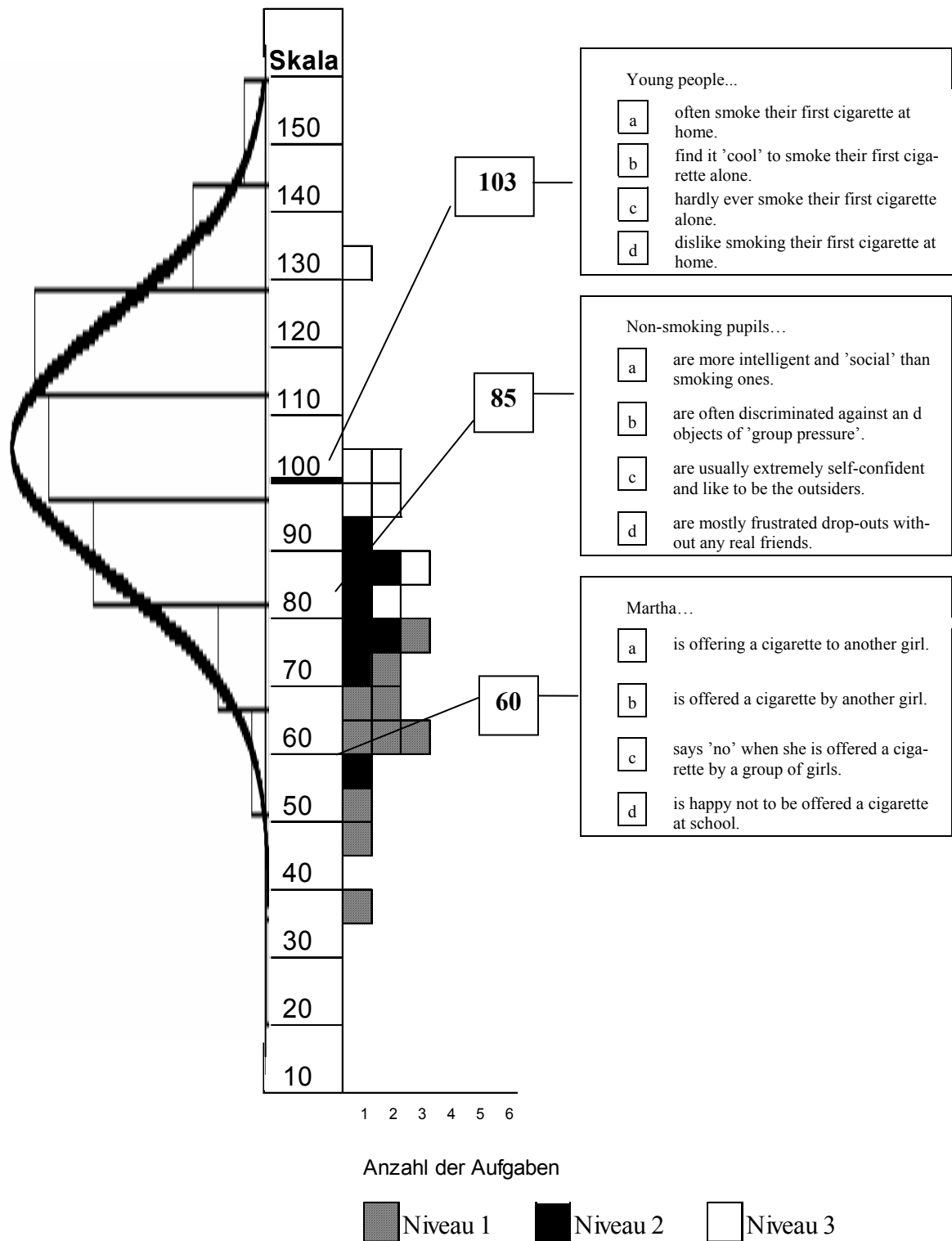
<sup>13</sup> Wie oben erwähnt, wurde der Leistungsmittelwert aus pragmatischen Gründen auf 100 und die Standardabweichung auf 25 festgesetzt. Die daraus resultierende Transformation erzeugt einige wenige negative Skalenwerte.

Abbildung 14 Verteilung der Schülerleistungen in Englisch im Vergleich mit den Schwierigkeiten der Testaufgaben im *SL-HAM 10/11-Englisch: Leseverständnis*

Verteilung der Schülerleistungen

Testaufgaben

Aufgaben aus dem Kerntext



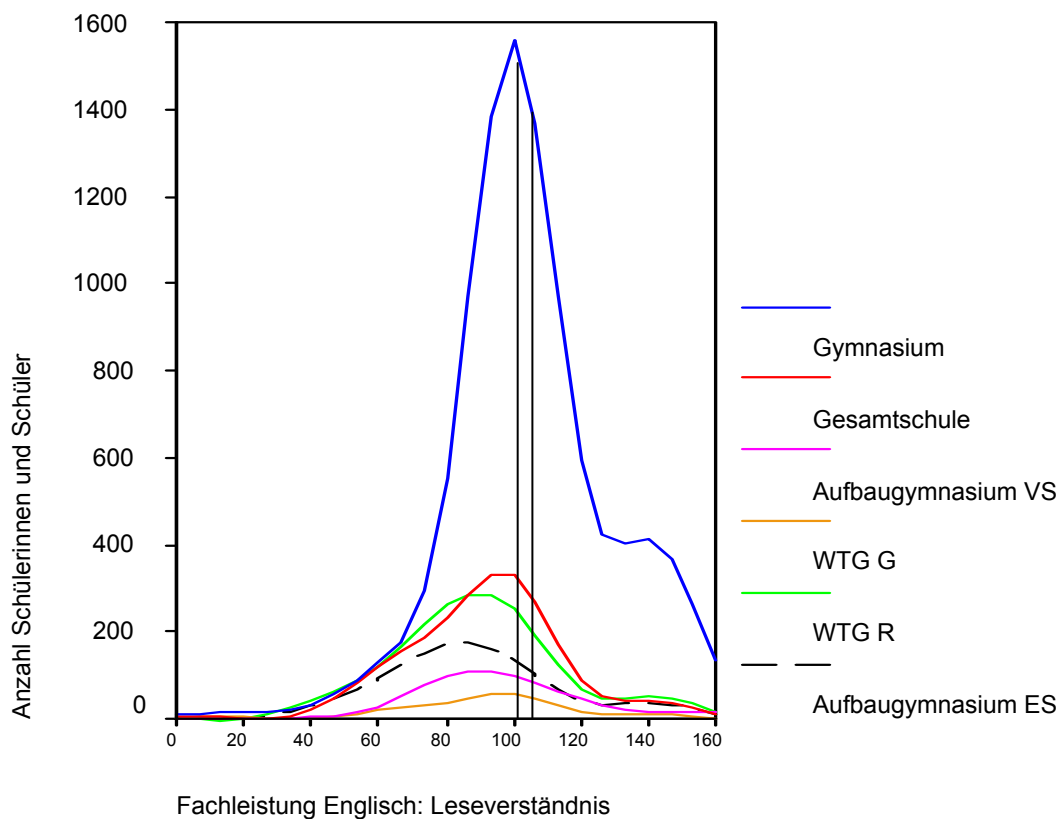
Gesamtstichprobe (N = 6.072); Leistungsmittelwert aller Schülerinnen und Schüler: 100,0. Die Kurve zeigt die Aufteilung der Fachleistungen, wenn sie normalverteilt wären.

Das erste Beispiel „*Martha ...*“ entspricht bei einem Skalenwert von 60 einem Leistungsniveau, das mit einer Wahrscheinlichkeit von 50 Prozent von 94,6 Prozent aller getesteten Jugendlichen erreicht oder übertroffen wurde. Das zweite Beispiel „*Non-smoking pupils ...*“ (Skalenwert 85) steht für das nächsthöhere Leistungsniveau 2; mit einer Wahrscheinlichkeit von 50 Prozent haben 75,3 Prozent aller getesteten Jugendlichen dieses Niveau erreicht oder übertroffen. Das dritte Beispiel „*Young people ...*“, das einem Skalenwert von 103 zuzuordnen ist, haben mit einer Wahrscheinlichkeit von 50 Prozent immerhin noch 41,6 Prozent aller Schülerinnen und Schüler der gymnasialen Oberstufe in Hamburg erreicht oder übertroffen.

### Fachleistung Englisch: Leseverständnis, differenziert nach Bildungsganggruppen

Abbildung 15 zeigt die Verteilungskurven für die 6.072 Schülerinnen und Schüler, die an diesem Untertest teilgenommen haben; die linke Senkrechte markiert den Mittelwert aller Vorstufenschülerinnen und -schüler, die rechte Senkrechte den Mittelwert der grundständigen Gymnasiastinnen und Gymnasiasten.

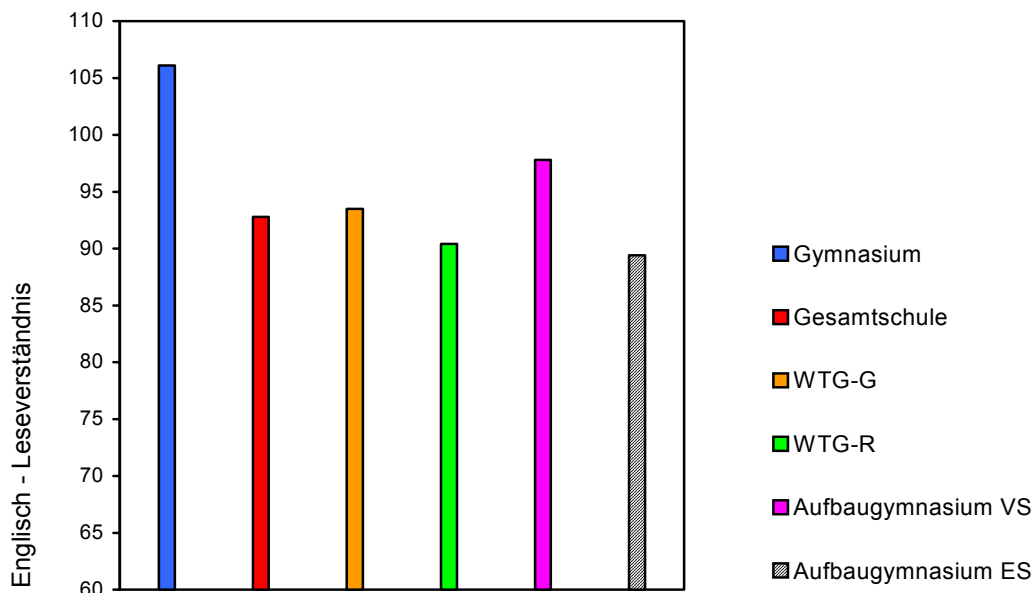
Abbildung 15 Fachleistung Englisch-Leseverständnis: Verteilung der Lernstände zu Beginn der Klassenstufe 11 nach Bildungsgang (absolut)



Auch hier sind wieder die Anteile der verschiedenen Bildungsganggruppen an der Gesamtstichprobe zu beachten: 61,9 Prozent befinden sich in der Gruppe Gymnasium und nur 2,7 Prozent in der Gruppe WTG-G (vgl. dazu Tabelle 18, S. 64). Auch in diesem Test liegen die Leistungen der grundständigen Gymnasiastinnen und Gymnasiasten größtenteils – im Unterschied zu allen anderen Bildungsganggruppen – über dem allgemeinen Mittel (linke Senkrechte in Abbildung 14). Wie bereits im Deutsch-Leseverständnistest gibt es in diesem Bildungsgang eine größere Gruppe von Schülerinnen und Schülern, die sich mit (fast) perfekten Lösungen von den übrigen absetzt (der zweite kleine Nebengipfel in der Verteilung). Die Leistungsverteilung der grundständigen Gesamtschülerinnen und Gesamtschüler ist leicht rechtssteil. Der Hauptteil der in dieser Gruppe getesteten Jugendlichen befindet sich demnach im rechten Bereich der Kurve, also über dem bildungsgangspezifischen Mittelwert von 92,8. Die Leistungen der Testteilnehmerinnen und Testteilnehmer aller anderen Bildungsgänge sind annähernd normalverteilt. Auch außerhalb der grundständigen Gymnasialschülerschaft gibt es indessen erkennbare Gruppen von Schülerinnen und Schülern mit nahezu oder vollständig fehlerfrei gelösten Testaufgaben.

Abbildung 16 veranschaulicht die Mittelwertunterschiede in der Fachleistung Englisch-Leseverständnis in einem Säulendiagramm.

Abbildung 16 Fachleistung Englisch-Leseverständnis: Verteilung der Lernstände zu Beginn der Klassenstufe 11 nach Bildungsgang



Im Gegensatz zum Englisch-C-Test, bei dem nur ein geringer Teil der anderen Gruppen den Mittelwert der grundständigen Gymnasiastinnen und Gymnasiasten übertroffen hatte, ist im Englisch-Leseverständnistest dieser Anteil höher:

Von den grundständigen Gesamtschülerinnen und Gesamtschüler haben 28,8 Prozent den Mittelwert der grundständigen Gymnasiasten von 106,1 (rechte Senkrechte in Abbildung 15) erreicht oder übertroffen, in der Gruppe WTG-G waren es 31,3 Prozent und in der Gruppe Aufbaugymnasium VS 38,1 Prozent, während es 25,6 Prozent aus der Gruppe WTG-R und 25,0 Prozent der Schülerinnen und Schüler aus der Einführungsstufe des Aufbaugymnasiums waren.

Bei der genaueren Betrachtung der statistischen Kennwerte der einzelnen Bildungsgänge werden erneut die erheblichen Leistungsunterschiede deutlich. Mit einer Effektstärke von  $d = 0,21$  unterscheiden sich die Jugendlichen der Gruppe Gymnasium mit 5,1 Skalenpunkten vom Gesamtmittelwert aller getesteten Vorstufenschülerinnen und -schüler. Die leistungsschwächsten Bildungsganggruppen stellen die mit Version II getesteten Schülerinnen und Schüler aus der Gruppe WTG-R und erwartungsgemäß die Schülerinnen und Schüler aus der Einführungsstufe des Aufbaugymnasiums. Mit 10,6 bzw. 11,6 Skalenpunkten liegen ihre Leistungen im Englisch-Leseverständnis um  $d = -0,43$  bzw.  $d = -0,47$  unter dem allgemeinen Mittelwert. Bei einem direkten Vergleich mit den grundständigen Gymnasiastinnen und Gymnasiasten ergibt sich eine Leistungsdifferenz von mehr als zwei Drittel einer Standardabweichung.

Tabelle 18 Fachleistung Englisch-Leseverständnis zu Beginn der Klassenstufe 11: Mittelwerte, Standardabweichungen und Effektstärken, insgesamt und getrennt nach Bildungsgang

	Mittelwert	Standardabweichung	Effektstärke $d$	$N$
<b>Gymnasium</b>	106,1	23,4	0,21	3.440 (61,9 %)
<b>Gesamtschule</b>	92,8	23,7	-0,33	848 (15,3 %)
<b>WTG-G</b>	93,5	26,9	-0,30	150 (2,7 %)
<b>WTG-R</b>	90,4	24,9	-0,43	816 (14,7 %)
<b>Aufbaugymnasium VS</b>	97,8	24,2	-0,13	302 (5,4 %)
<b><i>insgesamt</i></b>	<i>101,0</i>	<i>24,8</i>	–	<i>5.556</i> <i>(100 %)</i>
<b>Aufbaugymnasium ES</b>	89,4	25,2	-0,47	516



Auffällig ist der vergleichsweise hohe Leistungsstand der Schülerinnen und Schüler aus der Vorstufe des Aufbaugymnasiums, die – anders als in den übrigen Leistungsbereichen – im Englisch-Leseverständnistest sogar das zweitbeste Ergebnis erzielt haben und mit 5,0 bzw. 4,3 Skalenpunkten und Effektstärken von  $d = 0,20$  bzw.  $d = 0,17$  vor den grundständigen Gesamtschülerinnen und Gesamtschülern bzw. den Schülerinnen und Schülern aus der Gruppe WTG-G liegen. Gleichfalls anders als in den übrigen Leistungsbereichen ist der vergleichsweise geringe Unterschied zwischen den beiden Gruppen WTG-G und WTG-R, der 3,1 Skalenpunkte beträgt mit einer Effektstärke von lediglich  $d = 0,13$ .

### 3.2.7 Zusammenfassung der allgemeinen Fachleistung zu Beginn der Klassenstufe 11

Zum Abschluss der Diskussion der fachbezogenen Lernstände zu Beginn der Klassenstufe 11 soll die Frage nach den erreichten Lernständen noch einmal in allgemeinerer Form thematisiert werden, und zwar vor allem deshalb, weil die Übergangsberechtigungen zu Beginn der Sekundarstufe II nicht an das Leistungsniveau in einzelnen Schulfächern, sondern an den allgemeinen Leistungsstand eines Schülers bzw. einer Schülerin geknüpft sind.

Wie schon im Bericht zur LAU 9 wird der Index für die allgemeine Fachleistung gebildet, indem zunächst die durchschnittliche Testleistung für die Fächer Mathematik (zu gleichen Teilen gewichtet aus den Untertests *Mathematik I* und *Mathematik II*), Deutsch (zu gleichen Teilen gewichtet aus den Untertests *Leseverständnis* und *passives Rechtschreibwissen*) und Englisch (zu gleichen Teilen gewichtet aus den Untertests *Leseverständnis* und *C-Test*) ermittelt und anschließend auf eine geeignete Skala mit einem Mittelwert von 50 und einer Standardabweichung von 10 projiziert<sup>14</sup> wird.

Die Verteilung der so bestimmten allgemeinen Fachleistung der Schülerinnen und Schüler an Hamburger Schulen zu Beginn der gymnasialen Oberstufe bzw. zu Beginn der Einführungsstufe des Aufbaugymnasiums entspricht für alle Gruppen in recht guter Näherung einer Normalverteilung (vgl. Abbildung 17).

Die Unterschiede zwischen den Bildungsganggruppen bestätigen das aus den fächerspezifischen Analysen bereits gewohnte Bild. Es werden erwartungsgemäß wieder große Überschneidungsbereiche zwischen den Gruppen deutlich. So haben 19,3 Prozent der grundständigen Gesamtschülerinnen und Gesamtschüler den Mittelwert der grundständigen Gymnasiastinnen und Gymnasiasten erreicht oder übertroffen, in der Gruppe WTG-G sind es 20,9 Prozent und in der Vorstufe der Aufbaugymnasien 17,2 Prozent gegenüber 10,6 Prozent der Schülerinnen und Schüler aus der Gruppe WTG-R und 7,0 Prozent der Schülerinnen und Schüler aus der Einführungsstufe des Aufbaugymnasiums.

---

<sup>14</sup> Hierzu waren die Ausgangswerte zunächst über eine sog. „z-Transformation“ zu standardisieren.

Abbildung 17 Allgemeine Fachleistung: Verteilung der Lernstände zu Beginn der Klassenstufe 11 nach Bildungsgang (absolut)

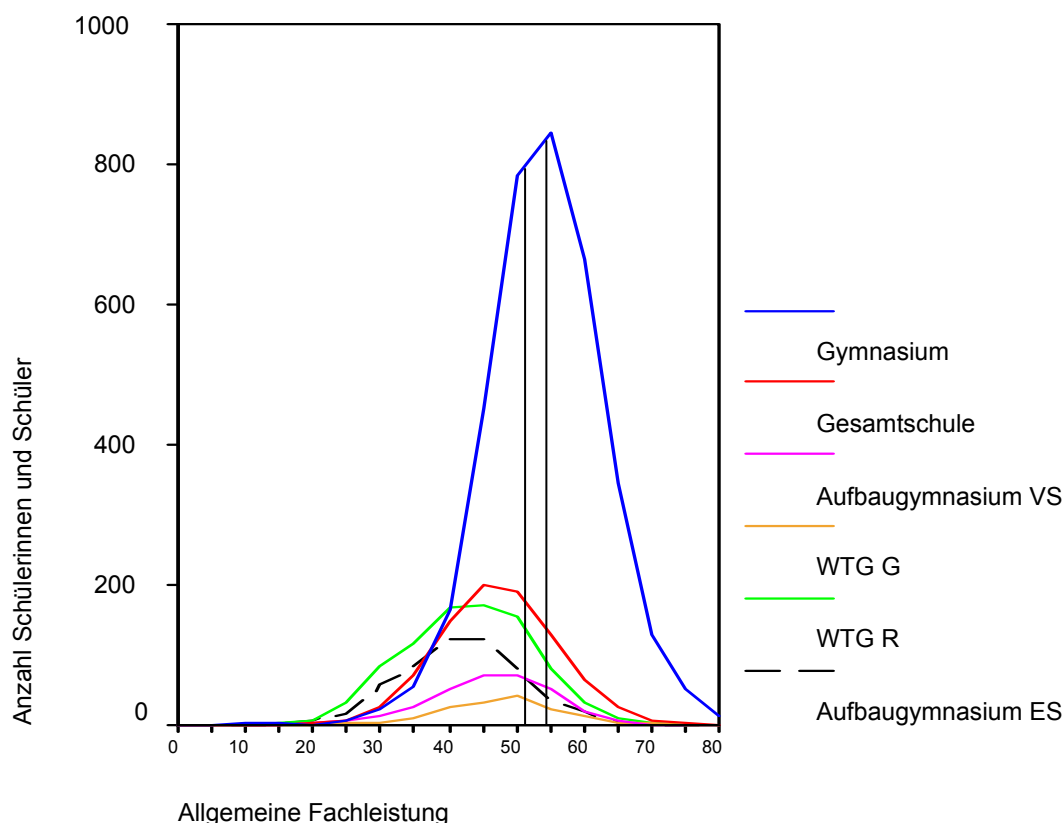
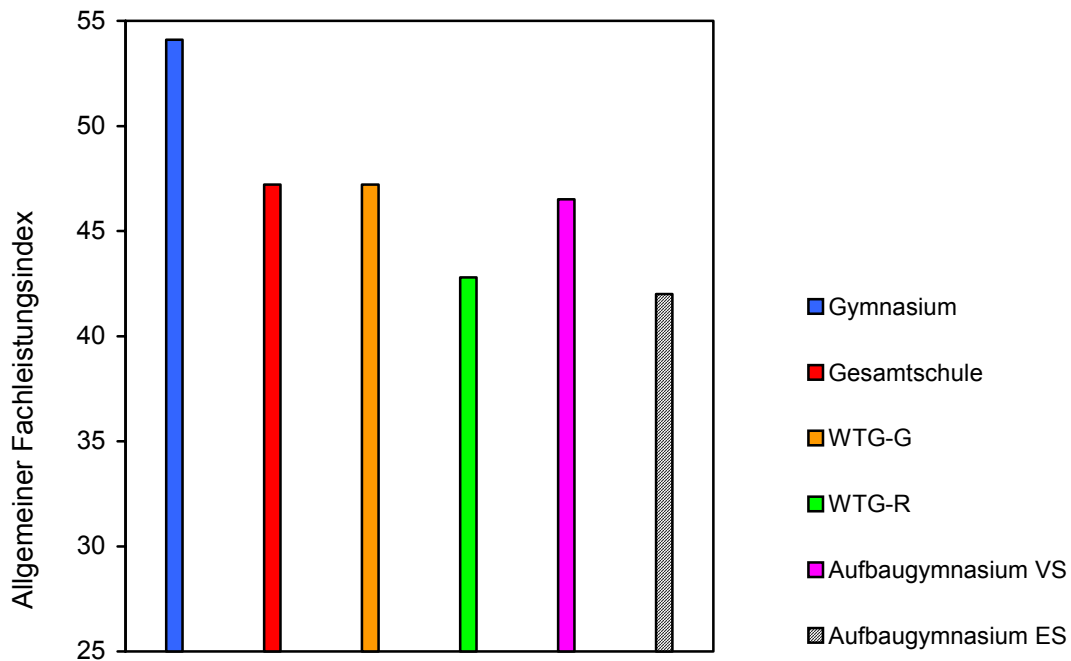


Abbildung 18 zeigt, analog den entsprechenden Grafiken für die einzelnen Fachleistungstests, die Leistungen der Schülerinnen und Schüler in einer anderen Form: Hier werden die Mittelwerte des allgemeinen Fachleistungsindex für die sechs Bildungsganggruppen einander gegenübergestellt.

Klar erkennbar ist auch hier der deutliche Vorsprung der grundständigen Gymnasiastinnen und Gymnasiasten vor den anderen Gruppen, der rund 7 Skalenpunkte bzw. mehr als 70 Prozent einer Standardabweichung vor den Gruppen Gesamtschule, WTG-R und Aufbaugymnasium VS und rund 12 Skalenpunkte bzw. annähernd 1,2 Standardabweichungen vor den beiden Gruppen WTG-R und Aufbaugymnasium ES beträgt (vgl. dazu Tabelle 19). Und wie bereits in den Fachleistungstests, so zeigt sich erneut, dass die Schülerinnen und Schüler, die mit einem Mittleren Schulabschluss einerseits in die Einführungsstufe des Aufbaugymnasiums, andererseits direkt in die Vorstufe eines Wirtschafts- oder technischen Gymnasiums eintreten, mit sehr ähnlichen Lernausgangslagen starten. Doch während die Schülerinnen und Schüler aus der Gruppe Aufbaugymnasium ES eine einjährige Vorbereitungsphase vor sich haben, um die leistungsbezogenen Voraussetzungen für die Versetzung in die Vorstufe zu erwerben, müssen die Schülerinnen und Schüler aus der Gruppe WTG-R einen nahezu identi-

sehen Lernrückstand auf ihre Mitschülerinnen und Mitschüler in der Vorstufe selbst aufholen.

Abbildung 18 Allgemeine Fachleistung: Verteilung der Lernstände zu Beginn der Klassenstufe 11 nach Bildungsgang



Auch für die allgemeine Fachleistung ist es von Bedeutung, wie sich die jeweiligen Leistungsverteilungen zueinander verhalten. Wie bei den entsprechenden fachspezifischen Tabellen sind daher in Tabelle 19 die Mittelwerte, Standardabweichungen und Effektstärken zusammengestellt. Die hier aufgelisteten Effektstärken  $d$  bezeichnen wiederum die relative Position der Mittelwerte für die sechs Bildungsganggruppen bezogen auf den Mittelwert und die Streuung der Gesamtheit aller untersuchten Schülerinnen und Schüler der gymnasialen Oberstufe.

Dem Anschein nach sind die Werte generell denen für die einzelnen Untertests ähnlich. Wegen der Möglichkeit eines kumulativen Zusammenwirkens der einzelnen Untertests lohnt es sich aber, zum Vergleich auch das Bestimmtheitsmaß  $Eta^2$  heranzuziehen, das angibt, welche Varianzanteile mit der Schulform bzw. Bildungsgangzugehörigkeit verbunden sind, und so die Effektstärken zu einem einzigen Kennwert zusammenfasst. Für die allgemeine Fachleistung beträgt  $Eta^2 = 0,24$ . Demgegenüber sind die Effekte der Untertests niedriger. Die entsprechenden Werte betragen für die Untertests *Mathematik I*  $Eta^2 = 0,13$ , *Mathematik II*  $Eta^2 = 0,12$ , *Deutsch-Leseverständnis*  $Eta^2 = 0,07$ , *passives Rechtschreibwissen*  $Eta^2 = 0,19$ , *Englisch (C-Test)*  $Eta^2 = 0,22$  und *Englisch-Leseverständnis*  $Eta^2 = 0,08$ .

Die schulformspezifischen Fachleistungsunterschiede zu Beginn der Sekundarstufe I waren mit  $Eta^2 = 0,34$  höher, und sie nahmen mit der Anzahl und Wirksamkeit äußerer Differenzierungen zwischen den Bildungsgängen noch zu. Die Aufteilung der Schülerschaft aus der Beobachtungsstufe der Haupt- und Realschulen in getrennte Hauptschul- und Realschulklassen auf der Grundlage der Fachleistungen insgesamt bzw. mit der Zuordnung der Schülerinnen und Schüler in den integrierten Schulformen in fachleistungsdifferenzierte Kursniveaus ließ den Anteil der damit verbundenen Leistungsdifferenzen in den Klassenstufen 7 bis 9 erheblich ansteigen, so dass in der LAU 9 das Bestimmtheitsmaß  $Eta^2 = 0,61$  betrug (vgl. LEHMANN & PEEK 1997, S. 105; vgl. LEHMANN, PEEK, GÄNSFUß und HUSFELDT 2002, 71 f.). Wenn nun ein wesentlich niedrigeres Differenzierungsniveau festzustellen ist, so hängt dies mit dem Übertritt eines großen, durchschnittlich weniger leistungsstarken Teils der Schülerschaft in berufliche Bildungsgänge zusammen.

Tabelle 19 Allgemeine Fachleistung zu Beginn der Klassenstufe 11: Mittelwerte, Standardabweichungen und Effektstärken, insgesamt und getrennt nach Bildungsgang

	Mittelwert	Standardabweichung	Effektstärke $d$	$N$
<b>Gymnasium</b>	54,1	8,6	0,34	3.540 (61,6 %)
<b>Gesamtschule</b>	47,2	8,5	-0,37	871 (15,2 %)
<b>WTG-G</b>	47,2	10,2	-0,37	158 (2,7 %)
<b>WTG-R</b>	42,8	9,3	-0,82	859 (15,0 %)
<b>Aufbaugymnasium VS</b>	46,5	8,4	-0,44	314 (5,5 %)
<b>insgesamt</b>	50,8	9,8	–	5.742 (100 %)
<b>Aufbaugymnasium ES</b>	42,0	8,8	-0,90	556

Aufgrund der starken positiven Auslese beim Übergang in die gymnasiale Oberstufe sinkt der Anteil der damit verbundenen Leistungsdifferenzen erheblich. Auch auf der Ebene der Untertests lässt sich dies zeigen. Um die Auswirkungen der äußeren Differenzierung aus einer anderen Perspektive zu charakte-

risieren, sind in Tabelle 20 noch einmal die Werte für das Bestimmtheitsmaß  $Eta^2$  aus den verschiedenen Phasen der Untersuchung zusammengestellt.

Tabelle 20 Unterschiede in den Fachleistungen zwischen den Schulformen und Kursniveaus bzw. den Bildungsgängen in Klassenstufe 11 (Bestimmtheitsmaß  $Eta^2$ ) nach Untertests<sup>15</sup>

	<b>Anfang Klassen- stufe 5</b>	<b>Ende Klassen- stufe 6</b>	<b>Anfang Klassen- stufe 7</b>	<b>Ende Klassen- stufe 8</b>	<b>Anfang Klassen- stufe 11</b>
<b>Mathematik</b>	0,22	0,36	0,43	0,48	0,13
<b>Mathematik II</b>	–	–	–	–	0,12
<b>Deutsch: Lesever- ständnis</b>	0,24	0,27	0,37	0,32	0,07
<b>Deutsch: Sprache</b>	0,26	0,42	0,44	0,44	–
<b>Rechtschreibwissen</b>	0,23	–	–	0,39	0,19
<b>Rechtschreibkönnen</b>	0,23	0,26	0,37	0,37	–
<b>Englisch: C-Test</b>	–	0,27	0,61	0,46	0,22
<b>Englisch: Lesever- ständnis</b>	–	–	–	–	0,08
<b>allgemeine Fach- leistung</b>	0,34	0,54	0,64	0,61	0,24

Danach haben die Schulformunterschiede während der Klassenstufen 5 und 6 zunächst in allen untersuchten Leistungsbereichen zugenommen. Insbesondere an der teilweise sprunghaften Zunahme des Bestimmtheitsmaßes  $Eta^2$  beim Übergang vom Ende der Klassenstufe 6 zum Anfang der Klassenstufe 7 wird ersichtlich, dass die größeren Schulformunterschiede ganz überwiegend dadurch zustande gekommen sind, dass mit der äußeren Leistungsdifferenzierung an den Haupt- und Realschulen sowie mit der Fachleistungsdifferenzierung an den Gesamtschulen zu Beginn der Klassenstufe 7 Niveauzuordnungen vorgenommen wurden, die sich im Folgenden keineswegs durchgängig als Grundlage divergen- ter Lernentwicklungen darstellen (LEHMANN, PEEK, GÄNSFUß UND HUSFELDT 2002, 73).

<sup>15</sup> Für die integrierten Schulformen erfolgte die Gruppenbildung fachspezifisch. Um auch bei wechselnder Kursniveauezugehörigkeit Vergleichbarkeit zwischen der LAU 7 und der LAU 9 herzustellen, wurde (in einzelnen Fällen kontrafaktisch) für den Anfang der Klassenstufe 7 dasselbe Kursniveau unterstellt wie für das Ende der Klassenstufe 8. Für Klassenstufe 11 wurde als Prädiktor die Bildungsganggruppe herangezogen, da die bisherigen Schulformen bzw. Kursniveaus in der Sekundarstufe II nicht mehr relevant sind.

### 3.3 Zusammenfassung

Die Analyse der Fachleistungen von Schülerinnen und Schülern an Hamburger Schulen zu Beginn der Klassenstufe 11 hat vor allem zwei Ergebnisse geliefert:

Einerseits sind erwartungsgemäß in der gymnasialen Oberstufe des Hamburger Bildungssystems erhebliche Überschneidungsbereiche zwischen den einzelnen Bildungsganggruppen vorhanden. In allen Formen der gymnasialen Oberstufe gibt es Schülerinnen und Schüler mit hervorragenden Fachleistungen. Bemerkenswert ist gleichwohl der immense durchschnittliche Leistungsunterschied zwischen den grundständigen Gymnasiastinnen und Gymnasiasten und allen anderen Bildungsganggruppen in der gymnasialen Oberstufe. Für nicht wenige der Schülerinnen und Schüler, die mit einem Mittleren Schulabschluss in die gymnasiale Oberstufe wechseln, begründet der feststellbare Leistungsrückstand Zweifel an der Erreichbarkeit eines Fachleistungsniveaus, das der Allgemeinen Hochschulreife in der Substanz entspricht. Umgekehrt findet sich jedoch auch in der Schülerschaft der grundständigen Gymnasien ein nennenswerter Anteil solcher Jugendlicher, deren Fachleistungen eher dem Anforderungsniveau entsprechen, das an Realschulen erreicht wird. 9,3 Prozent der grundständigen Gymnasiastinnen und Gymnasiasten haben allgemein geringere Fachleistungen nachgewiesen, als es dem Durchschnitt der getesteten Schülerinnen und Schüler mit einem Mittleren Schulabschluss (43,2 Skalenpunkte)<sup>16</sup> entspricht. Demgegenüber findet sich unter diesen Schülerinnen und Schülern mit 10,6 Prozent ein durchaus nennenswerter Anteil mit einer allgemeinen Fachleistung, die den Mittelwert der grundständigen Gymnasiastinnen und Gymnasiasten erreicht oder übertroffen hat.

Betrachtet man die einzelnen Fachleistungen getrennt, so zeigen sich deutlichere Überschneidungen: Für die einzelnen Fachleistungen gibt Tabelle 21 darüber Aufschluss, wie viel Prozent der grundständigen Gymnasiastinnen und Gymnasiasten unterhalb des entsprechenden Mittelwerts der Schülerinnen und Schüler mit einem Mittleren Schulabschluss liegen. Für einen beträchtlichen Teil der grundständigen Gymnasiastinnen und Gymnasiasten ist ein Vorsprung in den einzelnen Fachleistungsbereichen *nicht* festzustellen. Ein Viertel der Schülerinnen und Schüler dieser Gruppe liegt in den Fachleistungen Mathematik II und Deutsch: Leseverständnis unter dem Mittelwert der Schülerinnen und Schüler mit einem Mittleren Schulabschluss. In den anderen Fachleistungen schwanken die Anteile zwischen 9,3 und 21,5 Prozent.

---

<sup>16</sup> Der Mittelwert für die Testversion I beträgt 52,4 Skalenpunkte.

Tabelle 21 Prozentanteil der Schülerinnen und Schüler an grundständigen Gymnasien, deren Fachleistungen unterhalb des Mittelwertes der mit Testversion II Getesteten liegen, sowie Prozentanteil der Schülerinnen und Schüler mit einem Mittlerem Schulabschluss (Testversion II), deren Fachleistungen über dem Mittelwert der grundständigen Gymnasiastinnen und Gymnasiasten liegen

	Mathe- matik I	Mathe- matik II	Deutsch Lese- ver- ständ- nis	Deutsch Recht- schrei- ben	Eng- lisch C-Test	Eng- lisch Lese- verständ- nis	Allge- meine Fach- leistung
<b>Mittelwert der Schüler mit Mittlerem Schulabschluss (TV II)<sup>17</sup></b>	131,8	11,1	137,2	13,2	140,6	91,5	43,2
Standardabweichung	15,8	3,9	17,3	5,1	15,4	25,0	9,1
N	1.701	1.700	1.702	1.644	1.636	1.634	1.729
Effektstärke*	-0,51	-0,50	-0,39	-0,59	-0,64	-0,34	-0,68
<b>Mittelwert der Gymnasiasten</b>	145,4	14,5	146,8	18,3	158,1	106,1	54,1
Standardabweichung	15,7	4,3	15,1	4,6	14,5	23,4	8,6
N	3.471	3.469	3.491	3.407	3.441	3.440	3.540
Effektstärke*	0,31	0,27	0,20	0,35	0,40	0,24	0,41
<b>Anteil der grundständigen Gymnasiasten, deren Fachleistung <u>unter</u> dem Mittelwert für Schülerinnen und Schüler mit Mittlerem Schulabschluss liegt</b>	<b>19,1%</b>	<b>24,4 %</b>	<b>25,4 %</b>	<b>15,4 %</b>	<b>8,1 %</b>	<b>21,5 %</b>	<b>9,3 %</b>
<b>Anteil der Schülerinnen und Schüler mit Mittlerem Schulabschluss, deren Fachleistung <u>über</u> dem Mittelwert der grundständigen Gymnasiasten liegt</b>	<b>18,6 %</b>	<b>19,6 %</b>	<b>27,8 %</b>	<b>15,8 %</b>	<b>10,4 %</b>	<b>27,7 %</b>	<b>10,6 %</b>

\* Berechnet auf der Basis der Kennwerte für die Gesamtpopulation.

<sup>17</sup> Dazu gehören sowohl die Schülerinnen und Schüler aus der Einführungsstufe und der Vorstufe des Aufbaugymnasiums als auch die Schülerinnen und Schüler aus den und technischen Gymnasien mit einem Mittlerem Schulabschluss.



Andererseits ist nicht zu übersehen, dass in den durch Durchschnittswerte charakterisierten ‚typischen Fällen‘ erhebliche Abstände in den Leistungsständen zwischen den verschiedenen Formen der gymnasialen Oberstufe bestehen, die das Äquivalent von mehreren Lernjahren darstellen können. Insbesondere gelten entsprechende Rückstände im Durchschnitt für diejenigen, die mit einem qualifizierten Realschulabschluss nach den geltenden Übergangsregelungen die Berechtigung erworben haben, ihren Bildungsgang in der gymnasialen Oberstufe fortzusetzen. Schon dieser Abstand erweist die Praxis als Illusion, durch Verlangsamung des Lerntempos in der Klassenstufe 11 der Mehrheit in dieser Gruppe den Anschluss an die Schülerinnen und Schüler grundständiger Gymnasien zu ermöglichen. Aber auch die Differenz in den Fachleistungen zwischen der gymnasialen Oberstufe der grundständigen Gymnasien und der grundständigen Gesamtschulen sowie der Fachgymnasien erscheint größer, als es die formale Äquivalenzbehauptung zulässt.

## **4 Entwicklung der Fachleistungen in den Klassenstufen 9 und 10**

### **4.1 Methodische Vorüberlegungen**

Klassenwiederholungen, das Überspringen einer Klassenstufe sowie Abgänge (Fortzüge bzw. Wechsel in eine Berufsvorbereitungsmaßnahme oder in einen beruflichen Bildungsgang) haben – als Konsequenz der Anlage der Längsschnittuntersuchung – dazu geführt, dass davon betroffene Schülerinnen und Schüler in den jeweils folgenden Erhebungswellen nicht mehr erfasst werden konnten. Andererseits sind auch viele Schülerinnen und Schüler auf entsprechenden Wegen neu in die Untersuchungsgruppe gelangt. Insgesamt ist durch diese Fluktuation – bezogen auf die Gesamtheit der in der ursprünglichen Stichprobe von 1996 untersuchten Schülerinnen und Schüler – eine Überschätzung der jeweils erreichten Lernstände eingetreten, die sich auch auf die im Folgenden dargestellten Lernzuwachsrechnungen auswirkt. Der Grund dafür liegt darin, dass die Abgänge überwiegend aus dem unteren Bereich des Leistungsspektrums stammen; allerdings haben auch die Zugänge in der LAU 11 – insgesamt betrachtet – ebenfalls eine geringere Leistungsposition innerhalb der Klassenstufe (vgl. Kapitel 3, Tabelle 10). Demnach kann im Hinblick auf den Einfluss von Zu- und Abgängen auf das Leistungsniveau für die Klassenstufe 11 von sich gegenseitig zumindest teilweise ausgleichenden Phänomenen gesprochen werden. Die Stärke der betreffenden Effekte kann indessen nicht abgeschätzt werden.

Von den insgesamt 12.620 Schülerinnen und Schülern, die an der LAU 9 teilgenommen haben, wurden 4.614 in der LAU 11 erneut getestet. Annähernd 63 Prozent (8.006 Jugendliche) haben die Schule nach der neunten oder zehnten Klasse verlassen, haben eine Klasse wiederholt oder sind mit ihren Familien weggezogen. 2.346 Schülerinnen und Schüler, die in der LAU 9 getestet wurden, haben 2002 an der Paralleluntersuchung ULME („Untersuchung von Leistung, Motivation und Einstellungen zu Beginn der beruflichen Ausbildung in Hamburg“) teilgenommen.

Angenommen wird, dass die Nichtteilnehmer an der Erhebungswelle 2002 überwiegend durch Beendigung der Schulausbildung am Ende der Sekundarstufe I<sup>18</sup> und nur zum Teil durch Klassenwiederholung aus der Untersuchungs-

---

<sup>18</sup> Gemeint sind Schülerinnen und Schüler, die in eine Berufsausbildung übergegangen sind bzw. eine solche anstreben, Jugendliche, die ein freiwilliges soziales oder ökologisches Jahr absolvieren oder ein Praktikum ablegen etc.

gruppe ausgeschieden sind. Die Neuzugänge (1.725 Schülerinnen und Schüler)<sup>19</sup>, die einerseits durch Klassenwiederholungen in den Klassenstufen 10 und 11 oder durch Zuzug in die Untersuchungsgruppe der LAU 11, andererseits über den erfolgreichen Besuch der Einführungsstufe des Aufbaugymnasiums oder den Erwerb eines qualifizierten Mittleren Schulabschlusses im Rahmen beruflicher Bildungsgänge gekommen sind, dürfen ebenfalls im Längsschnitt nicht berücksichtigt werden, da deren Lernstände allenfalls teilweise den Lerngruppen in Klassenstufe 9 und 10 zugeordnet werden können.

Eine umfassende und unverzerrte Abschätzung der Lernzuwächse bis zum Ende der Klassenstufe 10, die sowohl die abgebende Schulform der Sekundarstufe I als auch die Form der aufnehmenden gymnasialen Oberstufe berücksichtigt, ist demnach nicht möglich. Immerhin können Ausleseeffekte beim Übergang in die Sekundarstufe II grob abgeschätzt werden, wenn man innerhalb der Schulformen der Sekundarstufe I die Fachleistungsunterschiede zu Beginn der Klassenstufe 9 zwischen denjenigen Jugendlichen, die an der LAU 11 teilgenommen haben, und den übrigen vergleicht (vgl. Tabelle 22).

Tabelle 22 Allgemeine Fachleistung der Schülerinnen und Schüler (einschließlich Einführungsstufe des Aufbaugymnasiums) nach Schulform zu Beginn der Klassenstufe 9 im Vergleich zwischen Teilnehmern und Nichtteilnehmern an der Erhebung 2002

ursprüngliche Schulform- zugehörigkeit in Klassen- stufe 9	Teilnahme an den Erhebungen 2002 (LAU 11)						Effekt- stärke <i>d</i>	Anteil Nicht- teilneh- mer in 2002
	nein		ja		insgesamt			
	<i>N</i> = 6.061		<i>N</i> = 4.446		<i>N</i> = 10.507			
	$\bar{x}$	<i>s</i>	$\bar{x}$	<i>s</i>	$\bar{x}$	<i>s</i>		
<b>Realschule</b>	48,2	6,0	49,9	6,0	48,5	6,0	0,28	80,7 %
<b>IHR-Schule</b>	42,9	7,0	49,0	6,0	43,7	7,2	0,85	86,4 %
<b>Gesamtschule</b>	43,0	7,0	50,6	7,2	44,9	7,8	0,97	74,9 %
<b>Gymnasium</b>	58,2	6,7	59,1	6,0	58,8	6,2	0,15	30,9 %

<sup>19</sup> 1.725 Neuzugänge und 4.614 Schüler, die aus der LAU 9 bekannt sind, ergeben zusammen 6.339 Jugendliche. Insgesamt wurden in der LAU 11 aber 6.411 Schülerinnen und Schüler getestet. Die Differenz von 72 ergibt sich dadurch, dass weitere Jugendliche in den Klassenstufen 5 und/oder 7 sowie 11, aber nicht in Klasse 9 getestet worden waren.

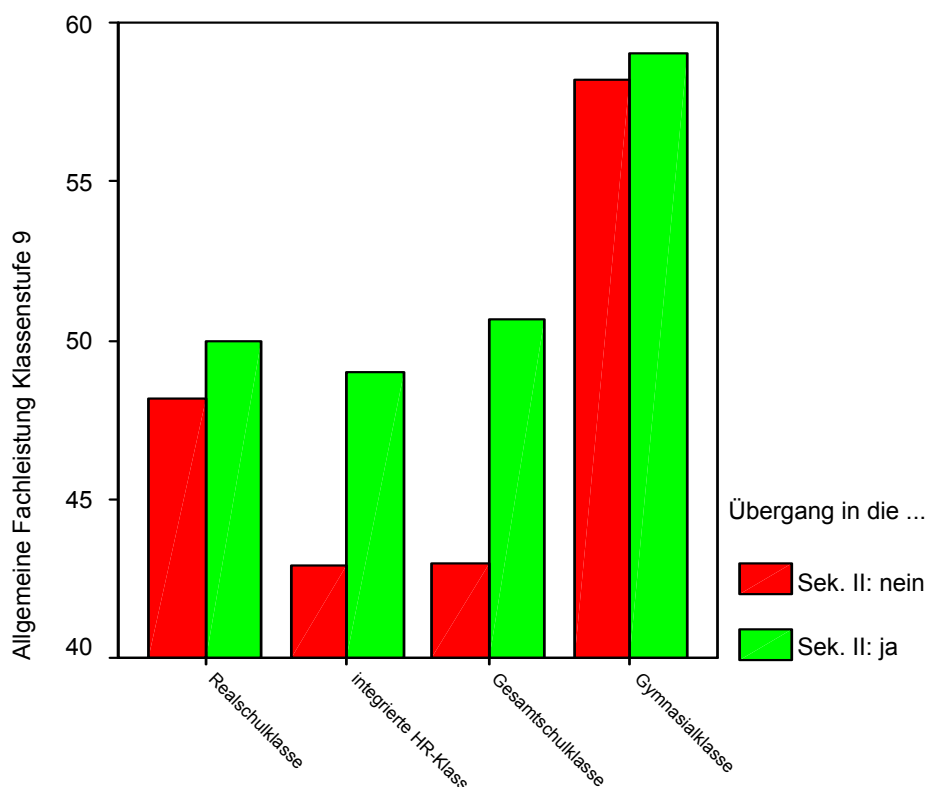
Tabelle 22 zeigt die allgemeine Fachleistung der Schülerinnen und Schüler im Jahr 2000 (LAU 9), getrennt nach Teilnahme an der LAU 11 unter Einschluss der integrierten Schulformen, jedoch ohne Rücksicht auf darin gegebene Kursniveaudifferenzierungen. Die zum Teil hohen Effektstärken belegen den erwarteten großen Unterschied zwischen beiden Gruppen. Diese Angaben beziehen sich auf alle Schülerinnen und Schüler, die an der LAU 9 teilgenommen haben bzw. in der LAU 9 und in der LAU 11 getestet wurden und für die ein allgemeiner Fachleistungsindex berechnet werden konnte. Deshalb weichen die angegebenen Schülerzahlen von den zuvor genannten ab.

Zu beachten ist, dass sich die Prozentanteile der Schülerinnen und Schüler, die an der LAU 11 nicht teilgenommen haben, je nach Schulform erheblich unterscheiden. An den Realschulen sowie an den integrierten Haupt- und Realschulen hat erwartungsgemäß der Großteil der Jugendlichen (ca. 82 Prozent) nach der 10. Klasse die Schule verlassen. An den Gesamtschulen ist die Situation differenzierter: Hier sind es 25 Prozent, die in den verschiedenen Bildungsganggruppen der LAU 11 angetroffen wurden, eine weitere Gruppe ist zwar in die gymnasiale Oberstufe versetzt worden, besuchte aber zum Zeitpunkt der LAU-11-Erhebung eine Schule im Ausland (nach der amtlichen Schulstatistik: 75 Jugendliche). Erwartungsgemäß erreicht der größte Teil der Gymnasiastinnen und Gymnasiasten die Sekundarstufe II; dass dennoch ca. 30 Prozent der Schülerschaft nicht in der Untersuchungsgruppe der LAU 11 angetroffen wurden, ist einerseits auf die Wiederholerquoten in den Klassenstufen 9 und 10 (laut amtlicher Schulstatistik 3,0 Prozent im Schuljahr 2001/02, 2,3 Prozent im Schuljahr 2002/03) zurückzuführen, andererseits auf den hohen Anteil von Schülerinnen und Schülern, die in der Klassenstufe 11 eine Schule im Ausland besuchen (zum Erhebungszeitpunkt waren es nach amtlicher Schulstatistik 549 Vorstufenschülerinnen und -schüler).

Abbildung 19 verdeutlicht grafisch die Leistungsunterschiede in der Klassenstufe 9 für diejenigen Schülerinnen und Schüler, die an der LAU 11 teilgenommen haben, im Vergleich zu den Jugendlichen, die nach der neunten oder zehnten Klasse die Schule verlassen haben, eine Klasse wiederholten oder aus sonstigem Grund nicht an der LAU 11 teilgenommen haben. Dabei entsprechen die großen Lernstandsunterschiede zwischen den Schulabgängern und jenen, die in der LAU 11 erfasst wurden, in den integrierten Schulformen den schul- bzw. schulformspezifischen Selektionsentscheidungen. Deren Gewicht wird über die numerische Darstellung der Effektstärke  $d$  (vgl. Tabelle 22) und in der grafischen Darstellung gleichermaßen deutlich. Für die Realschulen ist die Geringfügigkeit des Ausleseprozesses beim Übergang in die gymnasiale Oberstufe überraschend (Effektstärke  $d = 0,28$ ). Dass der Lernstand derjenigen Schülerinnen

und Schüler des Gymnasiums, die nicht in der Klassenstufe 11 angetroffen wurden, fast gleich hoch ist wie der Lernstand jener Schülerinnen und Schüler, die an der LAU 11 teilgenommen haben ( $d = 0,15$ ), dürfte vor allem auf diejenigen Schülerinnen und Schüler zurückgehen, die eine Schule im Ausland besuchen und eine eher leistungsstarke Gruppe darstellen.

Abbildung 19 Allgemeine Fachleistung der Schülerinnen und Schüler (einschließlich Einführungsstufe des Aufbaugymnasiums) zu Beginn der Klassenstufe 9 nach Schulform und getrennt nach Teilnahme und Nichtteilnahme an der Erhebung 2002



Mit den Lernstandswerten für die Fachleistungen Mathematik: I, Deutsch: Leseverständnis, Deutsch: Rechtschreiben und Englisch: C-Test stehen Grundlagen für eine Einschätzung der Lernentwicklungen von 1996 (für Englisch: C-Test von 1998) bis 2002 in den verschiedenen Schulformen bzw. Bildungsganggruppen zur Verfügung.

In diesem Kapitel soll in jeweils vier Schritten vorgegangen werden. Zunächst wird – ausgehend von der Klassenstufe 9 – für jede Schulform der durchschnittliche Lernzuwachs in der Teilgruppe bestimmt, die durch die Teilnahme sowohl an der LAU 9 als auch an der LAU 11 definiert ist, und vergleichenden Betrachtungen unterzogen. Sodann sollen die differenziellen Lernzuwachsraten in den Klassenstufen 9 und 10 mit Bezug auf die Lernausgangslage 2000 unter-

sucht werden. Im Anschluss folgt die Betrachtung der Lernentwicklung an einzelnen Schulen. Analysen der Lernentwicklung über den Gesamtzeitraum der Längsschnittuntersuchung schließen jeweils die Darstellungen ab.

Aus der neuen Gruppenbildung mit unterschiedlichen Testversionen, aber einheitlicher Metrik kann im Einzelfall darauf geschlossen werden, mit welchen Eingangsvoraussetzungen die Schülerinnen und Schüler in die Sekundarstufe II – mit einem qualifizierten Mittleren Schulabschluss oder aus der Mittelstufe eines grundständigen Gymnasiums, alternativ aus einer Gesamtschule – bzw. in die Einführungsstufe des Aufbaugymnasiums eingetreten sind. Somit informiert diese Unterscheidung auch darüber, wo die Jugendlichen zu Beginn der Klassenstufe 9 unterrichtet wurden. Da, wie mehrfach dargelegt, zahlreiche Teilnehmerinnen und Teilnehmer der LAU 9 aus systematischen Gründen nicht in der Zielgruppe der LAU 11 enthalten waren, werden die Leistungen der getesteten Schülerinnen und Schüler zusätzlich getrennt nach Teilnahme an der LAU 11 analysiert. Diese Vorgehensweise ermöglicht eine Interpretation, die von der ehemaligen Schulform (Klassenstufe 9) ausgeht. Gleichzeitig gibt sie Einblicke in die Selektionsprozesse beim Übergang aus der Sekundarstufe I in die Sekundarstufe II nach Maßgabe der Fachleistung zu Beginn der Klassenstufe 9 in den einzelnen Fächern.

Da die Lernzuwächse in den Klassenstufen 9 und 10 untersucht werden sollen, ist es naheliegend, die Analysen primär der alten Schulformgruppierung entsprechend vorzunehmen. Mithin beziehen sich die Berechnungen des Lernzuwachses für die beiden letzten Schuljahre auf die Schulformzugehörigkeit der LAU 9 in Kombination mit der Bildungsganggruppierung der LAU 11.

Die Berechnung der Lernzuwächse für den Gesamtzeitraum des Längsschnitts kann hingegen nur aus der Perspektive der Schulformgruppierung der LAU 11 erfolgen, da Schulform- und Kursniveauwechsel über den erfassten Zeitraum von sechs Jahren keinesfalls vernachlässigbare Größen darstellen. Sie würden die Ergebnisse verzerren, wollte man die Schülerinnen und Schüler einheitlich einer bestimmten Schulform- bzw. Kursniveaugruppe zuordnen. In diese Analyse werden deshalb nur die Jugendlichen einbezogen, für die sowohl Leistungsdaten aus der LAU 11 als auch zu *allen* übrigen Testzeitpunkten der Längsschnittuntersuchung vorliegen.

## 4.2 Entwicklung der Fachleistung Deutsch: Leseverständnis

Für die Fachleistung Deutsch – differenziert für *Leseverständnis* und *Rechtschreibung* – werden im Folgenden die Lernentwicklungen der Schülerinnen und Schüler, die in die gymnasiale Oberstufe oder in die Einführungsstufe des Aufbaugymnasiums eingetreten sind, vom Beginn der Klassenstufe 9 bis zum Beginn der Klassenstufe 11 dargestellt. Grundlage sind alle Schülerinnen und Schüler, für die Lernstandswerte sowohl aus der Erhebung des Jahres 2000 als auch aus derjenigen des Jahres 2002 zur Verfügung stehen.

Zunächst sollen für die Ergebnisse im Untertest *Leseverständnis* die statistischen Kennwerte für den Beginn der Klassenstufe 9 und für den Beginn der Klassenstufe 11 sowie die daraus gebildeten Differenzwerte als Indikatoren für die Lernfortschritte berichtet werden. Die vorletzte Zeile der Tabelle 23 (siehe unten, S. 81) gibt Aufschluss über den Mittelwert und die Standardabweichung der gesamten Schülerschaft der Klassenstufe 9 (ohne die Schülerinnen und Schüler in der Einführungsstufe des Aufbaugymnasiums) – also für 10.106 Jugendliche. Diese Schülerzahl weicht von der vorher berichteten ( $N = 12.620$ ) ab, weil zum einen die Hauptschülerinnen und Hauptschüler in dieser Tabelle unberücksichtigt bleiben, zum anderen nicht alle Schülerinnen und Schüler den Deutsch-Leseverständnistest bearbeitet haben und zum Dritten aufgrund von Schulformwechsel bzw. des Eintritts in die Einführungsstufe des Aufbaugymnasiums nicht alle Jugendlichen in die neue Gruppierung eingeordnet werden können. Der letzten Zeile der Tabelle ist zu entnehmen, dass für das Leseverständnis über alle Schulformen hinweg eine Erhöhung der Lernstandswerte um 12,9 Skalenpunkte oder 0,90 Standardabweichungen festzustellen ist. Berechnet man das Wachstum für ein Schuljahr, so ergeben sich 0,45 Standardabweichungen – ein Wert, der deutlich über dem für die Klassenstufen 5 bis 8 ermittelten liegt, jedoch auf eine Teilpopulation beschränkt ist. Nur für 3.948 Schülerinnen und Schüler liegen sowohl für die Klassenstufe 9 als auch für die Klassenstufe 11 Leistungsdaten für Deutsch-Leseverständnis vor.

Insgesamt hat sich in der hier untersuchten Teilgruppe die Leistungsstreuung erhöht. Das Gegenteil wäre vielleicht zu erwarten gewesen, hatte sich doch in den bisherigen Analysen die Gesamtstreuung wie auch die Streuung innerhalb der Schulformen und Kursniveaus von Untersuchung zu Untersuchung stetig verringert. Nunmehr aber haben sich die Rahmenbedingungen entscheidend verändert: Die Schülerschaft, deren Lernentwicklung vom Beginn der Klassenstufe 9 bis zum Beginn der Klassenstufe 11 rekonstruiert werden kann, stellt auch innerhalb der verschiedenen Bildungsgänge eine positive, vergleichsweise homogene Auswahl von Jugendlichen dar. Das bedeutet, dass die Varianz einer Teil-

menge der in LAU 9 getesteten Schülerinnen und Schüler hier Ausgangswert für die Berechnungen der Lernentwicklung ist.

Gleichzeitig ist zu vermuten, dass sich auch die Bildungsaspirationen in dieser Gruppe schon in der Klassenstufe 9 relativ stark auf den Eintritt in die gymnasiale Oberstufe konzentriert haben. Indessen zeigen sich im Hinblick auf die Lernentwicklung und hier insbesondere unter der Perspektive veränderter Heterogenität Unterschiede zwischen den Bildungsganggruppen, die differenziell gedeutet werden müssen.

Hatte sich die Gesamtstreuung der grundständigen Gymnasiastinnen und Gymnasiasten von der LAU 5 zur LAU 9 stetig verringert, so hat sie sich in den Klassenstufen 9 und 10 wieder deutlich – um 28,7 Prozent – erhöht, was darauf hindeutet, dass die Schülerinnen und Schüler in recht ungleichem Maße vom Unterricht profitieren konnten. Möglicherweise war es in dieser Schulform eine Begleiterscheinung der erkennbaren Beschleunigung des Lerntempos, dass ein Teil der Schülerinnen und Schüler den Anschluss nicht halten konnten, womit sich die Leistungsstreuung erhöht hätte.

Demgegenüber hat sich die Streuung in den beiden Gruppen mit den höchsten Lernzuwachsen – den Gesamtschülerinnen und Gesamtschülern, die in die gymnasiale Oberstufe versetzt wurden, sowie den Realschülerinnen und Realschülern, die mit einem qualifizierten Mittleren Schulabschluss in die Vorstufe des Aufbaugymnasiums bzw. des Wirtschafts- oder technischen Gymnasiums eingetreten sind – um 25,8 bzw. 5,0 Prozent vermindert. Die beträchtliche Streuungsminderung an den Gesamtschulen bei gleichzeitig hohem Leistungswachstum scheint ein Indiz dafür zu sein, dass das Bemühen, Anschluss an das übliche Gymnasialniveau zu finden, recht erfolgreich und jedenfalls nicht um den Preis der Vernachlässigung des vergleichsweise weniger lernstarken Segments in diesem Teil des Leistungsspektrums erfolgt ist.

Betrachtet man das *relative* Leistungswachstum – gemessen an der *Effektstärke*  $d^{20}$  –, so ergibt sich folgendes Bild: Bei Berücksichtigung der Ausgangsvarianz zu Beginn der Klassenstufe 9 verzeichnen die grundständigen Gymnasiastinnen und Gymnasiasten noch einen etwas größeren Lernzuwachs als die Gesamtschülerinnen und Gesamtschüler mit Versetzung in die gymnasiale Oberstufe und als die Realschülerinnen und Realschüler, die in eine gymnasiale Oberstufe eingetreten sind.

---

<sup>20</sup> Hier wird die *Effektstärke*  $d$  so berechnet, dass die Differenz aus der Fachleistung Deutsch-Leseverständnis der Klasse 11 und derjenigen aus Klasse 9 durch die Standardabweichung der Ausgangslage – also zu Beginn der Klassenstufe 9 – dividiert wird.



Tabelle 23 Fachleistung Deutsch-Leseverständnis: Lernstandswerte zu Beginn der Klassenstufe 9 und zu Beginn der Klassenstufe 11 sowie Lernzuwachs, nach Schulform der Klasse 9

Schulform Klasse 9	Bildungsgang Klasse 11	Beginn Klasse 9		N	Beginn Klasse 11		Lernzuwachs		
		$\bar{x}$	SD		$\bar{x}$	SD	Differenz	<i>d</i>	Varianz- änderung
IHR-Schule	<i>Abgänge</i>	112,7	17,6	587	–	–	–	–	–
IHR-Schule	Aufbaugymn. VS, WTG-R	123,2	15,6	43	133,7	19,3	10,5	0,67	+52,6 %
Realschule	<i>Abgänge</i>	121,1	16,1	1.635	–	–	–	–	–
Realschule	Aufbaugymn. VS, WTG-R	122,6	16,5	200	136,7	16,1	<b>14,1</b>	0,85	-5,0 %
Gesamt- schule	<i>Abgänge</i>	112,4	19,3	2.517	–	–	–	–	–
Gesamt- schule	Aufbaugymn. VS, WTG-R	121,9	16,5	72	134,6	13,3	12,7	0,77	+22,9 %
Gesamt- schule	Gymn., GS, WTG-G	128,7	16,8	612	143,7	14,4	<b>15,0</b>	<b>0,89</b>	-25,8 %
Gymnasium	<i>Abgänge</i>	133,2	13,6	1.419	–	–	–	–	–
Gymnasium	Gymn., GS, WTG-G	134,4	13,0	3.021	146,8	14,8	12,4	<b>0,95</b>	+28,7 %
<i>insgesamt</i> <i>N = 10.106</i>		124,6	18,5		–	–	–	–	–
<i>insgesamt</i> <i>N = 3.948<sup>21</sup></i>		<b>132,6</b>	<b>14,4</b>		<b>145,5</b>	<b>15,2</b>	<b>12,9</b>	<b>0,90</b>	<b>+11,7 %</b>

<sup>21</sup> Nicht alle der 4.614 Schülerinnen und Schüler (vgl. Kap. 2, Tabelle 8) konnten den Gruppen zugeordnet werden. So bleiben hier die Schülerinnen und Schüler unberücksichtigt, die sich in der Einführungsstufe des Aufbaugymnasiums befanden. Des Weiteren konnten Jugendliche nicht eingeordnet werden, die offensichtlich während oder nach der neunten bzw. zehnten Klasse die Schulform oder das Kursniveau gewechselt haben (N = 40). An dieser Stelle ist ebenfalls zu berücksichtigen, dass nicht für alle Schülerinnen und Schüler, die sowohl an der LAU 9 als auch an der LAU 11 teilgenommen haben, für Deutsch-Leseverständnis Leistungsdaten für beide Erhebungszeitpunkte vorliegen.

Mit sehr hohen Effektstärken von  $d = 0,95$  (Gymnasium) bzw.  $d = 0,89$  (Gesamtschule) und  $d = 0,85$  (Realschule) haben alle drei Schülergruppen innerhalb von zwei Schuljahren auffällig viel dazugelernt. Ein Deckeneffekt kann dabei ausgeschlossen werden: Lernstandswerte über 190 wurden nur von wenigen Schülerinnen und Schülern erreicht. Ebenso ist ein Bodeneffekt auszuschließen.

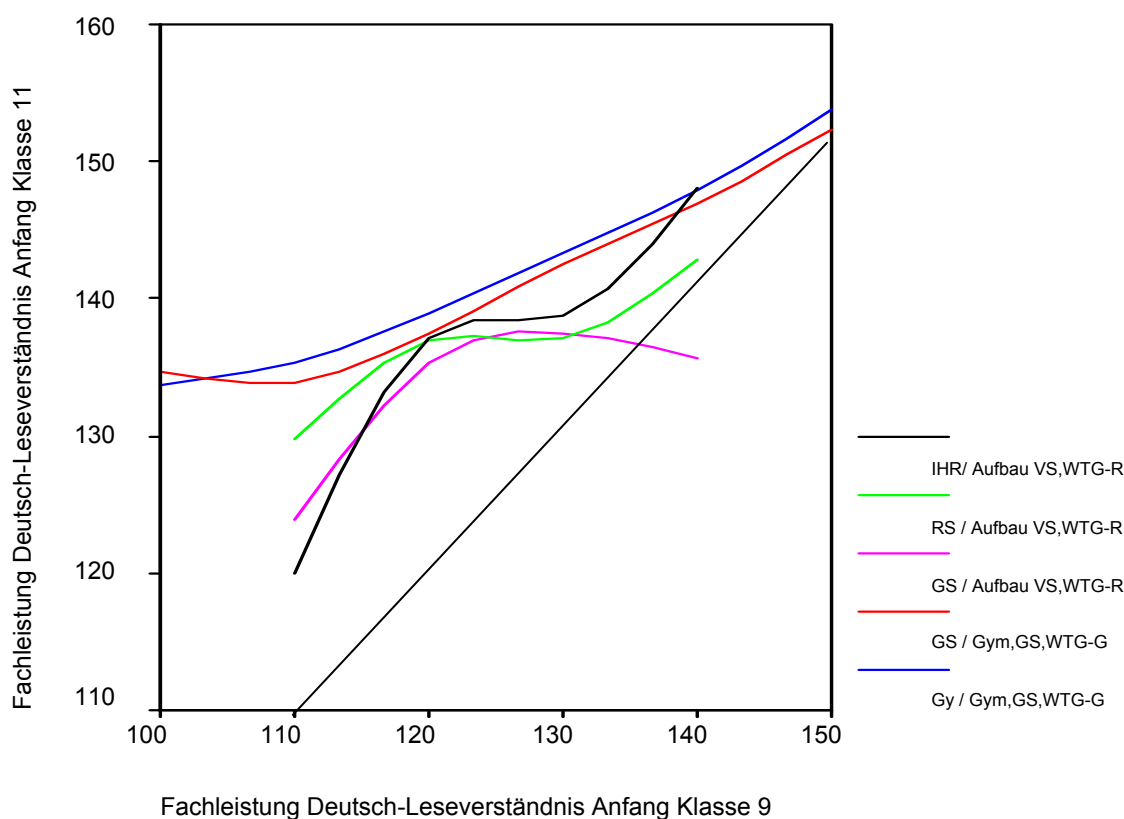
Angesichts der im Jahre 2000 beobachteten Lernstände schien es, als verlangsamten sich die Lernzuwächse im Bereich des Leseverständnisses, sobald in der mittleren Adoleszenz ein gewisses Niveau erreicht sei. Dies wurde durch die Beobachtung nahe gelegt, dass damals auf dem Hauptschulniveau die größten Lernzuwächse zu verzeichnen waren, während neben den Gymnasien auch für das Kursniveau I der Gesamtschulen eine niedrigere Wachstumsrate festgestellt wurde. Einen Sonderfall stellte das Kursniveau II der Gesamtschulen dar, auf dem die Lernentwicklung absolut und relativ gesehen durchschnittlich langsamer verlaufen ist als auf dem sonst zumeist vergleichbaren Hauptschulniveau. Auffällig waren insbesondere die äußerst selten auftretenden guten Testergebnisse (LEHMANN, PEEK, GÄNSFUß UND HUSFELDT 2002, S. 98). Offenbar waren zu Beginn der Klassenstufe 9 verhältnismäßig wenige Schülerinnen und Schüler im obersten Leistungsbereich anzutreffen, weil die Förderung und Entwicklung leistungsstärkerer Schülerinnen und Schüler eher schwach ausgeprägt war, während die Förderung von Jugendlichen im unteren und mittleren Leistungsbereich Priorität hatte. Diese Entwicklung hat sich im Verlauf der Klassenstufen 9 und 10 erheblich geändert. Für die Fachleistung Deutsch-Leseverständnis kann daher insgesamt ein Lernzuwachs diagnostiziert werden, der äußerst zufriedenstellend ist.

An dieser Stelle sollen die Lernstände zu Beginn der Klassenstufe 11 in Abhängigkeit von der Schulform und den dort gegebenen Lernständen zu Beginn der Klassenstufe 9 betrachtet werden (vgl. Abbildung 20). Auch hier gilt wieder, dass der Lernstand zu Beginn der Klassenstufe 11 umso höher ist, je günstiger die jeweilige *Leistungsverteilung* zu Beginn der Klassenstufe 9 war, und zwar im Vergleich der Schulformen.

So ist bei einer Schülerin oder einem Schüler mit einer bestimmten Lesekompetenz am Anfang der Klassenstufe 9 auch und gerade, wenn diese eher gering war, der Lernstand zu Beginn der Klassenstufe 11 generell am Gymnasium und an der Gesamtschule am günstigsten. Die Gesamtschülerinnen und Gesamtschüler mit Versetzung in die gymnasiale Oberstufe bleiben – bis auf den untersten Leistungsbereich – durchgängig etwa zwei Skalenpunkte unter der „Gymnasiallinie“. Die Realschülerinnen und Realschüler wie auch die Schülerinnen und Schüler aus den IHR-Schulen und die Gesamtschülerinnen und Gesamtschüler

mit einem Mittleren Schulabschluss verzeichnen im unteren Leistungsbereich recht hohe, dagegen im anteilstarken mittleren Leistungsbereich nur geringe Lernzuwächse. Doch während die IHR-Schülerinnen und -Schüler, vor allem aber die Realschülerinnen und Realschüler im oberen Leistungsbereich wieder günstigere Werte verzeichnen, flacht die Lernentwicklung der Gesamtschülerinnen und Gesamtschüler im oberen Leistungsbereich weiter ab.

Abbildung 20 Fachleistung Deutsch-Leseverständnis: Lernstände zu Beginn der Klassenstufe 11 nach Testleistung zu Beginn der Klassenstufe 9 und Schulform (Lernstandswerte)



Bei diesen Vergleichen ist darauf hinzuweisen, wie der scheinbare Widerspruch zu den in Tabelle 23 berichteten Befunden hinsichtlich der durchschnittlichen Lernzuwächse zustande kommt: Während Letztere die Ausgangsverteilung implizit durch Gewichtung der einzelnen Gruppen berücksichtigen, wird in der Abbildung 20 gerade von der Ausgangsverteilung abstrahiert.

Die Leistungen der einzelnen Bildungsganggruppen – bis auf die Gruppe der Gesamtschülerinnen und Gesamtschüler mit einem Mittleren Schulabschluss – scheinen im oberen Bereich zu konvergieren. Konnten zu Beginn der Klassenstufe 9 für die Schülerinnen und Schüler, die später den Übergang in die Sekundarstufe II bewältigt haben, 6,3 Prozent der Leistungen im Deutsch-Lese-

verständnisstest mit der Schulform in Zusammenhang gebracht werden, so werden für dieselben Jugendlichen zu Beginn der Klasse 11 nur noch 4,1 Prozent durch die Schulform<sup>22</sup> erklärt. Offenbar scheinen sich die Unterschiede zwischen den einzelnen Schulformen bzw. Kursniveaus tatsächlich zu verringern.

Ehe auf die Lernentwicklung auf der Ebene der Einzelschulen eingegangen wird, sollen die Lernstände zu Beginn der Klassenstufe 9 und zu Beginn der Klassenstufe 11, getrennt nach Realschulen (einschließlich IHR-Schulen), Gesamtschulen und Gymnasien, berichtet werden (vgl. Tabelle 24).

Tabelle 24 Fachleistung Deutsch-Leseverständnis zu Beginn der Klassenstufe 9 und zu Beginn der Klassenstufe 11 nach Schulform zu Beginn der Klassenstufe 9

	N	Beginn Klassenstufe 9		Beginn Klassenstufe 11		Lernzuwachs	
		$\bar{x}$	SD	$\bar{x}$	SD	Diffe- renz	Effekt- stärke d
<b>IHR- und Realschule</b>	243	122,7	16,3	136,2	16,7	13,5	0,83
<b>Gesamtschule</b>	684	128,0	16,8	142,7	15,1	14,7	0,88
<b>Gymnasium</b>	3.021	134,4	13,0	146,8	14,8	12,4	0,95
<b><i>insgesamt</i></b>	<b><i>3.948</i></b>	<b><i>132,6</i></b>	<b><i>14,4</i></b>	<b><i>145,5</i></b>	<b><i>15,2</i></b>	<b><i>12,9</i></b>	<b><i>0,9</i></b>

Im Bereich des Leseverständnisses sind die *absoluten* Lernzuwächse in den Gesamtschulen am größten, gefolgt von den Realschulen<sup>23</sup> und den Gymnasien. Hingegen liegen bei Betrachtung des *relativen* Leistungswachstums und bei Be-

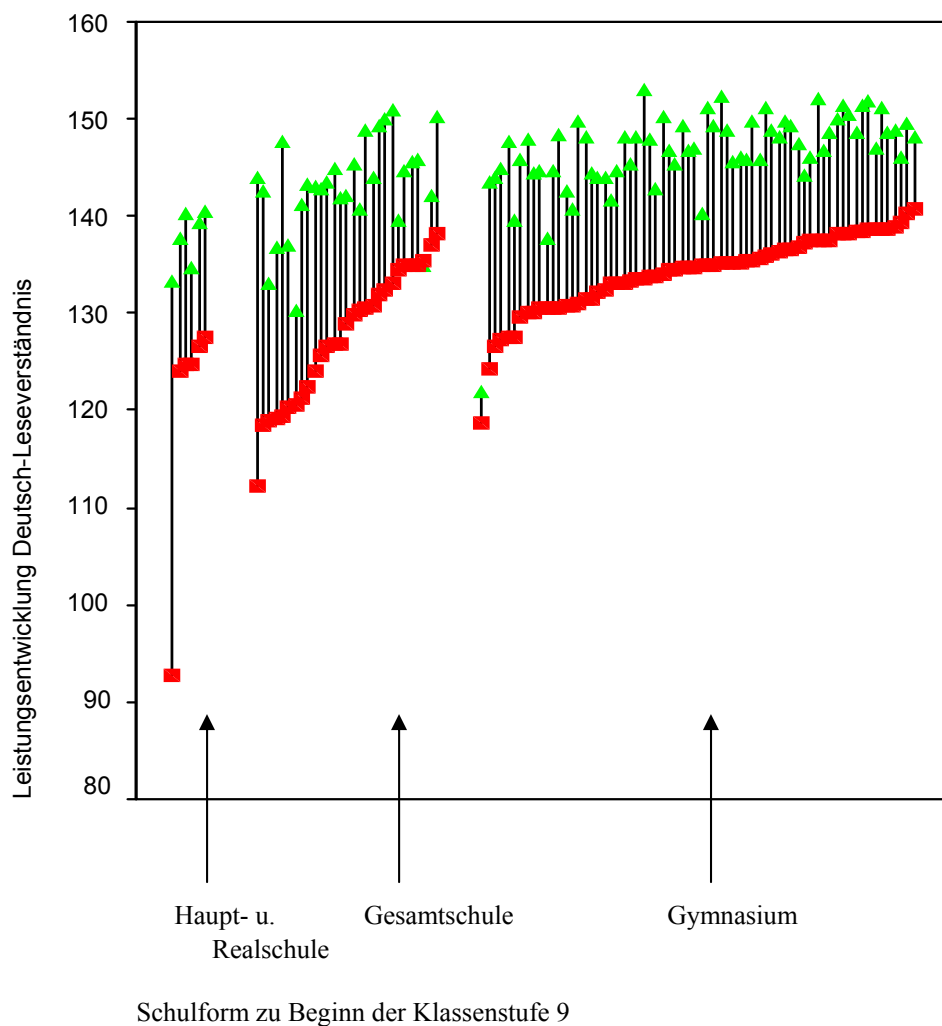
<sup>22</sup> Für die Schulform- und Kursniveauezugehörigkeit wurde – wie schon erwähnt – eine neue Variable gebildet, die sowohl die Schulformzugehörigkeit aus Klassenstufe 9 als auch die Testversion der Klassenstufe 11 berücksichtigt (vgl. die neun Gruppen in Tabelle 23). So bleiben Schülerinnen und Schüler unberücksichtigt, die sich in diese Struktur aufgrund eines Schulformwechsels nicht einfügen. Für die Klassenstufe 9 konnten 27 Prozent der Varianz der Leistung im Deutsch-Leseverständnisstest mit der normalen Schulformvariable erklärt werden. Berücksichtigt man hingegen nur diejenigen Schülerinnen und Schüler, die in die gymnasiale Oberstufe übergetreten sind und sowohl 2000 als auch 2002 getestet wurden, so können nur mehr 6,3 Prozent der Varianz der Deutschleistung in LAU 9 durch die neue Schulformvariable erklärt werden.

<sup>23</sup> Realschulen einschließlich IHR-Schulen.

rücksichtigung der Ausgangsvarianz die Gymnasien leicht vorn. Von Bedeutung ist diese Betrachtungsweise dann, wenn man die Einzelschule in den Blick nimmt.

In Abbildung 21 ist die Lernentwicklung im Bereich des Leseverständnisses innerhalb einzelner Schulen, getrennt nach Schulform und durchschnittlichem Lernstand, dargestellt. Durch die Übergangsauslese wie auch durch die teilweise hohe Zahl von Schülerinnen und Schülern, die zum Testzeitpunkt eine Schule im Ausland besuchten, sind einige Schulen nur noch mit sehr geringen Schülerzahlen vertreten. Dies gilt vor allem für die IHR- und Realschulen, aber auch für einzelne Gymnasien. In der Grafik werden nur diejenigen Schulen berücksichtigt, bei denen mindestens 10 Schülerinnen und Schüler in der gymnasialen Oberstufe angetroffen wurden.

Abbildung 21 Fachleistung Deutsch-Leseverständnis: Lernentwicklung vom Beginn der Klassenstufe 9 bis zum Beginn der Klassenstufe 11 nach Schulform und Einzelschule



Die Schulen werden hier jeweils separat innerhalb der eigenen Schulform miteinander verglichen; die schulspezifischen Lernausgangslagen entsprechen den untersten Punkten der Pfeile, den roten Quadraten. Die oberen grünen Dreiecke entsprechen dem durchschnittlich erzielten Lernstand zu Beginn der Klassenstufe 11. Die Länge der Pfeile repräsentiert den schulspezifischen Lernzuwachs. Dabei wird für den Bereich des Leseverständnisses deutlich, dass es in allen drei Schulformen einzelne Schulen gibt, die bis zu Beginn der Klassenstufe 11 im Mittel das Niveau nicht erreicht haben, das für andere die Lernausgangslage zu Beginn der Klassenstufe 9 darstellt.

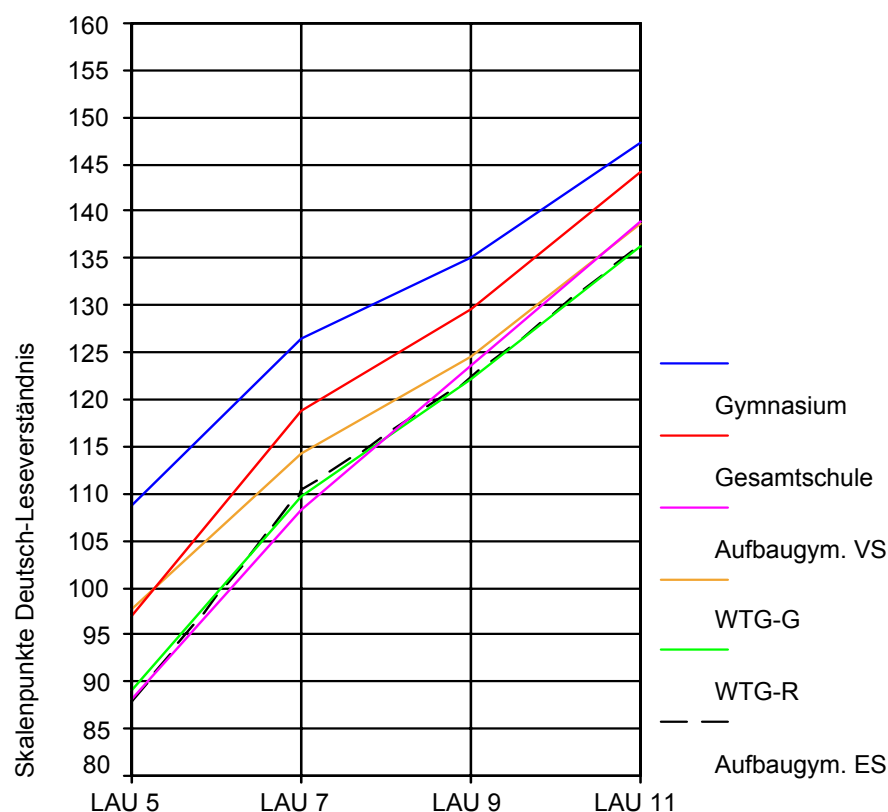
Für alle Schulformen legt die Grafik den Eindruck nahe, dass die Leistungszuwächse zwischen den Schulen zum Teil erheblich variieren, es hat den Anschein, dass insbesondere die Gymnasien im oberen Leistungsbereich im Mittel ein geringeres Leistungswachstum verzeichnen. Dieser Eindruck lässt sich auch quantitativ belegen: Der schulspezifische Zuwachs korreliert eindeutig negativ mit dem Schulmittelwert zu Beginn der Klassenstufe 9 ( $r = -0,68$ ). Das bedeutet: Je besser die Schülerinnen und Schüler einer Schule zu Beginn der Klassenstufe 9 in der Fachleistung Deutsch-Leseverständnis abgeschnitten hatten, desto ungünstiger fällt der Lernzuwachs in dieser Schule aus.

Damit haben sich insgesamt die Unterschiede zwischen den einzelnen Gymnasien verringert: Konnten zu Beginn der Klassenstufe 9 noch 8,0 Prozent der Varianz durch die Schulzugehörigkeit erklärt werden, so beträgt die Zwischenschulvarianz in der LAU 11 nur noch 5,5 Prozent an der Gesamtvarianz. Zwischen den Schulklassen hat sich der spezifische Varianzanteil an den Gymnasien in demselben Zeitraum von 16,0 Prozent auf 12,1 Prozent verringert. Im Hinblick auf das Leseverständnis verdankt sich der festgestellte Lernzuwachs in den Klassenstufen 9 und 10 am Gymnasium also allem Anschein nach zu einem gewissen Teil den kompensatorischen Leistungen von Schulen mit vergleichsweise niedrigen Lernständen zu Beginn des Betrachtungszeitraums. Denkbar ist, dass auch differenzielle Quoten für die Klassenwiederholung, den Schulformwechsel oder den Übergang in die Berufsbildung beteiligt sind. Entsprechende Untersuchungen müssen einem späteren Zeitpunkt vorbehalten bleiben.

Zum Abschluss dieses Unterkapitels zeigt Abbildung 22 die Lernzuwächse über den gesamten Zeitraum der Längsschnittuntersuchung von 1996 bis 2002. Dabei ist zu beachten, dass sich – wie gerade für die Gelenkstelle zwischen den Sekundarstufen I und II angesprochen – über den Untersuchungszeitraum von sechs Schuljahren Änderungen in der Zielgruppendefinition durch Klassenwiederholung, Schulform- und Kursniveauwechsel und nicht zuletzt Beendigung des Besuchs der allgemeinbildenden Schule ergeben haben, die rechnerisch

nicht ausgeglichen werden können. Deshalb wird an dieser Stelle die Entwicklung des Leseverständnisses im Deutschen für eine bestimmte Teilpopulation gleichsam retrospektiv, d. h. nach Maßgabe der für die Klassenstufe 11 gebildeten Bildungsgänge berechnet. Es wurden also nur solche Jugendliche berücksichtigt, die zu allen vier Testzeitpunkten der Längsschnittuntersuchung den Deutsch-Leseverständnistest bearbeitet haben. Insgesamt 3.422 Schülerinnen und Schüler haben sowohl in LAU 5 als auch in LAU 7, LAU 9 und LAU 11 am Deutsch-Leseverständnistest teilgenommen.

Abbildung 22 Fachleistung Deutsch-Leseverständnis: Lernentwicklung vom Beginn der Klassenstufe 5 bis zum Beginn der Klassenstufe 11 nach Bildungsgang der Klassenstufe 11 (nur Teilnehmer an allen vier Erhebungswellen)



Verzeichneten die grundständigen Gymnasiastinnen und Gymnasiasten in der LAU 5 noch einen Leistungsvorsprung von über elf Skalenpunkten vor den grundständigen Gesamtschülerinnen und Gesamtschülern, so waren es in der LAU 7 noch ca. acht, in der LAU 9 ungefähr fünf und in der LAU 11 nur noch annähernd drei Skalenpunkte. Damit bestätigt sich nochmals der bereits in LAU 7 und LAU 9 festgestellte Befund, dass die Hamburger Gymnasien, entge-

gen der ‚Systemlogik‘ des gegliederten Schulsystems, im Durchschnitt keineswegs eine besonders rasche Lernentwicklung ermöglichen.

Ähnliches gilt auch, vergleicht man Schülerinnen und Schüler der Gruppe WTG-G mit den grundständigen Gesamtschülerinnen und Gesamtschülern. Waren beide Gruppen von einem annähernd gleichen Leistungsstand von 97,0 (Gesamtschulen) bzw. 97,7 Skaleneinheiten (WTG-G) in der LAU 5 ausgegangen, so hatten im Jahr 1998 die Gesamtschülerinnen und Gesamtschüler bereits einen Vorsprung von 4,5 Punkten, der in den folgenden Untersuchungen noch leicht auf 5,5 Skaleneinheiten ausgebaut werden konnte. Dies zeigt, dass an den Gesamtschulen die Auslese derer, die in die gymnasiale Oberstufe versetzt werden, offenbar sehr eng mit den tatsächlichen Lernfortschritten, d. h. mit dem Lernpotenzial der betreffenden Schülerinnen und Schüler zusammenhängt.

Die Schülerinnen und Schüler aus dem Längsschnitt der LAU, die der Bildungsganggruppe WTG-R zugeordnet und die mit einem Mittleren Schulabschluss in die Vorstufe des Wirtschafts- oder technischen Gymnasiums eingetreten sind, und die Schülerinnen und Schüler aus der Gruppe Aufbaugymnasium ES, die mit einem Mittleren Schulabschluss das Vorbereitungsjahr in der Einführungsstufe absolvieren, haben sich über den gesamten Zeitraum der Längsschnittuntersuchung sehr ähnlich entwickelt, während die (zahlenmäßig kleine) Gruppe der Schülerinnen und Schüler, die mit einem Mittleren Schulabschluss unmittelbar in die Vorstufe des Aufbaugymnasiums eingetreten sind, zwischen den beiden Testzeitpunkten der LAU 7 und der LAU 11 eine etwas günstigere Lernentwicklung aufweist. Vergleicht man diese Schülerinnen und Schüler mit der (ebenfalls kleinen) Gruppe der Schülerinnen und Schüler aus der Bildungsganggruppe WTG-G, so haben sie einen Rückstand von anfänglich 9,6 Skaleneinheiten in den folgenden sechs Jahren vollständig aufholen können.

Bemerkenswert sind auch die unterschiedlichen Entwicklungsverläufe der Schülergruppen zwischen den verschiedenen Testzeitpunkten. Während die grundständigen Gymnasiastinnen und Gymnasiasten, die grundständigen Gesamtschülerinnen und Gesamtschüler wie auch die Schülerinnen und Schüler aus der Gruppe WTG-G den größten Lernzuwachs in den Klassenstufen 5 und 6, dann im Verlaufe der Klassenstufen 7 und 8 einen deutlich geringeren Lernzuwachs verzeichnen, um anschließend in den Klassenstufen 9 und 10 wieder zuzulegen, verläuft die Lernentwicklung der Gruppe Aufbaugymnasium VS auffallend kontinuierlich. Auch in den Gruppen WTG-R und Aufbaugymnasium ES verläuft die Lernentwicklung vergleichsweise stetig, jeweils mit dem höchsten Lernzuwachs in den Klassenstufen 9 und 10.



Die über den Gesamtzeitraum größten Lernfortschritte erzielen mit 48,6 bzw. 50,9 Skalenpunkten die Schülerinnen und Schüler, die entweder in die Einführungsstufe oder in die Vorstufe des Aufbaugymnasiums eingetreten sind, gefolgt von den Schülerinnen und Schülern aus der Gruppe WTG-R und den grundständigen Gesamtschülerinnen und Gesamtschülern mit einem Zuwachs von 47,2 bzw. 47,1 Skalenpunkten, während die Schülerinnen und Schüler aus der Gruppe WTG-G und die grundständigen Gymnasiastinnen und Gymnasiasten mit deutlichem Abstand die geringsten Lernfortschritte über den Gesamtzeitraum von sechs Jahren verzeichnen (40,9 bzw. 38,5 Skalenpunkte). Beträgt der mittlere Lernzuwachs der Gymnasiastinnen und Gymnasiasten rechnerisch 6,4 Skalenpunkte pro Schuljahr, erreicht er bei den Schülerinnen und Schülern, die unmittelbar in die Vorstufe des Aufbaugymnasiums eingetreten sind, 8,5 Skalenpunkte pro Schuljahr. Über die Mittelwerte, die Standardabweichungen sowie die Anteile der Schülerinnen und Schüler in den einzelnen Bildungsganggruppen gibt Tabelle 25 Aufschluss.

Tabelle 25 Fachleistungsentwicklung Deutsch-Leseverständnis: Lernstandswerte zu Beginn der Klassenstufen 5, 7, 9 und 11: Mittelwert, Standardabweichung und Anzahl der Schülerinnen und Schüler, die die Einzeltests bearbeitet haben, nach Bildungsgang der LAU 11

<b>Bildungsgang LAU 11</b>		<b>Lesen LAU 5</b>	<b>Lesen LAU 7</b>	<b>Lesen LAU 9</b>	<b>Lesen LAU 11</b>
<b>Gymnasium</b> N = 2.352	Mittelwert	108,8	126,5	135,0	147,3
	Standardabweichung	20,2	15,3	12,6	14,8
<b>Gesamtschule</b> N = 495	Mittelwert	97,0	118,7	129,7	144,1
	Standardabweichung	23,1	18,0	15,9	13,7
<b>WTG-G</b> N = 50	Mittelwert	97,7	114,2	124,6	138,6
	Standardabweichung	26,9	19,5	17,8	18,2
<b>WTG-R</b> N = 198	Mittelwert	89,2	109,7	122,2	136,4
	Standardabweichung	23,6	17,2	16,9	17,7
<b>Aufbaugymnasium VS</b> N = 54	Mittelwert	88,1	108,2	123,6	139,0
	Standardabweichung	21,5	17,2	15,8	15,4
<b>insgesamt</b> N = 3.149	Mittelwert	105,2	123,7	133,0	145,8
	Standardabweichung	22,0	16,8	14,1	15,2
<b>Aufbaugymnasium ES</b> N = 273	Mittelwert	87,9	110,3	122,3	136,5
	Standardabweichung	21,7	18,6	16,2	17,0

### 4.3 Entwicklung der Fachleistung Deutsch: Rechtschreiben

Wie bereits aufgezeigt wurde, lässt der Untertest *Deutsch-Rechtschreiben* ebenso wenig wie die anderen fachbezogenen Untertests einen Deckeneffekt erkennen, auch nicht für die Schülerinnen und Schüler an den Gymnasien. Angesichts des hier gegebenen Datenformats erschien eine probabilistische Skalierung der entsprechenden Testwerte nicht angezeigt. Die im Folgenden berichteten Lernzuwächse in diesem Untertest beziehen sich somit nicht – wie bei Deutsch-Leseverständnis – auf Skalenpunkte, sondern auf die Gesamtpunktzahl der gefundenen Rechtschreibfehler unter Berücksichtigung der fälschlich unterstrichenen Richtigschreibungen.

In Tabelle 26 sind die statistischen Kennwerte für das Leistungswachstum der einzelnen Bildungsganggruppen angegeben, wiederum mit der Maßgabe, dass nur diejenigen Schülerinnen und Schüler berücksichtigt werden können, für die Leistungsdaten aus beiden Erhebungswellen, 2000 und 2002, vorliegen. Für insgesamt 3.816 Schülerinnen und Schüler ist dies im Bereich Deutsch-Rechtschreiben der Fall.

Bei einem Gesamtzuwachs von 3,9 Lernstandspunkten bzw. einer Effektstärke von  $d = 0,74$  (also rechnerisch 0,37 Standardabweichungen pro Jahr) ist die Leistungsentwicklung im Bereich des passiven Rechtschreibenwissens etwas geringer ausgeprägt als im Fach Deutsch-Leseverständnis ( $d = 0,90$ ). Differenziert man zwischen den Schulformen, so zeigen sich ähnliche Relationen wie die schon zuvor beobachteten: Am höchsten sind – absolut betrachtet – die Zuwächse der grundständigen Gesamtschülerinnen und Gesamtschüler sowie der ehemaligen Realschülerinnen und Realschüler mit jeweils 4,3 Lernstandspunkten. Es folgen mit geringem Abstand die Schülerinnen und Schüler aus den IHR-Schulen und die grundständigen Gymnasiastinnen und Gymnasiasten (3,9 Punkte). Mit deutlicherem Abstand verzeichnen die Gesamtschülerinnen und Gesamtschüler mit einem Mittleren Schulabschluss den geringsten absoluten Lernzuwachs (3,1 Punkte).

Betrachtet man hingegen die relativen Lernzuwächse, die wieder durch die *Effektstärke*  $d$  angegeben werden und die die Ausgangsvarianz zu Beginn der Klassenstufe 9 mitberücksichtigen, so weisen die IHR-Schulen mit einer Effektstärke von mehr als einer Standardabweichung ( $d = 1,03$ ) die höchste Lernentwicklung auf, gefolgt von den Realschulen ( $d = 0,96$ ). Die Gesamtschülerinnen und Gesamtschüler (beide Gruppen) und die Gymnasiastinnen und Gymnasiasten liegen gleichauf mit einigem Abstand zurück.

Tabelle 26 Fachleistung Deutsch-Rechtschreiben: Lernstandswerte zu Beginn der Klassenstufe 9 und zu Beginn der Klassenstufe 11 sowie Lernzuwachs, nach Schulform der Klassenstufe 9

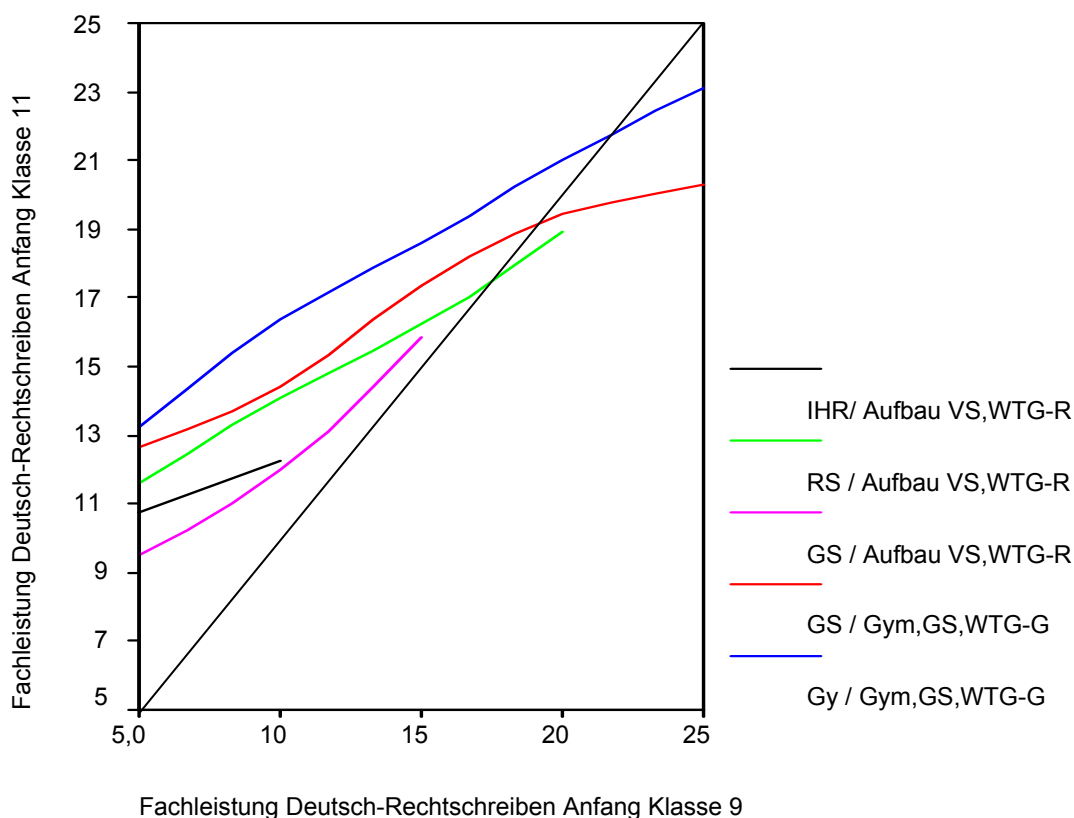
Schulform Klasse 9	Bildungsgang Klasse 11	Beginn Klasse 9			Beginn Klasse 11		Lernzuwachs		
		$\bar{x}$	SD	N	$\bar{x}$	SD	Differenz	<i>D</i>	Varianz- änderung
<b>IHR-Schule</b>	<i>Abgänge</i>	7,0	4,1	570	–	–	–	–	–
<b>IHR-Schule</b>	Aufbaugymn. VS, WTG-R	9,0	3,8	41	12,9	4,7	3,9	<b>1,03</b>	+55,9
<b>Realschule</b>	<i>Abgänge</i>	9,2	4,3	1.589	–	–	–	–	–
<b>Realschule</b>	Aufbaugymn. VS, WTG-R	9,7	4,5	195	14,0	4,6	<b>4,3</b>	<b>0,96</b>	+6,4
<b>Gesamt- schule</b>	<i>Abgänge</i>	7,3	4,4	2.410	–	–	–	–	–
<b>Gesamt- schule</b>	Aufbaugymn. VS, WTG-R	7,9	3,9	74	11,0	4,5	3,1	0,79	+37,1
<b>Gesamt- schule</b>	Gymn. GS, WTG-G	11,0	5,3	587	15,3	5,1	<b>4,3</b>	0,81	-6,3
<b>Gymnasium</b>	<i>Abgänge</i>	14,3	5,2	1.399	–	–	–	–	–
<b>Gymnasium</b>	Gymn., GS, WTG-G	14,6	5,0	2.919	18,5	4,6	3,9	0,78	-17,5
<i>insgesamt</i> <i>N = 9.784</i>		11,0	5,7		–	–	–	–	–
<i>insgesamt</i> <i>N = 3.816</i>		<b>13,6</b>	<b>5,3</b>		<b>17,5</b>	<b>5,0</b>	<b>3,9</b>	<b>0,74</b>	<b>-13,4</b>

Im Unterschied zum Deutsch-Leseverständnistest hat sich in diesem Deutsch-Untertest die Varianz insgesamt zwar verringert, bei Betrachtung der einzelnen Schulformen zeigt sich hingegen, dass die Varianzminderung auf die beiden zahlenmäßig stärksten Gruppen, die grundständigen Gymnasiastinnen und Gymnasiasten und die grundständigen Gesamtschülerinnen und Gesamtschüler, beschränkt ist. In den übrigen Gruppen hat die Streuung dagegen zugenommen: geringfügig bei den ehemaligen Realschülerinnen und Realschülern, erheblich demgegenüber bei den beiden zahlenmäßig kleinsten Gruppen der Gesamtschü-

lerinnen und Gesamtschüler mit einem Mittleren Schulabschluss und am stärksten bei den ehemaligen IHR-Schülerinnen und -Schülern, was auf eine sehr ungleichmäßige Leistungsnetwicklung hinweist. Insgesamt kann für die Fachleistung Deutsch-Rechtschreiben ein zufriedenstellender Lernzuwachs festgestellt werden.

An dieser Stelle ist es aufschlussreich, wiederum die Lernstände zu Beginn der Klassenstufe 11 in Abhängigkeit von der Schulform zu Beginn der Klassenstufe 9 und den dort gegebenen Lernständen zu betrachten (vgl. Abbildung 23). Die schwarze Diagonale repräsentiert wieder ein hypothetisches durchgängiges Nullwachstum und dient lediglich der optischen Orientierung.

Abbildung 23 Fachleistung Deutsch-Rechtschreiben: Lernstände zu Beginn der Klassenstufe 11 nach Testleistung zu Beginn der Klassenstufe 9 und Schulform (Lernstandswerte)



Die größten Lernzuwächse gibt es in allen Schulformen vor allem im unteren Leistungsbereich. Dabei ist bei gegebener Anfangsleistung der zu Beginn der Klassenstufe 11 erreichte Lernstand der grundständigen Gymnasiastinnen und Gymnasiasten durchgehend am höchsten. Anders als im Bereich des Leseverständnisses ist im Bereich des passiven Rechtschreibwissens keine Konvergenz im oberen Leistungsbereich zu erkennen. Innerhalb der Gesamtschulen sind bei

gleicher Anfangsleistung die Lernstände der in die gymnasiale Oberstufe versetzten Schülerinnen und Schüler zu Beginn der Klassenstufe 11 vor allem im unteren Leistungsbereich deutlich höher als die Lernstände ihrer Mitschülerinnen und Mitschüler, die mit einem Mittleren Schulabschluss in die Vorstufe des Aufbau-, Wirtschafts- oder technischen Gymnasiums eingetreten sind. Im mittleren und oberen Leistungsbereich ist über alle Gruppen hinweg kein Leistungswachstum mehr vorhanden.

Um eine angemessene Interpretation der Befunde für die Einzelschulen zu ermöglichen, seien wieder die schulformspezifischen Kennwerte ohne Differenzierung nach Bildungsganggruppen angeführt (vgl. Tabelle 27). Im Gegensatz zu Deutsch-Leseverständnis sind im Untertest Deutsch-Rechtschreiben sowohl die absoluten als auch die relativen Lernzuwächse für die Gymnasien am geringsten, allerdings sind die Unterschiede zwischen den Schulformen nicht sehr groß. Die Integrierten Haupt- und Realschulen sowie die Realschulen verzeichnen die größten absoluten, vor allem aber relativen Lernentwicklungsraten.

Tabelle 27 Fachleistung Deutsch-Rechtschreiben zu Beginn der Klassenstufe 9 und zu Beginn der Klassenstufe 11 nach Schulform der Klassenstufe 9

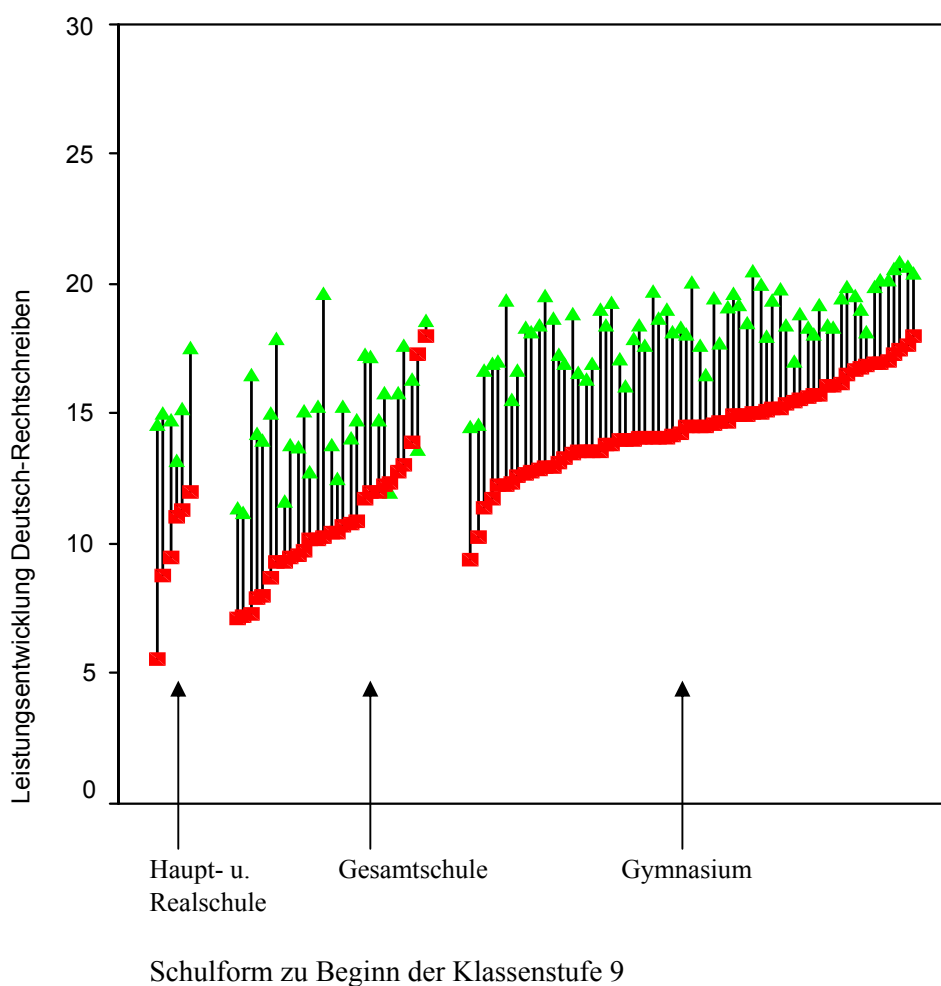
	Beginn Klassenstufe 9			Beginn Klassenstufe 11		Lernzuwachs	
	N	$\bar{x}$	SD	$\bar{x}$	SD	Differenz	Effektstärke $d$
<b>IHR- und Realschule</b>	236	9,5	4,3	13,8	4,6	4,3	1,00
<b>Gesamtschule</b>	661	10,6	5,2	14,8	5,2	4,2	0,81
<b>Gymnasium</b>	2.919	14,6	5,0	18,5	4,6	3,9	0,78
<b><i>insgesamt</i></b>	<b><i>3.816</i></b>	<b><i>13,6</i></b>	<b><i>5,3</i></b>	<b><i>17,5</i></b>	<b><i>5,0</i></b>	<b><i>3,9</i></b>	<b><i>0,74</i></b>

Bei der Differenzierung nach Schulformen und Einzelschulen, geordnet nach dem durchschnittlichen Lernstand zu Beginn der Klassenstufe 9 (vgl. Abbildung 24), sind ähnliche Zusammenhänge erkennbar wie im Deutsch-Leseverständnis<sup>24</sup>. Die Pfeillänge kennzeichnet wieder den Lernzuwachs ausgehend vom

<sup>24</sup> Es wurden wieder nur die Schulen in die Grafik einbezogen, die mit mindestens zehn Schülerinnen und Schülern in der Sekundarstufe II vertreten sind.

Lernstand zu Beginn der Klassenstufe 9 (rote Quadrate) bis zum Beginn der Klassenstufe 11 (grüne Pfeilspitze). Die Lernausgangslagen unterscheiden sich zum Teil erheblich. Ebenso sind die Unterschiede zu Beginn der Klassenstufe 11 sehr groß. Wiederum weisen einzelne Schulen zu diesem Zeitpunkt Lernstände auf, die von anderen Schulen – schulformübergreifend, teilweise aber auch innerhalb derselben Schulform – bereits zu Beginn der Klassenstufe 9 erreicht wurden. Bei den Gymnasien ist – deutlicher als bei Deutsch-Leseverständnis – zu erkennen, dass Schulen mit vergleichsweise besserer Ausgangsposition in der neunten Klasse geringere Lernzuwächse nachweisen als solche, die zu Beginn der Klassenstufe 9 von einer ungünstigeren Position ausgegangen sind. Dieser optische Eindruck lässt sich dadurch bestätigen, dass der schulspezifische Zuwachs eindeutig negativ mit dem Schulmittelwert zu Beginn der Klassenstufe 9 (*Pearsons*  $r = -0,42$ ) korreliert. Je besser die Schülerschaft einer Schule zu Beginn der Klassenstufe 9 in der Fachleistung Deutsch-Rechtschreiben war, desto geringer ist das Leistungswachstum dieser Schule für die Klassenstufe 9 und 10.

Abbildung 24 Fachleistung Deutsch-Rechtschreiben: Lernentwicklung vom Beginn der Klassenstufe 9 bis zum Beginn der Klassenstufe 11 nach Schulform und Einzelschule



Insgesamt haben sich aber auch in diesem Deutsch-Untertest die Unterschiede zwischen den einzelnen Gymnasien im Vergleich zu LAU 9 etwas verringert: Betrachtet man wieder nur die Schülerinnen und Schüler, die sowohl an der LAU 9 als auch an der LAU 11 teilgenommen haben, so erklärte die Schulzugehörigkeit in der Klassenstufe 9 noch 10,3 Prozent der Varianz bei den Gymnasiastinnen und Gymnasiasten. In Klassenstufe 11 hingegen beträgt die Zwischenschulvarianz nur noch 7,6 Prozent an der Gesamtvarianz. Der bezeichnende Varianzanteil an den Gymnasien zwischen den Schulklassen hat sich von 18,1 Prozent auf 15,1 Prozent in der elften Klasse vermindert. Der festgestellte Lernzuwachs in den Klassenstufen 9 und 10 in der Fachleistung Deutsch-Rechtschreiben wurde möglicherweise auch hier zu einem gewissen Teil durch die kompensatorischen Leistungen solcher Gymnasien bewirkt, die zu Beginn des Betrachtungszeitraums vergleichsweise niedrige Lernstände zeigten.

Uneinheitlicher ist das Bild bei den Gesamtschulen. Zwar verzeichnen auch hier Schulen mit eher ungünstigen Ausgangslagen die höchsten Lernzuwächse und die drei Schulen mit den günstigsten Ausgangslagen weisen die mit Abstand ungünstigsten Lernentwicklungen auf (in einer Schule haben die Rechtschreibleistungen sogar eine negative Entwicklung erfahren). Doch finden sich auch im unteren Bereich Schulen mit auffallend geringen Lernzuwächsen. Insgesamt haben sich auch hier die Unterschiede zwischen den einzelnen Schulen verringert. Betrug in Klassenstufe 9 die Varianz zwischen den Schulen noch 16,8 Prozent an der Gesamtstreuung, so waren es zu Beginn der Klassenstufe 11 nur noch 13,7 Prozent. Der Varianzanteil zwischen den einzelnen Klassen ist aber hier erheblich höher, wenn auch eine Verringerung von 33,1 Prozent in der Klassenstufe 9 auf 30,4 Prozent an der Gesamtstreuung in Klassenstufe 11 festzustellen ist.

Die wenigen Realschulen, die mit mindestens zehn Schülerinnen und Schülern in der Untersuchung vertreten sind, können eine Annäherung der Leistungsstände bei recht hohen Wachstumsraten verzeichnen. Bei den IHR-Schulen und den Realschulen erklärte die Schulzugehörigkeit in der Klassenstufe 9 noch 21,1 Prozent der Varianz, in Klassenstufe 11 hingegen beträgt die Zwischenschulvarianz nur noch 18,9 Prozent. Der bezeichnende Varianzanteil an den IHR-Schulen und den Realschulen zwischen den Schulklassen hat sich demgegenüber von 28,2 Prozent auf 28,8 Prozent in der Klassenstufe 11 erhöht. Auch hier scheint es – ähnlich wie bei den Gesamtschulen – in den einzelnen Schulklassen noch erhebliche Leistungsunterschiede zu geben.

Lernzuwachsrechnungen für den Gesamtzeitraum der Längsschnittuntersuchung sind für den Untertest Deutsch-Rechtschreiben nicht möglich, da dieser Test erstmals im Jahr 2000, also in Klassenstufe 9 eingesetzt wurde.

## 4.4 Entwicklung der Fachleistung Mathematik

Wie für die Fachleistung Deutsch soll auch für die Fachleistung Mathematik die Lernentwicklung der Schülerinnen und Schüler aus dem Längsschnitt vom Beginn der Klassenstufe 9 bis zum Beginn der Klassenstufe 11 untersucht werden. In Tabelle 28 sind die statistischen Kennwerte für den Untertest Mathematik I zu Beginn der Klassenstufe 9 und zu Beginn der Klassenstufe 11 sowie die daraus gebildeten Differenzwerte aufgeführt, die für die einzelnen Schulformen die durchschnittlichen Lernfortschritte wiedergeben.

In der vorletzten Zeile der Tabelle 28 ist die Gesamtschülerzahl für die in Klassenstufe 9 Getesteten angegeben ( $N = 9.514$ ). Diese Schülerzahl weicht wieder von der vorher berichteten ( $N = 12.620$ ) ab, weil die aus den oben dargestellten Gründen ausgeschiedenen Schülerinnen und Schüler in dieser Tabelle unberücksichtigt bleiben müssen. Der letzten Zeile der Tabelle ist zu entnehmen, dass es in den beiden Schuljahren zwischen den Erhebungen von 2000 und 2002 im Durchschnitt über alle Schulformen hinweg einen Lernzuwachs gegeben hat, der mit 0,43 Standardabweichungen, d. h. rechnerisch 0,21 Standardabweichungen pro Jahr, unter dem der Klassenstufen 7 und 8 ( $d = 0,27/\text{Jahr}$ ) liegt und im Übrigen nur etwa der Hälfte des Zuwachses entspricht, der für den Untertest Deutsch-Leseverständnis zu beobachten war.

Allerdings verteilen sich die Zuwächse ungleich auf die verschiedenen Gruppen. Vergleicht man die Werte zu Beginn der Klassenstufe 9 mit denen zu Beginn der Klasse 11 auf Basis der ehemaligen Schulform, so erkennt man das größte Wachstum in den Realschulen und in den Gesamtschulen. Sowohl absolut als auch relativ sind die Lernfortschritte in diesen beiden Schulformen am größten. Um einen Eindruck davon zu vermitteln, wie der Leistungsstand derjenigen Jugendlichen zu Beginn der Klassenstufe 9 ist, die nicht in die gymnasiale Oberstufe übergegangen sind, werden wieder die statistischen Kennwerte der Klassenstufe 9 auch für diese Gruppen angegeben.

Des Weiteren ist Tabelle 28 zu entnehmen, dass sich die Streuung der Mathematikleistungen der Gesamtschülerinnen und Gesamtschüler, die in die gymnasiale Oberstufe versetzt worden sind, sowie der (kleinen) Gruppe der ehemaligen IHR-Schülerinnen und -Schüler verringert hat, während sie sich insbesondere in den Gymnasien, aber auch in der Gruppe der Gesamtschülerinnen und Gesamtschüler mit einem Mittleren Schulabschluss und in den Realschulen erhöht hat. Letzteres mag zunächst wieder überraschen, hatte sich doch die Streuung in den Klassenstufen 5 bis 8 kontinuierlich vermindert, und zwar von  $SD = 24,1$  auf  $SD = 19,6$  Punkte, d. h. um 33,6 Prozent oder mehr als ein Drittel der Anfangsvarianz (vgl. LEHMANN, PEEK, GÄNSFUS und HUSFELDT 2002, S. 87).



Tabelle 28 Fachleistung Mathematik I: Lernstandswerte zu Beginn der Klassenstufe 9 und zu Beginn der Klassenstufe 11 sowie Lernzuwachs, nach Schulform der Klassenstufe 9

Schulform Klasse 9	Bildungsgang Klasse 11	Beginn Klasse 9			Beginn Klasse 11		Lernzuwachs		
		$\bar{x}$	SD	N	$\bar{x}$	SD	Differenz	$d$	Varianz- änderung
IHR-Schule	<i>Abgänge</i>	112,2	15,2	551	–	–	–	–	–
IHR-Schule	Aufbaugymn. VS, WTG-R	127,9	14,0	39	134,7	13,3	6,8	0,49	-9,6 %
Realschule	<i>Abgänge</i>	120,6	14,1	1.551	–	–	–	–	–
Realschule	Aufbaugymn. VS, WTG-R	124,8	14,8	193	134,2	15,3	<b>9,4</b>	<b>0,64</b>	+7,3 %
Gesamt- schule	<i>Abgänge</i>	112,5	15,7	2.298	–	–	–	–	–
Gesamt- schule	Aufbaugymn. VS, WTG-R	122,2	15,4	72	128,9	16,2	6,7	0,44	+10,0 %
Gesamt- schule	Gymn., GS, WTG-G	128,9	15,8	558	137,8	14,9	<b>8,9</b>	<b>0,56</b>	-11,5 %
Gymnasium	<i>Abgänge</i>	138,1	14,9	1.351	–	–	–	–	–
Gymnasium	Gymn., GS, WTG-G	140,1	14,2	2.901	146,1	15,5	6,0	0,42	+20,1 %
<i>insgesamt</i> <i>N = 9.514</i>		127,2	19,0		–	–	–	–	–
<i>insgesamt</i> <i>N = 3.763</i>		<b>137,2</b>	<b>15,5</b>		<b>143,8</b>	<b>16,0</b>	<b>6,6</b>	<b>0,43</b>	<b>+7,2 %</b>

Berücksichtigt man jedoch die Tatsache, dass sich die Ausgangsvarianz der Klassenstufe 9 in Tabelle 28 nur auf diejenigen bezieht, die den Übergang in die gymnasiale Oberstufe bewältigt haben, und geht man davon aus, dass die Leistungsanforderungen und damit bei manchen auch die Lernmotivation in den Klassenstufen 9 und 10 erheblich angestiegen sind, damit der Übertritt in die gymnasiale Oberstufe gelingt, dann wird erklärbar, dass die Individualvarianz in dieser Phase wieder ansteigt. Nachdem in keiner der eingesetzten Testversionen und in keiner der untersuchten Schulformen ein ausgeprägter Decken- oder Bo-

deneffekt festzustellen war, insbesondere auch nicht in der gymnasialen Version, ist davon auszugehen, dass hinter diesen Werten tatsächlich eine Spreizung der Leistungsverteilung steht.

Insoweit, als sich die Gymnasiastinnen und Gymnasiasten sowohl absolut als auch relativ in ihrer Lernentwicklung (deutlich) von den anderen Gruppen abheben, ist zu prüfen, ob hier nicht ein Artefakt vorliegt. Wie bereits angedeutet, kann für die grundständigen Gymnasiastinnen und Gymnasiasten ein Deckeneffekt ausgeschlossen werden. Lernstandswerte über 185 wurden bei einem Testumfang von 40 Items<sup>25</sup> bereits mit 35 (oder 87,5 Prozent) korrekt gelösten Aufgaben erzielt; der Anteil der Schülerinnen und Schüler in diesem Bereich besonders guter Fachleistungen umfasst an den Gymnasien jedoch nur 1,7 Prozent. Auch zu knapp bemessene Testzeit käme für die Erklärung der Seltenheit besonders hoher Lernstandswerte unter den Gymnasiastinnen und Gymnasiasten in Betracht. Nur je eines der selten korrekt gelösten Items befindet sich jedoch in der Nähe des jeweiligen Testendes (Item Nr. 38 in Form A, Item Nr. 37 in Form B), so dass die Anwesenheit einer unbeabsichtigten „Speed“-Komponente nahezu ausgeschlossen werden kann. Eine dritte denkbare Erklärung für die unerwartet niedrigen Lernfortschritte könnte in der mangelnden curricularen Einschlägigkeit einiger Testaufgaben in der Gymnasialversion liegen. Die Lösungshäufigkeiten liegen indessen nur bei zwei der 40 Aufgaben der Gymnasialversion unterhalb der Ratewahrscheinlichkeit. Dabei handelt es sich um eine Geometrie- und eine Algebraaufgabe. Charakteristisch für diese Items ist weniger ein ungewohnter als vielmehr ein besonders anspruchsvoller Inhalt, beispielsweise die Frage nach derjenigen linearen Funktion, die mit dem vorgegebenen Grafen der quadratischen Funktion  $f$  genau zwei Schnittpunkte bildet. Diese Frage (Aufgabe 33 in der Form A und Aufgabe 23 in der Form B der Testversion I) ist nur durch genaue Kenntnisse über quadratische Funktionen zu lösen. Um die Aufgabe grafisch zu bewältigen, muss man wissen, unter welchen Bedingungen sich eine quadratische Funktion staucht bzw. streckt und auf den Achsen verschiebt. Möglicherweise ist eine zusätzliche Zeichnung erforderlich. Alternativ kommt eine algebraische Lösung in Betracht, vorausgesetzt, die Beziehung zwischen der Frage nach „Schnittpunkten“ und der Ableitung einer Gleichung für die Lösungsmenge aus den beiden Funktionen ist verstanden. Aufgabe 38 wiederum (Testversion I, Form A; entspricht Aufgabe 28 in Form B) ist vermutlich deshalb so schwierig, weil zur Lösung der Aufgabe nicht nur der Satz des Pythagoras herangezogen werden muss, sondern zusätzlich ein sicherer Umgang mit Potenzregeln erforderlich ist.

---

<sup>25</sup> Eine Aufgabe wurde aus teststatistischen Gründen nicht mit in die Analyse einbezogen (Aufgabe 35 in der Form A der Testversion I).

Für die Erklärung der vergleichsweise geringen Lernzuwächse für die Gymnasiastinnen und Gymnasiasten käme prinzipiell auch ein Argument in Frage, das sich auf die Verzahnung der Testversionen von 2000 und 2002 bezieht. Wenn gezeigt werden könnte, dass die Ankeritems für die in den Klassenstufen 9 und 10 erfolgten Lernprozesse nicht wachstumssensitiv sind, also bei der zweiten Erhebung gleiche oder sogar geringere Lösungswahrscheinlichkeiten aufweisen als bei der ersten, wäre die neue Gesamtmessung angreifbar. Insbesondere käme es zu einer Unterschätzung der auf der Grundlage des neuen Tests ermittelten Lernstände.

Sorgfältige Überprüfungen haben jedoch ergeben, dass dies nur für eines der zehn Ankeritems zutrifft. Die durchschnittliche Lösungswahrscheinlichkeit dieser Aufgaben wurde – für die Schülerinnen und Schüler, die in LAU 9 und 11 getestet wurden – in der LAU 9 mit  $p = 0,61$  berechnet, während sie in LAU 11  $p = 0,69$  betrug. Aus der Differenz dieser Werte sowie der Standardabweichung für die Ankeritems zum ersten Messzeitpunkt ergibt sich ein Wert für den Lernzuwachs im Bereich der Ankeritems in den Klassenstufen 9 und 10 von insgesamt  $d = 0,17$ . Betrachtet man die Gruppe der Gymnasiastinnen und Gymnasiasten separat, so zeigt sich ein ähnliches Bild: Die durchschnittliche Lösungshäufigkeit der Ankeraufgaben, bezogen auf diese Gruppe, konnte für LAU 9 mit  $p = 0,65$  und für LAU 11 mit  $p = 0,73$  bestimmt werden. Der daraus resultierende Lernzuwachs in den Klassenstufen 9 und 10 ist mit einer Effektstärke von  $d = 0,17$  modellkonform genauso groß wie für alle Schülerinnen und Schüler zusammen.

Unter der Perspektive individueller Lernfortschritte scheinen in den Klassenstufen 9 und 10 die Leistungen im Fach Mathematik zwischen dem Gymnasium und den anderen Schulformen zu konvergieren (vgl. Abbildung 25, S. 101). Wie bei solchen Analysen schon häufiger beobachtet, profitieren vor allem Schülerinnen und Schüler mit ungünstiger Lernausgangslage von dem durch höhere Anforderungen charakterisierten Unterricht an den Gymnasien in unerwartet hohem Maße.

Ebenso unerwartet mag es sein, dass bei den Schülerinnen und Schülern mit vergleichsweise günstigeren Ausgangsleistungen kaum noch Unterschiede zwischen den Herkunftsbildungsgängen bestehen. Bei den Schülerinnen und Schülern mit sehr guten Ausgangsbedingungen (nur noch Gesamtschülerinnen und Gesamtschüler sowie Gymnasiastinnen und Gymnasiasten) sind ebenfalls keine Unterschiede mehr sichtbar. Die leistungsstärksten Jugendlichen sind indessen nur noch am Gymnasium anzutreffen.

Dies alles spricht – aus der Warte des Individuums – für die Durchlässigkeit des Hamburger Schulsystems –, hier insbesondere beim Übergang in die gymnasiale Oberstufe. Diese Aussage wird dadurch gestützt, dass sich die mit der Schulformzugehörigkeit verbundene Varianz in den Klassenstufen 9 und 10 von  $Eta^2 = 0,13$  (Klassenstufe 9) auf  $Eta^2 = 0,08$  (Klassenstufe 11) verringert hat<sup>26</sup>. Tatsächlich haben sich in den mit der LAU 11 erfassten Gruppen die Leistungen der Schülerinnen und Schüler aus unterschiedlichen Schulformen in den Klassenstufen 9 und 10 angenähert, allerdings vor allem dadurch, dass sich die durchschnittlichen Lernfortschritte an den grundständigen Gymnasien gegenüber den vorangehenden Klassenstufen nochmals verringert haben. Es scheint eine Tendenz an den grundständigen Gymnasien vor Beginn der Klassenstufe 11 zu geben, darauf zu verzichten, das Lerntempo der Schülerinnen und Schüler – insbesondere solcher im mittleren und oberen Leistungsbereich – entsprechend ihren Möglichkeiten zu forcieren oder auch nur aufrechtzuerhalten.

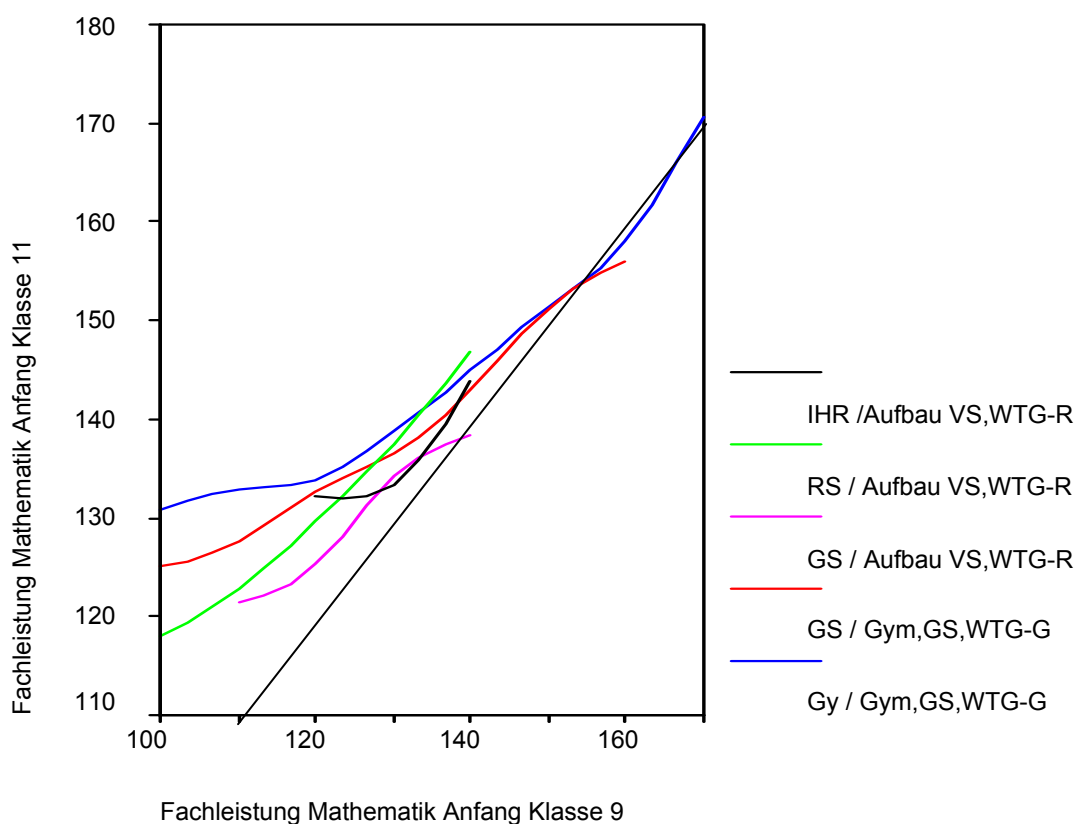
Gleichwohl bleibt zu fragen, wie die – gemessen an der Lernausgangslage – hohen *individuellen* Lernzuwächse der Gymnasiastinnen und Gymnasiasten aus dem unteren Leistungsbereich mit dem vergleichsweise niedrigen durchschnittlichen Lernfortschritt dieser Schulform in Einklang zu bringen sind. Die Antwort liegt (wie erstmals im Bericht zur LAU 7, vgl. LEHMANN, GÄNSFUß & PEEK, 1999, S. 87 ff. dargelegt) im Verhältnis der Anfangs- zur Endverteilung begründet. Gerade im mittleren und höheren Teil des Leistungsspektrums ist die erwartete Überlegenheit des Gymnasiums nicht nachzuweisen.

Abbildung 25 stellt diese Zusammenhänge grafisch dar. Die eingezeichnete schwarze Gerade repräsentiert zur Orientierung wieder eine *fiktive* Lerngruppe, die durchgängig die gleichen Leistungswerte in der neunten und der elften Klassenstufe, also keine Lernfortschritte zeigt.

---

<sup>26</sup> Auch hier wird noch einmal darauf hingewiesen, dass für die Schulformzugehörigkeit eine neue Variable gebildet wurde, die sowohl die Schulformzugehörigkeit aus Klassenstufe 9 als auch die Bildungsganggruppe der Klassenstufe 11 berücksichtigt. Für Klassenstufe 9 konnten 41 Prozent der Varianz der Mathematikleistung mit der normalen Schulformvariable für Klassenstufe 9 erklärt werden. Berücksichtigt man hingegen nur diejenigen Schülerinnen und Schüler, die sowohl in der LAU 9 als auch in der LAU 11 am Mathematiktest teilgenommen haben, so können 13 Prozent der Varianz der Mathematikleistung in LAU 9 durch die neue Schulform-Variable erklärt werden.

Abbildung 25 Fachleistung Mathematik: Lernstände zu Beginn der Klassenstufe 11 nach Testleistung zu Beginn der Klassenstufe 9 und Schulform (Lernstandswerte)



Indessen lässt die allgemeine Erhöhung der Leistungsvarianz am Gymnasium, die oben hervorgehoben wurde, angesichts verminderter Zwischen-Schul- und Zwischen-Klassen-Varianzen darauf schließen, dass *innerhalb der Schulklassen* in den Jahrgangsstufen 9 und 10 die Streuung zugenommen haben muss. Trotz der durchaus erfolgreichen kompensatorischen Bemühungen im unteren Leistungsbereich müssen sich also angesichts steigender Leistungsanforderungen am Gymnasium in der neunten und zehnten Klassenstufe Leistungsdifferenzen in dem Maße verstärken, wie sich der mittlere und obere Leistungsbereich zahlenmäßig verstärkt.

Bei der Zusammenfassung der Differenzierungsniveaus zeigt sich in den absoluten Mittelwertdifferenzen, dass die Lernentwicklung in den Klassenstufen 9 und 10 an den Realschulen (einschließlich IHR-Schulen) insgesamt etwas günstiger verlaufen ist als an den Gesamtschulen und, deutlicher, als an den Gymnasien. Betrachtet man die relativen Lernzuwächse, so hat sich an den auffallend niedrigen Lernzuwächsen an den Gymnasien nichts geändert.

Tabelle 29 Fachleistung Mathematik zu Beginn der Klassenstufe 9 und zu Beginn der Klassenstufe 11 nach Schulform der Klassenstufe 9

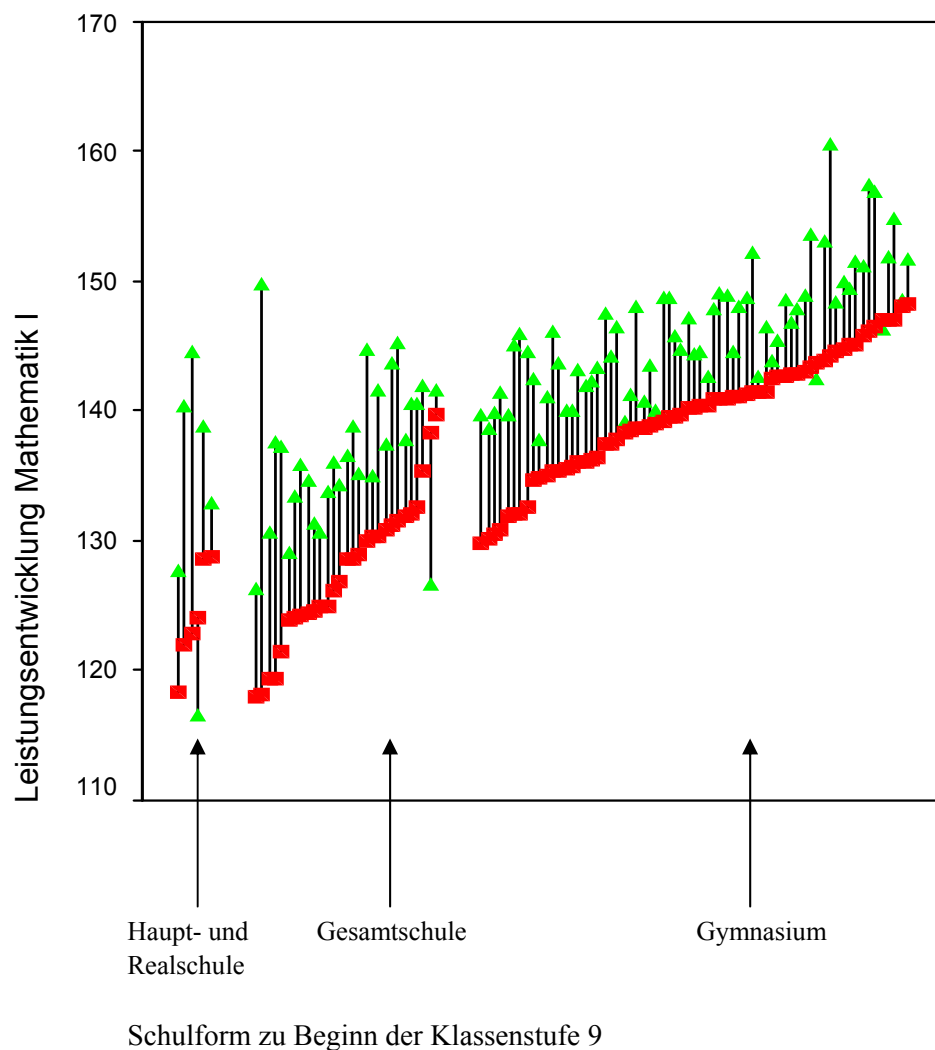
	Beginn Klassenstufe 9			Beginn Klassenstufe 11		Lernzuwachs	
	N	$\bar{x}$	SD	$\bar{x}$	SD	Diffe- renz	Effekt- stärke $d$
<b>IHR und Realschule</b>	232	125,3	14,6	134,3	14,9	9,0	0,62
<b>Gesamtschule</b>	630	128,1	15,9	136,8	15,3	8,7	0,55
<b>Gymnasium</b>	2.901	140,1	14,2	146,1	15,5	6,0	0,42
<b>insgesamt</b>	<b>3.763</b>	<b>137,2</b>	<b>15,5</b>	<b>143,8</b>	<b>16,0</b>	<b>6,6</b>	<b>0,43</b>

Zum Abschluss ist auch für die Fachleistung Mathematik zu prüfen, wie sich die durchschnittlichen Lernzuwächse zwischen den Einzelschulen unterscheiden. Es war gezeigt worden, dass es in dieser Hinsicht große Unterschiede zwischen den Realschulen einerseits und den Gymnasien andererseits gibt. Aber auch die Analyse auf Schulebene liefert interessante Aufschlüsse, wenn auch wiederum nur diejenigen Schülerinnen und Schüler berücksichtigt werden können, für die zu beiden Messzeitpunkten Lernstandswerte zur Verfügung stehen.

Bei der Untersuchung der Lernstände und der Zuwächse auf Einzelschulebene werden freilich die großen Unterschiede sichtbar, die auch in den früheren Erhebungen zu erkennen waren (vgl. Abbildung 26). Diese Abbildung ist genauso zu lesen wie die analogen Grafiken der Untertests *Deutsch: Leseverständnis* und *Deutsch: Rechtschreiben*. Die Lernfortschritte sind wieder daran zu erkennen, dass jeweils das Quadrat am unteren Ende eines ‚Pfeils‘ den durchschnittlichen Lernstand an einer Schule zu Beginn der Klassenstufe 9 als Lernausgangslage markiert. Die Pfeilspitze bezeichnet den zugehörigen Lernstand zu Beginn der Klassenstufe 11. Die Länge des Pfeils repräsentiert somit den jeweiligen Lernzuwachs einer Schule. Die verschiedenen Schulformen sind – da sie in der Grafik voneinander abgesetzt sind – deutlich voneinander zu unterscheiden. Innerhalb der jeweiligen Schulformen sind die Einzelschulen entsprechend ihrer Lernausgangslage geordnet. Dabei wird – wie schon für die Klassenstufen 7 und 8 und für die anderen Leistungsbereiche – sichtbar, dass in allen Schulformen der Fall auftritt, dass der durchschnittliche Lernstand einer Einzelschule zu Beginn der Klassenstufe 11 noch unter der Lernausgangslage einer anderen Schule

zu Beginn der Klassenstufe 9 liegt (vgl. analog LEHMANN, PEEK, GÄNSFUß und HUSFELDT 2002, S. 92). Erneut ist zu beachten, dass solche Schülerinnen und Schüler, die während der neunten oder zehnten Klasse die Schulform oder das Kursniveau gewechselt haben bzw. nicht in eine gymnasiale Oberstufe übergewechselt sind, in dieser Grafik unberücksichtigt bleiben.

Abbildung 26 Fachleistung Mathematik: Lernentwicklung vom Beginn der Klassenstufe 9 bis zum Beginn der Klassenstufe 11 nach Schulform und Einzelschule

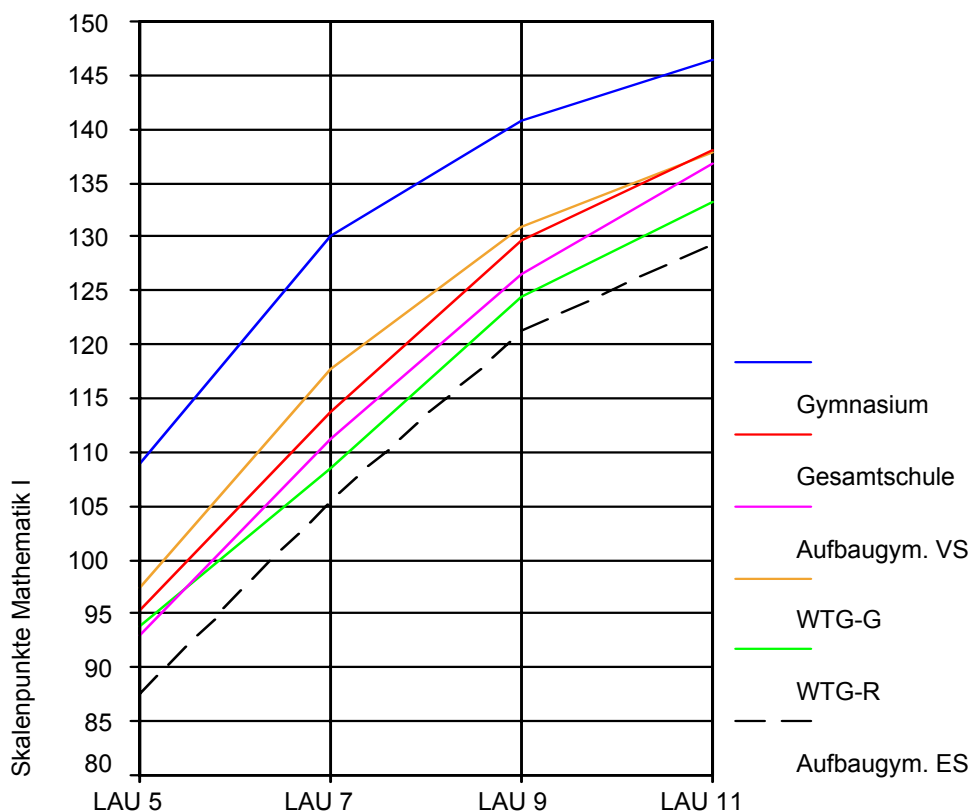


Anders als an den Integrierten Haupt- und Realschulen und den Gesamtschulen stellt sich das Problem der äußeren Differenzierung, die zwischen den Einzelschulen unterschiedliche Gewichte aufweisen kann, an den Gymnasien nicht. Umso bemerkenswerter ist es, wenn die Abbildung 26 für die Fachleistung Mathematik die Vermutung nahe legt, dass an den Gymnasien der Mathematikunterricht auf Schulebene sehr unterschiedlich effektiv gewesen sei und zu einer

divergenten Schulentwicklung geführt habe. Diese Vermutung hält indessen einer näheren quantitativen Überprüfung nicht stand: Der Anteil der Zwischen-Schul-Varianz an der Gesamtstreuung ist mit 10,5 Prozent (Klassenstufe 9) bzw. 9,0 Prozent (Klassenstufe 11) sogar etwas geschrumpft. Demnach kann sich die Heterogenisierung der mathematischen Fachleistung an den Gymnasien nicht auf Schulebene vollzogen haben. Auch auf Klassenebene hat sich der Anteil von 17,7 auf 15,4 Prozent etwas verringert. Allerdings zeigt der immer noch nicht unerhebliche Varianzanteil, dass offensichtlich die unterschiedliche Qualität des Mathematikunterrichts in den Schulklassen zu einem Teil zu einer Heterogenisierung der mathematischen Fachleistungen an den Gymnasien beigetragen haben kann.

Abbildung 27 schließlich zeigt für die Fachleistung Mathematik die Lernzuwächse über den gesamten Zeitraum der Längsschnittuntersuchung von 1996 bis 2002.

Abbildung 27 Fachleistung Mathematik: Lernentwicklung vom Beginn der Klassenstufe 5 bis zu Beginn der Klassenstufe 11 nach Bildungsgang der Klassenstufe 11 (nur Teilnehmer an allen vier Erhebungswellen)





An dieser Stelle wird die Entwicklung der Mathematikleistung wieder für die in Klassenstufe 11 gebildeten Untersuchungsgruppen berechnet. Dabei ist zu beachten, dass nur diejenigen Jugendlichen berücksichtigt werden können, für die zu allen *vier* Testzeitpunkten der Längsschnittuntersuchung Daten für den Mathematiktest vorliegen. Für insgesamt 3.264 Schülerinnen und Schüler liegen Daten für die Fachleistung Mathematik aus der LAU 5, LAU 7, LAU 9 und der LAU 11 vor.

Kann für den Zeitraum von Beginn der Klassenstufe 5 bis zu Beginn der Klassenstufe 9 für die Gruppen Gesamtschule, WTG-R sowie für die Schülerinnen und Schüler, die entweder in die Einführungs- oder in die Vorstufe des Aufbaugymnasiums eingetreten sind, ein noch recht gleichmäßiges Wachstum festgestellt werden, so hat sich die Leistungsentwicklung für die Bildungsganggruppen Gymnasium und WTG-G schon von Beginn der siebten bis zur neunten Klassenstufe deutlich verlangsamt. In dem Zeitraum vom Beginn der Klassenstufe 9 bis Anfang Klassenstufe 11 muss nun für alle Gruppen ein abnehmendes Leistungswachstum konstatiert werden.

Auffällig ist, dass im Jahre 2002 von keiner der anderen Bildungsganggruppen die Kompetenz – d. h.: das Anspruchsniveau und die Lösungssicherheit – erreicht werden konnte, dass die grundständigen Gymnasiastinnen und Gymnasiasten bereits zu Beginn der Klassenstufe 9 im Mittel erreicht hatten. Dabei entspricht der Lernstand, den die Schülerinnen und Schüler erreicht haben, die in die Einführungsstufe des Aufbaugymnasiums eingetreten sind, knapp dem Lernstand der grundständigen Gymnasiastinnen und Gymnasiasten zu Beginn der Klassenstufe 7.

Über den Gesamtzeitraum von sechs Jahren betrachtet, verzeichnen die (wenigen) Schülerinnen und Schüler, die unmittelbar in die Vorstufe des Aufbaugymnasiums eingetreten sind, mit 43,9 Skalenpunkten den größten Lernzuwachs, gefolgt von den grundständigen Gesamtschülerinnen und Gesamtschülern mit 42,8 Skalenpunkten und den Schülerinnen und Schülern in der Einführungsstufe des Aufbaugymnasiums mit 41,8 Skalenpunkten. Die vergleichsweise kontinuierlichste Lernentwicklung zeigt sich bei den Schülerinnen und Schülern, die direkt in die Vorstufe des Aufbaugymnasiums eingetreten sind.

Über die Mittelwerte, die dieser Abbildung zu Grunde liegen, sowie über die Standardabweichungen und die Anteile der Schülerinnen und Schüler in den einzelnen Bildungsganggruppen gibt Tabelle 30 Aufschluss.

Tabelle 30 Fachleistungsentwicklung Mathematik: Lernstandswerte zu Beginn der Klassenstufen 5, 7, 9 und 11 (Mittelwert, Standardabweichung und Anzahl der Schülerinnen und Schüler, die die Einzeltests bearbeitet haben) nach Bildungsgang der LAU 11

<b>Bildungsgang LAU 11</b>		<b>Mathe LAU 5</b>	<b>Mathe LAU 7</b>	<b>Mathe LAU 9</b>	<b>Mathe LAU 11</b>
<b>Gymnasium</b> N = 2.266	Mittelwert Standardabweichung	109,0 22,4	130,0 22,1	140,7 14,3	146,5 15,6
<b>Gesamtschule</b> N = 463	Mittelwert Standardabweichung	95,3 22,8	113,7 19,1	129,7 15,4	138,1 14,9
<b>WTG-G</b> N = 47	Mittelwert Standardabweichung	97,5 21,9	117,7 23,8	130,9 13,3	137,9 14,8
<b>WTG-R</b> N = 185	Mittelwert Standard- abweichung	93,8 23,1	108,4 19,7	124,5 15,2	133,2 16,2
<b>Aufbaugymnasium VS</b> N = 49	Mittelwert Standardabweichung	93,0 21,2	111,2 14,3	126,6 13,4	136,9 11,5
<i>insgesamt</i> N = 3.010	<i>Mittelwert</i> <i>Standardabweichung</i>	<i>105,5</i> <i>23,3</i>	<i>125,6</i> <i>22,8</i>	<i>137,6</i> <i>15,5</i>	<i>144,1</i> <i>16,0</i>
<b>Aufbaugymnasium ES</b> N = 254	Mittelwert Standardabweichung	87,6 21,0	105,5 16,5	121,2 13,7	129,4 15,2

## 4.5 Entwicklung der Fachleistung Englisch

Im Folgenden soll aufgezeigt werden, wie sich die Lernentwicklungen im Bereich der Fremdsprache Englisch vom Beginn der Klassenstufe 7 bzw. 9 bis zum Beginn der Klassenstufe 11 darstellen. Für die Schulleistung im Fach Englisch konnten erstmals zu Beginn des Jahres 2000 (Klassenstufe 9) Lernzuwächse berechnet werden, da die Erhebungen zu Beginn der Klassenstufe 5 wegen des damals in Hamburg noch nicht flächendeckend eingeführten Englischunterrichts an Grundschulen dieses Fach noch nicht berücksichtigen konnten. Nun aber liegen Testergebnisse für C-Tests aus den Jahren 1998, 2000 und 2002 (Klassenstufen 7, 9 und 11) vor, die miteinander so verzahnt sind, dass mit einer gemeinsamen Skala („Metrik“) für alle drei Erhebungszeitpunkte gearbeitet werden kann.

Zunächst soll die Lernentwicklung vom Beginn der Klassenstufe 9 bis zum Beginn der Klassenstufe 11 für den Englisch-C-Test betrachtet werden. Die entsprechenden Lernstandswerte und weitere Kenngrößen sind in Tabelle 31 wiedergegeben. Hier wird ersichtlich, dass die Lernentwicklung in Englisch anders verlaufen ist als in Deutsch und Mathematik. Der durchschnittliche Lernfortschritt in den Klassenstufen 9 und 10 über alle Schulformen hinweg ist mit 0,32 Standardabweichung bzw. 0,16 Standardabweichungen pro Schuljahr ausgesprochen niedrig, eine Beobachtung, die umso bemerkenswerter ist, als hier nur noch diejenigen Schülerinnen und Schüler erfasst werden, die in die gymnasiale Oberstufe eingetreten sind. Dieser Befund ist vor allem auf den geringen Lernzuwachs in der anteilsstärksten Gruppe der Gymnasiastinnen und Gymnasiasten zurückzuführen, der die wesentlich höheren Lernfortschritte in den übrigen Schulformen verdeckt.

Ein erster wichtiger Unterschied zu den übrigen Fachleistungen liegt darin, dass sich im Englischen die anfängliche Leistungsstreuung im Gesamtsystem auffällig verringert hat. Zweitens haben sich die Unterschiede zwischen den Schulformen – insbesondere zwischen den Gymnasien und den anderen Schulformen – deutlich vermindert: Lagen die Wachstumsraten bei den anderen Schulformen zwischen 0,52 und 0,72 Standardabweichungen, so konnten die Gymnasien nur ein Wachstum von  $d = 0,27$  Standardabweichungen erreichen.

Im Durchschnitt hat also die Schülerschaft der Gymnasien, gemessen an den im Englisch-C-Test geforderten Leistungen, im Untersuchungszeitraum weniger als die Hälfte dessen an Sprachkompetenz hinzugewonnen, was die Schülerinnen und Schüler der anderen Schulformen an Lernzuwächsen in den Klassenstufen 9 und 10 zu verzeichnen hatten.

Tabelle 31 Fachleistung Englisch (C-Test): Lernstandswerte zu Beginn der Klassenstufe 9 und zu Beginn der Klassenstufe 11 sowie Lernzuwachs, nach Schulform der Klasse 9

Schulform Klasse 9	Bildungsgang Klasse 11	Beginn Klasse 9			Beginn Klasse 11		Lernzuwachs		
		$\bar{x}$	SD	N	$\bar{x}$	SD	Differenz	$d$	Varianz- änderung
<b>IHR-Schule</b>	<i>Abgänge</i>	116,1	18,1	536	–	–	–	–	–
<b>IHR-Schule</b>	Aufbaugymn. VS, WTG-R	128,8	12,4	39	137,7	13,2	8,9	0,72	+12,8
<b>Realschule</b>	<i>Abgänge</i>	129,2	16,7	1.511	–	–	–	–	–
<b>Realschule</b>	Aufbaugymn. VS, WTG-R	132,7	17,2	190	141,7	14,4	9,0	0,52	-29,9
<b>Gesamt- schule</b>	<i>Abgänge</i>	115,6	19,3	2.176	–	–	–	–	–
<b>Gesamt- schule</b>	Aufbaugymn. VS, WTG-R	126,4	17,3	72	135,4	13,5	9,0	0,52	-38,6
<b>Gesamt- schule</b>	Gymn., GS, WTG-G	135,6	17,9	571	147,5	13,5	<b>11,9</b>	<b>0,66</b>	-43,1
<b>Gymnasium</b>	<i>Abgänge</i>	152,3	15,8	1.239	–	–	–	–	–
<b>Gymnasium</b>	Gymn., GS, WTG-G	154,0	14,0	2.780	157,8	14,1	3,8	0,27	+1,6
<b>insgesamt N = 9.114</b>		136,3	23,3		–	–	–	–	–
<b>insgesamt N = 3.652</b>		<b>149,2</b>	<b>17,2</b>		<b>154,7</b>	<b>15,2</b>	<b>5,5</b>	<b>0,32</b>	<b>-22,3</b>

Dass an den Gymnasien sowohl absolut als auch relativ betrachtet die Lernstandswerte am wenigsten gestiegen sind, kann nicht im Zusammenhang mit der Anfangsstreuung der Englischleistungen in dieser Schulform stehen, da diese im Vergleich zu den anderen Gruppen gering war. Welchen Einfluss der Umstand hatte, dass viele Schülerinnen und Schüler in der Jahrgangsstufe 11 wegen eines Auslandsschulbesuchs nicht erfasst wurden, kann hier nicht untersucht werden. Denkbar ist jedoch, dass der fremdsprachliche Schulunterricht bis etwa zur Klassenstufe 9 ein Plateau erreicht, auf dem bei weitgehend fortbestehenden

interindividuellen Kompetenzunterschieden kaum noch Lernfortschritte hinsichtlich der globalen Sprachkompetenz gemacht werden. Damit wäre auch der Befund vereinbar, dass in dieser Domäne die nichtgymnasialen Schulformen einen solchen Plateaueffekt in abgeschwächter Form aufweisen: hier könnte er zeitverzögert aufgetreten sein.

Auffällig ist der Umstand, dass sich – anders als in den anderen Fachleistungen – die Varianz mit Ausnahme der Gymnasien und der kleinen Gruppe der ehemaligen IHR-Schülerinnen und -Schüler durchgehend reduziert hat. Hervorzuheben ist dabei, dass gerade die beiden Gruppen, die das höchste absolute und relative Leistungswachstum zeigen – nämlich die Gesamtschulen und, wenn auch etwas weniger ausgeprägt, die Realschulen – gleichzeitig die größte Streuungsverringerung der Schülerleistungen aufweisen. Auch dies könnte ein Hinweis darauf sein, dass der Fremdsprachenunterricht *de facto* jeweils nur ein bestimmtes individuelles Niveau erreicht, auch wenn Unterschiede zwischen den Schülerinnen und Schülern fortbestehen und insoweit kein Deckeneffekt im üblichen Sinne festzustellen ist.

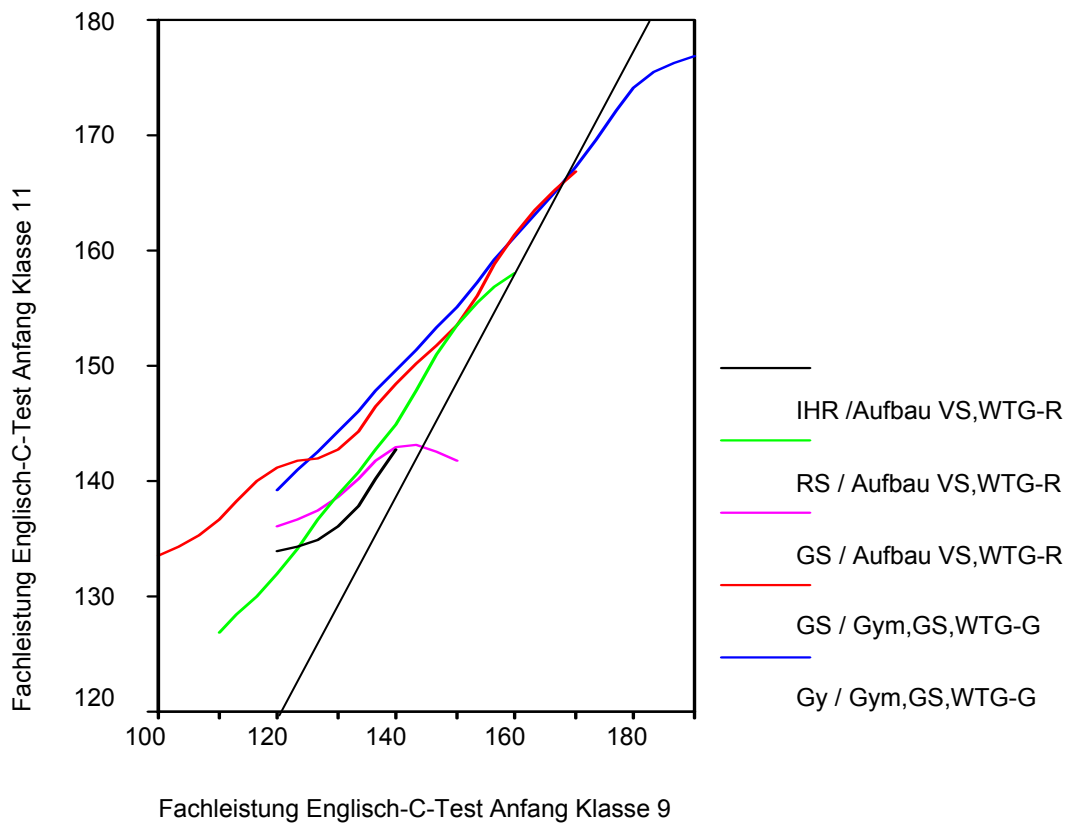
Angesichts dieser speziellen Situation im Fach Englisch ist die Aufschlüsselung der Lernstände zu Beginn der Klassenstufe 11 nach Schulformen und Lernstand Anfang Klassenstufe 9 besonders aufschlussreich (vgl. Abbildung 28).

Hier zeigt sich wieder, wie in den anderen untersuchten Domänen, dass die in allen Schulformen festzustellenden Lernfortschritte im Fach Englisch einhergehen mit einer Förderung der Schülerschaft, die sich im unteren und mittleren Leistungsspektrum befindet. Im oberen Leistungsspektrum findet über alle Schulformen hinweg im Hinblick auf die gemessenen sprachlichen Kompetenzen kein Leistungswachstum mehr statt.

Auffallend ist, dass die individuellen Lernfortschritte bei gegebener Anfangsleistung am Gymnasium und bei den grundständigen Gesamtschülerinnen und Gesamtschülern nahezu gleich verlaufen. Offenbar ist der größere Leistungszuwachs der Gesamtschule gegenüber dem Gymnasium durch die wesentlich stärkere Repräsentanz von Schülerinnen und Schülern mit anfänglich niedrigen Lernständen bedingt.

Die Realschulen ermöglichen offensichtlich im mittleren Leistungsspektrum ein höheres Wachstum und nähern sich damit den Wachstumsraten der Gymnasien an.

Abbildung 28 Fachleistung Englisch (C-Test): Lernstände zu Beginn der Klassenstufe 11 nach Testleistung zu Beginn der Klassenstufe 9 und Schulform (Lernstandswerte)



In Vorbereitung einer differenzierteren Betrachtung der Lernentwicklung einzelner Schulen ist es nützlich, zunächst die Lernausgangslagen für die Schülerschaften, die zu Beginn der Klassenstufe 9 den Schulformen Realschule (einschließlich IHR-Schule), Gesamtschule und Gymnasium zuzuordnen waren, tabellarisch zusammenzustellen (vgl. Tabelle 32).

Tabelle 32 Fachleistung Englisch (C-Test) zu Beginn der Klassenstufe 9 und zu Beginn der Klassenstufe 11 nach Schulform der Klassenstufe 9

	Beginn Klassenstufe 9			Beginn Klassenstufe 11		Lernzuwachs	
	N	$\bar{x}$	SD	$\bar{x}$	SD	Differenz	Effektstärke <i>d</i>
<b>IHR- und Realschule</b>	229	132,0	16,5	141,0	14,2	9,0	0,55
<b>Gesamtschule</b>	643	134,5	18,1	146,1	14,0	<b>11,6</b>	<b>0,64</b>
<b>Gymnasium</b>	2.780	154,0	14,0	157,8	14,1	3,8	0,27
<b>insgesamt</b>	<b>3.652</b>	<b>149,2</b>	<b>17,2</b>	<b>154,7</b>	<b>15,2</b>	<b>5,5</b>	<b>0,32</b>

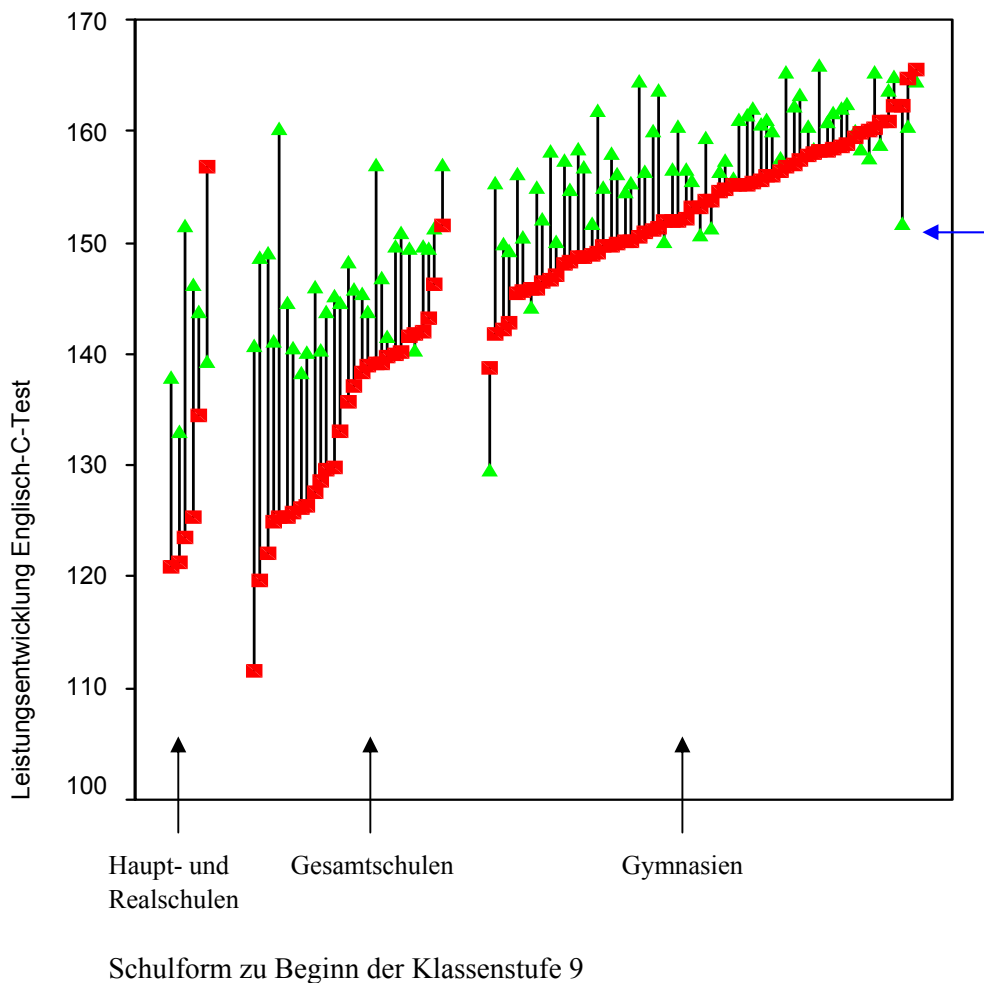
Es wird nochmals ersichtlich, dass die Lernzuwächse in den drei Ausgangsschulformen sehr unterschiedlich sind, auch wenn hier nur diejenigen Schülerinnen und Schüler betrachtet werden, die den Weg in die gymnasiale Oberstufe gefunden haben. Wegen der unterschiedlichen Anfangsverteilungen – große Ausgangsvarianz an den Gesamtschulen und geringere Ausgangsstreuung an den Gymnasien – sind wieder die absoluten und die relativen Zuwächse angegeben.

Auch in dieser Zusammenstellung sind – absolut und relativ gesehen – die Gesamtschulen mit großem Abstand gegenüber den Gymnasien die Schulform mit der günstigsten Lernentwicklung in den Klassenstufen 9 und 10. Allerdings ist in der Schülerschaft, die aus einer Gesamtschule in die gymnasiale Oberstufe eingetreten ist, der durchschnittliche Lernstand zu Beginn der Klassenstufe 11 immer noch um knapp 12 Lernstandspunkte, d. h.: um das Äquivalent von mehr als zwei Schuljahren, hinter dem der Gymnasien zurück.

Auf der Ebene der Einzelschulen (vgl. Abbildung 29, S. 112) sind im Bereich Englisch ähnlich ungleichmäßige Lernzuwächse zu verzeichnen wie bei dem Untertest Mathematik I. Deutlich zu erkennen sind auch in dieser Darstellung die mit Abstand geringsten Lernzuwächse an den Gymnasien. Überhaupt gibt es zahlreiche Beispiele dafür, dass an einzelnen Schulen, gemessen an ihrer Lernausgangslage, eine unerwartet günstige oder auch unerwartet ungünstige Lernentwicklung stattgefunden hat.

In allen drei Schulformen gibt es Schulen, die kein bzw. sogar ein negatives Wachstum aufweisen, darunter auffällig viele Gymnasien. Andererseits ist zu erkennen, dass es einer Realschule und mehreren Gesamtschulen gelungen ist, ausgehend von einer für die jeweilige Schulform keineswegs untypischen Lernausgangslage *im Durchschnitt* den Mittelwert eines Gymnasiums zu erreichen. Im Unterschied dazu lagen an einigen Gesamtschulen von Anfang an besonders günstige und insoweit einigen Gymnasien vergleichbare oder sogar bessere Anfangsbedingungen vor; diese Schulen weisen vergleichsweise geringe Lernzuwächse auf. Unter den Gymnasien fällt eine Schule heraus, die nach der Lernausgangslage am Beginn der Klassenstufe 9 zu den leistungsstärksten zählte, für den Beginn der Klassenstufe 11 aber eines der niedrigsten Ergebnisse in dieser Schulform verzeichnet (blauer Pfeil rechts in der Abbildung 29) – dies könnte auf eine größere Gruppe besonders leistungsstarker Schülerinnen und Schüler zurückzuführen sein, die zum Zeitpunkt der LAU 11 eine Schule im Ausland besuchten. Insgesamt stehen die hier berichteten Befunde unter dem Vorbehalt, dass hier ebenso wie in den analogen Analysen zu den anderen Untertests wegen des Designs der LAU 11 die Zahl der Probanden aus einer Einzelschule sehr klein sein kann.

Abbildung 29 Fachleistung Englisch (C-Test): Lernentwicklung vom Beginn der Klassenstufe 9 bis zum Beginn der Klassenstufe 11 nach Schulform und Einzelschule



Die Anfangs- und Endleistungen an den Gymnasien korrelieren auf Schulebene im Fach Englisch mit  $r = 0,66$  etwas weniger miteinander als im Rechtschreiben ( $r = 0,71$ ), in Mathematik ( $r = 0,72$ ) und im Leseverständnis ( $r = 0,67$ ). Der schulspezifische Zuwachs ist wie in den anderen Fachleistungen negativ mit dem Schulmittelwert zu Beginn der Klassenstufe 9 korreliert:  $r = -0,50$  (vgl. Deutsch-Leseverständnis:  $r = -0,31$ ; Deutsch-Rechtschreiben:  $r = -0,59$ ; Mathematik I:  $r = -0,34$ ). Demnach ist auch hier die Lernentwicklung an den Gymnasien mit ursprünglich eher niedrigen Lernständen insgesamt offenbar günstiger verlaufen als an solchen, die bereits zu Beginn des Untersuchungszeitraums hohe Lernstände zu verzeichnen hatten. So bestätigen sich für das Fach Englisch bei der differenziellen Analyse auf Schulebene im Wesentlichen die Zusammenhänge, die auch in den anderen Fächern sichtbar waren.

Aufgrund der insbesondere an den Gymnasien anzutreffenden ungleichmäßigen Entwicklung der einzelnen Schulen erscheint es angezeigt, an dieser Stelle



wieder die Zwischen-Schul-Varianz zu analysieren. Waren zu Beginn der Klassenstufe 9 noch 18,9 Prozent der Varianz durch die Schulzugehörigkeit bedingt, so ging dieser Betrag bis zum Beginn der Klasse 11 auf 13,7 Prozent zurück. Auch zwischen den Klassen hat sich der Varianzanteil von 31,1 Prozent in Klasse 9 auf 23,2 Prozent in Klassenstufe 11 verringert. Demnach sind es vor allem die individuellen Lernfortschritte von Schülerinnen und Schülern mit vergleichsweise ungünstiger Ausgangslage zu Beginn der Klassenstufe 9, die die Lernentwicklung auf den Klassenstufen 9 und 10 prägen. Für die besondere Gruppe derjenigen, die im Anschluss an die Klassenstufe 10 den Übergang in eine gymnasiale Oberstufe erreicht haben, ist eine solche konvergente Lernentwicklung charakteristisch.

Auf Grund der Verankerung der Fachleistungstests für je zwei benachbarte Messzeitpunkte durch so genannte Ankeraufgaben ist es auch für die Fachleistung Englisch-C-Test möglich, alle Aufgaben auf einer gemeinsamen Skala zu verorten. So kann über einen Zeitraum von vier Jahren hinweg die Lernentwicklung der Schülerinnen und Schüler dokumentiert werden, für die seit 1998 Leistungsdaten in Englisch vorliegen (vgl. Abbildung 30, S. 114).

Erneut ist an dieser Stelle darauf hinzuweisen, dass die Analyse wegen der Schülerfluktuationen nur anhand der in Klassenstufe 11 gebildeten Bildungsganggruppen vorgenommen werden kann. Es sind wiederum nur diejenigen Schülerinnen und Schüler in der Grafik berücksichtigt, die zu allen *drei* Erhebungszeitpunkten am Englisch-C-Test teilgenommen haben.

In der LAU 7 konnten die grundständigen Gymnasiastinnen und Gymnasiasten einen Leistungsvorsprung von 17,7 Skaleneinheiten vor den Schülerinnen und Schülern aus der Gruppe WTG-G verbuchen. In der LAU 9 hatte sich dieser Vorsprung auf 12,6 Skaleneinheiten verringert und in der LAU 11 auf 11,8 Skaleneinheiten. Bemerkenswert ist hier, dass die Verlangsamung der Kompetenzzuwächse in diesen beiden Gruppen ab Klassenstufe 9, wenn auch auf sehr unterschiedlichem Niveau, fast parallel erfolgt ist. Diese Erscheinung ist demnach schulformspezifisch und nicht primär von dem zu Beginn der Klassenstufe 9 erreichten Leistungsstand abhängig.

In den übrigen Untersuchungsgruppen dagegen ist, wie bereits angedeutet, diese Verlangsamung schwächer ausgeprägt, wobei die grundständigen Gesamtschülerinnen und Gesamtschüler bis zum Beginn der Klassenstufe 11 die Schülerinnen und Schüler aus der Gruppe WTG-G sogar überholt hat.

Insgesamt verzeichnen die grundständigen Gesamtschülerinnen und Gesamtschüler mit 45,8 Skaleneinheiten die günstigste Lernentwicklung in dem Vierjahreszeitraum, gefolgt von den Schülerinnen und Schülern, die in die Einführungs-

stufe des Aufbaugymnasiums eingetreten sind, und den Schülerinnen und Schülern, die mit einem Mittleren Schulabschluss in die Vorstufe des Wirtschafts- oder technischen Gymnasiums übergegangen sind, wobei die Schülerinnen und Schüler aus der Einführungsstufe sogar einen etwas höheren Lernstand erreicht haben (140,8 gegenüber 138,9 Skalenpunkte).

Festzuhalten bleibt allerdings, dass alle Bildungsganggruppen außerhalb der Gymnasien zu Beginn der Klassenstufe 11 nicht den Lernstand erreicht haben, den die grundständigen Gymnasiastinnen und Gymnasiasten bereits zu Beginn der Klassenstufe 9 erreicht hatten.

Abbildung 30 Fachleistung Englisch-C-Test: Lernentwicklung vom Beginn der Klassenstufe 7 bis zu Beginn der Klassenstufe 11 nach Bildungsgang der Klassenstufe 11 (nur Teilnehmer an allen drei Erhebungswellen)

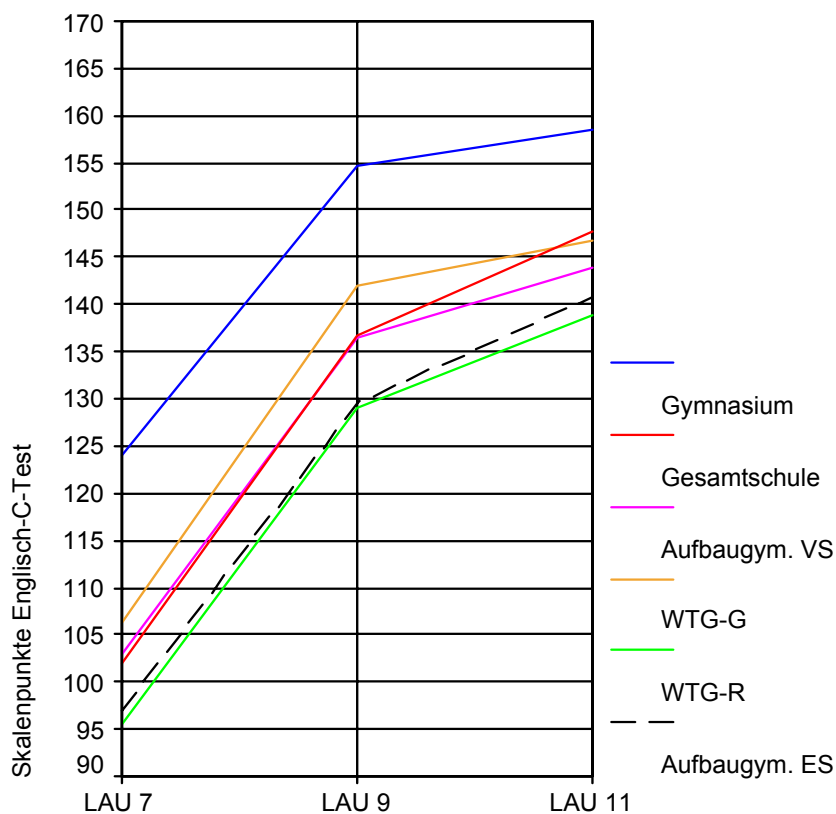


Tabelle 33 informiert wieder über Mittelwerte und Standardabweichungen der Untersuchungsgruppe der LAU 11 sowie über den jeweiligen Anteil der Schülerinnen und Schüler in diesen Segmenten der Studie.

Tabelle 33 Fachleistungsentwicklung Englisch-C-Test: Lernstandswerte zu Beginn der Klassenstufen 7, 9 und 11 (Mittelwert, Standardabweichung und Anzahl der Schülerinnen und Schüler, die die Einzeltests bearbeitet haben) nach Bildungsgang der LAU 11

Bildungsgang LAU 11		Englisch LAU 7	Englisch LAU 9	Englisch LAU 11
<b>Gymnasium</b> N = 2.183	Mittelwert	124,1	154,6	158,5
	Standardabweichung	16,5	13,6	13,7
<b>Gesamtschule</b> N = 502	Mittelwert	102,0	136,7	147,8
	Standardabweichung	16,2	18,0	13,1
<b>Aufbaugymnasium VS</b> N = 50	Mittelwert	102,9	136,5	144,0
	Standardabweichung	15,2	15,7	13,9
<b>WTG-G</b> N = 47	Mittelwert	106,4	142,0	146,7
	Standardabweichung	20,3	14,3	12,7
<b>WTG-R</b> N = 192	Mittelwert	95,4	129,1	138,9
	Standardabweichung	18,4	17,9	14,8
<b>insgesamt</b> N = 2.974	Mittelwert	117,9	149,4	155,0
	Standardabweichung	19,7	17,2	15,0
<b>Aufbaugymnasium ES</b> N = 250	Mittelwert	96,9	129,6	140,8
	Standardabweichung	15,6	17,1	13,4

## 4.6 Zusammenfassung

Abschließend sollen die Befunde zur Lernentwicklung in den Klassenstufen 9 und 10 anhand eines allgemeinen Fachleistungsindex zusammengefasst werden. Im Unterschied zu dem in Kapitel 3 gebildeten Index werden an dieser Stelle für die allgemeine Fachleistung – sowohl für die Klassenstufe 9 als auch für die Klassenstufe 11 – nur die Leistungsdaten für die Bereiche Deutsch-Leseverständnis, Deutsch-Rechtschreibung, Mathematik I sowie Englisch-C-Test herangezogen; Mathematik II und Englisch-Leseverständnis bleiben unberücksichtigt. Beide Indizes beruhen zu gleichen Teilen auf den Leistungen in den Fächern Mathematik, Deutsch und Englisch. Zwar liegen sie nicht auf derselben Skala, so dass direkte Vergleiche und daraus abgeleitete Aussagen über allgemeine Lernfortschritte nicht gemacht werden können; sie erlauben jedoch Einblicke in die relative Lernentwicklung von Schülerinnen und Schülern verschiedener Schulformen und mit unterschiedlichen Leistungsständen zu Beginn der Klassenstufe 9.

Um die allgemeine Fachleistung zu Beginn der Klassenstufe 9 mit derjenigen zu Beginn der Klassenstufe 11 in Beziehung setzen zu können, wurden die beiden Fachleistungsindizes standardisiert, und zwar nur für die Teilmenge derjenigen Schülerinnen und Schüler, die sowohl an der LAU 9 als auch an der LAU 11 teilgenommen haben. Hierzu wurde aus pragmatischen Gründen das Verfahren der so genannte „T-Noten“ nach MCCALL gewählt, nach dem für jeden der beiden Indizes der Mittelwert auf 50 Punkte und die Standardabweichung auf 10 Punkte festgelegt wurde<sup>27</sup>.

Die waagerechte Linie in Abbildung 31 kennzeichnet den Mittelwert zu Beginn der Klassenstufe 11 für diese Teilgruppe der Schülerinnen und Schüler, die sowohl an LAU 9 als auch an LAU 11 teilgenommen haben und für die für beide Testzeitpunkte ein allgemeiner Fachleistungsindex gebildet werden konnte.

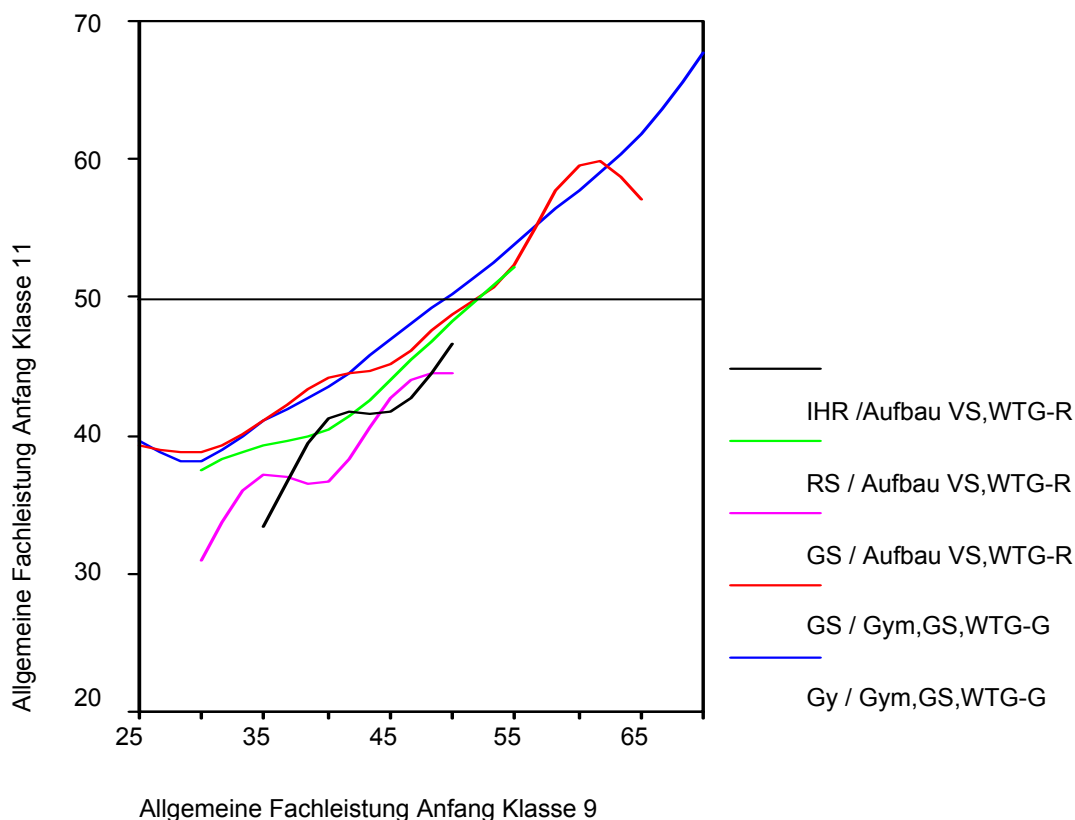
Aus Abbildung 31 wird, generalisiert für die allgemeine Fachleistung, ersichtlich, was sich bereits aus der Analyse der Ergebnisse der verschiedenen Untertests ergab: Hinsichtlich der Lernfortschritte in den Klassenstufen 9 und 10 verläuft die Entwicklung bei Berücksichtigung der Anfangsleistung zwischen den Gymnasien und den Gesamtschulen recht ähnlich, zumal in beiden Schulformen Schülerinnen und Schüler mit anfangs ungünstigem Leistungsbild überdurchschnittlich hohe Zuwächse erzielen. Unterschiede zwischen diesen beiden Bildungsgängen ergeben sich aus geringfügig niedrigeren Lernfortschritten an den Gesamtschulen im mittleren und oberen Teil des Leistungsspektrums, vor allem

---

<sup>27</sup> Zu diesem Verfahren vgl. HARTUNG *et al.*, 1993, S. 376.

aber aus den unterschiedlichen Anfangsverteilungen hinsichtlich der Fachleistung. In den anderen Schulformen sind demgegenüber, aus der individuellen Perspektive betrachtet, nahezu durchgängig geringere Lernfortschritte festzustellen, insbesondere bei den Gesamtschülerinnen und Gesamtschülern mit einem Mittleren Schulabschluss und bei den Schülerinnen und Schülern aus den IHR-Schulen. Selbst die besten Schülerinnen und Schüler dieser Schulformen – mit Ausnahme einer kleinen Spitzengruppe an den Realschulen – konnten in ihrer Lernentwicklung nicht mit dem Lerntempo Schritt halten, das für Schülerinnen und Schüler gleicher Ausgangsleistung an Gymnasien und Gesamtschulen typisch war.

Abbildung 31 Allgemeine Fachleistung: Lernstände zu Beginn der Klassenstufe 11, nach Testleistung zu Beginn der Klassenstufe 9 und Schulform (Lernstandswerte)



Unterhalb des Durchschnitts aller später in die gymnasiale Oberstufe übergetretenen Schülerinnen und Schüler befanden sich zum Erhebungszeitpunkt der LAU 9 35,5 Prozent der grundständigen Gymnasiastinnen und Gymnasiasten, während zu Beginn der Klassenstufe 11 40,5 Prozent unter dem dann gültigen Mittelwert lagen. Auch in den anderen Schulformen hat es leichte Verschiebungen zwischen den beiden Erhebungszeitpunkten gegeben. Während von den

grundständigen Gesamtschülerinnen und Gesamtschülern in der Klassenstufe 9 noch 77,7 Prozent unter dem Mittelwert aller später in die gymnasiale Oberstufe übergetretenen Schülerinnen und Schüler lagen, verringerte sich der Anteil in der Klassenstufe 11 auf 71,1 Prozent. Bei den Gesamtschülerinnen und Gesamtschülern mit einem Mittleren Schulabschluss lagen 96,1 Prozent der Schülerinnen und Schüler mit ihren Leistungen zu Beginn der Klassenstufe 9 unter dem Mittelwert aller Jugendlichen, die später den Weg in die gymnasiale Oberstufe gefunden haben, 94,7 Prozent waren es zu Beginn der Klassenstufe 11. Ähnlich hat sich an den Realschulen der Anteil unterhalb des relevanten Mittelwerts von Klassenstufe 9 nach Klassenstufe 11 verringert: Lagen in LAU 9 noch 89,1 Prozent unter dem Mittelwert aller später in die gymnasiale Oberstufe eingetretenen Jugendlichen, so waren es in LAU 11 87,1 Prozent. Demgegenüber konnten die Schülerinnen und Schüler, die eine integrierte Haupt- und Realschule besucht und den Übergang in die gymnasiale Oberstufe bewältigt haben, eine positivere Entwicklung verzeichnen: Zu Beginn der Klassenstufe 9 befanden sich 97,7 Prozent und zu Beginn der Klassenstufe 11 noch 90,7 Prozent unter dem Gesamtdurchschnitt der in der gymnasialen Oberstufe getesteten Schülerinnen und Schüler.

Komplementär zu den hier genannten Prozentanteilen gibt Tabelle 34 darüber Auskunft, wie viele Schülerinnen und Schüler aus den einzelnen Untersuchungsgruppen zu beiden Messzeitpunkten über dem jeweiligen Durchschnittswert lagen.

Tabelle 34: Anteil der Schülerinnen und Schüler der gymnasialen Oberstufe, deren allgemeine Fachleistung zu Beginn der Klassenstufen 9 bzw. 11 über dem Durchschnitt der Gesamtgruppe lag (Angaben in Prozent)

<b>Schulform Klassenstufe 9 / Bildungsganggruppe Klassenstufe 11</b>	<b>Beginn Klassenstufe 9</b>	<b>Beginn Klassenstufe 11</b>
<b>IHR / Aufbau VS, WTG-R</b>	2,3 %	9,3 %
<b>RS / Aufbau VS, WTG-R</b>	10,9 %	12,9 %
<b>GS / Aufbau VS, WTG-R</b>	3,9 %	5,3 %
<b>GS / Gy/GS/WTG</b>	22,3 %	28,9 %
<b>Gy / Gy/GS/WTG</b>	64,5 %	59,5 %

Diese Befunde sind deshalb bedeutsam, weil sich – mit Ausnahme der Gymnasien – in den Klassenstufen 9 und 10 in allen anderen Schulformen der Anteil derer erhöht hat, die Fachleistungen über dem Durchschnitt der späteren Schülerschaft der gymnasialen Oberstufe erreichen. Auch wenn die beobachteten Verschiebungen nur zwischen 1,4 und 7,0 Prozent liegen, so kann darin gleichwohl ein Hinweis darauf gesehen werden, dass außerhalb der grundständigen Gymnasien die Aussicht auf die Fortsetzung des schulischen Bildungsgangs in der gymnasialen Oberstufe mit dem Ziel des Erwerbs der allgemeinen Hochschulreife Kräfte und Anstrengungen freisetzt, die die erfolgreiche Umsetzung solcher Aspirationen wahrscheinlich macht.

Umgekehrt gilt für die Gymnasien offenbar, dass der faktische Automatismus des Übergangs von der Sekundarstufe I in die Sekundarstufe II vergleichsweise wenig Leistungsmotivation auslöst. Mit aller Vorsicht könnte man deshalb die Erwartung formulieren, dass bei der Einführung einer aussagekräftigen Fachleistungsprüfung am Ende der Sekundarstufe I auch an grundständigen Gymnasien positive Auswirkungen auf die Leistungsmotivation und die faktische Leistungsentwicklung der davon betroffenen Schülerinnen und Schüler eintreten werden. Übrigens würde man ähnlich positive Auswirkungen auch auf Lehrer- und Elternseite erwarten.

Bei Betrachtung der Fachleistungsentwicklung in Deutsch-Leseverständnis über den gesamten Zeitraum von 1996 bis 2002 konnte entgegen der üblichen Erwartung keineswegs eine besonders rasche Lernentwicklung der grundständigen Gymnasiastinnen und Gymnasiasten festgestellt werden. So verringerte sich der anfängliche Leistungsvorsprung dieser Gruppe vor den grundständigen Gesamtschülerinnen und Gesamtschülern von knapp 12 Skalenpunkten in Klassenstufe 5 bis Klassenstufe 11 auf etwa drei Skalenpunkte. Insgesamt verzeichnen die Gymnasiastinnen und Gymnasiasten einen deutlich geringeren Lernzuwachs als alle anderen Schülergruppen.

Auch in Mathematik verringerte sich während des Beobachtungszeitraums der anfängliche Leistungsvorsprung derer, die die Sekundarstufe I an einem Gymnasium durchlaufen haben, allerdings nicht in dem Ausmaß wie im Deutsch-Leseverständnis. Im Unterschied zu den nichtgymnasialen Bildungsgängen hat sich bei den grundständigen Gymnasiastinnen und Gymnasiasten und den Jugendlichen, die aus der gymnasialen Mittelstufe in ein Wirtschafts- oder technisches Gymnasium übergetreten sind, die Lernentwicklung im Fach Mathematik im Durchschnitt bereits in den Klassenstufen 7 und 8 merklich verlangsamt.

Im Bereich der Englisch-Leistungen, die im Format des Englisch-C-Tests erhoben wurden, und zwar erst ab 1998, ist in den Klassenstufen 9 und 10 eine deutliche Stagnation in der Entwicklung der fremdsprachlichen Kompetenzen zu

erkennen. Konnten in den Klassenstufen 7 und 8 alle Untersuchungsgruppen – von unterschiedlichen durchschnittlichen Leistungsständen ausgehend – ein annähernd gleichmäßiges Leistungswachstum verzeichnen, so hat sich die Leistungsentwicklung in den Klassenstufen 9 und 10 – wiederum vor allem für die grundständigen Gymnasiastinnen und Gymnasiasten sowie für die Jugendlichen, die aus der gymnasialen Mittelstufe in ein Wirtschafts- oder technisches Gymnasium eingetreten sind – negativ entwickelt. Gleichwohl ist der Leistungsvorsprung der grundständigen Gymnasiastinnen und Gymnasiasten von 10,7 Skalenpunkten bzw.  $d = 0,70$  Standardabweichungen vor dem nächstfolgenden Bildungsgang (grundständige Gesamtschülerinnen und Gesamtschüler) immer noch beträchtlich.

Zusammenfassend lässt sich sagen, dass anregungsreiche Lernmilieus mit leistungsstarken Schülerinnen und Schülern für *alle* dort Lernenden Vorteile bieten. Nach wie vor können die Schülerinnen und Schüler an grundständigen Gymnasien mit Abstand die besten Leistungen erzielen. In welchem Maße dagegen Entwicklungskapazitäten von Schülerinnen und Schülern, die aus Realschulen und IHR-Schulen in die gymnasiale Oberstufe eingetreten sind, dort zur Entfaltung gelangen, wäre noch zu ermitteln. Gleiches gilt für die Auswirkungen, die mit dem quantitativ bedeutenden Eintritt von Schülerinnen und Schülern aus nicht gymnasialen Bildungsgängen in die allgemeinbildende gymnasiale Oberstufe verbunden sind; diese Gruppe umfasst in der Klassenstufe 11 an Gymnasien einen Anteil von immerhin 10,7 Prozent<sup>28</sup>.

---

<sup>28</sup> Wegen der uneinheitlichen Übergangsregelungen an Integrierten Gesamtschulen bleibt diese Schulform hier außer Betracht, ebenso die Neuzugänge, für die nur unvollständige Daten zur abgebenden Schulform vorliegen.



## 5 Individueller und sozialer Kontext der Schulleistungen

Die zu Beginn der Klassenstufe 11 erhobenen Fachleistungen sind nicht beeinflusst von individuellen Voraussetzungen und Bedingungen, die mit dem Lebensumfeld der Schülerinnen und Schüler zusammenhängen. In diesem Kapitel werden daher zunächst die kognitiven Voraussetzungen der Schülerinnen und Schüler – insbesondere die Fähigkeit zum schlussfolgernden Denken – diskutiert. Einflussfaktoren, die auf das Alter und Geschlecht der Lernenden zurückgehen, werden im nachfolgenden Unterkapitel betrachtet. Unterschiede in den Schulleistungen, die mit einem Migrationshintergrund der Jugendlichen einhergehen, werden abschließend thematisiert.

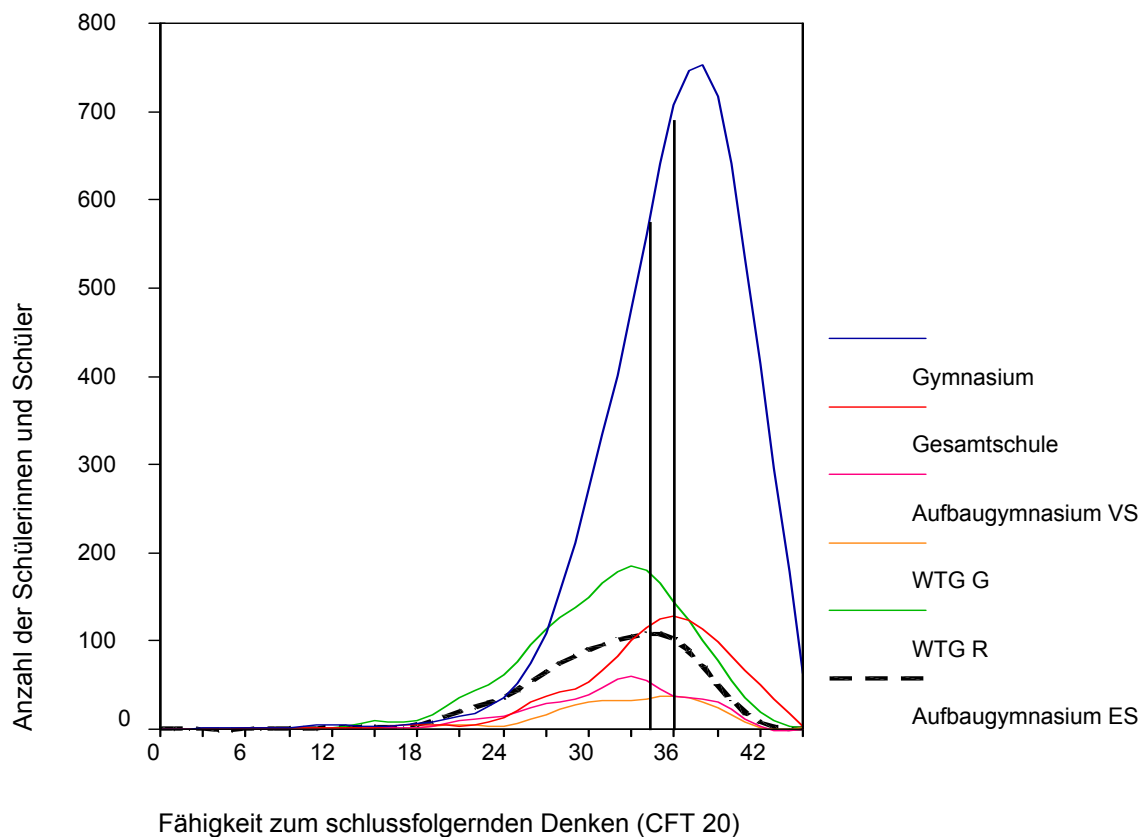
### 5.1 Kognitive Lernvoraussetzungen

Die Fähigkeit zum schlussfolgernden Denken gilt als wichtige Voraussetzung für den Schulerfolg. Sie wurde mit Hilfe des *CFT 20*-Tests (vgl. WEIß 1997) bestimmt. Da entsprechende Werte für einen Großteil der Schülerinnen und Schüler bereits aus der LAU 9 vorlagen, wurde dieser Test nur in den Aufbaugymnasien sowie in den Wirtschafts- und technischen Gymnasien durchgeführt und die Testwerte wurden auf die Jahrgangsstufe 9 zurückgerechnet.

Die Ergebnisse zeigen, dass es in der Fähigkeit zum schlussfolgernden Denken praktisch keinen Unterschied zwischen den Geschlechtern gibt. Allerdings lassen sich beträchtliche Unterschiede zwischen den einzelnen Bildungsgängen in der LAU-11-Erhebung erkennen. Abbildung 32 verdeutlicht die entsprechenden Verteilungen der Testergebnisse; sie ist analog zu den Grafiken in Kapitel 3 aufgebaut.

Die Grafik zeigt, dass sich die Maxima aller Verteilungen im obersten Drittel der Punkteskala befinden. Die Schülerinnen und Schüler erreichten in allen Bildungsgängen verhältnismäßig hohe Testergebnisse („rechtssteile Verteilung“). Dieser Befund stimmt mit den von WEIß erhobenen Daten einer Eichstichprobe überein (1997, S. 53), wobei die Resultate der Hamburger Jugendlichen – über alle Bildungsgänge der Klassenstufe 11 hinweg – mit einem Mittelwert von 34,7 Punkten und einer Standardabweichung von 5,4 Punkten generell besser ausfielen. In der bundesweiten Eichstichprobe mit Schülerinnen und Schülern der Klassenstufen 9 und 10, die dem Alter entsprechend in die gleiche Altersnormtabelle fallen, lag der Mittelwert bei 30,9 Punkten und die Standardabweichung bei 6,5 Punkten.

Abbildung 32 Fähigkeit zum schlussfolgernden Denken (Ergebnisse im *CFT 20*), nach Bildungsgang



Auffällig sind allerdings die Unterschiede zwischen den Bildungsganggruppen. Wie in den Abbildungen in Kapitel 3 markiert die rechte Senkrechte den Mittelwert der grundständigen Gymnasiastinnen und Gymnasiasten, während die linke Senkrechte den Gesamtmittelwert abbildet.

Um die Relationen zwischen den Bildungsganggruppen numerisch nachvollziehbar zu machen, sind in Tabelle 35 die wichtigsten Kennzahlen zusammengefasst.

Vergleicht man die hier berichteten Effektstärken mit denen für den allgemeinen Fachleistungsindex, so ergibt sich, dass die Zugehörigkeit zu einem bestimmten Bildungsgang offensichtlich sowohl mit der Fachleistung als auch mit den kognitiven Lernvoraussetzungen zusammenhängt. Eine Ausnahme bilden die grundständigen Gesamtschülerinnen und Gesamtschüler: Während der Zusammenhang zwischen der allgemeinen Fachleistung und der Bildungsgangzugehörigkeit einer Effektstärke von  $d = -0,37$  entspricht, zeigt sich für diese Schülerinnen und Schüler kein vergleichbarer Effekt zwischen der Bildungsgangzugehörigkeit und der Fähigkeit zum schlussfolgernden Denken ( $d = -0,06$ ). Das deutet darauf hin, dass in den Gesamtschulen – verglichen mit dem intellektuel-

len Potenzial – den tatsächlichen Fachleistungen ein höheres Gewicht bei den Übergangsentscheidungen zugemessen wird.

Tabelle 35 Fähigkeit zum schlussfolgernden Denken (Ergebnisse im *CFT 20*) zu Beginn der Klassenstufe 11 und in der Einführungsstufe des Aufbaugymnasiums, nach Bildungsgang (Mittelwerte, Standardabweichungen, Effektstärken)

	Mittelwert	Standardabweichung	Effektstärke <i>d</i>
<b>Gymnasium</b>	36,0	4,8	0,24
<b>Gesamtschule</b>	34,4	5,5	-0,06
<b>WTG-G</b>	33,1	4,8	-0,30
<b>WTG-R</b>	31,5	5,5	-0,61
<b>Aufbaugymnasium VS</b>	32,2	5,1	-0,47
<b><i>insgesamt</i></b>	<b><i>34,7</i></b>	<b><i>5,4</i></b>	<b><i>–</i></b>
<b>Aufbaugymnasium ES</b>	31,5	5,7	-0,60

Auffällig sind die nahezu identischen Werte der Schülerinnen und Schüler aus der Einführungsstufe des Aufbaugymnasiums und der Schülerinnen und Schüler, die mit einem Mittleren Abschluss in die Vorstufe des Wirtschafts- oder technischen Gymnasiums eingetreten sind (WTG-R). Offenbar sind die Differenzen zwischen beiden Gruppen nicht nur hinsichtlich der Fachleistungen gering, sondern auch hinsichtlich der kognitiven Grundfähigkeiten.

Schließlich bleibt festzuhalten, dass es – wie für die Fachleistungen berichtet – auch in Bezug auf die kognitiven Lernvoraussetzungen nennenswerte Überschneidungen zwischen allen Bildungsganggruppen gibt.

Tabelle 36 fasst die Korrelationen zwischen den einzelnen Fachleistungen und der Fähigkeit zum schlussfolgernden Denken zusammen. Diesen Werten zufolge korrelieren die Fähigkeiten, die mit dem sprachfreien *CFT 20* erhoben wurden, insgesamt am höchsten mit den Ergebnissen der Fachleistungen *Mathematik* und mit der allgemeinen Fachleistung. Niedrigere Korrelationen sind in den sprachbasierten Fächern festzustellen. Dennoch gibt es auch hier einige Unterschiede zwischen den Bildungsganggruppen. Verglichen mit den Werten für die grundständigen Gymnasiastinnen und Gymnasiasten zeigen die anderen Bil-

dungsganggruppen deutlich geringere Korrelationen im *Englisch-C-Test*. Bei den grundständigen Gesamtschülerinnen und Gesamtschülern stehen die Fachleistungen generell in vergleichsweise geringem Zusammenhang mit den kognitiven Lernvoraussetzungen. Besonders auffällig sind hier die im Widerspruch zum Gesamtbild stehenden niedrigen Korrelationen zwischen dem *Mathematik-I-Test* sowie der allgemeinen Fachleistung mit den Ergebnissen aus dem CFT 20 ( $r = 0,29$  bzw.  $r = 0,28$ ).

Tabelle 36 Korrelationen<sup>29</sup> der Fähigkeit zum schlussfolgernden Denken (Ergebnisse im *CFT 20*) mit den Ergebnissen der Fachleistungstests, nach Bildungsgang

	Mathe- matik I	Mathe- matik II	Lesever- ständnis	Recht- schrei- bung	Eng- lisch: C-Test	Eng- lisch: Lesen	allgem. Fach- leistung
<b>Gymnasium</b>	0,35**	0,34**	0,22**	0,19**	0,26**	0,20**	0,39**
<b>Gesamtschule</b>	0,29**	0,31**	0,14**	0,11*	0,09*	0,17**	0,28**
<b>WTG-G</b>	0,48**	0,40**	0,12 n.s.	0,20*	0,17*	0,22**	0,39**
<b>WTG-R</b>	0,43**	0,40**	0,33**	0,14**	0,16**	0,20**	0,41**
<b>Aufbaugy. VS</b>	0,39**	0,39**	0,37**	0,18**	0,15*	0,26**	0,44**
<b>insgesamt</b>	0,44**	0,42**	0,30**	0,29**	0,34**	0,27**	0,50**
<b>Aufbaugy. ES</b>	0,40**	0,30**	0,32**	0,17**	0,19**	0,25**	0,41**

Eine weitere Besonderheit bilden die Korrelationen in der Gruppe WTG-G. Während hier der Zusammenhang zwischen den Mathematikleistungen und den kognitiven Lernvoraussetzungen besonders eng ist, bestehen für die sprachbasierten Untertests *Passives Rechtschreibwissen* und *Englisch-C-Test* nur geringe Korrelationen auf signifikanten Niveau ( $p \leq 0,05$ ), im Untertest *Deutsch-Leseverständnis* lässt sich die Korrelation sogar nicht einmal zufallskritisch sichern. Da die entsprechenden Testergebnisse in dieser Gruppe überdurchschnittlich hoch streuen, die Ergebnisse des *CFT 20* dagegen eine vergleichsweise niedrige Streuung zeigen, muss vorerst offen bleiben, wie sich die Unterschiede in den Fachleistungen für diese Gruppe erklären lassen.

<sup>29</sup> Bei den Korrelationen bedeuten:  
 \*\* hochsignifikant ( $p \leq 0,01$ ) und  
 \* signifikant ( $p \leq 0,05$ ).

## 5.2 Alter und Geschlecht

In den früheren LAU-Erhebungen korrelierten die Fachleistungen der Schülerinnen und Schüler innerhalb einer Klassenstufe negativ mit ihrem Alter. Dieser Effekt zeigt sich auch in der LAU 11. Insbesondere Klassenwiederholungen gehen noch nach mehreren Schuljahren mit einer höheren Wahrscheinlichkeit niedrigerer Fachleistungen einher. So beträgt die Korrelation zwischen allgemeiner Fachleistung und Lebensalter der im Rahmen der LAU 11 getesteten Schülerinnen und Schüler zum Testzeitpunkt  $r = -0,26$ . Dieser Zusammenhang tritt vor allem bei den grundständigen Gymnasiastinnen und Gymnasiasten auf, während er in den anderen fünf Bildungsganggruppen weniger ausgeprägt ist.

Die insgesamt negative Korrelation lässt sich teilweise darauf zurückführen, dass sich die Schülerinnen und Schüler der verschiedenen Bildungsganggruppen in ihrem Durchschnittsalter unterscheiden (vgl. Tabelle 37). Dies ist zum Teil durch Unterschiede in der Häufigkeit von Klassenwiederholungen bedingt wie auch durch verspätete Einschulung oder eine nicht altersentsprechende Klasseneinstufung aus dem Ausland zugezogener Schülerinnen und Schüler; allerdings liegen die entsprechenden Daten nicht für alle Schülerinnen und Schüler, die an der LAU 11 teilgenommen haben, vor (vgl. Kapitel 3, S. 26). Hinzu kommen diejenigen Schülerinnen und Schüler, die im Anschluss an einen beruflichen Bildungsgang mit einem Mittleren Schulabschluss in ein Aufbau-, Wirtschafts- oder technisches Gymnasium eingetreten sind.

Tabelle 37 Lebensalter der Schülerinnen und Schüler, nach Bildungsgang (Mittelwerte, Standardabweichungen)

	Mittelwert	Standardabweichung	<i>N</i>	Effektstärke <i>d</i>
<b>Gymnasium</b>	16,9	0,63	3.503	-0,23
<b>Gesamtschule</b>	17,0	0,61	834	-0,18
<b>WTG G</b>	17,7	1,35	144	0,57
<b>WTG R</b>	18,1	1,67	746	0,91
<b>Aufbaugymnasium VS</b>	17,8	0,95	267	0,69
<b>Insgesamt</b>	17,2	0,98	5.494	–
<b>Aufbaugymnasium ES</b>	17,3	0,83	534	0,13

Entsprechend groß sind die Altersunterschiede zwischen den einzelnen Bildungsganggruppen. Diese betragen bis zu 13 Monate, eine Spanne, die mehr als einer ganzen Standardabweichung entspricht. Zudem gibt es beträchtliche Unterschiede in den Streuungen der Bildungsganggruppen: Die Altersspannen sind, bedingt durch die unterschiedlichen vorausgegangenen Bildungswege, an den Wirtschafts- und technischen Gymnasien wesentlich größer als an den Gesamtschulen und den grundständigen Gymnasien.

Tabelle 38 zeigt, dass das Durchschnittsalter derjenigen Schülerinnen und Schüler, die in der Klassenstufe 11 bzw. in der Einführungsstufe des Aufbau-gymnasiums erstmalig an der LAU-Erhebung teilgenommen haben, um etwa ein Jahr und einen Monat höher liegt als bei den Schülerinnen und Schülern, die schon vorher mindestens einmal im Längsschnitt erfasst wurden (die Effektstärke für diesen Altersunterschied entspricht  $d = 1,18$ ). Insgesamt 3.811 Schülerinnen und Schüler nahmen an allen vier LAU-Erhebungen seit der Klassenstufe 5 teil, haben die Sekundarstufe I also ohne zeitliche Verzögerung durchlaufen. Umgekehrt müssen bei den Neuzugängen in vielen Fällen leistungsbedingte Zeitverluste (z. B. Klassenwiederholungen) eine wesentliche Komponente darstellen.

Tabelle 38 Lebensalter der Schülerinnen und Schüler, nach erstmaliger Teilnahme an der LAU-Erhebung in LAU 11 und Teilnahme an mindestens einer der vorherigen LAU-Erhebungen (Mittelwerte, Standardabweichungen)

	Mittelwert	Standardabweichung	N
<b>Neuzugänge</b>	18,1	1,43	1.342
<b>Längsschnitt</b>	16,9	0,56	4.686
<b><i>insgesamt</i></b>	<i>17,2</i>	<i>0,96</i>	<i>6.028</i>

Betrachtet man das Alter der Schülerinnen und Schüler nach Geschlecht, so fällt auf, dass die männlichen Jugendlichen mit einem Durchschnittsalter von 17,2 Jahren etwas älter sind als die weiblichen Jugendlichen mit durchschnittlich 17,1 Jahren.

Im nächsten Schritt wird dem Zusammenhang zwischen Fachleistung und Geschlecht nachgegangen. An der Untersuchung zu Beginn des Schuljahres 2002/03 nahmen 53 Prozent weibliche und 47 Prozent männliche Jugendliche teil. Tabelle 39 zeigt deren Verteilung in den verschiedenen Bildungsgängen.

Tabelle 39 Verteilung der Jugendlichen in den Bildungsgängen nach Geschlecht (Fallzahlen in Klammern)

	männliche Jugendliche		weibliche Jugendliche		<i>N</i>
<b>Gymnasium</b>	47,0 %	(1.638)	53,0 %	(1.845)	3.483
<b>Gesamtschule</b>	39,7 %	(330)	60,3 %	(501)	831
<b>WTG-G</b>	68,7 %	(99)	31,3 %	(45)	144
<b>WTG-R</b>	58,8 %	(439)	41,2 %	(307)	746
<b>Aufbaugymnasium VS</b>	35,5 %	(93)	64,5 %	(169)	262
<b><i>insgesamt</i></b>	47,5 %	(2.599)	52,5 %	(2.867)	5.466
<b>Aufbaugymnasium ES</b>	40,9 %	(219)	59,1 %	(316)	535

Es ist erkennbar, dass die weiblichen Jugendlichen in den Gymnasien und Gesamtschulen sowie an den Aufbaugymnasien beider Stufen überrepräsentiert sind, während in den Wirtschafts- und technischen Gymnasien die männlichen dominieren.

Wie sich die Schulleistungen der weiblichen und der männlichen Jugendlichen unterscheiden, zeigt Tabelle 40 (siehe unten, S. 128).

Bezogen auf die allgemeine Fachleistung ist lediglich in der Gruppe WTG-G ein knapper Vorsprung der weiblichen Jugendlichen ( $d = 0,14$ ) zu verzeichnen. Demgegenüber zeigen die männlichen Jugendlichen aus den grundständigen Gymnasien und der Einführungsstufe des Aufbaugymnasiums geringfügig bessere Fachleistungen. Das Gesamtbild über die Bildungsganggruppen hinweg weist dementsprechend einen minimalen Vorsprung der männlichen Jugendlichen bezüglich ihrer allgemeinen Fachleistung aus (die Effektstärke für den Geschlechterunterschied beträgt  $d = -0,01$ ).

Tabelle 40 *Allgemeine Fachleistung* nach Geschlecht und Bildungsgang (Mittelwerte, Effektstärke, Fallzahlen in Klammern)

	männliche Jugendliche		weibliche Jugendliche		Effektstärke <i>d</i>
<b>Gymnasium</b>	54,5	(1.614)	53,9	(1.806)	-0,08
<b>Gesamtschule</b>	47,5	(327)	47,4	(485)	-0,01
<b>WTG-G</b>	46,5	(98)	48,0	(45)	0,14
<b>WTG-R</b>	43,2	(437)	42,9	(306)	-0,03
<b>Aufbaugymnasium VS</b>	46,9	(89)	46,4	(165)	-0,06
<i>insgesamt</i>	51,1	(2.565)	51,0	(2.807)	-0,01
<b>Aufbaugymnasium ES</b>	42,6	(213)	41,8	(310)	-0,09

Aufgeschlüsselt nach den zentralen Leistungsbereichen ergibt sich ein anderes Bild. Die nachfolgende Tabelle 41 zeigt die Leistungsentwicklung der jungen Frauen und Männer zunächst für den Leistungsbereich *Deutsch-Leseverständnis*.

Tabelle 41 *Fachleistung Deutsch: Leseverständnis* nach Erhebungszeitpunkt (Klassenstufe), Geschlecht und Bildungsgang (Mittelwerte; Effektstärken)

Bildungsgang	Klasse 5		Klasse 7		Klasse 9		Klasse 11		Effektstärke <i>d</i>
	Jun-gen	Mäd-chen	Jun-gen	Mäd-chen	Jun-gen	Mäd-chen	Jun-gen	Mäd-chen	
<b>Gymnasium</b> (N = 2.352)	109,0	108,7	126,0	127,0	133,5	136,3	146,6	148,0	0,09
<b>Gesamtschule</b> (N = 495)	96,5	97,4	118,0	119,2	128,1	131,0	144,3	144,0	-0,03
<b>WTG G</b> (N = 50)	100,6	92,9	112,4	117,1	122,0	129,0	137,0	141,2	0,23
<b>WTG R</b> (N = 198)	88,4	90,4	110,0	109,5	122,4	121,8	137,8	134,1	-0,21
<b>Aufbaugy. VS</b> (N = 54)	83,9	90,2	106,7	109,0	119,9	125,4	138,0	139,6	0,10
<i>insgesamt</i> (N = 3.149)	105,1	105,3	123,0	124,3	131,4	134,4	145,2	146,4	0,07
<b>Aufbaugy. ES</b> (N = 273)	85,3	89,7	108,1	112,0	119,8	124,0	138,4	135,1	-0,19

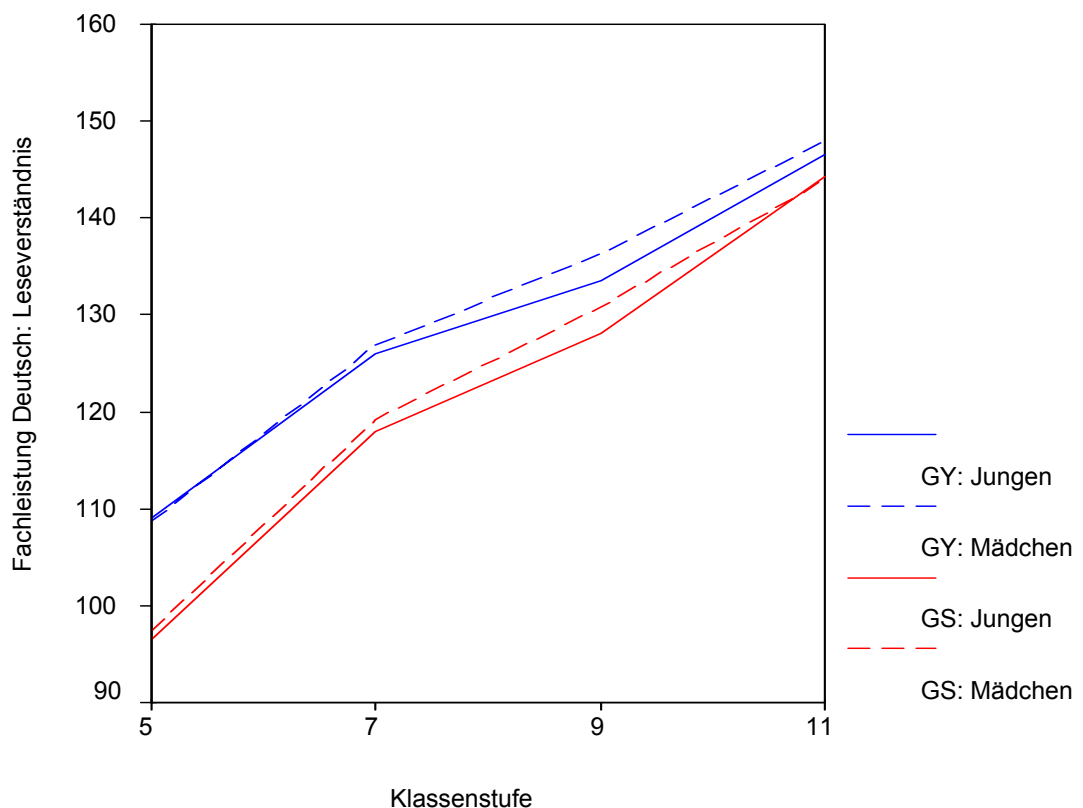


Hier hat sich der Unterschied zwischen den Geschlechtern mit einer Effektstärke von  $d = 0,01$  zu Beginn der Klassenstufe 5 über  $d = 0,08$  zu Beginn der Klassenstufe 7 auf  $d = 0,21$  zu Beginn der Klassenstufe 9 zu Gunsten der Mädchen erhöht, um dann in den beiden Folgejahren bis zum Beginn der Klassenstufe 11 wieder auf  $d = 0,07$  zurückzugehen.

Mit Ausnahme der Schülerinnen und Schüler aus der Gruppe WTG-R, die einen minimalen Unterschied von 0,6 Skaleneinheiten ( $d = -0,03$ ) zu Gunsten der Jungen verzeichnet, weisen die Mädchen in allen anderen Bildungsganggruppen zu Beginn der Klassenstufe 9 einen teilweise deutlichen Leistungsvorsprung vor den Jungen auf. Zwei Schuljahre später hat sich dieser Unterschied wieder erheblich verringert, in der Gruppe WTG-R und in der Einführungsstufe des Aufbaugymnasiums liegen die Jungen sogar recht deutlich vor den Mädchen.

Abbildung 33 veranschaulicht dieses eher heterogene Leistungsbild für die beiden zahlenmäßig stärksten Bildungsganggruppen, die grundständigen Gymnasiastinnen und Gymnasiasten und die grundständigen Gesamtschülerinnen und Gesamtschüler.

Abbildung 33 Fachleistung Deutsch *Leseverständnis* nach Erhebungszeitpunkt (Klassenstufe), Geschlecht und Bildungsgang (Mittelwerte)



In beiden Bildungsganggruppen unterscheiden sich die Mittelwerte der Jungen und Mädchen zu Beginn der Klassenstufe 5 nur minimal: in den Gymnasien zu Gunsten der Jungen ( $d = -0,02$ ), in den Gesamtschulen zu Gunsten der Mädchen ( $d = 0,04$ ). Zu Beginn der Klassenstufe 7 übertreffen die Leseleistungen der Mädchen in beiden Gruppen diejenigen der Jungen leicht, zu Beginn der Klassenstufe 9 deutlicher. Während sich die Gymnasiastinnen im Verlauf der Klassenstufen 7 und 8 um 9,4 Skalenpunkte und die Gesamtschülerinnen um 11,6 Skalenpunkte verbessern konnten, steigerten ihre Mitschüler ihre Leseleistung lediglich um 7,6 bzw. 10,1 Skalenpunkte (Geschlechterunterschied an den Gymnasien  $d = 0,22$ , an den Gesamtschulen  $d = 0,17$ ). Im Verlauf der Klassenstufen 9 und 10 haben die Jungen diesen Leistungsvorsprung der Mädchen jedoch wieder eingeholt, in den Gesamtschulen erreichen sie sogar einen etwas höheren Leistungsstand als die Mädchen ( $d = -0,03$ ).

In Tabelle 42 werden die entsprechenden Werte für den Mathematiktest wiedergegeben, wiederum aufgeschlüsselt nach Erhebungszeitpunkt (Klassenstufe), Bildungsgang und Geschlecht.

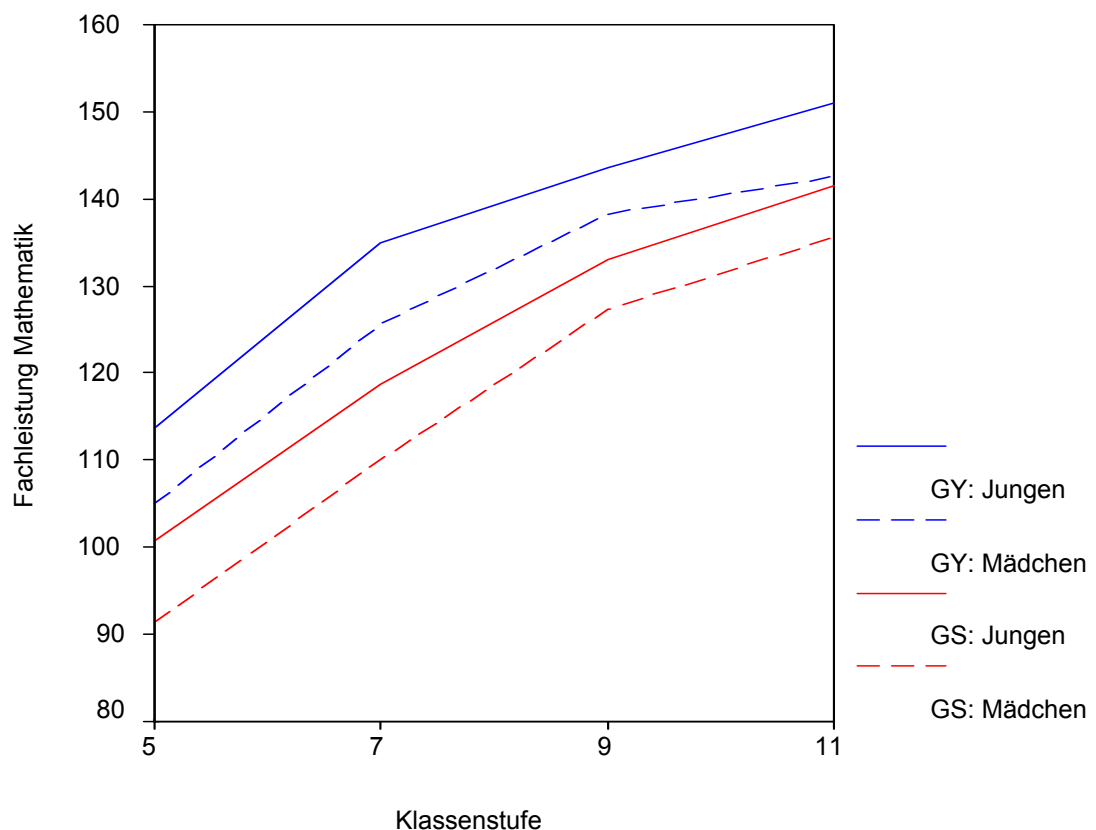
Tabelle 42 Fachleistung *Mathematik* nach Erhebungszeitpunkt (Klassenstufe), Geschlecht und Bildungsgang (Mittelwerte; Effektstärken)

<b>Bildungs- gang</b>	<b>Klasse 5</b>		<b>Klasse 7</b>		<b>Klasse 9</b>		<b>Klasse 11</b>		<b>Effekt- stärke <math>d</math></b>
	Jun- gen	Mäd- chen	Jun- gen	Mäd- chen	Jun- gen	Mäd- chen	Jun- gen	Mäd- chen	
<b>Gymnasium</b> (N = 2.266)	113,7	105,0	135,0	125,6	143,5	138,2	151,1	142,6	-0,55
<b>Gesamtschule</b> (N = 463)	100,8	91,4	118,6	110,1	133,1	127,3	141,5	135,6	-0,40
<b>WTG-G</b> (N = 47)	101,7	91,0	121,6	111,3	134,2	125,5	141,2	132,4	-0,59
<b>WTG-R</b> (N = 185)	97,4	88,4	112,1	102,9	126,9	120,8	136,2	128,8	-0,45
<b>Aufbaugy. VS</b> (N = 49)	99,2	89,6	117,2	108,0	131,7	124,0	141,7	134,3	-0,65
<b>insgesamt</b> (N = 3.010)	110,2	101,5	130,4	121,5	140,4	135,2	148,3	140,5	-0,48
<b>Aufbaugy. ES</b> (N = 254)	92,0	84,6	109,1	103,0	122,5	120,2	132,8	127,2	-0,37

Für die Mathematikleistungen finden sich deutliche Unterschiede zwischen männlichen und weiblichen Jugendlichen. In der gymnasialen Oberstufe sind die Leistungen der jungen Frauen mit einem Mittelwertunterschied von 7,8 Skaleneinheiten und einer Effektstärke von  $d = -0,48$  erheblich niedriger als die der jungen Männer. Am stärksten ausgeprägt ist der Unterschied in der Vorstufe der Aufbaugymnasien ( $d = -0,65$ ), gefolgt von der Gruppe WTG-G ( $d = -0,59$ ) und den grundständigen Gymnasien ( $d = -0,55$ ), am geringsten in der Einführungsstufe des Aufbaugymnasiums ( $d = -0,37$ ).

Abbildung 34 zeigt die unterschiedlichen Lernentwicklungen der grundständigen Gymnasiastinnen und Gymnasiasten im Vergleich mit den grundständigen Gesamtschülerinnen und Gesamtschülern, die aufgrund ihrer Gruppenstärken das Gesamtbild im Wesentlichen prägen, vom Eintritt in die Sekundarstufe I bis zum Übergang in die gymnasiale Oberstufe.

Abbildung 34 Fachleistung Mathematik nach Erhebungszeitpunkt (Klassenstufe), Geschlecht und Bildungsgang (Mittelwerte)



Während die Leistungsentwicklung der Schülerinnen und Schüler an den Gesamtschulen tendenziell parallel verläuft, zeigt sich an den Gymnasien eine divergente Entwicklung zwischen den Geschlechtern. Verzeichnen die Jungen bis zu Beginn der Klassenstufe 7 einen Lernfortschritt von 21,3 Skaleneinheiten

und die Mädchen einen Lernfortschritt von 20,7 Skalenpunkten, beträgt der Lernzuwachs der Jungen in den Klassenstufen 7 und 8 nur noch 8,5 Skalenpunkte, während sich die Mädchen um 12,6 Skalenpunkte verbessern konnten und sich deutlich dem Leistungsstand der Jungen annäherten ( $d = -0.37$ ), so dass in der Analyse der geschlechtsspezifischen Ergebnisse der LAU-9-Erhebung davon ausgegangen wurde, dass die Jungen an den Gymnasien eine viel stärkere Abschwächung des kumulativen Lernens zeigen als die Mädchen. Zu Beginn der Klassenstufe 11 hat sich die Leistungsdifferenz zwischen den Gymnasiasten und den Gymnasiastinnen jedoch wieder vergrößert ( $d = -0.55$ ): Die männlichen Jugendlichen verzeichneten eine durchschnittliche Leistungssteigerung um 7,6 Skalenpunkte, die weiblichen Jugendlichen verbessern ihre Mathematikleistung nur um 4,3 Skalenpunkte<sup>30</sup>.

Da die Mädchen an den Gesamtschulen mit einer Leistungssteigerung von 18,8 Skalenpunkten von Klassenstufe 5 zu Klassenstufe 7 (Jungen: 17,6), 17,1 Skalenpunkten von Klassenstufe 7 zu Klassenstufe 9 (Jungen: 14,5) und 8,3 Skalenpunkten von Klassenstufe 9 zu Klassenstufe 11 (Jungen: 8,4) eine den Jungen vergleichbare Lernentwicklung vollziehen, findet sich hier kein Hinweis auf geschlechtsspezifische Effekte des Mathematikunterrichts.

Abschließend soll die geschlechterspezifische Differenz für den *Englisch-C-Test* dargestellt werden; hier liegen, wie bereits erläutert, Leistungsdaten erst ab der Klassenstufe 7 vor.

Tabelle 43 ist zu entnehmen, dass der Leistungsunterschied zwischen den Schülerinnen und Schülern im Fach Englisch richtungsgleich und etwas größer ist als im Deutsch-Leseverständnis. Die Geschlechterdifferenz im Fach Englisch hat sich seit Beginn der Klassenstufe 7 ( $d = 0,17$ ) über die Klassenstufe 9 ( $d = 0,11$ ) bis zum Übergang in die Klassenstufe 11 ( $d = 0,13$ ) anders als in den beiden zuvor berichteten Leistungsbereichen nur unwesentlich verändert. Betrachtet man die Unterschiede innerhalb der Bildungsgänge, so fällt vor allem die vergleichsweise geringe Effektstärke in der Gruppe der grundständigen Gymnasiastinnen und Gymnasiasten ( $d = 0,09$ ) auf.

---

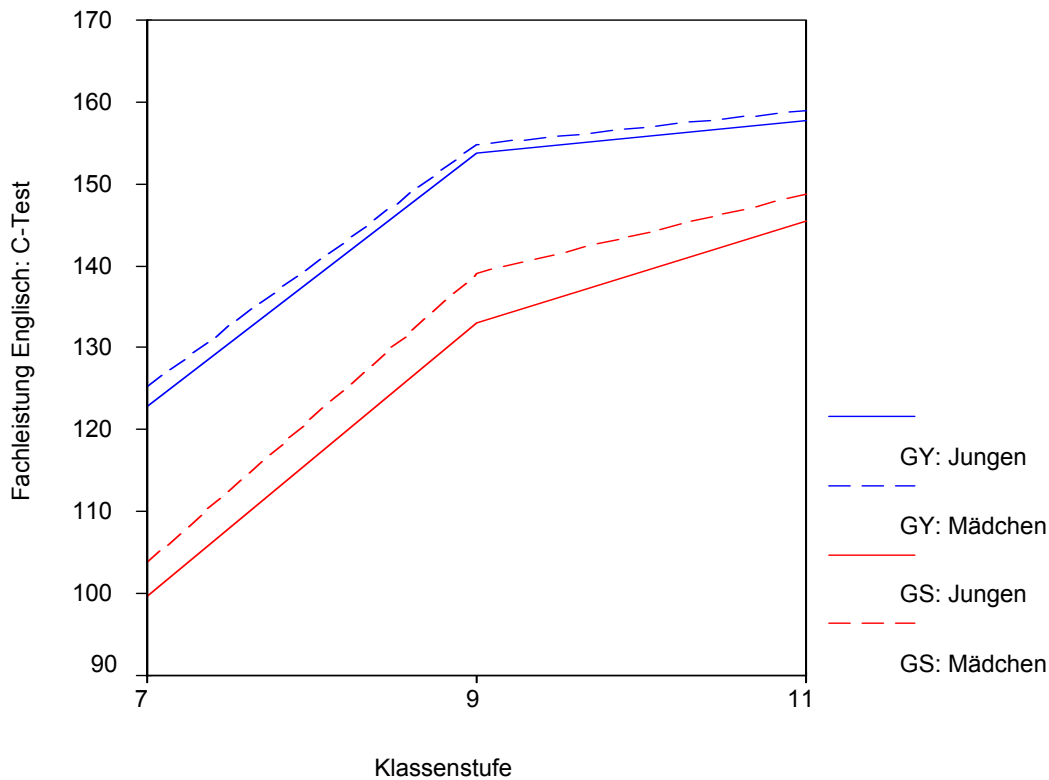
<sup>30</sup> Die zu beobachtende Diskontinuität der Lernentwicklung der Gymnasiastinnen in Mathematik könnte auf besondere Schwierigkeiten dieser Gruppe beim Übergang in die gymnasiale Oberstufe hinweisen.

Tabelle 43 Fachleistung Englisch: C-Test nach Erhebungszeitpunkt (Klassenstufe), Geschlecht und Bildungsgang (Mittelwerte; Effektstärken)

Bildungs- Gang	Klasse 7		Klasse 9		Klasse 11		Effekt- stärke <i>d</i>
	Jun- gen	Mäd- chen	Jun- gen	Mäd- chen	Jun- gen	Mäd- chen	
<b>Gymnasium</b> (N = 2.183)	122,8	125,3	154,1	155,0	157,8	159,1	0,09
<b>Gesamtschule</b> (N = 502)	99,4	103,8	132,9	139,3	145,9	149,1	0,24
<b>WTG G</b> (N = 47)	105,6	107,6	141,3	143,0	146,3	147,5	0,10
<b>WTG R</b> (N = 192)	93,1	98,8	129,2	128,9	137,9	140,4	0,17
<b>Aufbaugy. VS</b> (N = 50)	99,8	104,3	138,5	135,6	143,2	144,4	0,08
<i>insgesamt</i> (N = 2.974)	<i>116,1</i>	<i>119,4</i>	<i>148,3</i>	<i>150,3</i>	<i>154,0</i>	<i>155,9</i>	<i>0,13</i>
<b>Aufbaugy. ES</b> (N = 250)	95,2	98,1	127,1	131,2	140,0	141,4	0,11

Abbildung 35 (siehe unten, S. 134) veranschaulicht – analog zu den Leistungsbereichen Mathematik und Leseverständnis – die geschlechterspezifische Lernentwicklung der grundständigen Gymnasiastinnen und Gymnasiasten im Vergleich mit den grundständigen Gesamtschülerinnen und Gesamtschülern für das Fach Englisch. In beiden Gruppen findet von Klassenstufe 9 zu Klassenstufe 11 eine stark verlangsamte Lernentwicklung statt. Sowohl bei den Jungen als auch bei den Mädchen kommt es im Vergleich zu den beachtlichen Lernfortschritten im Verlauf der Klassenstufen 7 und 8 zu einer erheblichen Einschränkung des Lerntempos. Während die Lernentwicklung der Jungen und der Mädchen in den Gymnasien weitgehend parallel mit konstant geringem Leistungsunterschied verläuft, haben die Jungen an den Gesamtschulen in den Klassenstufen 7 und 8 nicht mit den Lernfortschritten der Mädchen Schritt halten können (Leistungsvorsprung der Mädchen:  $d = 0,36$ ), konnten sich dann aber in den Klassenstufen 9 und 10 wieder den Leistungen der Mädchen annähern ( $d = 0,24$ ). Trotz der mit 46,5 Skalenpunkten (Jungen) bzw. 45,3 Skalenpunkten (Mädchen) deutlich größeren Lernzuwächse gegenüber den Gymnasiasten (35,1 Skalenpunkte) und den Gymnasiastinnen (33,8 Skalenpunkte) ist der Leistungsabstand der Gesamtschülerinnen und Gesamtschüler zu Beginn der Klassenstufe 11 in Englisch beträchtlich und stärker ausgeprägt als in den Leistungsbereichen Mathematik und Leseverständnis.

Abbildung 35 Fachleistung Englisch (C-Test) nach Erhebungszeitpunkt (Klassenstufe), Geschlecht und Bildungsgang (Mittelwerte)



Zusammenfassend ergibt sich für die drei Leistungsbereiche Deutsch, Mathematik und Englisch, dass an den Gesamtschulen die weiblichen Jugendlichen insgesamt die größeren Leistungszuwächse in den Klassenstufen 5 bzw. 7 (Englisch) bis 8 verzeichnen. Die männlichen Jugendlichen an den Gesamtschulen verzeichnen dagegen die größeren Lernfortschritte in den Klassenstufen 9 und 10. Allerdings gelingt es dem jeweils leistungsstärkeren Geschlecht an den grundständigen Gesamtschulen in keinem der untersuchten Leistungsbereiche, die Ergebnisse der leistungsschwächeren Geschlechtergruppe der grundständigen Gymnasien zu erreichen.

### 5.3 Migrationshintergrund

Eine der wichtigsten Aufgaben des Hamburger Schulwesens ist die Integration von Schülerinnen und Schülern mit einem Migrationshintergrund. Nicht nur die soziale Integration, sondern vor allem auch die Eröffnung von Lebenschancen, die an die von der Schule vermittelten Fachkenntnisse und Kompetenzen geknüpft sind, gehört zu dieser Aufgabe. Vor diesem Hintergrund soll im Folgenden der Frage nachgegangen werden, inwieweit es den Hamburger Schulen gelungen ist, diese Teilgruppe der Schülerschaft im Verlauf der Sekundarstufe I an die leistungsbezogenen Anforderungen der gymnasialen Oberstufe bzw. der Einführungsstufe des Aufbaugymnasiums trotz oftmals bestehender sprachlicher und kultureller Barrieren heranzuführen.

Der Migrationshintergrund wird im Rahmen dieser Untersuchung über die Kriterien *Staatsangehörigkeit* (Nationalität) der Schülerinnen und Schüler, ihre *Muttersprache* und über die *Familiensprache* erfasst, zumal diese Merkmale mit Hilfe des Schülerbogens, des in der LAU 11 eingesetzten Schülerfragebogens und des in der LAU 9 eingesetzten Elternfragebogens erhoben wurden.

Nach *Staatsangehörigkeit* (ermittelt über den Schülerbogen) haben 16,7 Prozent der im Rahmen der LAU 11 getesteten Schülerinnen und Schüler eine ausländische Staatsangehörigkeit. Tabelle 44 gibt eine Übersicht über die Herkunftsländer bzw. -regionen der Schülerinnen und Schüler; die Reihenfolge richtet sich nach der Größe der jeweiligen Gruppe.

Tabelle 44 Staatsangehörigkeit der ausländischen Schülerinnen und Schüler, teilweise zusammengefasst zu Regionen (in Prozent)

Türkei	26,1%	Ost- und Südostasien	4,8%
Afghanistan	11,5%	Afrika: Subsahara	2,9%
Iran	10,5%	Südasien (Indien und Pakistan)	2,0%
Polen <sup>31</sup>	8,7%	sonstiges Europa	1,9%
ehemaliges Jugoslawien	8,0%	Nordafrika und Naher Osten	0,5%
Europäische Union	7,0%	<i>sonstige Staatangehörigkeit</i>	4,2%
Russische Föderation	6,1%	<i>doppelte (inkl. deutsche) Staatsangehörigkeit</i>	6,0%

<sup>31</sup> Obwohl Polen seit dem 1. Mai 2004 Mitgliedsstaat der Europäischen Union ist, wird es hier – um die Vergleichbarkeit mit der LAU 9 zu ermöglichen – gesondert aufgeführt.

Im Vergleich zur LAU 9 haben sich einige Verschiebungen in der Zusammensetzung der nichtdeutschen Schülerschaft ergeben. Zwar ist nach wie vor die Gruppe der Türkinnen und Türken die zahlenmäßig stärkste Gruppe unter den ausländischen Schülerinnen und Schülern, doch ist ihr prozentualer Anteil an allen ausländischen Schülerinnen und Schülern von rund einem Drittel zu Beginn der Klassenstufe 9 auf rund ein Viertel zu Beginn der Klassenstufe 11 bzw. der Einführungsstufe des Aufbaugymnasiums gesunken. Die Gruppe der Schülerinnen und Schüler aus dem Iran ist hingegen prozentual deutlich größer geworden.

Freilich wird über das Kriterium Staatsangehörigkeit nur eine Teilmenge der Schülerschaft mit Migrationshintergrund erfasst. Etliche Schülerinnen und Schüler leben in der zweiten oder dritten Generation in Deutschland und sind im Besitz eines deutschen Passes. Hinzu kommen die Jugendlichen aus Aussiedlerfamilien mit deutscher Staatsangehörigkeit. Diese Gruppen wurden über die Angaben der Schülerinnen und Schüler zu ihrer Muttersprache bzw. über die Angaben der Eltern zu der in der Familie gesprochenen Sprache gebildet. Stützt man sich auf diese Angaben, so haben insgesamt 31,6 Prozent der im Rahmen der LAU 11 getesteten Schülerinnen und Schüler einen Migrationshintergrund.

Dieser Wert ist allerdings nicht unbeeinflusst von den jeweiligen Rücklaufquoten der Fragebögen. Generell gilt, dass Ausfälle sowohl bei den Schülerinnen und Schülern als auch den Eltern umso geringer sind, je höher deren Bildungsaspirationen sind. Andererseits sind die Rücklaufquoten sowohl bei den Schülerinnen und Schülern mit einem Migrationshintergrund als auch bei deren Eltern geringer als bei den entsprechenden Gruppen ohne Migrationshintergrund. Mit der hier zugrunde gelegten Kombination der Angaben im Schülerfragebogen bzw. Elternfragebogen und über die Kenntnis der Staatsangehörigkeit aus dem Schülerbogen ist aber eine gute Näherung an die tatsächlichen Verhältnisse anzunehmen.

Mit diesen Einschränkungen werden nunmehr folgende Kategorisierungen vorgenommen, um Aussagen über die Lernstände und Lernentwicklungen von Schülerinnen und Schülern mit einem Migrationshintergrund im Vergleich zu Schülerinnen und Schülern ohne Migrationshintergrund treffen zu können:

1. Schülerinnen und Schüler mit deutscher Staatsangehörigkeit ohne einen an der *Muttersprache* oder *Familiensprache* erkennbaren Migrationshintergrund,
2. Schülerinnen und Schüler mit deutscher Staatsangehörigkeit und Migrationshintergrund, definiert entweder über die Schülerangaben zur *Muttersprache* (Schülerinnen und Schüler mit mindestens einer nichtdeutschen Mutterspra-



chen<sup>32</sup>) oder über die Elternangaben zur *Familiensprache* (Schülerinnen und Schüler, in deren Familien mindestens eine andere Sprache als Deutsch bzw. – etwa in binationalen Familien – neben Deutsch gesprochen wird<sup>33</sup>) und

3. Schülerinnen und Schüler ohne die deutsche Staatsangehörigkeit.

In Tabelle 45 werden diese drei Gruppen nach Bildungsgangzugehörigkeit aufgeschlüsselt.

Tabelle 45 Migrationshintergrund, definiert über die Kriterien Staatsangehörigkeit, Muttersprache und Familiensprache, nach Bildungsgang (Migrantenanteil in der jeweiligen Schulform und Bildungsbeteiligung), in Prozent<sup>34</sup>

	Ausländer		Deutsche mit Migrationshintergrund		Deutsche ohne Migrationshintergrund	
	Anteil	Bildungsbeteilig.	Anteil	Bildungsbeteilig.	Anteil	Bildungsbeteilig.
<b>Gymnasium</b> (N = 3.524)	12,9	47,9	15,4	64,7	71,7	65,2
<b>Gesamtschule</b> (N = 859)	17,6	16,0	12,5	12,7	70,0	15,5
<b>Aufbaugy. VS</b> (N = 283)	34,6	10,4	16,3	5,5	49,1	3,6
<b>WTG-G</b> (N = 159)	11,3	1,9	11,3	2,1	77,4	3,2
<b>WTG-R</b> (N = 835)	27,1	23,9	15,1	15,0	57,8	12,5
<b>insgesamt</b> (N = 5.660)	16,7 (N = 946)	100	14,9 (N = 841)	100	68,4 (N = 3.873)	100
<b>Aufbaugy. ES</b> (N = 544)	29,4	–	21,9	–	48,7	–

<sup>32</sup> Englisch, Französisch und Spanisch wurden nur dann berücksichtigt, wenn sie im Abgleich mit der Staatsbürgerschaft als mögliche Muttersprachen erschienen. Dagegen kommen in Familien mit gemischtem ethnischen Hintergrund auch mehrere Sprachen vor.

<sup>33</sup> Hier gelten die gleichen Einschränkungen wie bei dem Kriterium der Muttersprache.

<sup>34</sup> In dieser Tabelle ist zu unterscheiden nach „Anteil“ der jeweiligen Schülergruppe, der sich bezogen auf die jeweilige Schulform (zeilenweise) berechnet, und nach „Bildungsbeteiligung“, die sich auf die Gesamtzahl der Schülerinnen und Schüler, die einer bestimmten Gruppe angehören, (spaltenweise) berechnet..

Die Schülerinnen und Schüler mit ausländischer Staatsangehörigkeit sind im Vergleich zu den beiden anderen Gruppen an den Gymnasien unterrepräsentiert (47,9 Prozent gegenüber 64,7 bzw. 65,2 Prozent) und sowohl an den Aufbaugymnasien als auch den Wirtschafts- und technischen Gymnasien (in der Gruppe der Schülerinnen und Schüler mit einem Mittleren Schulabschluss) überrepräsentiert. Annähernd ausgewogen ist die Bildungsbeteiligung der drei Gruppen an den Gesamtschulen (16,0, 12,7 und 15,5 Prozent). Bezogen auf den jeweiligen Bildungsgang beträgt der Anteil der Schülerinnen und Schüler mit einem Migrationshintergrund an der Schülerschaft der gymnasialen Oberstufe des Gymnasiums immerhin 28,3 Prozent und ist damit fast genauso groß wie der entsprechende Anteil an der Schülerschaft der gymnasialen Oberstufe der Gesamtschule (30 Prozent). Deutlich höher ist dieser Anteil an der Schülerschaft der Wirtschafts- und technischen Gymnasien und insbesondere an der Schülerschaft der Aufbaugymnasien; hier beträgt er sowohl in der Einführungsstufe als auch in der Vorstufe über 50 Prozent.

Tabelle 46 gibt die allgemeine Fachleistung wieder, mit der die drei nach ihrem Migrationshintergrund unterschiedenen Schülergruppen in den jeweiligen Bildungsgang eingetreten sind, sowie die Effektstärken als Maß für die Mittelwertdifferenz bezogen auf die jeweilige Gesamtgruppe.

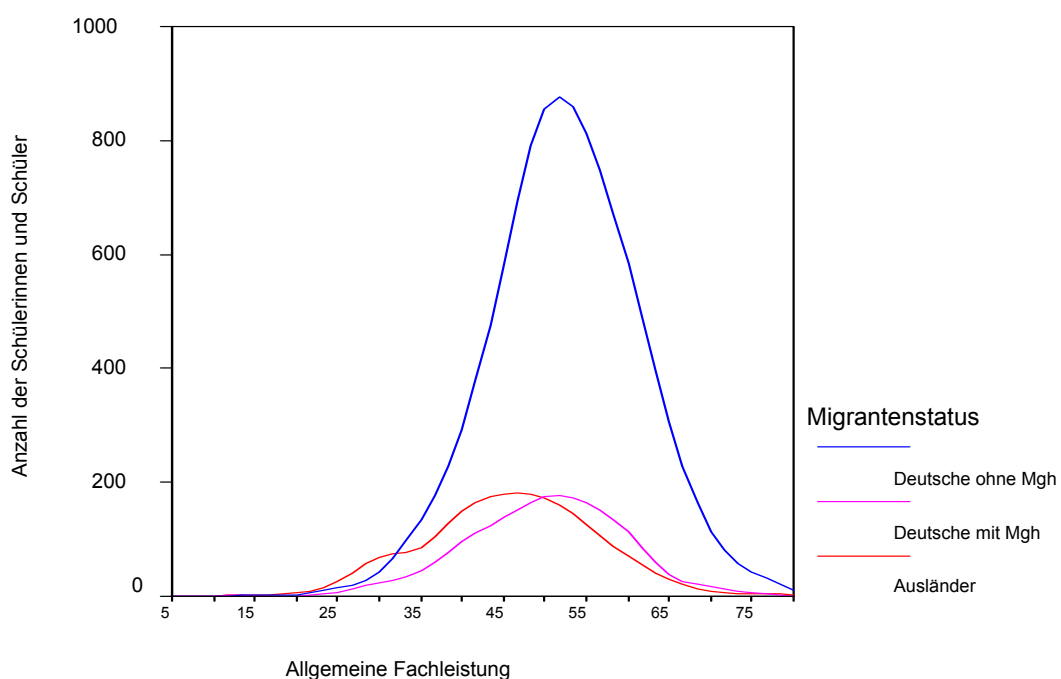
Tabelle 46 Allgemeine Fachleistung nach Migrationshintergrund und Bildungsgang (Mittelwerte und Effektstärken  $d$ )

	Ausländer		Deutsche mit Migrationshintergrund		Deutsche ohne Migrationshintergrund	
	Mittelwert	Effektstärke $d$	Mittelwert	Effektstärke $d$	Mittelwert	Effektstärke $d$
<b>Gymnasium</b>	51,0	-0,36	53,2	-0,12	54,9	0,09
<b>Gesamtschule</b>	43,7	-0,43	46,6	-0,08	48,3	0,12
<b>Aufbaugy. VS</b>	43,4	-0,35	47,0	0,09	48,0	0,22
<b>WTG G</b>	42,7	-0,44	43,3	-0,38	48,4	0,12
<b>WTG R</b>	38,0	-0,55	41,5	-0,17	45,8	0,30
<b>insgesamt</b>	45,8	-0,52	50,0	-0,09	52,3	0,15
<b>Aufbaugy. ES</b>	37,1	-0,57	42,6	0,06	44,9	0,32

Über alle Bildungsgänge hinweg verzeichnet die Gruppe der ausländischen Schülerinnen und Schüler mit 45,8 Skaleneinheiten die geringste durchschnittliche Fachleistung ( $d = -0,52$ ), während die deutschen Schülerinnen und Schüler ohne Migrationshintergrund mit 52,3 Skaleneinheiten ( $d = 0,15$ ) durchgängig den höchsten Leistungsstand aufweisen. Die deutschen Schülerinnen und Schüler mit Migrationshintergrund erreichen den für die Gesamtgruppe festgelegten Mittelwert von 50 Skaleneinheiten ( $d = -0,09$ ). Der Mittelwertunterschied der Schülerinnen und Schüler mit ausländischer Staatsangehörigkeit entspricht bezogen auf den durchschnittlichen Leistungsstand der Gesamtgruppe mit rund einer halben Standardabweichung der in der LAU 9 für die damalige Gesamtschülerenschaft ermittelten Leistungsdifferenz ( $d = -0,49$ ; vgl. LEHMANN, PEEK, GÄNSFUB & HUSFELDT 2002, S. 152).

Bezogen auf die verschiedenen Bildungsgänge verzeichnen die Gymnasien die geringsten Leistungsunterschiede zwischen den drei nach ihrem Migrationshintergrund unterschiedenen Gruppen, während sie in den Wirtschafts- und technischen Gymnasien (vor allem in der Teilgruppe derjenigen Schülerinnen und Schüler, die mit einem Mittleren Schulabschluss in die Vorstufe eingetreten sind) und in der Einführungsstufe des Aufbaugymnasiums besonders stark ausgeprägt sind. Abbildung 36 veranschaulicht die Verteilung der Fachleistungen nach Migrationshintergrund.

Abbildung 36 Allgemeine Fachleistung nach Migrationshintergrund



Während die erreichten Lernstände in der Gruppe der deutschen Schülerinnen und Schüler ohne Migrationshintergrund nahezu normalverteilt sind, zeigt sich in der Gruppe der deutschen Schülerinnen und Schüler mit Migrationshintergrund eine tendenziell rechtssteile Verteilung und in der Gruppe der Schülerinnen und Schüler ausländischer Staatsangehörigkeit eine tendenziell linkssteile Verteilung, die auf eine Dominanz leistungsschwächerer Schülerinnen und Schüler in dieser Gruppe hinweist.

Die vorausgegangenen LAU-Erhebungen hatten darauf hingewiesen, dass die geringeren Fachleistungen der Schülerinnen und Schüler ausländischer Staatsangehörigkeit mit einer ungünstigeren sozialen Lage dieser Gruppe einhergeht. Eine entsprechende Analyse auf der Grundlage der Angaben zu den bildungsrelevanten Ressourcen im Elternhaus, die im Rahmen der LAU 11 erhoben wurden, kann diesen Zusammenhang vorerst nicht bestätigen. Aus Tabelle 47 geht hervor, dass lediglich die Merkmale „Buchbestand im Elternhaus“, „eigener Schreibtisch“ und „eigene Bücher“ mit der allgemeinen Fachleistung korrelieren und lediglich der „Buchbestand im Elternhaus“ einen wesentlichen Effekt auf die Leistung ausübt. Diese bildungsrelevanten Ressourcen spielen in der insgesamt leistungsschwächeren Gruppe der ausländischen Schülerinnen und Schüler eine stärkere Rolle als in den beiden Vergleichsgruppen.

Tabelle 47 Einfluss der bildungsrelevanten Ressourcen im Elternhaus auf die allgemeine Fachleistung nach Migrationshintergrund (Ergebnisse von Regressionsanalysen)

<b>bildungsrelevante Ressourcen im Elternhaus</b>	<b>Deutsche ohne Migrationshintergrund</b> (Beta-Koeffizient)	<b>Deutsche mit Migrationshintergrund</b> (Beta-Koeffizient)	<b>Ausländer</b> (Beta-Koeffizient)	<b>insgesamt</b> (Beta-Koeffizient)
<b>Buchbestand im Elternhaus</b>	0,207	0,276	0,169	0,279
<b>eigener Schreibtisch</b>	0,037	0,033	0,063	0,050
<b>eigene Bücher</b>	0,027	0,031	0,037	0,031
$r^2$	0,049	0,084	0,039	0,089

Ein weiterer Analyseschritt gilt der Frage, ob sich die in der LAU 9 vorgefundenen Unterschiede zwischen den Jugendlichen mit ausländischer Staatsangehörigkeit wiederfinden, die im Zusammenhang mit ihrer Nationalität bzw. Herkunftsregion stehen. Dies ist, wie die in der Tabelle 48 wiedergegebenen Leis-

tungsdaten belegen, auch in der mit der LAU 11 erfassten Schülerpopulation der Fall. Die Reihenfolge der Herkunftsländer entspricht wiederum dem Anteil der jeweiligen Teilgruppe an der Gesamtheit der Schülerschaft mit ausländischer Staatsangehörigkeit; der maßgebliche Gesamtmittelwert für diese Schülergruppe beträgt 45,8 Punkte auf der Skala für die allgemeine Fachleistung.

Tabelle 48 Allgemeine Fachleistung ausländischer Schülerinnen und Schüler nach Herkunftsland bzw. Herkunftsregion (Mittelwerte, Standardabweichungen, Effektstärken innerhalb der Analysegruppe)

<b>Herkunftsland</b>	<b>Mittelwert</b>	<b>Standard- abweichung</b>	<b>Effektstärke <i>d</i></b>
<b>Türkei</b>	42,56	9,27	-0,31
<b>Afghanistan</b>	44,50	9,43	-0,13
<b>Iran</b>	44,71	10,62	-0,11
<b>Polen</b>	49,25	9,03	0,33
<b>ehem. Jugoslawien</b>	44,84	9,80	-0,09
<b>Europäische Union</b>	47,59	11,54	0,17
<b>Russische Föderation</b>	51,36	11,12	0,54
<b>Sonstige</b>	47,77	10,38	0,19
<b><i>insgesamt</i></b>	<b><i>45,80</i></b>	<b><i>10,31</i></b>	<b><i>–</i></b>

Die für jede Gruppe ermittelte durchschnittliche allgemeine Fachleistung erreicht eine maximale Differenz von annähernd 9 Skalenpunkten, was einer Effektstärke von  $d = 0,85$  entspricht. Die Schülerinnen und Schüler mit türkischer Staatsangehörigkeit liegen knapp eine drittel Standardabweichung unter dem für die Gesamtgruppe errechneten Mittelwert, während die durchschnittliche allgemeine Fachleistung der Schülerinnen und Schüler aus der Russischen Föderation mehr als eine halbe Standardabweichung über dem Mittelwert liegt, dicht gefolgt von den polnischen Schülerinnen und Schülern, deren mittlerer Leistungsstand eine drittel Standardabweichung oberhalb des Mittelwertes der Gesamtgruppe liegt. Die afghanischen und iranischen Jugendlichen sowie die Schülerinnen und Schüler aus dem ehemaligen Jugoslawien erreichen Leistungen, die etwas unter dem Mittelwert liegen, die Schülerinnen und Schüler, deren Familien aus einem Land der Europäischen Union stammen, sowie die Schülerinnen und Schüler aus „sonstigen“ Ländern erreichen Leistungen, die etwas deutlicher über dem Mittelwert liegen.

Abschließend werden die mittleren Lernstände der drei nach ihrem Migrationshintergrund unterschiedenen Schülergruppen differenziert nach Leistungsbereichen berichtet. In Tabelle 49 sind die entsprechenden Daten zusammengestellt.

Tabelle 49 Fachleistungen in den Bereichen *Mathematik*, *Deutsch* und *Englisch* nach Migrationshintergrund (Mittelwerte, Effektstärken  $d$ )

	Ausländer		Deutsche mit Migrationshintergrund		Deutsche ohne Migrationshintergrund	
	Mittelwert	Effektstärke $d$	Mittelwert	Effektstärke $d$	Mittelwert $d$	Effektstärke $d$
<b>Mathematik</b> <sup>35</sup>	47,19	-0,39	50,01	-0,07	51,67	0,11
Mathematik I	136,14	-0,33	140,68	-0,05	143,03	0,09
Mathematik II	11,90	-0,38	13,15	-0,09	14,03	0,11
<b>Deutsch</b> <sup>36</sup>	46,71	-0,48	49,63	-0,12	51,78	0,14
Leseverständnis	137,22	-0,45	141,93	-0,15	146,60	0,14
Fehlersuche	14,97	-0,35	16,59	-0,05	17,33	0,10
<b>Englisch</b> <sup>37</sup>	46,90	-0,41	50,22	-0,03	51,39	0,11
C-Test	145,68	-0,40	151,86	-0,03	154,15	0,10
Leseverständnis	93,00	-0,32	100,44	-0,02	102,91	0,08
<b>Allgemeine Fachleistung</b>	45,80	-0,52	50,00	-0,09	52,30	0,15

Die deutschen Schülerinnen und Schüler ohne Migrationshintergrund liegen in allen Leistungsbereichen etwa eine zehntel Standardabweichung über dem jeweiligen Gesamtdurchschnitt, während die Fachleistungen der deutschen Schülerinnen und Schüler mit Migrationshintergrund knapp unter dem Gesamtdurchschnitt liegen. Am stärksten ausgeprägt ist hier die Differenz in dem Untertest Deutsch-Leseverständnis mit  $d = -0,15$ . Demgegenüber liegen die Fachleistungen der ausländischen Schülerinnen und Schüler mindestens eine drittel Standardabweichung unter dem jeweiligen Mittelwert, wiederum ist dieser Effekt am deutlichsten in den Fachleistungen Deutsch.

<sup>35</sup> Index, gebildet aus den Untertests *Mathematik I* und *Mathematik II*. Der Mittelwert wurde nach der so genannten „z“-Transformation der Untertests – ebenso wie bei der Bildung des allgemeinen Fachleistungsindex – auf 50 gesetzt, die Standardabweichung auf 10.

<sup>36</sup> Index, gebildet aus den Untertests *Leseverständnis* und *Fehlersuche*.

<sup>37</sup> Index, gebildet aus den Untertests *Englisch: C-Test* und *Englisch: Leseverständnis*.

Angesichts dieser Befunde und unter der Voraussetzung, dass mit den eingesetzten Tests für die weiteren Qualifikationsprozesse wesentliche Kompetenzen überprüft werden, stellt sich auch auf der Grundlage der LAU-11-Erhebung vor allem für die Gruppe der Schülerinnen und Schüler ausländischer Staatsangehörigkeit an den Aufbau-, Wirtschafts- und technischen Gymnasien, aber auch an den gymnasialen Oberstufen der Gymnasien und Gesamtschulen die Herausforderung, nach geeigneten pädagogischen Möglichkeiten zu suchen, die beobachteten beträchtlichen Leistungsrückstände dieser Schülergruppe – insbesondere im Bereich der deutschen Sprache – zu verringern, um dem Risiko eines andernfalls absehbaren Scheiterns wirksam zu begegnen.

## 6 Schlussbetrachtung

Die Untersuchung der Fachleistungen von Schülerinnen und Schülern, die zum Beginn des Schuljahres 2002/03 den Weg in die gymnasiale Oberstufe gefunden haben, hat es ermöglicht, eine Reihe von Leitfragen miteinander zu verschränken:

- Auf welchem Wege sind die Schülerinnen und Schüler an diese Gelenkstelle zwischen den Sekundarstufen I und II gelangt und welche Hintergrundfaktoren haben sich dabei als wirksam erwiesen?
  - Welche Lernstände sind in den verschiedenen Bildungsgängen bis zum Übergang in die allgemein bildende Sekundarstufe II erreicht worden und welche Bedeutung besitzen diese Markierungen der jeweiligen ‚Lernausgangslage‘ für die verschiedenen Formen der gymnasialen Oberstufe?
  - Welche Folgerungen ergeben sich insbesondere unter dem Aspekt der Vergleichbarkeit von Zensuren, Übergangs- bzw. Zugangsberechtigungen sowie von Schulabschlüssen?
  - Was lässt sich der Ertrag der unterrichtlichen Arbeit in den verschiedenen Bildungsgängen ermitteln, insbesondere bei Berücksichtigung der jeweiligen Lernvoraussetzungen?
1. Der traditionelle Weg in die gymnasiale Oberstufe – Entscheidung für die Beobachtungsstufe des Gymnasiums, möglichst reibungsloser Durchgang durch die Sekundarstufe I, prüfungsfreier, gleichsam automatischer und bedingungsloser Eintritt in die gymnasiale Oberstufe – stellt zwar qualitativ und quantitativ das dominante Muster dar. Rechnet man die hier nicht untersuchten, weil zum Zeitpunkt der Erhebungen im Ausland befindlichen Schülerinnen und Schüler hinzu, so sind ca. 70 Prozent diesen Weg gegangen. Immerhin 14 Prozent sind jedoch über das anspruchsvollere Kursniveau in der Mittelstufe der Gesamtschule in die gymnasiale Oberstufe gelangt und rund 15 Prozent haben zunächst an einer Real- oder Gesamtschule den Mittleren Schulabschluss erworben. Bezogen auf die Chance, eine Form der Sekundarstufe II zu besuchen, die zur Allgemeinen Hochschulreife führt, weist das Hamburger Schulsystem also ein wesentlich höheres Maß an Durchlässigkeit auf, als von den Kritikern gemeinhin unterstellt wird.
  2. Damit ist indessen nicht gesagt, dass die (mit einer sogleich zu diskutierenden Ausnahme) formal als äquivalent betrachteten Formen der gymnasialen Oberstufe auch substanziell unter Bedingungen arbeiten, die



untereinander vergleichbar sind. Sowohl, was die kognitiven Lernvoraussetzungen anbetrifft, als auch in Anbetracht außerschulischer Hintergrundfaktoren unterscheiden sich die Schülerschaften beim Einstieg in die gymnasiale Oberstufe deutlich. Die in dieser Hinsicht ‚privilegierteste‘ Gruppe ist jeweils an den grundständigen Gymnasien anzutreffen. Es folgen die Gesamtschüler mit direkter Übergangsberechtigung in die gymnasiale Oberstufe, danach diejenigen mit direkter Übergangsberechtigung an ein Fachgymnasium und etwa gleichauf die Schülerinnen und Schüler der Vorstufe des Aufbaugymnasiums sowie wiederum etwa gleichauf die Schülerinnen und Schüler mit Mittlerem Schulabschluss und die Schülerschaft der Einführungsstufe des Aufbaugymnasiums.

3. Diesem Muster folgen, nicht unerwartet, die durchschnittlichen schulformspezifischen Fachleistungen. Die entsprechenden Unterschiede sind in vielen Fällen sehr erheblich; sie können dem Äquivalent von zwei und mehr Lernjahren entsprechen. Angesichts dessen ist es fraglich, ob ein reduziertes Lerntempo in der Klassenstufe 11, das Schülerinnen und Schülern, die außerhalb der traditionellen Gymnasialaufbahn die Zugangsberechtigung zur gymnasialen Oberstufe erworben haben, den Anschluss ermöglichen soll, auf realistischen Erwartungen beruht. Dabei wurde die Frage der unterstellten Äquivalenzbeziehungen von Zensuren (etwa der so genannten A- und B-Noten in den Gesamtschulen) im vorliegenden Zusammenhang noch nicht untersucht.
4. Angesichts der starken sozialräumlichen Gliederung Hamburgs, die nach den bisher vorliegenden Analysen jenseits der Grundschulen auch die Erträge der Arbeit an den Sekundarschulen beeinflusst, war mit Überschneidungen in den Leistungsspektren der einzelnen Bildungsgänge zu rechnen, erst recht, wenn diese formal als gleichwertig betrachtet werden. Im Hinblick auf diesen in der einschlägigen Literatur häufig betonten Aspekt ist es eher überraschend, wie gering diese Überschneidungsbereiche ausgeprägt sind: Etwa 10 Prozent der Schülerschaft von grundständigen Gymnasien hat Testwerte erzielt, die unter dem Durchschnitt derer liegen, die mit dem Mittleren Schulabschluss in die gymnasiale Oberstufe eingetreten sind, während umgekehrt 10 Prozent der letztgenannten Gruppe den grundständig-gymnasialen Mittelwert übertroffen haben.
5. Wie schon in den bereits vorliegenden Analysen zur Längsschnittuntersuchung LAU legen die Daten der LAU 11 erneut den Schluss nahe, dass unter den eindeutig bevorzugten Arbeitsbedingungen der gymnasialen Mittelstufe (hier der Klassenstufen 9 und 10) geringere Lernzu-

wächse zu beobachten sind als eigentlich zu erwarten. Es entspräche der ‚Systemrationalität‘ eines gegliederten Schulwesens, wenn in den positiv ausgelesenen Schulformen deutlich raschere Lernfortschritte realisiert würden als an den weniger begünstigten Schulen.

6. Von besonderem Interesse sind die Lernerfolge derjenigen Jugendlichen, die aus Migrantenfamilien stammen. Dabei ist es zweckmäßig zu unterscheiden zwischen Jugendlichen, die die deutsche Staatsbürgerschaft besitzen und für die kein Indiz für einen Migrationshintergrund vorliegt, solchen, die zwar Deutsche sind, deren sprachliches Umfeld (d. h., deren Mutter- oder Familiensprache) auf Einwanderung oder Aussiedlung schließen lässt, und denjenigen, die keine deutsche Staatsbürgerschaft besitzen. Dabei zeigt sich, dass diese drei Gruppen überraschenderweise an den Gesamtschulen und bei besonders qualifiziertem Mittlerem Schulabschluss auch an den Fachgymnasien in der Oberstufe anteilmäßig nahezu repräsentativ vertreten sind. In den Gymnasien der Regelform sind Schülerinnen und Schüler mit Migrationshintergrund unterrepräsentiert, in den Fach- und Aufbaugymnasien ohne besonders qualifizierten Mittleren Bildungsabschluss aber überrepräsentiert. Letztere stellen also offenbar – trotz der beobachteten Defizite im Bereich der Fachleistung – für die Jugendlichen aus Migrantenhaushalten zumindest formal ein beachtliches Integrationsangebot dar. Allerdings ist zu berücksichtigen, dass bei den Schülerinnen und Schülern mit Migrationshintergrund dasjenige Segment hier außer Betracht bleiben musste, das durch vorzeitigen Schulabgang oder frühen Einstieg in die Berufsausbildung aus dem Längsschnitt ausgeschieden ist. Seine Berücksichtigung würde die tatsächliche Situation dieser Gruppe von Jugendlichen in einem problematischeren, aber auch realistischeren Licht erscheinen lassen.

## Literaturverzeichnis

- BAUMERT, J. / BOS, W. / LEHMANN, R. H. (Hrsg.) (2000): TIMSS/III. Dritte Internationale Mathematik- und Naturwissenschaftsstudie – Mathematische und naturwissenschaftliche Bildung am Ende der Schullaufbahn. Band I: Mathematische und naturwissenschaftliche Grundbildung am Ende der Pflichtschulzeit. Opladen.
- BEHÖRDE FÜR SCHULE, JUGEND UND BERUFSBILDUNG, AMT FÜR SCHULE, HAMBURG (Hrsg.) (1998): Der Hamburger Schulleistungstest für sechste und siebte Klassen – SL-HAM 6/7. Hamburg.
- BEHÖRDE FÜR SCHULE, JUGEND UND BERUFSBILDUNG, AMT FÜR SCHULE, HAMBURG (Hrsg.) (1998): Der Hamburger Schulleistungstest für achte und neunte Klassen – SL-HAM 8/9. Hamburg.
- BEHÖRDE FÜR BILDUNG UND SPORT, AMT FÜR BILDUNG (Hrsg.) (1998): Der Hamburger Schulleistungstest für zehnte und elfte Klassen – SL-HAM 10/11. Hamburg.
- FISCHER, G. H. / MOLENAAR, I. H. (1995): Rasch Models – Foundations, Recent Developments, and Applications. New York.
- GROTJAHN, R. (1995): Der C-Test. State of the Art. In: Zeitschrift für Fremdsprachenforschung. 6. Jahrgang. Heft 2. S. 37 - 60.
- HAMBLETON, R. K. / SWAMINATHAN, H. / ROGERS, H. J. (Eds.) (1991): Fundamentals of Item Response Theory. Newbury Park, CA.
- HARTUNG, J., EPELT, B., KLÖSENER, K.-H. (1993): Statistik. Lehr- und Handbuch der angewandten Statistik. München.
- HORSTKEMPER, M. (1987): Schule, Geschlecht und Selbstvertrauen. Eine Längsschnittstudie über Mädchensozialisation in der Schule. München.
- LEHMANN, R. H. / PEEK, R. (1997): Aspekte der Lernausgangslage von Schülerinnen und Schülern der fünften Klassen an Hamburger Schulen. Bericht über die Untersuchung im September 1996 (unveröffentlichter Forschungsbericht). Hamburg.
- LEHMANN, R. H. / GÄNSFUß, R. / PEEK, R. (1999): Aspekte der Lernauslagen und der Lernentwicklung von Schülerinnen und Schülern an Hamburger Schulen – Klassenstufe 7. Bericht über die Untersuchung im September 1998 (unveröffentlichter Forschungsbericht). Hamburg.

- LEHMANN, R. H. / PEEK, R. / GÄNSFUß, R. / HUSFELDT, V. (2002): Aspekte der Lernauslage und der Lernentwicklung – Klassenstufe 9. Ergebnisse einer Längsschnittuntersuchung in Hamburg. Hamburg.
- LEHMANN, R. H. / IVANOV, S. / HUNGER, S. / GÄNSFUß, R. (2004, im Druck): Untersuchung der Leistungen, Motivation und Einstellungen zu Beginn der beruflichen Ausbildung. Hamburg.
- NIKOLOVA, R. / LEHMANN, R. L. (2003): On the Dimensionality of the Cognitive Test Used in the IEA Civic Education Study: analyses and implications. In: European Educational Research Journal, Volume 2, Number 3. 370-383.
- OECD & STATISTICS CANADA (1995): Literacy, Economy and Society. Results of the First International Adult Literacy Survey. Ontario.
- RAATZ, U. / KLEIN-BRALEY, C. (1983): Ein neuer Ansatz zur Messung der Sprachleistung. Der C-Test. Theorie und Praxis. In: R. Horn / K. Ingenkamp / R. S. Jäger (Hrsg.). Tests und Trends 3. Jahrbuch der Pädagogischen Diagnostik. Weinheim und Basel. S. 107-137.
- WEIß, R. (1997): Grundintelligenztest Skala 2 (CFT 20) mit Wortschatztest (WS) und Zahlenfolgentest (ZF). 4. überarbeitete Auflage. Göttingen.

## Glossar

### Alpha

Alpha (genauer: „Cronbachs Alpha“) ist eine Kennzahl für die interne Konsistenz einer  $\Rightarrow$  Skala. Werden inhaltlich zusammenpassende Aufgaben oder Fragen zu einer Skala zusammengefasst, z. B. zehn Fragen aus dem Schülerfragebogen zur Skala „Schulzufriedenheit“, so gibt Alpha Auskunft darüber, wie gut sich die Antworten zu einer Gesamttendenz zusammenfügen. Alpha kann maximal den Wert 1 annehmen. Als Faustregel gilt, dass bei Skalen ab  $\alpha = 0,75$  eine befriedigende und ab  $\alpha = 0,85$  eine gute Skalenqualität vorliegt.

### Ankeraufgaben

Ankeraufgaben sind Aufgaben, die in vorangegangenen Untersuchungen eingesetzt und ausgewertet wurden. Sie dienen als Berechnungsgrundlage für Längsschnittanalysen und ermöglichen Aussagen zur Lernentwicklung.

### arithmetisches Mittel, arithmetischer Mittelwert, Durchschnittswert

$\Rightarrow$  Mittelwert.

### Beta (-Gewicht)

$\Rightarrow$  Regressionsanalyse.

### *d*, Effektstärke *d*, standardisierte Effektstärke *d*

ist ein standardisiertes Maß für Merkmalsunterschiede zwischen zwei Gruppen. Sie wird berechnet, indem die Differenz der  $\Rightarrow$  Mittelwerte der Gruppen durch die gemeinsame  $\Rightarrow$  Standardabweichung dividiert wird. Sollen Merkmalsunterschiede einer Untersuchungsgruppe herausgestellt werden, die sich aufgrund beispielsweise von Lernentwicklung ergeben, so wird die Differenz der Mittelwerte durch die Standardabweichung der Ausgangslage dividiert.

### diskontinuierliche Texte

Diskontinuierliche Texte sind Grafiken, Tabellen oder Listen, die im Unterschied zu kontinuierlichen Texten (z. B. Beschreibungen) keinen unterbrochenen Textfluss aufweisen.

**erklärte Varianz**  $\Rightarrow$  Kovarianzanalyse,

$\Rightarrow$  Regressionsanalyse,  $\Rightarrow$  Varianzanalyse

## ***Eta*<sup>2</sup>, Bestimmtheitsmaß *Eta*<sup>2</sup>**

ist die Maßzahl für die erklärte Varianz bei der  $\Rightarrow$  Kovarianzanalyse und der  $\Rightarrow$  Varianzanalyse. *Eta*<sup>2</sup> bezeichnet den Varianzanteil, den man erhält, wenn man die Einzelwerte durch den jeweiligen Gruppenmittelwert ersetzt, daraus die Varianz berechnet und durch die ursprüngliche Varianz teilt; es variiert zwischen 0 und 1. Z. B. gibt das mit der Schulform verbundene *Eta*<sup>2</sup> für ein Leistungsmerkmal Auskunft darüber, welchen Anteil Schulformunterschiede (Mittelwertdifferenzen) an den Leistungsunterschieden zwischen den Schülerinnen und Schülern überhaupt haben.

## **Interkorrelation**

$\Rightarrow$  Korrelation

## **interne Konsistenz**

$\Rightarrow$  Alpha.

## **intervallskalierte Variable**

Wenn Gleichheit der Abstände zwischen zwei benachbarten Werten einer Variable angenommen werden darf, spricht man von einer intervallskalierten Variable. Ist dies nicht der Fall, liegt eine nominalskalierte (keine Rangfolge der Werte) oder eine ordinalskalierte (Rangfolge der Werte, aber keine Abstandsgleichheit) Variable vor.

## **Item**

$\Rightarrow$  Variable

## **Kernaufgaben**

Kernaufgaben sind Aufgaben, die bei Einsatz verschiedener Testversionen (RS- / Gy-Version) den übergreifenden Kernbereich in den versionsspezifischen Aufgabenpools bilden und somit die Vergleichbarkeit der einzelnen Schulformen gewährleisten.

## **Koeffizient**

Ein Koeffizient ist ein statistischer, mathematischer Kennwert.  $\Rightarrow$  Pearsons *r* ist z. B. ein Korrelationskoeffizient, d. h. ein statistisches Zusammenhangsmaß.

## **Korrelation, korrelieren**

Zusammenhang zwischen zwei  $\Rightarrow$  Variablen (Merkmalen). Für  $\Rightarrow$  intervallskalierte Daten ist das Korrelationsmaß der Pearsonsche Produkt-Moment-Korrelationskoeffizient *r* (kurz „Pearsons *r*“ oder nur „*r*“). Er variiert zwischen minus 1 und plus 1. Ein hohes negatives *r* besagt: Je

höher das eine Merkmal ausgeprägt ist, desto niedriger das andere Merkmal, und je niedriger das eine Merkmal, desto höher das andere Merkmal. Ein hohes positives  $r$  besagt sinngemäß entsprechend: Je höhere Werte das eine Merkmal annimmt, desto höhere auch das andere (bzw. je niedriger, desto niedriger). Ein  $r$  nahe Null sagt aus, dass zwischen den beiden Merkmalen kein Zusammenhang besteht.  $r^2$  gibt direkt die so  $\Rightarrow$  erklärte Varianz an.

### **Kovarianzanalyse**

Die Kovarianzanalyse ist eine Mischform aus  $\Rightarrow$  Regressionsanalyse und  $\Rightarrow$  Varianzanalyse. Die abhängige Variable (Kriterium) wird durch unabhängige Variablen erklärt, von denen mindestens eine wie bei der Varianzanalyse nominalskaliert ist (der so genannte „Faktor“), andere dagegen wie bei der Regressionsanalyse intervallskaliert (Kovariaten). Die Maßzahl für die insgesamt erklärte Varianz ist  $\Rightarrow R^2$ ; für den Faktor allein kann auch ein  $\Rightarrow \text{Eta}^2$  bestimmt werden.

### **Kriterium (s-Variable)**

abhängige Variable,  $\Rightarrow$  Regressionsanalyse.

### **linkssteile / rechtssteile Verteilung**

Wenn sich die Werte im Gegensatz zu einer Normalverteilung nicht gleichmäßig um den Mittelwert streuen, gibt die Schiefe einer Verteilung an, in welchem Ausmaß die Werte vermehrt in eine Richtung der Skala tendieren. Man unterscheidet nach linkssteiler und rechtssteiler Verteilung. Eine linkssteile Verteilung (= positiv schief) liegt dann vor, wenn der Hauptanteil der Verteilung auf der linken Seite liegt. Umgekehrt wird eine rechtssteile Verteilung (= negativ schief) dann vorliegen, wenn der Hauptteil der Verteilung auf der rechten Seite liegt. Links- und rechtssteile Verteilungen sind immer asymmetrisch.

### **Lösungswahrscheinlichkeit**

Die Lösungswahrscheinlichkeit einer Aufgabe gibt an, wie groß die Wahrscheinlichkeit ist, dass irgendein Schüler bzw. irgendeine Schülerin diese Aufgabe löst. Die Lösungswahrscheinlichkeit wird mit dem Wert  $p$  (vom englischen „probability“) angegeben und liegt zwischen 0 und 1. Eine Lösungswahrscheinlichkeit von  $p = 0,47$  beispielsweise besagt, dass „auf lange Sicht“ 47 Prozent der Schülerinnen und Schüler einer definierten Gruppe diese Aufgabe lösen.

## **Mittelwert**

Kurzbezeichnung für den arithmetischen Mittelwert. Der Mittelwert darf berechnet werden, wenn mindestens  $\Rightarrow$  intervallskalierte Daten vorliegen. Er ist die Summe der Einzelwerte aller Fälle, dividiert durch die Fallzahl.

## **multiple Regressionsanalyse**

$\Rightarrow$  Regressionsanalyse.

## **multiple $R$ , multiple $R^2$**

$\Rightarrow$  Regressionsanalyse.

## **Normalverteilung**

Die Normalverteilung ist ein Verteilungsmodell für kontinuierliche Zufallsvariablen. Die Normalverteilung unterstellt eine symmetrische, eingipflige Verteilung in Form einer Glocke (Gauß-Verteilung), die sich auf beiden Seiten (Minimum und Maximum) der x-Achse schnell asymptotisch annähert. Die Normalverteilung ist gegeben durch die Wahrscheinlichkeitsdichte, diese wiederum ist durch die Standardabweichung und den Erwartungswert definiert. Ist der Erwartungswert 0 und die Standardabweichung 1, spricht man von einer Standardnormalverteilung. Im mittleren Bereich (plus/minus eine Standardabweichung) liegen etwa 68 Prozent der Fälle; im Bereich Mittelwert plus/minus zwei Standardabweichungen liegen etwa 95 Prozent der Fälle ( $\Rightarrow$  Standardabweichung).

## **Objektivität**

Objektivität ist ein Gütekriterium für sozialwissenschaftliche Messungen. Objektivität in Schulleistungsuntersuchungen ist gegeben, wenn für alle Schülerinnen und Schüler gleiche Aufgabenstellungen, Bearbeitungszeiten, Erläuterungen der Aufgaben, Arbeitsmaterialien u. Ä. gelten; die Auswertung und die Interpretation erfolgen nach klaren Kriterien, die unabhängig von der Person des Auswerter sind.

## **Prädiktor (-variable)**

unabhängige Variable,  $\Rightarrow$  Regressionsanalyse.

$r$  Abkürzung für Pearsons  $r$ ,  $\Rightarrow$  Korrelation.

## **(lineare) Regressionsanalyse**

Die (multiple) Regressionsanalyse ist ein Analyseverfahren, das den Zusammenhang zwischen einer  $\Rightarrow$  intervallskalierten abhängigen (zu erklärenden) Variable (dem so genannten Kriterium) und mehreren, ebenfalls intervallskalierten unabhängigen (erklärenden) Variablen (den so genannten



Prädiktoren) aufdeckt. Bei der Berechnung der Regressionsgleichung werden die  $\Rightarrow$  Korrelationen der Prädiktoren untereinander berücksichtigt. Die Maßzahl für den Zusammenhang zwischen allen Prädiktoren einerseits und dem Kriterium andererseits ist das „multiple R“. Es kann wie  $\Rightarrow$  Pearsons  $r$  interpretiert werden, kann allerdings keine negativen Werte annehmen. Das quadrierte multiple R entspricht der erklärten Varianz. Der eigenständige Beitrag jedes einzelnen Prädiktors (bei Konstanthaltung der anderen Prädiktoren) zur Aufklärung der Unterschiede im Kriterium wird mit den  $\Rightarrow$  Beta-Gewichten angegeben.

### **Reliabilität**

Reliabilität ist ein Gütekriterium für sozialwissenschaftliche Messungen. Sie entspricht der Zuverlässigkeit einer Messung. Reliabel ist ein Test oder eine Skala, wenn nur geringe Messfehler auftreten. Es gibt verschiedene Methoden, die Reliabilität zu überprüfen, z. B.  $\Rightarrow$  Cronbachs Alpha.

### **Signifikanz**

In der Statistik werden Unterschiede als signifikant bezeichnet, wenn sie mit einer bestimmten Wahrscheinlichkeit und nicht durch Zufall zustande gekommen sind. Der Grad der zu überprüfenden Unwahrscheinlichkeit wird vorher festgelegt und mit  $\alpha$  bezeichnet, beispielsweise  $\alpha = 0.05$  für 5 Prozent Irrtumswahrscheinlichkeit, 0,01 für 1 Prozent Irrtumswahrscheinlichkeit und 0,001 für 0,1 % Irrtumswahrscheinlichkeit.

### **Skala**

1. Kurzbezeichnung für die Ausprägungen einer Einschätzskala (Rating-skala). Bei der Vorgabe der Antwortmöglichkeiten von 1 = „trifft völlig zu“ bis 4 = „trifft überhaupt nicht zu“ im Schülerfragebogen spricht man z. B. von einer vierstufigen Skala.
2. Inhaltlich zusammenpassende Einzelitems können (z. B. durch Aufsummieren oder Mittelwertbildung) zu einer Skala zusammengefasst werden. Ein Beispiel ist die Skala „Schulzufriedenheit“, die auf dem Wege der individuellen Mittelwertbildung über elf Fragen aus dem Fragebogen für Schülerinnen und Schüler ermittelt wurde.

### **Standardabweichung, SD**

Die Standardabweichung ist ein so genanntes Streuungsmaß, das für intervallskalierte Daten Auskunft darüber gibt, wie homogen oder heterogen eine Merkmalsverteilung ist. Je kleiner die Standardabweichung ist, desto enger gruppieren sich die Werte der einzelnen Fälle um den Mittelwert, je größer sie ist, desto weiter streuen sie um den Mittelwert.

Liegt eine Normalverteilung vor, so lässt sich über die Verteilung Folgendes sagen: Im Bereich Mittelwert plus/minus eine Standardabweichung liegen etwa 68 Prozent der Fälle; im Bereich Mittelwert plus/minus zwei Standardabweichungen liegen etwa 95 Prozent der Fälle.

### **Streuung**

⇒ Standardabweichung.

### **Validität**

Validität ist ein Gütekriterium für sozialwissenschaftliche Messungen. Sie gibt die Gültigkeit eines Messinstruments (z. B. eines Tests) an. Ein Test ist valide, wenn er das misst, was er zu messen vorgibt. Validität lässt sich an Außenkriterien überprüfen; z. B. lassen sich die Aufgaben der fachbezogenen Leistungstests mit den Curricula der jeweiligen Schulform und Klassenstufe validieren (curriculare Validität).

### **Variable**

Ein erhobenes Merkmal, das mehrere Ausprägungen haben kann, wird als Variable bezeichnet, z. B. Alter oder erreichter Rasch-Skalenwert im Mathematiktest. Die Ausprägungen nennt man auch Merkmalsausprägungen.

### **Varianz**

Die Varianz ist das Quadrat der ⇒ Standardabweichung. Mathematisch ist die Varianz der Durchschnitt aus den quadrierten Abweichungen aller Einzelwerte vom Gesamtmittelwert.

### **Varianzanalyse**

Dieses Analyseverfahren ist mit der ⇒ Regressionsanalyse verwandt. Auch hier geht es um die Aufklärung individueller Unterschiede in einer abhängigen Variable durch eine oder mehrere unabhängige Variablen. Der Hauptunterschied zur Regressionsanalyse besteht darin, dass die unabhängigen Variablen nicht ⇒ intervallskaliert zu sein brauchen. Die Maßzahl für die erklärte Varianz heißt ⇒  $\eta^2$ .