



LANDESZENTRALE  
VERGLEICHSARBEITEN  
IM SAARLAND

Schuljahre 2003/04 – 2007/08

ABSCHLUSSBERICHT

Dr. Christoph Paulus  
Manuela Leidunger, M.A.  
Universität des Saarlandes

Unter Mitarbeit von Beatrice Egner, Lisa Dörr, Liane Fiore, Oliver Michaely, Steffen Oberheim, Kathrin Philippi, Angelica Ragusa, Melane Scazzari, Eva Schumacher, Judith Therre und Daniela Wagner

Impressum

Autoren:

Dr. Christoph Paulus

Manuela Leidinger, M.A.

Universität des Saarlandes

Empirische Humanwissenschaften

Fachrichtung 5.1 Erziehungswissenschaft

Projektträger:

Ministerium für Bildung, Familie, Frauen und Kultur

Hohenzollernstraße 60

66117 Saarbrücken

Dieses Projekt wurde finanziert durch das Ministerium für Bildung, Familie, Frauen und Kultur des Saarlandes



## Inhaltsverzeichnis

<b>1</b>	<b><i>EINLEITUNG</i></b>	<b>5</b>
1.1	Hintergrund und Bedeutung von Vergleichsarbeiten	5
1.2	Zum pädagogischen Nutzen der Vergleichsarbeiten	7
1.3	Zum Umgang mit Schulleistungsstudien	8
<b>2</b>	<b><i>LANDESZENTRALE VERGLEICH SARBEITEN (LZV) IN DER SEKUNDARSTUFE I DES SAARLANDES</i></b>	<b>10</b>
2.1	Konzeption und Umsetzung	10
2.2	Aufgabenerstellung und Itemanalysen	11
2.3	Beispielaufgaben	15
2.4	Korrekturverfahren	17
2.5	Rückmeldungen	17
2.5.1	Schulrückmeldung	19
<b>3</b>	<b><i>DIE ERGEBNISSE DER EINZELNEN JAHRE IM ÜBERBLICK</i></b>	<b>23</b>
<b>4</b>	<b><i>TESTLEISTUNGEN 2003/04</i></b>	<b>24</b>
4.1	Gymnasium	24
4.2	Erweiterte Realschule	24
4.3	Gesamtschule	24
<b>5</b>	<b><i>TESTLEISTUNGEN 2004/05</i></b>	<b>25</b>
5.1	Gymnasium	25
5.2	Erweiterte Realschule	25
5.3	Gesamtschule	25
<b>6</b>	<b><i>TESTLEISTUNGEN 2005/06</i></b>	<b>26</b>
6.1	Gymnasium	26
6.2	Erweiterte Realschule	26
6.3	Gesamtschule	26
<b>7</b>	<b><i>TESTLEISTUNGEN 2006/07</i></b>	<b>27</b>
7.1	Gymnasium	27
7.2	Erweiterte Realschule	27
7.3	Gesamtschule	27
<b>8</b>	<b><i>TESTLEISTUNGEN 2007/08</i></b>	<b>28</b>
8.1	Gymnasium	28
8.2	Erweiterte Realschule	28
8.3	Gesamtschule	28
<b>9</b>	<b><i>FOLGEN UND ENTWICKLUNGEN</i></b>	<b>29</b>
9.1	Bildungspolitische Folgen	29
9.1.1	Einstellung und Bereitschaft zur externen Evaluation geschaffen	29
9.1.2	Unterstützung von Problemschulen	29
9.1.3	„Ranking“	30

Landeszentrale Vergleichsarbeiten in der Sekundarstufe I	4
<b>9.2 Pädagogische Folgen</b>	<b>30</b>
9.2.1 Vergleichsarbeiten können Schwächen aufdecken, zu Diskussionen führen und in der Folge zu Verbesserungen	30
9.2.2 Teaching for the test	31
<b>10 LITERATUR</b>	<b>33</b>
<b>11 ANHANG</b>	<b>35</b>
<b>11.1 Leistungsübersicht ERS (Geschlecht, Migrationsstatus)</b>	<b>35</b>
<b>11.2 Leistungsübersicht GYM (nach Geschlecht und Migrationsstatus)</b>	<b>38</b>

# 1 Einleitung

Dieser Abschlussbericht bezieht sich auf den Zeitraum der Schuljahre 2003/04 – 2007/08. Der vorher durchgeführte Testdurchlauf im Fach Mathematik der Klassenstufe 8 wird hier nicht berücksichtigt.

## 1.1 Hintergrund und Bedeutung von Vergleichsarbeiten

Mit der Teilnahme an der dritten internationalen Mathematik- und Naturwissenschaftsstudie – The Third International Mathematics and Science Study (TIMSS) – im Jahr 1995 hat die Bundesrepublik Deutschland eine Wende hin zur systematischen Überprüfung der Erträge schulischer Bildungsprozesse eingeleitet. Zuvor gab es seit mehr als 20 Jahren keine deutsche Beteiligung an der internationalen Schulleistungsforschung.

Die TIMS-Studie untersuchte auf internationaler Ebene (mehr als 40 Nationen) die mathematischen und naturwissenschaftlichen Leistungen von Schülerinnen und Schülern der Primarstufe sowie der Sekundarstufe I und II, wobei sich die deutsche Beteiligung auf die Sekundarstufe I und II beschränkte. Es zeigte sich, dass die Testleistungen der deutschen Schülerinnen und Schüler im Bereich der mathematisch-naturwissenschaftlichen Grundbildung in der Gruppe vergleichbarer Länder im unteren Bereich lag (siehe Baumert et al., 1997), was eine enorme Resonanz in der Öffentlichkeit hervorrief. Im Jahre 1997 verabschiedete die Kultusministerkonferenz den Beschluss zur weiteren Teilnahme an übergreifenden Vergleichsuntersuchungen zur Erfassung des Lern- und Leistungsstandes der Schülerinnen und Schüler (siehe Konstanzer Beschlüsse). Hierbei wurden Schwerpunkte für die Bereiche der muttersprachlichen, mathematischen, naturwissenschaftlichen und fremdsprachlichen Kompetenzen formuliert.

Die Ergebnisse der ersten im Jahr 2000 durchgeführten PISA-Studie bestätigten die Ergebnisse der TIMS-Studie.

Als unmittelbare Konsequenz auf die Veröffentlichung dieser Ergebnisse formulierten die Kultusminister sieben Handlungsfelder, in denen die Länder als auch die Kultusministerkonferenz vorrangig tätig werden sollen. Diese sieben Handlungsfelder reichen von der Verbesserung der Sprachkompetenz der Migranten über die Förderung Benachteiligter bis zum Ausbau der Ganztagschulen (Abbildung 1).



Abbildung 1: Die sieben Handlungsfelder

In Hinblick auf die Einführung standardisierter Lernstandserhebungen ist Handlungsfeld 5 von maßgeblicher Bedeutung, da damit eine Forderung nach schulübergreifenden, vergleichenden Tests einhergeht.

Als vordergründigstes Ziel galt in diesem Kontext die Erarbeitung und Überprüfung gemeinsamer Standards im Rahmen der Schulbildung. Im Oktober 2004 erfolgte schließlich die Verabschiedung der Bildungsstandards für die Primarstufe in den Fächern Deutsch und Mathematik sowie für den Hauptschulabschluss in den Fächern Deutsch, Mathematik und Erste Fremdsprache<sup>1</sup>.

Die Bildungsstandards sind als „abschlussbezogene Regelstandards definiert und zielen bei einem mittleren Anforderungsniveau, das durch Aufgabenbeispiele veranschaulicht wird, auf systematisches und vernetztes Lernen und folgen so dem Prinzip des kumulativen Kompetenzerwerbs - Kompetenzen, die fachbezogen formuliert sind, die zugrunde liegende Wissensbestände nennen und die Schülerinnen und Schüler bis zu einem bestimmten Zeitpunkt ihres Bildungsgangs erreicht haben sollen.“ (Hofmann-Göttig et al., 2005). Während die Bil-

<sup>1</sup> Die Verabschiedung der Bildungsstandards für den mittleren Schulabschluss in den Fächern Biologie, Chemie und Physik erfolgte im Dezember desselben Jahres.

dungsstandards also die erwarteten Lernergebnisse definieren, überprüfen zentrale, standardorientierte Tests systematisch das Erreichen der Vorgaben. Diese Überprüfung stellt eine notwendige Ergänzung zu Bildungsstandards und Kernlehrplänen dar. Als Instrument der Überprüfung dienen Vergleichsarbeiten.

Im Saarland wurden im Schuljahr 2004/2005 die so genannten Landeszentralen Vergleichsarbeiten eingeführt, die ausgehend von testtheoretischen Standards erstmalig eine objektive und systematische Lernstandserhebung gewährleisteten. Sie geben Aufschluss darüber, ob und in welchem Umfang die formulierten Lernergebnisse zu einem bestimmten Zeitpunkt erreicht wurden, erheben dabei aber nicht den Anspruch kausal-analytisches Wissen zu erbringen. Ihre Aufgabe liegt vielmehr in der Erzeugung deskriptiven Wissens (Beschreibung der Schülerleistungen), weshalb aus den Ergebnissen der Landeszentralen Vergleichsarbeiten keine unmittelbaren Handlungshinweise abgeleitet werden können.

Dennoch ist jede Rückgabe von Wissen an Schulen als eine Eingabe von externem Wissen in Schulen zu verstehen. Wie mit diesem Wissen vor Ort, also in den Schulen selbst, umzugehen ist, wird im nächsten Kapitel erläutert.

## 1.2 Zum pädagogischen Nutzen der Vergleichsarbeiten

An dieser Stelle soll kurz erläutert werden, was Schulleistungsstudien, und somit Vergleichsarbeiten, leisten können und wollen. Generell gilt Schulleistung als ein Produkt einer Vielzahl von Bedingungsfaktoren, die unterschiedliche Ebenen umfassen und in komplexer Weise zusammenwirken (Hosenfeld, Schrader & Helmke, 2006). Die Komplexität dieses Zusammenspiels ist in Abbildung 2.1 exemplarisch anhand der Einflussfaktoren Persönlichkeitsmerkmale, sozioökonomischer Status, Vorwissen sowie elterlicher Erziehungsstil graphisch aufbereitet. Es handelt sich dabei um Faktoren, die sowohl die schulische Leistung als auch sich selbst gegenseitig beeinflussen.

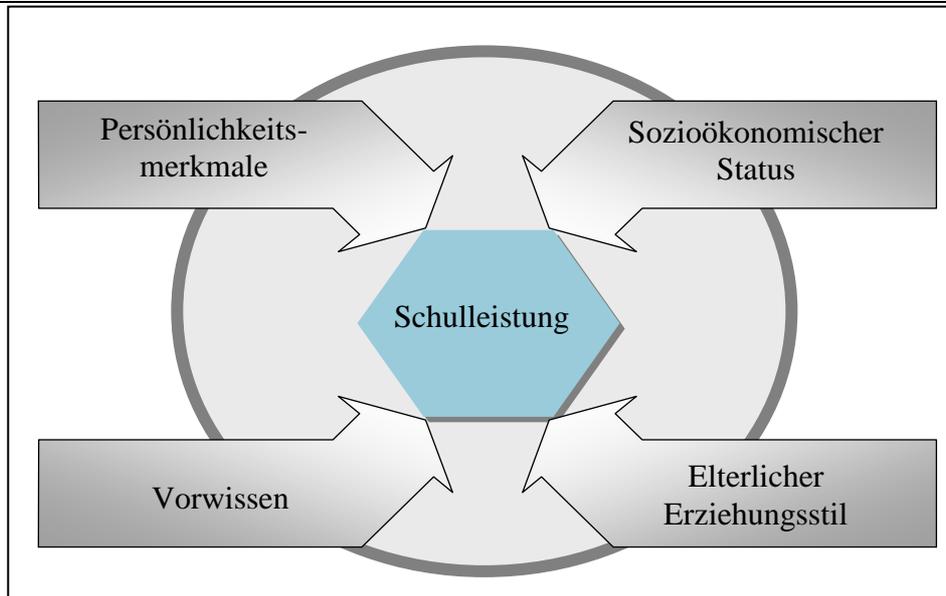


Abbildung 2: Bedingungsfaktoren von Schulleistung

Daher können mittels Vergleichsarbeiten keine Aussagen über die Qualität einzelner Schulen oder Klassen gemacht werden, da solche Urteile den Erfassungsradius dieser Instrumente übersteigen würde. Vergleichsarbeiten sind vielmehr als momentane Bestandsaufnahme zu verstehen, die Impulse für Schul- und Unterrichtsentwicklung geben und somit Veränderungsprozesse einleiten sollen/können (Hosenfeld, Schrader & Helmke, 2006). Anweisungen zum sinnvollen Umgang mit den gewonnenen Ergebnissen, werden im Folgenden gegeben.

### 1.3 Zum Umgang mit Schulleistungsstudien

Die gewonnenen Ergebnisse durchlaufen wichtige Stationen auf ihrem Weg von der Information (über Leistungen von Schülern, Klassen, Schulen) bis hin zum Ertrag für die Verbesserung des Lehrens und Lernens. Dabei gilt es zunächst, eine optimale Ausgangsbasis zu schaffen, indem die als Rückmeldung angebotenen Informationen in einem angemessenen Format dargestellt werden; d.h. die zurückgemeldeten Informationen sollten klassenspezifisch, reichhaltig, verständlich und zeitnah sein (Hosenfeld, 2005). Diese an das Rückmeldeformat gestellten Forderungen schaffen eine förderliche Ausgangssituation zur Rezeption der Ergebnisse. In einem zweiten Schritt gilt es dann, diese Informationen in den Schulen zu interpretieren und zu reflektieren. In dieser Phase spielen die Identifikation der Stärken und Schwächen

einer Klasse, die Bildung von Hypothesen sowie die Überprüfung der formulierten Annahmen eine zentrale Rolle. Im Anschluss an diese Prozesse muss die Aktion, also die konkrete Handlung erfolgen; d.h. Schritte zu Verbesserung und Optimierung eingeleitet werden.

Es handelt sich dabei nicht um automatisch ablaufende Prozesse. Persönliches Engagement und Interesse von Seiten der Lehrerschaft sowie der Schulleitung sind in diesem Zusammenhang gefragt. Dabei gilt zu beachten, dass eine Motivation zur Veränderung dann zu erwarten ist, wenn Ergebnisse unbefriedigend ausfallen, aber als bedeutsam eingeschätzt und Erfolge als in der Kontrolle des Handelnden liegend wahrgenommen werden und der antizipierte Nutzen höher als die damit verbundenen materiellen oder psychischen Kosten. Änderungsmaßnahmen werden dann eingeleitet, wenn der Nutzen die Kosten übersteigt (Kosten-Nutzen-Bilanz). Als Fazit lässt sich festhalten, dass die „Rezeption und Nutzung von rückgemeldeten Evaluationsergebnissen ein sozial, organisatorisch und psychologisch hochkomplexer Vorgang ist, der offenbar nur dann zum Erfolg führt, wenn verschiedene Voraussetzungen gegeben sind“ (Hofmann-Göttig, Eschmann & Daumen, 2005, S. 35). Vergleichsarbeiten als Instrument der Schulentwicklung liefern bezogen auf die Bildungsstandards einen fremden Blick auf die standortgebundenen Leistung einzelner Schulklassen und Schulen.

## 2 Landeszentrale Vergleichsarbeiten (LZV) im Saarland

### 2.1 Konzeption und Umsetzung

Studien wie PISA und IGLU operieren primär auf internationaler Ebene und streben Leistungsvergleiche zwischen Nationen an. Dies impliziert, dass mit Stichproben gearbeitet werden kann. In der hier vorliegenden Lernstandserhebung wird ein anderer Weg beschritten, indem alle Klassen dreier Klassenstufen im Rahmen mehrerer Totalerhebung verpflichtend an dieser Untersuchung teilnehmen. Dies bringt erhebliche Vorteile mit sich:

- *Durch die Testung der Grundgesamtheit sind, methodisch betrachtet, keine statistischen Tests zur Generalisierung von Stichproben erforderlich.*
- *Die Auswertungen sind zwar (nur) landesbezogen interpretierbar, lassen aber wegen der hohen Teilnehmerzahl Rückschlüsse zu, die in vielerlei Hinsicht auch auf andere Bundesländer übertragbar sein können.*
- *Selektionseinflüsse sind ausgeschlossen. Bei freiwilliger Teilnahme wird eine Stichprobe oftmals dadurch verzerrt, dass sich „schlechte“ Schulen in der Regel nicht an Vergleichsstudien beteiligen, die Ergebnisse überschätzen eher den Bildungsstand der Schülerinnen und Schüler.*

Die Landeszentralen Vergleichsarbeiten wurden im Saarland zwischen 2004 bis 2008 in folgenden Schulformen und Fächern geschrieben (Tabelle 1).

Tabelle 1: Übersicht Landeszentrale Vergleichsarbeit

Klassenstufe	Schulform	Fach
7	Erweiterte Realschule	Deutsch
	Gesamtschule	
	Gymnasium	
8	Erweiterte Realschule hier: H- und M-Zweig	Mathematik
	Gesamtschule Grund- und Erweiterungskurs	
	Gymnasium	
9	Erweiterte Realschule: M-Zweig	erste Fremdsprache
	Gesamtschule: Erweiterungskurs	
	Gymnasium	

Die Testergebnisse wurden statistisch ausgewertet und den Schulen zurückgemeldet.

Für die Konstruktion der Aufgaben verantwortlich waren Fachkommissionen, die sich zusammensetzten aus Lehrkräften der jeweiligen Fächer. Die inhaltliche Konzeption der Lan-

deszentralen Vergleichsarbeiten basierte auf den jeweiligen Lehrplänen für die Klassenstufen 7 bis 9. Die Aufgabenkonstruktion durch die Fachkommissionen gewährleistet eine sehr hohe inhaltliche Validität der Messinstrumente, was den Aussagegehalt der Ergebnisse dieser Studie untermauert.

Die Intention der Untersuchung lag darin, den Schulen ein Diagnoseinstrument zur Verfügung zu stellen, das als Orientierungshilfe dienen soll, anhand derer Lehrkräfte fundierte Aussagen über die Leistungen der Schülerinnen und Schüler erhalten und mittels dessen sie den Lern- und Leistungsstand ihrer Klasse in einem überregionalen Vergleich einordnen können. Das Ergebnis der Landeszentralen Vergleichsarbeiten soll die Lehrkräfte bei der Gestaltung ihres Unterrichts unterstützen und ihnen die Möglichkeit bieten, die Leistungen der einzelnen Schülerinnen und Schüler in einem breiteren Vergleichsrahmen zu beurteilen.

## 2.2 Aufgabenerstellung und Itemanalysen

Mit der Vorbereitung der Landeszentralen Vergleichsarbeiten 2004/05 wurde die Prätistung der Testaufgaben („Items“) eingeführt. Dies sollte dazu dienen, die Aufgaben- und damit die Testschwierigkeit der zu erstellenden Vergleichsarbeit in einem ausgewogenen Schwierigkeitsgrad zu konzipieren. Ausgewogen bedeutet dabei die gleichmäßige Menge an leichten („Eisbrecher-Items“) und schwierigen Testaufgaben. Insbesondere unter Berücksichtigung der Tatsache, dass bis zum Schuljahr 2006/07 die LZVen als eine von fünf Klassenarbeiten gewertet wurde, sollte sich die zu erwartende Bewertung nicht allzu weit von den bisherigen Leistungsergebnissen der Schüler wegbewegen, um deren fachliche Leistungsmotivation nicht allzu sehr zu vermindern.

Der Itempool, bestehend auch ca. der doppelten Anzahl von Testaufgaben des später zu realisierenden Umfangs wurde uns von den Fachkommissionen zur Verfügung gestellt. Anhand einer von uns aufgrund der Ergebnisse der letztjährigen Vergleichsarbeiten vorausgewählten, repräsentativen Stichprobe wurden alle Aufgaben von uns auf ihre Itemschwierigkeit, Trennschärfe (auf Wunsch) und Verstehbarkeit geprüft. Dabei wurde nach folgendem Muster (Beispiel Fach Mathematik) verfahren:

- Testpopulation für Mathematik sind Schüler der jetzigen Klasse 8; wissenschaftlich gesehen bestehen keine Bedenken an diesem Vorgehen.
- Je nach Zahl der zu testenden Aufgaben werden je 18 GYM- bzw. ERS/GeS-Klassen benötigt (entspricht ca. 100 Schülern pro Aufgabe)
- Durchführung wird unterstützt durch Testleiter der Universität

- alle Aufgabenentwürfe der Kommission werden getestet (z.B. 30 Aufgaben)
- der geschätzte Bearbeitungsaufwand pro Aufgabe sollte mitgeliefert werden
- jeder Schüler erhält ca. 1/6 der Aufgabenentwürfe (z.B. 5 Aufgaben) plus 5 Füllaufgaben (s. Abbildung)
- Coverstory: Die Schüler sollen die subjektive Aufgabenschwierigkeit und den Anreizcharakter der Aufgabe einschätzen sowie die Aufgaben lösen:
  - *„Aufgabenstellungen und Aufgaben sind für Schüler/-innen oftmals wenig abwechslungsreich und realitätsfern. Aus diesem Grunde haben wir neue Aufgabenformen entwickelt, die wir heute mit deiner Hilfe erproben möchten. Für uns ist deine Meinung wichtig, wie du mit diesen Aufgaben zurecht kommst und ob sie für dich spannend und interessant sind. Wir geben dir jetzt diese neuen Aufgaben und bitten dich darum, dass du jeweils vor und auch nach der Bearbeitung deine persönliche Meinung / Einschätzung zur Schwierigkeit und Qualität der Aufgabenform abgibst. Bitte achte darauf, dass du keine Frage bzw. Aufgabe vergisst! Alle deine Angaben werden natürlich anonym und vertraulich behandelt. Niemand außer uns, auch nicht dein Lehrer / deine Lehrerin, wird Einblick in deine Antworten haben. Diese werden weder bewertet noch benotet, uns interessiert lediglich deine Meinung!“*

Nach der Vorauswertung konnten den Fachkommissionen somit sehr exakte Aussagen über den Schwierigkeitsgrad der von ihnen entwickelten Aufgaben sowie teilweise Verbesserungsvorschläge in Aufbau und Formulierung gegeben werden. Zudem wurden Gesamteinschätzungen der Aufgabenblöcke erstellt, die zur Unterstützung des Verständnisses zudem noch farblich gekennzeichnet waren: rot für schwere, gelb für mittelschwere und grün für leichte Aufgaben (s.u.).

Die folgenden Beispiele zeigen Ausschnitte aus Prätest-Rückmeldungen:

### Detailanalysen<sup>2</sup>

Mögliche Lösungen	40,00 €	45,00 €	50 €	60,00 €	nicht bearbeitet
Aufgabe 1.3	0 %	65,3 %	0%	32,7 %	
Gesamtschwierigkeit	<b>0,45</b>				

Keine Probleme der Schüler/-innen

---

<sup>2</sup> Anhand der Farbdarstellung konnte die Aufgabenschwierigkeit leicht klassifiziert werden: grün = leicht ( $p > .75$ ), gelb = mittelschwer ( $.74 < p < .35$ ) und rot = schwer ( $p < .35$ )

## Aufgabe 2: Stammbrüche

	richtig gelöst	teilweise richtig <sup>3</sup>	falsch gelöst	nicht bearbeitet
Aufgabe 2.1	75,5 %	0 %	22,4 %	2,0 %
Aufgabe 2.2	28,6 %	0 %	44,9 %	26,5 %
Aufgabe 2.3	40,8 %	4,1 %	38,8 %	16,3 %
Gesamtschwierigkeit	0,483			

Probleme der Schüler/-innen:

- ❖ 34,7 % der Schüler/-innen gaben an, mit der Aufgabe Probleme gehabt zu haben. Problembeschreibung: Der Begriff „Stammbruch“ wurde nicht verstanden.

Probleme der Schüler/-innen:

- ❖ 17,5 % gaben an, Probleme mit dieser Aufgabe gehabt zu haben.
  - Verständnisprobleme bei b und c

## Aufgabe 1: Dialogues au téléphone

Mögliche Lösungen	Anrufer wird weiter verbunden	Anrufer wird sich wieder melden	Anrufer wird zurückgerufen	Anrufer hinterlässt eine Nachricht	nicht bearbeitet
Aufgabe 1.1	7,7 %	0	15,4 %	76,9 %	-
Aufgabe 1.2	50 %	17,3 %	21,2 %	5,8 %	5,8
Aufgabe 1.3	5,8 %	80,8 %	11,5 %	0	1,9
Aufgabe 1.4	9,7 %	0	78,8 %	11,5 %	-
Aufgabe 1.5	3,8 %	28,8 %	46,2 %	21,2 %	-

Probleme der Schüler/-innen:

- ❖ Text war schwer zu verstehen;
- ❖ Vokabelprobleme;
- ❖ CD hat nicht richtig funktioniert (technischer Fehler vor Ort);
- ❖ Es wurde zu schnell und undeutlich gesprochen

<sup>3</sup> Aufgabe 2.3 : Richtig überprüft, aber falsch ausgerechnet → falsches Ergebnis.

**Trennschärfen Aufgaben 1-3 (Französisch):**

	<i>Fr1bis3_Summe</i>
A1a_rf	<b>,582(**)</b>
A1b_rf	<b>,647(**)</b>
A1c_rf	<b>,231</b>
A1d_rf	<b>,677(**)</b>
A1e_rf	<b>,480(**)</b>
A1f_rf	<b>,470(**)</b>
A2.1_rf	<b>,475(**)</b>
A2.2_rf	<b>,604(**)</b>
A2.3_rf	<b>,392(*)</b>

	<i>Fr1bis3_Summe</i>
A2.4_rf	<b>,579(**)</b>
A2.5_rf	<b>,282</b>
A2.6_rf	<b>,193</b>
A2.7_rf	<b>,193</b>
A3.1_rf	<b>,349(*)</b>
A3.2_rf	<b>,521(**)</b>
A3.3_rf	<b>,320</b>
A3.4_rf	<b>,340(*)</b>

\* Die Korrelation ist auf dem Niveau von 0,01 (2-seitig) signifikant.

\* Die Korrelation ist auf dem Niveau von 0,05 (2-seitig) signifikant.

Die Prätistung führte zu einer sehr hohen Genauigkeit der Vorhersage der Schülerleistungen, wie die folgende Tabelle beispielhaft demonstriert:

Tabelle 2: Gegenüberstellung der Prättest-Schätzungen und der tatsächlichen Aufgabenschwierigkeit

Aufgabennummer	Aufgabenname	Themenbereich	Max. Punkte	Mittelwert	Streuung	Prättest-Einschätzung	Test-Schwierigkeit
Part 1	Good News for Grizzlies	Leseverstehen	6	3,21	1,72	0,51	0,54
Part 2	NASA Planning Travels to Moon and Mars		6	4,00	1,49	0,65	0,67
Part 3	Drowning the money		6	3,24	1,81	0,59	0,54
Part 4	Short news headings		6	4,95	1,04	0,78	0,82
Aufgabe 1.1.a	ganze Zahlen	K5-L1-A1	0,5	0,30	0,25	0,63	0,59
Aufgabe 1.1.b	gewöhnliche Brüche	K5-L1-A1	1	0,78	0,39	0,73	0,78
Aufgabe 1.1.d	Dezimalbruch	K5-L1-A1	1	0,31	0,36	0,29	0,31
Aufgabe 3.2	Bruchteil DIN-A4-Blatt	K2-L1-A1	0,5	0,31	0,24	0,65	0,62

## 2.3 Beispielaufgaben

Art und Aussehen der Aufgaben haben im Testzeitraum immer wieder Veränderungen erfahren. So wurden die Kommissionen zunehmend erfahrener und kreativer in der Konstruktion von Testitems. Einige besonders gelungene Beispiele seien hier aufgezeigt:

**Aufgabe 2****3 Punkte**

12 x 0,75 l  
**2,16 €**  
 zuzüglich 3,30 € Pfand

Du kaufst in einem Einkaufszentrum einen Kasten Sprudel.

- a) Wie viel Geld musst du an der Kasse bezahlen?

- b) Sprudel wird häufig in unterschiedlichen Flaschengrößen verkauft. Um die Preise vergleichen zu können, muss immer der Preis für 1 Liter angegeben werden. Wie viel kostet 1 Liter dieser Sprudelsorte?



9 €



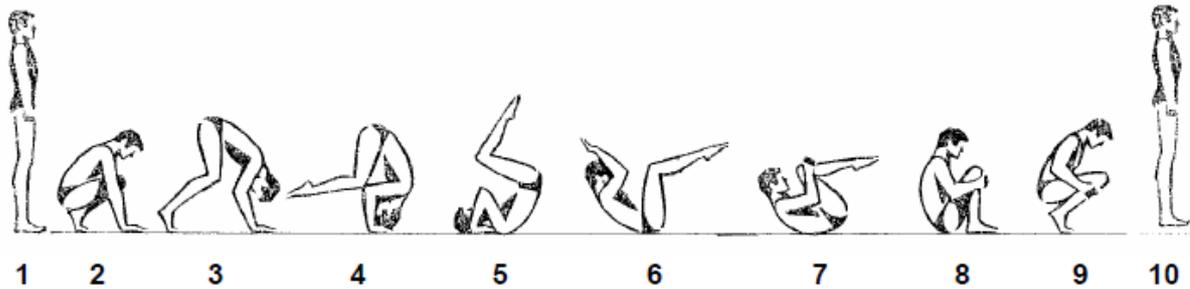
8 €



7 €


 €

Welcher Gesamtpreis gehört in das Kästchen unter der vierten Abbildung?  
 Schreibe deine Überlegungen auf.

**Aufgabe****40 Punkte**

Die Abbildungen zeigen den Ablauf einer „Rolle vorwärts“.

Schreibe anhand der Bilder eine Vorgangsbeschreibung des gesamten Ablaufs dieser Sportübung. Achte auf die Haltung bzw. Stellung der verschiedenen Körperteile.

Hinweis: Beschränke die Darstellung auf den Hauptteil der Beschreibung, du musst also keine Einleitung und keinen Schluss verfassen.

Noémie voit son père toutes les deux semaines.		
	vrai	<input type="checkbox"/>
	faux	<input type="checkbox"/>
	ce n'est pas dit dans le texte	<input type="checkbox"/>
Elle voudrait vivre avec son père.		
	vrai	<input type="checkbox"/>
	faux	<input type="checkbox"/>
	ce n'est pas dit dans le texte	<input type="checkbox"/>
Sa mère est contre les copains de Noémie.		
	vrai	<input type="checkbox"/>
	faux	<input type="checkbox"/>
	ce n'est pas dit dans le texte	<input type="checkbox"/>

## 2.4 Korrekturverfahren

Die Auswertung der Landeszentralen Vergleichsarbeiten umfasste die Erfassung und Bearbeitung der Schülerleistungen in den Bereichen Deutsch, Mathematik und Erste Fremdsprache. Dies bedeutete die computerunterstützte Eingabe und Auswertung von jährlich ca. 25.000 Schülerprofilen (Erfassung von Schülerdaten und Leistungsergebnissen mittels SPSS sowie Erstellung von Schul- und Klassenrückmeldungen). Die Vergleichsarbeiten wurden wie in den Vorjahren auch von den jeweiligen Lehrkräften der Klassen und Kurse korrigiert. Die Testergebnisse wurden in dafür vorgesehen Ergebnislisten eingetragen und an unsere Fachrichtung weitergeleitet. Diese Listen waren in großem Maße korrekt und vollständig. Dennoch gab es einzelne Lehrkräfte, die Teilergebnisse summierten oder in ihrer Punktevergabe vom Lösungsschlüssel abwichen und beispielsweise Teilpunkte vergaben. Hier musste postalisch nachgefragt und auf die Zusendung der verbesserten oder ergänzten Listen gewartet werden, wodurch ein unnötiges Zeitfenster in der eigentlichen Auswertungsphase entstand, welches die Rückmeldung an die Schulen zeitlich verzögerte.

## 2.5 Rückmeldungen

Rückmeldungen wurden sowohl für die Schulleiter/-innen als auch für die jeweiligen Klassenlehrer/-innen angefertigt. Dabei beinhaltete die Rückmeldung an die Schulleitung<sup>4</sup> zunächst allgemein-deskriptive Informationen über die teilnehmenden Schülerpopulationen. Im Ergebnisteil wurden die Leistungen der einzelnen Klassen der Schule je Fach dargestellt sowie eine Einordnung der Leistungen nach den Kriterien „deutlich unter dem Durchschnitt“, „unter dem Durchschnitt“, „im Durchschnitt“, „über dem Durchschnitt“ und „deutlich über dem Durchschnitt“ vorgenommen.

Pädagogisch relevant an dieser Positionierung erscheint der Punkt, dass somit objektive und unabhängige Leistungseinschätzungen ermöglicht werden. Diese Resultate bedürfen allerdings eines Vergleichsmaßes, ohne das die erhobene Leistungsrückmeldung keine wirkliche Aussagekraft besitzt. Aus pädagogischer Sicht sollten damit die Überlegungen angestoßen werden, dass Lernerfolg als ein Produkt aus Lehrleistung, Schülermerkmalen und lokalen sowie situativen Bedingungen (Schulgröße, Schülerpopulation, Klassengröße, Lehrerzahl

---

<sup>4</sup> Beispielhafte Rückmeldungen für Schulleitung und Klassenlehrer/-in sind im Anhang beigefügt.

etc.) gesehen werden muss. Die Auswerter können nur objektiv Leistungsergebnisse zurückmelden, die Interpretation der Ergebnisse obliegt den Schulen und Lehrkräften vor Ort. Allerdings wird damit nicht automatisch die Variable Lehrerleistung unabhängig von allen anderen Variablen gedeutet. Es können aber beispielsweise ungünstige Ergebnisse zum Anlass genommen werden, bei den Schulträgern Verbesserungen einzufordern. Dazu sollte eine schulinterne Ist-Soll-Analyse durchgeführt werden, die Fragen wie die folgenden zu beantworten versucht:

- *Was funktioniert gut?*
- *Was kann verbessert werden?*
- *Wo ist noch Potential vorhanden?*
- *Wo wird evtl. Hilfe von außen (LPM, Ministerium) benötigt?*

Vergleiche zwischen verschiedenen Klassen werden prinzipiell als unfair erlebt, da das individuelle Leistungsergebnis der Klasse sowohl von Merkmalen des Lehrmaterials und Kompetenz der unterrichtenden Lehrkraft als auch von Merkmalen der Schülerschaft beeinflusst wird. Es ist daher nahe liegend, auch die sozialen Hintergründe der Schülerschaft zu berücksichtigen, um einen „fairen Vergleich“ zu gewährleisten. Die Landeszentralen Vergleichsarbeiten werden diesem Anspruch gerecht, indem für die Population der Erweiterten Realschulen und Gesamtschulen Bezugsgruppen generiert werden, innerhalb derer leistungsrelevante Rahmenbedingungen erfasst werden, die von der unterrichtenden Lehrkraft nicht unmittelbar verändert werden können. Im Rahmen der Landeszentralen Vergleichsarbeiten erfolgt auf Schulebene sowohl ein Vergleich der einzelnen Schulen mit der Gesamtgruppe (Landesergebnis) als auch ein Vergleich mit einer Gruppe homogener Schulen (Schulcluster), deren durchschnittliche Testleistungen einen Referenzrahmen bilden und einen Orientierungspunkt für die einzelnen Schulen. Auf Klassen- bzw. Kursebene erfolgt ein Vergleich mit so genannten Vergleichsklassen (-kursen). Diese Clusterung erfolgte anhand der Variablen Bildungsbackground, Migrationsanzahl und geschätzter Sprachfähigkeiten der Schüler.

Eine Übersicht über die verschiedenen Cluster zeigt die folgende Tabelle:

Tabelle 3: Übersicht über die Clusterinhalte zur Bildung von Vergleichsschulen

<i>Cluster</i>	<i>Interpretation: Klassen mit ...</i>	<i>Anteil C-Kinder</i>	<i>Bildungsbackground</i>	<i>Migrantenzahl</i>
1	... Schülern mit guten D-Kenntnissen, überdurchschnittlichem Bildungsbackground, aber mit hohem bis sehr hohen Migrantenanteilen	< 10%	Stufe 4	> 10%
2	... Schülern mit guten bis befriedigenden D-Kenntnissen, durchschnittlichem Bildungsback-	< 20%	Stufe 3	< 20%

Cluster	Interpretation: Klassen mit ...	Anteil C-Kinder	Bildungsbackground	Migrantenzahl
	ground und hohen Migrantenanteilen			
3	... Schülern mit guten bis befriedigenden D-Kenntnissen, durchschnittlichem Bildungsbackground und hohen Migrantenanteilen	< 20%	Stufe 3	> 20%
4	.... Schülern mit guten bis befriedigenden D-Kenntnissen, hohem bis durchschnittlichem Bildungsbackground und niedrigen Migrantenanteilen	< 20%	Stufe 3-4	< 10%
5	.... Schülern mit schlechten D-Kenntnissen, hohem Bildungsbackground und hohen Migrantenanteilen	> 20%	Stufe 4	< 20%
6	.... Schülern mit guten bis befriedigenden D-Kenntnissen, hohem Bildungsbackground und hohen bis sehr hohen Migrantenanteilen	< 20%	Stufe 4	> 10%
7	.... Schülern mit überwiegend guten D-Kenntnissen, sehr hohem Bildungsbackground und niedrigen Migrantenanteilen	< 20%	Stufe 5	< 10%
8	.... Schülern mit schlechten D-Kenntnissen, mittlerem Bildungsbackground und sehr hohen Migrantenanteilen	> 20%	Stufe 3	> 20%
9	.... Schülern mit überwiegend schlechten D-Kenntnissen, niedrigem Bildungsbackground und sehr hohen Migrantenanteilen	> 10%	Stufe 2	> 20%

### 2.5.1 Schulrückmeldung

Die Informationen für die **Schulleiter** beinhalteten neben allgemeinen schulspezifischen Informationen (vgl. Abbildung 3) auch die Leistungsergebnisse der Schule als Übersicht (vgl. Abbildung 4).

#### Allgemeine schulbezogene Informationen:

	Deutsch	Mathematik (H-Bildungsgang)	Mathematik (M-Bildungsgang)	Französisch	Englisch
Anzahl der Klassen	2	2	2	1	2
Anzahl der Schüler	50	31	34	15	44
Anzahl der Jungen	26	20	21	6	18
Anzahl der Mädchen	24	11	13	9	26
Anzahl der Schüler mit nicht-deutscher Muttersprache	4	3	0	3	3

Abbildung 3: Allgemeine Informationen der Schulrückmeldung

**Testleistungen Ihrer Schule:**

	Erreichbare Punkte	Durchschnittliche Testleistung	Durchschnittliche Testleistung in %	Standardabweichung	„Normalbereich“ <sup>1</sup>
Ihre Schule – Deutsch	58	40,58	69,97	7,74	[32,84 ; 48,32]
Landesdurchschnitt		38,60	66,55	8,10	[30,50 ; 46,70]
Ihre Schule – Mathematik (H)	25	14,50	58,00	3,75	[10,75 ; 18,25]
Landesdurchschnitt		13,61	54,44	3,94	[9,67 ; 17,55]
Ihre Schule – Mathematik (M)	32	17,06	53,31	3,73	[13,33 ; 20,79]
Landesdurchschnitt		17,05	53,28	4,48	[12,57 ; 21,53]
Ihre Schule – Französisch	60	30,70	51,17	4,96	[25,74 ; 35,66]
Landesdurchschnitt		33,09	55,15	8,99	[24,10 ; 42,08]
Ihre Schule – Englisch	90	45,59	50,66	13,62	[31,97 ; 59,21]
Landesdurchschnitt		54,47	60,52	13,08	[41,39 ; 67,55]

Abbildung 4: Testleistungen der Schule in der Schulrückmeldung

Zusätzlich erhielten sie Informationen über alle an den Vergleichsarbeiten beteiligten Klassen und deren Leistungsergebnisse:

**Testleistungen der einzelnen Klassen Ihrer Schule:**

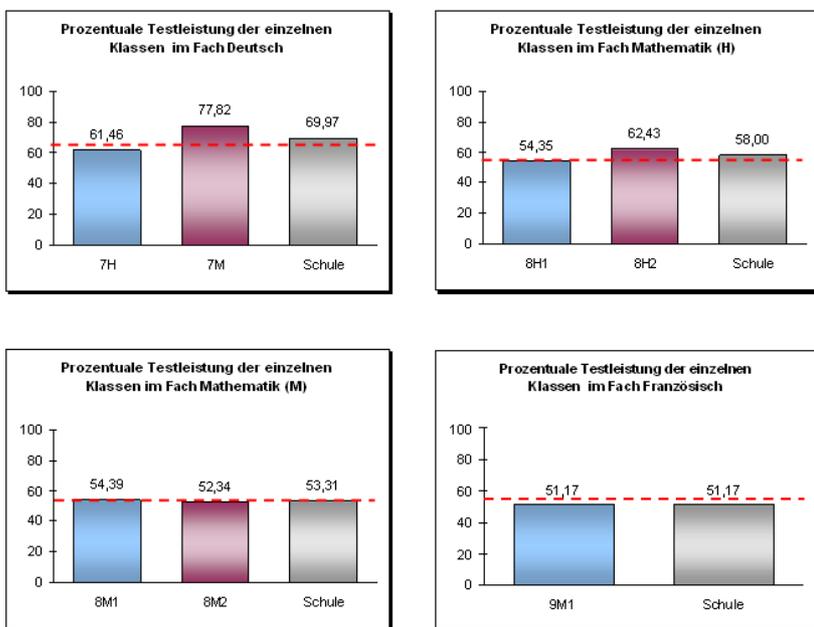


Abbildung 5: Leistungsübersicht über Klassenergebnisse in der Schulrückmeldung (Ausschnitt)

Die über alle Klassenergebnisse gemittelten Leistungswerte wurden dann in der sog. Positionierung der Schule beschrieben, in der der Schulleiter die relative Position seiner Schule im Landesvergleich

erkennen konnte. Da ein Schulranking weder gewünscht noch methodisch sinnvoll erstellt werden konnte (die Heterogenität der Fächer, Klassenstufen und Schülerstrukturen lässt eine sinnvolle Interpretation eines solchen Rankings inhaltlich nicht zu), wurde lediglich die Position im Gesamtergebnis aller vergleichbaren Schulen dargestellt.

**Positionierung Ihrer Schule:**

Im Gesamtergebnis aller vergleichbaren Erweiterten Realschulen / Gesamtschulen (Schulcluster) liegt Ihre Schule ...

	Deutsch	Mathematik (H)	Mathematik (M)	Französisch	Englisch
deutlich unter dem Durchschnitt					
unter dem Durchschnitt				x	x
im Durchschnitt	x	x	x		
über dem Durchschnitt					
deutlich über dem Durchschnitt					

Abbildung 6: Positionierung der Schule

Die Rückmeldung an **den/die Klassenlehrer/-in** beinhaltet im ersten Teil eine kurze deskriptive Darstellung der Schüler sowie eine grafische Darstellung der Notenverteilung, die im letzten Jahr (weil keine Bewertung stattfand) in prozentualer Zielerreichung ausgedrückt wurde.

In Ihrer Klasse nahmen an der diesjährigen Landeszentralen Vergleichsarbeit im Fach Mathematik (H-Bildungsgang) 17 Schüler<sup>1</sup> teil. Davon waren 6 weiblich und 11 männlich<sup>2</sup>.

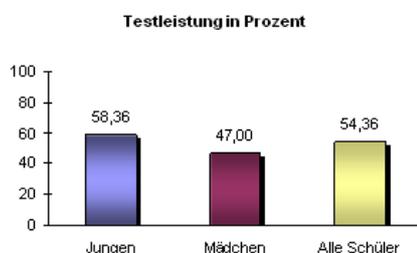


Abbildung 7: Testleistungen der Schüler einer Klasse

Darstellung der Ergebnisse Ihrer Klasse in Prozent

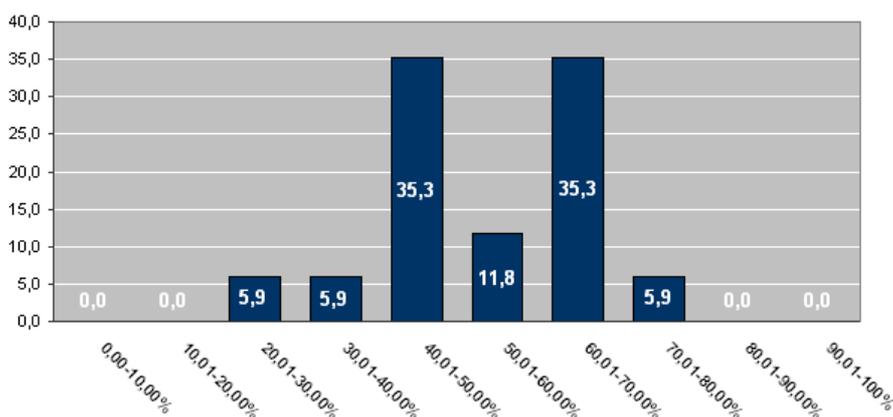


Abbildung 8: Prozentuale Leistungsergebnisse einer Klasse

Daran anschließend wurde die durchschnittliche Klassenleistung für jede der gestellten Aufgaben in Form einer Stärken-Defizit-Analyse abgebildet. Als „Stärken“ wurden Leistungen definiert, bei denen mindestens 75% der maximalen Punktzahl erreicht wurden, als „Schwächen“ galten Leistungen, die unter 50% der maximalen Punktzahl lagen. Leistungen zwischen 50% und 75% entsprechen dem Durchschnitt und wurden nicht explizit gekennzeichnet.

Kompetenzen	Aufgaben	Max. Punkte	Stärke	Defizit	$\bar{x}$ Klasse / Kurs	$\bar{x}$ Schule	$\bar{x}$ Vgl. Klassen <sup>2</sup>	$\bar{x}$ ERS / GeS
K 1 Mathematisch argumentieren	1b,1c	2			1,00	1,03	1,09	1,13
K 2 Probleme mathematisch lösen	2a, 2b, 3, 7	5,5	x		4,47	4,40	4,20	4,14
K 3 Mathematisch modellieren	4, 6b, 9, 11a, 11b	8,5		x	3,74	4,39	4,19	4,22
K 4 Mathematische Darstellungen verwenden	6a, 6c, 8	3			1,68	1,71	1,51	1,47
K 5 Mit symbolischen, formalen und technischen Elementen der Mathematik umgehen	1a, 5, 10a, 10b, 10c	6		x	2,71	2,97	2,70	2,66
$\Sigma$		25			13,59	14,50	13,69	13,61

Abbildung 9: Stärken-Defizit-Analyse einer Klasse (Ausschnitt)

Neben der inhaltlichen Zuordnung der Einzelaufgaben zu ihrem Kompetenz- bzw. Anspruchsniveau gemäß der Definition der Bildungsstandards des Faches wurde die Klassenleistung in Bezug zur Schul- und Clusterleistung („Vergleichsklassen“) sowie im Vergleich zu den Resultaten der Schule, der Vergleichsklassen sowie dem Landesdurchschnitt gesetzt (s. Abbildung 10).

### Zusammenfassung der Klassenleistungen

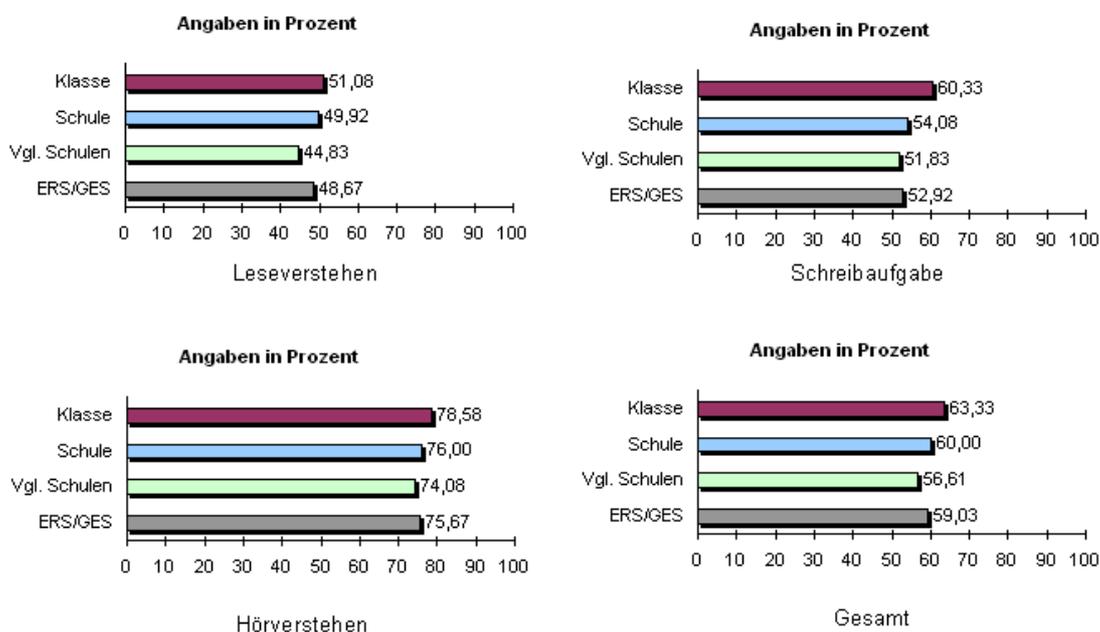


Abbildung 10: Klassenleistung mit Vergleichswerten

### 3 Die Ergebnisse der einzelnen Jahre im Überblick

Für die Vergleichsarbeiten im Zeitraum 2003/04 bis 2007/08 liegen diesbezüglich folgende allgemein deskriptive Daten vor:

<b>Schuljahr</b> <b>Deskriptive Daten</b>	<b>I</b> <b>2003/04</b>	<b>II</b> <b>2004/05</b>	<b>III</b> <b>2005/06</b>	<b>IV</b> <b>2006/07</b>	<b>V</b> <b>2007/08</b>
<b>Anzahl teilnehmender Schulen</b>					
Gymnasien	33	33	33	33	32
Erweiterte Realschulen	54	53	50	54	51
Gesamtschule	16	15	14	14	14
Gesamt	103	101	97	101	97
<b>Anzahl teilnehmender Klassen Gymnasium</b>					
Deutsch	135	125	129	126	128
Mathematik	137	134	131	130	124
Französisch	81	74	71	65	66
Englisch	49	48	55	51	49
Latein	7	11	10	11	10
Gesamt	482	392	396	383	377
<b>Anzahl teilnehmender Klassen Erweiterte Realschule</b>					
Deutsch	230	222	194	192	184
Mathematik	246	238	215	219	195
Französisch	82	85	71	70	67
Englisch	51	49	44	52	48
Gesamt	609	594	524	533	494
<b>Anzahl teilnehmender Klassen Gesamtschule</b>					
Deutsch	67	67	65	66	63
Mathematik	96	92	96	91	87
Französisch	24	21	21	22	18
Englisch	29	30	30	32	29
Gesamt	216	210	212	211	197
<b>Anzahl teilnehmender Schüler:</b>					
Jungen	14.520	14.239	12.964	12.969	12.207
Mädchen	14.532	14.112	12.957	12.925	11.998
Schüler mit nicht-deutscher Muttersprache		2.460	2.150	2.146	2.016
Gesamt <sup>5</sup>	29.052	28.464	26.314	26.017	24.308

<sup>5</sup> Bei der Gesamtzahl wurden auch die Schüler berücksichtigt, bei denen die Angaben zu Geschlecht fehlten.

## 4 TESTLEISTUNGEN 2003/04

### 4.1 Gymnasium

	Anzahl Aufgaben	Maximale Punktzahl	Mittelwert	SD
Deutsch	6	65	46,41	8,20
Mathematik	17	30	16,70	5,54
Französisch	10	36	30,02	3,76
Englisch	9	36	30,25	3,23
Latein	17	61	42,98	9,41

### 4.2 Erweiterte Realschule

	Anzahl Aufgaben	Maximale Punktzahl	Mittelwert	SD
Deutsch	12	45	32,14	5,64
Mathematik (H-Bildungsgang)	13	30	14,90	5,49
Mathematik (M-Bildungsgang)	15	30	14,94	4,55
Französisch	10	36	23,13	5,29
Englisch	8	36	23,82	5,04

### 4.3 Gesamtschule

	Anzahl Aufgaben	Maximale Punktzahl	Mittelwert	SD
Deutsch	12	45	32,16	5,49
Mathematik (G-Kurs)	13	30	12,22	5,10
Mathematik (E-Kurs)	15	30	13,72	4,52
Französisch	10	36	24,16	5,72
Englisch	8	36	24,04	5,54

## 5 TESTLEISTUNGEN 2004/05

### 5.1 Gymnasium

	Anzahl Auf- gaben	Maximale Punktzahl	Mittelwert	SD
Deutsch	11	50	3,31	6,65
Mathematik	25	28	4,04	4,29
Französisch	12	90	4,46	12,93
Englisch	19	54	1,89	6,72
Latein	16	84	3,62	11,86

### 5.2 Erweiterte Realschule

	Anzahl Auf- gaben	Maximale Punktzahl	Mittelwert	SD
Deutsch	15	60	41,00	8,27
Mathematik (H-Bildungsgang)	23	35	16,86	6,22
Mathematik (M-Bildungsgang)	23	32	16,33	5,20
Französisch	14	36	21,37	5,55
Englisch	16	54	36,24	8,49

### 5.3 Gesamtschule

	Anzahl Auf- gaben	Maximale Punktzahl	Mittelwert	SD
Deutsch	15	60	41,45	8,06
Mathematik (G-Kurs)	23	35	14,94	5,59
Mathematik (E-Kurs)	23	32	16,10	5,46
Französisch	14	36	20,60	6,41
Englisch	16	54	36,16	9,01

## 6 TESTLEISTUNGEN 2005/06

### 6.1 Gymnasium

	Anzahl Auf- gaben	Maximale Punktzahl	Mittelwert	SD
Deutsch	11	80	49,02	10,35
Mathematik	26	28	15,61	4,59
Französisch	12	54	40,16	6,56
Englisch	13	72	47,11	8,91
Latein	15	80	54,34	13,93

### 6.2 Erweiterte Realschule

	Anzahl Auf- gaben	Maximale Punktzahl	Mittelwert	SD
Deutsch	10	60	36,85	8,92
Mathematik (H-Bildungsgang)	15	30	16,76	5,05
Mathematik (M-Bildungsgang)	21	33	16,94	5,07
Französisch	9	40	26,23	5,40
Englisch	11	36	22,20	5,20

### 6.3 Gesamtschule

	Anzahl Auf- gaben	Maximale Punktzahl	Mittelwert	SD
Deutsch	10	60	38,01	8,97
Mathematik (G-Kurs)	15	30	14,84	4,68
Mathematik (E-Kurs)	21	33	16,89	5,31
Französisch	9	40	25,77	5,74
Englisch	11	36	22,46	5,17

## 7 TESTLEISTUNGEN 2006/07

### 7.1 Gymnasium

	Anzahl Auf- gaben	Maximale Punktzahl	Mittelwert	SD
Deutsch	21	66	42,03	8,84
Mathematik	22	25	12,92	5,43
Französisch	12	54	38,54	7,17
Englisch	14	72	53,47	7,07
Latein	16	90	55,62	16,82

### 7.2 Erweiterte Realschule

	Anzahl Auf- gaben	Maximale Punktzahl	Mittelwert	SD
Deutsch	11	60	43,12	8,12
Mathematik (H-Bildungsgang)	20	30	14,54	4,38
Mathematik (M-Bildungsgang)	17	30	13,89	4,84
Französisch	9	60	38,83	8,51
Englisch	9	90	58,36	13,31

### 7.3 Gesamtschule

	Anzahl Auf- gaben	Maximale Punktzahl	Mittelwert	SD
Deutsch	11	60	42,58	8,16
Mathematik (G-Kurs)	20	30	13,47	4,38
Mathematik (E-Kurs)	17	30	13,75	4,73
Französisch	9	60	37,21	9,23
Englisch	9	90	57,69	13,75

## 8 TESTLEISTUNGEN 2007/08

### 8.1 Gymnasium

	Anzahl Auf- gaben	Maximale Punktzahl	Mittelwert	SD
Deutsch	11	40	26,48	4,95
Mathematik	23	29	14,87	5,31
Französisch	12	54	37,26	7,75
Englisch	14	72	53,12	8,28
Latein	18	69	35,52	11,98

### 8.2 Erweiterte Realschule

	Anzahl Auf- gaben	Maximale Punktzahl	Mittelwert	SD
Deutsch	11	58	38,27	8,21
Mathematik (H-Bildungsgang)	19	25	13,84	3,97
Mathematik (M-Bildungsgang)	18	32	16,97	4,56
Französisch	11	60	33,58	8,81
Englisch	9	90	54,36	13,33

### 8.3 Gesamtschule

	Anzahl Auf- gaben	Maximale Punktzahl	Mittelwert	SD
Deutsch	11	58	39,28	7,84
Mathematik (G-Kurs)	19	25	13,02	3,84
Mathematik (E-Kurs)	18	32	17,22	4,31
Französisch	11	60	31,26	9,44
Englisch	9	90	54,66	12,65

## 9 Folgen und Entwicklungen

### 9.1 Bildungspolitische Folgen

#### 9.1.1 *Einstellung und Bereitschaft zur externen Evaluation geschaffen*

Aller Anfang ist schwer – das gilt insbesondere für Veränderungen im Ablauf des „Tagesgeschäftes“ in Schulen und hier ganz besonders, was die Bewertung und Beurteilung von Unterricht betrifft. Dies kann im Sinne der output-Orientierung, wie sie nach der schrittweisen Einführung der Bildungsstandards in den Unterricht vorgesehen ist, nur über standardisierte Leistungstest geschehen. In diesem Rahmen haben die Vergleichsarbeiten den Boden für eine bundesweite Leistungsmessung, wie sie durch Schaffung des IQB vorgesehen ist, bereitet. Durch ihre praktische Durchführung wurde Schulen, Eltern, Schüler und Lehrer aufgezeigt, welche Möglichkeiten sich durch überregionale Testungen ergeben, welche Vorteile daraus auf den verschiedenen Ebenen gezogen werden können und wie man mit – teilweise auch kritischen Ergebnissen – umgehen kann. Dies alles führt dazu, dass die Vorbehalte, die bei vielen Lehrern im Zusammenhang mit dem Begriff der externen Evaluation noch existieren, abgebaut werden konnten.

#### 9.1.2 *Unterstützung von Problemschulen*

Schulen, die in problematischen Einzugsgebieten liegen oder mit einer schwierigen Schülerpopulation zu tun haben, benötigen häufig bessere Unterstützung in Form von Lehrerstunden, Ausstattung usw. Häufig ist allen am Bildungsprozess beteiligten Institutionen die manchmal unzureichende Ausstattung in diesen Fällen nicht bekannt. Im Rahmen der Vergleichsarbeiten hatten Schulen, die mit diesen Problemen über Jahrgänge hinweg zu tun haben, jetzt Informationen auf der Hand, die einerseits einen Bedarf, andererseits aber auch möglicherweise Erfolge aufzeigen können und so gezielter unterstützt werden können. Dies betrifft nicht nur finanzielle Unterstützung, sondern auch personelle oder pädagogische Stärkung bei schulspezifischen pädagogischer Maßnahmen (z.B. Förderung von Migranten). Beispielhaft ist das auch Tabelle 4 zu erkennen: Es gibt Schulen, deren Schüler die aufgrund erschwelter Umstände in jedem Jahr große Leistungsdefizite aufweisen u.u.

### 9.1.3 „Ranking“

Ein häufig befürchtetes Schul-Ranking wurde nicht erstellt und ist auch methodisch und inhaltlich nicht sinnvoll, wie die folgende Beispieltabelle für Gymnasien (Mathematik) deutlich macht. Es konnte lediglich die jeweilige Position des Leistungsstandes des jeweiligen Jahrgangs dargestellt werden. Die numerischen Abstände können dabei zwischen den einzelnen Positionen sehr unterschiedlich sein und werden in dieser Übersicht nicht gewichtet. Ein Ranking vorzunehmen wäre inhaltlich und methodisch auch nicht aussagekräftig, wie die Veränderungen der Position über die Jahre hinweg klar erkennen lässt. Für die Position spielt auch neben der veränderten Schülerpopulation auch die jeweilige Testschwierigkeit eine entscheidende Rolle. Zur besseren Veranschaulichung wurden die Positionen der Schulen im oberen Leistungsdrittel grün, im mittleren Leistungsdrittel gelb und im unteren Leistungsdrittel rot markiert.

	2004	2005	2006	2007
Schule 1	Yellow	Red	Green	Yellow
Schule 2	Green	Yellow	Yellow	Green
...	Green	Yellow	Yellow	Red
...	Green	Green	Yellow	Green
...	Green	Yellow	Yellow	Yellow
...	Yellow	Yellow	Yellow	Green
...	Yellow	Yellow	Yellow	Red
...	Red	Red	Green	Red
...	Yellow	Green	Yellow	Red
...	Red	Red	Green	Yellow
...	Red	Red	Red	Red

Tabelle 4: Positionsveränderungen (Beispiel)

## 9.2 Pädagogische Folgen

### 9.2.1 Schwächen aufdecken und diskutieren

Die übliche Basis für schulinterne Vergleiche sind gemeinsame Klassenarbeiten in Parallelklassen. Diese werden aber nach subjektiven Kriterien und nach Absprache der Fachlehrer erstellt und bilden insofern keinen tatsächlichen Leistungsstand ab. Die LVA wurden nun in vielen Schulen dazu genutzt, die Leistungsergebnisse einer Klassenstufe in den Fachkonferenzen untereinander zu vergleichen und insbesondere die Stärken-Defizit-Analysen gegeneinander zu stellen. Daraus lassen sich neben an-

deren Effekten auch Folgen möglicher unterschiedlicher pädagogischer Ansätze und Vorgehensweisen erkennen, was zur Qualitätsverbesserung in Schulen und Unterricht beitragen kann.

Verbesserungen sind auch auf Schülerseite durch Analyse der LVA möglich. Die Ergebnisse sollen für Lehrer und Eltern die Stärken und Schwächen der Kinder in den getesteten Bereichen widerspiegeln. Lehrer können den Eltern gezielt Hinweise geben, wie sie ihren Kindern beim Lernen helfen können und Schullaufbahnempfehlungen erhalten eine zusätzliche Entscheidungsgrundlage<sup>6</sup>.

### 9.2.2 *Teaching for the test*

Ein häufig genannter Kritikpunkt an den LVA war das sog. „teaching for the test“, also das gezielte Üben und Vorbereiten der Schüler auf die Testung. Da die LVA aber als Gegenstand die zentralen Themen und Fertigkeiten der getesteten Fächer prüfen, kann eine solche gezielte Wiederholung und Übung nur im Sinne der Schüler gewesen sein. Der als negativ geäußerte Einwand, hier würde nur versucht, Schulen in ihrer Gesamtleitung höher zu positionieren, mag dabei als Motivation in Einzelfällen gedient haben – geholfen wurde damit aber letztlich den Schülern, die durch Wiederholung gerade zu Beginn eines Schuljahres einen guten Einstieg in die Fächer erhalten haben. Teaching for the test in diesem Sinne eingesetzt ist also eher wünschenswert als abzulehnen.

### 9.2.3 *Diagnostische Kompetenz der Lehrer*

Vergleichsarbeiten (Lernstandserhebungen) bieten vielfältige Möglichkeiten zur Stärkung der diagnostischen Kompetenzen von Lehrkräften, wie beispielsweise der zuverlässigen Einschätzung von Aufgabenschwierigkeiten oder der zutreffenden Beschreibung von Lernständen einzelner Schüler(innen). Gerade die diagnostische Kompetenz wurde in internationalen Studien bei deutschen Lehrern noch als problematisch angesehen. So wurden beispielsweise in PISA 2000 nur rund 14% der schwachen Leser von ihren Lehrern auch als solche vorab klassifiziert (Frederking et al,

---

<sup>6</sup> <http://bildungsklick.de/pm/52658/vergleichsarbeiten-2007-in-den-fachern-deutsch-mathematik-und-englisch/>

2005). „Die Bedeutung der diagnostischen Kompetenz für das schulische Lernen ergibt sich daraus, dass die Schwierigkeit von Unterrichtsmaßnahmen, Fragen und Aufgaben auf die Lernvoraussetzungen der Schüler abgestimmt sein muss. Lehrkräfte, die über Fähigkeitsunterschiede, über Aufgabenschwierigkeiten und/oder über Stärken und Schwächen ihrer Schüler nicht im Bilde sind, können keine ausreichende Passung zwischen Lernvoraussetzungen und Anforderungen herstellen. Individuelle Förderung setzt deshalb zwingend ein ausreichendes diagnostisches Wissen der verantwortlichen Lehrkraft voraus.“ (Helmke, 2008). Lehrer sind entgegen vieler Meinungen zwar gut in der Lage, Unterschiede zwischen Schülern ihrer Klasse(n) zu bestimmen, es fehlt ihnen allerdings an objektiven Leistungsergebnissen, bei denen der Standard nicht von ihnen selbst gesetzt wurde.

## 10 Literatur

- Arnold, K.-H. (2005). Mehr Fairness im Bildungssystem. Fragen zu Standards und Vergleichsarbeiten. In G. Becker, A. Bremerich-Vos, M. Demmer, K. Maag Merki, B. Pribe, K. Schwippert, L. Stäudel & K.-J. Tillmann (Hrsg.), *Standards. Unterrichten zwischen Kompetenzen, zentralen Prüfungen und Vergleichsarbeiten*, S. 25–27. Friedrich Jahresheft.
- Baumert, J., Lehmann, R., Lehrke, M., Schmitz, B., Clausen, M., Hosenfeld, I., Köller, O. & Neubrand, J. (1997). *TIMSS. Mathematisch-naturwissenschaftlicher Unterricht im Vergleich. Deskriptive Befunde*. Opladen: Leske + Budrich.
- Frederking, V., Heller, H. & Scheunpflug, A. (Hrsg.) (2005). Nach PISA: Konsequenzen für Schule und Lehrerbildung nach zwei Studien. Wiesbaden: VS Verlag.
- Helmke, A. (2008). Mehr Diagnosekompetenz an Schulen. URL: <http://aqs.rlp.de/aktuelles/mitteilungen/mehr-diagnosekompetenz-an-schulen.html> (08.07.09).
- Hofmann-Göttig, J., Eschmann, W. & Daumen, C. (2005). Und sie bewegt sich doch...Vom Umgang mit den Ergebnissen externer Evaluation aus Sicht von Bildungspolitik und Schulaufsicht. In G. Becker, A. Bremerich-Vos, M. Demmer, K. Maag Merki, B. Pribe, K. Schwippert, L. Stäudel & K.-J. Tillmann (Hrsg.), *Standards. Unterrichten zwischen Kompetenzen, zentralen Prüfungen und Vergleichsarbeiten*, S. 32-36. Friedrich Jahresheft.
- Hosenfeld, I. (2005). Rezeption – Reflexion – Aktion. Wie lassen sich Lernstandserhebungen und Vergleichsarbeiten pädagogisch nutzen? In G. Becker, A. Bremerich-Vos, M. Demmer, K. Maag Merki, B. Pribe, K. Schwippert, L. Stäudel & K.-J. Tillmann (Hrsg.), *Standards. Unterrichten zwischen Kompetenzen, zentralen Prüfungen und Vergleichsarbeiten*, S. 112–114. Friedrich Jahresheft.
- Hosenfeld, I., Schrader, W.-H. & Helmke, T. (2006). Von der Rezeption zur Ergebnisrückmeldung: Leistungsevaluation im Spannungsfeld von System-Monitoring und Schulentwicklung. In I. Hosenfeld & F.-W. Schrader (Hrsg.), *Schulische Leistung. Grundlagen, Bedingungen, Perspektiven*, S. 289 – 313. Münster: Waxmann.

- Isaac, K., Halt, A. C., Hosenfeld, I., Helmke, A. & Groß Ophoff, J. (2006). VERA: Qualitätsentwicklung und Lehrerprofession durch Vergleichsarbeiten. Kommentar zur Diskussion um VERA. *Die Deutsche Schule*, 98, 107-111.
- Köller, O. (2007). Bildungsstandards, einheitliche Prüfungsanforderungen und Qualitätssicherung in der Sekundarstufe II. In D. Benner (Hrsg.), *Bildungsstandards*, S. 3–28. Paderborn: Ferdinand Schöningh Verlag
- Rolff, H.-G. (2004). Was bringt die vergleichende Leistungsmessung für die pädagogische Arbeit in Schulen? In F.E. Weinert (Hrsg.), *Leistungsmessungen in Schulen*, S. 337-352. Weinheim: Beltz Verlag.
- Schwippert, K. (2005). Tests. Oder: Wie man Äpfel mit Birnen vergleicht. In G. Becker, A. Bremerich-Vos, M. Demmer, K. Maag Merki, B. Pribe, K. Schwippert, L. Stäudel & K.-J. Tillmann (Hrsg.), *Standards. Unterrichten zwischen Kompetenzen, zentralen Prüfungen und Vergleichsarbeiten*, S. 15–17. Friedrich Jahresheft.

## 11 Anhang

### 11.1 Leistungsübersicht ERS/GeS (Geschlecht, Migrationsstatus)

**2003/04**

		N	M	S
Deutsch	Jungen	3611	30,87	5,72
	Mädchen	3267	33,56	5,21
	Migranten	---		
	Nicht-Migranten			
Mathe H	Jungen	1886	15,45	5,43
	Mädchen	1478	12,73	5,24
	Migranten			
	Nicht-Migranten			
Mathe M	Jungen	1984	15,45	4,59
	Mädchen	2031	13,88	4,42
	Migranten			
	Nicht-Migranten			
Englisch	Jungen	880	23,84	5,24
	Mädchen	812	23,94	5,15
	Migranten			
	Nicht-Migranten			
Französisch	Jungen	1095	22,71	5,38
	Mädchen	1241	23,83	5,32
	Migranten			
	Nicht-Migranten			

**2004/05**

		N	M	S
Deutsch	Jungen	3471	39,78	8,21
	Mädchen	3028	42,67	7,96
	Migranten	936	36,11	9,70
	Nicht-Migranten	5565	41,97	7,62
Mathe H	Jungen	1769	17,53	6,19
	Mädchen	1380	15,02	5,77
	Migranten	484	15,19	6,06
	Nicht-Migranten	2665	16,68	6,13
Mathe M	Jungen	1890	17,53	5,14

<sup>7</sup> Im ersten Jahr lagen zur ethnischen Herkunft keine Informationen vor

		N	M	S
	Mädchen	1936	15,02	5,12
	Migranten	338	14,74	5,82
	Nicht-Migranten	3488	16,41	5,20
Englisch	Jungen	837	35,95	8,72
	Mädchen	873	36,60	8,62
	Migranten	203	35,79	9,10
	Nicht-Migranten	1566	36,27	8,62
Französisch	Jungen	1210	20,58	5,65
	Mädchen	1246	21,90	5,69
	Migranten	191	20,79	6,54
	Nicht-Migranten	2265	21,29	5,63

**2005/06**

		N	M	S
Deutsch	Jungen	2946	35,48	8,85
	Mädchen	2827	38,97	8,70
	Migranten	805	34,24	9,66
	Nicht-Migranten	4974	37,67	8,73
Mathe H	Jungen	1628	17,35	4,85
	Mädchen	1241	14,83	4,85
	Migranten	491	14,85	5,46
	Nicht-Migranten	2416	16,57	4,88
Mathe M	Jungen	1771	17,89	5,04
	Mädchen	1705	15,99	5,07
	Migranten	282	14,65	5,37
	Nicht-Migranten	3240	17,13	5,07
Englisch	Jungen	765	22,24	5,15
	Mädchen	760	22,35	5,23
	Migranten	140	21,57	4,96
	Nicht-Migranten	1441	22,37	5,21
Französisch	Jungen	983	25,67	5,43
	Mädchen	1102	26,56	5,47
	Migranten	144	27,41	5,99
	Nicht-Migranten	1941	26,05	5,42

**2006/07**

		N	M	S
Deutsch	Jungen	2987	41,13	8,33
	Mädchen	2746	44,99	7,36
	Migranten	814	39,72	8,73
	Nicht-Migranten	4934	43,50	7,89
Mathe H	Jungen	1450	15,25	4,36
	Mädchen	1247	13,12	4,17

		N	M	S
	Migranten	442	12,82	4,41
	Nicht-Migranten	2257	14,55	4,34
Mathe M	Jungen	1758	15,14	4,73
	Mädchen	1816	12,60	4,54
	Migranten	304	11,89	4,87
	Nicht-Migranten	3275	14,04	4,76
Englisch	Jungen	980	57,91	13,54
	Mädchen	917	58,34	13,42
	Migranten	207	53,91	14,43
	Nicht-Migranten	1693	58,63	13,26
Französisch	Jungen	1017	37,95	8,81
	Mädchen	1012	39,10	8,51
	Migranten	94	39,52	38,47
	Nicht-Migranten	1937	10,12	8,60

**2007/08**

		N	M	S
Deutsch	Jungen	2868	36,99	8,15
	Mädchen	2508	40,43	7,64
	Migranten	775	35,83	9,01
	Nicht-Migranten	4619	39,04	7,86
Mathe H	Jungen	1297	14,31	3,88
	Mädchen	977	12,73	3,83
	Migranten	390	13,16	4,25
	Nicht-Migranten	1923	13,71	3,88
Mathe M	Jungen	1565	17,95	4,34
	Mädchen	1617	16,14	4,45
	Migranten	316	15,77	4,62
	Nicht-Migranten	2899	17,18	4,45
Englisch	Jungen	870	54,62	12,85
	Mädchen	820	54,31	13,33
	Migranten	177	52,28	13,43
	Nicht-Migranten	1513	54,73	13,02
Französisch	Jungen	853	31,78	8,85
	Mädchen	987	34,19	8,98
	Migranten	114	32,79	9,24
	Nicht-Migranten	1735	33,11	8,98

## 11.2 Leistungsübersicht GYM (nach Geschlecht und Migrationsstatus)

**2003/04**

		N	M	S
Deutsch	Jungen	1759	44,91	8,40
	Mädchen	1943	47,77	7,76
	Migranten	--- <sup>8</sup>		
	Nicht-Migranten			
Mathe	Jungen	1696	17,74	5,39
	Mädchen	1882	15,77	5,49
	Migranten			
	Nicht-Migranten			
Englisch	Jungen	594	30,00	3,20
	Mädchen	639	30,48	3,24
	Migranten			
	Nicht-Migranten			
Französisch	Jungen	934	29,54	3,79
	Mädchen	1153	30,40	3,70
	Migranten			
	Nicht-Migranten			
Latein	Jungen	81	43,68	9,78
	Mädchen	86	43,33	9,35
	Migranten			
	Nicht-Migranten			

**2004/05**

		N	M	S
Deutsch	Jungen	1699	35,20	6,70
	Mädchen	1842	37,49	6,40
	Migranten	141	33,78	6,88
	Nicht-Migranten	3400	36,50	6,61
Mathe	Jungen	1681	17,60	4,20
	Mädchen	1893	16,19	4,26
	Migranten	60	15,83	4,58
	Nicht-Migranten	3514	16,87	4,28
Englisch	Jungen	802	52,13	12,59
	Mädchen	1026	54,66	13,09
	Migranten	27	50,72	14,19

<sup>8</sup> Vgl. FN 7

		N	M	S
	Nicht-Migranten	1801	53,59	12,91
Französisch	Jungen	568	35,79	6,50
	Mädchen	567	35,66	6,84
	Migranten	32	34,43	7,49
	Nicht-Migranten	1103	35,76	6,64
Latein	Jungen	100	58,25	12,10
	Mädchen	123	57,53	11,70
	Migranten	4	47,25	4,65
	Nicht-Migranten	219	58,05	11,87

**2005/06**

		N	M	S
Deutsch	Jungen	1707	47,22	9,97
	Mädchen	1796	50,98	10,29
	Migranten	133	42,53	10,35
	Nicht-Migranten	3457	49,27	10,28
Mathe	Jungen	1624	16,58	4,47
	Mädchen	1740	14,87	4,44
	Migranten	90	14,10	4,48
	Nicht-Migranten	3386	15,65	4,48
Englisch	Jungen	657	47,40	9,00
	Mädchen	674	46,93	8,85
	Migranten	47	48,59	9,03
	Nicht-Migranten	1304	47,06	8,90
Französisch	Jungen	790	39,35	6,43
	Mädchen	993	40,65	6,60
	Migranten	17	41,65	4,06
	Nicht-Migranten	1779	40,11	6,58
Latein	Jungen	93	52,10	14,10
	Mädchen	119	56,10	13,60
	Migranten	1	54,34	---
	Nicht-Migranten	211	54,31	13,95

**2006/07**

		N	M	S
Deutsch	Jungen	1708	40,82	4,51
	Mädchen	1817	43,16	9,01
	Migranten	173	40,00	9,49
	Nicht-Migranten	3384	42,13	8,79
Mathe	Jungen	1587	13,74	5,30
	Mädchen	1687	12,17	5,42
	Migranten	46	10,96	6,00
	Nicht-Migranten	3282	12,94	5,42

		N	M	S
Englisch	Jungen	631	53,70	6,89
	Mädchen	632	53,23	7,25
	Migranten	45	52,17	5,62
	Nicht-Migranten	1219	53,52	7,12
Französisch	Jungen	743	37,48	7,11
	Mädchen	915	39,40	7,11
	Migranten	20	39,13	6,68
	Nicht-Migranten	1638	38,53	7,18
Latein	Jungen	108	59,99	17,18
	Mädchen	135	56,45	16,42
	Migranten	1	27,50	k. A.
	Nicht-Migranten	242	58,15	16,74

**2007/08**

		N	M	S
Deutsch	Jungen	1689	26,13	4,86
	Mädchen	1788	26,83	5,00
	Migranten	114	25,18	4,87
	Nicht-Migranten	3365	16,53	4,95
Mathe	Jungen	1554	15,88	5,27
	Mädchen	1717	13,96	5,18
	Migranten	87	13,32	5,52
	Nicht-Migranten	3184	14,91	5,30
Englisch	Jungen	672	52,90	8,15
	Mädchen	576	53,37	8,45
	Migranten	22	53,50	9,36
	Nicht-Migranten	1227	53,11	8,27
Französisch	Jungen	746	36,78	8,01
	Mädchen	884	37,66	7,50
	Migranten	21	34,93	9,92
	Nicht-Migranten	1609	37,29	7,71
Latein	Jungen	93	34,17	10,94
	Mädchen	124	36,02	12,73
	Migranten	0	---	----
	Nicht-Migranten	218	35,21	11,98