



Seit 1558

Friedrich-Schiller-Universität Jena

Kompetenztest.de

Lehrstuhl für Methodenlehre und
Evaluationsforschung (Prof. Rolf Steyer)

Projektleiter: Dr. Christof Nachtigall

Landesbericht

Thüringer Kompetenztests

2008

Dr. Christof Nachtigall (Hrsg.)

Unter Mitarbeit von:

Marc Müller, Heike Scholze
und Jörg Wanjek (Universität Jena)

Kompetenztest.de 

Inhaltsverzeichnis

1	Die Kompetenztests 2008	3
1.1	Ziele der Kompetenztests	3
1.2	Chancen und Risiken der Kompetenztests	4
2	Testentwicklung und Gütekriterien	6
2.1	Testentwicklung	6
2.2	Zweitkorrekturen	7
2.2.1	Deutsch	8
2.2.2	Mathematik	10
2.2.3	Zusammenfassung und Aussichten	12
3	Durchführung und Datenerfassung	14
3.1	Anmeldung zu den Kompetenztests 2008 und zum Schülerlängsschnitt	14
3.2	Teilnahme an den Kompetenztests 2008	15
3.3	Webbasierte Datenerfassung	15
3.4	Auswertung der Hotline	18
4	Diagnostische Kompetenz	20
4.1	Einleitung	20
4.2	Auswertung der Diagnostischen Kompetenz für Klassenstufe 3	21
4.3	Auswertung der Diagnostischen Kompetenz für Klassenstufe 6	23
4.4	Auswertung der Diagnostischen Kompetenz für Klassenstufe 8	26
4.5	Zusammenhang der Ergebnisse der Diagnostischen Kompetenz mit Lehrervariablen	28
4.6	Fazit	28
5	Rückmeldung der Testergebnisse	30
5.1	Korrigierter Landesmittelwert und faire Vergleiche	33
6	Testergebnisse auf Landesebene	35
6.1	Erreichte Gesamtpunktzahlen	35
6.2	Erreichte Gesamtpunktzahlen getrennt nach Schulart	40
6.3	Aufgabenschwierigkeiten	43
7	Analyse leistungsrelevanter Faktoren	57
7.1	Einflussfaktoren auf Klassenebene	57
7.1.1	Mathematik	58
7.1.2	Deutsch	65
7.1.3	Englisch	72
7.1.4	Einfluss der Klassengröße	79
7.2	Einflussfaktoren auf Schülerebene	80
7.2.1	Geschlecht	80
7.2.2	Muttersprache	83
7.2.3	Sozioökonomischer Status (SES)	86
7.2.4	Jahr des Schulbesuchs	91

7.2.5	Besondere Lernschwierigkeiten	99
7.2.6	Zusammenhang Halbjahresnote und Testergebnisse	101
7.2.7	Vorhersage der Testergebnisse durch die Schülermerkmale	106
7.2.8	Entwicklung über die letzten Jahre	107
7.2.9	Zusammenfassung	109
8	Rezeption und Evaluation der Tests	111
8.1	Kurzbefragung nach der Dateneingabe	111
8.2	Ergebnisse der Verwendbarkeitsanalysen	111
8.2.1	Anmerkungen und Anregungen zu den Kompetenztests	112
8.3	Rezeption der Testergebnisse an den Schulen	115
8.4	Evaluationen zu den Kompetenztests	117
8.4.1	Gesamtergebnis	117
8.4.2	Vergleich der Schulleiter und Lehrer aller Schularten	119
8.4.3	Schularttypische Besonderheiten	121
8.4.4	Fazit	125
9	Zusammenfassung und Ausblick	127
9.1	Entwicklung und Durchführung der Kompetenztests 2008	127
9.2	Nutzen und Nutzung der Kompetenztests	128
9.3	Perspektiven	130
10	Literatur	131

1 Die Kompetenztests 2008

Die Thüringer Kompetenztests fanden im Schuljahr 2007 / 2008 zum sechsten Mal statt. Neben den in den vergangenen Jahren optimierten Verfahren zur Durchführung, Datenauswertung und Rückmeldung der Ergebnisse ergaben sich eine Reihe von Neuerungen im Bereich der Entwicklung und des Einsatzbereichs der Kompetenztests. So wurden in Klassenstufe 3 erstmals die an der Universität Landau im Projekt VerA entwickelten Tests für Deutsch und Mathematik in Thüringen eingesetzt. In Klassenstufe 6 erfolgte—ebenfalls ein Novum—eine gemeinsame Testentwicklung in einem Verbundprojekt der Länder Hamburg, Hessen, Mecklenburg-Vorpommern, Schleswig-Holstein und Thüringen. Getestet wurden die Fächer Deutsch, Mathematik und Englisch, wobei im Bereich Englisch neben dem Hörverstehen erstmals auch die Lesekompetenz geprüft wurde. In Klassenstufe 8 kam ein vom Institut zur Qualitätsentwicklung im Bildungswesen (IQB) in Berlin entwickelter Mathematiktest zum Einsatz. Durch die bei der Testentwicklung erfolgte Zusammenarbeit war es möglich, die inhaltlichen und methodischen Qualitätskriterien weiter zu erhöhen und zu vereinheitlichen.

Die Durchführung der Tests verlief wie in den Vorjahren weitestgehend reibungslos, sowohl von Seiten der Schulen als auch von Seiten des Projektes *kompetenztest.de* der Universität Jena, welches die Datenerfassung und Auswertung durchführte und den Schulen die Ergebnisberichte auf dem passwortgeschützten Schulportal zur Verfügung stellte. Die einzelnen Schritte von der Testentwicklung bis zur Nutzung der Testergebnisse durch die Schulen werden in diesem Landesbericht dokumentiert.

Seitdem die Durchführung der Tests an den Schulen weitgehend problemlos erfolgt, steht verstärkt die Nutzung der Tests durch die Schulen im Blickfeld. Die Entwicklung der letzten Jahre hat gezeigt, dass die Tests zunehmend als Chance und weniger als unerwünschte Zusatzbelastung gesehen werden. Dennoch ist es wichtig, die Akzeptanz der Tests und die praktische Nutzung ihrer Ergebnisse an den Schulen nicht selbstverständlich voraus zu setzen, sondern diese müssen weiter unterstützt und gefördert werden. Entsprechende Befragungen der Lehrkräfte und Schulleiter zeigen zudem, dass es bei der Einschätzung der Tests große Unterschiede zwischen den Schularten gibt. Die Grundschulen stehen den Tests deutlich positiver gegenüber als die weiterführenden Schulen. Das mag zum Teil an den im Jahre 2007 durchgeführten Regionalkonferenzen liegen, bei denen die Nutzungsmöglichkeiten der Tests für viele Teilnehmer deutlich wurden. Auch bewerten die Schulleiter die Nützlichkeit der Tests höher als die Fachlehrer, was mit einem besseren Informationsstand der Schulleiter im Hinblick auf die Nutzung der Tests zu tun hat. Aus diesem Grunde werden immer wieder von Schulleitern Referenten zum Thema Nutzung der Kompetenztests für schulinterne Fortbildungen angefragt.

1.1 Ziele der Kompetenztests

Mit den jährlichen Kompetenztests wird in einer landesweiten Erhebung der Leistungsstand der Thüringer Schüler gemessen. Im Schuljahr 2007/2008 fanden die Tests in den Fächern Deutsch und Mathematik in den Klassenstufen 3 und 6 aller Schularten statt. Zusätzlich wurden Tests im Bereich Englisch Hör- und Leseverstehen in Klassenstufe 6 sowie im Fach Mathematik in Klassenstufe 8 geschrieben. Die so erhobenen Daten erlauben Leistungseinschätzungen

und Leistungsvergleiche auf Schüler-, Klassen- und Schulebene, sowohl bezogen auf das Gesamtergebnis im Test als auch bezogen auf Teilbereiche des Faches bis hin zu den Ergebnissen der einzelnen Aufgaben. Zentrales Ziel der Kompetenztests ist es, durch die Ergebnisse und Vergleiche auf Klassen- und Schulebene Analysen der Stärken und Schwächen des eigenen Unterrichts sowie der schulischen Rahmenbedingungen zu ermöglichen und einen Anstoß für Maßnahmen zur Unterrichts- und Schulentwicklung an den Thüringer Schulen zu geben. Dazu gibt es eine Reihe von Berichten zu jedem Test. Der Sofortbericht, der unmittelbar nach Ende der Dateneingabe dem Fachlehrer zur Verfügung steht, dient primär zur Arbeit mit den einzelnen Schülern. Ergebnis- und Ergänzungsberichte stehen ca. 6 Wochen nach Ende der Dateneingabe zur Verfügung. Sie enthalten Vergleichswerte von Schulen bzw. Klassen mit ähnlicher Schülerschaft, die zur Reflektion über die Stärken und Schwächen des eigenen Unterrichts genutzt werden sollen. Die gleichzeitig erscheinenden Schulberichte dokumentieren das Abschneiden der ganzen Schule mit entsprechenden Vergleichswerten. Sie dienen vorrangig der Schulentwicklung. Der Landesbericht und die im Jahr 2007 neu konzipierten Schulamtsberichte geben die Ergebnisse auf der entsprechenden Aggregatebene wieder und dienen dem Bildungsmonitoring. So enthält der Landesbericht auch die Ergebnisse zur Untersuchung der diagnostischen Kompetenz. Hier können die am Test beteiligten Fachlehrer auf anonymer und freiwilliger Basis am Vortag des Tests eine Reihe von Testaufgaben einsehen und die Lösungshäufigkeiten in ihrer Klasse prognostizieren. Weiterhin zeigt der Landesbericht die Resultate der jährlichen Befragung der Lehrkräfte, in denen z. B. die derzeit angewendeten Unterrichtsmethoden thematisiert werden. Auch hier liefern die Ergebnisse Vergleichsmöglichkeiten und können Anstoß für die Weiterentwicklung des eigenen Unterrichts sein.

1.2 Chancen und Risiken der Kompetenztests

Die Kompetenztests sind ein ausgereiftes, etabliertes und akzeptiertes Instrument, welches in vielfältiger Weise an den Schulen genutzt wird. Sowohl im Hinblick auf die Arbeit mit dem einzelnen Schüler, als auch zur Evaluation und Entwicklung des eigenen Unterrichts bzw. der eigenen Schule liefern die Ergebnisse wichtige Hinweise. Besonders die Vergleichsdaten ermöglichen eine Standortbestimmung und sollten Anstoß für Maßnahmen der Unterrichts- und Schulentwicklung sein. Das umfangreiche Unterstützungsangebot für diese Prozesse umfasst z. B. testspezifische Materialien zur Weiterarbeit in der Klasse, eine Reihe von Fortbildungen für Fachlehrer und Schulleiter oder weiterführende Materialien wie das ThILLM-Heft Nr. 115 (vgl. auch die Hinweise im Downloadbereich von www.kompetenztest.de).

Diese Chancen sind auch nach 6 Jahren der Durchführung und der damit gewonnenen Erfahrungen zu betonen. Eine Vielzahl von Lehrern nimmt diese Chancen wahr. Doch zeigen die Evaluationsdaten zu den Kompetenztests, dass noch immer nur ein Teil des Potenzials der Tests genutzt wird. Für eine Reihe von Lehrern sind die Tests etwas, das unter dem Etikett der „Mehrbelastung“ eher abgelehnt und mit dem sich nicht auseinander gesetzt wird. Das ist durchaus nachvollziehbar, wenn man bedenkt, dass die Ergebnisse eben nicht nur den Blick auf andere – die Schüler – schärfen, sondern auch zu einer Betrachtung der Stärken und Schwächen des eigenen Unterrichts auffordern. Dieser Blick erfordert Überwindung und wird nur dann zum selbstverständlichen pädagogisch-professionellen Handeln, wenn ein erfahrbarer Nutzen resultiert. Hier bleibt die Aufgabe, nachhaltig und unterstützend den Nutzen der Tests für den Fachlehrer weiter zu verbessern und dies für die Kollegen auch erfahrbar zu machen. Hier kommen den didaktischen Materialien, welche die Tests begleiten und mit denen sehr konkret Unterstützung bei der Unterrichtsentwicklung gegeben wird, große Bedeutung zu.

Eine nach wie vor bedenkliche Entwicklung besteht im „teaching to the test“, dem systematischen Trainieren von Testaufgaben mit dem alleinigen Ziel eines möglichst guten Abschneidens bei den Tests. Dies konterkariert das eigentliche Ziel einer nachhaltigen Unterrichtsentwicklung. Wenn in wertvoller Unterrichtszeit lediglich alte Testaufgaben mit dem Ziel gesteigerter Performanz beim Test geübt werden, dürfte der Kompetenzgewinn auf Seiten der Schüler gering ausfallen. Hingegen kann es sehr sinnvoll sein, den Hintergrund und die Ideen von alten Testaufgaben für den eigenen Unterricht zu nutzen. Daher ist das Schreiben von Vorjahrestests nicht automatisch zu verdammen, allein der Nutzen muss im Kompetenzgewinn der Schüler und in positiven Impulsen für den eigenen Unterricht liegen.

2 Testentwicklung und Gütekriterien

2.1 Testentwicklung

Die Testentwicklung im Schuljahr 2007 / 2008 wies große Veränderung im Vergleich zu den Vorjahren auf. Wurden bisher die Kompetenztests für Klassenstufe 6 ausschließlich im Land Thüringen entwickelt und erprobt, so fand dieser Prozess in diesem Jahr gemeinsam in einem Verbund der Bundesländer Hamburg, Hessen, Mecklenburg-Vorpommern, Sachsen, Schleswig-Holstein und Thüringen statt. Dabei waren die Länder Hamburg und Mecklenburg-Vorpommern federführend bei der Testentwicklung im Fach Englisch, das Land Schleswig-Holstein im Fach Deutsch und das Land Thüringen im Fach Mathematik. Die Aufgabenentwicklung orientierte sich im Hinblick auf Vorgehensweise und methodische Richtlinien an den Arbeiten zur Normierung der Bildungsstandards, welche vom Institut zur Qualitätsentwicklung im Bildungswesen (IQB) in Berlin durchgeführt werden. Auf diese Weise wird sichergestellt, dass die Tests wissenschaftlichen Gütekriterien genügen. Das Erstellen der Aufgaben (Itemwriting) und eine erste Vortestung (Präpilotierung) erfolgte durch speziell geschulte Fachlehrer aus allen beteiligten Bundesländern. Die entwickelten Aufgaben mussten sich dann mehreren Rückmelderunden unterziehen, in denen sie mit Hilfe von fachdidaktischer und psychometrischer Expertise auf ihre Tauglichkeit für die Kompetenztests geprüft und optimiert wurden. Die Arbeitsgruppe im Fach Mathematik wurde dabei von Frau Dr. Herwig vom Thillm geleitet, die fachdidaktische Betreuung erfolgte durch Prof. Dr. Blum von der Universität Kassel und die psychometrischen Analysen erstellte Dr. Nachtigall vom Projekt kompetenztest.de der Universität Jena. Anders als die früheren Kompetenztests basierte die finale Auswahl der Testitems auf der Item Response Theorie¹ Konkret wurden die Testaufgaben gemäß dem sogenannten Raschmodell ausgewählt und geprüft. Dies erfordert einen höheren Aufwand bei der Testentwicklung, als es in früheren Jahren der Fall war. Es muss eine Vielzahl von Aufgaben an einer großen Stichprobe von Schülern unter realistischen Einsatzbedingungen getestet werden. Auf der anderen Seite bringt diese Form der Testentwicklung eine Reihe von Vorteilen. Beispielsweise können die Schwierigkeit der Tests über unterschiedliche Jahre verglichen und annähernde Konstanz gesichert werden. Zudem eignen sich auf diese Weise entwickelte Tests in besonderer Weise für die Entwicklung von Kompetenzstufenmodellen.²

Auch bei den Kompetenztests für Klassenstufe 3 gab es grundlegende Änderungen. Statt wie bisher die Tests in Thüringen zu entwickeln bzw. Tests aus Bayern zu übernehmen, werden ab dem Schuljahr 2007 / 2008 die vom Projekt VerA an der Universität Landau im Auftrag der 16 Bundesländer entwickelten Tests von Thüringen übernommen. In Klassenstufe 8 kam ein vom Institut zur Qualitätsentwicklung im Bildungswesen in Berlin entwickelter Test im Fach Mathematik zum Einsatz. Alle Tests werden nach ähnlichen und methodisch hohen Qualitätsstandards entwickelt.

Die Inhalte der Tests orientieren sich wie in den Jahren zuvor an den in den Lehrplänen festgeschriebenen Anforderungsbereichen und beinhalten in den Fächern Deutsch und Mathematik jeweils Aufgaben zu Reproduktion, zur Anwendung und zum Transfer von Wissen. Diese sind

¹ Zum Hintergrund dieser Theorie vgl. z. B. das Lehrbuch von Rost(1994).

² Zur Bedeutung von Kompetenzstufenmodellen im Kontext der Bildungsstandards vergleiche Klieme et al. (2003), Kap. 5-7.

spezifisch auf die Anforderungen in den einzelnen Fächern und Klassenstufen ausgerichtet. Es werden Aufgaben gestellt, die im Lehrplan festgelegte, grundlegende Fähigkeiten der Schüler überprüfen sollen. Hierzu zählen im Fach Deutsch beispielsweise das sinnerfassende Lesen sowie Grundlagen der Sprachbetrachtung und Rechtschreibung. Im Fach Mathematik werden zum Beispiel Aufgaben zu den Basisrechenoperationen und zur Geometrie oder textgebundene Aufgaben gestellt.

Neben dieser curricularen Validität orientieren sich die Testaufgaben an den in den Bildungsstandards festgelegten Kompetenzen. In Mathematik sind dies z. B. die allgemeinen mathematischen Kompetenzen wie z. B. "mathematisch argumentieren" oder "mathematisch modellieren" sowie verschiedene Leitideen wie z. B. "Zahl" oder "Messen". Die Aufgaben des Englischtests in Klassenstufe 6 orientieren sich darüber hinaus an den Kompetenzstufen A1 und A2 des gemeinsamen europäischen Referenzrahmens.

2.2 Zweitkorrekturen

Seit dem Schuljahr 2003/2004 finden Zweitkorrekturen der Kompetenztests in den Fächern Deutsch und Mathematik³ an Stichproben statt. Diese wurden 2008 analog zu den Vorjahren durch die Fachberater durchgeführt. Mit Hilfe der Zweitkorrekturen wird geprüft, ob die Auswertungshinweise eindeutig umsetzbar waren und die Testergebnisse in diesem Sinne verlässlich sind. Die Durchführung durch die Fachberater hat zudem den Vorteil, dass diese damit eine Expertise hinsichtlich der Tests gewinnen, die für ihre Beratungsarbeit hilfreich sein kann.

Für die Stichprobe 2008 wurden 30 Grundschulen, 41 Regelschulen, 2 Gesamtschulen und 29 Gymnasien zufällig ausgewählt.

In diesen Schulen wurde nach einem festgelegten Schlüssel eine Stichprobe von Testheften bestimmt und an die Fachberater des jeweiligen Schulamtsbezirkes geschickt. Die zuständigen Fachberater kontrollierten die Testhefte und sendeten die Ergebnisse per E-Mail an die Universität Jena. Die Rücklaufquote der Zweitkorrekturen war insgesamt sehr gut, den Fachberatern sei an dieser Stelle ein Dank ausgesprochen.

Im Folgenden werden die Mittelwertsunterschiede der Gesamtergebnisse für die Erst- und Zweitkorrektur dargestellt. Positive Mittelwertsdifferenzen deuten darauf hin, dass die Erstkorrektoren, also die Fachlehrer, wohlwollender bewertet haben als die Zweitkorrektoren, also die entsprechenden Fachberater. Negative Differenzwerte, deuten eher auf eine zurückhaltendere, konservativere Bewertung des Fachlehrers hin. Zusätzlich wird untersucht, ob diese Differenzwerte in Zusammenhang zu verschiedenen Schülermerkmalen, der Schulart oder dem Schulamtsbereich stehen.

Als Gütemaß wird die Produkt-Moment-Korrelation r aufgeführt, um eine

Vergleichbarkeit mit den verwendeten Werten der Landesberichte aus vergangenen Jahren sicherzustellen. Dieser Korrelationskoeffizient gibt Auskunft über die relative Übereinstimmung zwischen den beiden Korrektoren. Er kann Werte zwischen -1 und +1 annehmen und liegt nahe +1, wenn die von den Fachlehrern als gut bewerteten Schüler auch vom Zweitkorrektor gut bewertet wurden und umgekehrt.

³ Für den Kompetenztest im Fach Englisch wurden keine Zweitkorrekturen durchgeführt, da dieser Test fast ausschließlich aus Aufgaben im Multiple-Choice-Format besteht. Bei diesem Aufgabentypus sind die Auswertungskriterien von Natur aus sehr eindeutig, so dass davon ausgegangen werden kann, dass die Auswertungsobjektivität hinreichend gegeben ist.

Zusätzlich wird der Koeffizient Cohens κ verwendet. Dieses Maß ist darauf ausgelegt, die absolute Übereinstimmung verschiedener Bewerter zu erfassen – dies entspricht genau dem Anliegen der Zweitkorrekturen. Eine hohe Übereinstimmung zwischen Erst- und Zweitkorrektur ist tendenziell mit höheren Kennwerten verbunden, wobei man diese wie folgt einschätzen kann: Bei statistisch signifikanten Reliabilitätsmaßen kann eine Größenordnung von 0,7 bereits als „gut“ angesehen werden (Wirtz & Caspar, 2002). Landis und Koch (1977) schlagen vor, ein κ von 0,41 bis 0,60 als „mittelmäßige“, ein κ von 0,61 bis 0,80 als „beachtliche“ und ein κ von 0,81 bis 1,00 als „(fast) vollkommene Übereinstimmung“ anzusehen.

Im Anhang werden die Analysen auf Aufgabenebene aufgeführt. Zur besseren Übersicht und um Fehlerschwerpunkte bei der Korrektur zu identifizieren sei an dieser Stelle angeraten, die entsprechenden Anhänge anzusehen.

2.2.1 Deutsch

DK3 - Kompetenztest Deutsch der Klassenstufe 3

Für den Kompetenztest Deutsch der Klassenstufe 3 ergibt sich für das aktuelle Jahr mit $r = 0,99$ ein sehr hoher Korrelationskoeffizient. Ähnlich hohe Koeffizienten fanden sich auch in den vergangenen Jahren für die Kompetenztests Deutsch in der Klassenstufe 3, bei denen zwei unabhängige Tests für die Teilbereiche Rechtschreiben/Sprachbetrachtung und Lesen durchgeführt wurden.

Im Mittel liegt die Differenz der Gesamtergebnisse von Erst- und Zweitkorrektur bei 0,33 Punkten. Dies bedeutet, dass die Fachlehrer die Ergebnisse ihrer Schüler leicht besser bewerten als die unabhängigen Zweitgutachter. Die Spanne dieser Abweichungen liegt zwischen -4 und +7 Punkten. In knapp der Hälfte der Fälle ($n = 47$, 49,5 %) gab es eine 100 %-ige Übereinstimmung in der Gesamtpunktzahl, während bei 17,9 % der Schüler ($n = 17$) das Ergebnis der Zweitkorrektur insgesamt positiver ausfiel als die Korrektur durch den Fachlehrer. Bei den verbleibenden 31 Testheften (32,6 %) bewertete der Fachlehrer insgesamt besser.

Betrachtet man, wie eben geschehen, das Gesamtergebnis der Tests, so ist die Differenz zwischen Erst- und Zweitkorrektur als Maß für die Genauigkeit dahingehend unscharf, als dass gegenläufige Unterschiede bei einzelnen Aufgaben sich ausgleichen können. So ist es möglich, dass durch eine gleiche Anzahl positiver und negativer Abweichungen bei den Bewertungen insgesamt herauskommt, dass kein Unterschied existiert, obwohl die Bewertungen einzelner Aufgaben höchst unterschiedlich ausfallen.

Dieses Problem stellt sich nicht, wenn man Koeffizienten auf Aufgabenebene betrachtet. Für die einzelnen Aufgaben wurde wie auch im letzten Jahr Cohens κ berechnet.

Die κ -Werte für die einzelnen Aufgaben liegen fast ausschließlich im Bereich zwischen 0,81 und 1. Drei Aufgaben bewegen sich außerhalb dieses Bereiches mit Koeffizienten zwischen $\kappa = 0,70$ und $\kappa = 0,77$. Zwar sind diese Aufgaben tatsächlich diejenigen, bei denen die Bewertungen der beiden Gutachter am häufigsten nicht übereinstimmen.⁴ Allerdings wird der geringere Koeffizient auch durch ein statistisches Artefakt mitbedingt: Wenn die Fachlehrer und die Zweitgutachter sich systematisch darin unterscheiden, ab welchem Schwellenwert sie eine Aufgabe als gelöst ansehen, wird der κ -Wert systematisch unterschätzt. Dies ist bei diesen Aufgaben der Fall. Die Fachlehrer bewerten die Aufgaben statistisch bedeutsam häufiger als

⁴ Jeweils bei 11 von insgesamt 98 zweitkorrigierten Testheften stimmen die Bewertungen von Fachlehrer und Zweitgutachter nicht überein. Betroffen sind die Aufgaben 1.7c, 2.2 und 3.8b.

gelöst als die Fachberater. Letztere scheinen also insbesondere bei nicht eindeutigen Bewertungskriterien einen höheren Bewertungsmaßstab anzulegen. Der für alle Aufgaben des Tests gemittelte κ -Wert liegt bei 0,93 und ist äußerst zufrieden stellend. Die Reliabilitätskoeffizienten für die Tests des vergangenen Schuljahres waren nahezu identisch.

Es sind keine Auffälligkeiten festgestellt worden, die die Richtung der Abweichung zwischen Erst- und Zweitkorrektur in Zusammenhang mit verschiedenen Schülermerkmalen bringen.⁵ Auch zeigten sich keine bedeutsamen Unterschiede zwischen verschiedenen Schulämtern. Es kann davon ausgegangen werden, dass der Deutschtest der Klassenstufe 3 wie in den vergangenen Jahren sehr objektiv bewertet werden konnte und dies auch entsprechend pflichtbewusst durch die Fachlehrer durchgeführt wurde.

DK6 - Kompetenztest Deutsch der Klassenstufe 6

Die Korrelation zwischen den Gesamtergebnissen von Erst- und Zweitkorrektur ist auch im Kompetenztest Deutsch der Klassenstufe 6 mit $r = 0,96$ als sehr hoch zu bewerten.

Ein differenzierteres Bild zeigt sich, wenn man die κ -Werte betrachtet. Über alle Aufgaben gemittelt ergibt sich ein Koeffizient in Höhe von $\kappa = 0,77$. Dieser Wert verweist zwar immer noch auf eine gute bzw. beachtliche Übereinstimmung zwischen den verschiedenen Korrektoren, liegt aber deutlich unterhalb der Werte, die für die anderen Kompetenztests gefunden wurden und auch unterhalb des Wertes aus dem vergangenen Schuljahr. Bei genauerer Betrachtung zeigt sich, dass Bewertungsunsicherheiten vor allem im ersten Teil des Deutschtests bei der Aufgabe Brieffreundschaften auftraten, bei der die Schüler dazu aufgefordert waren, einen Text frei zu produzieren. Dieser Text musste vom Lehrer nach verschiedenen Kriterien beurteilt werden. Hier hatten vier von 20 Aufgaben κ -Werte unter 0,41. Nur drei Aufgaben des ersten Teils hatten κ -Werte über 0,70. Der über alle Aufgaben des ersten Teils gemittelte Koeffizient beträgt $\kappa = 0,51$. Im zweiten Teil des Deutschtests waren die Koeffizienten hingegen sehr zufriedenstellend. Sie bewegten sich im Bereich zwischen $\kappa = 0,88$ und 1. Für diesen Teil ergab sich ein Durchschnittswert von $\kappa = 0,95$.

Die Abweichungen zwischen Erst- und Zweitkorrektur in Bezug auf das Gesamtergebnis eines Schülers im Test reichten von -13 bis +13 Punkten. Diese erscheinen auf den ersten Blick vor allem im Vergleich zur Differenz bei anderen Tests groß. Allerdings war die Maximalpunktzahl für diesen Test mit 96 Punkten auch am größten. Negative Differenzen und damit eine bessere Bewertung durch den Zweitgutachter zeigten sich bei 77 Schülern (29,4 %). Bei 102 Schülern (38,9 %) urteilte der Fachlehrer besser. Bei 83 Schülern stimmten die Ergebnisse der Erst- und Zweitkorrektur überein. Insgesamt beträgt der Mittelwert für die Differenz 0,01 Punkte. Man kann also davon ausgehen, dass weder die Fachlehrer noch die Fachberater systematisch andere Bewertungsmaßstäbe ansetzen.

Auch für den Deutschtest der Klassenstufe 6 wurde untersucht, ob sich Zusammenhänge zwischen verschiedenen Schülermerkmalen und der Differenz zwischen den Ergebnissen von Erst- und Zweitkorrektur zeigen. Im letzten Jahr lag hier ein Zusammenhang zur Halbjahresnote vor. Schüler mit besserer Note wurden von ihrem Fachlehrer tendenziell besser bewertet als von den Fachberatern, während Schüler mit schlechterer Note vom Zweitkorrekteur besser bewertet wurden. Solch ein Effekt konnte dieses Jahr für die Halbjahresnote nicht nachgewiesen werden. Allerdings zeigte sich, dass im Mittel unterschiedliche Bewertungsdifferenzen auftreten in Abhängigkeit davon, ob bei einem Schüler eine besondere Lernschwierigkeit oder

⁵ In die Analysen wurden Halbjahresnote, Muttersprache, Geschlecht, besondere Lernschwierigkeiten, Beschuldungsdauer und die Bücherfrage einbezogen.

sonderpädagogischer Förderbedarf (BLSF) vorliegt. Ist dies der Fall, dann bewertete der Zweitkorrekteur im Mittel 2 Punkte besser als der Fachlehrer. Anderenfalls fiel die Bewertung des Fachlehrers im Mittel um 0,9 Punkte besser aus. Dieser Effekt sollte allerdings mit Vorsicht interpretiert werden, da unter den 262 Schülern, deren Testhefte einer Zweitkorrektur unterzogen wurden, nur bei 14 Schülern eine BLSF angegeben wurde. Auch zeigte sich, dass die mittleren Differenzen zwischen verschiedenen Schulämtern variieren. Es konnten allerdings keine einzelnen Schulämter ausgemacht werden, die sich signifikant von anderen abgrenzen.

2.2.2 Mathematik

Betrachtet man den Korrelationskoeffizienten r , so zeigt sich für die Kompetenztests im Fach Mathematik ein fast perfekter Zusammenhang zwischen der Einschätzung der Schülerleistungen durch den Fachlehrer und die unabhängigen Zweitgutachter. Für jede der drei Klassenstufen beträgt $r = 0,99$. Ähnlich hohe Koeffizienten fanden sich auch in den Vorjahren.

MK3 - Kompetenztest Mathematik der Klassenstufe 3

Der Reliabilitätskoeffizient Cohens κ liegt für die Aufgaben des Mathematiktests in der Klassenstufe 3 zwischen $\kappa = 0,75$ und $\kappa = 1$. Der über alle Aufgaben gemittelte Wert beträgt $\kappa = 0,94$. Auch dieser Koeffizient bestätigt also, dass die Bewertungen von Erst- und Zweitkorrekturen insgesamt sehr gut übereinstimmen und damit eine objektive und eindeutige Bewertung der Testleistungen möglich ist.

Eine 100 %-ige Übereinstimmung liegt allerdings nicht für jeden Schüler vor. Bei 13 Schülern (9,9 %) ist die Differenz zwischen Erstkorrektur und Zweitkorrektur negativ, hier bewertet also der Zweitkorrekteur das Testheft besser. Bei der großen Mehrheit der Schüler gibt es keine Unterschiede im Gesamttestergebnis (86 Schüler, 65,6 %) und bei 32 Schülern (24,4 %) fällt das Ergebnis der Erstkorrektur besser aus als das der Zweitkorrektur. Der Prozentsatz der Schüler, bei denen keine Abweichungen vorliegen, ist zwar in diesem Jahr geringer als im Letzten. Allerdings enthielt der Kompetenztest des letzten Jahres auch deutlich weniger Aufgaben.

Die Spanne der Abweichungen zwischen Erst und Zweitkorrektur reicht von -5 bis +3 Punkten. Insgesamt konnten in diesem Test maximal 43 Punkte erreicht werden. Im Mittel weichen die Ergebnisse um 0,19 Punkte voneinander ab. Damit bewerten auch in diesem Test die Fachlehrer die Ergebnisse Ihrer Schüler tendenziell etwas besser.

Auf Seiten möglicher Einflussvariablen auf die Differenz zwischen Erst- und Zweitkorrektur zeigt sich auf Schülerebene, dass das Geschlecht der Schüler einen bedeutsamen Einfluss hat. So gibt es im Mittel praktisch keine Unterschiede zwischen Erst- und Zweitkorrektur bei Mädchen (MW = 0,03, $n = 66$). Bei Jungen hingegen bewerten die Fachlehrer die Tests im Mittel etwas besser als die unabhängigen Zweitgutachter (MW = 0,34, $n = 64$). Andere Variablen auf Schülerebene zeigten keine bedeutsamen Zusammenhänge mit der Differenz. Auch zeigten sich keine bedeutsamen Unterschiede zwischen verschiedenen Schulämtern.

MK6 - Kompetenztest Mathematik der Klassenstufe 6

Cohens κ bewegt sich für die Aufgaben des Mathematiktests der Klassenstufe 6 zwischen $\kappa = 0,87$ und $\kappa = 1$. Im Mittel ergibt sich ein sehr guter Wert von $\kappa = 0,96$, der identisch ist mit dem des Vorjahres.

Insgesamt konnten im Test 28 Punkte erreicht werden. Die mittlere Differenz zwischen Erst- und Zweitkorrektur beträgt 0,17 Punkte. Individuell reichen die Abweichungen von -3 bis +4 Punkten. Bei 141 Schülern (69,5 %) zeigten sich für die Gesamtergebnisse von Erst- und Zweitkorrektur keine Unterschiede. In 22,2 % der Fälle (45 Schüler) fiel die Bewertung in der Erstkorrektur besser aus, für 17 Schüler (8,4 %) in der Zweitkorrektur.

Keine der betrachteten Schülervariablen zeigte einen bedeutsamen Zusammenhang zur Differenz zwischen Erst- und Zweitkorrektur. Auch zeigen sich keine bedeutsamen Unterschiede zwischen verschiedenen Schularten und Schulämtern.

Ein objektiv bewertbarer Test mit wenigen geringen Unterschieden zwischen den Korrekturen liegt demnach auch im Mathematiktest der Klassenstufe 6 vor.

MK8 - Kompetenztest Mathematik der Klassenstufe 8

Bis auf zwei Items befinden sich die κ -Koeffizienten für den Mathematiktest der Klassenstufe 8 in einem Bereich oberhalb der 0,8-Grenze. Bei Aufgaben, die unterhalb dieser Grenze lagen⁶, handelt es sich um Items, in denen eine Begründung bewertet werden sollte. Der über alle Aufgaben gemittelte κ -Wert ist ähnlich hoch wie der des Vorjahres und beträgt 0,94.

Bei 53,4 % ($n = 111$) der Schüler gab es keine Unterschiede zwischen den Gesamtergebnissen von Erst- und Zweitkorrektur. In 44,1 % der Fälle (71 Schüler) bewertete der Fachlehrer die Testhefte besser, in 12,5 % der Fälle (26 Schüler) der Zweitkorrekteur. Insgesamt liegt der Mittelwert der Differenz zwischen Erst- und Zweitkorrektur bei 0,33 Punkten. Die Differenz auf Schülerebene reichte dabei von -4 bis +5 Punkten. Angesichts der Tatsache, dass in diesem Test ein Maximum von 58 Punkten erreichbar war, erscheint diese Spannweite als gering.

Wie im Kompetenztest Deutsch in Klassenstufe 6 zeigte sich, dass ein Zusammenhang besteht zwischen dem Vorliegen einer besonderen Lernschwäche oder von sonderpädagogischem Förderbedarf und der Differenz zwischen Erst- und Zweitkorrektur. Schüler mit BLSF werden dabei vom Fachlehrer kritischer bewertet als vom Zweitkorrekteur ($MW = -0,29$), während Schüler ohne BLSF ($MW = 0,35$) im Mittel vom Zweitgutachter kritischer bewertet werden. Auch hier gilt aber, dass der Effekt mit Vorsicht zu interpretieren ist, da nur sehr wenige Schüler die Diagnose BLSF erhielten. Um auszuschließen, dass der Effekt nur zufällig zustande kam, sollte seine Existenz in den nächsten Jahren überprüft werden. Ebenso analog zum Deutschtest der Klassenstufe 6 zeigten sich bedeutsame Unterschiede zwischen Schulämtern, die jedoch nicht durch ein Herausstechen einzelner oder gar eines einzelnen Schulamtes charakterisiert sind.

Außerdem zeigte sich ein bedeutsamer Unterschied zwischen den Schularten. Die mittlere Differenz zwischen Erst- und Zweitkorrektur war bei Schülern des Gymnasiums ($MW = 0,62$, $n = 91$) größer als bei Regelschülern ($MW = 0,14$, $n = 110$).

Trotz dieser Zusammenhänge kann davon ausgegangen werden, dass auch in diesem Test eine objektive Bewertung möglich war, die durch die Fachlehrer pflichtbewusst durchgeführt wurde.

⁶ Aufgabe 9 mit $\kappa = 0,68$ und Aufgabe 24,2 mit $\kappa = 0,77$

2.2.3 Zusammenfassung und Aussichten

Zusammenfassend ist zu sagen, dass sowohl in den Kompetenztests in Mathematik, als auch in Deutsch beachtliche Übereinstimmungen zwischen den Ergebnissen der Erstkorrektur und der Zweitkorrektur erzielt wurden. Dies spricht insgesamt für eine objektive Korrigierbarkeit in den Tests.

Die hohen Korrelationen, die bei allen Tests in allen Klassenstufen zu verzeichnen sind, machen deutlich, dass die Objektivität der Auswertung auf Ebene des Gesamtests gegeben ist.

Die hohen κ -Koeffizienten bezeugen zudem, dass auch die absoluten Testergebnisse gut interpretierbar sind. Eine Ausnahme bilden hier einzelne Teilaufgaben des Deutschtests der Klassenstufe 6. Es ist damit zu rechnen, dass die Leistung des einzelnen Schülers im ersten Teil des Tests besonders bei der Aufgabe Brieffreundschaften nur sehr ungenau gemessen werden konnte. Das Gesamtergebnis sollte allerdings interpretierbar sein.

Die Ursachen für die zum Teil sehr unbefriedigenden κ -Koeffizienten für diese Aufgabe sind zum einen im Aufgabenformat zu sehen. Es handelt sich hierbei um eine freie Textproduktion, die vom Lehrer nach verschiedenen Kriterien bewertet werden soll. Auch in vorherigen Kompetenztestjahrgängen hat sich immer wieder gezeigt, dass eine wirklich objektive Bewertung solcher Aufgaben nur schwer möglich ist. Um aber ein größeres Spektrum an Inhalten und Anwendungen in die Leistungserhebung einzubeziehen, sind solche Aufgabenformate unverzichtbar. Daher sollte das Ziel bei der künftigen Anwendung solcher Aufgaben sein, die Kriterien so eindeutig wie möglich zu gestalten und, wie z. T. bereits verwirklicht, mit vielen Auswertungsbeispielen zu versehen. Zum anderen mussten in diesem Jahr Änderungen am Auswertungsmanual vorgenommen werden. Dies geschah zu einem Zeitpunkt, zu dem die Tests bereits geschrieben und zum Teil auch schon durch die Fachlehrer ausgewertet waren. Die Änderungen bezogen sich auf eine Teilaufgabe der Aufgabe Brieffreundschaften. Es ist daher nicht auszuschließen, dass einige der Diskrepanzen in den Auswertungen darauf zurückzuführen sind, dass Erst- und Zweitgutachter unterschiedliche Manuale verwendeten. Diese Möglichkeit hat für die Diskrepanzen in den Korrekturen allerdings nur einen kleinen Erklärungswert.

Bei der Betrachtung der Differenzwerte zwischen den Gesamtergebnissen von Erst- und Zweitkorrektur zeigte sich durchgängig, dass der Fall, dass die Fachlehrer die Testhefte besser bewerten als die Zweitgutachter, häufiger vorkam als der umgekehrte. Man kann daraus schließen, dass die Fachberater die Testhefte allgemein etwas kritischer bewerten. Der Effekt dieser Bewertungstendenz fällt allerdings nur gering aus. Dies wird deutlich, wenn man die mittleren Differenzwerte betrachtet. Diese sind entweder klein oder praktisch nicht vorhanden, wie im Deutschtest der Klassenstufe 6.

Systematische Zusammenhänge zwischen den Differenzwerten und anderen Faktoren konnten allenfalls testspezifisch gefunden werden, ein über alle Tests und Klassenstufen einheitliches Muster zeigte sich nicht. Keinerlei Einflussfaktoren konnten für den Deutschtest der Klassenstufe 3 und den Mathematiktest der Klassenstufe 6 gefunden werden. Für den Mathematiktest der Klassenstufe 3 zeigte sich, dass die Fachlehrer nur bei Jungen im Mittel leicht besser bewerteten als die Zweitgutachter, nicht aber bei Mädchen. Im Deutschtest der Klassenstufe 6 und im Mathematiktest der Klassenstufe 8 gab es gewisse Unterschiede zwischen Schulämtern und zwischen Schülern mit und ohne besonderen Lernschwierigkeiten bzw. sonderpädagogischem Förderbedarf. Möglicherweise lassen Fachlehrer bei der Bewertung von Aufgaben, bei denen die Auswertungskriterien weniger eindeutig sind, ihre vorherigen Erfahrungen mit den Schülern einfließen. Bei Schülern mit BLSF ist es wahrscheinlicher, dass sie im bisherigen Unterricht

eher dadurch aufgefallen sind, dass sie bestimmte Leistungsanforderungen nicht bewältigen können. Dies könnte erklären, warum die Fachlehrer Schüler mit BLSF kritischer bewerten als die Zweitkorrekteure, die über keine Informationen zum bisherigen Lernverlauf der Schüler verfügen. Aus bereits erwähnten Gründen sollten diese Effekte allerdings mit Vorsicht interpretiert werden. Auch die Tatsache, dass die Variablen, die im vergangenen Jahr einen statistisch bedeutsamen Zusammenhang zur Differenz zwischen Erst- und Zweitkorrektur aufwiesen, dieses Jahr keine Rolle spielten, verweist darauf, dass es sich um zufällige auftretende und keinesfalls stabile Effekte handeln könnte.

3 Durchführung und Datenerfassung

Die Kompetenztests 2008 wurden in der Klassenstufe 3 am 6. Mai (Deutsch) und am 8. Mai (Mathematik) geschrieben. In der Klassenstufe 6 erfolgte die Testdurchführung am 21. Februar (Deutsch), 26. Februar (Mathematik) und 28. Februar (Englisch). Der Mathematiktest in Klassenstufe 8 wurde am 4. März durchgeführt.

3.1 Anmeldung zu den Kompetenztests 2008 und zum Schülerlängsschnitt

Im Jahr 2008 wurde fast jeder Schüler, der an den Thüringer Kompetenztests teilnahm, auch am Thüringer Schülerlängsschnitt angemeldet. Dazu wurde ein Erhebungsprogramm entwickelt. Mit diesem Programm wurden die Stammdaten der Schüler für den Thüringer Schülerlängsschnitt erfasst und an das Thüringer Landesrechenzentrum gesendet. Die daran angeschlossene Anmeldung der Schüler zu den Kompetenztests erfolgte in anonymisierter Form. Für die Dateneingabe und die Ergebnisrückmeldung wurde ein Schülercode erstellt¹. Durch dieses Vorgehen wurde der Thüringer Schülerlängsschnitt realisiert. Dieser Schülerlängsschnitt ist dadurch gekennzeichnet, dass weder die Universität Jena (Datenauswertung) noch das Thüringer Landesrechenzentrum (Treuhand der Stammdaten aller Schüler) Leistungsdaten zusammenführen können, ohne dass dieses Vorgehen durch das Thüringer Kultusministerium autorisiert wird. Eine solche Autorisierung kann vom Thüringer Kultusministerium nur für konkrete Analysezwecke erteilt werden und ist zeitlich begrenzt. Eine gemeinsame Speicherung von Stammdaten und Leistungsdaten erfolgt nicht zentral sondern ausschließlich in den Schulen. Dies war nur mit Hilfe einer offline Datenerfassung möglich, da auf diese Weise die Stammdaten getrennt von den Schülercodes bleiben und so eine Zuordnung durch Dritte verhindert wird.

Das Erhebungsprogramm wurde allen Thüringer Schulen auf CD zur Verfügung gestellt (Verteilung über die Thüringer Schulämter). Das Programm kann in den folgenden Jahren erneut verwendet werden. Aktualisierungen werden den Schulen durch ein Update über das Internet ermöglicht.

Die Datenerfassung der Ergebnisse der Thüringer Kompetenztests erfolgte wie in den vergangenen Jahren webbasiert. Zur Identifikation der Schüler wurde ein Schülercode verwendet, dessen Zuordnung zu einzelnen Schülern nur der Schule bekannt war. Das webbasierte Verfahren ist im Speziellen in Abschnitt 3.3 beschrieben.

Auch dieses Jahr konnte wieder die Anzahl der Schulen, die das Erhebungsprogramm nicht benutzen konnten, verringert werden: Nur wenige Schulen mussten ihre Klassen online anmelden und erhielten auf diese Weise für jede Klasse eine Liste mit Schülercodes. Eine Zuordnung der Codes zu Schülern sollte dann selbst im Vorfeld der Tests vorgenommen werden. Diese Schüler können nicht am Schülerlängsschnitt teilnehmen, eine Zuordnung der Ergebnisse der verschiedenen Fächer eines Schülers wird jedoch ermöglicht.

¹ Für den Datenschutz ist wichtig, dass dieser Schülercode nur für die Durchführung der Kompetenztests 2008 und nur für die betreffende Schule eindeutig ist. Eine Zuordnung von Schülern über diesen Schülercode ohne das Thüringer Landesrechenzentrum ist nicht möglich.

3.2 Teilnahme an den Kompetenztests 2008

Es nahmen insgesamt 40 551 Schüler aus 849 Schulen an den Kompetenztests teil. Die genauen Teilnehmerzahlen, gegliedert nach Klassenstufe und Schulart, zeigt Tabelle 3.1.

Tabelle 3.1: Teilnehmerzahlen an den Kompetenztests 2008

Klassenstufe	Schulart	Anzahl Schulen	Anzahl Klassen	Anzahl Schüler
3	Förderschule	29	38	337
	Grundschule	463	817	15 013
6	Förderschule	38	42	361
	Gymnasium	94	251	5 733
	Regelschule	249	405	7 248
8	Förderschule	37	45	346
	Gymnasium	95	224	4 633
	Regelschule	249	457	6 880

Die Teilnehmezahlen insgesamt haben die gleiche Größenordnung wie im Vorjahr. In Klassenstufe 3 sank in den Grundschulen trotz geringer Erhöhung der Anzahl der Klassen die Zahl der Schüler leicht ab (die Klassengröße² sank im Durchschnitt auf 18,37 Schüler im Vergleich zu 19,1 im Vorjahr). In Klassenstufe 6 stieg dagegen die Zahl der Schüler im Vergleich zum Vorjahr leicht an.

3.3 Webbasierte Datenerfassung

Analog zu den letzten Jahren wurden die Testhefte von den Fachlehrern gemäß der vorgegebenen Auswertungsmanuale korrigiert und die Ergebnisse anschließend im Schulportal auf *kompetenztest.de* für jeden Schüler itemweise eingegeben.

Das im letzten Jahr entwickelte AJAX-Eingabesystem wurde aufgrund der guten Resonanz mit nur minimalen Anpassungen auch in diesem Jahr eingesetzt. Die Neuprogrammierung 2007 hatte eine Beschleunigung sowie einfachere Gestaltung der Eingabe zum Ziel. Beides konnte eindrucksvoll erreicht werden, wie auch die Analysen der Logdateien und Lehrerbeurteilungen zeigten. Im Landesbericht Thüringer Kompetenztests 2007 findet sich eine ausführlichere Darstellung der Unterschiede und Vorteile des neuen Systems.

Obwohl die Systemvoraussetzungen zum Einsatz der AJAX-Eingabe bewusst niedrig gewählt wurden (Microsoft® Internet Explorer® ab Version 5.5 oder Firefox ab Version 1.5, seit 2001 bzw. 2005 auf dem Markt), wurden wieder alle Eingaben zusätzlich in der alten Version programmiert, so dass die Dateneingabe auch auf PCs mit älterer Softwareausstattung vorgenommen werden konnten. Die Auswahl der Eingabeversion erfolgte dabei wie im letzten Jahr automatisch beim Start im Portal. In diesem Jahr konnten 94 % der fast 100 000 Datenzeilen mit der neuen AJAX-Eingabe eingegeben werden, erwartungsgemäß eine Steigerung gegenüber 90 % im Vorjahr. Dabei war 2008 der Anteil der neuen Version in allen Schularten ähnlich hoch.

² Die Zahl der Testeingaben pro Klasse wird hier als Maß für die Klassengröße verwendet

Die Dateneingabe der Test- und soziodemographischen Daten der Schüler erfolgte in beiden Versionen wie in den letzten Jahren zusammen mit den Schülercodes. Alle Lehrer mussten dazu die bei der Anmeldung der Klassen erzeugten Schülercode-Listen erhalten, ohne die die Eingabe der Testdaten nicht möglich war. Pro Klasse konnten zwei nicht angemeldete Schüler mit dem Code „999“ eingegeben werden. In diesem Jahr wurde ca. 540 mal davon Gebrauch gemacht, was fast einer Verdopplung der Anzahl im Vergleich zum letzten Jahr entspricht. Erklärbar wird diese Zunahme sicherlich mit dem frühen Anmeldetermin der Kompetenztests 2008 (Dezember/Januar) und der langen Zeitspanne bis zum letzten Test. Angesichts von ca. 40 000 Schülern spricht jedoch auch diese Zahl unter diesen Umständen noch immer für die Güte des verwendeten Kodiervorgangs.

Die Begleitbefragung der Lehrer wurde auch 2008 zeitgleich mit der Eingabe durchgeführt. Es wurden wieder mehrere, überlappende Versionen der Begleitfragebögen erstellt, um die Anzahl der Items pro Lehrer zu verringern. Ebenso wurde sichergestellt, dass jeder Lehrer nur eine Version der Begleitfragebögen ausfüllen musste, aber von jeder Version eine für wissenschaftliche Aussagen ausreichende Zahl angefordert wurde. Die Items aller Begleitfragebögen sind im Anhang aufgeführt.

Die zeilenweise Übermittlung der Daten an die Universität Jena erlaubt eine grobe Analyse der Eingabedauer einer Zeile. Minimal verfälscht werden diese durch die nicht messbaren Zeiten der Übertragungen über das Internet. In die Zeiten fließen auch Pausen und andere Aktivitäten (zu große Zeiten) sowie nicht bearbeitete oder nur korrigierte Zeilen (zu kleine Zeiten) mit ein; daher werden hier jeweils die kleinsten sowie größten 15 % der Zeiten-Werte nicht betrachtet. Tabelle 3.2 gibt einen Überblick über die verbleibenden 70 % der Eingabezeiten in Sekunden, getrennt nach bearbeitetem Test. Dargestellt sind jeweils die 50. Perzentile (Median), die der mittleren Dauer der Eingabe einer Zeile entsprechen. Ebenfalls sind die 25. und 75. Perzentile dargestellt. Die Hälfte aller analysierten Zeilen-Eingaben hatte eine Dauer im Bereich zwischen diesen beiden Werten. Eine getrennte Auswertung für die beiden Versionen des Eingabesystems wird aufgrund der unbedeutend werdenden Nutzung der alten Version nicht mehr vorgenommen.

Tabelle 3.2: Zusammenfassung der Eingabezeiten pro Zeile in Sekunden

	Klassenstufe 3		Klassenstufe 6				MK8
	DK3	MK3	DK6	MK6	EK6H	EK6L	
25. Perzentil	54	51	63	39	48	41	68
50. Perzentil	66	62	78	47	60	49	86
75. Perzentil	82	77	98	58	76	60	109
Anzahl Items	51	47	54	34	49	52	66

Wie in den letzten Jahren auch ist die Zunahme der Eingabezeit mit Zunahme der Zahl der Items deutlich erkennbar. Ausnahme bildet hierbei der Teil Leseverständnis des Englischtests, der trotz einer höheren Itemzahl als der Teil Hörverstehen im Mittel kürzere Eingaben ermöglichte. Erklärt werden könnte dies damit, dass schülerspezifische Stammdaten nur im Teil Hörverstehen eingegeben werden mussten und diese Eingaben (eventuell durch Hinzunahme anderer Quellen) mehr Zeit kosten.

Die Eingabezeiten sind ohne Berücksichtigung des Tests bei ähnlicher Itemanzahl mit den Zeiten des Vorjahres vergleichbar. Allerdings fällt auf, dass in diesem Jahr deutlich mehr Items eingegeben werden mussten. Eine Ausnahme bildet hierbei nur der Mathematiktest in Klassenstufe 6. Bei den anderen Tests mussten in diesem Jahr 38 % bis 68 % mehr Items eingegeben werden; im Fach Englisch hat sich die Zahl der einzugebenden Items sogar fast verdoppelt.

Dies hat natürlich auch verlängerte Eingabezeiten zur Folge. Durchschnittliche Eingabezeiten einer Zeile von unter einer Minute konnten in diesem Jahr nur noch für den MK6-Test erreicht werden. Die mittlere Eingabezeit für eine ganze Klasse von knapp 20 Schülern liegt damit deutlich über dem letztjährigen Wert von 15 bis 20 Minuten und erreicht bei Mathematik Klassenstufe 8 oder Englisch Klassenstufe 6 die 30 Minuten.

Die Analyse der Zeitpunkte der Dateneingaben zeigt eine deutliche Häufung der Eingaben von Montag bis Donnerstag in der Zeit von 10:00 Uhr bis 14:00 Uhr. Abbildung 3.1 stellt die Verteilung der Eingaben der Datenzeilen über die Wochentage und Tageszeiten grafisch dar. Es ist hier auch zu erkennen, dass ein großer Anteil der Eingaben von den Lehrern zu nicht schultypischen Arbeitszeiten vorgenommen wird. So wurden fast 20 % aller Datenzeilen wochentags nach 19:00 Uhr oder am Wochenende eingegeben. Darunter hatten die Deutsch- und Englischtests in Klassenstufe 6 einen deutlich erhöhten Anteil von Wochenendeingaben.

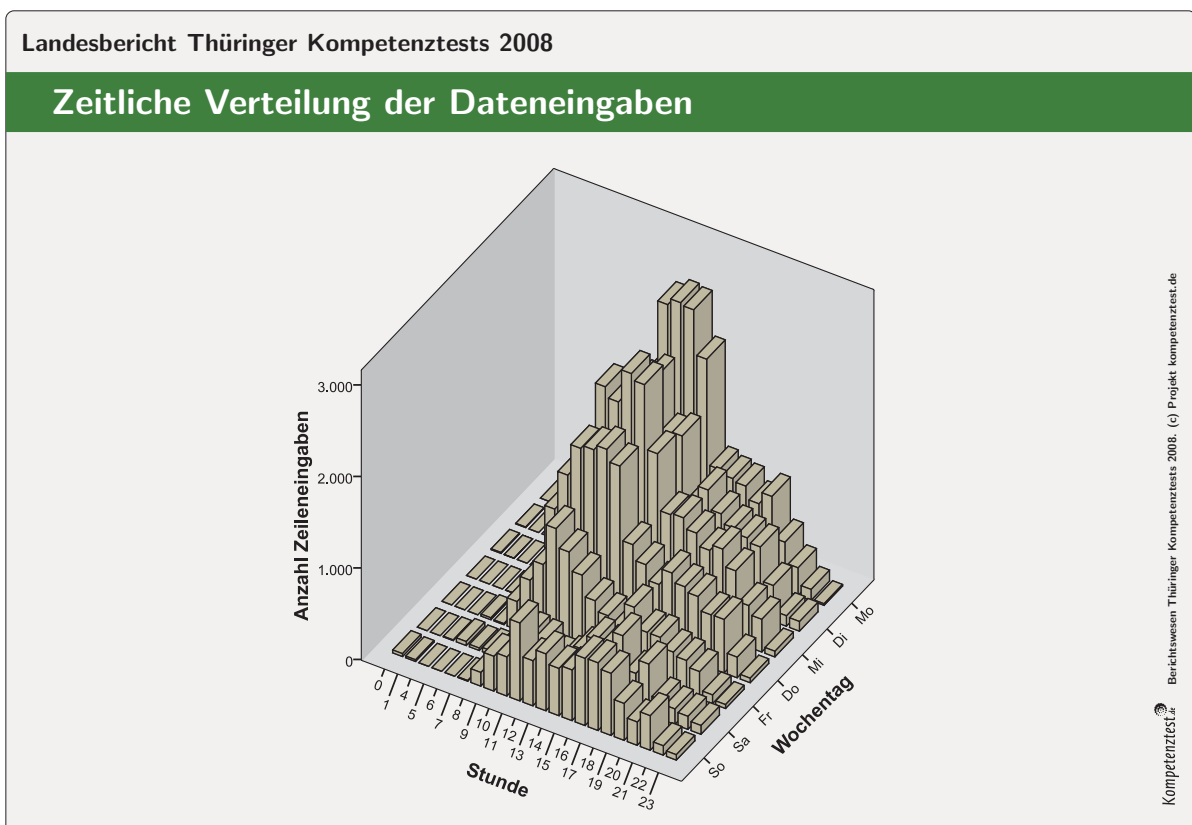


Abbildung 3.1: Verteilung der Dateneingaben (Zeilen) über die Wochentage und Tageszeiten

Die Verteilung der Eingaben zeigt über den gesamten Eingabezeitraum einen erstaunlich konstanten Anstieg. Von den Lehrern wird also der gesamte Zeitraum der Eingabe (ca. 4-5 Wochen) gleichermaßen genutzt. Es ist lediglich eine leicht verzögerte Eingabe der Deutschtests bemerkbar, die sicherlich durch eine längere Korrekturzeit der Aufgaben erklärt werden kann. Dieser Effekt zeigt sich besonders deutlich in der Klassenstufe 6.

Direkt nach der Eingabe hatten die Lehrer auch in diesem Jahr die Möglichkeit, die Dateneingabe zu bewerten. Tabelle 3.3 zeigt die Ergebnisse dieser Bewertung getrennt für die verschiedenen Tests. Während nur für die beiden Mathematiktests die Wertungen des letzten Jahres gleich gut oder besser ausgefallen sind, wurden die anderen Tests 0,1 bis 0,3 Notenpunkte schlechter bewertet. Sicherlich ist die Zunahme der Zahl der Items und damit auch der Dauer der Eingabe ein möglicher Grund hierfür.

Eine bessere Beurteilung der Eingaben durch die Grundschullehrer ist in diesem Durchgang nicht so deutlich erkennbar wie in den Vorjahren. Der Gesamtdurchschnitt von 1,8 über alle Tests spricht trotz geringer Verschlechterung im Jahresvergleich für eine große Zustimmung zum verwendeten Online-System.

Tabelle 3.3: Zusammenfassung der Bewertungen in Schulnoten 1–6 („Welche 'Note' würden Sie der Dateneingabe im Internet geben?“) und die Anzahl der jeweils zu tätigen Eingaben (Spalten)

	Klassenstufe 3		Klassenstufe 6				MK8
	DK3	MK3	DK6	MK6	EK6H	EK6L	
Mittelwert (Streuung)	1,7 (0,8)	1,7 (0,7)	2,0 (0,8)	1,7 (0,7)	1,9 (0,9)	1,9 (0,8)	1,9 (0,8)
Anzahl Items	51	47	54	34	49	52	66

Die in diesem Jahr zum Teil deutlich gestiegene Zahl von Testitems stellte für die Lehrer eine zusätzliche Belastung dar, die z. B. zu längeren Eingabezeiten führte. Natürlich steigt damit auch immer der Aufwand für die Korrektur der Aufgaben und Analyse der Ergebnisse, der hier nicht erfasst werden konnte. Zum Teil haben sich diese Aspekte auch in den leicht gesunkenen Bewertungen durch die Lehrer gezeigt. In den Folgejahren sollte der Abwägung von Nutzen und Kosten einer möglichst umfangreichen Erfassung der Ergebnisse auf der Ebene von Teilaufgaben mehr Beachtung geschenkt werden.

3.4 Auswertung der Hotline

Auch in diesem Jahr hat sich die Unterstützung der teilnehmenden Schulen durch die Hotline als hilfreich erwiesen. Die Lehrer konnten sich zu jeder Zeit per Telefon, Email oder Fax an das Projekt *kompetenztest.de* wenden, wo ihre Anfragen und Probleme möglichst zeitnah bearbeitet wurden. Die telefonische Hotline wurde zweimal wöchentlich für jeweils zwei Stunden angeboten. In besonders problematischen Phasen wurde die telefonische Erreichbarkeit pro Woche spontan verlängert, um alle auftretenden Schwierigkeiten lösen zu können. Außerhalb der Hotlinezeit wurde vor allem auch auf schriftliche Anfragen reagiert. Dabei wurden die Schulen häufig gesondert telefonisch kontaktiert, um möglichst schnell erste Problemlöseschritte einzuleiten.

Die Heterogenität der auftretenden Probleme hat sich auch in diesem Jahr nicht wesentlich verändert. Es gab Komplikationen mit dem Erhebungsprogramm bei der Anmeldung der Klassen und der erfolgreichen Aktualisierung des Programms über das Internet. Die Betreuung und Behebung der Probleme wurde dieses Jahr erfolgreich durch die Unterstützung der Hotline bewältigt. Technische Schwierigkeiten beim Upload der erzeugten Dateien an das Thüringer Landesrechenzentrum (TLRZ) traten leider auch dieses Jahr wieder auf, die Verzögerungen hielten sich aber erfreulicherweise in Grenzen. Gelegentlich kam es auch wieder vor, dass durch Änderungen von Schülerdaten im Erhebungsprogramm nach dem Upload der Dateien die Schülercode-Listen von der ersten Version abwichen. Die Mitarbeiter des Projekts konnten dann neue Schülercode-Listen erzeugen und per Mail den betreffenden Schulen zusenden.

In diesem Jahr gab es auch wieder die üblichen Passwortanfragen. Wenn Schulen ihr Passwort als Zugang zum Schulportal verlegt haben, konnte dies wie auch in den letzten Jahren, problemlos per Fax nach dem üblichen Prozedere mitgeteilt werden. Wegen teils mangelnder Ein-

deutigkeit bei den Korrektur- und Bewertungshinweisen gab es zu wenigen Testaufgaben auch Anfragen an die Hotline zur Punktevergabe. Die Mitarbeiter des Projekts *kompetenztest.de* empfahlen dann in schwierigen/nicht eindeutigen Fällen den für die Aufgabenentwicklung verantwortlichen Ansprechpartner im Thüringer Institut für Lehrerfortbildung, Lehrplanentwicklung und Medien (ThILLM).

Zusammenfassend lässt sich festhalten, dass sich die Hotline auch in diesem Jahr als ein zentraler Bestandteil bei der Durchführung der Kompetenztests erwiesen hat. Sie bildet die Schnittstelle zwischen der Durchführung vor Ort und den organisatorischen bzw. technischen Abläufen. Einerseits dient die Hotline den teilnehmenden Lehrern als Anlaufstelle bei Fragen und Problemen, andererseits konnten durch den direkten Kontakt mit den Beteiligten Hinweise und Verbesserungsmöglichkeiten aufgenommen und umgesetzt werden.

4 Diagnostische Kompetenz

4.1 Einleitung

Auch in diesem Jahr gab es für die am Test beteiligten Fachlehrer die Möglichkeit, bereits einen Tag vor den Kompetenztests die Schülerleistungen einzuschätzen. Diese Einschätzung der Diagnostischen Kompetenz soll im folgenden Kapitel beschrieben werden.

Daten, die eine solche Darstellung ermöglichen, wurden jeweils für die Kompetenztests in Deutsch und Mathematik in den Klassenstufen 3, 6 und 8 erhoben. Wie in den letzten Jahren wurde auf eine Erhebung der Lehrereinschätzung im Fach Englisch in Klassenstufe 6 verzichtet, da die Bereitstellung der Audiodateien (für den Testteil „Hörverstehen“) am Testvortrag sehr aufwändig wäre. Den teilnehmenden Lehrern wurde jeweils über die Schulportalseite angeboten, für durchschnittlich sechs ausgewählte Testaufgaben aus den Kompetenztests eine Einschätzung darüber abzugeben, wie viele der Schüler die jeweiligen Aufgaben eines Tests richtig lösen werden. Diese Einschätzung kann der Fachlehrer mit den tatsächlichen Ergebnissen aus der Sofortrückmeldung vergleichen.

Die Erhebung der Diagnostischen Kompetenz soll für den Fachlehrer ein nützliches Instrument sein. Dadurch wird die Möglichkeit geboten, die Genauigkeit der Einschätzung der Schülerleistungen zu überprüfen. Weiterhin kann der Lehrer dadurch Rückmeldung und Anregungen erhalten, sich intensiver mit dem Leistungsbild seiner Klassen zu beschäftigen. Eine realistische Einschätzung der Schülerinnen und Schüler einer Klasse bezüglich ihrer fachspezifischen Kompetenzen sind eine notwendige Bedingung für einen förderlichen Unterricht. Fehleinschätzungen der Leistungsfähigkeit können zu Über- oder Unterforderung der Schüler führen. Misserfolgserlebnisse können einen negativen Einfluss auf die Schüler- und Lehrermotivation haben. Kein Lehrer wird jedoch perfekte Leistungsaussagen über seine Schüler treffen. Ziel der Diagnostischen Kompetenz soll es demnach sein, den Lehrern eine Möglichkeit zu bieten, die eigenen Einschätzungen überprüfen zu können und die Genauigkeit bei der Kompetenzeinschätzung der Schüler auszubauen.

Die freiwillige Nutzung der Diagnostischen Kompetenz hat sich im Vergleich zum letzten Jahr nicht wesentlich verändert. Im Durchschnitt haben 27,9 % aller Lehrer an Thüringer Schulen, über alle Schularten, alle teilnehmenden Klassen und Fächer hinweg die Diagnostische Kompetenz genutzt. Das sind lediglich 1,5 % mehr als im Schuljahr 2006/2007. Der geringe Unterschied erklärt sich vor allem durch stark gegensätzliche Tendenzen in den Klassenstufen. Während 10 % weniger Mathematiklehrer in Klassenstufe 8 die Diagnostische Kompetenz dieses Jahr genutzt haben, zeigt sich in allen anderen Klassenstufen und Tests ein zum Teil deutlicher Anstieg in der Teilnahme. In der Klassenstufe 8 wurde die Diagnostische Kompetenz dieses Jahr erst zum zweiten, in allen anderen Stufen bereits zum vierten Mal angeboten. Es ist plausibel anzunehmen, dass der Neuheitseffekt für die Mathematiklehrer in Klassenstufe 8 verschwunden ist und die Teilnehmerzahlen deswegen gesunken sind.

In den nachfolgenden Unterkapiteln wird jeweils über die Diagnostische Kompetenz in den einzelnen Tests für die verschiedenen Klassenstufen und in den weiterführenden Schulen auch getrennt nach Schularten berichtet. Dargestellt werden dabei drei Maße, welche unterschiedliche Beurteilungen zulassen. Zum einen die mittleren Abweichungen aus der eingeschätzten

relativen Häufigkeit (den Lehreraussagen) und der tatsächlichen relativen Häufigkeit der richtigen Lösungen. Diese Differenz wird in Prozent dargestellt und gibt Auskunft darüber, ob die Lehrer in einem Test ihre Schüler im Mittel überschätzt oder unterschätzt haben. Ein positiver Wert bedeutet dabei eine mittlere Überschätzung der tatsächlichen Lösungshäufigkeiten, ein negativer Wert eine Unterschätzung der realen Schülerleistungen. Zum Zweiten werden jeweils die Standardabweichungen der Differenzwerte berichtet. Diese geben Auskunft über die Breite der Verteilung und sind ein Maß für die Streuung um den Wert der mittleren Abweichung. Die Standardabweichung soll hier als Genauigkeit der Lehrereinschätzungen interpretiert werden und zeigt, wie ähnlich die verschiedenen Lehrer innerhalb eines Tests geschätzt haben. Hohe Werte in den Standardabweichungen sind ein Hinweis darauf, dass es innerhalb der Gruppe der Lehrer für einen Test große Abweichungen in der Einschätzung der Schülerleistungen gibt. Des Weiteren sollen Extremwerte berichtet werden, welche zeigen, wie viel Prozent der Lehrer bei einem Test die Schüler im Mittel über 25 %-Punkte über- oder unterschätzt haben. Diese Grenze ist willkürlich aus den Erfahrungen mit der Diagnostischen Kompetenz in den letzten 4 Jahren festgelegt wurden.

Es soll darauf hingewiesen werden, dass zu den einzelnen Tests jeweils nur die Ergebnisse der Grundschulen (für DK3 und MK3), der Regelschulen und Gymnasien (für DK6, MK6, MK8) berichtet werden. Wegen der kleinen Fallzahlen der teilnehmenden Förderschullehrer an der Diagnostischen Kompetenz kann keine Anonymität gewährleistet werden. Aus Datenschutzgründen ist eine Auswertung daher leider nicht möglich.

Im Anschluss an die Darstellung der Ergebnisse der Diagnostischen Kompetenz in den einzelnen Tests soll versucht werden, Zusammenhänge der Einschätzungsleistungen mit Merkmalen aus den Lehrerfragebögen zu beschreiben.

4.2 Auswertung der Diagnostischen Kompetenz für Klassenstufe 3

Im Fach Deutsch zeigt sich bei den Grundschullehrern in diesem Jahr ein wesentlicher Anstieg in der Teilnahme an der Diagnostischen Kompetenz. Im Vergleich zu 2007 (22,3 %) haben dieses Jahr 29,6 % aller Deutschlehrer an Thüringer Grundschulen das Instrument genutzt. Kaum Veränderungen zeigten sich bei den Mathematiklehrern der Grundschulen, von 855 nahmen hier ähnlich wie im Vorjahr 199 Fachlehrer teil (23,3 %). Angesichts der Freiwilligkeit der Nutzung ist die Verwendung der Diagnostischen Kompetenz von etwa einem Viertel der Thüringer Grundschullehrer eine erfreulich hohe Rate.

Fach Deutsch

Im Mittel unterscheiden sich die von den Lehrern vorhergesagte und die tatsächliche Anzahl von Schülern, die die ausgewählten Aufgaben vollständig gelöst haben, um -1,8 %-Punkte. Dieser durchschnittliche Abweichungswert des Deutstests in Klassenstufe 3 ist klein und unterscheidet sich nicht wesentlich von der mittleren Abweichung für die Kompetenztests in Deutsch der Klassenstufe 3 aus dem Jahr 2007 (2,1 %). Durchschnittlich wurde also die Schülerleistungen leicht unterschätzt. Betrachtet man die Standardabweichung von 14,8 % für diese mittlere Abweichung oder graphisch die Streuung in Abbildung 4.1 so wird deutlich, dass die Lehrereinschätzungen voneinander abweichen. Man würde schlussfolgern, dass sich die Deutschlehrer im Mittel bei der Beurteilung der Schülerleistungen etwas verschätzen. Bei 11 % der Deutschlehrer der Grundschulen liegen die Unterschiede zwischen geschätzter und tatsächlicher Lösungshäufigkeit über 25 %-Punkten.

Aufgabenebene: Interessanter sind jedoch die mittleren Abweichungen in den einzelnen ausgewählten Aufgaben zur Diagnostischen Kompetenz, die ein weitaus differenzierteres Bild liefern. Auffälligkeiten hinsichtlich besonders starken Abweichungen in einzelnen Aufgaben gibt

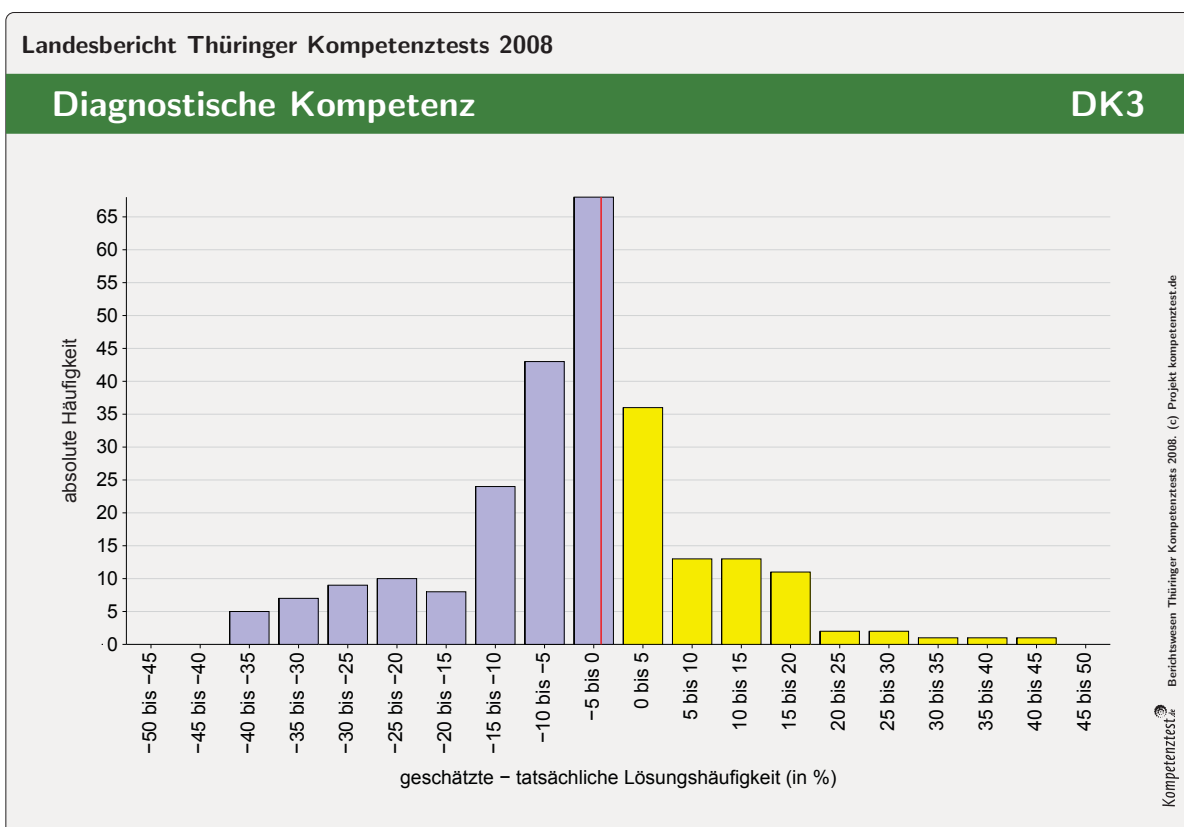


Abbildung 4.1: DK3: Abweichung der Lehrerschätzung von der tatsächlichen Lösungshäufigkeit

es beispielsweise bei Aufgabe 1-4. Ziel dieser Aufgabe war es, den Inhalt einer Geschichte mit eigenen Worten wiederzugeben. Hier zeigt sich eine mittlere Abweichung zwischen Lehrereinschätzung und Schülerleistung von -18,1 %, das bedeutet, im Mittel wurden die tatsächlichen Leistungen der Schüler eher unterschätzt. Eine noch größere mittlere Abweichung von 28,7 %-Punkten, findet sich bei Aufgabe 3-7A. Hier haben die Deutschlehrer die Leistungen ihrer Schüler bezüglich dieser Aufgabe im Mittel überschätzt. Die Schüler sollten hier, aus einer Wortreihe einer bestimmten Wortart versuchen, das falsche Wort herauszufinden. Vielen Drittklässlern ist diese Aufgabe nicht so leicht gefallen, wie ihre Lehrer das erwartet hätten.

Fach Mathematik

Ein recht ausgeglichenes Bild zeigt sich im mathematischen Bereich. Die mittlere Abweichung zwischen Lehrereinschätzung und Schülerleistung liegt hier lediglich bei 1,4 % und die Standardabweichung beträgt 13 %. Der Anteil der Lehrer, bei denen die Unterschiede zwischen geschätzter und tatsächlicher Lösungshäufigkeit größer als 25 %-Punkte sind, beträgt 6 %. Das bedeutet, ein Großteil der Grundschullehrer liegt mit ihren Einschätzungen zur Diagnostischen Kompetenz relativ nahe am Mittelwert und die Streuung um diese mittlere Abweichung ist moderat. Den Mathematiklehrern gelingt also eine etwas genauere Einschätzung. Im Vergleich zum letzten Jahr (mittlere Abweichung 2,7 %) zeigen sich auch hier wieder keine Unterschiede über die Zeit. Es gibt bei den Mathematiklehrern der Grundschule im Bereich der Diagnostischen Kompetenz also kaum wesentliche Abweichungen, die Einschätzungen der Lehrer spiegelt ziemlich gut die Schülerleistungen wider. Auf Aufgabenebene zeigen sich keinerlei bedeutende Abweichungen (also alle mittleren Abweichungen liegen unter 12 %-Punkten) für die einzelnen Aufgaben der Diagnostischen Kompetenz in diesem Jahr.

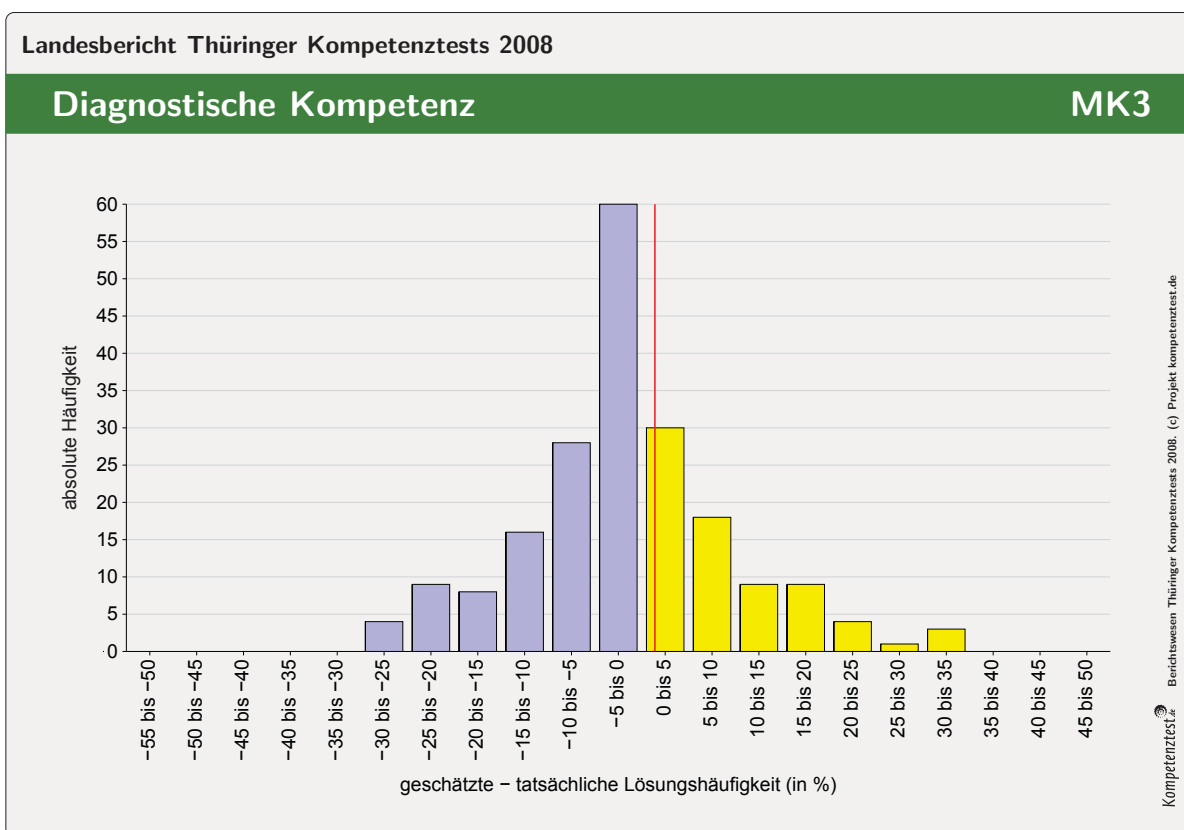


Abbildung 4.2: MK3: Abweichung der Lehrerschätzung von der tatsächlichen Lösungshäufigkeit

In den Abbildungen 4.1 und 4.2 sind die Verteilungen der Abweichungen der geschätzten von den tatsächlichen relativen Lösungshäufigkeit in Deutsch und im Bereich Mathematik wie oben beschrieben dargestellt. Die im Text erwähnten mittleren Differenzen lassen sich hier gut nachvollziehen und die Streuung (Standardabweichung von der mittleren Differenz) gibt Aufschluss darüber, wie genau geschätzt wurde. Beim Vergleich der beiden Abbildungen fällt auf, dass in der Klassenstufe 3 die Streuung in Deutsch leicht größer ist, als in Mathematik. Das bedeutet, dass Fehleinschätzungen im Fach Deutsch in etwas extremeren Bereichen liegen, die Lehrereinschätzungen also etwas stärker abweichen.

4.3 Auswertung der Diagnostischen Kompetenz für Klassenstufe 6

In Klassenstufe 6 beteiligten sich im Schuljahr 2007 / 2008 etwa 24,8 % der Deutschlehrer und 33,8 % der Mathematiklehrer der weiterführenden Schulen an der Einschätzung der Testleistungen ihrer Schüler. Im Vergleich zu den Teilnehmerquoten aus dem Vorjahr (Deutsch: 23,2 %, Mathe: 27,3 %) lässt sich hier auch ein positiver Zuwachs verzeichnen. Betrachtet man die Teilnahme nur in Abhängigkeit von der Schulform, so waren für den Deutschtest 60,1 % der teilnehmenden Lehrer Regelschul- und nur 31,8 % Gymnasiallehrer, für den Mathematiktest 59,3 % der teilnehmenden Lehrer Regelschul- und 32,2 % Gymnasiallehrer. Diese Verteilung entspricht der Basisrate in Thüringen. Über die Hälfte aller Deutsch- und Mathematiklehrer an weiterführenden Schulen sind Regelschullehrer.

Fach Deutsch

Schulebene: In den Regelschulen unterscheiden sich die von den Lehrern vorhergesagte und die tatsächliche Lösungshäufigkeit im Mittel um -7,3 %-Punkte, die Standardabweichung be-

trägt hier 15,9 %. Das bedeutet die Regelschüler werden im Mittel eher unterschätzt und die Einschätzungen sind dabei weniger genau. Im Vergleich zum letzten Jahr (13,5 %) lässt sich hier ein relativ großer Unterschied ausmachen. Die Abweichung geht in negative Richtung und die Differenz zwischen beiden Jahren ist recht hoch. Am Gymnasium unterscheidet sich die von den Lehrern vorhergesagte und die tatsächliche Lösungshäufigkeit um -14,2 %. Im Mittel unterschätzen die Deutschlehrer am Gymnasium ihre Schüler in ihren Leistungen im dies-jährigen Kompetenztest. Die Standardabweichung von 18,3 % deutet auf eine relativ große Streuung und den stark negativen Abweichungswert. Die Einschätzungen der Gymnasiallehrer sind also in diesem Jahr etwas ungenauer. Wie bei den Regelschullehrern zeigt sich auch hier beim Vergleich mit dem Vorjahr (8,6 %) eine relativ große Differenz und Richtungsänderung der Einschätzungen der Schülerleistungen. Der Anteil der Lehrer, bei denen die mittlere Differenz zwischen geschätzter und tatsächlicher Lösungshäufigkeit größer als 25 %-Punkte ist, liegt bei den Deutschlehrern der Regelschulen bei 20,2 %, bei den Gymnasien 25,5 %.

Beim Vergleich zwischen beiden Schulformen ist es auffällig, dass die Gymnasiallehrer ihre Schüler doppelt so stark unterschätzt haben, wie die Regelschullehrer. Im letzten Jahr war das Muster umgekehrt: die Regelschullehrer haben ihre Schüler wesentlich stärker überschätzt in ihren Deutschleistungen, als die Lehrer am Gymnasium. Dies sind deutliche Veränderungswerte, die Gründe dafür sind aber unklar. Bei der Interpretation solcher Veränderungen ist immer zu bedenken, dass jedes Jahr andere Aufgaben für die Diagnostische Kompetenz ausgewählt werden oder dass die Tests an sich andere Schwerpunkte haben können. Eine weitere Möglichkeit wäre, dass solche Ergebnisse nur zufällig zustande gekommen sind.

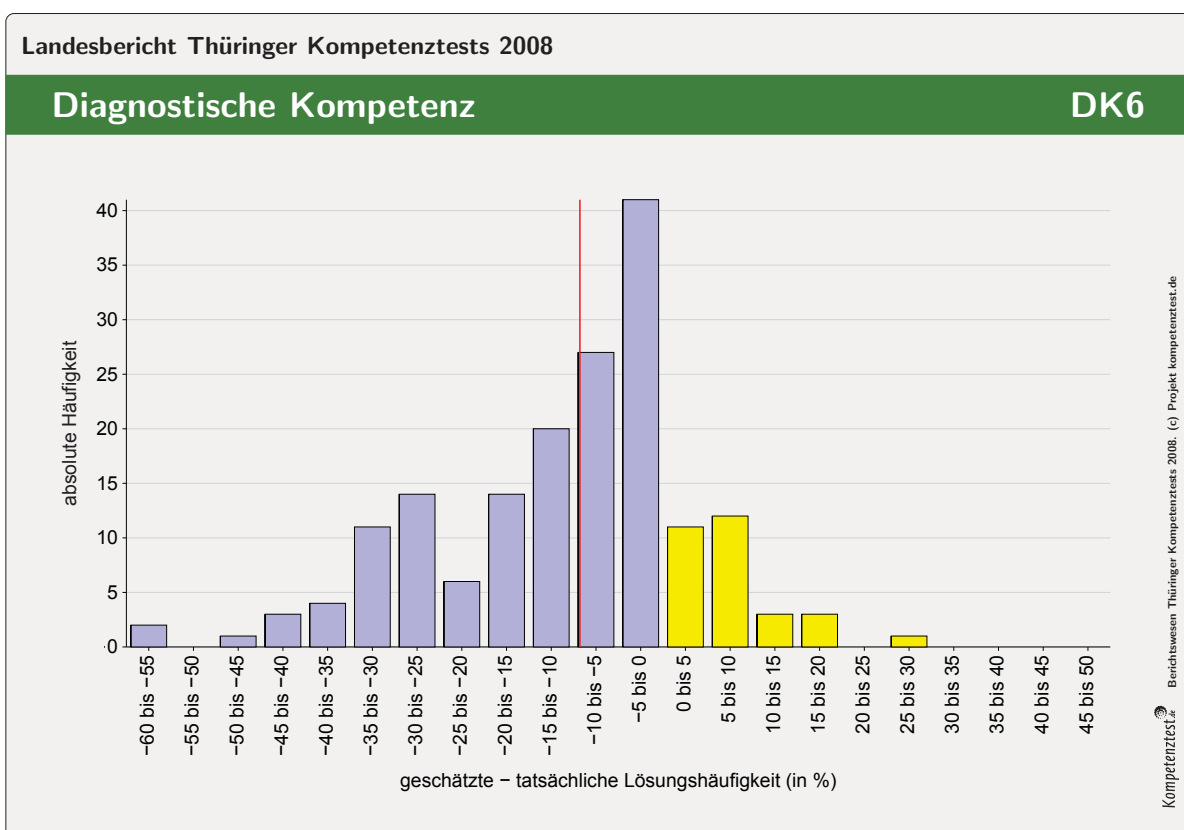


Abbildung 4.3: DK6: Abweichung der Lehrerschätzung von der tatsächlichen Lösungshäufigkeit

Aufgabenebene: Über alle Schulformen und Aufgaben hinweg ergibt sich eine mittlere Abweichungen von -9,3 %, also eine Unterschätzung der tatsächlichen Schülerleistungen. Diese stark negative Divergenz zwischen der Lehrereinschätzung und den Lösungshäufigkeiten zeigt sich bei den Aufgaben 23 (-28,8 %) und 39 (-23 %) aus dem Testheft Teil 2. Die Schüler sollten

hier aus einer Geschichte den genannten Tieren bestimmte Handlungen zuordnen (Aufgabe 23), bzw. die witzigste Lösung eines Witzes ankreuzen (Aufgabe 39). Eine starke Überschätzung der Schülerleistungen findet sich bei Aufgabe 22 aus dem Testheft Teil 2 (24,4 %). Schüler sollten bei dieser Aufgabe vorgegebene Konjunktionen in einem Lückentext ergänzen, was offensichtlich schwerer gefallen ist, als es ihre Deutschlehrer vermutet hätten. Der Abschätzung des richtigen Umgangs mit Konjunktionen könnte demnach ein weiterer Schwerpunkt im Deutschunterricht sein.

Fach Mathematik

Schulebene: In den Regelschulen unterscheiden sich die von den Lehrern vorhergesagte und die tatsächliche Lösungshäufigkeit im Mittel um 1,3 %. Im Vergleich zum Vorjahr (-0,8 %) gibt es hier keine wesentliche Veränderung in der mittleren Abweichung. Die Streuung um die mittlere Abweichung liegt hier bei 14 %-Punkten. Es scheint den Regelschullehrern im Fach Mathematik also relativ genau zu gelingen, die Schülerleistungen bei den Testaufgaben einzuschätzen. Am Gymnasium unterscheiden sich die vorhergesagten von den tatsächlichen Lösungshäufigkeiten um -3,5 %-Punkte, im Mittel wird die Leistung der Schüler also leicht unterschätzt. Betrachtet man die Standardabweichung von 9,8 %-Punkten für diese mittlere Abweichung so wird deutlich, dass die Lehrereinschätzungen hier weniger stark voneinander abweichen. Auch hier gelingt es den Mathematiklehrern im Vergleich zu den Deutschlehrern etwas präziser, die Schülerleistungen einzuschätzen. Es gibt keine bemerkenswerten Veränderungen zum Schuljahr 2006 / 2007 (-3,7 %). Die Differenz zwischen den mittleren Abweichungen der Regelschul- und Gymnasiallehrer ist bei den Mathematiktests geringer als bei den Deutschtests. Der Anteil der Lehrer, bei denen die mittlere Differenz zwischen geschätzter und tatsächlicher Lösungshäufigkeit größer als 25 %-Punkte ist, liegt bei den Mathematiklehrern der Regelschulen bei 8,6 %. Am Gymnasium gibt es in diesem Jahr keinen Mathematiklehrer, bei dem die mittlere Abweichung diesen Bereich überschreitet, was ein erfreuliches Ergebnis ist.

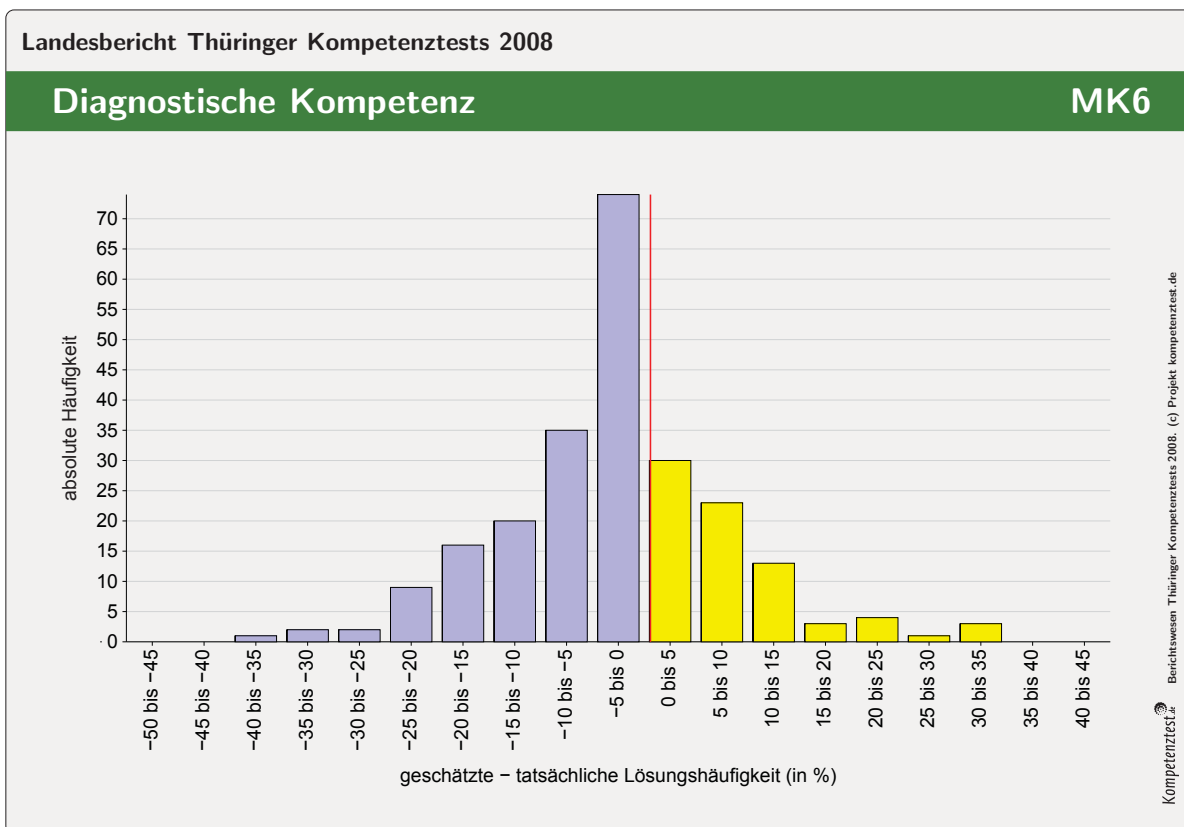


Abbildung 4.4: MK6: Abweichung der Lehrerschätzung von der tatsächlichen Lösungshäufigkeit

Aufgabenebene: Über alle Schulformen und Aufgaben hinweg ergibt sich eine mittlere Abweichung von 0,6 %, also eine ziemlich gute Leistungseinschätzung durch die Fachlehrer. Vergleicht man nur über alle Schularten hinweg die Einschätzungen der von den Mathematiklehrern vorhergesagten und der tatsächlichen Lösungshäufigkeiten auf der Aufgabenebene, so gibt es zwei Aufgaben, bei denen die mittleren Abweichungen in diesem Jahr auffällig sind. Bei Aufgabe 1 findet sich eine deutliche Überschätzung der tatsächlichen Schülerleistungen durch die Fachlehrer (23,2 %). Diese Aufgabe gehört zum Kompetenzbereich K4 („mathematische Darstellungen verwenden“) und der Leitidee L5 („Daten und Zufall“). Eine ebenfalls starke Abweichung der tatsächlichen Lösungshäufigkeiten von der Lehrereinschätzung zeigt sich bei Aufgabe 8. Hier liegt die mittlere Abweichung bei -25,9 %. Die Mathematiklehrer unterschätzten die eigentlichen Schülerleistungen also deutlich. Zu dieser Aufgabe gehören die Kompetenzbereiche 2 („Probleme mathematisch lösen“) und 5 („Umgang mit symbolischen, formalen und technischen Elementen der Mathematik“), sowie die Leitideen 1 („Zahl“) und 4 („Funktionaler Zusammenhang“). Die Zuordnung einzelner auffälliger Aufgaben zu den Kompetenzbereichen und Leitideen der Mathematik kann für den Fachlehrer eine Anregung sein, in welchen Bereichen Diskrepanzen zwischen der Sicht des Lehrers und der realen Schülerleistung bestehen, auf die zukünftig verstärkt geachtet werden sollte.

Die Abbildungen 4.3 und 4.4 stellen die Verteilung der Abweichungen der geschätzten von der tatsächlichen Lösungshäufigkeit in den Fächern Deutsch und Mathematik dar. Weiterhin lässt sich hier gut die Breite der Verteilung beobachten, was den angegebenen Standardabweichungen entspricht. Je weniger Lehrer sich in den Extrembereichen (über 25 %-Punkte) verschätzen, desto geringer wird die Standardabweichung und desto genauer sind die Einschätzungen der Lehrer im Mittel.

4.4 Auswertung der Diagnostischen Kompetenz für Klassenstufe 8

Schulebene: Die Diagnostische Kompetenz wurde im Schuljahr 2007/2008 erst zum zweiten Mal für die Klassenstufe 8 erhoben, weil der Kompetenztest Mathematik erst seit vorletztem Schuljahr durchgeführt wird. Die Beteiligung hat im Vergleich zur ersten Erhebung im Vorjahr (2007: 37,2 %) deutlich abgenommen (2008: 28,2 %) und erreicht das Level der Teilnahmequoten aus Klassenstufe 6. Von den 205 teilnehmenden Lehrern unterrichten 60 % an Regelschulen und 34,1 % an Gymnasien. Wie im Unterkapitel für die Klassenstufe 6 schon erwähnt wurde, entspricht dieses Verhältnis der Basisrate an Regelschul- und Gymnasiallehrern im Land Thüringen. Die mittlere Abweichung der eingeschätzten und tatsächlichen Ergebnisse liegt bei 7,5 %. Es findet sich also eine leichte Überschätzung der Schülerleistungen.

In den Regelschulen unterscheiden sich die von den Lehrern vorhergesagte und die tatsächliche Lösungshäufigkeit im Mittel um 6,6 %-Punkte. Die Standardabweichung beträgt hier 11,4 %-Punkte. Die mittlere Abweichung am Gymnasium beträgt 9,8 %-Punkte, die Standardabweichung liegt bei 10,5 %-Punkten. Im Mittel wird die Leistung der Schüler also etwas besser eingeschätzt, als sie tatsächlich war und diese Einschätzung der Gymnasiallehrer streut dabei moderat. Die Tendenz zur leichten Überschätzung der Schülerleistungen bei den Kompetenztests Mathematik der Klassenstufe 8 lässt sich in Abbildung 4.5 gut nachvollziehen. Weiterhin lässt sich hier die moderate Standardabweichung gut ablesen. Den Lehrern gelingt es also relativ gut, die tatsächlichen Lösungshäufigkeiten vorherzusagen.

Die Differenz zwischen der Einschätzung der Regelschul- und Gymnasiallehrer ist relativ gering. Im Vergleich zum Vorjahr findet man an beiden Schularten dieses Jahr deutliche Überschätzungen der Schülerleistungen. Im letzten Schuljahr wurden Regelschüler zu -10,6 %, Gymnasiasten mit -12,3 % von ihren Mathematiklehrern unterschätzt. Der Anteil der Lehrer, bei

denen die mittlere Differenz zwischen geschätzter und tatsächlicher Lösungshäufigkeit größer als 25 %-Punkte ist, liegt bei den Mathematiklehrern der Regelschulen bei 8,9 %. Am Gymnasium überschätzten sich 5,7 % der Lehrer in diesem Extrembereich.

Aufgabenebene: Betrachtet man die Einschätzungen zu den einzelnen Aufgaben für die Diagnostische Kompetenz über alle Schularten hinweg, so streuen diese im Bereich einer relativ starken Überschätzung. Dies wird vor allem deutlich bei den Aufgaben 9 und 20.1. Bei Aufgabe 9 schätzen 22,1 % der Mathematiklehrer die Schülerleistung besser ein, als dies in der tatsächlichen Lösungshäufigkeit der Fall ist. Diese Aufgabe „Das unmögliche Dreieck“ verlangte von den Schülern eine Begründung, warum die Seitenlängen und Innenwinkel des gezeigten Dreiecks miteinander unvereinbar sind. Eine noch deutlichere Überschätzung der Schülerleistungen findet sich bei Aufgabe 20.1 „Preisänderungen im Mobilfunk“, hier schätzen 30 % der Mathematiklehrer ein, dass die Schüler diese Aufgabe lösen werden.

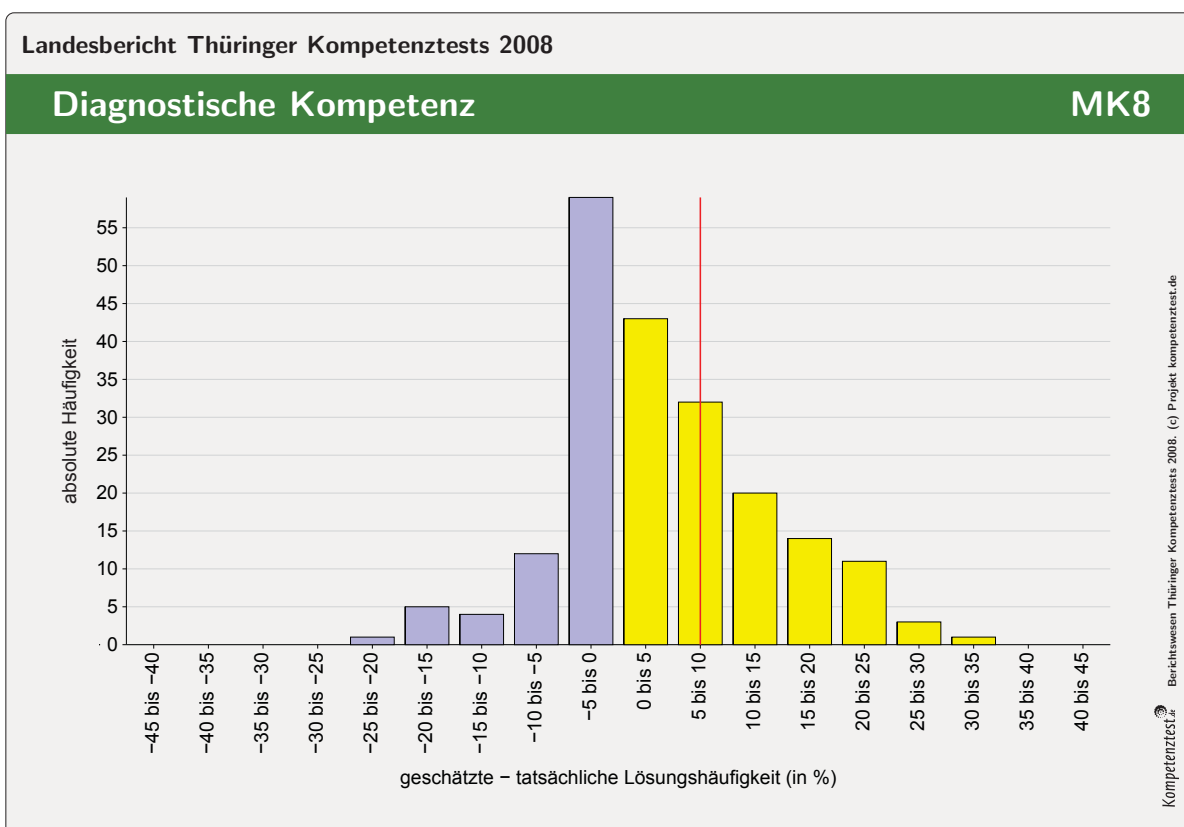


Abbildung 4.5: MK8: Abweichung der Lehrerschätzung von der tatsächlichen Lösungshäufigkeit

4.5 Zusammenhang der Ergebnisse der Diagnostischen Kompetenz mit Lehrervariablen

Im folgenden Kapitel sollen Personenmerkmale der Lehrer, welche in den Lehrerfragebögen erfasst wurden und im Zusammenhang mit den Ergebnissen der Diagnostischen Kompetenz stehen, berichtet werden. Dies soll zeigen, wodurch sich die Gruppe teilnehmenden Lehrer in diesem Schuljahr auszeichnet.

Als Variablen wurden dabei das Geschlecht der Lehrer und das Dienstalter in Dienstjahren näher betrachtet. Die Gruppe der teilnehmenden Lehrer wurde jeweils auf bedeutsame Unterschiede zu der Gesamtgruppe aller Lehrer für einen Test verglichen.¹

Beispielsweise wurde analysiert, ob überproportional weibliche Mathematiklehrer in Klassenstufe 6 die Diagnostische Kompetenz als Instrument zur Einschätzung der Schülerleistung nutzen. Dieser Anteil wurde dann mit dem Gesamtanteil an Mathematiklehrerinnen für die Klassenstufe 6 an weiterführenden Schulen verglichen.

Bei allen Analysen (Tests auf signifikante Unterschiede zwischen den teilnehmenden Lehrern verglichen mit der jeweiligen Gesamtpopulation der Lehrer) für die Variablen Geschlecht und Dienstalter über die Fächer und Klassenstufen fand sich nur ein statistisch bedeutsamer Unterschied. Lediglich für den Deutschtest in Klassenstufe 3 wurde gefunden, dass die an der Diagnostischen Kompetenz teilnehmenden Lehrer systematisch im Durchschnitt jünger (im Sinne des Dienstalters) sind, als die Vergleichsgruppe der Deutschlehrer in Klassenstufe 3. Alle weiteren Unterschiede wurden nicht signifikant.

In allen Tests fand sich der Trend, dass die Teilnehmer eher die jüngeren Lehrer sind. Die Stärke dieser Unterschiede ist jedoch als gering einzuschätzen und lediglich in DK3 wird diese Differenz statistisch signifikant. Dadurch ist die Generalisierbarkeit auf die Gesamtpopulation der Lehrer eingeschränkt. Für das Lehrermerkmal Geschlecht scheint es keine Unterschiede zu geben. Die Frage der Generalisierbarkeit der Ergebnisse der Diagnostischen Kompetenz auf die Gesamtgruppe aller Lehrer kann an dieser Stelle nicht beantwortet werden. Angesichts der Ergebnisse hinsichtlich des Dienstalters gibt es Unterschiede zwischen den teilnehmenden Lehrern und der Gesamtlehrerschaft. Das bedeutet jedoch nicht, dass die inhaltlichen Ergebnisse der Diagnostischen Kompetenz sich von der Gesamtpopulation unterscheiden müssen.

4.6 Fazit

Im letzten Abschnitt sollen die Ergebnisse der Diagnostischen Kompetenz für das Schuljahr 2007 / 2008 zusammenfassend betrachtet und einige auffällige Trends berichtet werden.

Teilnehmerzahlen:

Wie bereits in der Einleitung erwähnt gibt es im Vergleich zum Vorjahr im Durchschnitt kaum Veränderungen der Teilnehmerzahlen über alle Tests hinweg. Deutlicher Zuwachs lässt sich jedoch für Deutsch Klassenstufe 3 (plus 7,3 %-Punkte) und Mathematik Klassestufe 6 (plus 6,5 %-Punkte) bezeichnen, wohingegen die hohen Teilnehmerquoten für Mathematik in Klassenstufe 8 aus dem letzten Jahr deutlich zurück gegangen sind (minus 9 %-Punkte). Es bleibt zu hoffen, dass die Nutzung der Diagnostischen Kompetenz als ein praktisches Instrument für den Fachlehrer, weiterhin an Beliebtheit gewinnt. Dennoch ist der leichte Zugewinn an teilnehmenden Lehrern ein erfreuliches Ergebnis.

¹ Das hierfür verwendete statistische Verfahren ist der Signifikanztests. Das Signifikanzniveau lag bei $\alpha = 0.05$. Nähere Informationen zu dieser Methode finden Sie z. B. bei Nachtigall & Wirtz (2004).

Fächervergleiche:

Die mittleren Abweichungen der Lehrereinschätzungen von den tatsächlichen Lösungshäufigkeiten für die einzelnen Fächer sind beim Vergleich mit den letzten Jahren weniger stabil. Innerhalb einzelner Fächer gab es teilweise im Vergleich mit den Vorjahresergebnissen deutliche Veränderungen. Die Ergebnisse weisen aber wieder eine deutliche Heterogenität auf, die impliziert, dass eine testspezifische Interpretation notwendig ist. Es lassen sich verschiedene Vergleiche anstellen, deren Aussagen jedoch nicht unbedingt auf die Heterogenität der Fächerkultur, sondern eventuell auf weniger kontrollierbare Umstände zurückzuführen ist. Betrachtet man beispielsweise einen Vergleich zwischen Mathematik (mittlere Abweichung in Klassenstufe 3, 6 und 8) und Deutsch (mittlere Abweichung in Klassenstufe 3 und 6), so zeigt sich für Deutsch eine relative Unterschätzung der Schülerleistungen (-5,5 %-Punkte), wohingegen in den diesjährigen Mathematiktests die tatsächlichen Lösungshäufigkeiten geringfügig überschätzt wurde (3,2 %-Punkte). Hinsichtlich der Standardabweichungen zeigt sich ein leichter Fächertrend: diese sind für die Deutschttests (15,8 %) etwas höher als für die diesjährigen Kompetenztests im Fach Mathematik (12,3 %). Diese Differenz in der Streuung der Schätzungen um die mittlere Abweichung impliziert, dass die Lehrereinschätzungen im Fach Mathematik etwas präziser sind, als dies im Fach Deutsch der Fall ist. Im letzten Jahr der Diagnostischen Kompetenz lag beim Vergleich der Fächer der Schluss nahe, dass die Einschätzungen der Schülerleistungen im Fach Deutsch etwas schwieriger sind, es deswegen im Schnitt größere mittlere Abweichungen gab. Außerdem zeigt sich, dass bei den zwei Deutschttests der Anteil der Lehrer, bei denen die mittlere Differenz zwischen geschätzter und tatsächlicher relativer Lösungshäufigkeit größer als 25 %-Punkte ist, insgesamt höher ist, als bei den drei Mathematiktests. Es scheint also, dass es im Fach Mathematik genauer gelingt, die tatsächlichen Leistungen der Schüler zu beurteilen. Dies kann durchaus seine Ursache in der relativ klaren Struktur von Mathematikaufgaben haben, die in der Regel eindeutige Lösungen implizieren.

Schulvergleiche:

Beim Vergleich der mittleren Abweichungen zwischen den einzelnen Klassenstufen lässt sich als Trend feststellen, dass diese in den Grundschulen geringer sind als in den weiterführenden Schulen (Mittlere Abweichungen für die einzelnen Tests: DK3: -1,8 %, MK3: 1,4 %, DK6: -9,3 %, MK6: 0,6 %; MK8: 7,5 %). Für die Standardabweichungen zeigt sich kein bedeutender Trend, der auf schulspezifische Besonderheiten hinweist. Gründe dafür könnten an den höheren Aufgabeanforderungen an die Fachlehrer der weiterführenden Schulen liegen, oder auch daran, dass Grundschullehrer für Mathematik oder Deutsch eine Klasse vermutlich in mehr als nur einem Fach unterrichten und so ein breiteres Leistungsspektrum der Schüler abschätzen könnten. Eindeutige Ursachen lassen sich bei solchen Ergebnissen nur schwer festmachen und die Interpretation der Diagnostischen Kompetenz sollte auch nur testspezifisch erfolgen, da die Aufgaben für die einzelnen Tests jährlich verschiedene Schwerpunkte abdecken und zum Teil nicht direkt vergleichbar sind.

Im Hinblick auf die Weiterentwicklung der Diagnostischen Kompetenz seitens der Lehrkräfte sind die Ergebnisse auf Aufgabenebene entscheidend. Wie gezeigt wurde, finden sich zwischen den vielfältigen Aufgaben eines Tests oft große Unterschiede in der Genauigkeit der Leistungsabschätzungen. Dies kann beispielsweise an nicht eindeutigen Korrekturhinweisen, aber auch an praktisch bedeutsamen Fehleinschätzungen des Wissens- und Leistungsstandes der Schüler durch die Fachlehrer liegen. Im letzteren Fall kann die Einschätzung der Diagnostischen Kompetenz wertvolle Hinweise für die Weiterentwicklung des eigenen Unterrichts liefern. Die Hauptziele sollten dennoch sein, einerseits noch mehr Lehrer zu motivieren, die praktische Bedeutsamkeit der Diagnostischen Kompetenz als wertvolles Instrument für den eigenen Unterricht anzuerkennen und andererseits eine höhere Genauigkeit / Sicherheit im Hinblick auf die zu erwartenden Schülerleistungen zu erlangen.

5 Rückmeldung der Testergebnisse

Die Rückmeldung der Testergebnisse wurde über die Jahre hinweg im Hinblick auf Verständlichkeit und Nützlichkeit optimiert. Die Ergebnisberichte liegen mittlerweile in ausgereiften und von den Lehrkräften wertgeschätzten Formaten vor. Sie liefern Informationen auf verschiedenen Ebenen. Im Hinblick auf den einzelnen Schüler zeigen die Testergebnisse die individuellen Lernstände und liefern Hinweise auf Förderbedarf. Die zusammengefassten Ergebnisse auf Klassen- und Schulebene liefern wertvolle Informationen für die Unterrichts- und Schulentwicklung. Diese Berichte enthalten zusätzliche Vergleichsdaten und zeigen, wo die eigene Klasse/Schule im Vergleich zu Klassen/Schulen mit ähnlicher Schülerzusammensetzung „steht“. Als dritte Ebene erhält die Bildungsadministration, d. h. die Schulämter und das Kultusministerium eigene Berichte, in denen die Ergebnisse auf entsprechendem Aggregationsniveau zusammengefasst und analysiert werden.

Entsprechend diesen drei Ebenen erfolgt die Rückmeldungen zu den Kompetenztests in drei Wellen. In einer ersten Welle können die teilnehmenden Lehrer unmittelbar nach Ende der Dateneingabe die Ergebnisse der eigenen Klasse in einem Sofortbericht auf dem geschützten Schulportal abrufen. Der Sofortbericht zeigt die Klassenergebnisse in übersichtlicher Form. Er enthält die Lösungshäufigkeiten zu den einzelnen Aufgaben sowie entsprechende Angaben für die einzelnen Kompetenz- und/oder Anforderungsbereiche. Außerdem wird die Verteilung der Schülerleistungen in der Klasse dargestellt (vgl. Abb. 5.1).

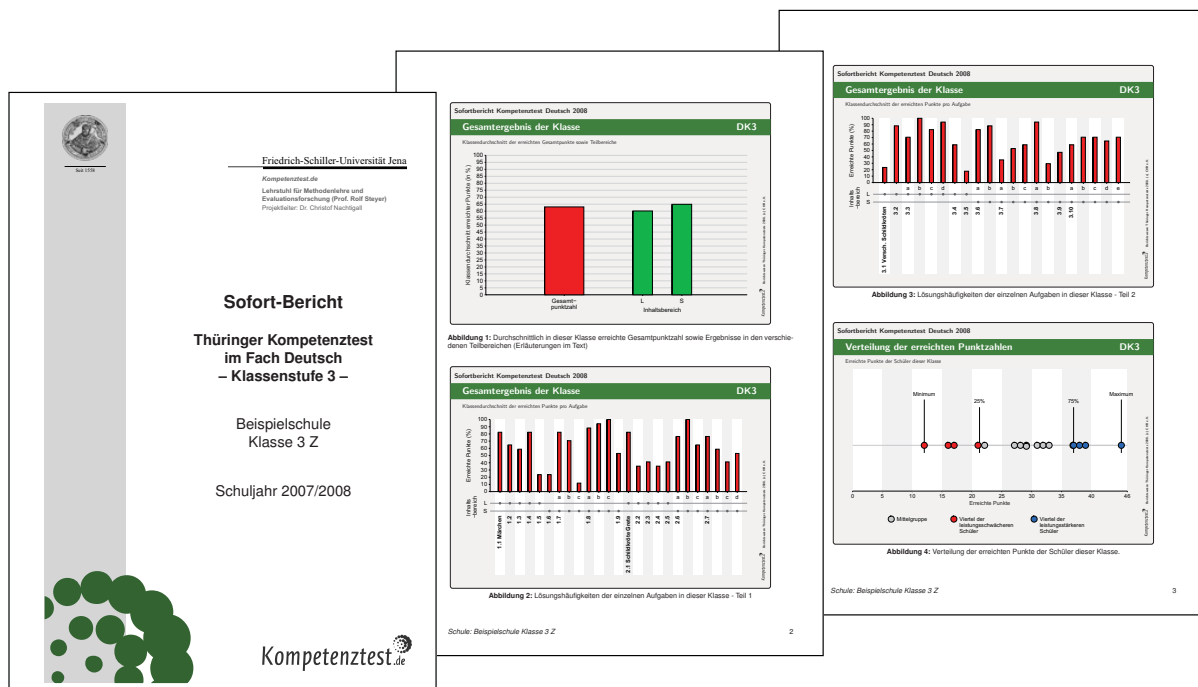


Abbildung 5.1: Beispielseiten eines Sofortberichts im Fach Deutsch, Klassenstufe 3

Die zweite Rückmeldewelle erfolgt ca. sechs Wochen nach Ende der Dateneingabe. Zu diesem Zeitpunkt stehen die landesweiten Vergleichsdaten zur Verfügung. Die Fachlehrer erhalten auf dem geschützten Schulportal klassenbezogene Ergebnis- und Ergänzungsberichte zu

den einzelnen Tests. Abb. 5.2 sowie 5.3 und 5.4 zeigen Ausschnitte aus diesen Berichten. Die Rückmeldungen enthalten die Klassenergebnisse zuzüglich landesweiter Vergleichswerte (korrigierte Landesmittelwerte). Diese Vergleichswerte werden berechnet, um möglichst faire Vergleiche zu ermöglichen und den Effekt des eigenen Unterrichts deutlicher zu machen. Die Idee hinter diesen Vergleichswerten und die mit ihnen verbundenen Analysemöglichkeiten werden in Abschnitt 5.1 beschrieben.

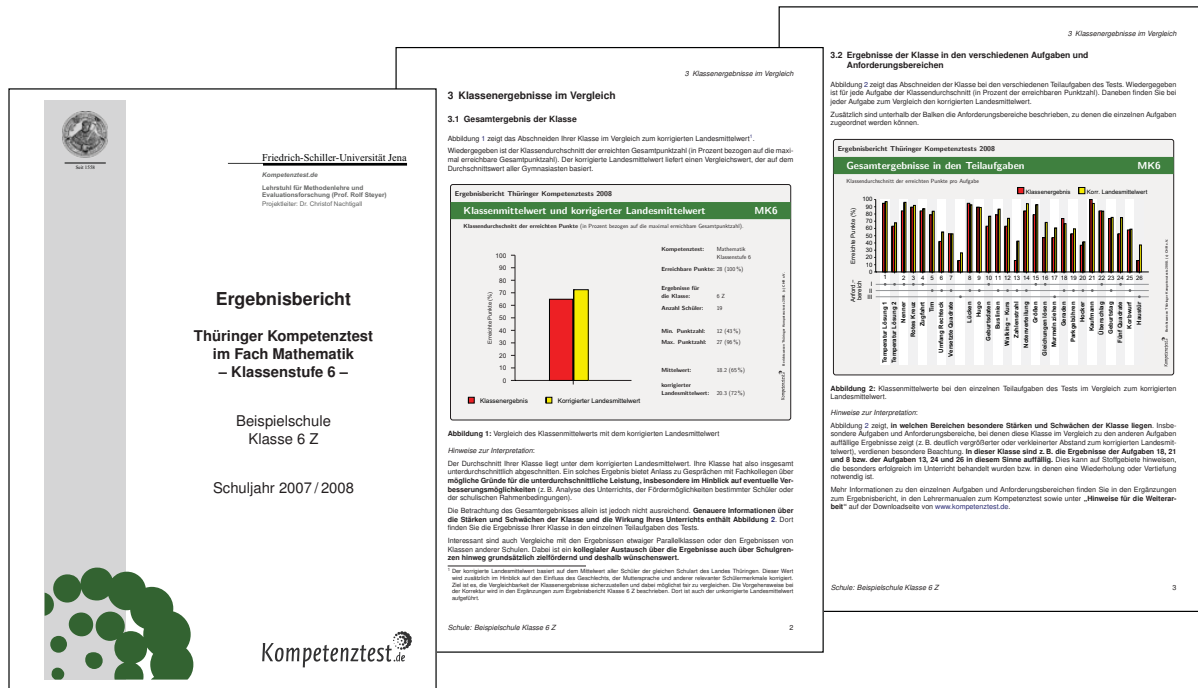
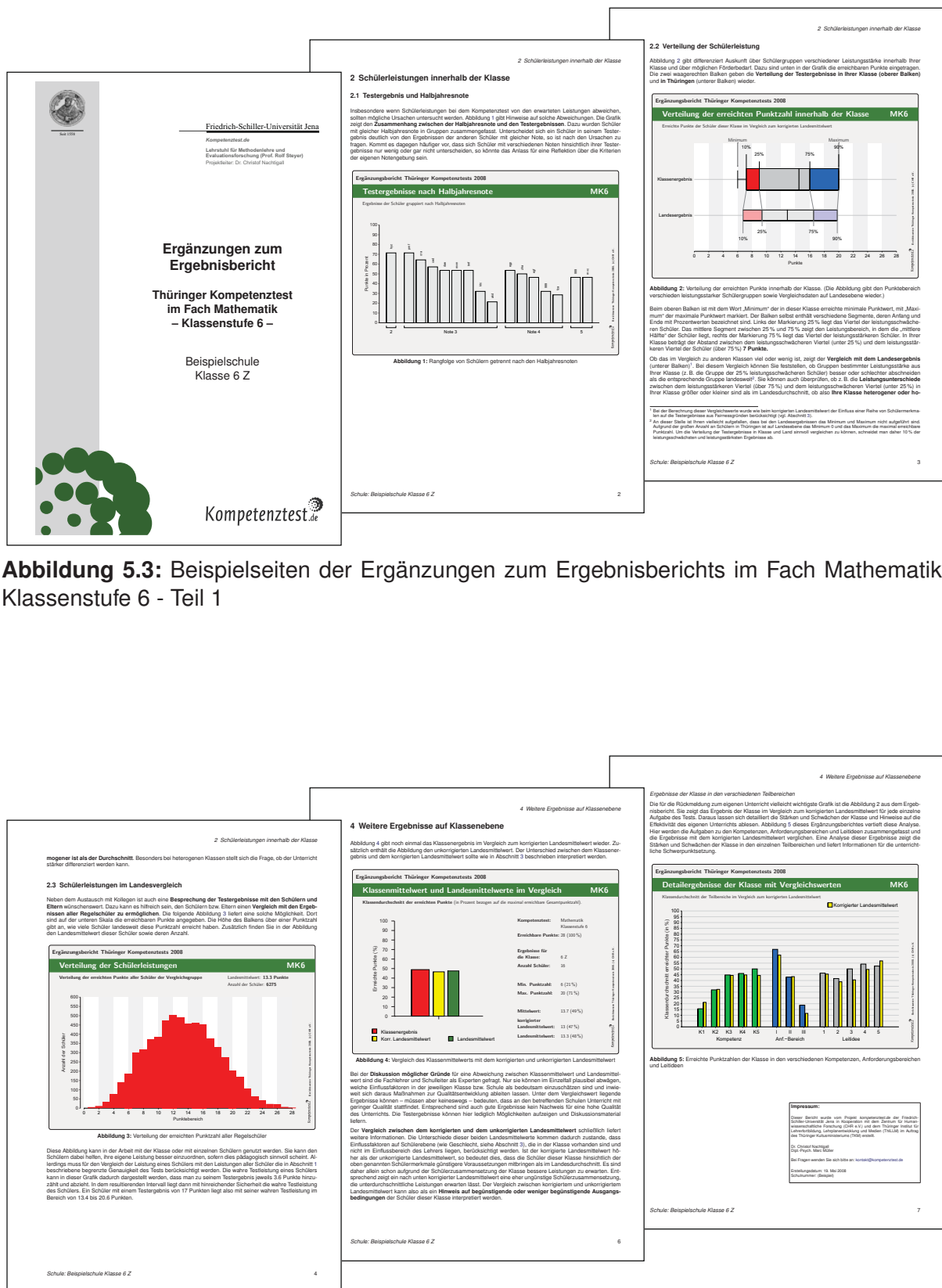


Abbildung 5.2: Beispielseiten eines Ergebnisberichts im Fach Mathematik, Klassenstufe 6

Speziell für die Schulleitungen werden seit dem Jahr 2006 Schulberichte erstellt und auf dem Schulportal bereitgestellt. Die Schulberichte enthalten die Testergebnisse der gesamten Schule. Auch diese Ergebnisse werden zusammen mit Vergleichsdaten wiedergegeben. Diese ermöglichen eine Einordnung der schulischen Leistung im Vergleich zu Schulen mit ähnlicher Schülerschaft. Zusätzlich enthalten die Schulberichte auch die Testergebnisse der früheren Kompetenztests an dieser Schule, so dass Entwicklungen ablesbar werden. Abbildung 5.5 zeigt Ausschnitte aus einem solchen Schulbericht.

Die dritte Rückmeldewelle umfasst Berichte für jedes Staatliche Schulamt sowie diesen Landesbericht. Sie erscheinen jeweils im November und enthalten die Testergebnisse in stark aggregierter Form. Ihre Funktion ist die Unterstützung bei Prozessen des Bildungsmonitoring. Die Schulamtsberichte wurden 2007 in Absprache mit Vertretern der Schulämter völlig neu konzipiert. Zentral ist jetzt eine Ergebnisdarstellung, die sich von Format und Inhalten an den Klassen- und Schulberichten orientiert. Dies soll insbesondere für die Fachberater zur Unterstützung dienen. Der Landesbericht liefert einen Überblick über die landesweiten Testergebnisse sowie eine Vielzahl von vertiefenden Analysen möglicher Einflussvariablen und Hintergrundbedingungen.



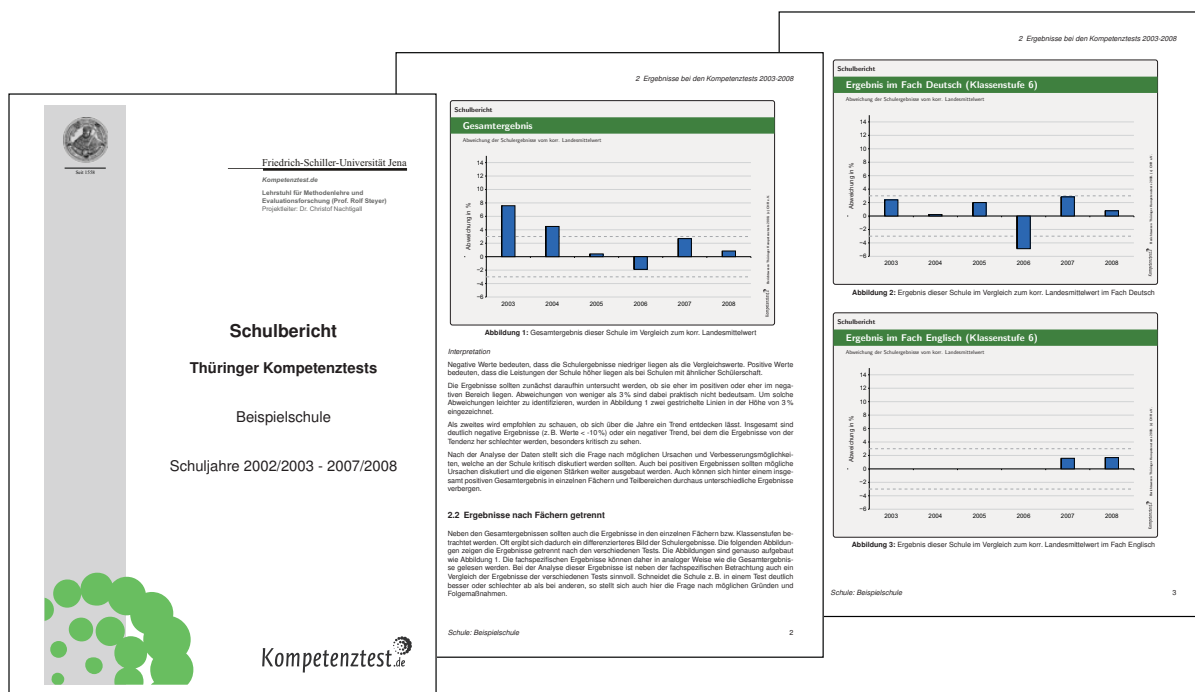


Abbildung 5.5: Beispielseiten eines Schulberichts

5.1 Korrigierter Landesmittelwert und faire Vergleiche

Neben den Testergebnissen der eigenen Klasse bzw. Schule werden bei den Ergebnismeldungen der zweiten und dritten Welle auch Vergleichswerte zurück gemeldet, die möglichst faire Vergleiche ermöglichen. Warum sind faire Vergleiche wichtig? Grundsätzlich kann es sehr irreführend sein, von Ergebnissen von Klassen bzw. Schulen direkt auf die Qualität des Unterrichts oder schulische Rahmenbedingungen zu schließen. In die Testergebnisse gehen nämlich immer eine Reihe von verschiedenen Faktoren ein:

- **Merkmale des Elternhauses der Schüler:** sozialer Status, Bildungsnähe, kulturelles Niveau, Berufsstand und Ausbildung der Eltern,
- **Voraussetzungen der Schüler:** Lerngeschichte und Vorwissen, Intelligenz, Lernmotivation und Tagesform,
- **Merkmale des Unterrichts und des Lehrers:** Motivation, didaktische Fähigkeit, Unterrichtsstil
- **Merkmale der Schule:** Schulart sowie äußere und innere schulische Bedingungen.

Leistungen der Schüler sind immer als ein Ergebnis der Interaktion dieser vier Einflussquellen zu sehen. Sie können nicht allein auf die Schule und den unterrichtenden Fachlehrer zurückgeführt werden. Daher sind Vergleiche der Testergebnisse im Sinne eines Schul- oder Lehrerrankings nicht zulässig. Vielmehr ist bei der Interpretation der Ergebnisse zu diskutieren, inwieweit sie auf den Unterricht oder andere, durch die Schule nicht beeinflussbare Gründe zurück zu führen sind.

Einige solcher Merkmale, welche die Voraussetzungen der Schüler bzw. das Elternhaus betreffen, konnten bei den Kompetenztests mit erhoben und bei der Berechnung der korrigierten Landesmittelwerte berücksichtigt werden. Es handelt sich um die Merkmale:

- Geschlecht
- Muttersprache
- Wiederholer einer Klassenstufe
- Diagnose besonderer Lernschwierigkeiten bzw. Sonderpädagogischer Förderbedarf
- Sozioökonomischer Status / Bildungsnähe¹

Auf welche Weise werden diese Merkmale bei den Vergleichen berücksichtigt? Das Verfahren soll an dieser Stelle kurz erläutert werden: Zunächst wurden für die Vergleichswerte nur Schüler der gleichen Schulart berücksichtigt, da es nicht sinnvoll ist, z. B. die Ergebnisse von Regelschülern mit Mittelwerten zu vergleichen, die auch von Gymnasiasten stammen. Zusätzlich wird berücksichtigt, in welcher Art und Weise die oben genannten Merkmale in der Klasse auftreten. Überwiegt beispielsweise in einer Klasse der Jungenanteil deutlich, so sind z. B. beim Deutschtest für diese Klasse allein aus diesem Grund schlechtere Ergebnisse zu erwarten. Diese „Verzerrung“ wird beim korrigierten Landesmittelwert berücksichtigt. Es wird der Mittelwert der Testergebnisse aller Jungen und aller Mädchen der gleichen Schulart berechnet. Diese Mittelwerte werden dann gemäß dem Anteil der Jungen und Mädchen in der betrachteten Klasse zum korrigierten Landesmittelwert zusammengefasst. Der korrigierte Landesmittelwert entspricht also dem Wert, der bei Schülern der gleichen Schulart und der vorliegenden Geschlechterverteilung im Mittel zu erwarten ist. In analoger Weise werden die anderen oben genannten Schülermerkmale berücksichtigt.

Hinweise zur Interpretation

Der korrigierte Landesmittelwert verdeutlicht einen Erwartungswert. Er entspricht dem Wert, der für Klassen der gleichen Schulart und der gleichen Verteilung von Geschlecht, Muttersprache etc. im Mittel zu erwarten ist. Das bedeutet für die Suche nach den Ursachen von positiven oder negativen Abweichungen: Die Abweichungen können nicht auf die berücksichtigten Schülermerkmale zurückgeführt werden, sondern es müssen andere Gründe dafür verantwortlich sein.

Dies bedeutet im Einzelnen: Die Abweichung des Mittelwerts Ihrer Klasse vom korrigierten Landesmittelwert wäre z. B. weder in der Anzahl der Wiederholer in der Klasse noch in den anderen berücksichtigten Schülermerkmalen begründet. Daher ist im Rahmen schulischer Qualitätsentwicklung zu überlegen, welche anderen potentiellen Einflussfaktoren relevant sein können. Solche Einflussfaktoren gibt es auf Klassenebene (z. B. Unterrichtsqualität, Unterrichtsausfall), auf Schulebene (z. B. Schulklima, Ausstattung der Schule) oder auf Schülerebene in Form von Schülermerkmalen, die nicht berücksichtigt werden konnten (z. B. kognitive Lernvoraussetzungen der Schüler). Kurz gesagt, es kann am Unterricht, aber auch an anderen Gründen liegen. Welche Gründe letztlich ausschlaggebend sind, kann nur mit dem Wissen über die Bedingungen vor Ort beurteilt werden. In diesem Sinne ist von einer „Deutungshoheit“ der Schulen zu sprechen. Durch die Korrektur bei den Vergleichswerten können allerdings bestimmte Gründe, die vom Lehrer nicht beeinflussbar sind, ausgeschlossen werden.

¹ Diese wichtigen Hintergrundmerkmale werden durch die sogenannte „Bücherfrage“ erfasst (vgl. Abschnitt 7.2.3).

6 Testergebnisse auf Landesebene

Dieser Abschnitt enthält die Ergebnisse der insgesamt sieben verschiedenen Kompetenztests 2008 auf Landesebene. In diesem Jahr fanden die Tests im Bereich Deutsch und Mathematik in der Klassenstufe 3, Deutsch, Mathematik, Englisch-Hörverstehen und Englisch-Leseverstehen in der Klassenstufe 6 und der Kompetenztest Mathematik in der Klassenstufe 8 statt. Die Ergebnisse werden zunächst hinsichtlich der Mittelwerte und Häufigkeitsverteilung des Gesamtergebnisses aller Schüler dargestellt, anschließend erfolgt die Auswertung getrennt nach Schularten. In Abschnitt 6.3 finden sich die Landesergebnisse auf Aufgabenebene. Weitere Analysen, z. B. zu einzelnen Kompetenzniveaus, Inhalts- oder Anforderungsbereichen, können dem Anhang zum Landesbericht 2008 entnommen werden.

6.1 Erreichte Gesamtpunktzahlen

In Tabelle 6.1 sind die Mittelwerte und Streuungen der erreichten Punktzahlen sowie die Teilnehmerzahlen der Kompetenztests 2008 wiedergegeben. Ein direkter Vergleich mit den Ergebnissen der Kompetenztests früherer Jahre ist derzeit noch nicht möglich, da dieses Jahr vollständig neue Aufgaben verwendet wurden, die nicht gemeinsam mit den Aufgaben früherer Jahre skaliert wurden.

Tabelle 6.1: Ergebnisse der Kompetenztests für die Klassenstufen 3 (Deutsch DK3, Mathematik – MK3), sowie 6 (Deutsch – DK6, Englisch-Hörverstehen – EK6H, Englisch-Leseverstehen – EK6L, Mathematik – MK6) und 8 (Mathematik – MK8). Angegeben sind die Mittelwerte und Streuungen der erreichten Gesamtpunktzahlen aller Thüringer Schüler

Test	DK3	MK3	DK6	MK6	EK6H	EK6L	MK8
Mittelwert	29,73	25,54	62,19	15,98	30,96	33,97	34,46
(Streuung)	(8,89)	(8,43)	(15,90)	(5,47)	(7,43)	(9,69)	(8,76)
Anzahl der Schüler	15 122	15 048	12 500	12 429	12 279	12 296	11 859

Die folgenden Grafiken geben die erreichte Punktzahl aller Thüringer Schüler für die verschiedenen Tests in Form einer Häufigkeitsverteilung wieder. Die Werte auf der waagerechten Achse geben die erreichte Punktzahl an. Die Höhe der Säulen entspricht der Anzahl von Schülern, welche diese Punktzahl im Test erreicht haben. Der senkrechte Strich markiert den Landesmittelwert. Die Grafiken zeigen, dass sich die Schülerleistungen in Thüringen über die gesamte Spannweite der erreichbaren Punkte verteilen. Alle Tests bieten auch leichte Aufgaben (siehe Abschnitt 6.3), so dass auch schwächere Schüler zu Punkten kommen.

Auch für die leistungsstarken Schüler bieten die Tests genügend Differenzierungsmöglichkeiten. Die Ausnahme stellen diesbezüglich die Englischtests Hörverstehen und Leseverstehen der Klassenstufe 6 dar: In diesen Tests wurde häufiger eine annähernd volle Punktzahl erreicht. Dies ist in der Konstruktion der Tests begründet, die in Abschnitt 2.1 auf Seite 6 nä-

her beschrieben sind. Dadurch werden systematisch bessere Ergebnisse erreicht, was jedoch nicht mit höherer Unterrichtsqualität gleichgesetzt werden kann - diese jedoch auch nicht ausschließt.

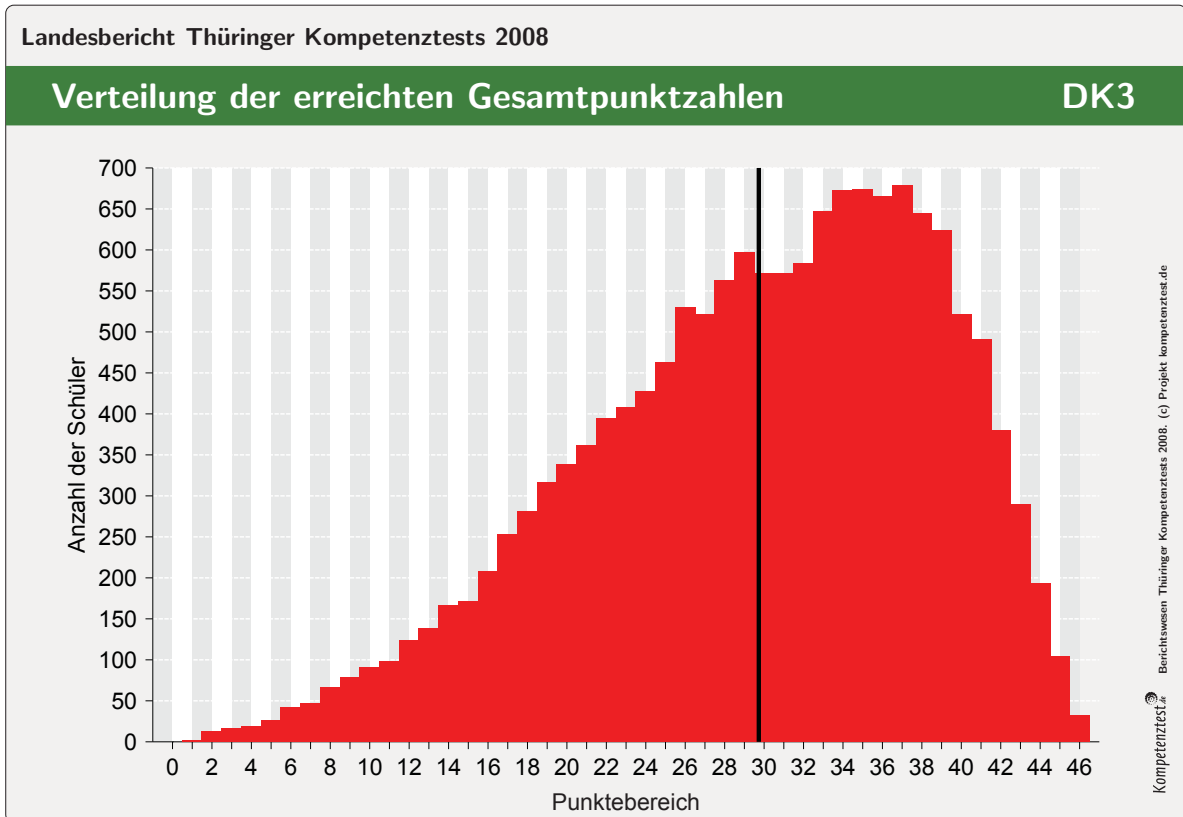


Abbildung 6.1: Verteilung der Gesamtpunktzahlen im Fach Deutsch, Klassenstufe 3

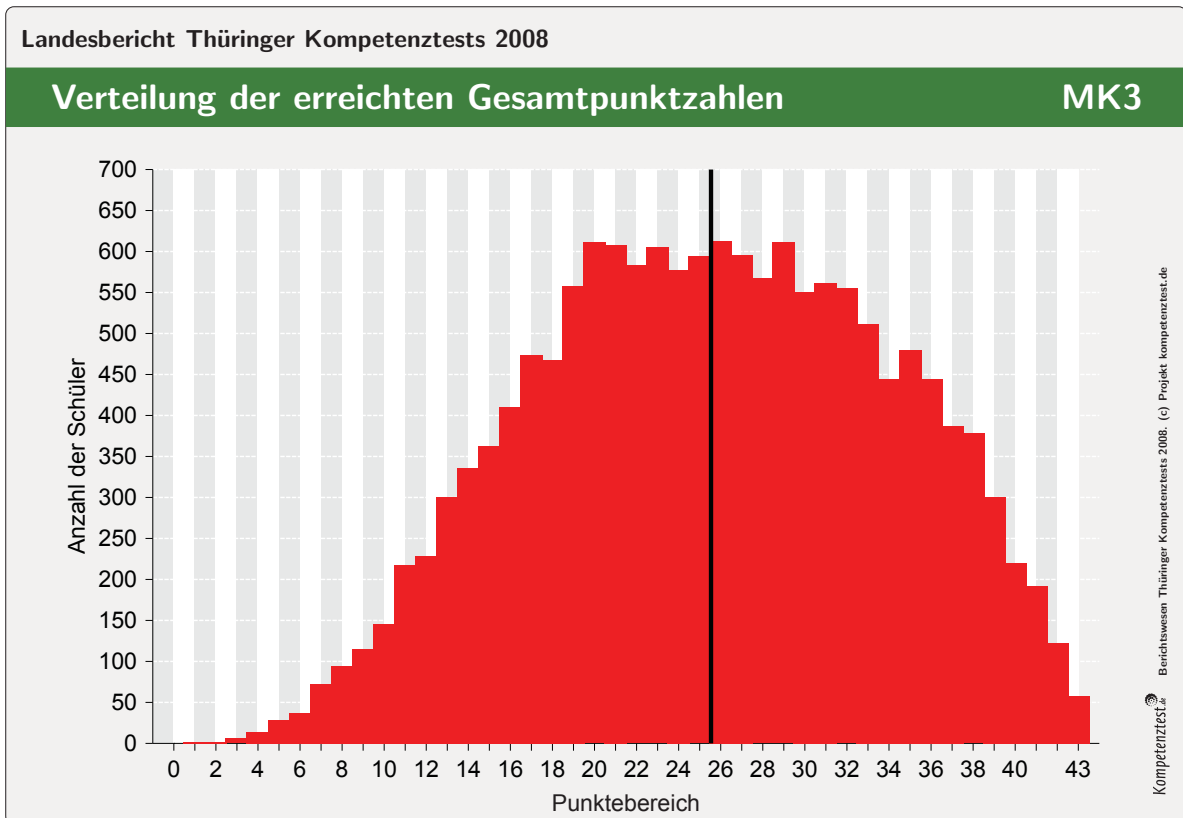


Abbildung 6.2: Verteilung der Gesamtpunktzahlen im Fach Mathematik, Klassenstufe 3

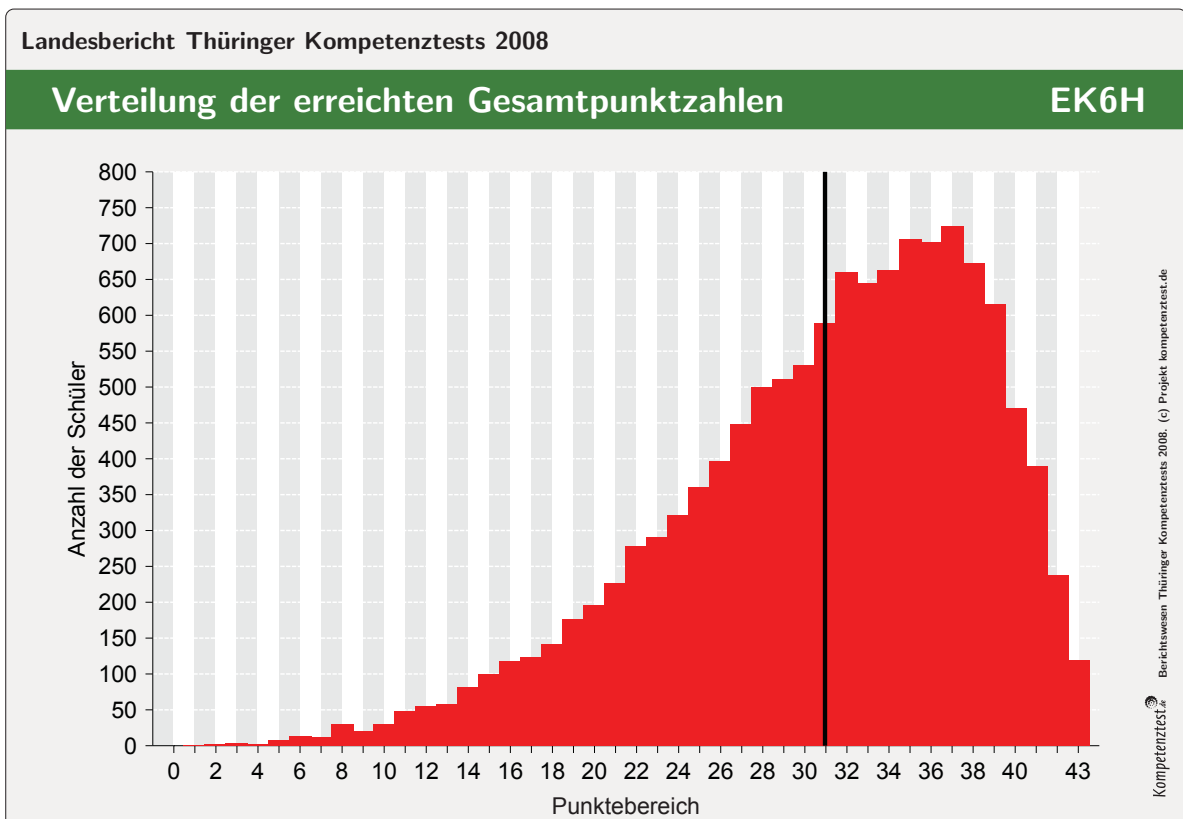


Abbildung 6.3: Verteilung der Gesamtpunktzahlen im Fach Englisch Hörverstehen, Klassenstufe 6

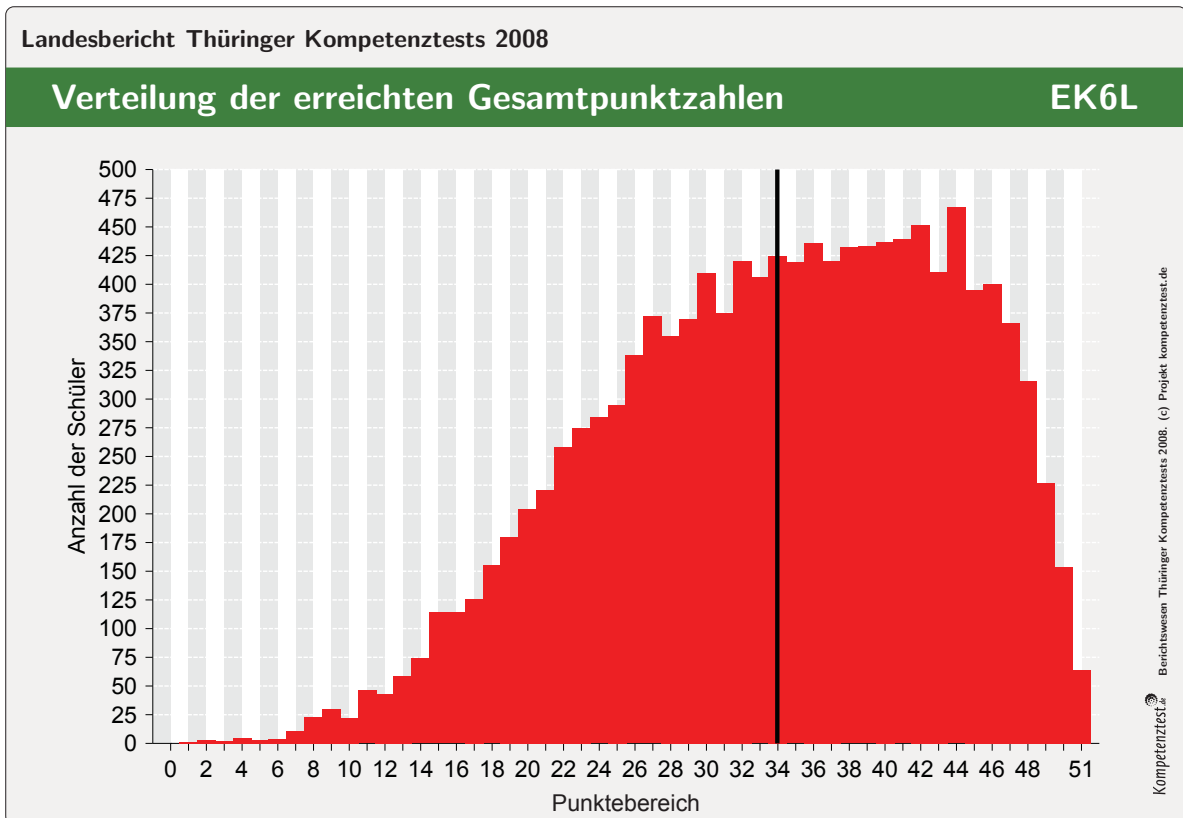


Abbildung 6.4: Verteilung der Gesamtpunktzahlen im Fach Englisch Leseverstehen, Klassenstufe 6

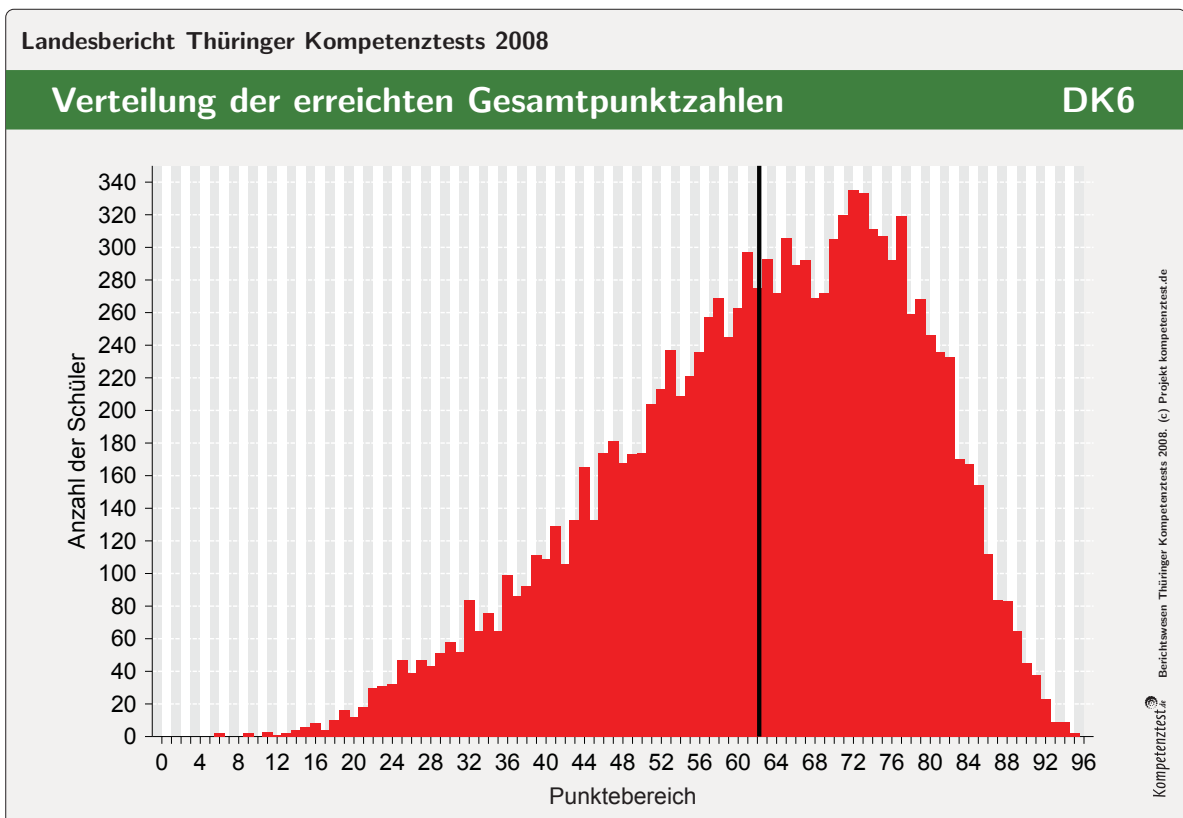


Abbildung 6.5: Verteilung der Gesamtpunktzahlen im Fach Deutsch, Klassenstufe 6

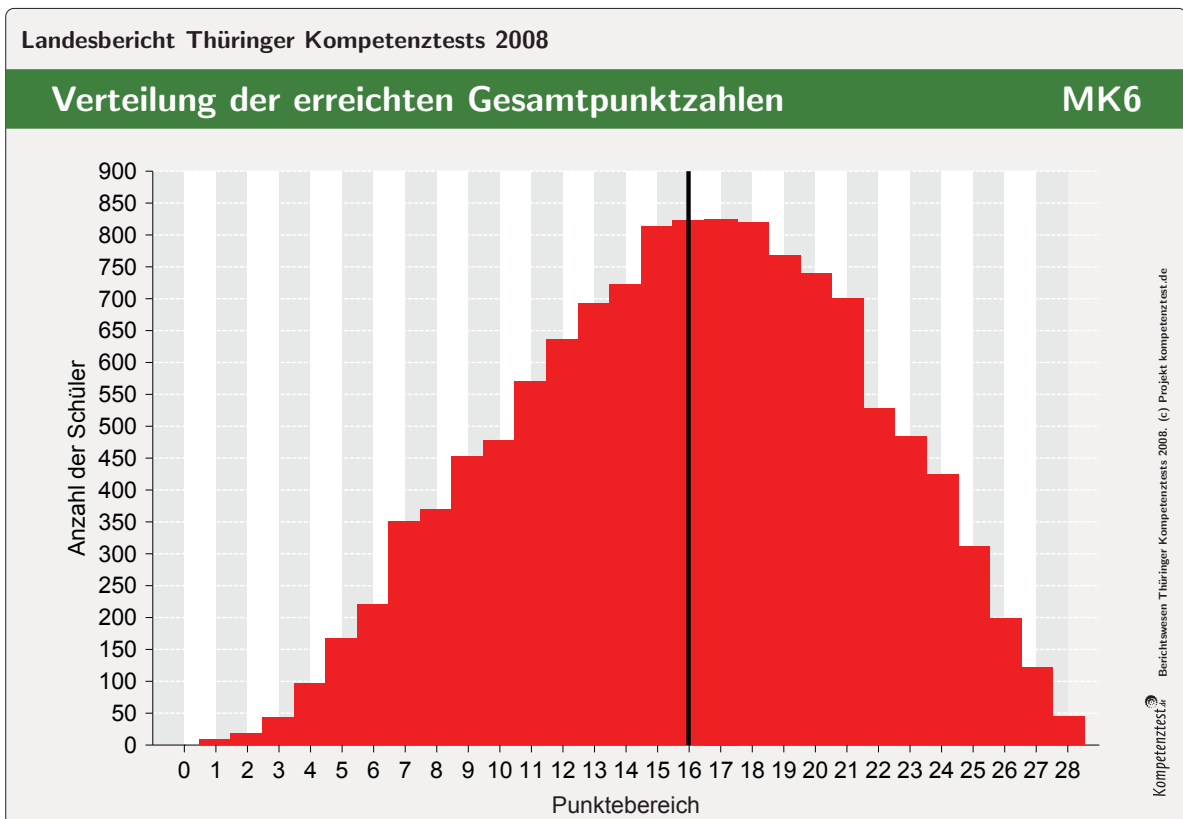


Abbildung 6.6: Verteilung der Gesamtpunktzahlen im Fach Mathematik, Klassenstufe 6

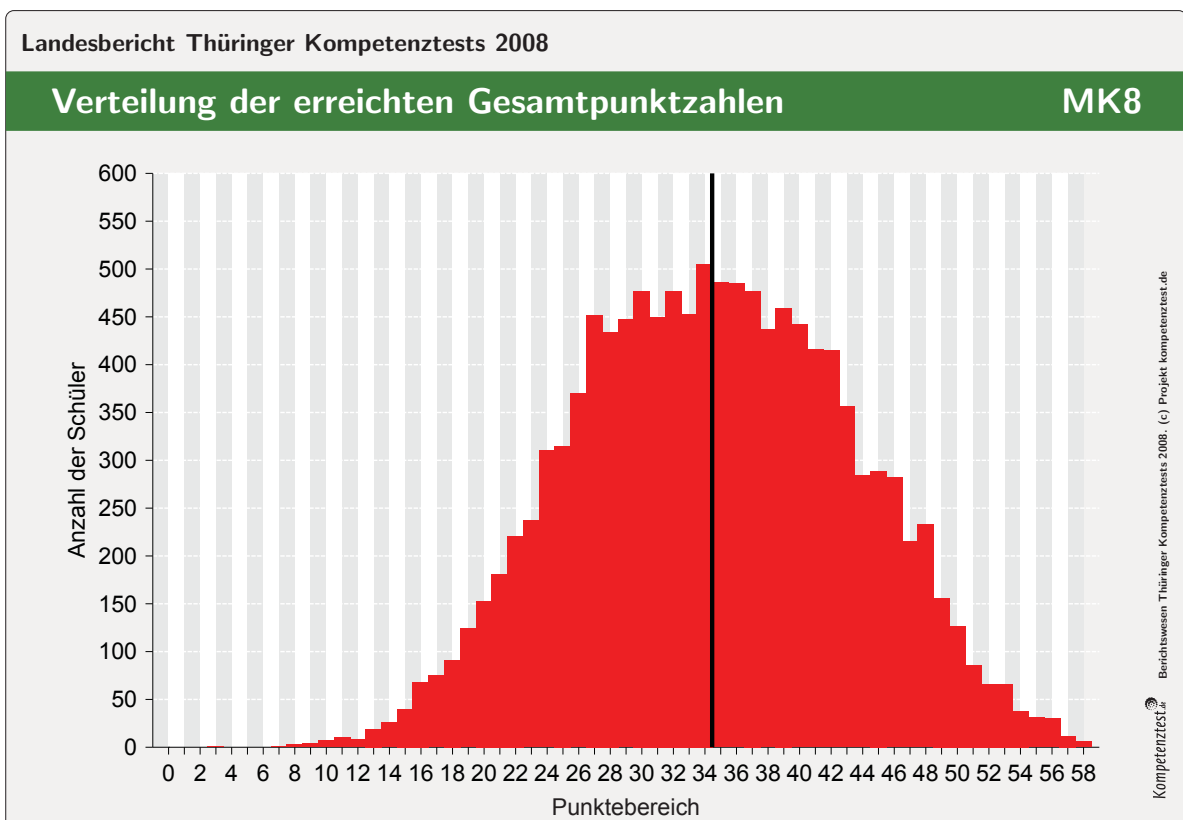


Abbildung 6.7: Verteilung der Gesamtpunktzahlen im Fach Mathematik, Klassenstufe 8

6.2 Erreichte Gesamtpunktzahlen getrennt nach Schulart

In diesem Abschnitt wird die Häufigkeitsverteilung der erreichten Gesamtpunktzahlen hinsichtlich der verschiedenen Schularten differenziert. Die Darstellung erfolgt durch so genannte Perzentilbänder, wie sie auch in den großen internationalen Studien wie PISA üblich sind. Die erreichten Punktzahlen der Schüler werden dabei durch einen waagerechten Balken gekennzeichnet. Dessen Abschnitte, die so genannten Perzentile, geben jeweils den Punktwert wieder, den ein bestimmter Prozentsatz der Schüler als Maximum erreicht hat.

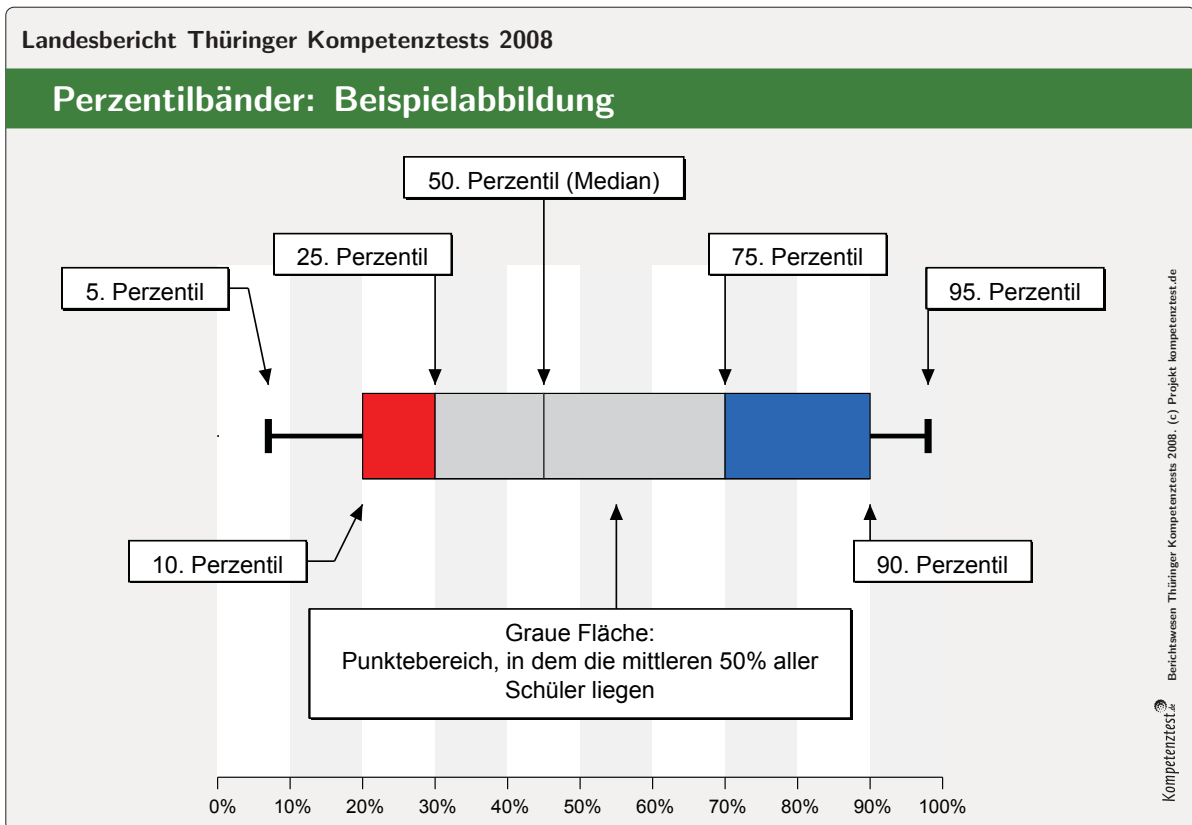


Abbildung 6.8: Beispiel für die Darstellung der erreichten Punktzahl in Form eines Perzentilbandes

Abbildung 6.8 zeigt dies an einem Beispiel. Wiedergegeben ist das 5., 25., 50., 75., 90. und 95. Perzentil. Der Grafik ist beispielsweise zu entnehmen, dass das leistungsschwächste Viertel der Schüler (25. Perzentil) maximal 30 % der Punkte erreicht hat.

Insgesamt zeigen sich die erwarteten Unterschiede zwischen den verschiedenen Schularten, mit Hilfe dessen man auch die Schularten nach Leistungsniveau differenzieren könnte. Besonders deutlich sind diese Unterschiede zwischen Gymnasien und Regelschulen: Das 25. Perzentil der Gymnasiasten liegt in allen Tests der Klassenstufe 6 mindestens auf Höhe des 75. Perzentils der Regelschüler; im Englischtest Hörverstehen der Klassenstufe 6 sowie im Mathematiktest der Klassenstufe 8 sind die Differenzen noch größer. Das 25. Perzentil der Gymnasialergebnisse liegt hierbei deutlich über dem 75. Perzentil der Regelschüler.

Im Anhang zum Landesbericht findet sich eine Aufschlüsselung der Testergebnisse zu den verschiedenen Inhalts- und Anforderungsbereichen. Diese wurden jedoch nur für die Fälle berechnet, wo der untersuchte Inhalts- bzw. Anforderungsbereich

genügend Punkte für eine Darstellung enthielt.

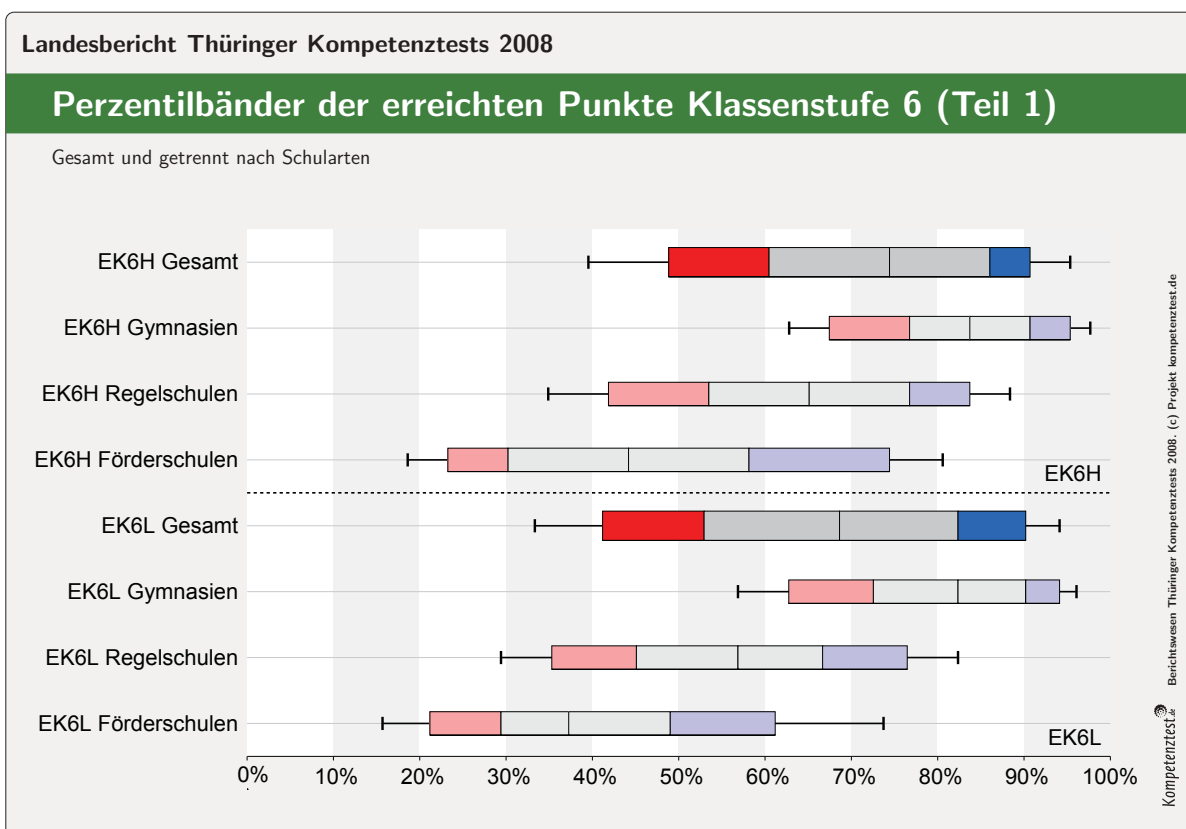


Abbildung 6.9: Ergebnisse bei den Tests Englisch in der Klassenstufe 6 getrennt nach Schulart. Die Abbildung gibt die erreichte Punktzahl in Prozent wieder.

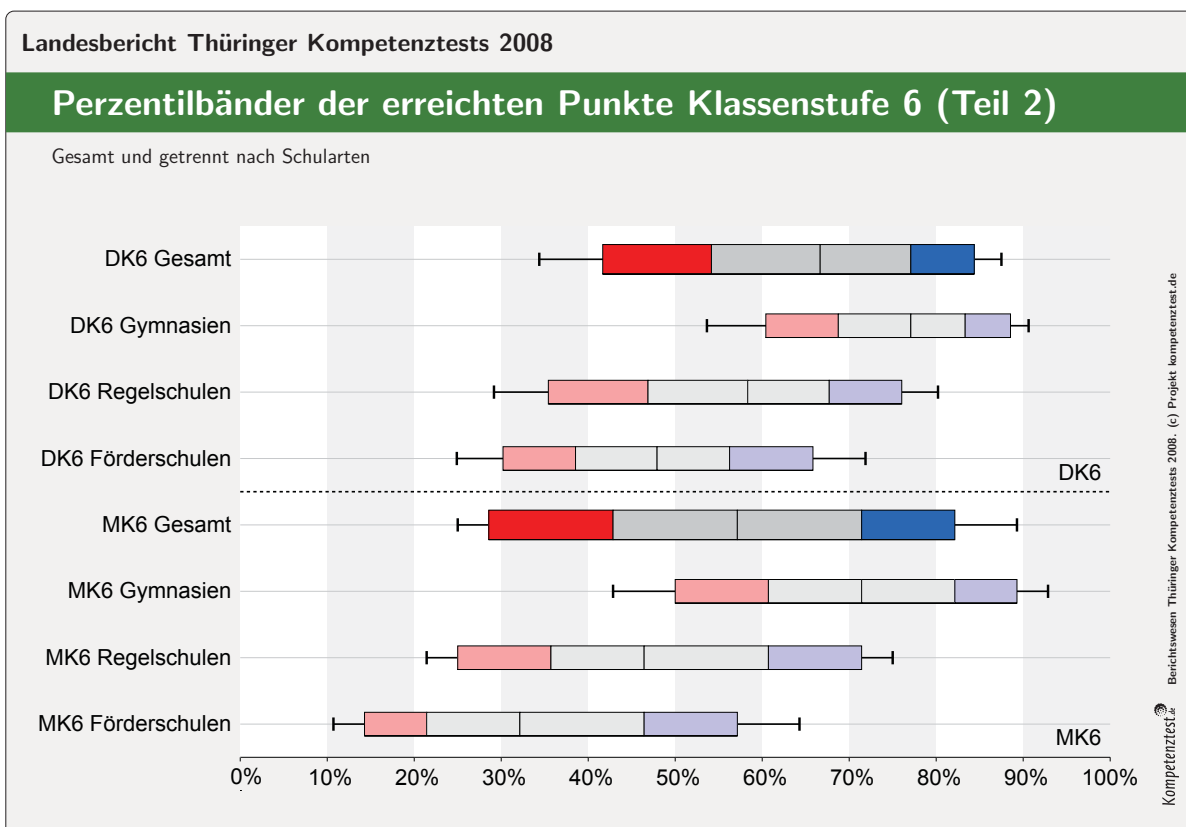


Abbildung 6.10: Ergebnisse bei den Tests Deutsch und Mathematik in der Klassenstufe 6 getrennt nach Schulart. Die Abbildung gibt die erreichte Punktzahl in Prozent wieder.

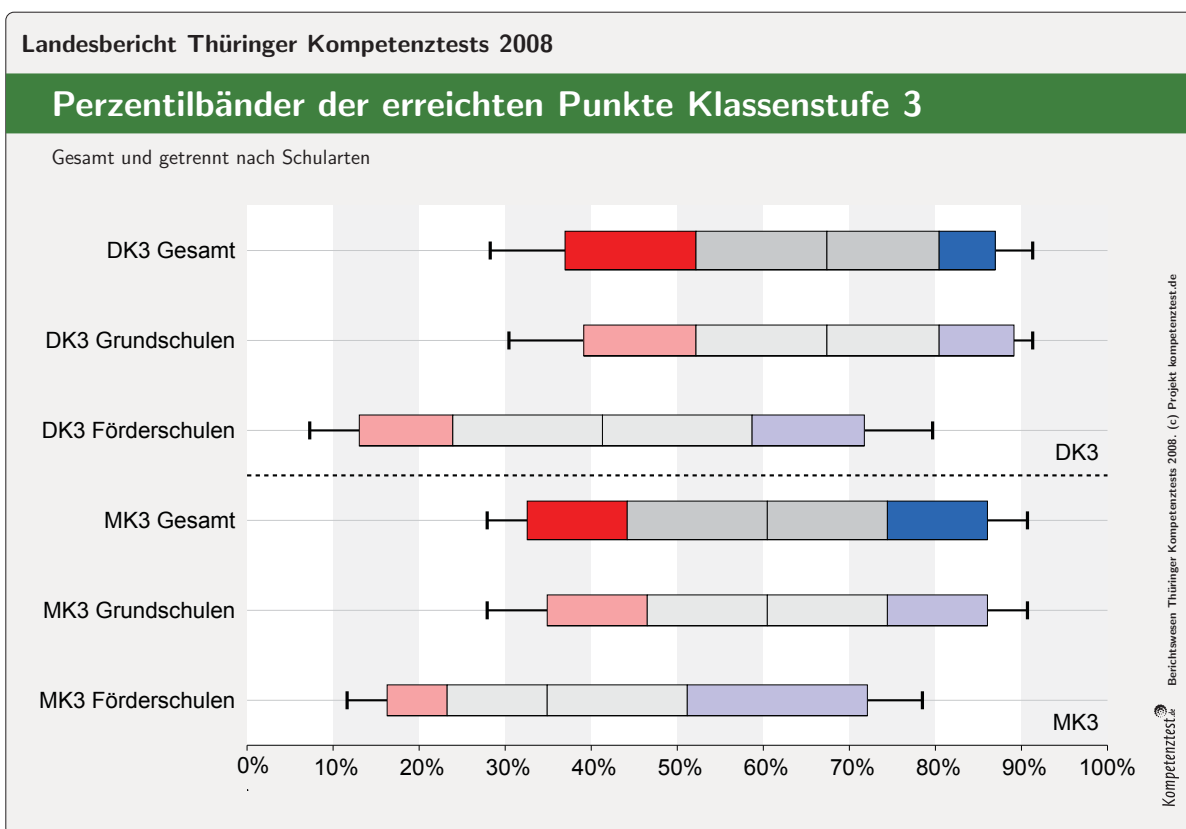


Abbildung 6.11: Ergebnisse bei den Tests Deutsch und Mathematik in der Klassenstufe 3 getrennt nach Schulart. Die Abbildung gibt die erreichte Punktzahl in Prozent wieder.

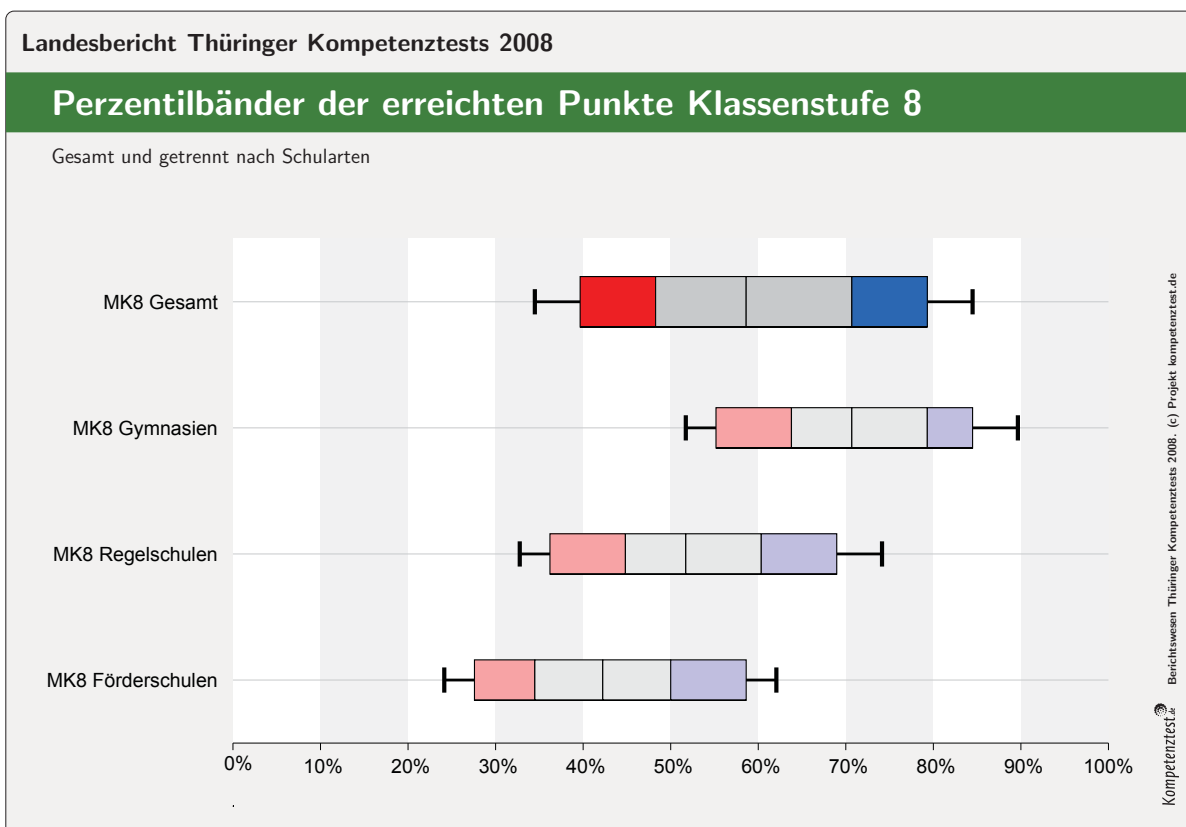


Abbildung 6.12: Ergebnisse bei dem Test Mathematik in der Klassenstufe 8 getrennt nach Schulart. Die Abbildung gibt die erreichte Punktzahl in Prozent wieder.

6.3 Aufgabenschwierigkeiten

Neben der Analyse der Gesamtergebnisse sind die Ergebnisse in den Teilbereichen der Tests und in den einzelnen Aufgaben von besonderem Interesse. Hier zeigt sich, in welchen Bereichen die Stärken und Schwächen der Schüler liegen.

Wie wird die Schwierigkeit einer Aufgabe bestimmt? Die Schwierigkeit berechnet sich als die im Durchschnitt von den Schülern erreichte Punktzahl im Verhältnis zur maximal erreichbaren Punktzahl. Sind zum Beispiel bei einer Aufgabe 5 Punkte maximal möglich, und die Schüler erreichen im Durchschnitt 3 Punkte, dann ist die Schwierigkeit dieser Aufgabe $\frac{3}{5} = 0,6$. Anders ausgedrückt erreichen die Schüler im Mittel 60 % der maximalen Punktzahl. Bei einem Prozentwert nahe 0 ist die Aufgabe als sehr schwierig zu interpretieren. Liegt der Prozentwert jedoch nahe 100, so kann (fast) jeder Schüler dieser Schulart die Aufgabe vollständig lösen, die Aufgabe ist folglich für die Schüler als sehr leicht einzustufen.

Die folgenden Abbildungen geben die Aufgabenschwierigkeit bei den verschiedenen Teilaufgaben der Tests wieder. Dargestellt ist der Landesdurchschnitt der erreichten Punkte in Prozent. Die Angaben sind außerdem nach der Schulart aufgeschlüsselt. Hier zeigen sich die erwarteten Unterschiede zwischen den Schularten. Zusätzlich sind in den Abbildungen die zugehörigen Anforderungsbereiche bzw. Kompetenzniveaus wiedergegeben.

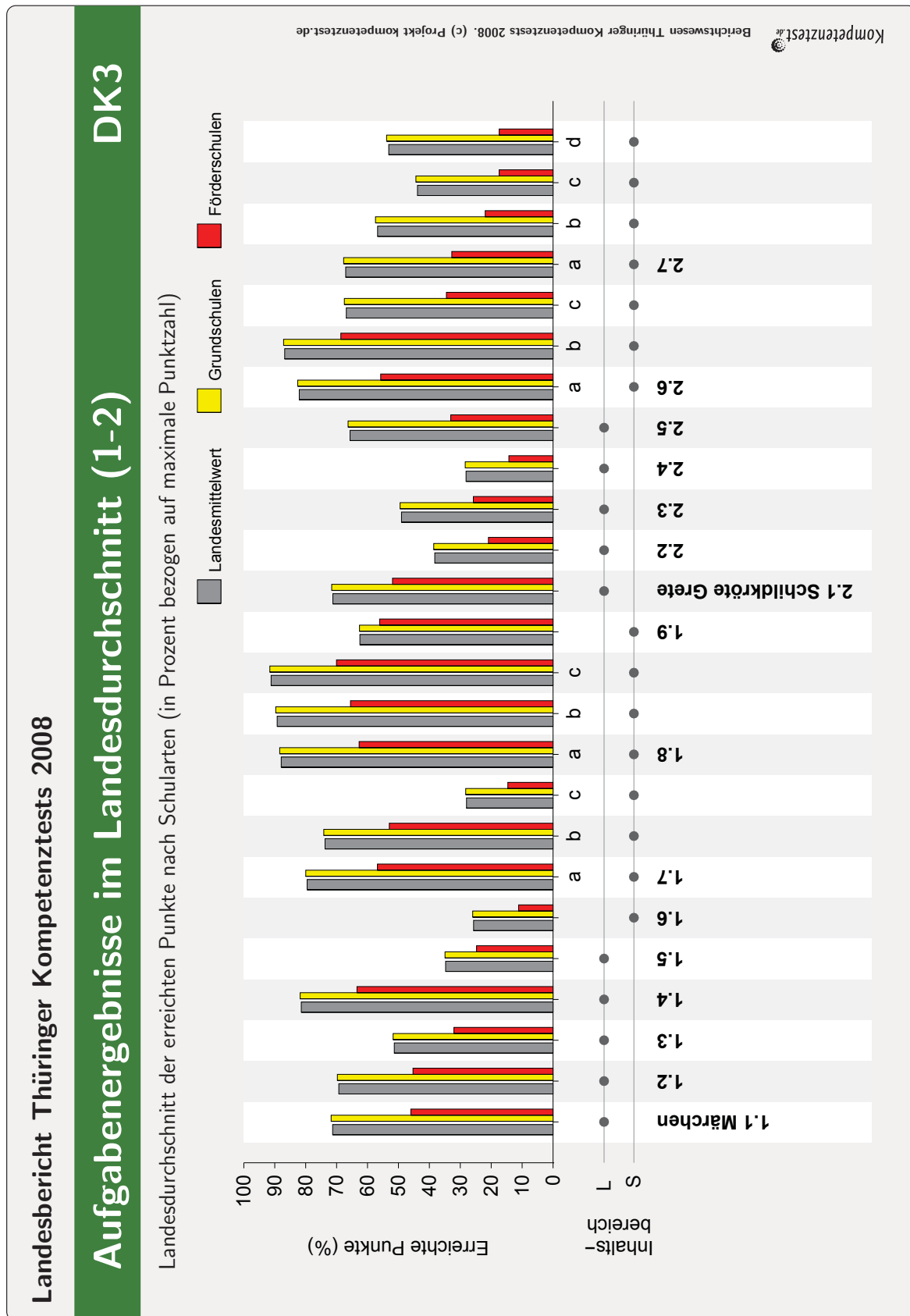


Abbildung 6.13: Aufgabenschwierigkeiten im Landesdurchschnitt sowie getrennt nach Schulart (Deutschtest, Klassenstufe 3, DK3 - Teil 1)

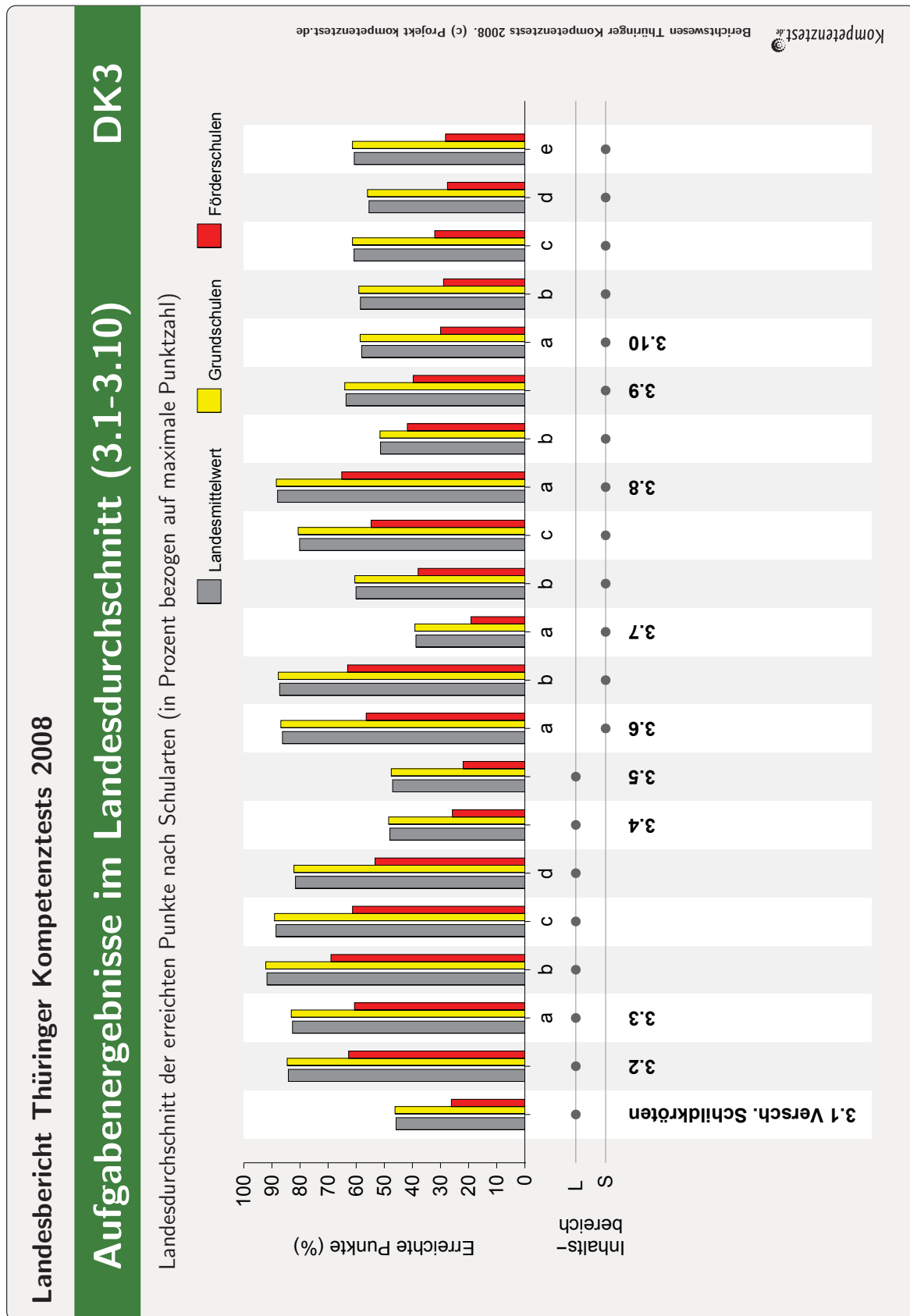


Abbildung 6.14: Aufgabenschwierigkeiten im Landesdurchschnitt sowie getrennt nach Schulart (Deutschtest, Klassenstufe 3, DK3 - Teil 2)

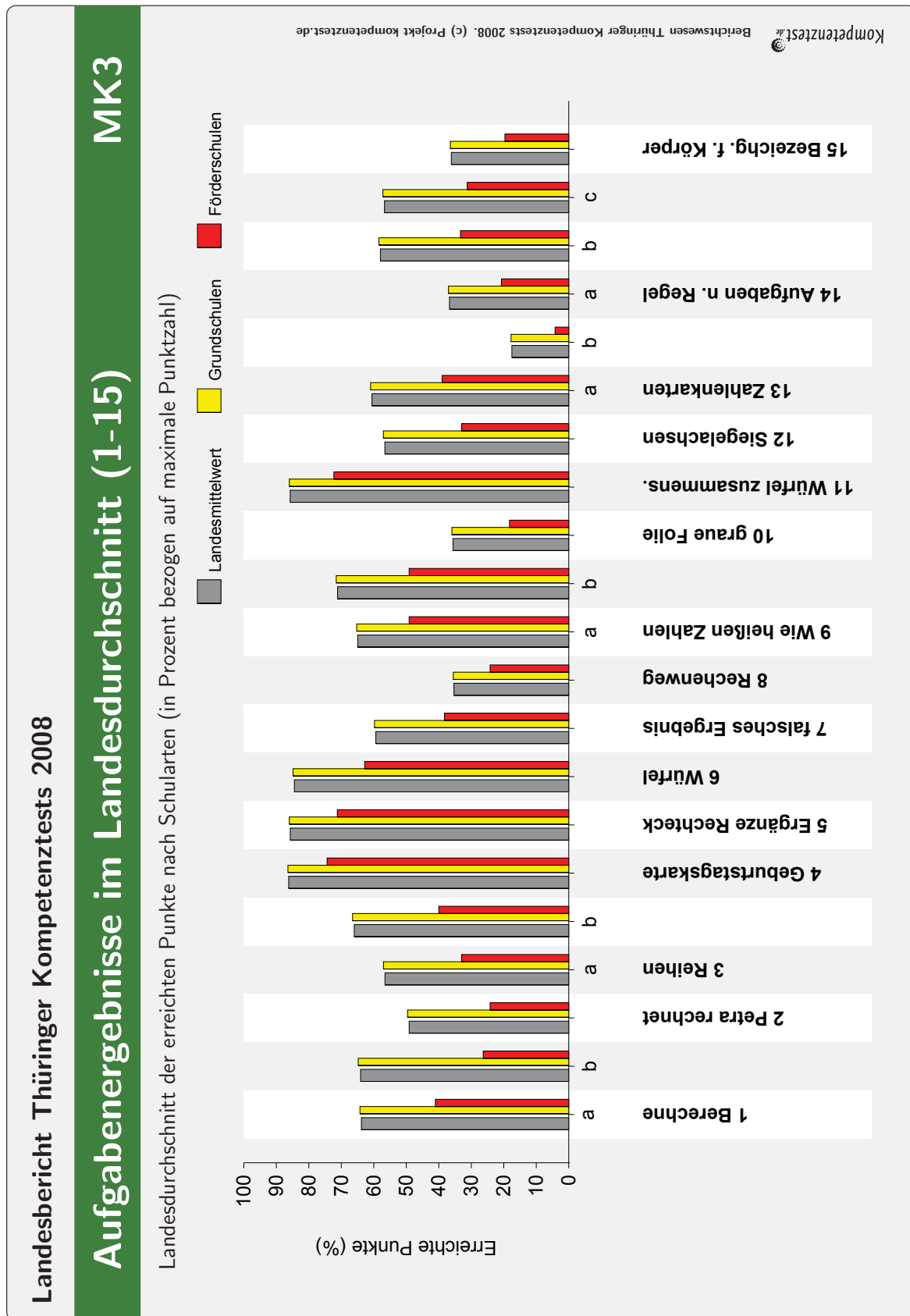


Abbildung 6.15: Aufgabenschwierigkeiten im Landesdurchschnitt sowie getrennt nach Schulart (Mathematiktest, Klassenstufe 3, MK3 - Teil 1)

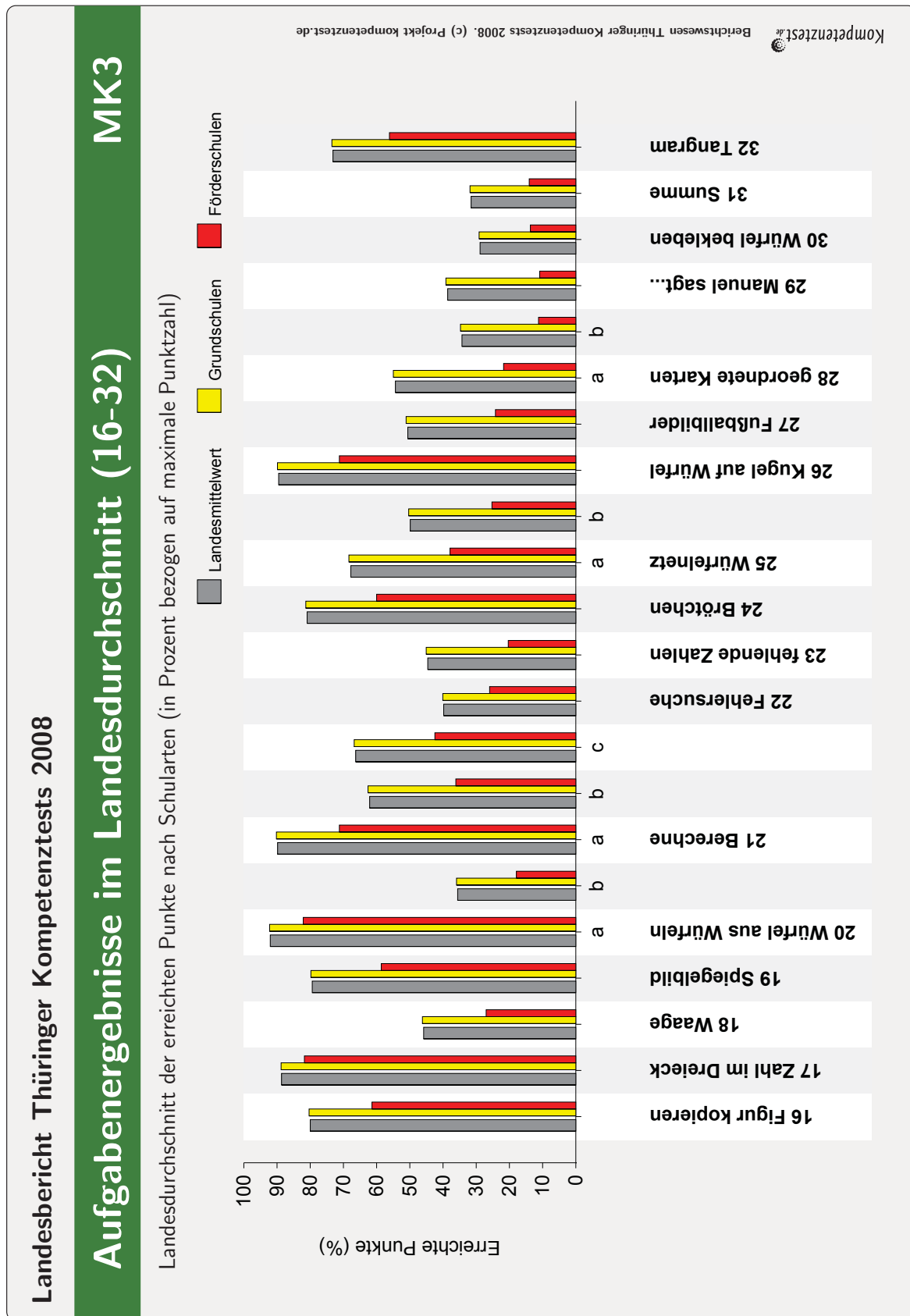


Abbildung 6.16: Aufgabenschwierigkeiten im Landesdurchschnitt sowie getrennt nach Schulart (Mathematiktest, Klassenstufe 3, MK3 - Teil 2)

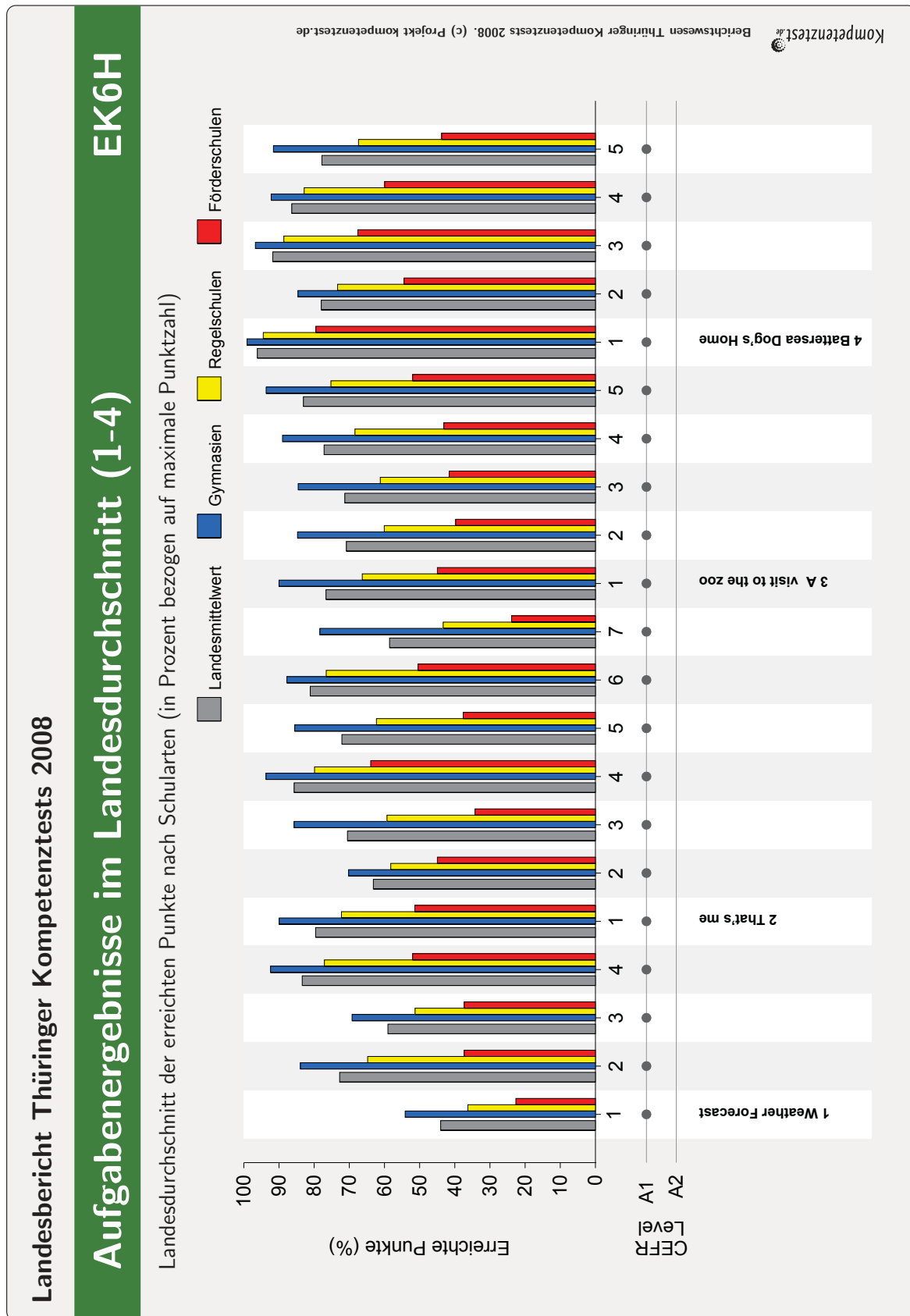


Abbildung 6.17: Aufgabenschwierigkeiten im Landesdurchschnitt sowie getrennt nach Schulart (Englishtest Hörverstehen, Klassenstufe 6, EK6H - Teil 1)

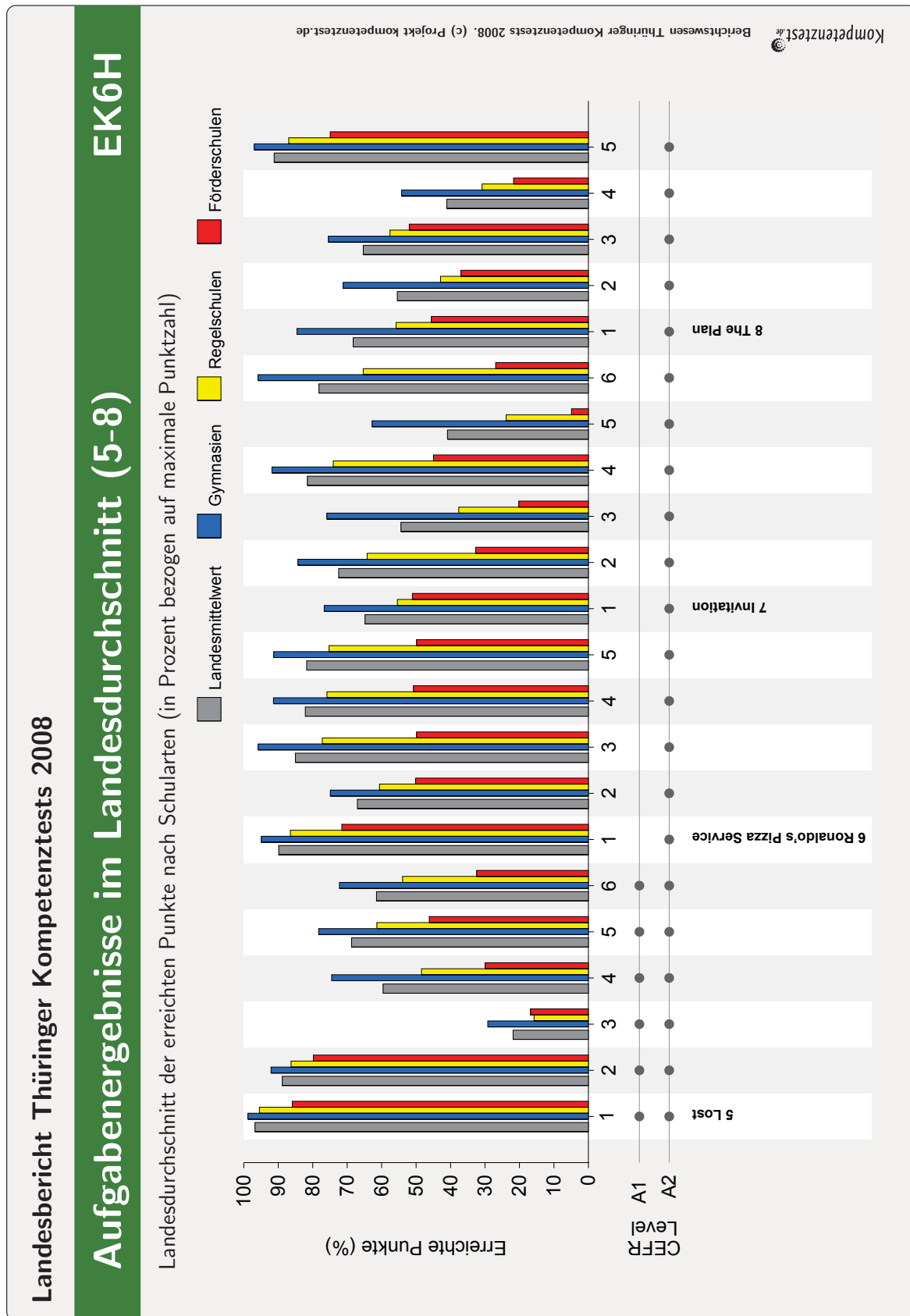


Abbildung 6.18: Aufgabenschwierigkeiten im Landesdurchschnitt sowie getrennt nach Schulart (Englishtest Hörverstehen, Klassenstufe 6, EK6H - Teil 2)

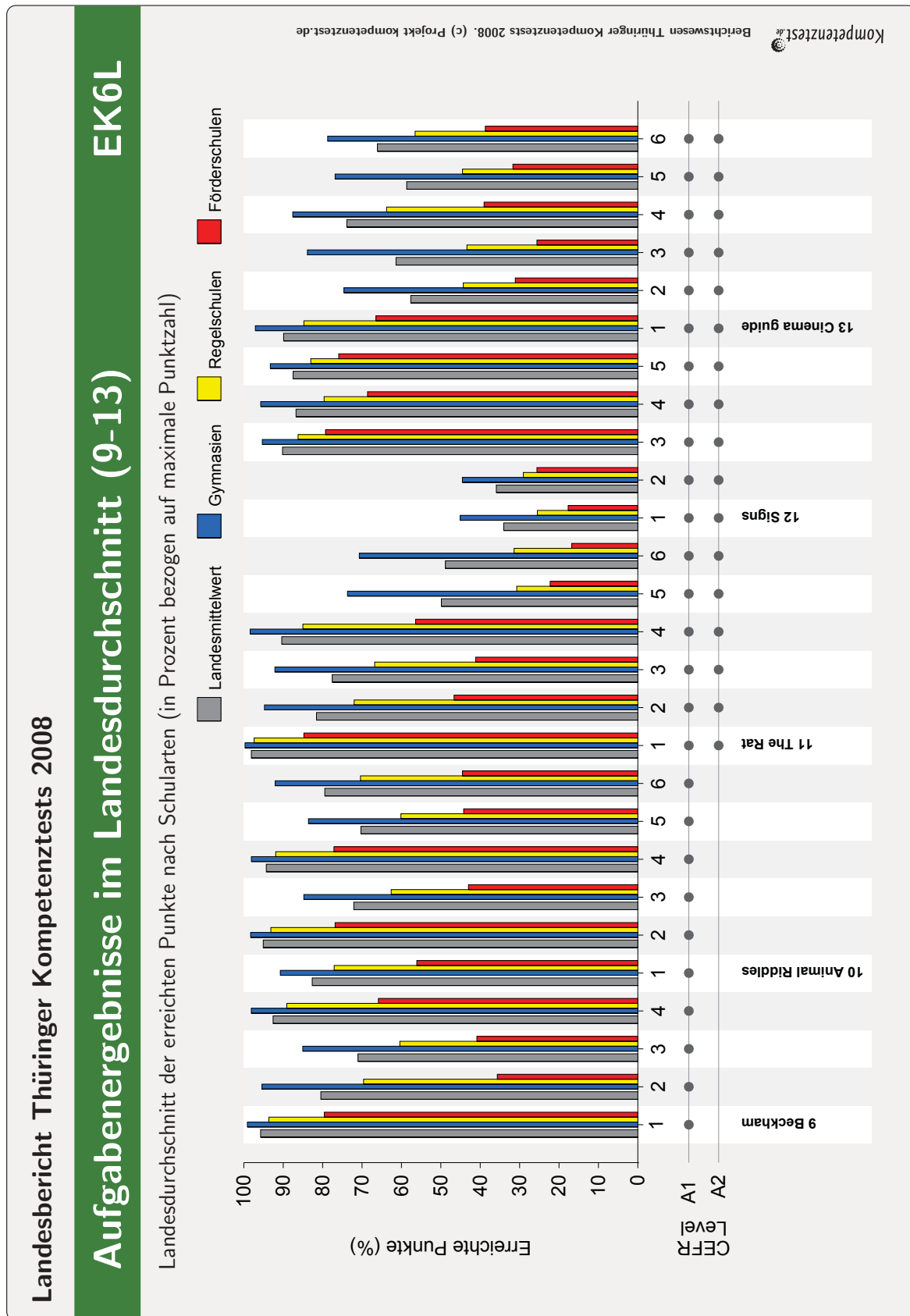


Abbildung 6.19: Aufgabenschwierigkeiten im Landesdurchschnitt sowie getrennt nach Schulart (Englishtest Leseverstehen, Klassenstufe 6, EK6L - Teil 1)

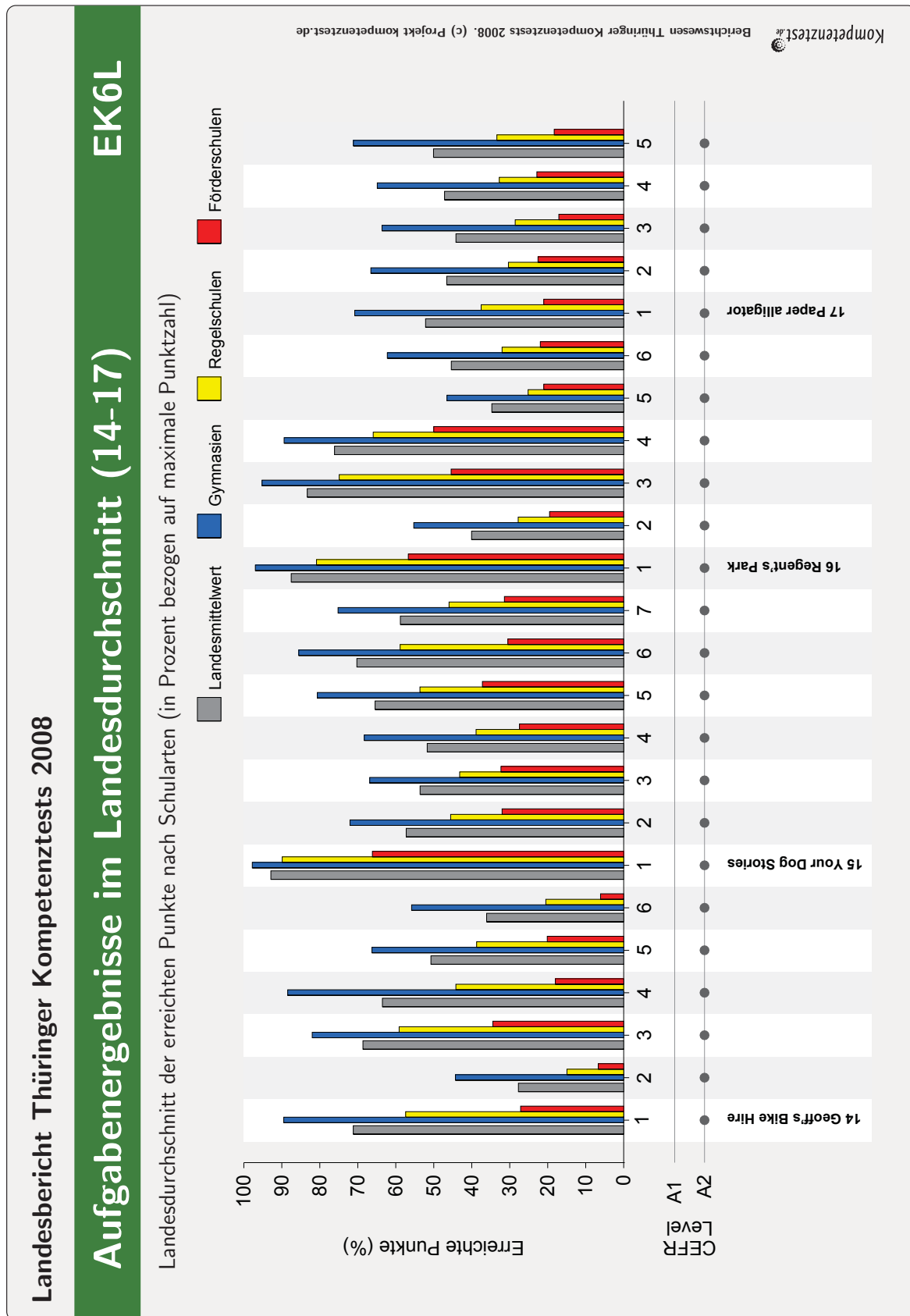


Abbildung 6.20: Aufgabenschwierigkeiten im Landesdurchschnitt sowie getrennt nach Schulart (Englishtest Leseverstehen, Klassenstufe 6, EK6L - Teil 2)

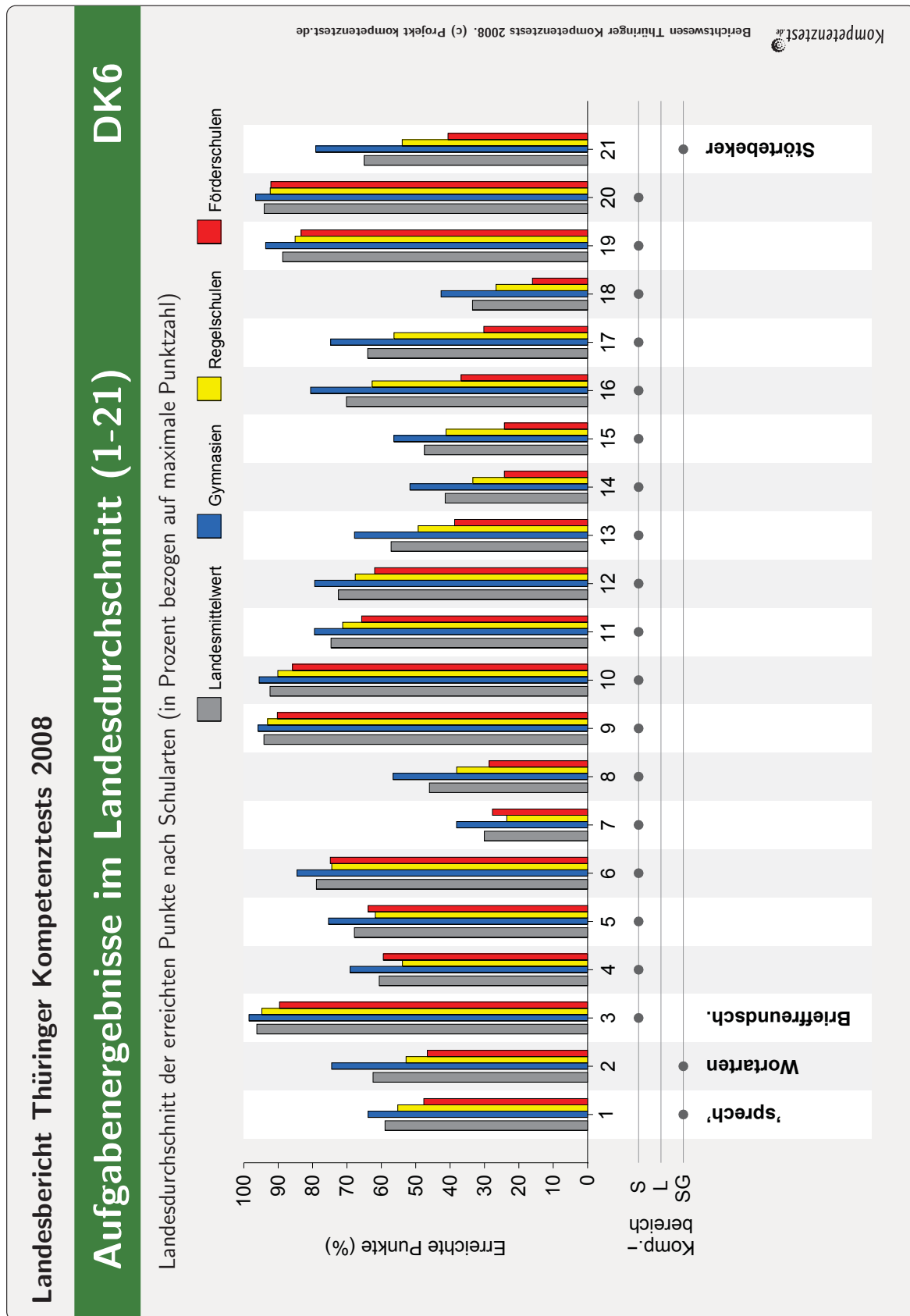


Abbildung 6.21: Aufgabenschwierigkeiten im Landesdurchschnitt sowie getrennt nach Schulart (Deutschtest, Klassenstufe 6, DK6 - Teil 1)

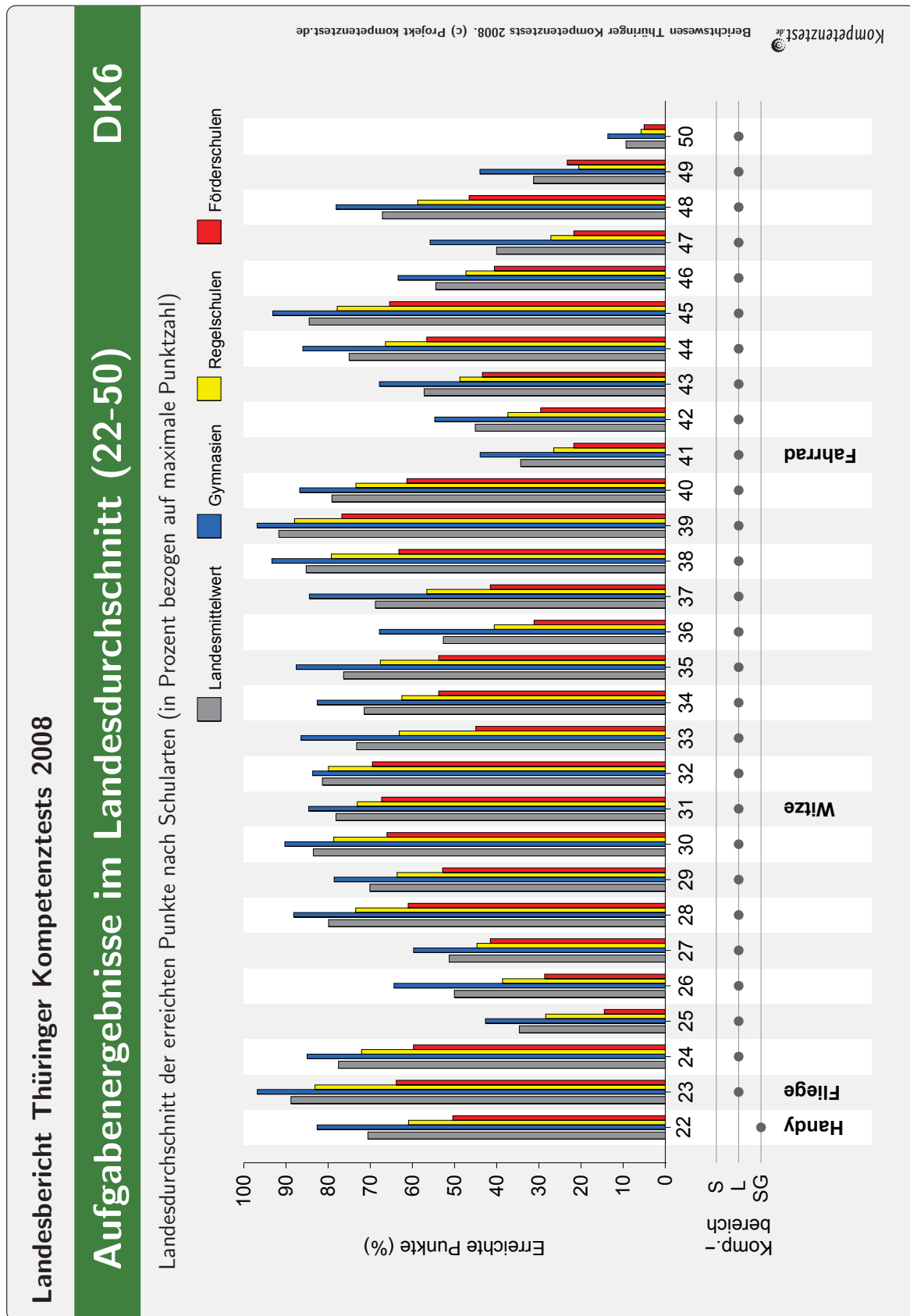


Abbildung 6.22: Aufgabenschwierigkeiten im Landesdurchschnitt sowie getrennt nach Schulart (Deutschtest, Klassenstufe 6, DK6 - Teil 2)

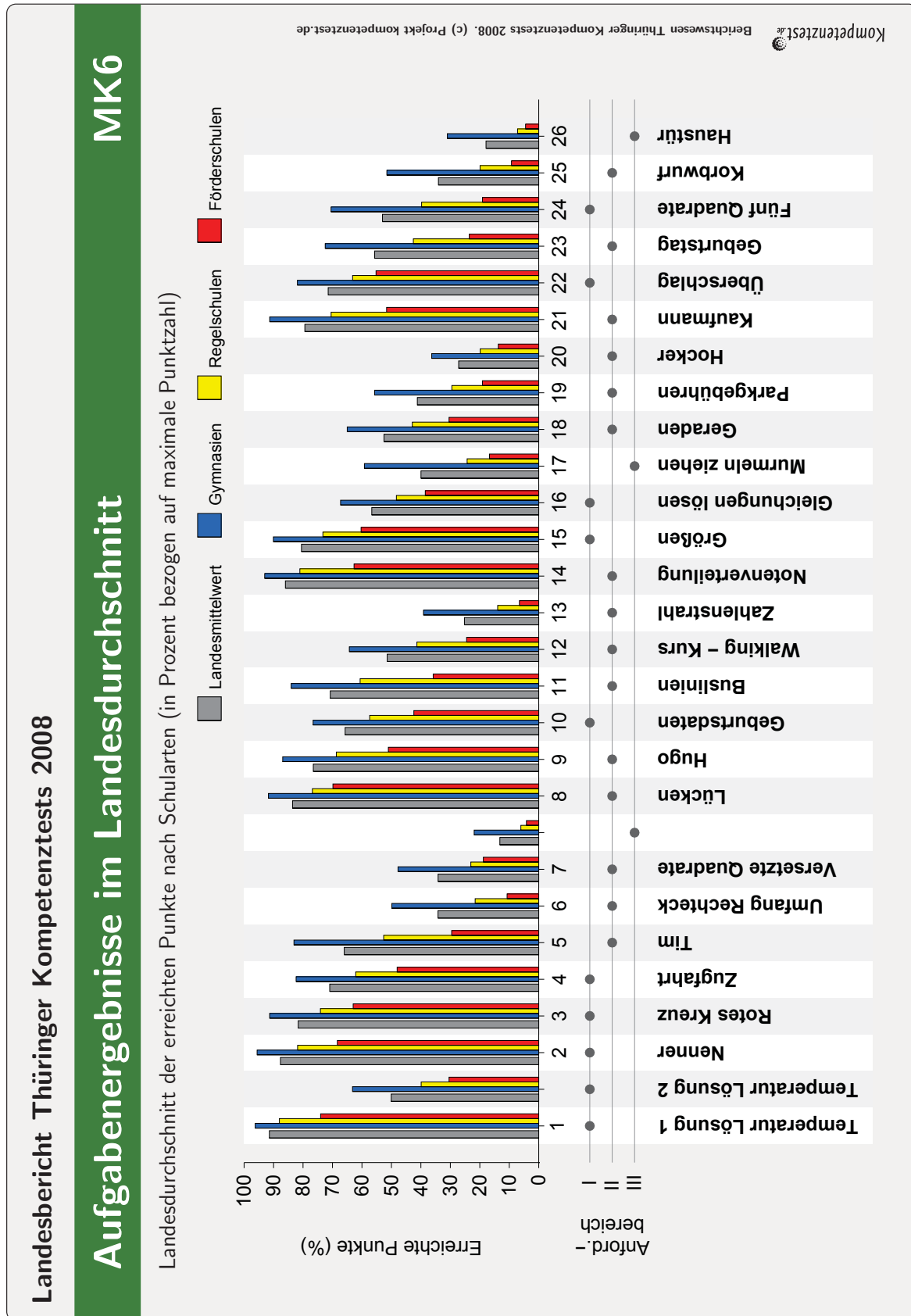


Abbildung 6.23: Aufgabenschwierigkeiten im Landesdurchschnitt sowie getrennt nach Schulart (Mathematiktest, Klassenstufe 6, MK6)

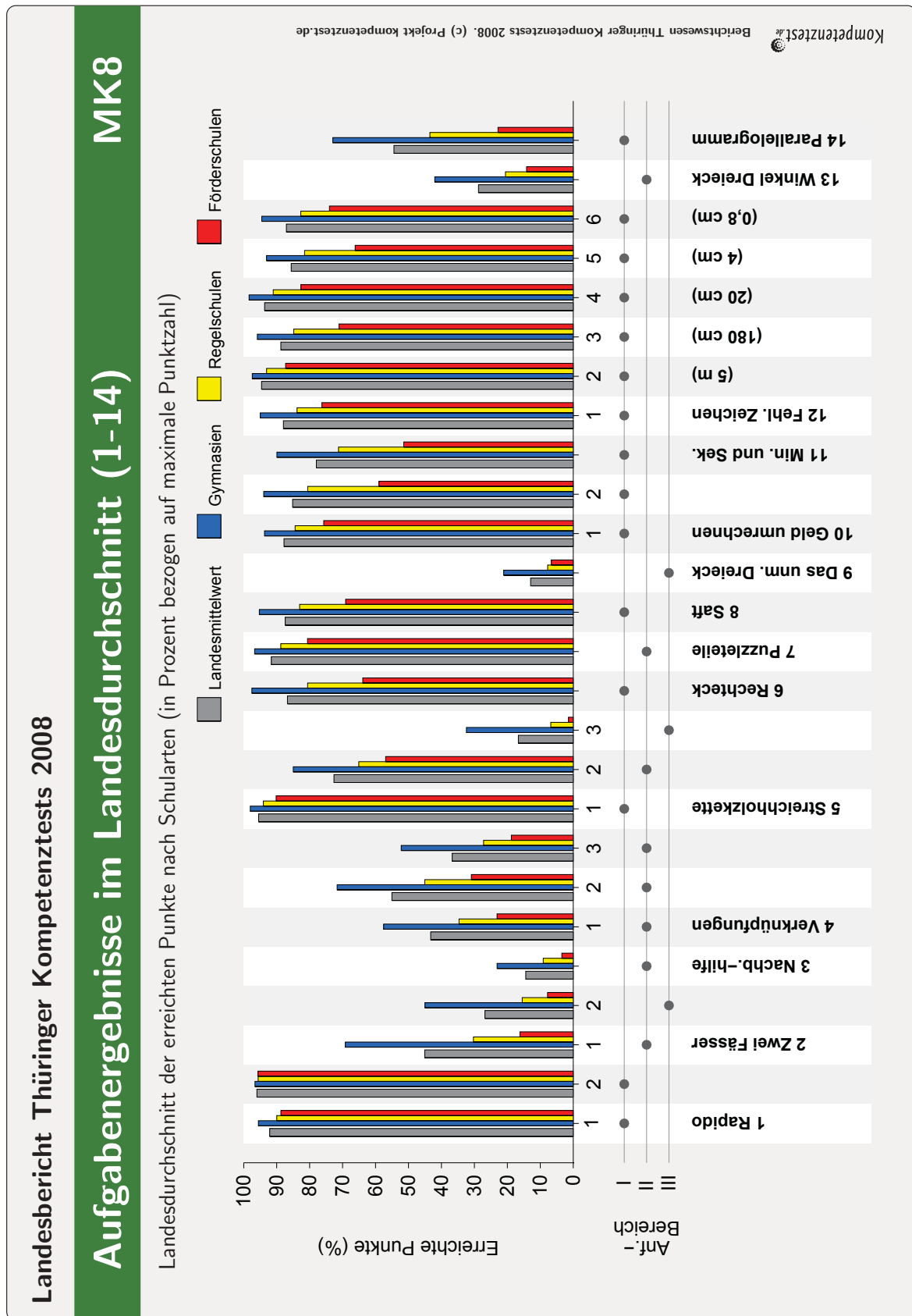
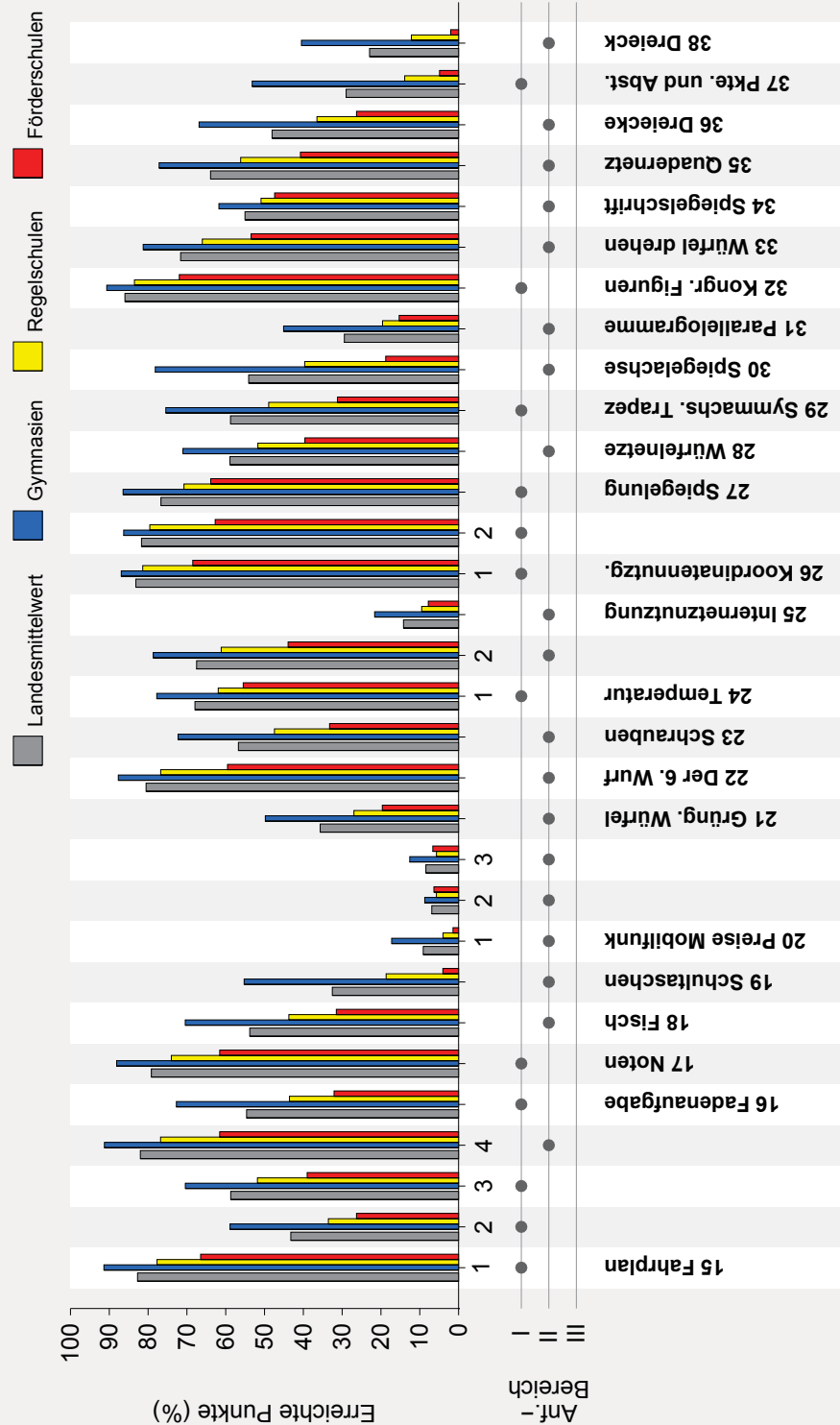


Abbildung 6.24: Aufgabenschwierigkeiten im Landesdurchschnitt sowie getrennt nach Schulart (Mathematiktest, Klassenstufe 8, MK8 - Teil 1)

Aufgabenergebnisse im Landesdurchschnitt (15-38) MK8

Landesdurchschnitt der erreichten Punkte nach Schularten (in Prozent bezogen auf maximale Punktzahl)



Kompetenztest.de
Berichtswesen Thüringer Kompetenztests 2008. (c) Projekt kompetenztest.de

Abbildung 6.25: Aufgabenschwierigkeiten im Landesdurchschnitt sowie getrennt nach Schulart (Mathematiktest, Klassenstufe 8, MK8 - Teil 2)

7 Analyse leistungsrelevanter Faktoren

7.1 Einflussfaktoren auf Klassenebene

Auf Klassenebene gibt es eine Vielzahl von Faktoren, welche die Schülerleistung unterschiedlich beeinflussen können. Neben allgemeinen Rahmenbedingungen wie Klassengröße und materielle Ausstattung sind hier vor allem Aspekte der Unterrichtsgestaltung sowie der Art und Weise der Bearbeitung fachlicher Inhalte zu nennen. Im Rahmen der Analyse wurde den teilnehmenden Fachlehrern zu diesem Themenkreis ein Online-Fragebogen auf dem Schulportal zur Bearbeitung bereitgestellt, der hauptsächlich auf die Reflexion der fachlichen und didaktischen Unterrichtsgestaltung Bezug nimmt. Das sind zum Beispiel Fragen zum Einsatz neuer Medien, zu Sozialformen bei der Gestaltung des Unterrichts sowie zu Formen und Kriterien der Leistungsbewertung. Der vollständige Fragebogen befindet sich im Anhang zum Landesbericht unter www.kompetenztest.de.

Nachfolgend sind nur die prägnantesten Ergebnisse dargestellt. Zu jedem Teilabschnitt werden die Ergebnisse des Lehrerfragebogens in Form von Häufigkeiten (in Prozent) sowie Zusammenhänge zwischen der Lehrerreflexion auf die Unterrichtsgestaltung und den Testergebnissen der Schülerleistungen aufgezeigt. Ein bestehender Zusammenhang wird durch die Produkt-Moment-Korrelation beziehungsweise durch Mittelwertunterschiede angegeben. Hierbei muss beachtet werden, dass der Zusammenhang nicht zwingend im Sinne einer Ursache und deren Wirkung interpretiert werden darf. Es kann also beispielsweise bei Übereinstimmungen von hohen Testergebnissen der Schüler und bevorzugter Anwendung des Frontalunterrichts nicht kausal geschlussfolgert werden, dass Frontalunterricht die Ursache für die hohen Testergebnisse ist. Vielmehr weisen die Korrelationen auf mögliche Einflussfaktoren für die Testleistung hin und müssen in ihrer Gesamtheit mit weiteren Einflussgrößen gesehen werden.

In der Klassenstufe 3 wurden die beiden Schularten Grund- und Förderschule gemeinsam ausgewertet; in der Klassenstufe 6 erfolgte die Auswertung nach Schularten getrennt (Regelschule, Gymnasium, Förderschule). Für die Förderschulen wurden aufgrund der geringen Teilnehmerzahlen keine Korrelationen berechnet.

Insgesamt haben 96,94 % der Lehrer den Lehrerfragebogen ausgefüllt. Im Vergleich zum letzten Jahr (92,88 %) fiel die Teilnahmequote damit höher aus. Am höchsten war die Beteiligung unter den Englischlehrern (98,7 %). Der entsprechende Wert für die Mathematiklehrer lag bei 97,6 % und für die Deutschlehrer bei 95,2 %. Schulartspezifische Besonderheiten konnten nicht festgestellt werden. Insgesamt gingen 3 513 Fragebögen in die Auswertung ein.

7.1.1 Mathematik

Fragenkomplex: „Medieneinsatz“

Im Jahresvergleich 2007 und 2008 gibt es nur wenige Veränderungen im Medieneinsatz. Wie aus der Abbildung 7.1 ersichtlich, sind die mit Abstand am meisten genutzten Medien die Tafel und das Mathematiklehrbuch. Von wesentlich weniger Lehrern werden Arbeitsblätter, OHP-Folien und Magnettafeln in fast jeder Stunde genutzt. Weiter zugenommen hat in diesem Schuljahr die Nutzung von vorgedruckten und vom Lehrer erstellten Arbeitsblättern. Besonders im Grund- und Förderschulbereich stellt das Arbeitsblatt ein sehr flexibles Medium dar, um Aufgabenstellungen individuell anzupassen.

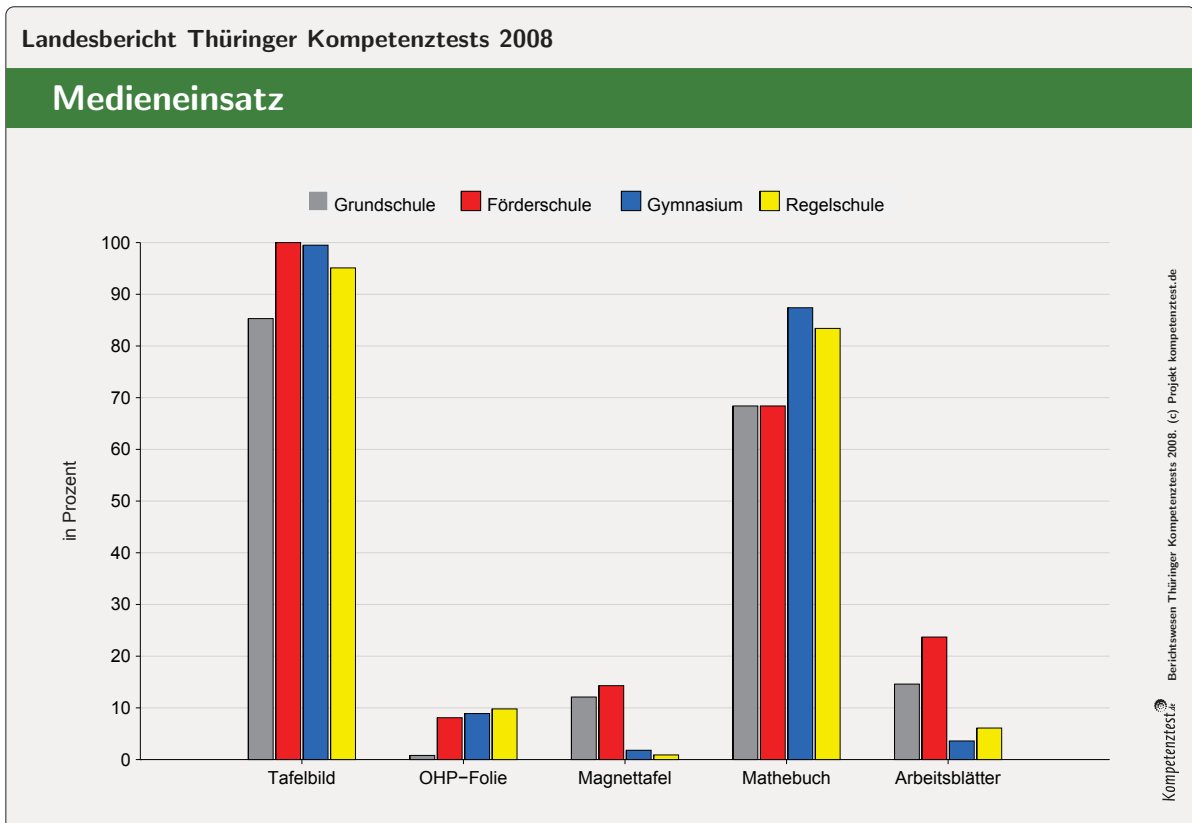


Abbildung 7.1: Folgende Medien werden in jeder Stunde oder in fast jeder Stunde im Unterricht eingesetzt

Interessant ist in diesem Zusammenhang außerdem, welche Medien keine Verwendung im Unterricht finden (siehe Abbildung 7.2). Über 70 % der Lehrer geben an, dass sie kein Flipchart im Unterricht einsetzen. Aber auch vom Schüler erstellte Arbeitsblätter werden von wenigstens 35 % der Lehrkräfte nicht im Unterricht verwendet. Im Förderschulbereich ist diese Zahl mit 65 % noch höher, aber auf Grund der Schülerschaft auch nachvollziehbar.

Die neuen Medien (PC, Internet und Lernsoftware) setzt der überwiegende Teil der Lehrkräfte mehr oder weniger häufig im Unterricht ein (siehe Abbildung 7.3). Wie auch in den letzten Jahren erfolgt der Einsatz der neuen Medien im Mathematikunterricht innerhalb der Grundschule häufiger als in den anderen Schularten. Auch die Anzahl der Lehrkräfte, die angeben, die neuen Medien gar nicht im Unterricht einzusetzen, ist in der Grundschule mit 15,9 % am geringsten. Auf der anderen Seite nutzen 26,1 % der Gymnasiallehrer und 25,4 % der Regelschullehrer nicht die Möglichkeiten, die die neuen Medien im Unterricht bieten.

Landesbericht Thüringer Kompetenztests 2008

Medien, die nie eingesetzt werden

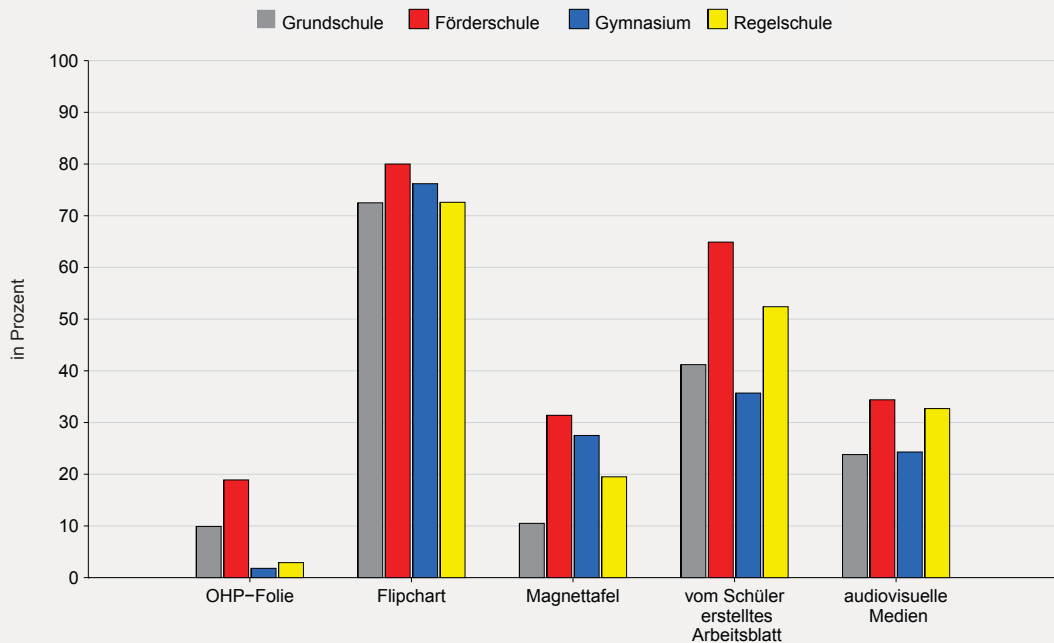

Kompetenztest.de
Berichtswesen Thüringer Kompetenztests 2008. (c) Projekt kompetenztest.de

Abbildung 7.2: Folgende Medien werden in keiner Stunde im Unterricht eingesetzt

Landesbericht Thüringer Kompetenztests 2008

Einsatz neuer Medien (PC, Internet, Lernsoftware)

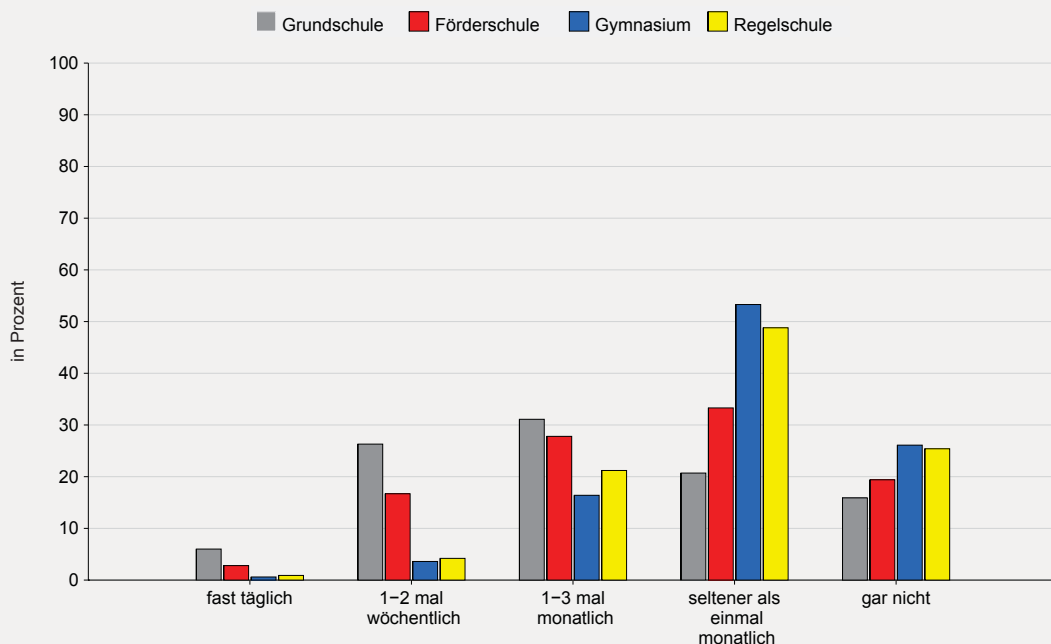

Kompetenztest.de
Berichtswesen Thüringer Kompetenztests 2008. (c) Projekt kompetenztest.de

Abbildung 7.3: Setzen Sie die neuen Medien (PC, Internet, Lernsoftware) in Ihrem Mathematikunterricht ein?

Fragenkomplex: „Methodische Gestaltung des Unterrichts“*Festigung von Grundwissen*

Das sichere Beherrschen von Grundwissen ist eine Voraussetzung für erfolgreiches Lernen. Viele Grundlagen in der Mathematik, wie die Grundrechenarten und das kleine Einmaleins werden in der Grundschule gelehrt. Eine ständige Wiederholung erhöht das Behalten des Gelernten. Dieses Grundwissen ist Voraussetzung für den nachfolgenden Stoff in den weiterführenden Schulen. Das könnte ein Grund sein, dass in der Grundschule die meisten Lehrer (68,1 %) angaben, in jeder beziehungsweise fast jeder Stunde Grundwissen zu festigen. Dieser Wert entspricht etwa dem Ergebnis von 2007 (67,1 %). Auch in den Förderschulen liegen diese Werte auf Vorjahresniveau (52,9 %, 2007 50 %). Anders sieht es in den Gymnasien und Regelschulen aus. Hier wurde sogar ein Rückgang der ohnehin schon kleineren Werte des letzten Jahres beobachtet. So verringerte sich die Anzahl der Gymnasiallehrer, die angaben, Grundwissen, welches nicht zum laufenden Stoff gehört, in jeder beziehungsweise fast jeder Stunde zu sichern, von 45 % auf 38,8 % und bei den Regelschullehrern von 60,4 % auf 53,1 %.

Sozialformen

Aus der Abbildung 7.4 wird ersichtlich, dass die Einzelarbeit schulartübergreifend als eine Sozialform zur Unterrichtsgestaltung favorisiert wird. Im Grundschul-, Förderschul- und Regelschulbereich sind diese Zahlen etwas geringer ausgefallen als im Vorjahr. Die Partnerarbeit hat in der Grund- und Förderschule mittlerweile fast die gleiche Bedeutung wie die Einzelarbeit erreicht. Dagegen ist die Partnerarbeit im Gymnasium und in der Regelschule leicht zurück gegangen (um ca. 7 % gegenüber 2007). Obwohl die Gruppenarbeit in der Grundschule auch in diesem Berichtsjahr eine größere Rolle spielt als im Gymnasial- und Regelschulbereich, sind die Zahlen doch 12 % schlechter als 2007. Eine größere Steigerung von 31,3 % auf 42,4 % konnte nur im Förderschulbereich beobachtet werden.

Unterrichtsformen

Auch in diesem Schuljahr steht der Frontalunterricht an erster Stelle bei den Unterrichtsformen (siehe Abbildung 7.5). Fast alle Regelschul- und Gymnasiallehrer geben an, Frontalunterricht häufig beziehungsweise sehr häufig anzuwenden. Die anderen Unterrichtsformen werden eher selten praktiziert, wobei offene Unterrichtsformen und Stationenlernen im Regelschulbereich noch am häufigsten durchgeführt werden. In den Grund- und Förderschulen spielen auch offene Unterrichtsformen eine recht große Rolle, gefolgt von Tagesplan, Stationenlernen und Wochenplan. Im Förderschulbereich hat die Wochenplanarbeit zu Gunsten der Tagesplanarbeit um ca. 10 % abgenommen. Ganz leichte Zuwächse gab es im Gymnasium und in der Regelschule bei der fachbezogenen und fachübergreifenden Projektarbeit. Allerdings ist der Anteil der fachbezogenen Projektarbeiten, die in jeder oder fast jeder Stunde durchgeführt werden, im Grund- und Förderschulbereich im Vergleich zum letzten Erhebungszeitraum leicht zurück gegangen. Dies trifft auch auf die fachübergreifende Projektarbeit im Förderschulbereich zu.

Hausaufgaben

Mehr als 4/5 aller Lehrer der Grund- und Regelschulen sowie der Gymnasien geben in fast jeder Unterrichtsstunde Hausaufgaben auf. Im Förderschulbereich liegt dieser Anteil bei 47 %. Der Anteil der Lehrkräfte, die in jeder Unterrichtsstunde Hausaufgaben aufgeben, ist mit 41 % im Gymnasium am höchsten (siehe Abbildung 7.6).

Landesbericht Thüringer Kompetenztests 2008

Sozialformen im Unterricht

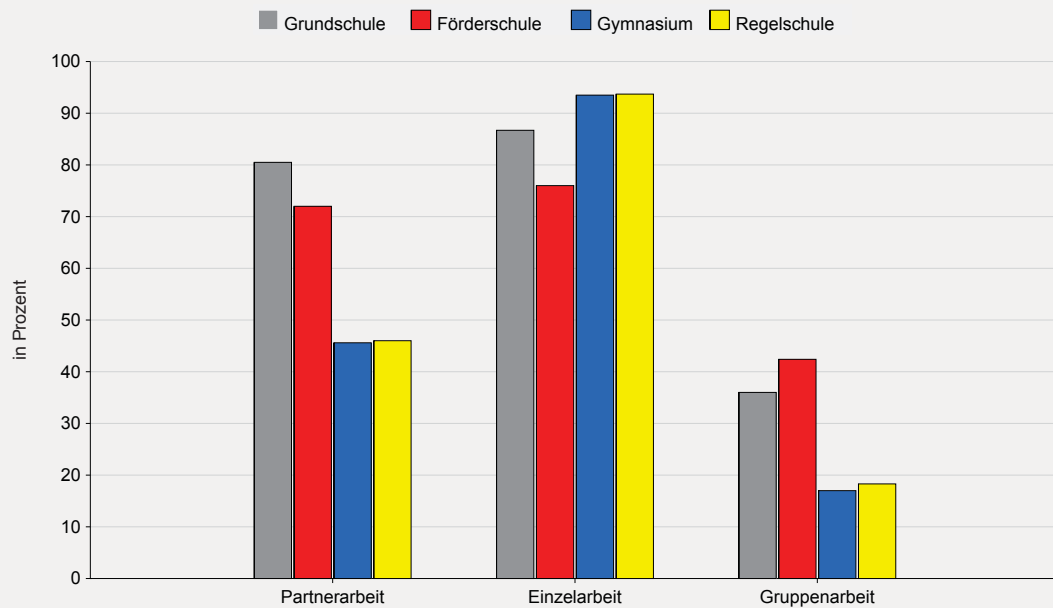
Kompetenztest.de
Berichtswesen Thüringer Kompetenztests 2008. (c) Projekt kompetenztest.de

Abbildung 7.4: Folgende Sozialformen werden sehr häufig oder häufig im Unterricht eingesetzt

Landesbericht Thüringer Kompetenztests 2008

Häufig eingesetzte Unterrichtsformen

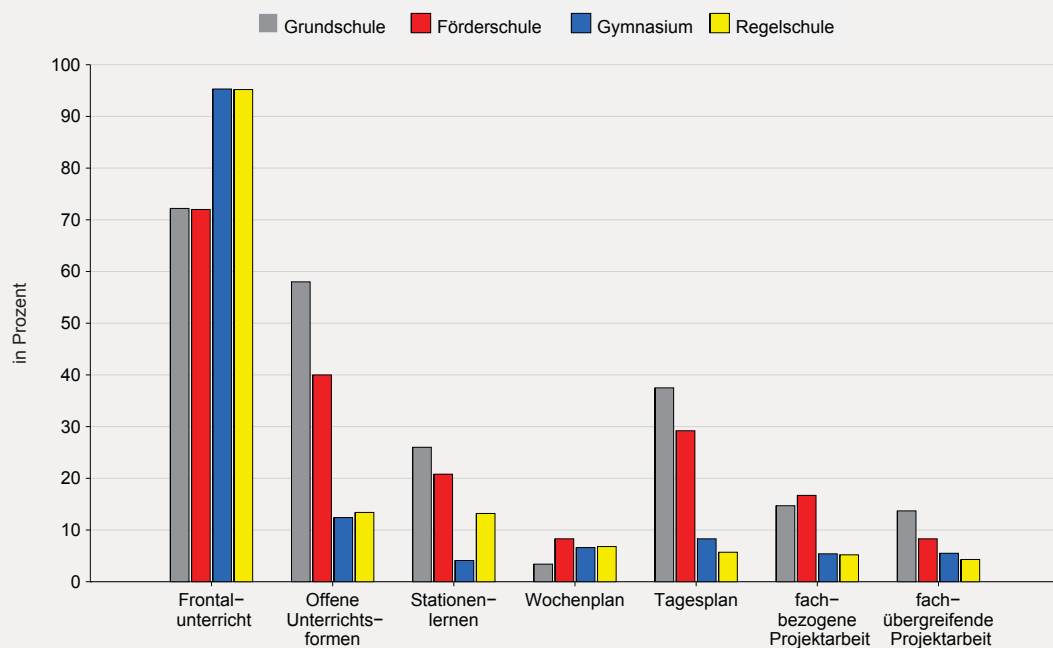
Kompetenztest.de
Berichtswesen Thüringer Kompetenztests 2008. (c) Projekt kompetenztest.de

Abbildung 7.5: Folgende Unterrichtsformen werden in jeder oder in fast jeder Stunde im Unterricht eingesetzt

Landesbericht Thüringer Kompetenztests 2008

Hausaufgabenhäufigkeit

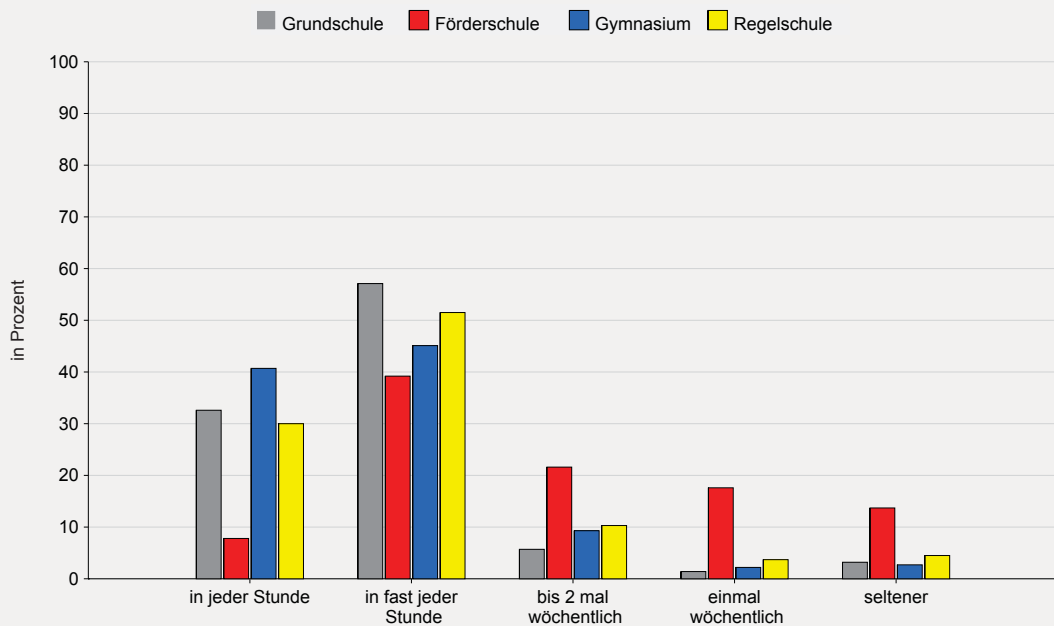

Kompetenztest.de
Berichtswesen Thüringer Kompetenztests 2008. (c) Projekt kompetenztest.de

Abbildung 7.6: Häufigkeit der erteilten Hausaufgaben im Mathematikunterricht

Landesbericht Thüringer Kompetenztests 2008

Funktionen der Hausaufgaben

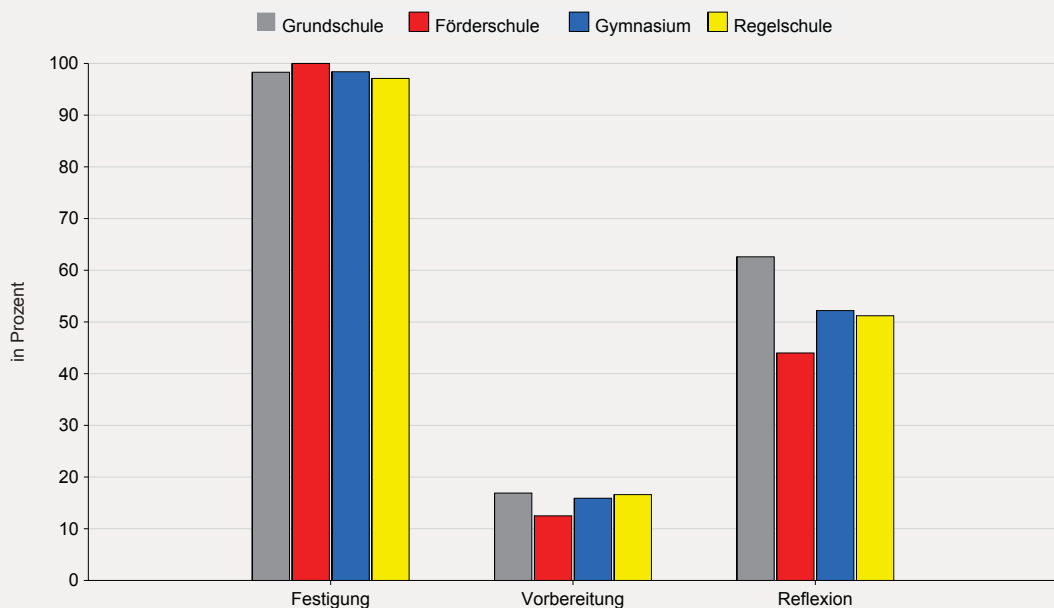

Kompetenztest.de
Berichtswesen Thüringer Kompetenztests 2008. (c) Projekt kompetenztest.de

Abbildung 7.7: Häufige oder sehr häufige Funktionen der Hausaufgaben

Der Abbildung 7.7 ist zu entnehmen, dass schulartübergreifend die Festigung der vermittelten Inhalte der Hauptgrund für das Aufgeben von Hausaufgaben ist. An zweiter Stelle steht dann die Reflexion über vorausgegangenen Unterricht. Eine untergeordnete Rolle spielt der Einsatz der Hausaufgaben zur Vorbereitung des folgenden Unterrichts. Diese Zahlen sind außerdem in allen Schulformen gegenüber dem letzten Jahr leicht rückläufig.

In Bezug auf die Auswertung der Hausaufgaben ergeben sich kaum Unterschiede zum letzten Befragungszeitraum. In erster Linie werden die Hausaufgaben mit den Schülern durchgesprochen beziehungsweise die Schüler können Fragen zu den Hausaufgaben stellen (siehe Tabellen 7.1 und 7.2). Etwa 1/4 der Lehrkräfte prüfen die Hausaufgaben häufig in nachfolgenden Kontrollen. Bis zu 18 % der Lehrer führen häufig mündliche oder schriftliche Leistungskontrollen zu den Hausaufgaben durch. In den einzelnen Schulformen sind diese Werte, bis auf den Förderschulbereich, leicht rückläufig. Dies ist positiv zu werten, da die Hausaufgaben die Funktionen des Übens und selbstständigen Handelns anregen sollen und nicht der Kontrolle dienen sollten.

Tabelle 7.1: In welcher Form werten Sie Hausaufgaben aus?

	Ich spreche sie mit den Schülern durch	Ich gebe Gelegenheit Fragen zu den HA zu stellen	Ich prüfe die HA in nachfolgenden Kontrollen ab
Grundschule	89,6 %	86,6 %	27,2 %
Förderschule	98,0 %	82,4 %	24,0 %
Gymnasium	96,1 %	95,0 %	18,3 %
Regelschule	93,8 %	93,0 %	26,3 %

Tabelle 7.2: In welcher Form werten Sie Hausaufgaben aus?

	Ich sammle schriftlich zu erledigende HA ein	Ich führe mündliche od. schriftliche Leistungs- kontrollen zu den HA durch
Grundschule	15,2 %	5,7 %
Förderschule	14,0 %	18,4 %
Gymnasium	4,4 %	10,0 %
Regelschule	8,8 %	15,9 %

Darstellung ausgewählter Unterrichtsaspekte in Ihrem Einfluss auf die durchschnittliche Testleistung der Schüler

Mit Hilfe von Regressionsmodellen wurde untersucht, ob und inwiefern bestimmte Unterrichtsmerkmale mit der Testleistung der Schüler im Fach Mathematik einhergehen. Dafür wurden die Einzelergebnisse der Schüler pro Klasse zu einem Klassenwert gemittelt. Da es drei verschiedene Versionen der Lehrerfragebögen gab (zur Aufwandsreduktion für den einzelnen Lehrer), liegen die Informationen zur Unterrichtsgestaltung immer nur anteilig vor und werden daher getrennt nach Merkmalsgruppen betrachtet. Folgende Merkmalsgruppen wurden hinsichtlich ihrer Bedeutung für die Testergebnisse auf Klassenebene hin untersucht: „Unterrichtsformate“ (z. B. Unterrichts- und Sozialformen), „Hausaufgabengestaltung“ (Häufigkeit, Auswertung) sowie der Einsatz von „Unterrichtsmedien“ (z. B. Flipchart). Die Ergebnisse werden so berichtet, dass jeweils die Ausprägung eines Unterrichtsmerkmals angegeben wird, die mit hohen Testergebnissen einhergeht.

Wie gut Unterschiede in der Testleistung von Klassen mit Unterschieden in der jeweiligen Unterrichtsgestaltung beschrieben werden können, lässt sich am Anteil aufgeklärter Varianz (=Varianzaufklärung) ablesen.

Wichtig ist es zu betonen, dass die im Folgenden berichteten Ergebnisse der Regressionsanalysen lediglich als Orientierung dienen können, da viele mögliche Störeinflüsse nicht kontrolliert werden können. Die notwendige Mittelung der Schülerleistungen zu einem Gesamtwert einer Klasse vernachlässigt beispielsweise Einflussfaktoren seitens der Schüler (z. B. Motivation, Fähigkeit). Zudem können die Zusammenhänge zwischen Unterrichtsmerkmalen und Testleistung aufgrund der Stichprobengröße nicht schulartspezifisch betrachtet werden, somit bleiben Einflüsse der Schulart unberücksichtigt. Auch sollten die Ergebnisse aufgrund der systematisch fehlenden Werte vorsichtig interpretiert werden. Nicht zuletzt muss natürlich festgehalten werden, dass nur Merkmale hinsichtlich möglicher Zusammenhänge zur Testleistung untersucht werden können, die auch im Fragebogen erhoben wurden.

Nachfolgend werden für die drei Klassenstufen getrennt die Ergebnisse der Regressionsanalysen berichtet.

In den dritten Klassen kann mit Hilfe der Unterrichtsmerkmale vergleichsweise recht wenig an der Testleistung erklärt werden. Die relativ höchste Aufklärung von 5,6 % wird mit folgenden Merkmalen der Gruppe „Unterrichtsmedien“ erzielt: (häufiger) Einsatz des Rechenprogramms (so vorhanden) sowie von OHP-Folien. Das bedeutet, höhere Testergebnisse werden in Klassen beobachtet, deren Lehrer angaben, oft mit dem Rechenprogramm bzw. mit dem Overheadprojektor zu arbeiten.

Merkmale der Gruppe „Unterrichtsformate“ klären lediglich 2,3 % der Testleistungsvarianz auf, während Merkmale der Hausaufgabengestaltung zusammen betrachtet keine signifikanten Beziehungen zum Testergebnis aufweisen.

Eine Verallgemeinerung von Vorhersagemustern über die sechsten und achten Klassen ist nur bedingt möglich, daher werden die Vorhersagemodelle getrennt für beide Klassenstufen betrachtet.

In den sechsten Klassen erzielen Merkmale der Gruppe „Unterrichtsmedien“ eine vergleichsweise hohe Varianzaufklärung von 18,1 %. Der stärkste Zusammenhang wird hier zum Einsatz des Rechenprogramms (so vorhanden) beobachtet. Entgegen der gefundenen Zusammenhänge in den dritten Klassen, werden höhere Testergebnisse eher in sechsten Klassen beobachtet, deren Lehrer angaben, das Rechenprogramm vergleichsweise selten einzusetzen. Darüber hinaus zeigen der (weniger häufige) Einsatz der Magnettafel und der (häufige) Einsatz des Mathebuches Zusammenhänge zur Testleistung.

Merkmale der Gruppe „Unterrichtsformate“ erzielen eine Varianzaufklärung von 7,5 %, hier steht insbesondere der (weniger häufige) Einsatz von Stationenlernen, (häufige) fächerübergreifende Projektarbeit und das (häufige) Reflektieren des Unterrichts mit den Schülern in Beziehung zum Testergebnis. Mit Merkmalen der „Hausaufgabengestaltung“ lässt sich das beobachtete Klassenergebnis im Test wiederum kaum erklären.

In den achten Klassen finden sich dagegen stärkere Zusammenhänge zwischen dem Testergebnis und der „Hausaufgabengestaltung“ (Varianzaufklärung von 11,8 %). Hohe Testergebnisse gehen insbesondere einher mit (weniger häufigem) Nachprüfen der Hausaufgaben in Form von Leistungskontrollen und (weniger häufigem) Einsammeln schriftlicher Hausaufgaben sowie dem (häufigen) Aufgeben fakultativer Hausaufgaben. Darüber hinaus erzielen die folgenden Merkmale der Gruppe „Unterrichtsmedien“ eine Varianzaufklärung von 9 % an den Testergebnissen: (weniger häufiger) Einsatz vorgedruckter Arbeitsblätter sowie (häufiger) Einsatz des Tafelbildes.

An dieser Stelle ist noch einmal zu betonen, dass diese Ergebnisse nicht bedeuten, dass andere Variablen keinerlei Einfluss auf die Schülerleistung haben. Vielmehr ist zu vermuten, dass die Effekte dieser Variablen über andere Variablen vermittelt werden. Zudem sollten die beobachteten Zusammenhänge zwischen Unterrichtsmerkmalen und Testergebnissen nicht kausal interpretiert werden (im Sinne einer Vorhersage), da offen bleibt, was Ursache und was Wirkung ist. So kann ein positiver Zusammenhang zwischen dem Einsatz des Rechenprogramms und den Testergebnissen einerseits bedeuten, dass sich ein häufiger Einsatz des Rechenprogramms tatsächlich förderlich auf die Leistung der Schüler auswirkt. Andererseits kann der Zusammenhang auch einfach aufzeigen, dass Lehrer in Klassen mit hohem Leistungsdurchschnitt in Mathematik öfter das Rechenprogramm einsetzen.

7.1.2 Deutsch

Fragenkomplex: „Einsatz der neuen Medien“

Ein großer Teil der Grund- und Förderschullehrer verwendet die neuen Medien regelmäßig im Unterricht (siehe Abbildung 7.8). In diesen beiden Schulformen werden die neuen Medien wesentlich häufiger eingesetzt als im Gymnasium oder der Regelschule. Immerhin 23,4 % der Gymnasial- und 12,8 % der Regelschullehrer geben an, nie diese Medien im Unterricht einzusetzen. Im Grundschulbereich beträgt der Anteil der Lehrer, die nie neue Medien einsetzen, nur 6,5 % und im Förderschulbereich 4,2 %, wobei auf Grund der wenigen beteiligten Förderschullehrer diese 4,2 % genau einem Lehrer entsprechen. Auch setzen mit ca. 44 % wesentlich mehr Gymnasial- und Regelschullehrer als Grundschullehrer die neuen Medien seltener als ein Mal im Monat ein.

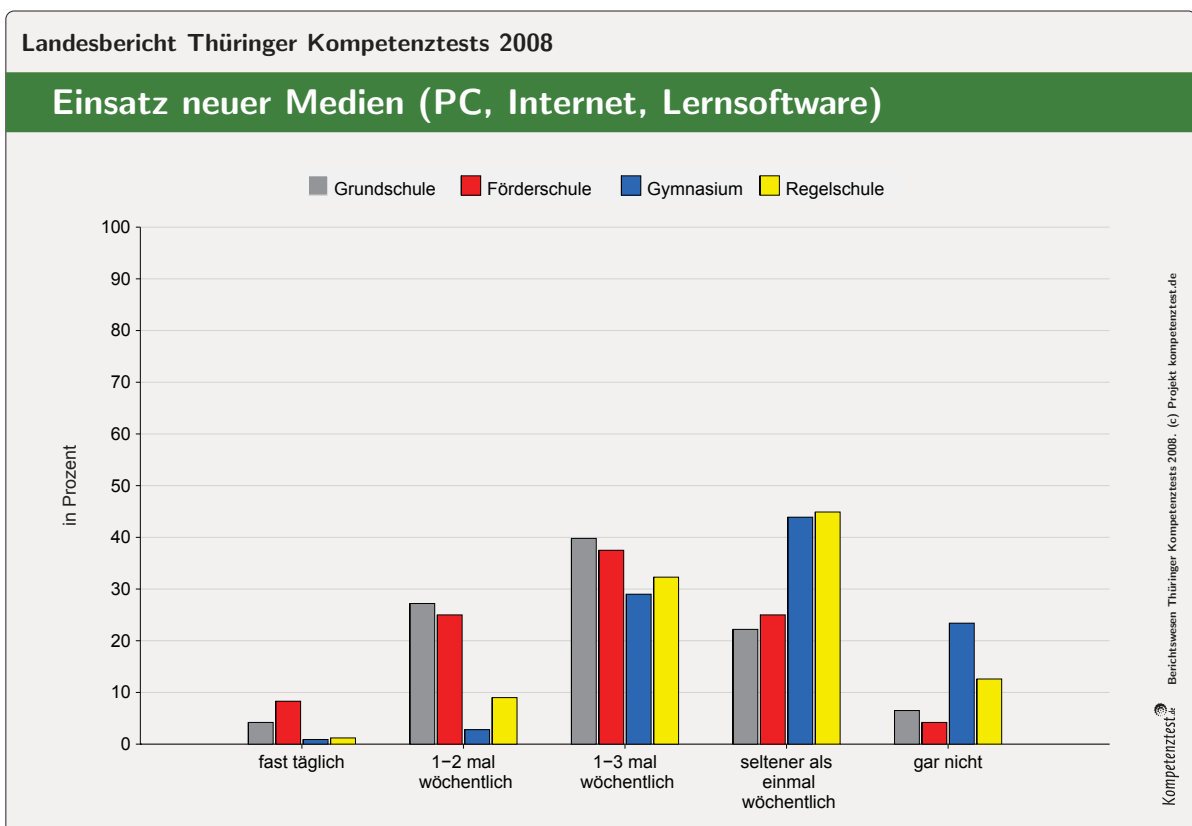


Abbildung 7.8: Setzen Sie die neuen Medien (PC, Internet, Lernsoftware) in Ihrem Deutschunterricht ein?

Fragenkomplex: „Fachkultur und methodische Gestaltung des Unterrichts“

Arbeit mit Medien im Deutschunterricht

In den Lehrerfragebögen wurden die Deutschlehrer über den Einsatz eines breiten Spektrums an Materialien befragt, welche sie im Unterricht einsetzen. Da sich alle Materialien entsprechend der Unterrichtsthematik für den Einsatz eignen, steht für die Nutzung der jeweiligen Medien nur ein begrenzter Zeitrahmen zur Verfügung. So verwundert es nicht, dass von 80 % bis 90 % der Lehrkräfte angegeben wird, dass sie die folgenden Materialien in manchen Unterrichtseinheiten einsetzen: aktuelle Zeitungen, Zeitschriften und Bücher, die die Schüler in ihrer Freizeit lesen. Häufiger wird das Lesebuch und das Sprachbuch eingesetzt. In der Grundschule spielen außerdem Arbeitshefte eine große Rolle. In den anderen Unterrichtsformen variiert der Einsatz recht stark zwischen nie und in jeder Stunde.

Sozialformen

Die am meisten praktizierte Sozialform ist wie in den vergangenen Jahren die Einzelarbeit der Schüler (siehe Abbildung 7.9). So geben ca. 90 % der Lehrkräfte aller Schulformen an, dass sie häufig bzw. sehr häufig diese Form anwenden. Auch Partnerarbeit spielt im Unterricht eine große Rolle. Im Grundschulbereich wird diese Sozialform von 89 % der Lehrkräfte häufig beziehungsweise sehr häufig eingesetzt. 66 % der befragten Regelschullehrer geben an, dass sie diese Form häufig oder sehr häufig anwenden. Die anderen beiden Schulformen liegen dazwischen. Nicht ganz so oft wird Gruppenarbeit im Unterricht eingesetzt. Hier geben über 90 % der Lehrkräfte aller Schulformen an, dass sie nur manchmal beziehungsweise häufig die Schüler in Gruppen arbeiten lassen.

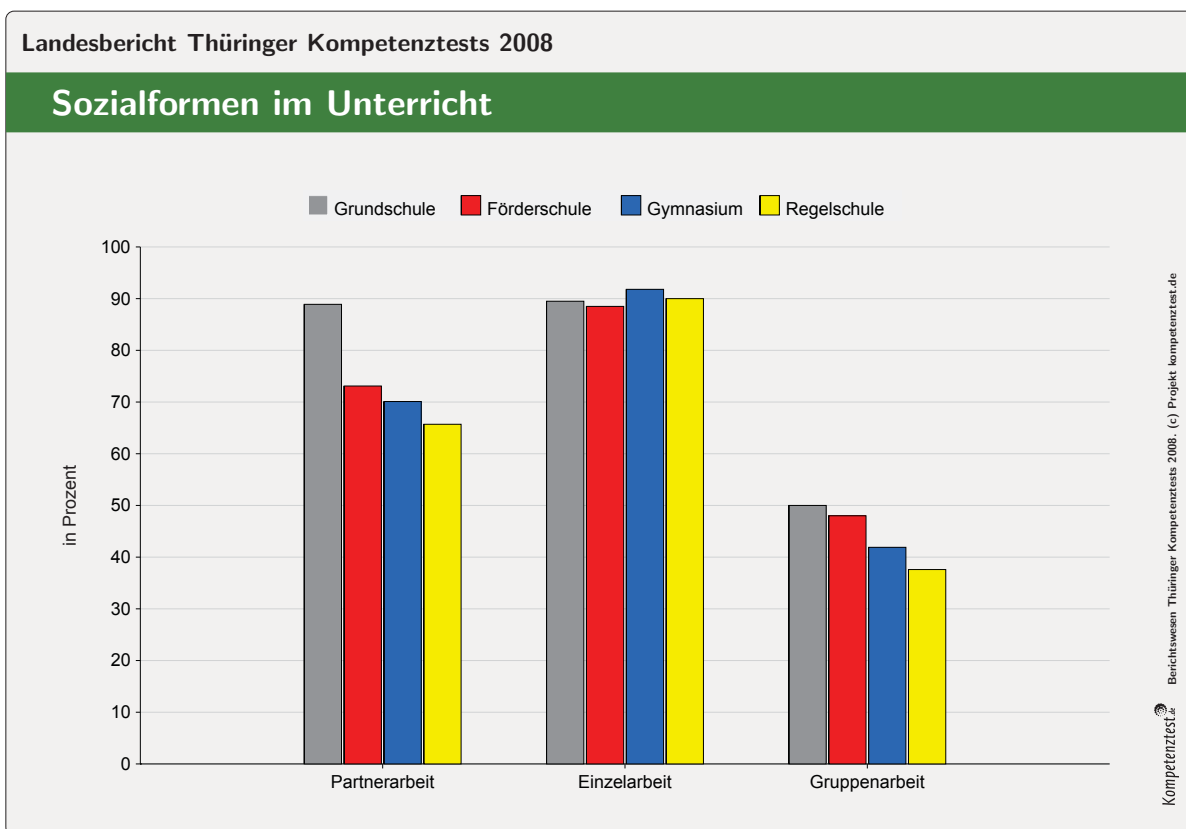


Abbildung 7.9: Folgende Sozialformen werden sehr häufig oder häufig im Unterricht eingesetzt

Unterrichtsformen

Der Abbildung 7.10 ist zu entnehmen, dass die Deutschlehrer auch in diesem Schuljahr den Frontalunterricht bevorzugten. Obwohl im Vergleich zum letzten Schuljahr die Werte leicht rückläufig sind. Der Anteil der Lehrer, die angeben, dass sie den Frontalunterricht sehr häufig bzw. häufig anwenden, beträgt im Gymnasium und der Regelschule 84 %, in der Grund- und Förderschule ca. 72 %. Nur sehr wenige Lehrer geben an, Frontalunterricht selten einzusetzen. Ergänzt wird der Frontalunterricht besonders im Grund- und Förderschulbereich durch Stationenlernen, Wochenplanarbeit, Tagesplanarbeit und offene Unterrichtsformen.

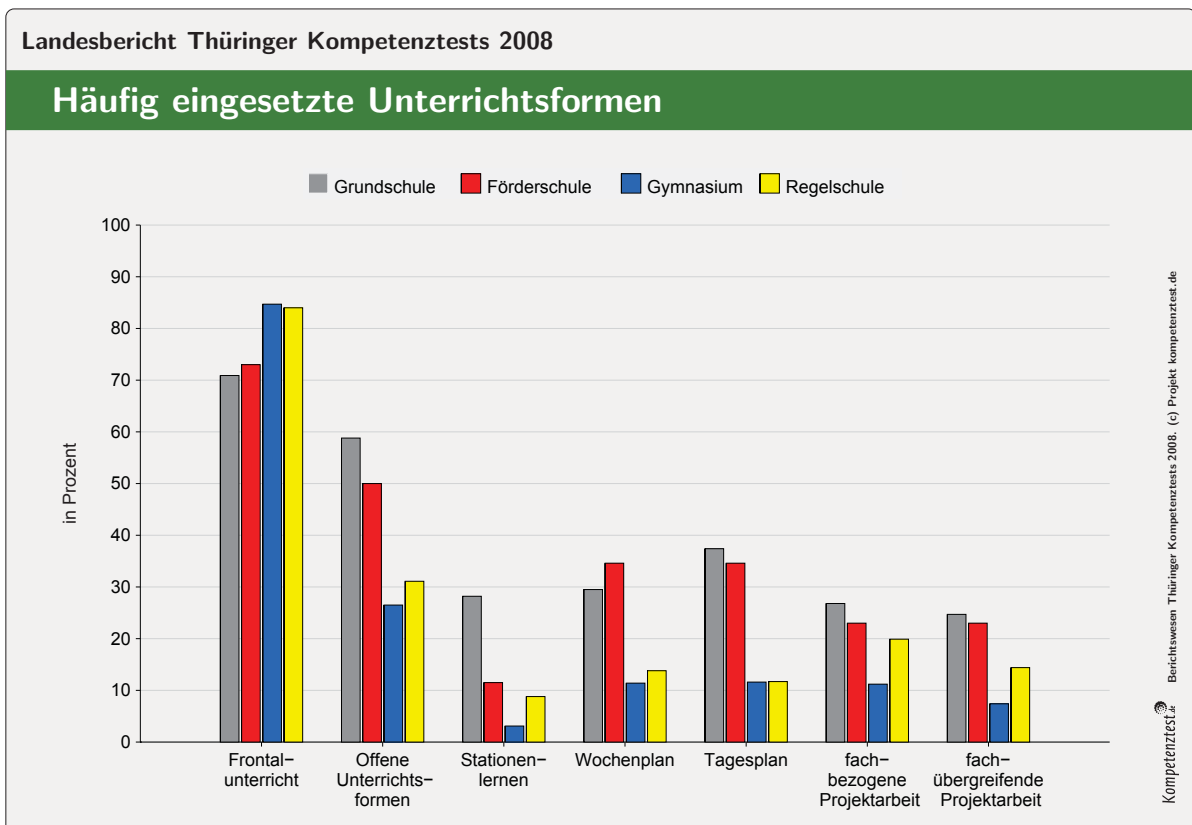


Abbildung 7.10: Folgende Unterrichtsformen werden in jeder oder in fast jeder Stunde im Unterricht eingesetzt

Schülerbeteiligung an der Unterrichtsstruktur

Auf die Frage: „Wie oft beteiligen Sie die Schüler an der Planung Ihres Unterrichts?“ geben über 90 % der Lehrkräfte aller Schulformen an, dass sie die Schüler innerhalb mancher Unterrichtseinheiten beziehungsweise innerhalb mancher Unterrichtsstunden an der Planung beteiligen (siehe Abbildung 7.11). Der Anteil der Lehrer, die ihre Schüler nie an der Planung des Unterrichts beteiligen, ist in der Regelschule mit 8,1 % am größten. Sehr klein ist dagegen die Zahl der Lehrer, die ihre Schüler in jeder Unterrichtsstunde an der Planung beteiligen (nur unter 2 % der Lehrer).

Beim Vergleich der Abbildungen 7.11 und 7.12 lässt sich erkennen, dass die Reflexion des Unterrichts häufiger als die Beteiligung der Schüler an der Planung stattfindet. Hier reflektieren zwischen 11 % und 25 % der Lehrer innerhalb jeder Unterrichtsstunde, wobei den größten Anteil die Förderschullehrer und den geringsten die Gymnasiallehrer ausmachen. Die Mehrzahl der Lehrer (50 % bis 58 %) reflektieren ihren Unterricht mit den Schülern innerhalb mancher Unterrichtsstunden.

Landesbericht Thüringer Kompetenztests 2008

Schülerbeteiligung an Unterrichtsplanung

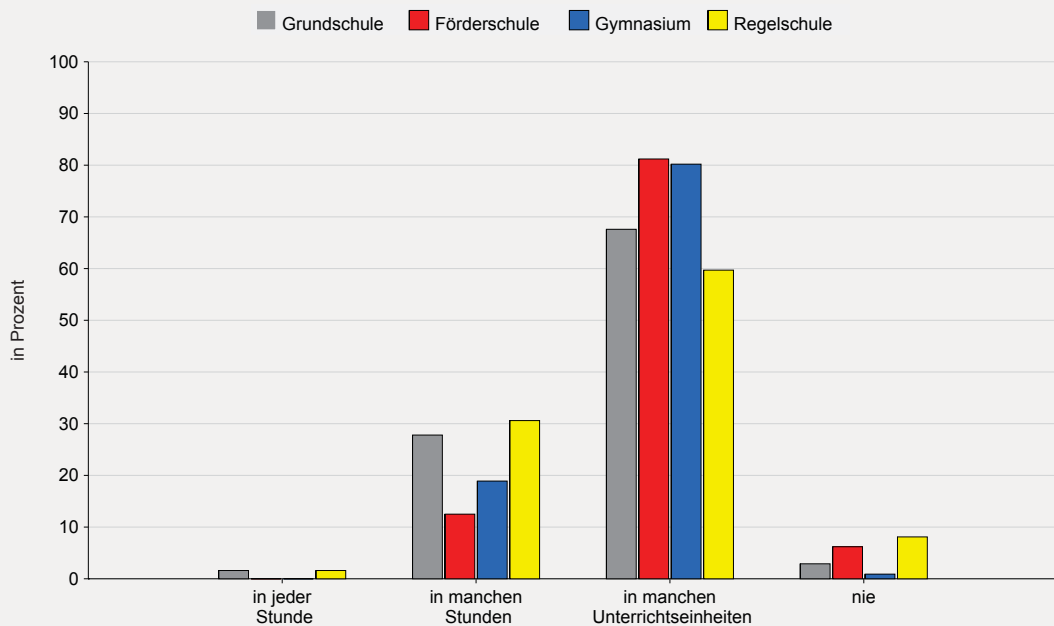

Kompetenztest.de
Berichtswesen Thüringer Kompetenztests 2008. (c) Projekt kompetenztest.de

Abbildung 7.11: Wie häufig beteiligen Sie die Schüler an der Planung des Unterrichts?

Landesbericht Thüringer Kompetenztests 2008

Unterrichtsreflektion

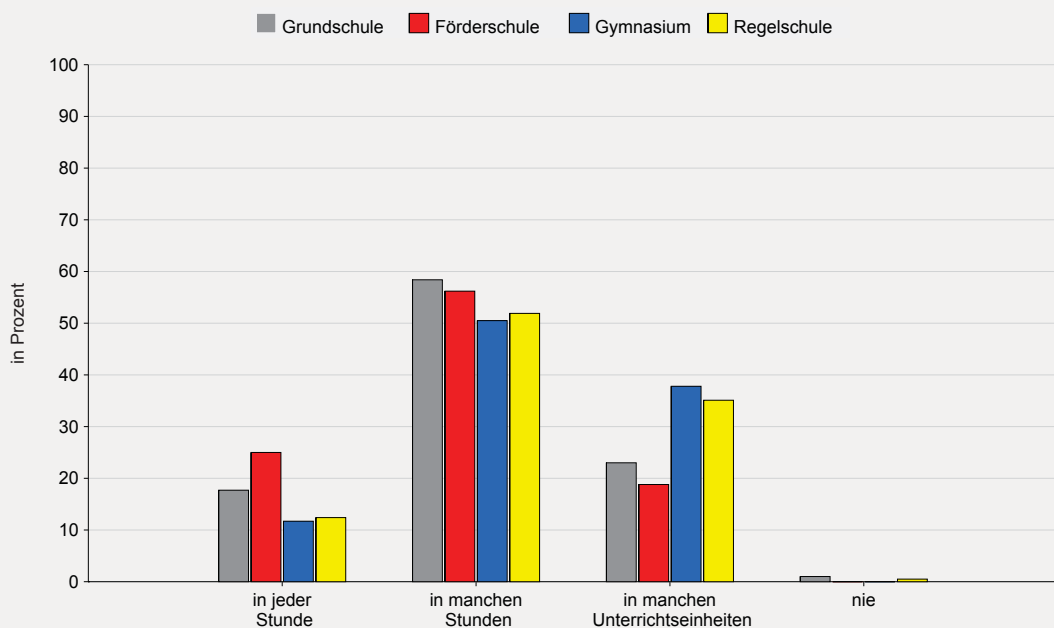

Kompetenztest.de
Berichtswesen Thüringer Kompetenztests 2008. (c) Projekt kompetenztest.de

Abbildung 7.12: Wie häufig reflektieren Sie mit Ihren Schülern Ihren Unterricht?

Einsatz von Hausaufgaben

In diesem Jahr wurden wiederum die Häufigkeit (Abbildung 7.13), die Art und die Funktion (Abbildung 7.14) von Hausaufgaben im Deutschunterricht erfragt. Hierbei gliederte sich der Itemblock in einen ersten Teil, der die Stellung der Hausaufgaben und in einen zweiten, der die Art ihrer Auswertung eruieren sollte. Dabei gaben 76 % der Gymnasiallehrer, 70 % der Regelschullehrer, 88,3 % der Grundschullehrer und 21,7 % der Förderschullehrer an, in jeder bzw. fast jeder Unterrichtsstunde Hausaufgaben zu stellen. Damit sind diese Werte

bis auf den Grundschulbereich etwas niedriger als im letzten Schuljahr. Trotzdem wird der hohe Stellenwert der Hausaufgaben auch in diesem Schuljahr deutlich. Neben diesem Faktum ist von Interesse, welche Funktion die Hausaufgaben im jeweiligen Unterricht einnehmen.

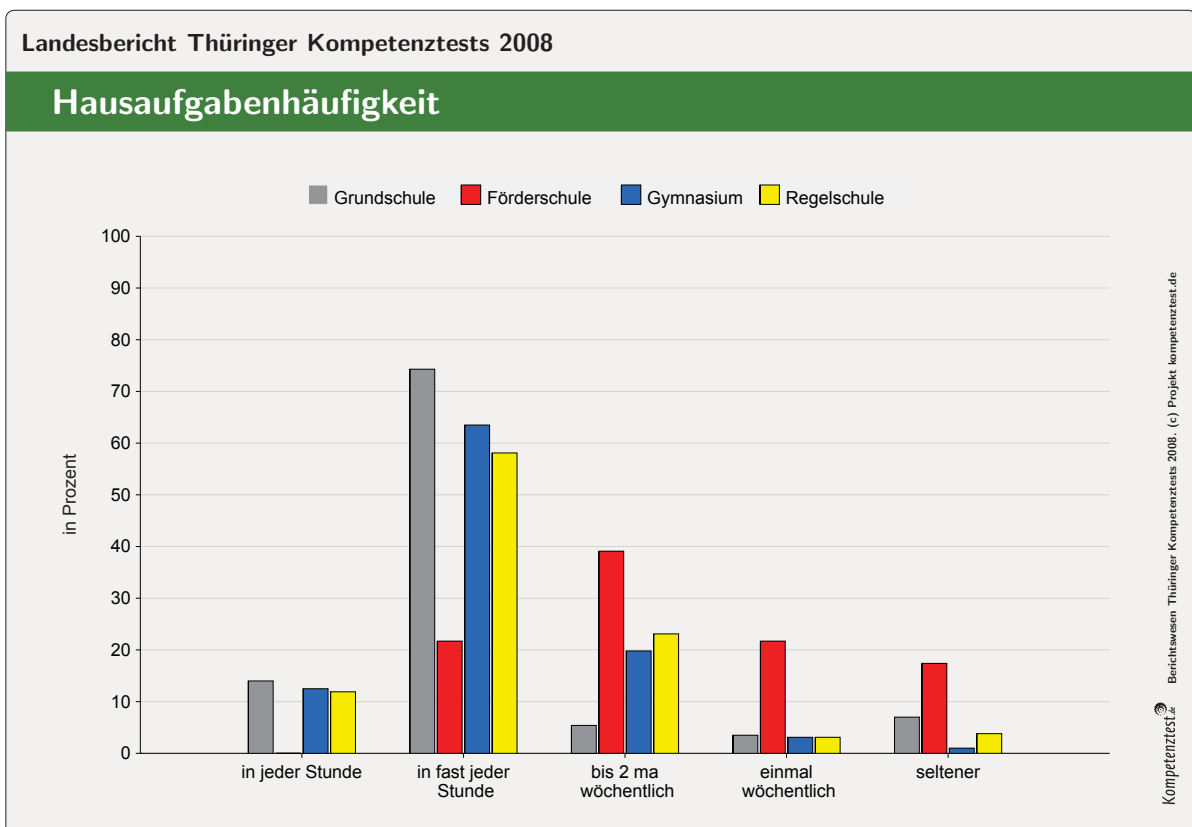


Abbildung 7.13: Häufigkeit der erteilten Hausaufgaben im Deutschunterricht

Mit den erhobenen Daten wird auch in diesem Jahr deutlich, dass schulartübergreifend die Festigung und Wiederholung der im Unterricht vermittelten Lerngegenstände und Inhalte die primäre Funktion der Hausaufgaben einnimmt. Diese Aussage korrespondiert in starkem Maße mit dem Wert für die Funktion der Hausaufgabe, Unterricht zu reflektieren. Dieser Wert schließt das Moment der Wiederholung ein. Somit kann übereinstimmend das Fazit gezogen werden, dass die Hauptfunktion aufgebener Hausaufgaben in Übungen methodischer und sachlicher Inhalte des Unterrichts besteht, die das Ziel die Wiederholung und Festigung aufweisen. Vorbereitende Hausaufgaben, das heißt die selbstständige Erarbeitung von Lerninhalten, rangiert dieser Funktion deutlich nachgeordnet. Trotzdem geben über 80 % der Grund-, Gymnasial- und Regelschullehrer an, zumindest manchmal die Hausaufgaben zur Vorbereitung des folgenden Unterrichts einzusetzen. Weiterhin wurden die Lehrer auch danach befragt, wie sie die Hausaufgaben im Unterricht auswerten. Hier gibt der größte Teil der Lehrkräfte an, die Aufgaben mit den Schülern durchzusprechen und Fragen zu den Hausaufgaben zu beantworten. Nur manchmal beziehungsweise selten werden die Hausaufgaben auch in nachfolgenden Kontrollen überprüft oder die schriftlich zu erledigenden Hausaufgaben werden eingesammelt.

Landesbericht Thüringer Kompetenztests 2008

Funktionen der Hausaufgaben

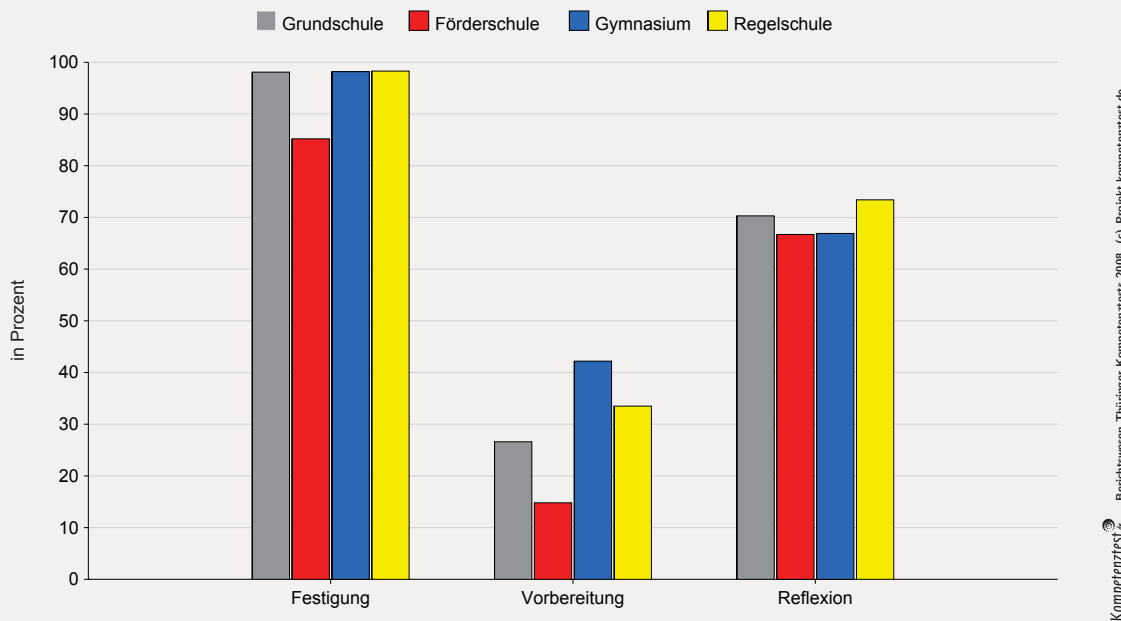


Abbildung 7.14: Häufige oder sehr häufige Funktionen der Hausaufgaben

Darstellung ausgewählter Unterrichtsaspekte in ihrem Einfluss auf die durchschnittliche Testleistung der Schüler einer Klasse

Analog zum Vorgehen für das Fach Mathematik wurde wieder mit Hilfe von Regressionsmodellen untersucht, ob und inwiefern bestimmte Unterrichtsmerkmale mit der Testleistung der Schüler im Fach Deutsch einhergehen. Dafür wurden wieder die Einzelergebnisse der Schüler pro Klasse zu einem Klassenwert gemittelt. Die Informationen zur Unterrichtsgestaltung liegen für jede Klasse immer nur anteilig vor (Aufwandsreduktion Lehrerfragebogen) und werden daher getrennt nach Merkmalsgruppen betrachtet. Folgende Merkmalsgruppen wurden hinsichtlich ihrer Bedeutung für die Testergebnisse auf Klassenebene hin untersucht: „Unterrichtsformate“ (z. B. Unterrichts- und Sozialformen), „Inhaltliche Schwerpunkte“ (Lesen, Schreiben), „Hausaufgabengestaltung“ (Häufigkeit, Auswertung), „Unterrichtsmaterialien“ (Lesebuch, Zeitschriften) sowie der Einsatz von „Unterrichtsmedien“ (z. B. Flipchart). Die Ergebnisse werden so berichtet, dass jeweils die Ausprägung eines Unterrichtsmerkmals angegeben wird, die mit hohen Testergebnissen in Deutsch einhergeht.

Wie gut Unterschiede in der Testleistung von Klassen mit Unterschieden in der jeweiligen Unterrichtsgestaltung beschrieben werden können, lässt sich am Anteil aufgeklärter Varianz (=Varianzaufklärung) ablesen.

Auch an dieser Stelle ist es wichtig, noch mal zu betonen, dass die im Folgenden berichteten Ergebnisse lediglich als Orientierung dienen können, da viele mögliche Störeinflüsse nicht kontrolliert werden können. Die notwendige Mittelung der Schülerleistungen zu einem Gesamtwert einer Klasse vernachlässigt beispielsweise Einflussfaktoren seitens der Schüler (z. B. Motivation, Fähigkeit). Zudem können die Zusammenhänge zwischen Unterrichtsmerkmalen und Testleistung aufgrund der Stichprobengröße nicht schulartspezifisch betrachtet werden, somit bleiben Einflüsse der Schulart unberücksichtigt. Auch sollten die Ergebnisse aufgrund der sys-

tematisch fehlenden Werte vorsichtig interpretiert werden. Nicht zuletzt muss natürlich festgehalten werden, dass nur Merkmale hinsichtlich möglicher Zusammenhänge zur Testleistung untersucht werden können, die auch im Fragebogen erhoben wurden.

Nachfolgend werden für die Klassenstufen 3 und 6 getrennt die Ergebnisse der Regressionsanalysen schulartübergreifend berichtet.

In den dritten Klassen wird für folgende Merkmale der Gruppe „Inhaltliche Schwerpunkte“ die vergleichsweise höchste Varianzaufklärung von 8,9 % erzielt: (weniger häufiges) lautes Lesen zur Übermittlung von Informationen sowie (jeweils häufiges) schriftliche Beschreiben und freies Sprechen. Das bedeutet, höhere Testergebnisse in Deutsch werden zum Beispiel in Klassen beobachtet, deren Lehrer angaben, in ihrem Unterricht seltener laut Lesen zu lassen.

Aus der Gruppe der „Unterrichtsmedien“ geht lediglich der (häufige) Einsatz der Flip-Chart mit einer höheren Testleistung einher, was einer Varianzaufklärung von 4,7 % entspricht. Mit „Unterrichtsmaterialien“ (häufiges Lesen von Büchern in der Freizeit, häufiger Einsatz von Fernsehsendungen) kann nur ein geringfügiger Anteil der Varianz (3,5 %) aufgeklärt werden, während Merkmale der „Hausaufgabengestaltung“ keine bedeutsamen Zusammenhänge zum Testergebnis der Klasse aufweisen.

In den sechsten Klassen wird eine relativ hohe Varianzaufklärung von 19,8 % mit Merkmalen der Gruppe „Inhaltliche Schwerpunkte“ beobachtet. Hierbei gehen insbesondere das (häufige) schriftliche Beschreiben, das (häufige) schriftlich Berichten, das (häufige) mündliche Äußern und Begründen eines Standpunktes sowie das (häufige) Einüben von Lesetechniken mit hohen Testleistungen einher. Die Bedeutung des schriftlichen Beschreibens zeigt sich also stabil über beide Klassenstufen. Mit den folgenden beiden Merkmalen der Gruppe „Unterrichtsmaterialien“ wird eine Varianzaufklärung von 6,6 % erzielt: (häufige) Verwendung von künstlerischen Schülertexten sowie der (weniger häufige) Rückgriff auf in der Freizeit gelesene Zeitschriften. Merkmale der Gruppe „Hausaufgabengestaltung“ liefern eine Varianzaufklärung von 6 %. Signifikante Zusammenhänge zu (hohen) Testergebnissen zeigen das (weniger häufige) Aufgeben von Hausaufgaben und eine (häufige) Nutzung der aufgegebenen Hausaufgaben zur Vorbereitung der nächsten Unterrichtsstunde. Schließlich werden durch die (häufige) Verwendung von OHP-Folien sowie den (weniger häufigen) Rückgriff auf das ggf. vorhandene Rechtschreibprogramm 5,6 % der Testleistungsvarianz aufgeklärt.

Wie bereits für Mathematik beschrieben, muss an dieser Stelle noch einmal betont werden, dass die Ergebnisse nicht bedeuten, dass andere Variablen keinerlei Einfluss auf die Schülerleistung in Deutsch haben. Es ist vielmehr zu vermuten, dass die Effekte mancher Merkmale über andere Merkmale vermittelt werden. Zudem sollten die beobachteten Zusammenhänge zwischen Unterrichtsmerkmalen und Testergebnissen nicht kausal interpretiert werden (im Sinne einer Vorhersage), da offen bleibt, was Ursache und was Wirkung ist. So kann ein positiver Zusammenhang zwischen dem Einsatz des Rechtschreibprogramms und den Testergebnissen einerseits bedeuten, dass sich ein häufiger Einsatz des Rechtschreibprogramms tatsächlich förderlich auf die Leistung der Schüler auswirkt. Andererseits kann der Zusammenhang auch einfach aufzeigen, dass Lehrer in Klassen mit hohem Leistungsdurchschnitt in Deutsch öfter das Rechtschreibprogramm einsetzen.

7.1.3 Englisch

Zum zweiten Mal erhielten in diesem Jahr auch die Englischlehrer der Klassen 6 Fragebögen zur Unterrichtsgestaltung. Die wichtigsten Ergebnisse dieser Befragung sind nachfolgend dargestellt. Dabei werden Regelschulen und Gymnasien im Vergleich zueinander und separat ausgewertet.

Fragenkomplex: „Einsatz der neuen Medien“

Aus Abbildung 7.15 ist ersichtlich, dass der überwiegende Teil der befragten Lehrkräfte der Gymnasien und Regelschulen die neuen Medien kontinuierlich im Unterricht einsetzt. Dabei gaben 37,1 % der Regelschullehrer und 32 % der Gymnasiallehrer an, die neuen Medien mindestens ein Mal monatlich im Englischunterricht einzusetzen. Hierbei zeigt sich eine Verbesserung zum letzten Schuljahr (31 % und 18 %). Allerdings gab auch ein Viertel der Gymnasiallehrer an, die neuen Medien nie im Unterricht einzusetzen.

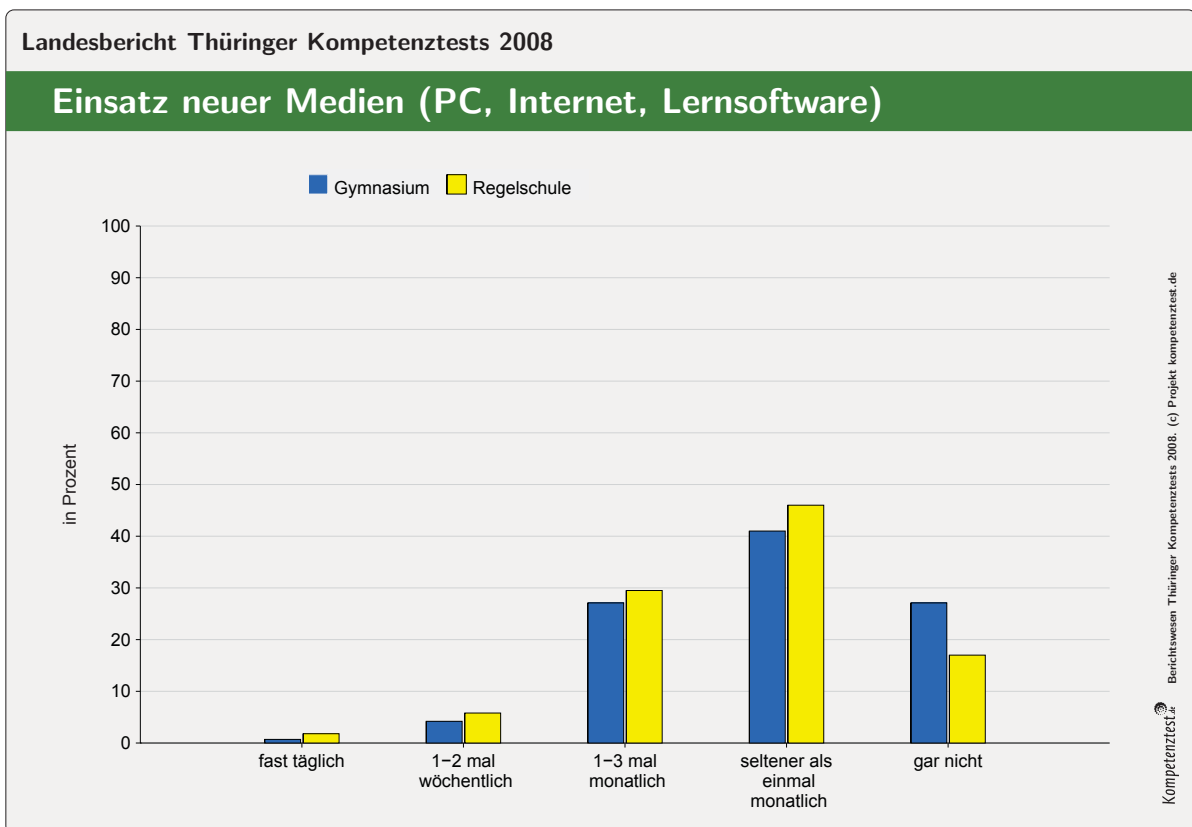


Abbildung 7.15: Setzen Sie die neuen Medien (PC, Internet, Lernsoftware) in Ihrem Englischunterricht ein?

Fragenkomplex: „Fachkultur und methodische Gestaltung des Unterrichts“

Arbeit mit Medien im Englischunterricht

Hinsichtlich der anderen Medien wurden die Lehrer ebenfalls befragt, wie häufig ein bestimmtes Medium im Englischunterricht eingesetzt wird. Einen ersten Überblick veranschaulicht die Abbildung 7.16.

Landesbericht Thüringer Kompetenztests 2008

Medieneinsatz

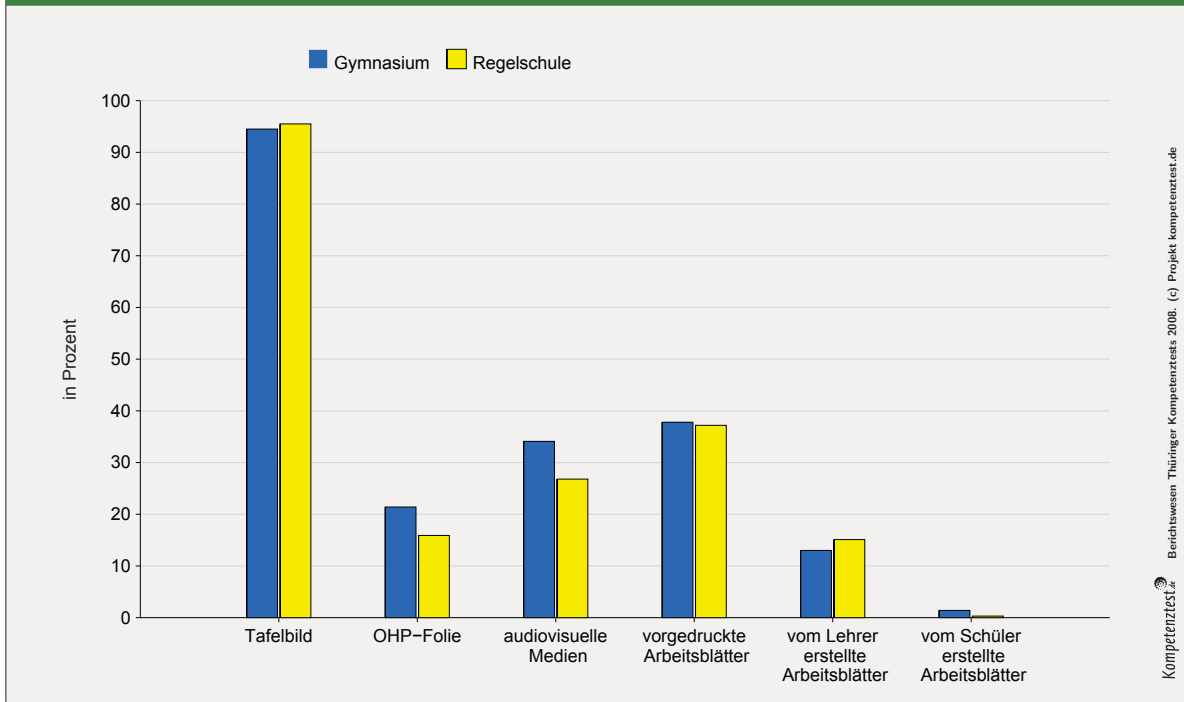


Abbildung 7.16: Folgende Medien werden in jeder Stunde oder in fast jeder Stunde im Unterricht eingesetzt

Zunächst wird mit diesen Werten deutlich, dass die traditionelle Tafel das Medium ist, das den Unterricht am deutlichsten prägt. Da es sich sowohl bei der Tafelarbeit als auch bei der Arbeit mit der OHP-Folie um Medien handelt, die vorwiegend im Frontalunterricht und einheitlich für alle Schüler einer Klasse angewandt werden, kann unter Beachtung der Werte, die auf die Frage nach den Sozialformen des Unterrichts vorliegen, gefolgert werden, dass der Unterricht schulartübergreifend in hohem Maße von Lehrerlenkung und für die Klasse einheitlichen Aspekten geprägt ist. Audiovisuelle Medien werden im Unterricht aller Schularten im Vergleich zu den Fächern Deutsch und Mathematik relativ häufig eingesetzt. Auch die Arbeitsblätter, die auf die Erarbeitung der Inhalte des Unterrichts durch Einzel- oder kooperatives

Lernen abzielen, nehmen als Medien einen vergleichsweise hohen Stellenwert ein. Hierbei rangieren schulartübergreifend wiederum die vordruckten, aus Lehrermaterialien entnommenen Arbeitsblätter an vorderster Position. Arbeitsblätter, die von dem jeweiligen Lehrer selbst erstellt wurden und damit gezielt auf die individuelle Klassensituation abgestimmt sind, besitzen demgegenüber eine weitaus geringere Bedeutung. Durch Schüler erstellte Arbeitsmaterialien besitzen einen nicht so großen Stellenwert als Medien im Unterricht, ca. 65 % aller Lehrer setzen diese Medien selten in ihrem Unterricht ein.

Sozialformen

Hinsichtlich der Präferenz bestimmter Sozialformen im Unterricht ergibt sich schulartübergreifend eine eindeutig häufigere Anwendung der Einzelarbeit und der Partnerarbeit gegenüber der Gruppenarbeit.

Tabelle 7.3: Folgende Sozialformen werden sehr häufig und häufig im Englischunterricht eingesetzt

	Partnerarbeit	Einzelarbeit	Gruppenarbeit
Gymnasium	90,9 %	91,7 %	36,8 %
Regelschule	76,6 %	93,9 %	30,9 %

Unterrichtsformen

Ebenso eindeutig wie in Bezug auf die Sozialformen ergibt sich schulartübergreifend eine eindeutige Präferenz für die Unterrichtsform des Frontalunterrichts. Damit besitzt der lehrerzentrierte Frontalunterricht schulartübergreifend nach wie vor die den Unterricht dominierende Bedeutung, wird aber schulartspezifisch durch andere Unterrichtsformen ergänzt. Projektarbeit gestaltet sich vorwiegend innerhalb des Rahmens des Fachunterrichts, nur gelegentlich sind Projektarbeiten fachübergreifend. Hierbei ist feststellbar, dass am Gymnasium wesentlich stärker projektorientiertes Arbeiten vorzufinden ist.

Tabelle 7.4: Folgende Unterrichtsformen werden in jeder Stunde oder in fast jeder Stunde im Unterricht eingesetzt

	Frontal- unterricht	Stationen- lernen	Wochen- plan	Tages- plan
Gymnasium	93,1 %	4,2 %	6,3 %	12,2 %
Regelschule	87,8 %	8,2 %	6,5 %	11,5 %

Tabelle 7.5: Folgende Unterrichtsformen werden in jeder Stunde oder in fast jeder Stunde im Unterricht eingesetzt

	fachbezogene Projektarbeit	fachübergreifende Projektarbeit
Gymnasium	23,1 %	7,0 %
Regelschule	11,8 %	3,1 %

Kommunikationsformen

Bei den am häufigsten eingesetzten Kommunikationsformen im Englischunterricht dominiert neben dem Lehrer-Schüler-Gespräch das Schüler-Schüler-Gespräch (siehe Abbildung 7.17). Damit wird deutlich, dass die kommunikative Ebene im Englischunterricht zur Erreichung des Unterrichtsziels eine bedeutende Rolle spielt.

Kommunikationsformen im Englischunterricht

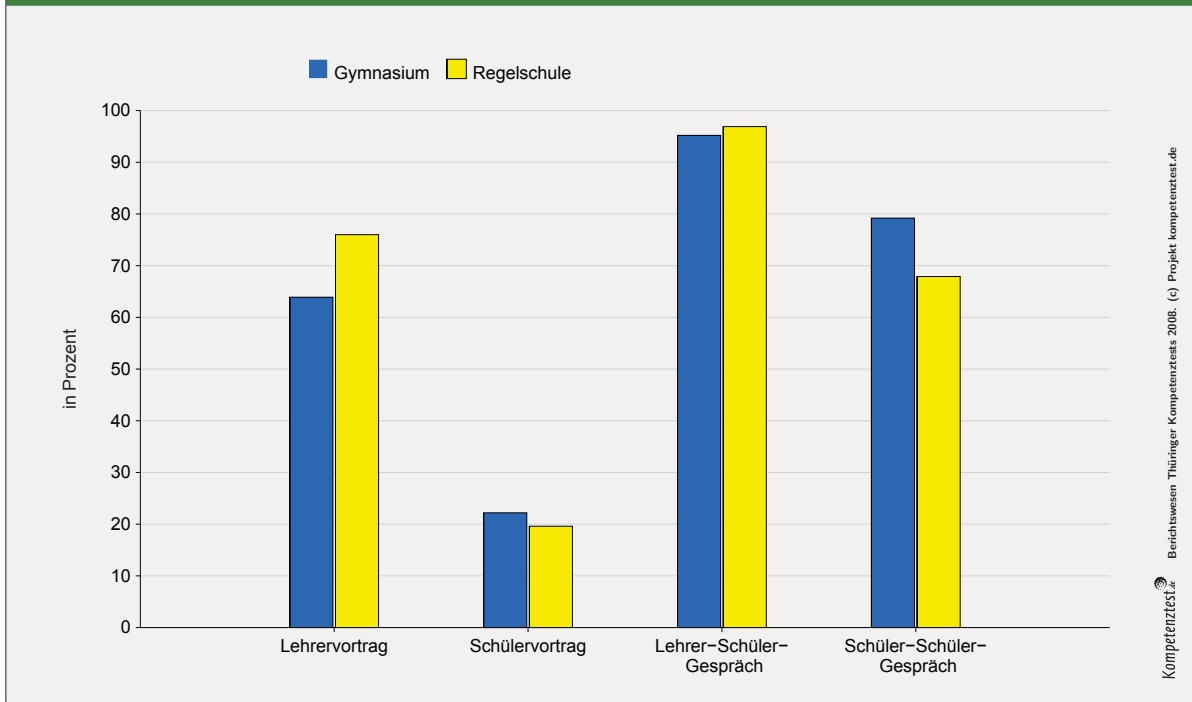


Abbildung 7.17: Folgende Kommunikationsformen werden häufig und sehr häufig im Englischunterricht eingesetzt

Hören als inhaltlicher Aspekt des Unterrichts

Entsprechend der Abbildung 7.18 signalisieren die Werte im Bereich des Hörens und Zuhörens, dass vor allem der Informationsvermittlungs- und -erarbeitungsaspekt im Zentrum des Unterrichts steht. Erstrangig ist dabei das Reagieren auf Aufforderungen, was auf das fremdsprachliche Prinzip, Englisch als Unterrichtssprache, in nahezu allen Unterrichtsstunden schließen lässt. Wenn mit den Schülern Hörübungen absolviert werden, dienen diese vorrangig dem Sinnerfassen und dem Herausarbeiten von Hauptinformationen. Hinsichtlich des Erfassens detaillierter Informationen sind deutlich geringere Werte zu konstatieren. Das Hören in Alltagssituationen ist gegenüber den Informationsvermittlungsaspekten deutlich nachrangig.

Sprechen als inhaltlicher Aspekt des Unterrichts

Im Bereich des Sprechens wird vor allem deutlich, dass die Gesprächsführung und das themenzentrierte Sprechen im Mittelpunkt des Unterrichts stehen (siehe Abbildung 7.19). Das sprachliche monologische Darlegen zusammenhängender Sachverhalte ist dem Üben von Gesprächen mit der typischen Form des Agierens und Reagierens nachrangig. Insgesamt ist feststellbar, dass das Sprechen im Englischunterricht in Klasse 6 am Gymnasium einen höheren Stellenwert einnimmt als an der Regelschule.

Lesen als inhaltlicher Aspekt des Unterrichts

Im Bereich des Lesens wird in Abbildung 7.20 sichtbar, dass die verschiedenen Aspekte des Lesens im Unterricht sehr wichtig sind. Schulartübergreifend spielt dabei das Lesen zur Unterhaltung im Unterricht kaum eine Rolle.

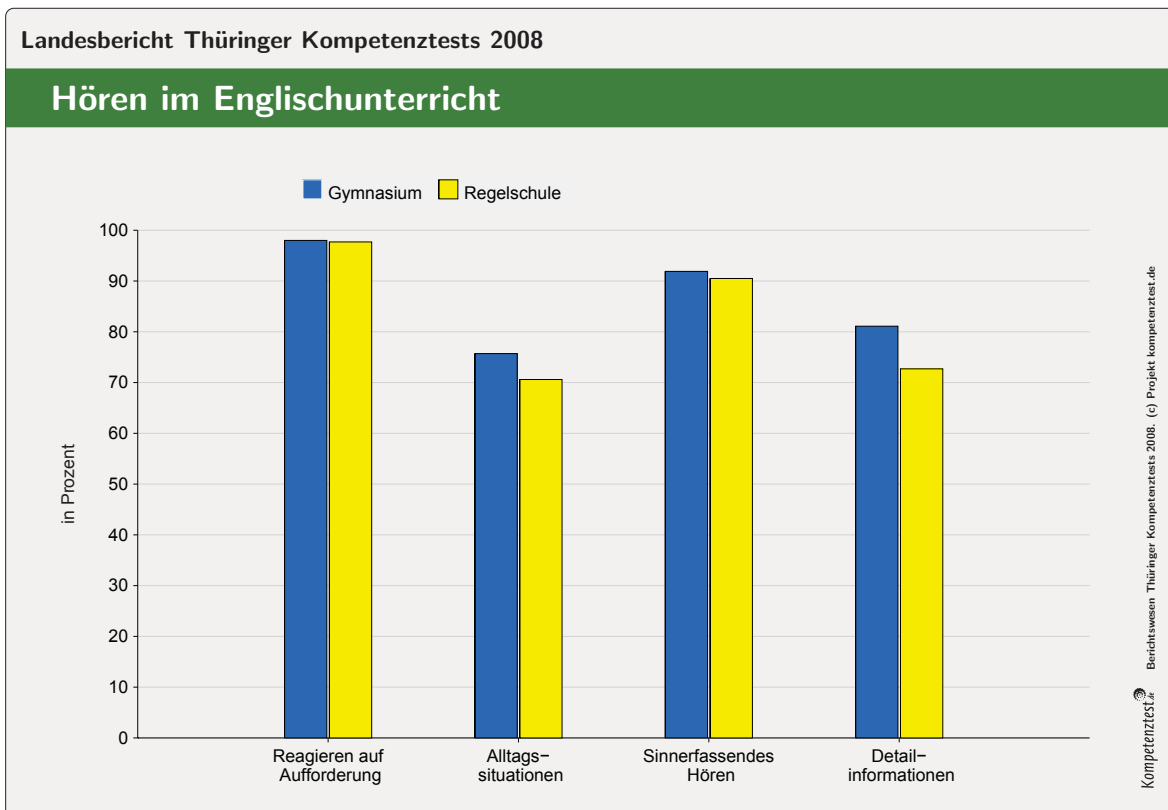


Abbildung 7.18: Häufig und sehr häufig im Englischunterricht umgesetzte Tätigkeiten des Hörens

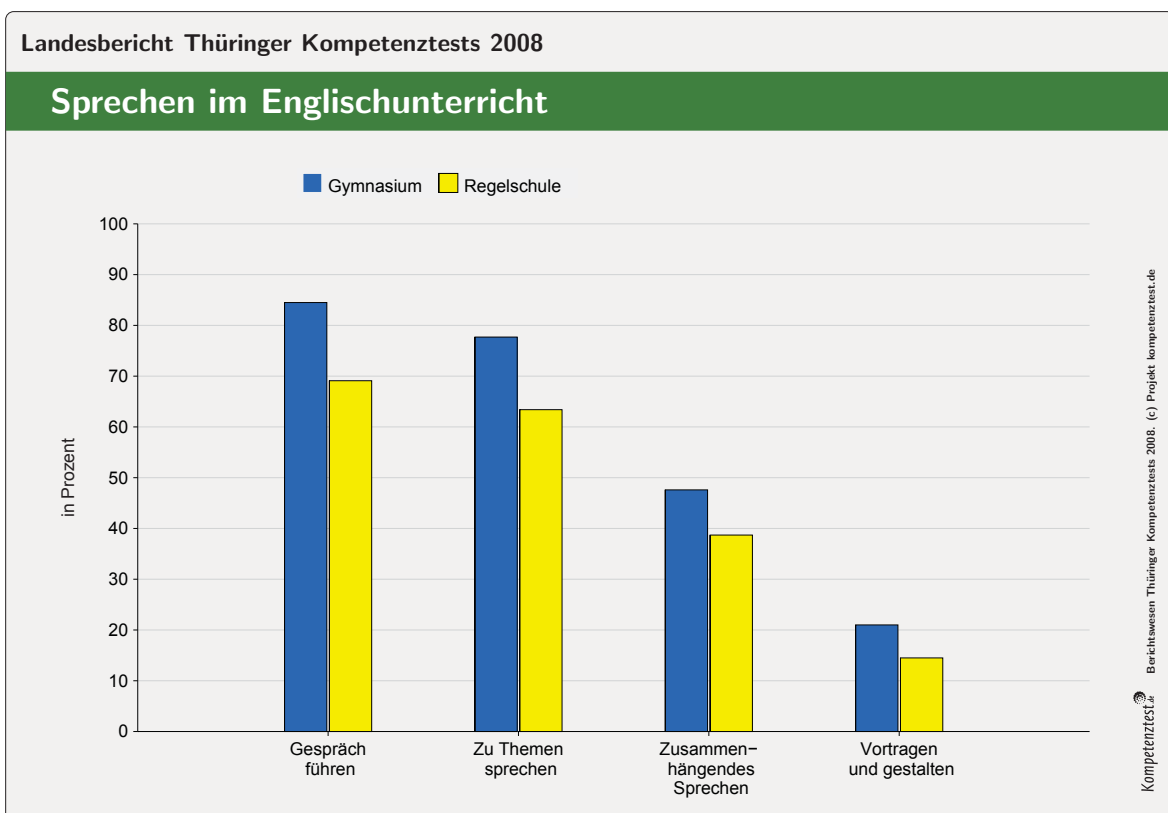


Abbildung 7.19: Häufig und sehr häufig im Englischunterricht umgesetzte Tätigkeiten des Sprechens

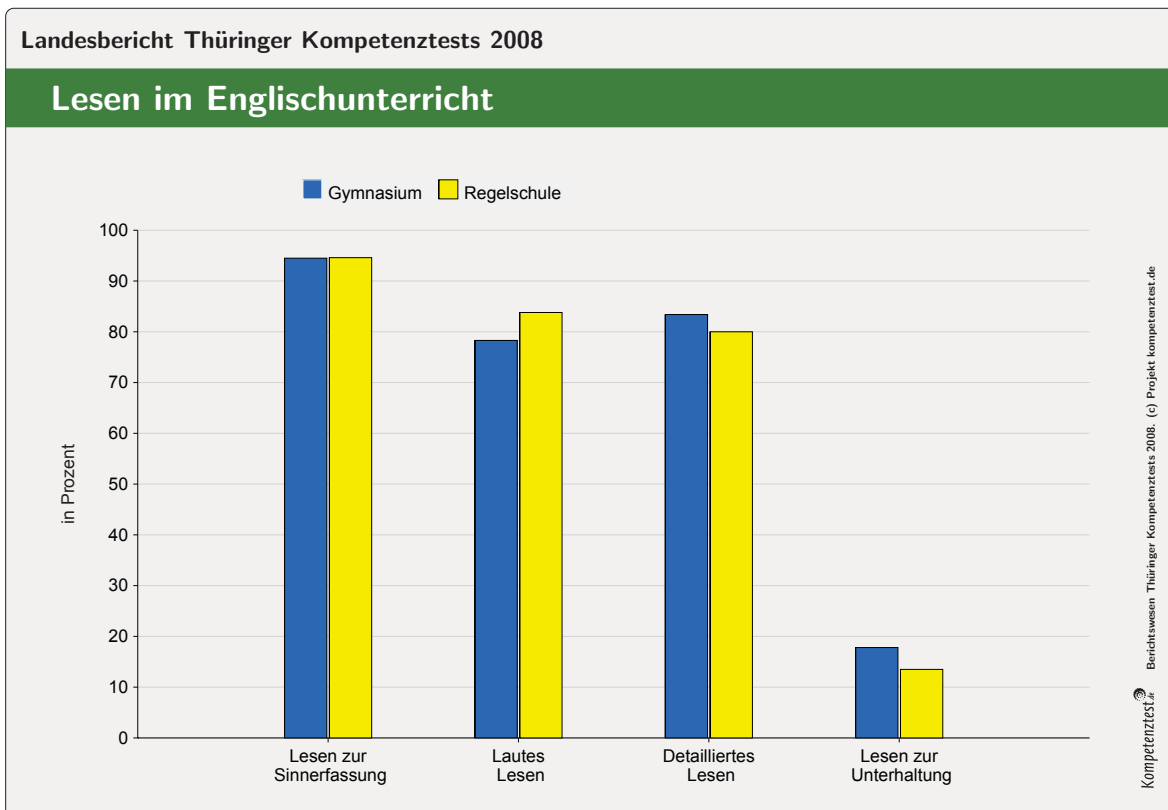


Abbildung 7.20: Häufig und sehr häufig im Englischunterricht umgesetzte Tätigkeiten des Lesens

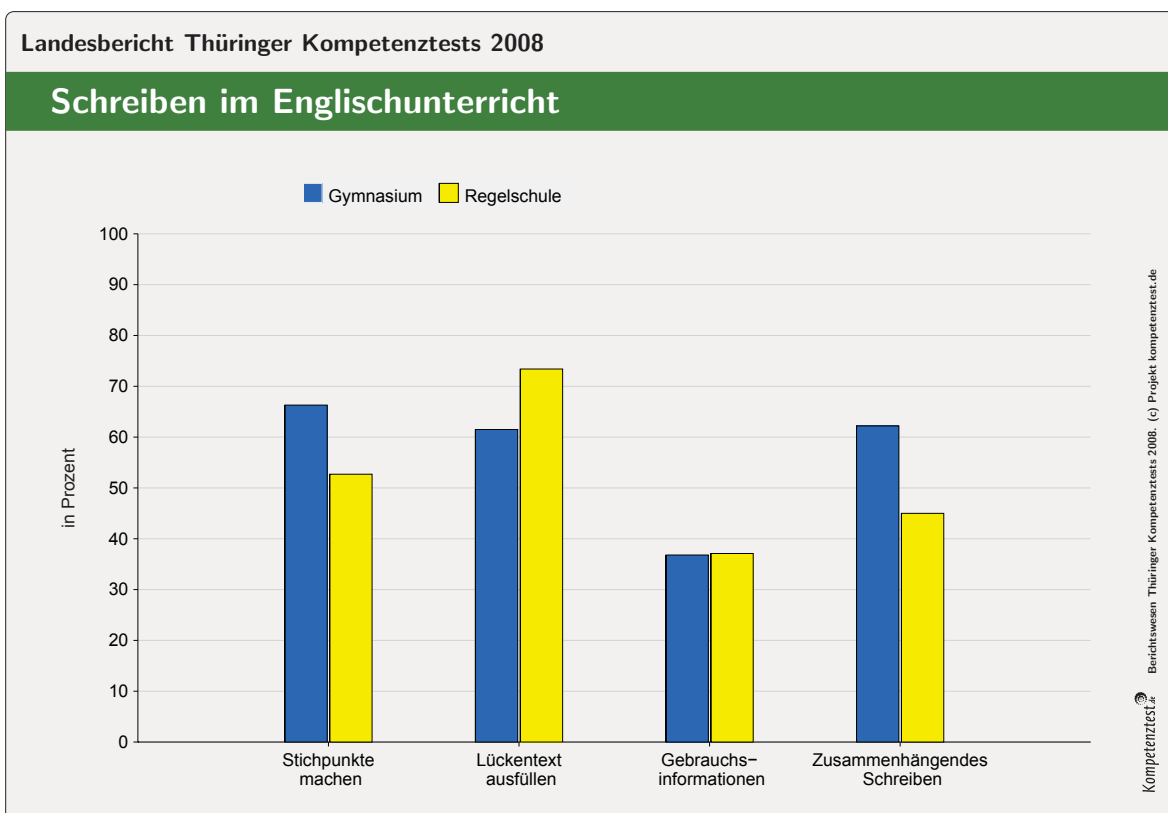


Abbildung 7.21: Häufig und sehr häufig im Englischunterricht umgesetzte Tätigkeiten des Schreibens

Schreiben als inhaltlicher Aspekt des Unterrichts

Hinsichtlich des Schreibens nehmen den höchsten Stellenwert die Formen Stichpunkte machen und das Ausfüllen von Lückentexten ein (siehe Abbildung 7.21). Dieser Aspekt wird nur am Gymnasium durch den Anteil des zusammenhängenden Schreibens übertroffen. Dem gegenüber ist feststellbar, dass schulartübergreifend das Verfassen von Gebrauchsinformationen eine deutlich nachrangige Rolle im Unterricht spielt. Hinsichtlich des Anteils des Schreibens am Gesamtunterricht im Fach Englisch kann ausgesagt werden, dass das Schreiben quantitativ den geringsten Teil ausmacht.

Zusammenfassend kann aus der Analyse der Daten hinsichtlich der im Unterricht eingesetzten Materialien folgendes abgeleitet werden: Die traditionellen Medien nehmen schulartübergreifend den breitesten Raum als Unterrichtsmaterialien ein. Sie fungieren als die am häufigsten eingesetzten Vermittler zwischen dem Lerngegenstand und der Schülergruppe bzw. dem Schüler. Dabei fällt auf, dass insbesondere das Sprachbuch, gemessen an der Häufigkeit seines Einsatzes, als äußerst geeignetes Medium erachtet wird. Hinsichtlich des Einbezugs von Materialien in den Unterricht, die die Schüler in ihrer Freizeit nutzen, zeigen sich deutlich geringere Werte. Individuell gestaltete Materialien werden nur selten im Unterricht eingesetzt.

Darstellung ausgewählter Unterrichtsaspekte in ihrem Einfluss auf die durchschnittliche Testleistung der Schüler einer Klasse

Analog zum Vorgehen für die Fächer Mathematik und Deutsch wurde wieder mit Hilfe von Regressionsmodellen untersucht, ob und inwiefern bestimmte Unterrichtsmerkmale mit der Testleistung der Schüler im Fach Englisch einhergehen. Dafür wurden wieder die Einzelergebnisse der Schüler pro Klasse zu einem Klassenwert gemittelt. Die Informationen zur Unterrichtsgestaltung liegen für jede Klasse immer nur anteilig vor (Aufwandsreduktion Lehrerfragebogen) und werden daher getrennt nach Merkmalsgruppen betrachtet. Folgende Merkmalsgruppen wurden hinsichtlich ihrer Bedeutung für die Testergebnisse auf Klassenebene hin untersucht: „Unterrichtsformate“ (Unterrichts-, Kommunikations- und Sozialformen), „Inhaltliche Schwerpunkte“ (Lesen, Schreiben), „Hausaufgabengestaltung“ (Häufigkeit, Auswertung) sowie der Einsatz von „Unterrichtsmedien“ (z. B. Flipchart). Die Ergebnisse werden so berichtet, dass jeweils die Ausprägung eines Unterrichtsmerkmals angegeben wird, die mit hohen Testergebnissen in Deutsch einhergeht.

Wie gut Unterschiede in der Testleistung von Klassen mit Unterschieden in der jeweiligen Unterrichtsgestaltung beschrieben werden können, lässt sich am Anteil aufgeklärter Varianz (=Varianzaufklärung) ablesen.

Auch an dieser Stelle ist es wichtig, noch mal zu betonen, dass die im Folgenden berichteten Ergebnisse lediglich als Orientierung dienen können, da viele mögliche Störeinflüsse nicht kontrolliert werden können. Die notwendige Mittelung der Schülerleistungen zu einem Gesamtwert einer Klasse vernachlässigt beispielsweise Einflussfaktoren seitens der Schüler (z. B. Motivation, Fähigkeit). Zudem können die Zusammenhänge zwischen Unterrichtsmerkmalen und Testleistung aufgrund der Stichprobengröße nicht schulartspezifisch betrachtet werden, somit bleiben Einflüsse der Schulart unberücksichtigt. Auch sollten die Ergebnisse aufgrund der systematisch fehlenden Werte vorsichtig interpretiert werden. Nicht zuletzt muss natürlich festgehalten werden, dass nur Merkmale hinsichtlich möglicher Zusammenhänge zur Testleistung untersucht werden können, die auch im Fragebogen erhoben wurden.

Aus der Gruppe „Inhaltliche Schwerpunkte“ stehen relativ viele Merkmale in signifikanter Beziehung zum Testergebnis (auch wenn die Zusammenhänge teilweise recht klein sind). Entsprechend wird eine recht hohe Varianzaufklärung von 24,2 % erzielt. Merkmale, die insbesondere mit einer hohen Testleistung einhergehen, sind (häufiges) Darlegen zusammenhängender Sachverhalte, (weniger häufiges) Arbeiten mit Nachschlagewerken, (häufiges) Führen von Gesprächen in unterschiedlichen Situationen, (häufiges) Führen themengebundener Gespräche,

(häufiges) Verstehen von/Reagieren auf Bitten, Fragen und Aufforderungen, (weniger häufiges) Ausfüllen von Lückentexten, (häufiges) Anfertigen von Notizen und Stichwortgerüsten sowie das (häufige) Erfassen und Verarbeiten von Detailinformationen.

Fast alle der „Unterrichtsmedien“ stehen in Zusammenhang zum Testergebnis in Englisch und klären insgesamt 8,4 % der Varianz der Testergebnisse auf. Die statistisch bedeutsamen Merkmale dieser Gruppe sind: (häufiger) Einsatz von OHP-Folien, schülergenerierten Arbeitsblättern und audiovisuellen Medien sowie (weniger häufiger) Einsatz von lehrergenerierten Arbeitsblättern. Das bedeutet, höhere Testergebnisse in Englisch werden zum Beispiel in Klassen beobachtet, deren Lehrerangaben, in ihrem Unterricht häufiger den Overheadprojektor, Schülerarbeitsblätter und audiovisuelle Medien und weniger häufiger eigene Arbeitsblätter einzusetzen.

Darüber hinaus können 9 % der Varianz der Testergebnisse durch Merkmale der Gruppe „Unterrichtsformate“ aufgeklärt werden, nämlich (häufige) Schüler-Schüler-Gespräche, (weniger häufigeres) Stationenlernen, (häufige) Partnerarbeit sowie (häufige) Schülervorträge.

Wie bereits für Mathematik und Deutsch geschehen, soll auch an dieser Stelle noch einmal betont werden, dass die Ergebnisse nicht bedeuten, dass andere Variablen keinerlei Einfluss auf die Schülerleistung in Deutsch haben. Es ist vielmehr zu vermuten, dass die Effekte mancher Merkmale über andere Merkmale vermittelt werden. Zudem sollten die beobachteten Zusammenhänge zwischen Unterrichtsmerkmalen und Testergebnissen nicht kausal interpretiert werden (im Sinne einer Vorhersage), da offen bleibt, was Ursache und was Wirkung ist. So kann ein positiver Zusammenhang zwischen dem Einsatz schülergenerierter Arbeitsblätter und den Testergebnissen einerseits bedeuten, dass sich der häufige Rückgriff auf Schülerarbeitsblätter tatsächlich förderlich auf die Leistung der Schüler in Englisch auswirkt. Andererseits kann der Zusammenhang auch einfach aufzeigen, dass Lehrer in Klassen mit hohem Leistungsdurchschnitt in Englisch öfter Schülerarbeitsblätter einsetzen.

7.1.4 Einfluss der Klassengröße

Neben den direkten Einflüssen des Unterrichts finden sich auf Klassenebene noch weitere Einflüsse auf die Schülerleistung. Vieldiskutiert ist der Effekt der Klassengröße. An diesem Beispiel zeigen sich allerdings auch besonders deutlich die Grenzen von korrelativen Analysen. Betrachtet man den Zusammenhang zwischen der Klassengröße und der mittleren Schülerleistung im Test in der Klasse 6 über alle Schularten hinweg, so ergibt sich jedes Jahr fachunabhängig ein Zusammenhang von mittlerer Größe. Im Jahr 2008 findet sich eine Korrelation von $r = 0,50$ in Deutsch und von $r = 0,51$ in Mathematik. In der Klassenstufe 3 hat die Korrelation immerhin noch einen Wert von $r = 0,14$ für den Deutschttest und $r = 0,11$ für das Fach Mathematik. Das Vorzeichen der Korrelation ist positiv, demnach haben die größeren Klassen tendenziell die *besseren* Ergebnisse erzielt. Eine kausale Interpretation dieser Zusammenhänge kann aber zu kapitalen Fehlschlüssen führen: Betrachtet man den Zusammenhang zwischen Klassengröße und Testergebnis nach Schularten getrennt, so ergibt sich in der Klassenstufe 6 für alle Schularten eine Korrelation von Null, es besteht also innerhalb der einzelnen Schularten kein Zusammenhang.

Der Grund für diese scheinbar völlig widersprüchlichen Ergebnisse besteht darin, dass Förderschulen naturgemäß die kleinsten Klassen, aber auch die niedrigsten Ergebnisse haben, während Gymnasien die größten Klassen und die höchsten Ergebnisse aufweisen. Größere Klassen haben daher tatsächlich tendenziell die besseren Ergebnisse, aber nicht aufgrund, sondern möglicherweise trotz ihrer Größe. Die Möglichkeit einer solchen „Verfälschung“ von Zusammenhängen durch nicht beachtete andere Einflussfaktoren ist daher bei der Interpretation der hier berichteten Zusammenhänge immer im Auge zu behalten (vgl. Nachtigall et al., 2001).

Die Frage, welchen *kausalen* Einfluss die Größe einer Klasse tatsächlich auf die erzielten Schülerleistungen hat, kann nur im Rahmen so genannter kontrollierter Studien ermittelt werden, bei denen unter sonst möglichst ähnlichen Bedingungen die gleichen Lehrer in Klassen verschiedener Größe unterrichten.

Dieses Beispiel ist deshalb so wichtig, weil es die Grenzen der Interpretation von korrelativen Zusammenhängen aufzeigt. Solche Zusammenhänge können, müssen aber nicht eine kausale Beziehung wieder geben (es besteht auch eine korrelative Beziehung zwischen der Anzahl der Störche und der Anzahl der Geburten pro Jahr in den Regionen – auch diese dürfte wohl nicht kausal interpretierbar sein). Entsprechend sind die in diesem Kapitel dargestellten Zusammenhänge zu interpretieren. Sie stellen Kandidaten für kausale Einflüsse dar, aber ob tatsächlich kausale Beziehungen vorliegen, kann in diesem Bericht nicht entschieden werden.

7.2 Einflussfaktoren auf Schülerebene

In diesem Abschnitt soll die Beziehung zwischen den Testergebnissen und den bei den Kompetenztests erhobenen Schülermerkmalen näher betrachtet werden. Schülermerkmale, wie beispielsweise Muttersprache oder der sozioökonomische Status, liegen nicht im Einflussbereich des Lehrers. Sie können aber einen großen Einfluss auf das Abschneiden im Kompetenztest haben. Deshalb ist es besonders wichtig die entsprechenden Schülermerkmale zu erfassen und in die Auswertung der Kompetenztests einzubeziehen, um Vergleiche zwischen Klassen und Schulen fairer zu gestalten. Es folgen detaillierte Analysen zu den Schülermerkmalen Geschlecht, Muttersprache, sozioökonomischer Status, Jahr des Schulbesuchs und besondere Lernschwierigkeiten. Dabei werden in der Regel Mittelwerte verglichen und die Stärke von Zusammenhängen mit Hilfe von Korrelationen beschrieben. Weiterhin wird die Beziehung zwischen Halbjahresnote und Testergebnissen dargestellt und in Form einer Regressionsanalyse versucht, den gemeinsamen Einfluss aller erfassten Merkmale auf die Testleistung abzuschätzen. Am Ende dieses Abschnittes wird es eine längsschnittliche Betrachtung der Ergebnisse des Kompetenztests Mathematik geben. Dazu wird der Einfluss der Schülermerkmale in einem Jahrgang über die Zeit hinweg betrachtet, nämlich der Einfluss der Schülermerkmale in der 6. Klasse von 2006 verglichen mit den Ergebnissen der 8. Klasse von 2008.

In den dargestellten Tabellen sind neben den einzelnen Mittelwerten (MW) auch die Standardabweichung (SD) und die Fallzahl (N) angegeben. Außerdem wurde zur besseren Interpretation der gefundenen Unterschiede die Effektstärke (d) berechnet. Diese gibt an, wie stark der gefundene Effekt ist. Dabei gilt nach Cohen (1988)¹ eine Effektstärke ab $d = 0,2$ als geringer, ab $d = 0,5$ als mittelstarker und ab $d = 0,8$ als starker Effekt. Ein negatives Vorzeichen besagt, dass die zweite Gruppe die höheren Werte hat. So bedeutet beispielsweise ein Unterschied zwischen Jungen und Mädchen von $d = -0,237$ beim Kompetenztest Deutsch, Gymnasium Klassenstufe 6, dass die Mädchen im Mittel besser abschneiden. Dieser Effekt würde als klein interpretiert werden.

7.2.1 Geschlecht

Der aus früheren Jahrgängen bekannte Geschlechtereffekt zeigt sich auch bei den diesjährigen Kompetenztests: Mädchen erreichten im Fach Deutsch und Jungen in Mathematik tendenziell ein höheres Gesamtergebnis. Beim Kompetenztest Englisch unterscheiden sich Jungen und Mädchen hingegen nicht.

¹ vergleiche Cohen (1988)

Tabelle 7.6: Mittelwertevergleich der Testergebnisse der Klassenstufe 3 zwischen Jungen und Mädchen unter Berücksichtigung der Schulart

	Männlich			Weiblich			Effektstärke
	MW	(SD)	(N)	MW	(SD)	(N)	(d)
Grundschule							
DK 3	29,28	(8,99)	(7 384)	30,74	(8,387)	(7 124)	(-0,180)
MK 3	26,65	(8,42)	(7 364)	24,86	(8,82)	(7 071)	(0,216)
Förderschule							
DK 3	19,00	(9,82)	(180)	19,67	(10,18)	(92)	(-0,060)
MK 3	17,39	(8,86)	(184)	16,17	(8,67)	(87)	(0,139)

In der Klassenstufe 3 unterscheiden sich die Leistungen von Jungen und Mädchen bei den Kompetenztests Mathematik und Deutsch nicht wesentlich voneinander. Bei der Analyse der Ergebnisse des Tests Mathematik der Thüringer Grundschulen lässt sich ein kleiner Effekt feststellen, welcher zeigt, dass Jungen im Mittel leicht besser abschneiden als die Mädchen. In Deutsch liegt mit $d = -0,18$ kein nennenswerter Geschlechtereffekt vor. Man kann lediglich schlussfolgern, dass Mädchen der Grundschule tendenziell eine bessere Leistung im Kompetenztest Deutsch erzielen. In den Förderschulen kann man weder beim Kompetenztest Mathematik noch beim Kompetenztest Deutsch einen geschlechtsspezifischen Effekt feststellen. Jungen und Mädchen schneiden im Mittel etwa gleich gut ab. Vergleicht man die Geschlechtereffekte dieses Jahres mit jenen der Vorjahre, könnte man vermuten, dass es eine Tendenz gibt, dass sich die Leistungen von Jungen und Mädchen aneinander angleichen. Man kann dabei aber noch nicht ausschließen, dass diese Entwicklung per Zufall entstanden ist. Es könnte auch sein, dass sich Mädchen und Jungen im nächsten Jahr wieder stärker voneinander unterscheiden.

In der Klassenstufe 6 lassen sich ebenfalls geschlechtsspezifische Unterschiede bei den Ergebnissen der Kompetenztests Deutsch und Mathematik finden. Sowohl in Gymnasien, als auch in Regelschulen und Förderschulen schneiden Mädchen im Kompetenztest Deutsch durchschnittlich etwas besser ab als die Jungen. Hingegen im Kompetenztest Mathematik erzielten Jungen der Gymnasien, Regelschulen und Förderschulen im Mittel eine bessere Testleistung. Für beide Fächer und über die verschiedenen Schularten hinweg sind die geschlechtsspezifischen Effekte als klein zu bewerten. Im Gegensatz zum letzten Jahr wurde der Kompetenztest Englisch 2008 getrennt für Lesen und Hören erfasst und ausgewertet. Wie auch im letzten Jahr zeigt sich, dass Jungen und Mädchen in Englisch im Durchschnitt gleiche Leistungen erzielten. Lediglich im Untertest Lesen in der Regelschule kann man bei einem Wert $d = -0,204$ von einem kleinen Effekt zugunsten der Mädchen sprechen. Damit ist das Abschneiden in den Kompetenztests bezogen auf das Geschlecht seit dem letzten Jahr soweit stabil geblieben. Eine Ausnahme bildet das Ergebnis des Kompetenztests Deutsch in der Förderschule. Während Jungen und Mädchen im letzten Jahr durchschnittlich gleich gut abschnitten, lässt sich 2008 ein kleiner Effekt finden, der anzeigt, dass Mädchen im Mittel bessere Leistungen im Test Deutsch zeigten.

Seit letztem Jahr nehmen die 8. Klassen des Freistaates Thüringen verpflichtend am Kompetenztest Mathematik teil. 2008 zeigt sich erneut, dass Jungen die Aufgaben tendenziell besser lösten als die gleichaltrigen Mädchen. Während die geschlechtsspezifische Differenz der Test-

Tabelle 7.7: Mittelwertevergleich der Testergebnisse der Klassenstufe 6 zwischen Jungen und Mädchen unter Berücksichtigung der Schulart

	Männlich			Weiblich			Effektstärke
	MW	(SD)	(N)	MW	(SD)	(N)	(d)
Gymnasium							
DK 6	70,69	(11,01)	(2 573)	73,34	(11,35)	(2 927)	(−0,237)
MK 6	20,19	(4,02)	(2 577)	18,72	(4,16)	(2 913)	(0,359)
EK 6 L	40,49	(6,46)	(2 556)	41,14	(5,98)	(2 857)	(−0,104)
EK 6 H	35,48	(4,623)	(2 556)	35,71	(4,55)	(2 856)	(−0,052)
Regelschule							
DK 6	52,41	(14,75)	(3 348)	56,81	(13,97)	(2 908)	(−0,306)
MK 6	14,04	(4,75)	(3 314)	12,48	(4,56)	(2 883)	(0,335)
EK 6 L	27,92	(8,37)	(3 281)	29,57	(7,79)	(2 864)	(−0,204)
EK 6 H	27,27	(7,01)	(3 271)	27,79	(6,74)	(2 858)	(−0,076)
Förderschule							
DK 6	44,22	(13,11)	(201)	48,57	(12,32)	(108)	(−0,342)
MK 6	10,53	(4,87)	(214)	8,42	(3,95)	(110)	(0,476)
EK 6 L	19,88	(8,38)	(210)	20,96	(7,94)	(108)	(−0,132)
EK 6 H	19,62	(8,50)	(210)	19,54	(7,88)	(108)	(0,010)

Tabelle 7.8: Mittelwertevergleich der Testergebnisse der Klassenstufe 8 zwischen Jungen und Mädchen unter Berücksichtigung der Schulart

	Männlich			Weiblich			Effektstärke
	MW	(SD)	(N)	MW	(SD)	(N)	(d)
Gymnasium							
MK 8	42,09	(7,13)	(2 150)	39,98	(6,46)	(2 459)	(0,325)
Regelschule							
MK 8	31,73	(6,51)	(3 594)	29,24	(7,10)	(3 127)	(0,350)
Förderschule							
MK 8	25,85	(6,62)	(219)	22,09	(6,86)	(121)	(0,558)

leistung mit kleinen Effekten in Gymnasien und Regelschulen eher gering ausfällt, zeigt sich der Geschlechtereffekt zugunsten der Jungen besonders deutlich in der Förderschule. Mit einem Wert von $d = 0,558$ liegt hier immerhin ein mittelstarker Effekt vor.

7.2.2 Muttersprache

Die PISA-Ergebnisse² von 2006 haben gezeigt, dass sich in vielen OECD-Staaten weiterhin große Kompetenzunterschiede zwischen Jugendlichen ohne Migrationshintergrund und Jugendlichen mit Migrationshintergrund finden lassen. Deutschland zählt dabei im internationalen Vergleich zu den Staaten mit besonders ausgeprägten, migrationsspezifischen Kompetenzunterschieden. Schüler, welche in einem anderen Land geboren sind bzw. als zweite Generation in Deutschland leben, schneiden deutlich schlechter ab als die deutschsprachigen Mitschüler.

Im Kompetenztest wurde ein möglicher Migrationshintergrund erfasst, indem die Lehrer angeben sollten, ob die Muttersprache ihrer Schüler Deutsch sei oder nicht. Die aus diesen Angaben resultierende Verteilung ist in Tabelle 7.9 dargestellt. Interessant ist dabei der Vergleich der Verteilung innerhalb einer Schulart mit der jeweiligen Gesamtverteilung.

Tabelle 7.9: Verteilung der Schüler mit deutscher und nicht-deutscher Muttersprache auf die verschiedenen Schularten

	Muttersprache Deutsch		Mutterspr. nicht Dt.	
	Anzahl	(in Prozent)	Anzahl	(in Prozent)
Klassenstufe 3				
Grundschulen	14 971	(96,3 %)	572	(3,7 %)
Förderschulen	319	(98,5 %)	5	(1,5 %)
Gesamt	15 290	(96,4 %)	577	(3,6 %)
Klassenstufe 6				
Gymnasien	5 696	(96,0 %)	235	(4,0 %)
Regelschulen	6 564	(95,8 %)	290	(4,2 %)
Förderschulen	348	(97,8 %)	8	(2,1 %)
Gesamt	12 929	(95,8 %)	565	(4,2 %)
Klassenstufe 8				
Gymnasien	4 914	(98,4 %)	80	(1,6 %)
Regelschulen	7 212	(95,8 %)	313	(4,2 %)
Förderschulen	380	(96,9 %)	12	(3,1 %)
Gesamt	12 911	(96,9 %)	548	(3,1 %)

In Klassenstufe 3 besuchten 2008 fast alle Schüler mit nicht-deutscher Muttersprache eine Grundschule. Während 3,7 % aller Schüler der Klassenstufe 3 eine nicht-deutsche Muttersprache hatten, waren es an Förderschulen nur 1,5 % der Schüler. Damit waren Schüler mit nicht-deutscher Muttersprache im Vergleich zur Gesamtverteilung an Förderschulen leicht unterrepräsentiert.

Über die verschiedenen Schularten hinweg waren in Klassenstufe 6 im Schuljahr 2007 / 2008 4,2 % der Schüler nicht-deutscher Muttersprache. Damit sind in Gymnasien, wie auch in Regelschulen so viele Schüler mit Migrationshintergrund zu finden, wie man es aufgrund der Gesamtverteilung erwarten könnte. An Förderschulen waren in diesem Jahr Schüler mit nicht-deutscher Muttersprache wie in der Klassenstufe 3 anteilig leicht unterrepräsentiert.

² vergleiche PISA-Konsortium Deutschland (2007)

Insgesamt gesehen, haben 3,1 % der Schüler der 8. Klassen nach Angabe der Lehrer einen Migrationshintergrund. Damit weist der Anteil an Schülern nicht-deutscher Muttersprache von 1,6 % an Gymnasien auf eine leichte Unterrepräsentation der eben genannten Gruppe hin. Dagegen sind Schüler mit Migrationshintergrund an Regelschulen mit 4,2 % leicht überrepräsentiert. Der Anteil Schüler nicht-deutscher Muttersprache von 3,1 % an den Thüringer Förderschulen entspricht der Erwartung, die man aufgrund des Gesamtanteils über die Schularten hinweg erwarten würde.

Mittelwertevergleich der Testergebnisse

In den Tabellen 7.10, 7.11 und 7.12 sind die Ergebnisse der Schüler in den Kompetenztests getrennt nach der Schulart und der Muttersprache der Schüler dargestellt. Dabei werden die Ergebnisse der Förderschulen nicht berichtet, da hier die Anzahl der Kinder mit nicht-deutscher Muttersprache zu gering für eigene statistische Analysen ist. In den Tabellen wird ersichtlich, dass die Muttersprache der Schüler einen klaren Bezug zu den Ergebnissen in den Kompetenztests Deutsch, Mathematik und Englisch aufweist. Schüler nicht-deutscher Muttersprache erreichen, wie auch in den letzten Jahren, durchschnittlich schlechtere Ergebnisse als die Mitschüler ohne Migrationshintergrund.

Tabelle 7.10: Mittelwertevergleich der Testergebnisse der Klassenstufe 3 zwischen Schülern mit Deutsch als Muttersprache und anderer Muttersprache

	Muttersprache Deutsch			Mutterspr. nicht Dt.			Effektstärke
	MW	(SD)	(N)	MW	(SD)	(N)	(d)
Grundschule							
DK 3	30,13	(8,65)	(14 299)	24,96	(9,65)	(536)	(0,564)
MK 3	25,85	(8,30)	(14 061)	22,88	(8,37)	(519)	(0,356)

Tabelle 7.11: Mittelwertevergleich der Testergebnisse in Klassenstufe 6 zwischen Schülern mit Deutsch als Muttersprache und anderer Muttersprache

	Muttersprache Deutsch			Mutterspr. nicht Dt.			Effektstärke
	MW	(SD)	(N)	MW	(SD)	(N)	(d)
Gymnasium							
DK 6	72,26	(11,16)	(5 309)	67,71	(12,94)	(221)	(0,377)
MK 6	19,47	(4,11)	(5 274)	18,01	(4,94)	(216)	(0,321)
EK 6 L	40,89	(6,169)	(5 199)	39,46	(7,19)	(214)	(0,214)
EK 6 H	35,66	(4,52)	(5 198)	34,23	(5,78)	(214)	(0,276)
Regelschule							
DK 6	54,85	(14,41)	(6 077)	45,33	(14,585)	(263)	(0,657)
MK 6	13,40	(4,71)	(5 936)	11,40	(4,62)	(259)	(0,429)
EK 6 L	28,84	(8,11)	(5 882)	26,11	(8,51)	(261)	(0,328)
EK 6 H	27,61	(6,84)	(4 866)	25,38	(7,59)	(261)	(0,309)

Auffallend ist im Vergleich zum letzten Jahr, dass die Effektstärke d zwischen Schülern deutscher und nicht-deutscher Muttersprache 2008 größer ausfällt. In der Klassenstufe 3 erzielen die Grundschüler mit Migrationshintergrund im Mittel deutlich schlechtere Ergebnisse bei

Tabelle 7.12: Mittelwertevergleich der Testergebnisse in Klassenstufe 8 zwischen Schülern mit Deutsch als Muttersprache und anderer Muttersprache

	Muttersprache Deutsch			Mutterspr. nicht Dt.			Effektstärke
	MW	(SD)	(N)	MW	(SD)	(N)	(d)
Gymnasium							
MK 8	41,00	(6,55)	(4 554)	38,72	(7,23)	(79)	(0,331)
Regelschule							
MK 8	30,73	(7,18)	(6 601)	26,78	(7,74)	(279)	(0,540)

den Kompetenztests als die Schüler ohne Migrationshintergrund. Im Test Deutsch liegt mit $d = 0,564$ ein mittelstarker Effekt vor. Es ist aber zu bedenken, dass sich die Schüler mit und ohne Migrationshintergrund im letzten Jahr nur geringfügig voneinander unterschieden. In Mathematik lässt sich eine ähnliche Entwicklung erkennen. Zwar ist der Effekt $d = 0,356$ als klein zu bewerten. Jedoch haben sich Schüler deutscher und nicht-deutscher Muttersprache im Kompetenztest letzten Jahres gar nicht voneinander unterschieden.

Die Gymnasiasten der Klassenstufe 6 deutscher und nicht-deutscher Muttersprache unterscheiden sich in allen Kompetenztests leicht voneinander, es liegen jeweils kleine Effekte vor. Im Gegensatz zum letzten Jahr zeigt sich auch bei der Mathematikkompetenz ein kleiner Effekt. 2007 erbrachten Schüler mit und ohne Migrationshintergrund im Mittel die gleiche Leistung in den Mathematikkompetenztests. In den sechsten Klassen der Regelschulen zeigt sich ein ähnliches Bild. Im Kompetenztest Deutsch unterscheiden sich deutsche und nicht-deutsche Muttersprachler in ihren Ergebnissen mit einem mittelstarken Effekt von $d = 0,657$. In Mathematik und Englisch liegen kleine Effekte vor, welche auf einen Leistungsunterschied der beiden Gruppen hinweisen, der im letzten Jahr nicht vorhanden war. Damit schneiden in der Klassenstufe 6 Schüler mit Migrationshintergrund im Mittel bei allen Kompetenztests schlechter ab als Schüler ohne Migrationshintergrund.

Dieser migrationspezifische Kompetenzunterschied ist auch in der Klassenstufe 8 vorhanden. Auch hier zeigen Schüler nicht-deutscher Muttersprache im Durchschnitt eine schlechtere Leistung als die deutschen Muttersprachler. Mit einem mittelstarken Effekt ist die Leistungsdivergenz in den Regelschulen stärker ausgeprägt als in den Gymnasien.

Betrachtet man die Ergebnisse im Hinblick auf die verschiedenen Fächer und Schularten, so fällt einerseits auf, dass die migrationsbedingten Leistungsunterschiede im Kompetenztest Deutsch größer sind als in Mathematik oder Englisch. Dieser Effekt erscheint logisch, wenn man bedenkt, dass sprachliche Fähigkeiten besonders wichtig sind, um Texte zu verstehen und zu verfassen. Dagegen sollten die sprachlichen Fähigkeiten in Mathematik zwar auch einen Einfluss haben, hier sollten es aber auch mathematische und logikbezogene Kompetenzen sein, die auf die Mathematikleistung wirken. Andererseits zeigt ein Blick auf die Ergebnisse der verschiedenen Schularten, dass migrationsbezogene Effekte in Gymnasium kleiner ausfallen als in den Regelschulen. Eine mögliche Erklärung hierfür wäre, dass Schüler mit nicht-deutscher Muttersprache am Gymnasium eine bessere Förderung ihrer Sprachfähigkeiten erhalten. Eine genauso plausible Erklärung wäre, dass unter den Schülern mit nicht-deutscher Muttersprache nur diejenigen das Gymnasium besuchen, die außerschulisch eine entsprechende Sprachförderung erfahren haben. Beispielsweise wäre es vorstellbar, dass die Gymnasiasten mit Migrationshintergrund aus Familienhäusern mit höherem sozioökonomischen Status stammen, in welchen es angesehen ist, mehrere Sprachen zu beherrschen. Sollte hier die

Ursache für die geringeren Leistungsdiskrepanzen liegen, wäre es z. B. sinnvoll, die Elternhäuser der entsprechenden Kinder stärker mit Informationen zur Bedeutsamkeit sprachlicher Integration ihrer Kinder zu versorgen. Außerdem sollte ausführlich über das Bildungssystem und dessen Möglichkeiten informiert werden.

Insgesamt ist festzuhalten, dass sich Schüler deutscher und nicht-deutscher Muttersprache in diesem Jahr über alle Schularten und Kompetenztests hinweg in ihrer Leistung unterscheiden haben. Schüler nicht-deutscher Muttersprache schnitten im Mittel immer schlechter ab. Dieses Ergebnis spricht dafür, dass die gezielte Sprachförderung von Kindern mit nicht-deutscher Muttersprache intensiviert werden sollte. Da nicht-deutsche Muttersprache schon in der Grundschule zu deutlichen Kompetenzunterschieden führt, sollte so früh wie möglich damit begonnen werden, die Sprachfähigkeiten der Schüler mit Migrationshintergrund zu trainieren. Immerhin besteht durch frühes Trainieren der Unterrichtssprache die Möglichkeit, dass Leistungsunterschiede in den folgenden Schuljahren nicht nur in Deutsch, sondern auch in anderen Unterrichtsfächern, abgemildert werden.

7.2.3 Sozioökonomischer Status (SES)

Im Rahmen der PISA-Studie wird regelmäßig der Zusammenhang zwischen sozialer Herkunft und Schülerkompetenz untersucht. Die PISA-Untersuchung von 2006³ konnte erneut zeigen, dass es in allen OECD-Staaten einen bedeutsamen Zusammenhang zwischen dem SES des Elternhauses der teilnehmenden Jugendlichen und den erreichten Kompetenzen gibt. Deutschland weist dabei einen durchschnittlichen Zusammenhang auf. Es hat sich gezeigt, dass sich der Zusammenhang zwischen sozialer Herkunft und Lesekompetenz für deutsche Schüler seit 2000 abgeschwächt hat. Obwohl die verbleibenden Unterschiede weiter als hoch zu bezeichnen sind, zeigt dieser positive Trend, dass es sich lohnt auf Bildungsungerechtigkeiten aufmerksam zu machen, um einen Anstoß zu entsprechenden Maßnahmen zu liefern.

Im Rahmen der PISA-Erhebung wird die soziale und kulturelle Herkunft der Schüler über eine Vielzahl von Indikatoren erfasst. Solch eine Vorgehensweise ist sehr wünschenswert, wäre aber im Rahmen der Kompetenztests mit einem zu großen Aufwand für Schüler und Lehrer verbunden. Daher muss hier auf sparsame Methoden zur Erfassung des sozioökonomischen Status (abgekürzt aus dem Englischen: SES) als Maß für die soziale Herkunft der Schüler zurückgegriffen werden. Im internationalen Grundschulvergleich IGLU⁴ wurde der sozioökonomische Status u. a. mit Hilfe der sogenannten „Bücherfrage“ erhoben. Die Bücherfrage wird auf einer fünfstufigen Skala gemessen und hat sich als ein nützlicher Indikator für den SES erwiesen. Dabei sind die Angaben auf der Ebene des einzelnen Schülers ungenau und dürfen nicht interpretiert werden. In aggregierter Form sind die Ergebnisse der Bücherfrage hingegen recht zuverlässige Indikatoren für den SES. Daher wurde die Bücherfrage seit dem Jahr 2004 im Rahmen der Kompetenztests gestellt und für Gruppenanalysen als Indikator für die soziale Herkunft genutzt.

Als Maß für den Zusammenhang zwischen dem so erhobenen SES und den Ergebnissen der Schüler in den Kompetenztests wurden zunächst Korrelationen berechnet. Diese sind in den folgenden Tabellen nach Schularten getrennt wiedergegeben.

³ vergleiche PISA-Konsortium Deutschland (2007)

⁴ vergleiche Bos (2003)

Tabelle 7.13: Korrelation des SES mit den Testergebnissen (Klassenstufe 3)

	Grundschule	Förderschule	Gesamt
$r_{\text{SES} - \text{DK3}}$	0,22	0,04	0,22
$r_{\text{SES} - \text{MK3}}$	0,25	0,13	0,25

Tabelle 7.14: Korrelation des SES mit den Testergebnissen (Klassenstufe 6)

	Gymnasium	Regelschule	Förderschule	Gesamt
$r_{\text{SES} - \text{DK6}}$	0,17	0,20	0,16	0,37
$r_{\text{SES} - \text{MK6}}$	0,17	0,19	0,16	0,36
$r_{\text{SES} - \text{EK6 H}}$	0,09	0,13	0,09	0,31
$r_{\text{SES} - \text{EK6 L}}$	0,14	0,16	0,14	0,36

Tabelle 7.15: Korrelation des SES mit den Testergebnissen (Klassenstufe 8)

	Gymnasium	Regelschule	Förderschule	Gesamt
$r_{\text{SES} - \text{MK 8}}$	0,21	0,21	0,16	0,40

Die Korrelationstabellen zeigen, dass die Testleistung in Abhängigkeit vom SES variiert. Zu beachten ist, dass aufgrund der ungenauen Messung des SES durch die Bücherfrage auf Schülerebene die angegebenen Werte eine Untergrenze der Stärke des Zusammenhangs darstellen. Um ein genaueres Bild davon zu bekommen, wie stark sich die Schüler unterschiedlicher sozialer Herkunft für die verschiedenen Schularten unterscheiden, werden im folgenden Abschnitt die Mittelwerte der Kompetenztestleistungen verglichen. Dabei wird die Gruppe von Schülern mit eher niedrigem SES der Gruppe von Schülern mit eher hohem SES gegenüber gestellt. Niedriger SES wird in diesem Fall so operationalisiert, dass bei der Bücherfrage eine der beiden niedrigsten Ausprägungen angekreuzt wurden, hoher SES durch die Angabe einer der beiden höchsten Kategorien.

Tabelle 7.16: Mittelwertevergleich der Testergebnisse der Klassenstufe 3 bezüglich des sozio-ökonomischen Status unter Berücksichtigung der Schulart

	SES niedrig			SES hoch			Effektstärke
	MW	(SD)	(N)	MW	(SD)	(N)	(d)
Grundschule							
DK 3	26,81	(8,76)	(3 839)	32,18	(8,33)	(4 456)	(−0,628)
MK 3	22,45	(7,94)	(3 988)	27,83	(8,19)	(4 551)	(−0,667)
Förderschule							
DK 3	19,29	(9,63)	(116)	20,74	(10,58)	(58)	(−0,143)
MK 3	15,65	(8,65)	(130)	18,32	(9,62)	(60)	(−0,296)

Die Analysen zeigen, dass sich besonders die Grundschüler abhängig vom SES in ihren Leistungen unterscheiden. Sowohl im Kompetenztest Mathematik als auch in Deutsch liegt ein mittlerer Effekt vor. Hingegen für die Förderschulen der Klassenstufe 3 liegt für Mathematik ein kleiner Effekt zugunsten der Schüler mit hohem SES vor, in Deutsch kann man nur von einem Trend sprechen. Die Effektstärke von $d = -0,143$ zeigt, dass bei Förderschülern die Leistung im Kompetenztest Deutsch nicht systematisch mit SES variieren. Jedoch ist an dieser Stelle erneut anzumerken, dass der SES mit der Bücherfrage auf Schülerebene nicht perfekt, sondern nur ungenau erfasst werden kann. Insbesondere in Klassenstufe 3 ist von nicht unbedeutenden Messfehlern hinsichtlich der Verwendung der Bücherfrage auszugehen. Trotzdem zeigen sich in der Grundschule deutliche Zusammenhänge, die auf den erhöhten Unterstützungsbedarf für Schüler aus sozial schwächeren Familien hinweisen.

Tabelle 7.17: Mittelwertevergleich der Testergebnisse der Klassenstufe 6 bezüglich des sozio-ökonomischen Status unter Berücksichtigung der Schulart

	SES niedrig			SES hoch			Effektstärke
	MW	(SD)	(N)	MW	(SD)	(N)	(d)
Gymnasium							
DK 6	69,28	(10,51)	(458)	73,68	(10,93)	(2 882)	(−0,410)
MK 6	18,11	(4,275)	(490)	19,95	(4,07)	(3 032)	(−0,441)
EK 6 L	39,35	(6,42)	(460)	41,67	(6,03)	(2 883)	(−0,373)
EK 6 H	34,92	(4,57)	(461)	35,98	(4,51)	(2 880)	(−0,234)
Regelschule							
DK 6	50,75	(14,34)	(1 910)	57,91	(14,46)	(1 326)	(−0,497)
MK 6	12,16	(4,64)	(2 062)	14,51	(4,713)	(1 391)	(−0,502)
EK 6 L	27,05	(7,98)	(1 947)	30,43	(8,12)	(1 326)	(−0,420)
EK 6 H	26,39	(7,1)	(1 945)	28,67	(6,65)	(1 321)	(−0,331)
Förderschule							
DK 6	44,19	(13,08)	(158)	48,15	(13,76)	(33)	(−0,295)
MK 6	9,16	(4,36)	(175)	10,39	(4,59)	(38)	(−0,275)
EK 6 L	19,28	(8,10)	(164)	21,17	(9,72)	(36)	(−0,211)
EK 6 H	18,45	(8,08)	(164)	18,67	(7,74)	(36)	(−0,028)

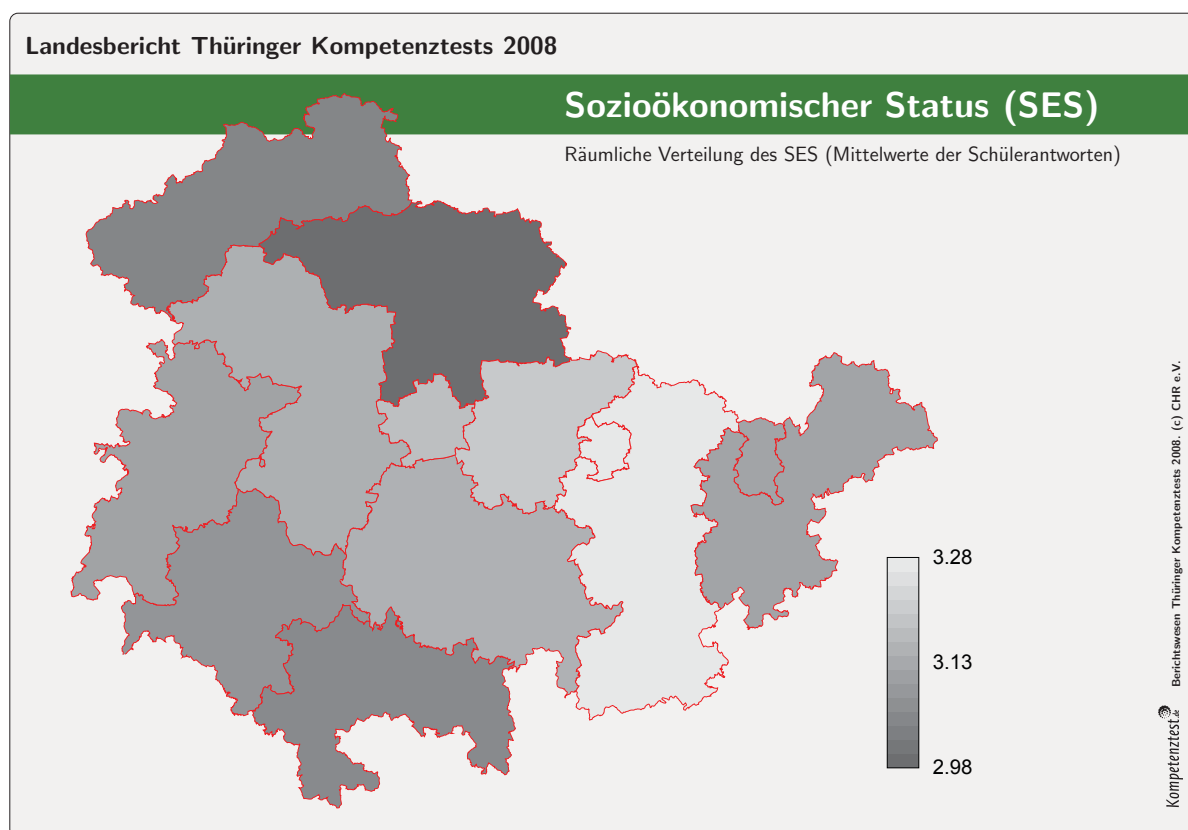
Der Mittelwertevergleich in der Klassenstufe 6 und 8 weist darauf hin, dass sich Schüler in Abhängigkeit vom SES insbesondere im Kompetenztest Mathematik und Deutsch in ihren Leistungen voneinander unterscheiden. Jugendliche aus Familienhäusern mit hohem SES schneiden durchschnittlich besser ab als Jugendliche mit niedrigem SES. Es fällt weiterhin auf, dass die Leistungsdifferenz in Gymnasien und Regelschulen im Mittel größer ist als in den Förderschulen. Möglicherweise unterscheiden sich Förderschüler mit niedrigem bzw. hohem SES deshalb nicht so stark voneinander, weil die Schüler aus Elternhäusern mit hohem SES besonders „schwierige“ Fälle sind. Es könnte sein, dass bei speziellen, stark ausgeprägten Leistungsschwächen der soziale Hintergrund weniger Einfluss hat, da es schwieriger ist, solche Schüler angemessen zu unterstützen und zu fördern.

Die Abbildungen 7.22, 7.23 und 7.24 veranschaulichen die geographische Verteilung des sozio-ökonomischen Status (SES) innerhalb Thüringens. Die dargestellten Mittelwerte wurden durch Umkodieren der fünf möglichen Antworten der Bücherfrage auf eine Skala von 1 bis 5 ermittelt und dienen nur der Veranschaulichung. Dabei bedeuten höhere Werte einen höheren SES. Dargestellt sind jeweils die aggregierten Werte auf Ebene der Schulamtsbereiche. Da

Tabelle 7.18: Mittelwertevergleich der Testergebnisse der Klassenstufe 8 bezüglich des sozioökonomischen Status unter Berücksichtigung der Schulart

	SES niedrig			SES hoch			Effektstärke (d)
	MW	(SD)	(N)	MW	(SD)	(N)	
Gymnasium							
MK 8	38,92	(6,38)	(492)	42,09	(6,50)	(2 593)	(–0,492)
Regelschule							
MK 8	28,84	(7,06)	(2 547)	32,78	(7,31)	(1 537)	(–0,548)
Förderschule							
MK 8	23,88	(6,72)	(198)	26,59	(7,77)	(37)	(–0,373)

der sozioökonomische Status der Eltern eine wichtige Einflussvariable auf die Schülerleistung darstellt, verdeutlicht die Abbildung die Problematik eines angestrebten fairen Vergleichs von Schul- und Klassenergebnissen. Merkmale wie der sozioökonomische Status sollten dabei unbedingt berücksichtigt werden.

**Abbildung 7.22:** Räumliche Verteilung des SES bei den Schülern der Klassenstufe 3

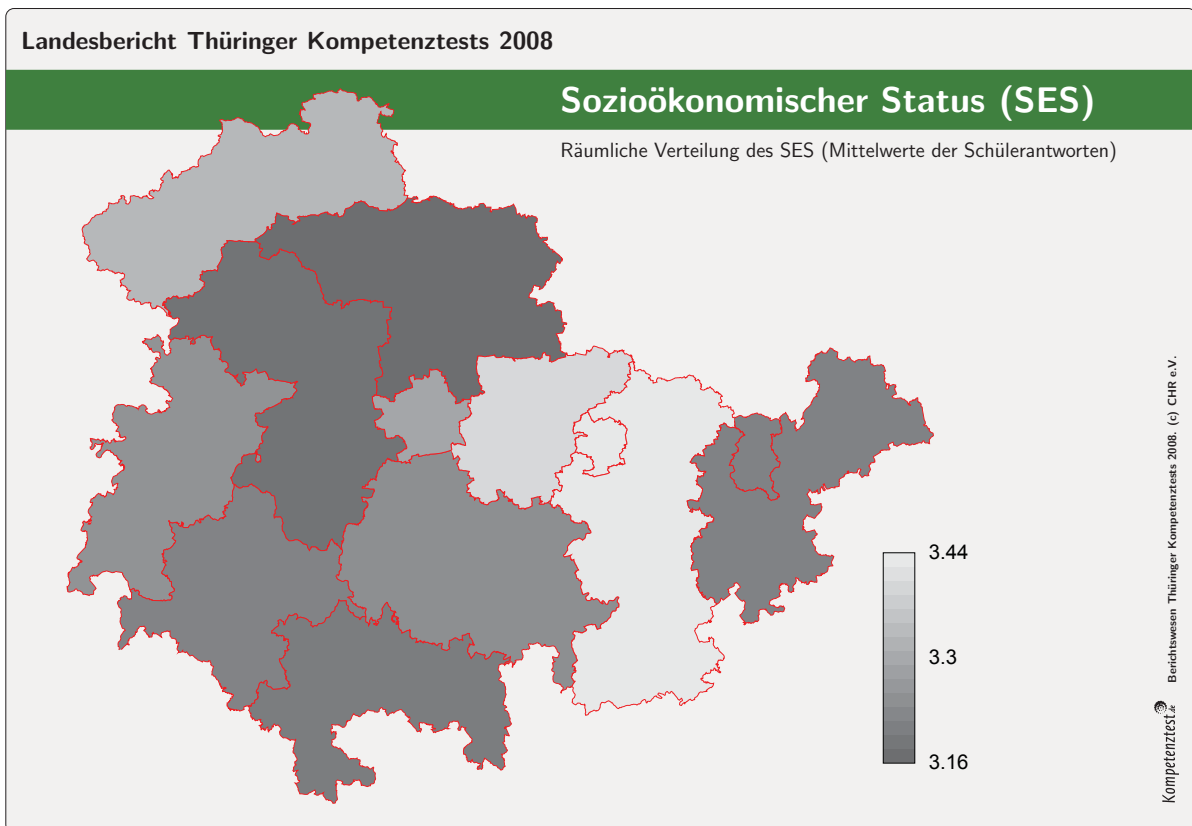


Abbildung 7.23: Räumliche Verteilung des SES bei den Schülern der Klassenstufe 6

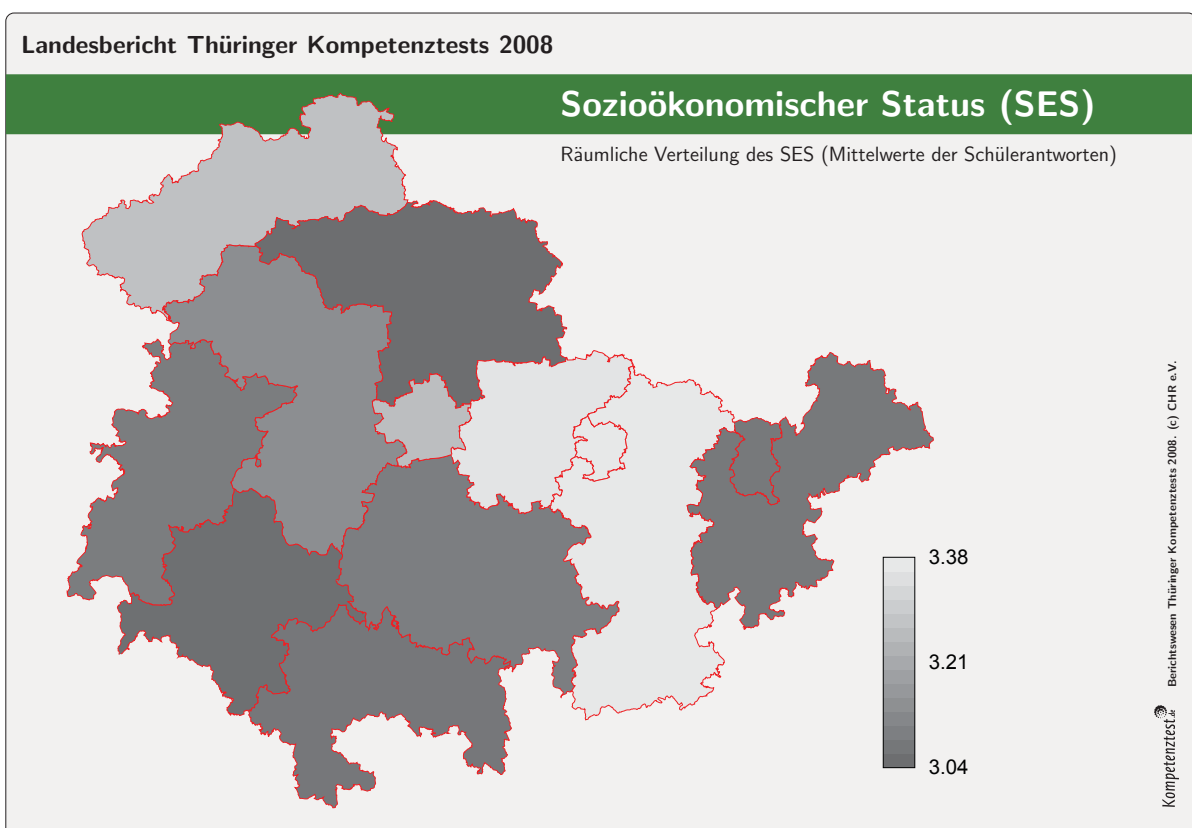


Abbildung 7.24: Räumliche Verteilung des SES bei den Schülern der Klassenstufe 8

7.2.4 Jahr des Schulbesuchs

Dieser Abschnitt beschäftigt sich damit, inwiefern sich die Kompetenz von Schülern unterscheidet, welche aufgrund guter oder schlechter Schulleistung ein Jahr übersprungen, ein Jahr wiederholt oder die Schule gewechselt haben. Außerdem soll näher darauf eingegangen werden, wie durchlässig die einzelnen Schularten sind, d. h. beispielsweise inwieweit haben Regelschüler das Potential mit der Leistung von Gymnasiasten mitzuhalten.

Überspringen einer Klassenstufe

Die Analysen zum Überspringen einer Klassenstufe beziehen sich im Rahmen dieses Berichts nur auf Schüler der Grundschulen. Es zeigt sich, dass Schüler, die eine Klassenstufe übersprungen haben, wie in den letzten Jahren bessere Ergebnisse im Kompetenztest erzielen als ihre Mitschüler. Es liegt sowohl im Kompetenztest Mathematik, als auch in Deutsch ein kleiner Effekt vor. Das bessere Abschneiden der Überspringer erscheint logisch, wenn man bedenkt, dass nur Schüler mit herausragender Leistung eine Klassenstufe überspringen können. Man könnte vermuten, dass jene Schüler außerschulisch, z. B. von den Eltern, eine zusätzliche Anregung und Motivation zum Lernen bekommen.

Tabelle 7.19: Mittelwertevergleich der Testergebnisse Klassenstufe 3 zwischen Schülern mit 3 Jahren Schulbesuch und Schülern, die eine Klassenstufe übersprungen haben

	Überspringer			3 Jahre Schulbesuch			Effektstärke
	MW	(SD)	(N)	MW	(SD)	(N)	(d)
Grundschule							
DK 3	33,40	(8,02)	(57)	30,56	(8,49)	(13 612)	(0,344)
MK 3	30,06	(9,61)	(49)	26,28	(8,16)	(13 425)	(0,424)

Wiederholung einer Klassenstufe

In der Klassenstufe 3 zeigt sich, dass Grundschüler, welche keine Klasse wiederholt haben, wesentlich bessere Ergebnisse im Kompetenztest erzielen als die Schüler, welche mindestens eine Klassenstufe wiederholten. Sowohl in Mathematik als auch in Deutsch liegen wie im letzten Jahr große Effekte vor. Die Effekte könnten insofern erklärbar sein, dass besonders leistungsschwächere Schüler in der Grundschule noch keinen ausreichenden Förderbedarf bekommen. Eine weitere Erklärung, warum die Effekte in der Grundschule verhältnismäßig groß sind, könnte an der Klassenzusammensetzung in Grundschulen ansetzen. In einer Klasse sitzen hier sowohl jene Schüler, welche später auf das Gymnasium gehen, als auch jene Schüler, welche später eine Förderschule besuchen. Demnach können die leistungsschwächeren Schüler eher in die Situation kommen, eine Klassenstufe wiederholen zu müssen. Dagegen schneiden die Wiederholer in Förderschulen wie auch 2007 nur leicht schwächer ab als ihre Mitschüler. Vielleicht liegt hier nur ein kleiner Effekt vor, weil der Förderbedarf einzelner Schüler in der Förderschule stärker im Mittelpunkt steht als in Grundschulen. Dass es in Förderschulen von besonderer Bedeutung ist, Wiederholer in den Unterricht einzubeziehen, zeigt die Verteilung der Schüler: In Klassenstufe 3 der Förderschulen sind wesentlich mehr Schüler, welche schon ein Jahr wiederholten, als Schüler, welche regulär seit drei Jahren zur Schule gehen.

Im Vergleich zu den anderen Schularten der Klassenstufe 6, ist im Gymnasium mit 40 Schülern eine eher geringe Zahl von Wiederholern zu verzeichnen. Eventuell ist dies dadurch erklärbar, dass Schüler der Klassenstufe 6, die auf dem Gymnasium eher schlechte Leistungen erzielen,

Tabelle 7.20: Mittelwertevergleich der Testergebnisse Klassenstufe 3 zwischen Schülern mit Klassenwiederholung und 3 Jahren Schulbesuch

	Wiederholer			3 Jahre Schulbesuch			Effektstärke
	MW	(SD)	(N)	MW	(SD)	(N)	(d)
Grundschule							
DK 3	22,46	(8,39)	(1 220)	30,60	(8,45)	(13 558)	(−0,966)
MK 3	19,02	(7,25)	(1 154)	26,31	(8,15)	(13 377)	(−0,945)
Förderschule							
DK 3	18,59	(9,69)	(233)	20,85	(11,03)	(54)	(−0,218)
MK 3	16,70	(8,67)	(225)	18,40	(9,25)	(48)	(−0,190)

Tabelle 7.21: Mittelwertevergleich der Testergebnisse Klassenstufe 6 zwischen Schülern mit und ohne Klassenwiederholung

	Wiederholer			6 Jahre Schulbesuch			Effektstärke
	MW	(SD)	(N)	MW	(SD)	(N)	(d)
Gymnasium							
DK 6	66,44	(12,51)	(36)	72,14	(11,25)	(5 464)	(−0,479)
MK 6	17,43	(5,26)	(40)	19,41	(4,15)	(5 485)	(−0,418)
EK 6 L	41,47	(6,87)	(36)	40,83	(6,21)	(5 377)	(0,098)
EK 6 H	36,28	(4,90)	(36)	35,60	(4,58)	(5 376)	(0,143)
Regelschule							
DK 6	47,00	(14,70)	(470)	55,06	(14,38)	(5 786)	(−0,554)
MK 6	10,92	(4,52)	(491)	13,51	(14,68)	(5 786)	(−0,563)
EK 6 L	24,95	(8,36)	(478)	29,04	(8,05)	(5 667)	(−0,498)
EK 6 H	24,70	(7,49)	(477)	27,75	(6,78)	(5 652)	(−0,427)
Förderschule							
DK 6	45,24	(12,57)	(68)	45,87	(13,144)	(240)	(−0,049)
MK 6	9,63	(3,99)	(70)	9,87	(4,93)	(265)	(−0,053)
EK 6 L	20,08	(6,56)	(65)	20,25	(8,62)	(252)	(−0,022)
EK 6 H	19,18	(7,445)	(65)	19,63	(8,45)	(252)	(−0,057)

es vorziehen auf eine Regelschule zu wechseln, anstatt die Klassenstufe zu wiederholen. Dafür könnte auch die später berichtete, erhöhte Anzahl von Schulwechslern in der Regelschule sprechen. Die verbleibenden Wiederholer im Gymnasium schneiden im Kompetenztest Mathematik und Deutsch im Mittel schwächer ab als die Mitschüler. Dagegen zeigt sich in Englisch, dass die Wiederholer tendenziell ein leicht besseres Ergebnis erzielen. Auffallend ist, dass die wiederholenden Schüler in Gymnasien in diesem Jahr deutlich schwächer abschneiden als die Mitschüler. Im letzten Jahr zeigten Wiederholer noch in etwa gleich gute Leistungen bei den Kompetenztests Deutsch und Mathematik. Aufgrund der geringen Anzahl der Schüler, welche schon eine Klassenstufe wiederholten, ist die Durchschnittsleistung der Wiederholer abhängig vom Abschneiden einzelner Schüler. Es wäre folglich denkbar, dass sich der Unterschied zwi-

Tabelle 7.22: Mittelwertevergleich der Testergebnisse Klassenstufe 8 zwischen Schülern mit und ohne Klassenwiederholung

	Wiederholer			8 Jahre Schulbesuch			Effektstärke
	MW	(SD)	(N)	MW	(SD)	(N)	(d)
Gymnasium							
MK 8	37,13	(6,60)	(92)	41,04	(6,54)	(4 541)	(–0,641)
Regelschule							
MK8	27,23	(6,86)	(984)	31,13	(7,14)	(5 896)	(–0,557)
Förderschule							
MK8	25,24	(7,84)	(118)	24,33	(6,59)	(228)	(0,126)

schen den wiederholenden und den Schülern in der regulären Schulzeit im nächsten Jahr anders gestaltet. In den Regelschulen zeichnet sich ab, dass Wiederholer wie auch 2007 in allen Kompetenztests schwächer abschnitten als Schüler, die regulär 6 Jahre die Schule besuchten. Dagegen unterscheiden sich Wiederholer in den Förderschulen in den Kompetenztests im Mittel nicht vom Rest der Klasse.

Ein ähnliches Bild findet man in der Klassenstufe 8 vor: In Gymnasien und Regelschulen schneiden Wiederholer mit einem mittleren Effekt schwächer ab als die Mitschüler. In Förderschulen unterscheidet sich die Leistung mit einem Trend insofern, als dass Wiederholer in Mathematik leicht besser abschnitten.

Die Ergebnisse weisen daraufhin, dass Wiederholer in Förderschulen eher in den Leistungsstand der Klasse integriert werden. Sie schneiden bei den Kompetenztests mindestens genauso gut ab wie Schüler, die keine Klassenstufe wiederholt haben, wenn nicht sogar etwas besser. Wiederholer der anderen Schularten schneiden auch nach ihrer Wiederholung schwächer ab als das Mittel der Klassenstufe, d. h. es besteht zum Teil anscheinend mehr Förderbedarf als tatsächlich angeboten wird.

In den Grafiken 7.25 bis 7.32 ist die räumliche Verteilung von Klassenwiederholern in Thüringen dargestellt. Wie aus den Abbildungen ersichtlich wird, unterscheiden sich die einzelnen Regionen zum Teil deutlich hinsichtlich des prozentualen Anteils von Wiederholern.

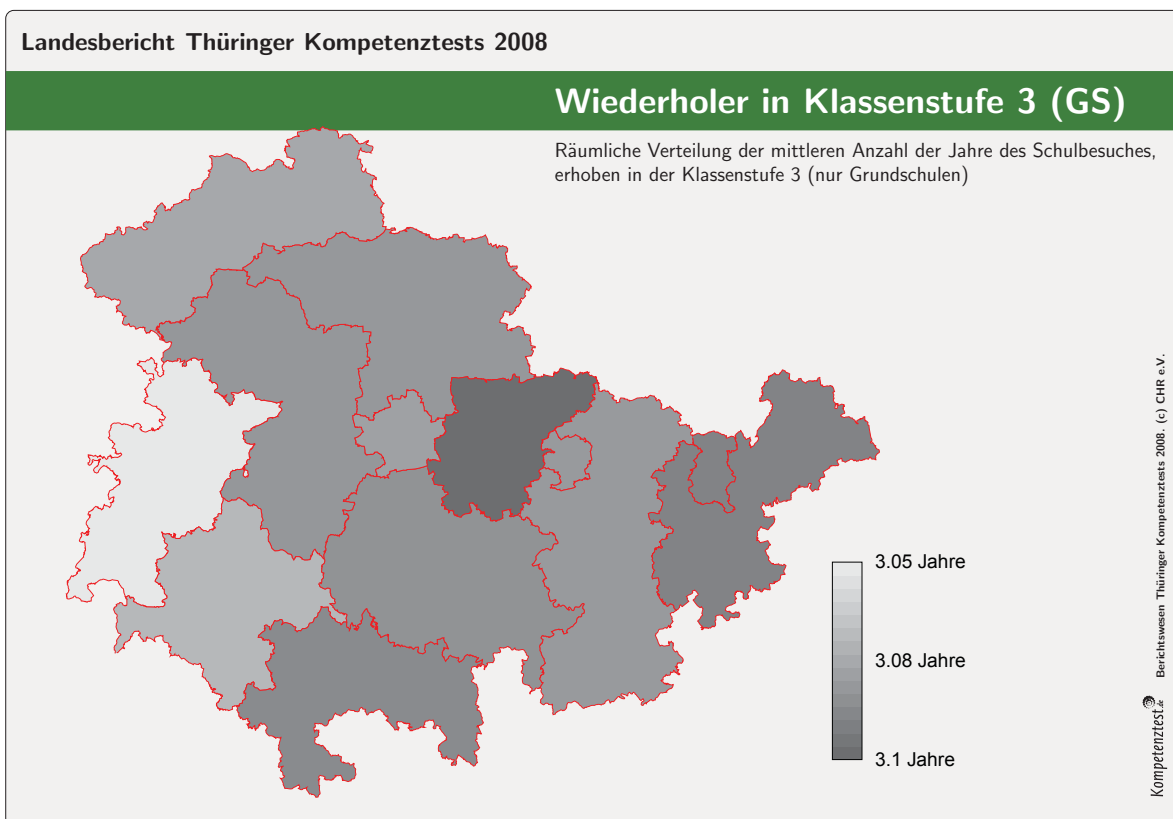


Abbildung 7.25: Räumliche Verteilung der mittleren Anzahl der Jahre des Schulbesuches, erhoben in der Klassenstufe 3 (nur Grundschulen)

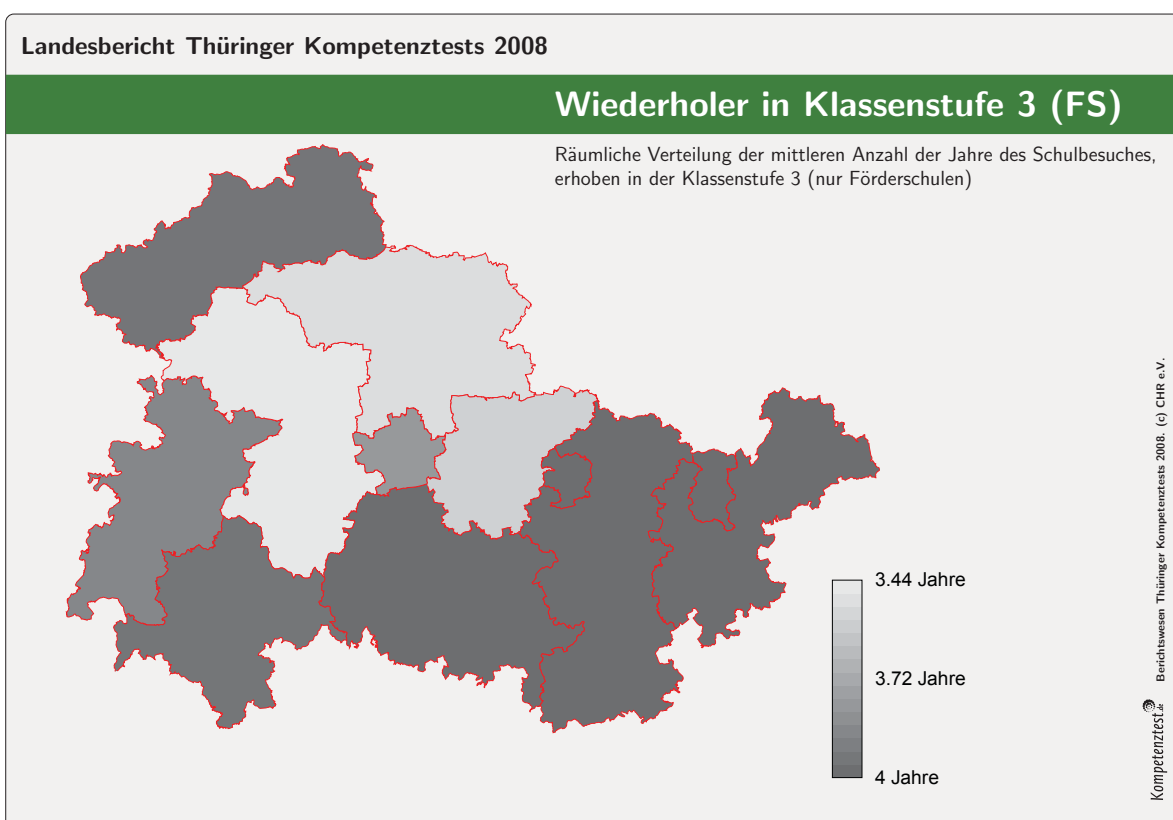


Abbildung 7.26: Räumliche Verteilung der mittleren Anzahl der Jahre des Schulbesuches, erhoben in der Klassenstufe 3 (nur Förderschulen). Für den Schulamtsbezirk Neuhaus können aufgrund der geringen Fallzahlen keine Angaben gemacht werden.

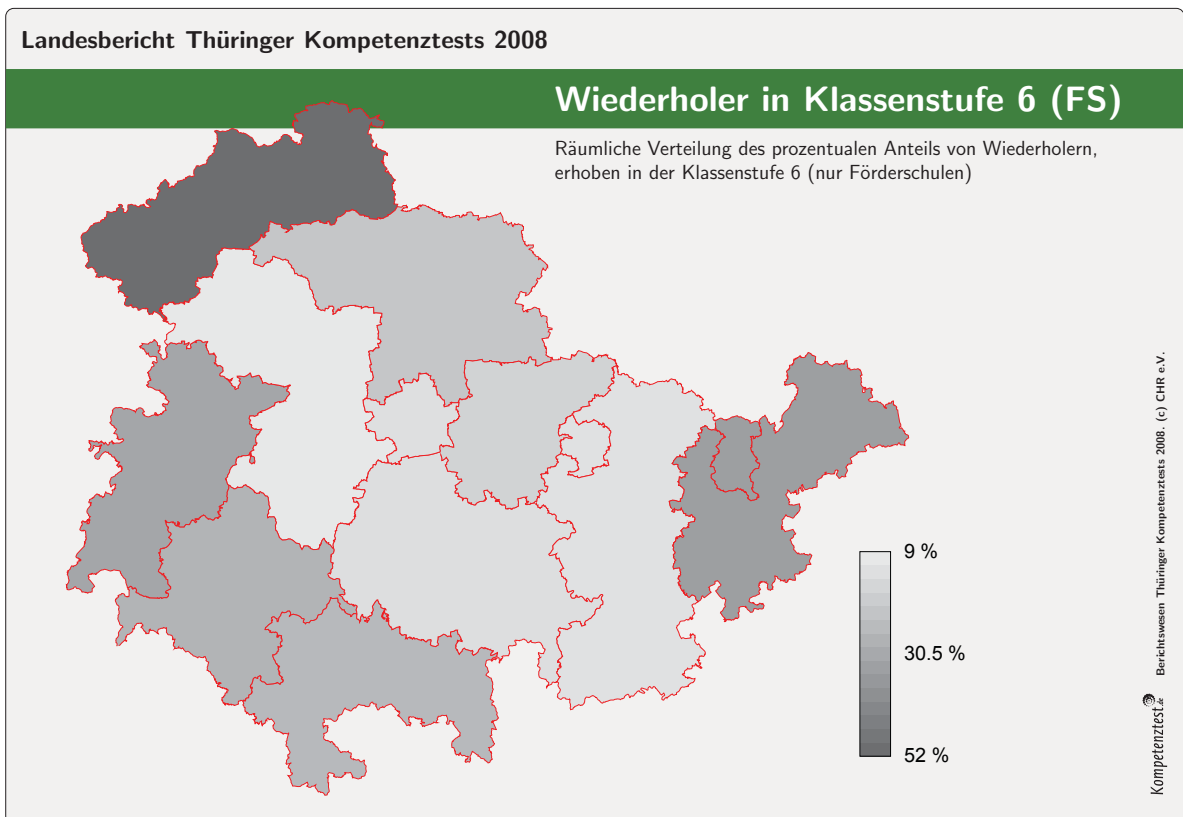


Abbildung 7.27: Räumliche Verteilung des prozentualen Anteils von Wiederholern, erhoben in der Klassenstufe 6 (nur Förderschulen)

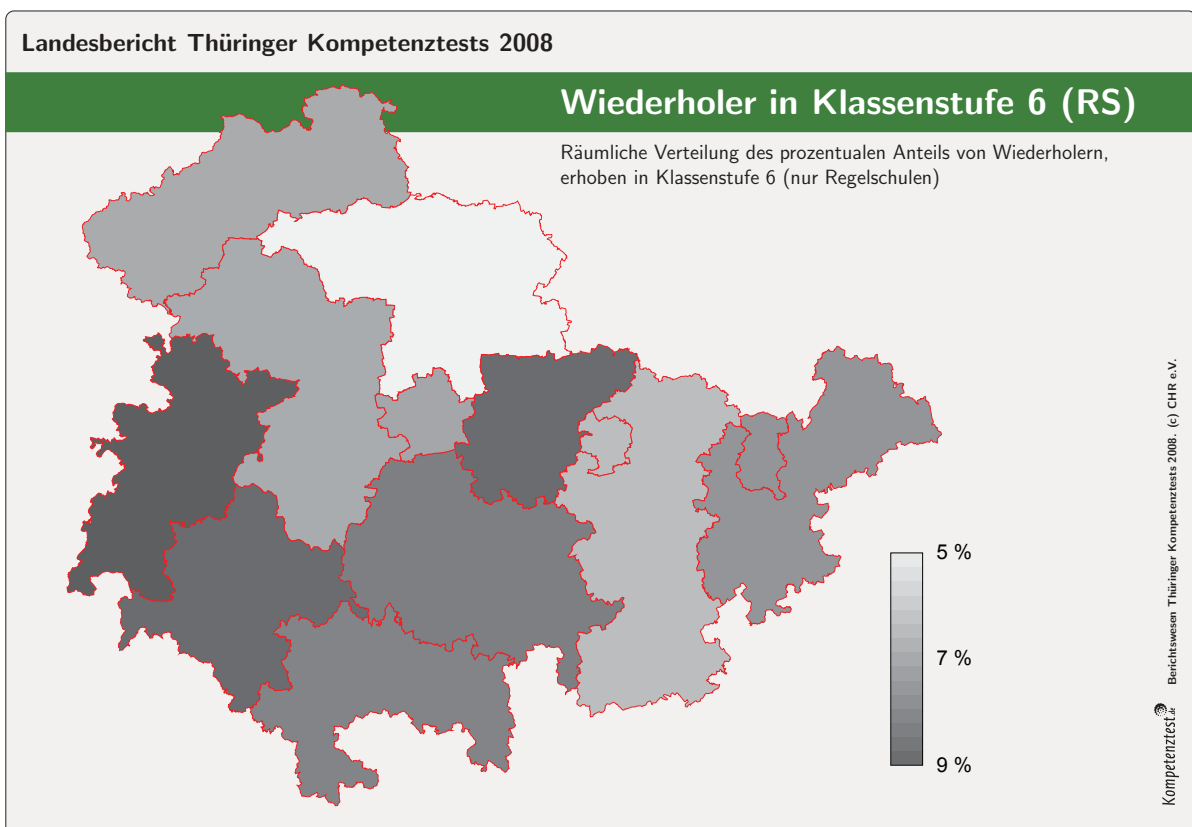


Abbildung 7.28: Räumliche Verteilung des prozentualen Anteils von Wiederholern, erhoben in Klassenstufe 6 (nur Regelschulen)

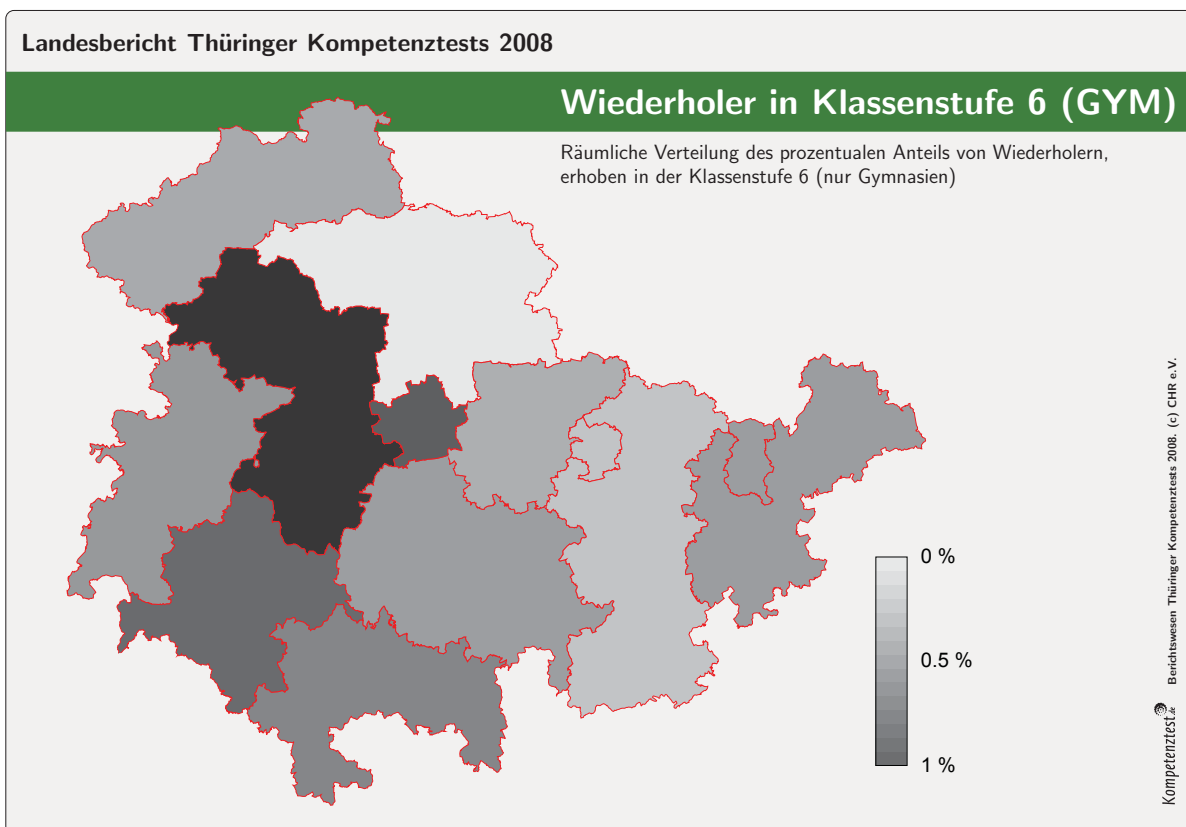


Abbildung 7.29: Räumliche Verteilung des prozentualen Anteils von Wiederholer, erhoben in der Klassenstufe 6 (nur Gymnasien)

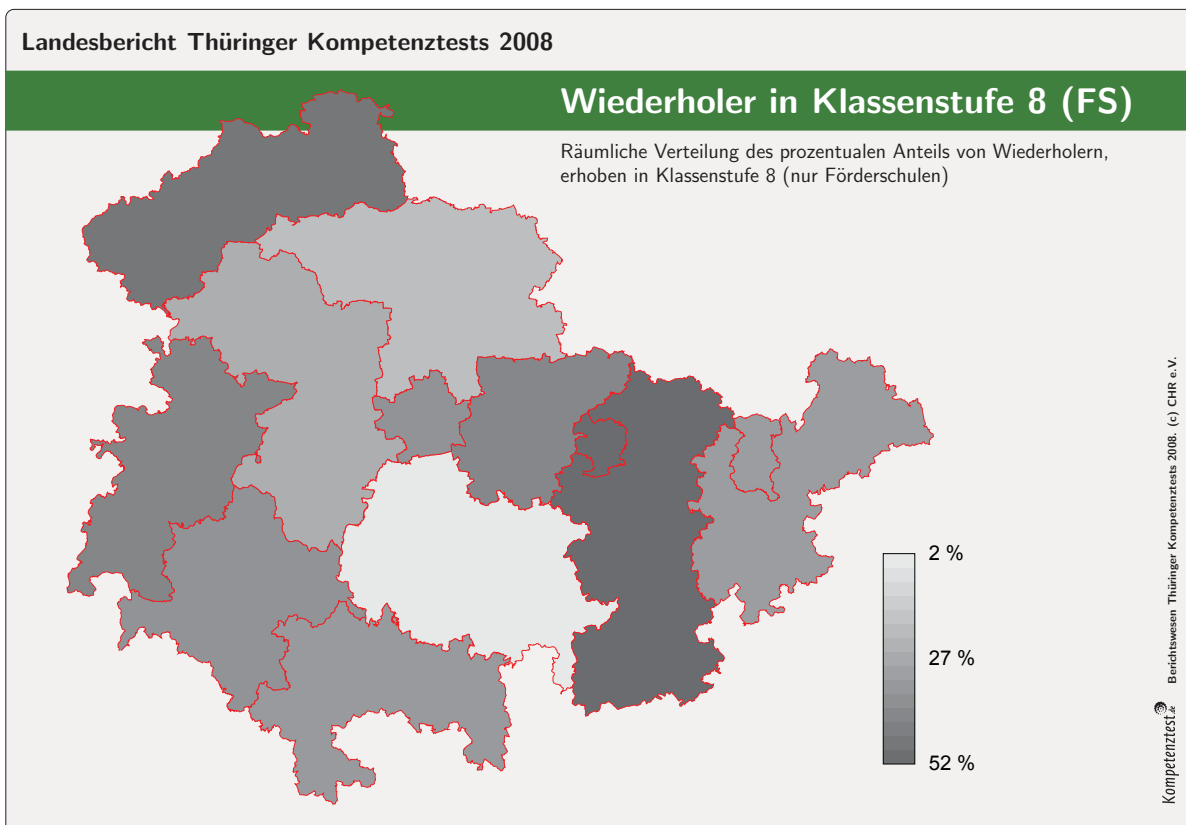


Abbildung 7.30: Räumliche Verteilung des prozentualen Anteils von Wiederholern, erhoben in Klassenstufe 8 (nur Förderschulen)

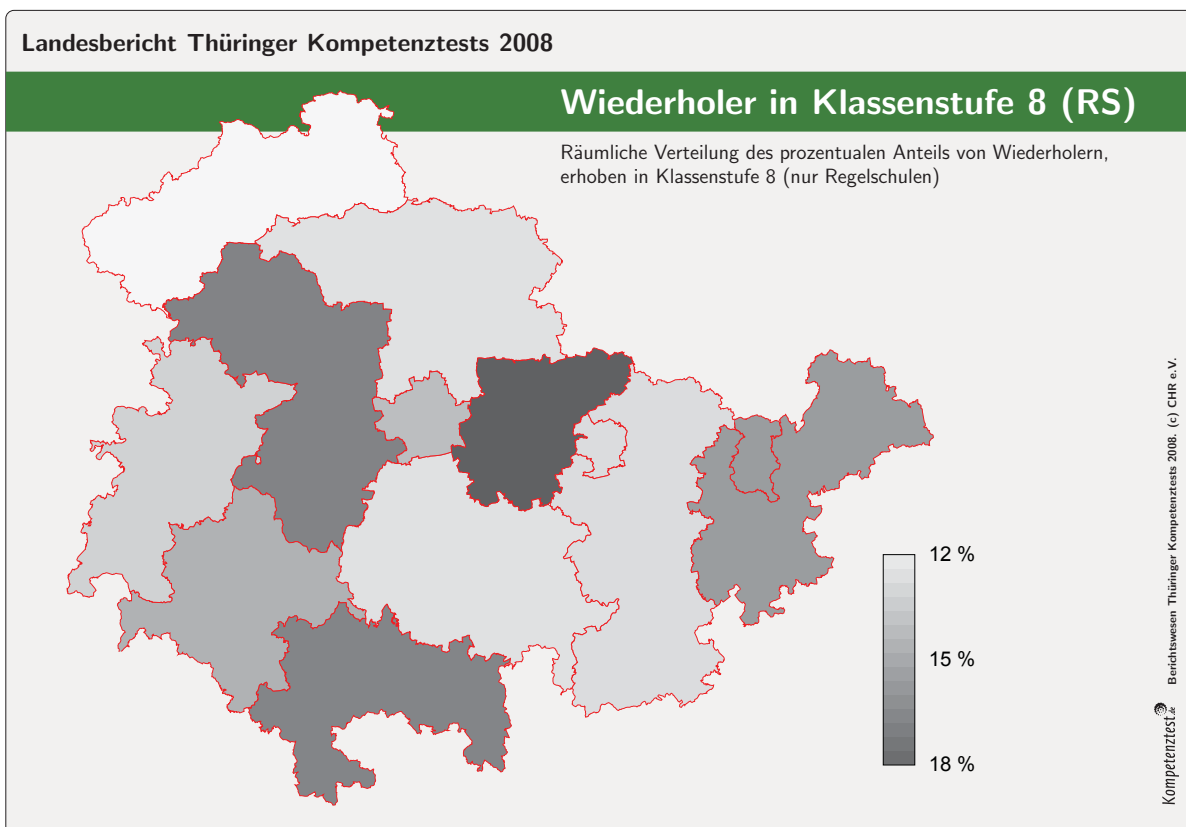


Abbildung 7.31: Räumliche Verteilung des prozentualen Anteils von Wiederholern, erhoben in Klassenstufe 8 (nur Regelschulen)

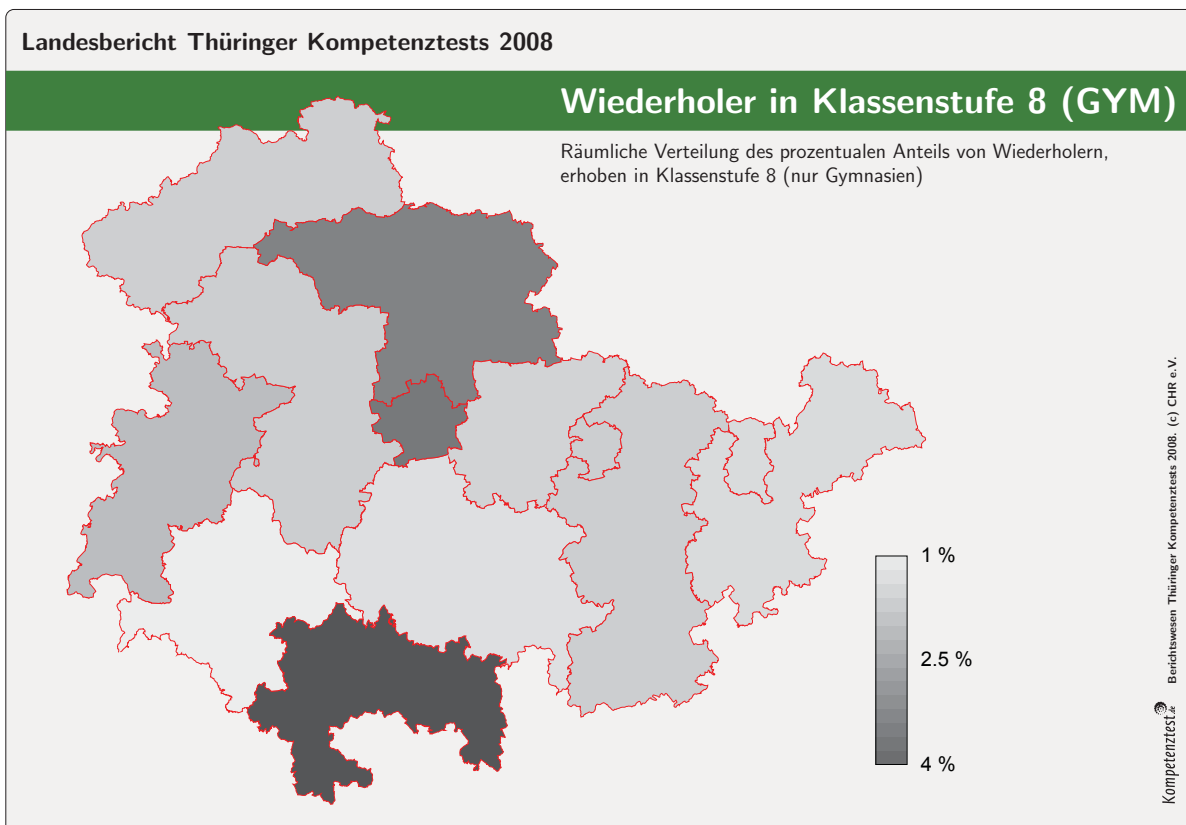


Abbildung 7.32: Räumliche Verteilung des prozentualen Anteils von Wiederholern, erhoben in Klassenstufe 8 (von Gymnasien)

Schulwechsel

Im Zuge des Kompetenztests der Klassenstufe 8 wird erhoben, wie viele Schüler der einzelnen Schularten die Schule mindestens einmal gewechselt haben. Im Gymnasium fällt auf, dass die Schüler, welche keinen Schulwechsel hinter sich haben und demzufolge seit der Klassenstufe 5 das Gymnasium besuchen, bessere Leistungen zeigen als Schüler, welche zum Gymnasium wechselten. Dies könnte an dem Anforderungsunterschied liegen, an welchen sich Regelschüler, die auf das Gymnasium wechseln, erst gewöhnen müssen. In Regelschulen unterscheiden sich Schulwechsler nicht bedeutend vom Rest der Schüler. Es gibt jedoch die Tendenz, dass Schulwechsler leicht besser abschneiden als die Mitschüler. Dies erscheint logisch, weil viele der gewechselten Schüler vom Gymnasium kommen und damit eventuell an eine höhere Leistungsanforderung gewöhnt sind. Schulwechsler in Förderschulen zeigen im Durchschnitt die gleiche Leistung wie die anderen Förderschüler. Das zeigt, dass sich die wechselnden Schüler in den Leistungsstand der neuen Schule gut integriert haben und entsprechend beim Lernen von Unterrichtsinhalten mithalten können.

Tabelle 7.23: Mittelwertevergleich der Testergebnisse Klassenstufe 8 zwischen Schülern mit und ohne Schulwechsel

	Schulwechsler			kein Schulwechsler			Effektstärke
	MW	(SD)	(N)	MW	(SD)	(N)	(d)
Gymnasium							
MK 8	38,18	(7,39)	(85)	41,01	(6,54)	(4 548)	(-0,406)
Regelschule							
MK8	31,49	(6,99)	(327)	30,52	(7,24)	(6 553)	(0,136)
Förderschule							
MK 8	24,73	(6,63)	(44)	24,63	(7,11)	(302)	(0,015)

Durchlässigkeit der Schularten

Eine weitere interessante Fragestellung ist, inwieweit es Schüler in Förder- und Regelschulen gibt, welche eventuell das Potential besitzen auch einen anderen Bildungsweg zu bewältigen. Dazu wird einmal betrachtet, wie viele Regelschüler besser abschneiden als das Drittel der Gymnasialschüler, welche im Kompetenztest die schwächste Leistung zeigten. Zum anderen wird analysiert wie viele Förderschüler prozentual eine bessere Kompetenztestleistung erzielen als das schwächste Drittel der Regelschule. Für welche Anzahl von Regelschülern bzw. Förderschülern wäre es also eventuell lohnenswert einen anderen Bildungsgang einzuschlagen bzw. dorthin zu wechseln?

Tabelle 7.24 gibt an wie viel Prozent von den über 6 200 Regelschülern und den 320 Förderschülern in den jeweiligen Fächern besser abschneiden als das schwächste Drittel der Schüler des jeweils anderen Bildungsgangs. Es zeigt sich, dass mindestens ein Zehntel der Regelschüler im Kompetenztest besser abschnitt als das leistungsschwächste Drittel der Gymnasiasten. In den Förderschulen sind es im Kompetenztest Deutsch sogar bis zu 37 % der Schüler, welche eine bessere Leistung erzielten als das leistungsschwächste Drittel der Regelschüler. Die Berechnungen in der Klassenstufe 8 ergeben, dass von den über 6 553 Regelschülern 14,3 % besser im Kompetenztest Mathematik abschneiden als das leistungsschwächste Drittel der Schüler im Gymnasium. Die Daten weisen darauf hin, dass einige Schüler durchaus die Kompetenz und das Potential zu einem anderen Bildungsgang haben. Dabei muss die Frage offen

bleiben, inwieweit der einzelne Schüler im betreffenden Fach und darüber hinaus in den anderen, nicht getesteten Fächern die Anforderungen des anderen Bildungsganges tatsächlich bewältigen würde.

Tabelle 7.24: Durchlässigkeit der Schularten für die Klassenstufe 6 in Prozent

	DK6	MK6	EK6 L	EK6 H
RS-Gym	15,6 %	14,5 %	10 %	15,9 %
FS-RS	37,7 %	13,4 %	23,5 %	24,5 %

Die Ergebnisse können ein Anlass sein, zu diskutieren, ob der Bildungsweg, welcher schon im Alter von etwa zehn Jahren eingeschlagen wird, für einige Schüler tatsächlich der richtige ist. Die Analysen zum Schulwechsel zeigen, dass nur wenige Schüler ins Gymnasium wechseln. Es kommt seltener zu einem Wechsel auf das Gymnasium als vom Gymnasium zur Regelschule. Die Analysen zur Durchlässigkeit der Schularten lassen aber vermuten, dass einige Schüler das Potential besitzen einen anderen Bildungsweg einzuschlagen.

7.2.5 Besondere Lernschwierigkeiten

Auch in diesem Jahr wurden diagnostizierte Lernschwierigkeiten in die Analyse der Kompetenztests einbezogen. Wie in den Jahren zuvor wird dabei differenziert zwischen: „besonderen Lernschwierigkeiten“ im Bereich Rechnen (BLR), im Bereich Schriftspracherwerb (BLS) sowie aufgrund von Problemen im Verhalten (BLV) und „Sonderpädagogischem Förderbedarf“ mit den Förderschwerpunkten: Hören (SF1), Sehen (SF2), körperliche und motorische Entwicklung (SF3), Lernen (SF4), Sprache (SF5) bzw. emotionale und soziale Entwicklung (SF6).

Die Fachlehrer für Mathematik, Deutsch und Englisch sollten bei der Dateneingabe angeben, ob bei einem Schüler eine besondere Lernschwierigkeit oder ein sonderpädagogischer Förderbedarf vorliegt. Die Angaben der Mathematik-, Deutsch- und Englischlehrer sind in den Kreuztabellen 7.25 und 7.26 gegenübergestellt. Die Darstellung beschränkt sich auf den Vergleich der Angaben von Mathematik- und Deutschlehrern in der Klassenstufe 3 und die Gegenüberstellung der Angaben der Fachlehrer für Englisch und Deutsch in der Klassenstufe 6. Der Vergleich der Angaben der Mathematik- und Deutschlehrer der Klassenstufe 6 unterscheidet sich in seiner Aussage nicht, deshalb wird aus Übersichtsgründen keine grafische Darstellung angeführt. In der hervorgehobenen Diagonale ist die Anzahl der Schüler abgetragen, bei denen die Angaben der Fachlehrer übereinstimmen. Zum Beispiel stimmen die Angaben der Deutsch- und Mathelehrer zu besonderen Lernschwierigkeiten im Bereich Schriftspracherwerb bei 341 Schülern der Klassenstufe 3 überein. Bei 139 Schülern der Klassenstufe 6, die laut den Angaben des Deutschlehrers eine besondere Lernschwierigkeit im Bereich Schriftspracherwerb haben, gab der entsprechende Englischlehrer an, dass weder eine besondere Lernschwierigkeit noch ein sonderpädagogischer Förderbedarf vorliegt.

Aus den Tabellen 7.25 und 7.26 geht hervor, dass besondere Lernschwierigkeiten einzelner Schüler nicht allen befragten Fachlehrern (Mathematik-, Deutsch- und Englischlehrern) gleichermaßen bekannt sind. Es ist sicherlich weniger gravierend, wenn ein Deutschlehrer nicht über die Rechenschwäche eines Schülers informiert ist. Bedenkt man aber, dass einige Englischlehrer von der Leistungsschwierigkeit beim Schriftspracherwerb ihrer Schüler wissen, jedoch nicht die Deutschlehrer, dann muss man vermuten, dass die Kommunikation unter den Fachlehrern noch verbesserungswürdig ist. Sowohl in Deutsch, als auch in Englisch spielt

Tabelle 7.25: Vergleich der Angaben von Deutsch- und Mathematiklehrern zu Besonderen Lernschwierigkeiten/ sonderpädagogischem Förderbedarf ihrer Schüler, Klassenstufe 3

		Deutsch					Gesamt
		keine	BLS	BLV	BLR	SF	
Mathematik	keine	13 326	300	34	14	82	13 756
	BLS	54	341	3	2	7	407
	BLV	13	17	125	1	8	164
	BLR	272	249	8	204	25	758
	SF	30	18	6	4	509	567
Gesamt		13 695	925	176	225	631	15 652

Tabelle 7.26: Vergleich der Angaben von Deutsch- und Englischlehrern zu Besonderen Lernschwierigkeiten/ sonderpädagogischem Förderbedarf ihrer Schüler, Klassenstufe 6

		Deutsch					Gesamt
		keine	BLS	BLV	BLR	SF	
Englisch	keine	12 022	139	41	13	52	12 267
	BLS	89	223	5	4	28	349
	BLV	43	10	87	–	10	150
	BLR	21	2	1	36	2	62
	SF	60	32	7	1	364	464
Gesamt		12 235	406	141	54	456	13 292

die Sprachkompetenz eine besondere Rolle. Es herrscht leider unter den genannten Fachlehrern keine Einigkeit darüber, inwieweit einzelne Schüler aufgrund einer besonderen Leistungsschwierigkeit nicht die gleichen Ausgangsvoraussetzungen haben wie die restlichen Schüler. Die einzelnen Fachlehrer sollten also unbedingt abgleichen, in welchen Fällen eine spezifische Lernschwierigkeit vorliegt, so dass die betroffenen Schüler gezielt gefördert werden können. Sogar bei Schülern mit sonderpädagogischem Förderbedarf gibt es, genau wie in den letzten Jahren, in vielen Fällen abweichende Angaben. Diese Ergebnisse sind bedenklich, da die Förderpläne von der Klassenkonferenz erarbeitet, sowie in regelmäßigen Zeitabständen überprüft werden sollten. Es wäre in diesem Zusammenhang also besonders wichtig, dass pädagogisches Handeln abgestimmt wird und unter den Fachlehrern ein einheitlicher Kenntnisstand zu Problembereichen einzelner Schüler vorherrscht.

In Anbetracht dieser unterschiedlichen Lehrerangaben wird bei der folgenden Darstellung von Gruppenunterschieden immer die Angabe des jeweiligen Fachlehrers als Differenzierungsgrundlage gewählt. Bei den Kompetenztests im Fach Mathematik werden die Schüler also beispielsweise nach den Angaben des Mathematiklehrers den Gruppen zugeordnet. In einigen Fällen ist die Anzahl der Schüler zu gering für eigene statistische Analysen. Dann werden keine Unterschiede berichtet. Dies betrifft zum Beispiel die Förderschüler der Klassenstufe 3.

Ein Mittelwertevergleich zwischen den Schülern mit besonderen Lernschwierigkeiten und den Schülern mit sonderpädagogischem Förderbedarf hat ergeben, dass sich die Leistungen der beiden Gruppen nicht bedeutsam voneinander unterscheiden. Aus diesem Grund wurden beide Gruppen zusammengefasst und mit den Kompetenztestleistungen der Schüler ohne besondere Lernschwierigkeiten verglichen. Die Ergebnisse dieser Analyse sind in Tabelle 7.27 zusammengefasst. Es zeigt sich, dass Grundschüler mit der Diagnose einer besonderen Lernschwierigkeit oder eines sonderpädagogischen Förderbedarfs deutlich schwächer abschneiden als die anderen Schüler. Die großen Effektstärken könnten dadurch erklärt werden, dass in der Grundschule sowohl spätere Gymnasiasten als auch spätere Förderschüler zusammen unterrichtet werden. Aufgrund der großen Leistungsspanne wäre es denkbar, dass nicht genug Zeit bleibt, um im Speziellen auf den Förderbedarf einzelner einzugehen. In den nachfolgenden sechsten und achten Klassen liegen zumeist mittlere Effekte vor, d. h. auch hier schneiden Schüler ohne Lernschwierigkeiten im Durchschnitt besser ab. Eine Ausnahme bildet hierbei die Klassenstufe 8 der Förderschulen. Die Fachlehrer haben hier angegeben, dass bei der Mehrzahl der Förderschüler eine besondere Lernschwierigkeit oder sonderpädagogischer Förderbedarf vorliegt. Das heißt, die Förderschule hat insbesondere den Auftrag die Schüler mit einer entsprechenden Diagnose speziell zu fördern. Das könnte den Trend erklären, dass jene Schüler mit Lernschwierigkeiten im Kompetenztest leicht besser abschneiden als die Schüler ohne entsprechende Diagnose.

Tabelle 7.27: Mittelwertevergleich der Testergebnisse zwischen Schülern ohne und mit besonderer Lernschwierigkeit oder sonderpädagogischem Förderbedarf

	MW	keine (SD)	(N)	MW	BL/SF (SD)	(N)	Effektstärke (d)
Grundschule							
DK 3	31,03	(8,14)	(13 211)	21,08	(8,36)	(1 624)	(1,206)
MK 3	26,61	(7,97)	(13 218)	18,02	(7,35)	(1 545)	(1,121)
Gymnasium Klassenstufe 6							
DK 6	72,17	(11,20)	(5 468)	63,63	(13,82)	(62)	(0,679)
MK 6	19,40	(4,16)	(5 487)	18,87	(4,28)	(38)	(0,126)
EK 6 L	40,87	(6,19)	(5 384)	36,34	(7,97)	(50)	(0,635)
DK 6 H	35,64	(4,54)	(5 392)	31,78	(7,05)	(50)	(0,651)
Regelschule Klassenstufe 6							
DK 6	55,38	(14,33)	(5 692)	46,25	(13,81)	(648)	(0,649)
MK 6	13,54	(4,67)	(5 665)	11,11	(4,64)	(612)	(0,522)
EK 6 L	29,24	(8,04)	(5 604)	23,88	(7,55)	(604)	(0,687)
EK 6 H	27,87	(6,78)	(5 604)	24,16	(6,97)	(606)	(0,539)
Regelschule Klassenstufe 8							
MK 8	30,70	(7,20)	(6 609)	27,49	(7,34)	(271)	(0,441)
Förderschule Klassenstufe 8							
MK 8	23,66	(6,02)	(35)	24,75	(7,15)	(311)	(-0,165)

7.2.6 Zusammenhang Halbjahresnote und Testergebnisse

Die Halbjahresnote als ein leistungsrelevantes Schülermerkmal kann in gewisser Weise als Ergebnis des Unterrichts gesehen werden. Sie ist kein dem Unterricht zeitlich vorgeordnetes Merkmal und wird daher nicht für die Berechnung der Vergleichswerte herangezogen. Von Interesse ist aber der Zusammenhang zwischen Halbjahresnote und Testergebnis, da aus deren Stärke die Ähnlichkeit bzw. Unähnlichkeit der Notengebungskriterien mit den im Test erhobenen Leistungen gezeigt werden kann. In den Abbildungen 7.33 bis 7.35 sind die Zusammenhänge zwischen der Halbjahresnote im jeweiligen Fach und der erreichten Punktzahl der Schüler dargestellt. Wiedergegeben sind Streudiagramme mit eingezeichneten Regressionsgeraden. Die Abbildungen 7.33 bis 7.39 zeigen, dass bei gleicher Halbjahresnote sehr unterschiedliche Punktzahlen im Kompetenztest erfolgten. Für diese unterschiedlichen Testergebnisse können sehr viele Faktoren verantwortlich sein. Einerseits müssen abgefragte Kompetenztestinhalte nicht notwendig mit den Unterrichtsinhalten übereinstimmen. Zum anderen setzt sich die Halbjahresnote aus mehreren Leistungsbewertungen zusammen, unter anderem auch mündlichen Leistungskontrollen. Folglich kann es sein, dass die erzielte Kompetenztestleistung nicht mit der Halbjahresnote übereinstimmt. Eine weitere Ursache hierfür könnte sein, dass Lehrer eine soziale Bezugsnormorientierung bevorzugen. Das heißt, dass sie die Leistungen der Schüler untereinander vergleichen und diesen Vergleich als Grundlage für die Notengebung verwenden. Damit würde ein Schüler mit einem mittleren Leistungsniveau in einer Klasse mit eher leistungsstärkeren Mitschülern schlechtere Notenbewertungen erhalten als in einer Klasse mit eher leistungsschwächeren Mitschülern. Trotzdem kann dieser Schüler im Kompetenztest überdurchschnittlich gut abschneiden. Es wäre andererseits auch denkbar, dass Schüler, die im Unterricht im Mittel recht gute Noten bekommt, tagesformabhängig ein schlechtes Ergebnis beim Kompetenztest erzielten. Dies sollte aber eher die Ausnahme als die Regel sein.

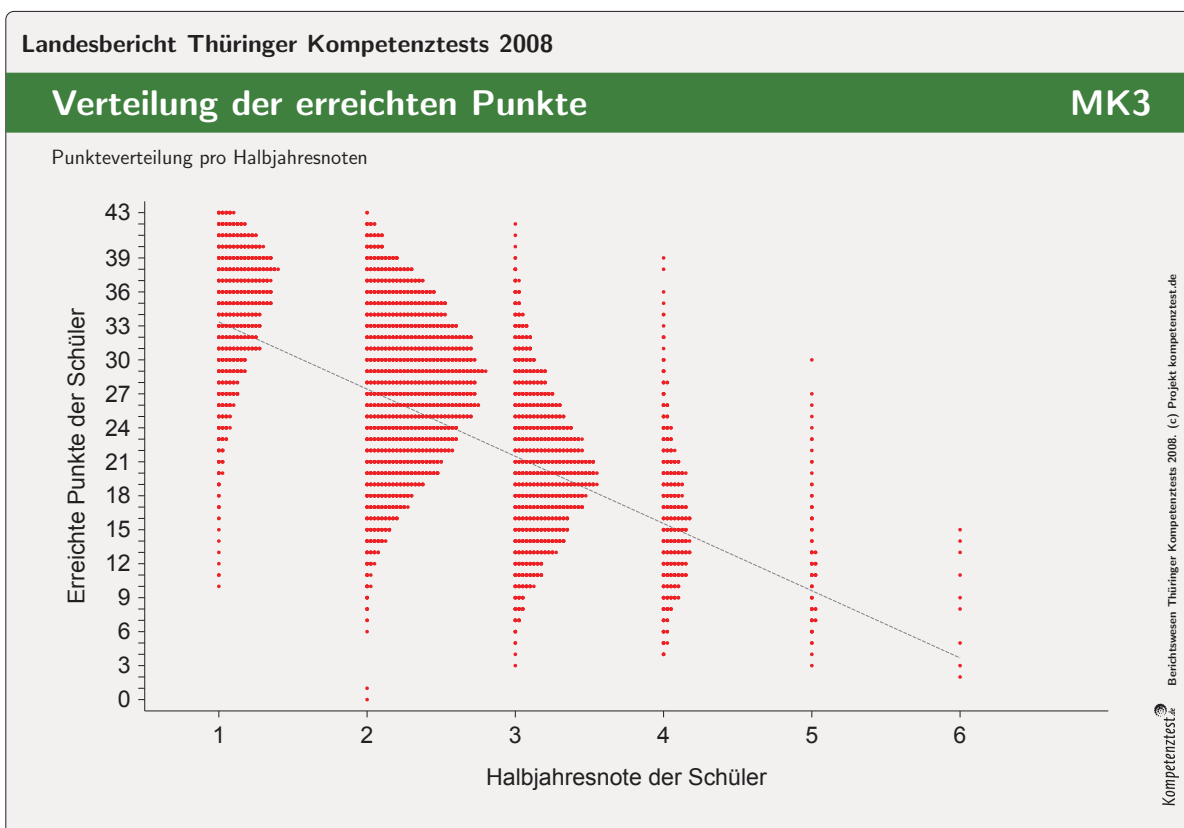


Abbildung 7.33: Zusammenhang zwischen Halbjahresnote und Testergebnis im Fach Mathematik, Klassenstufe 3, MK3

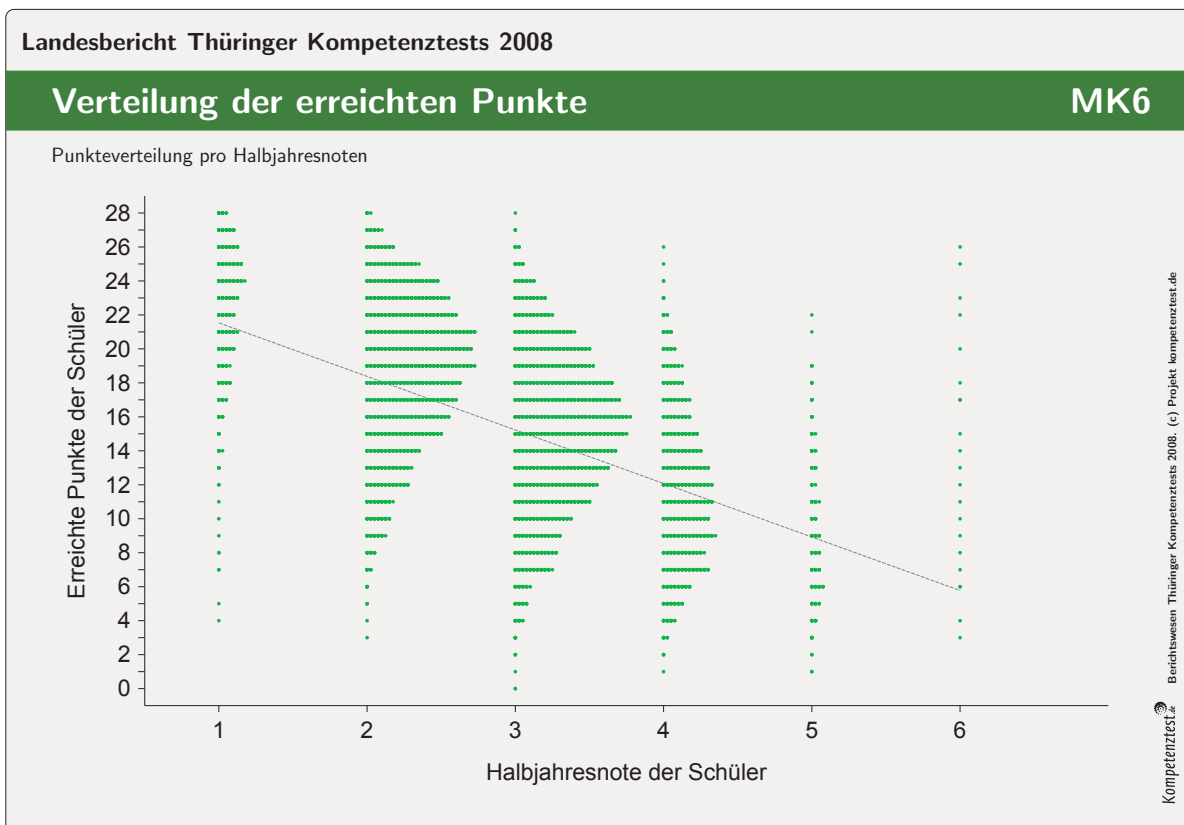


Abbildung 7.34: Zusammenhang zwischen Halbjahresnote und Testergebnis im Fach Mathematik, Klassenstufe 6, MK6

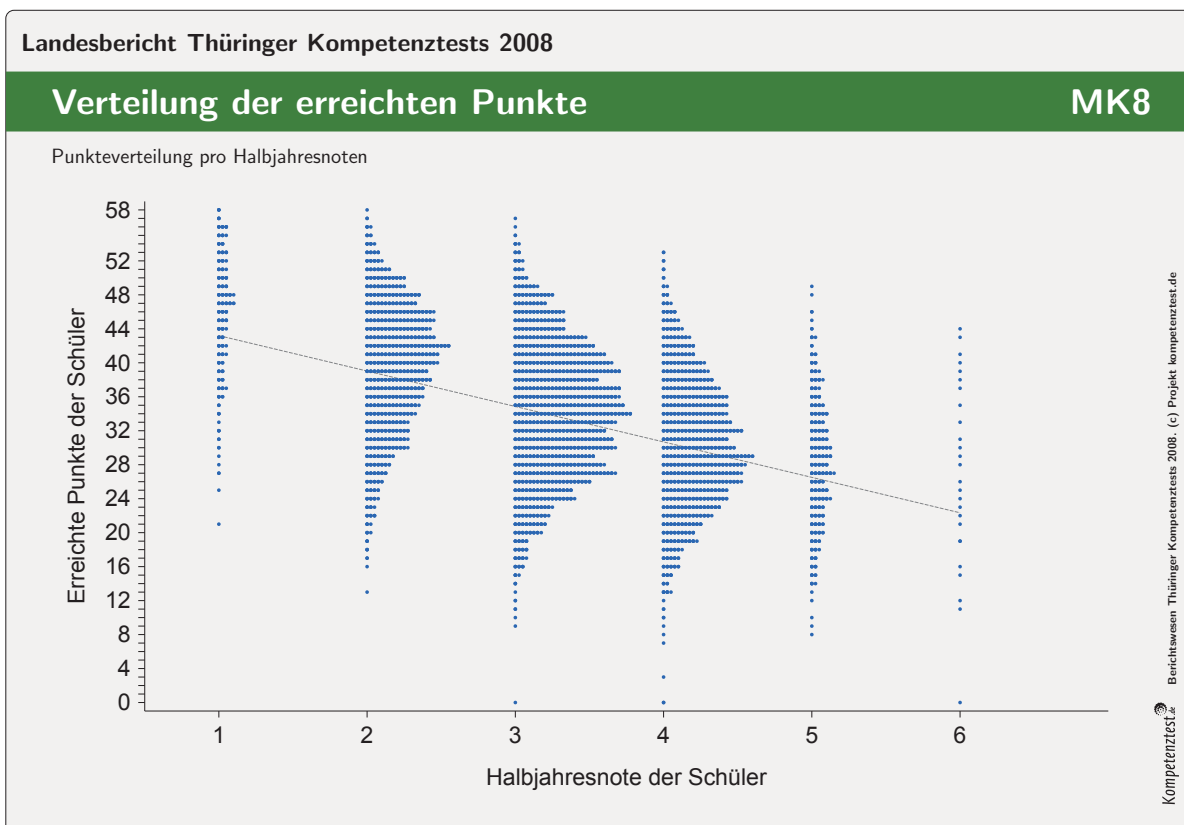


Abbildung 7.35: Zusammenhang zwischen Halbjahresnote und Testergebnis im Fach Mathematik, Klassenstufe 8, MK8

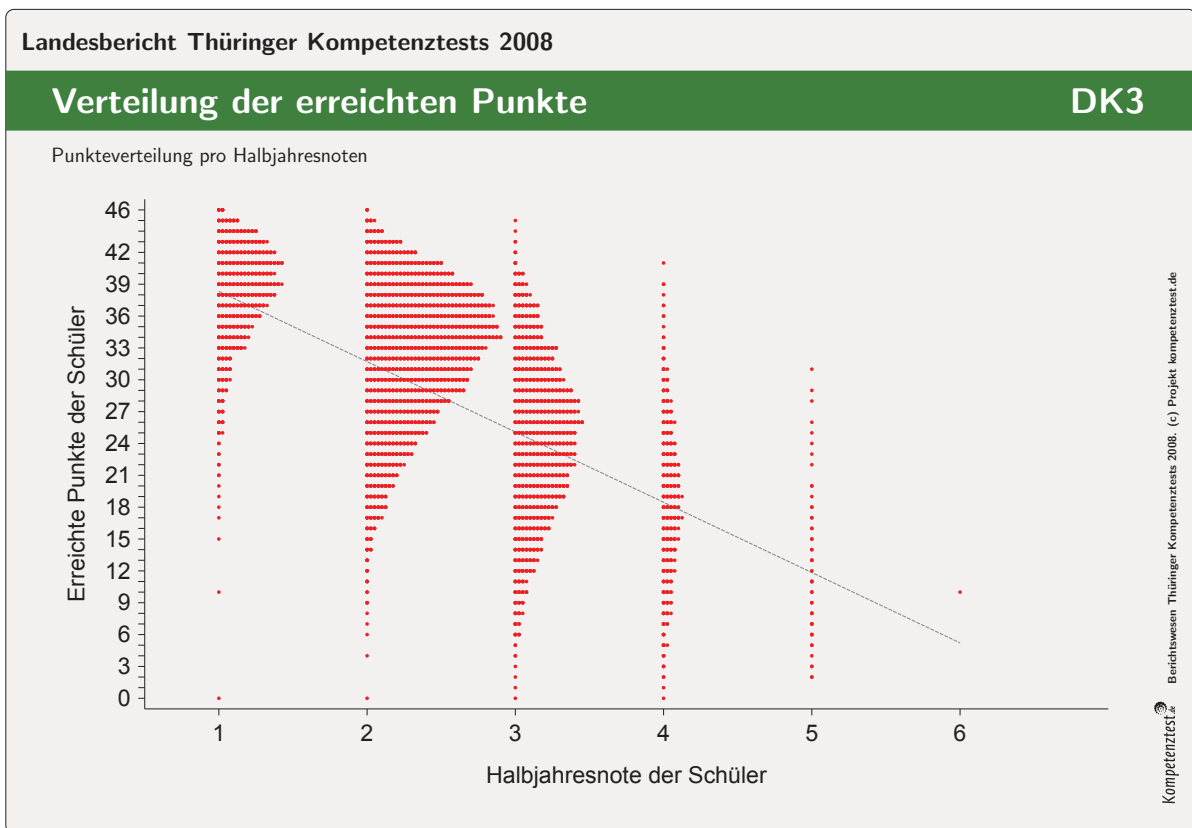


Abbildung 7.36: Zusammenhang zwischen Halbjahresnote und Testergebnis im Fach Deutsch, Klassenstufe 3, DK3

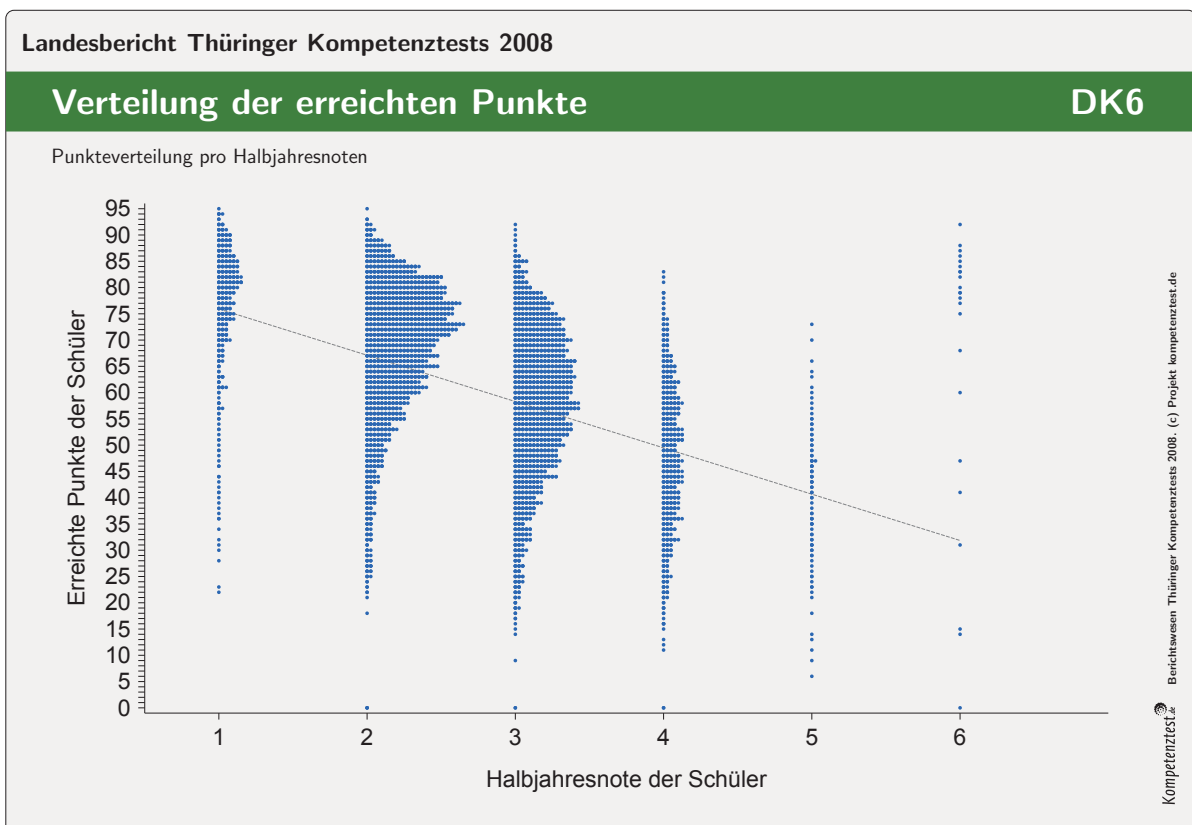


Abbildung 7.37: Zusammenhang zwischen Halbjahresnote und Testergebnis im Fach Deutsch, Klassenstufe 6, DK6

Landesbericht Thüringer Kompetenztests 2008

Verteilung der erreichten Punkte

EK6H

Punkteverteilung pro Halbjahresnoten

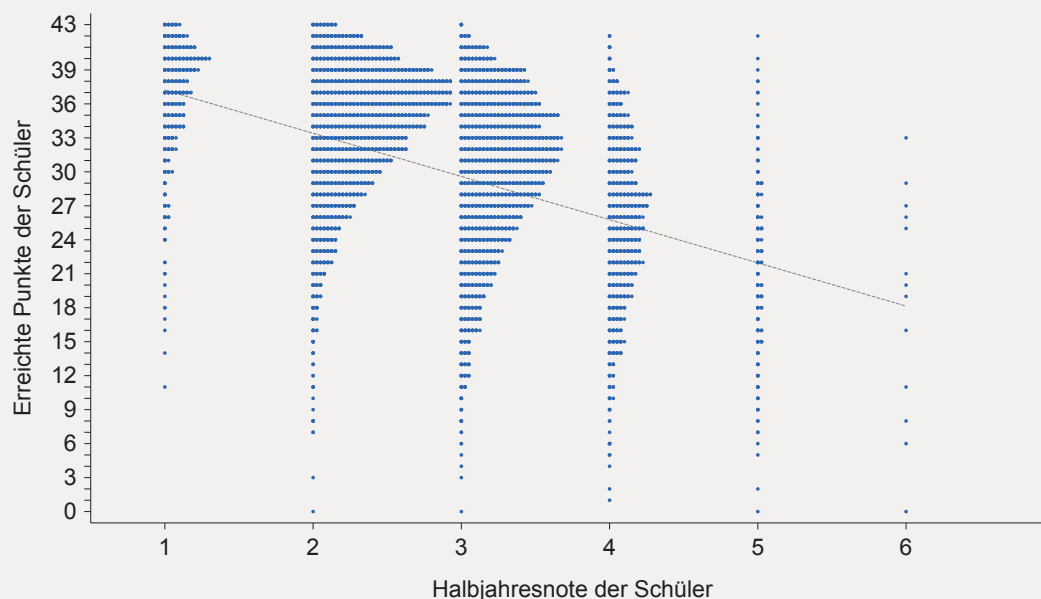
Kompetenztest.de
Berichtswesen Thüringer Kompetenztests 2008. (c) Projekt kompetenztest.de

Abbildung 7.38: Zusammenhang zwischen Halbjahresnote und Testergebnis im Fach Englisch Hörverstehen, Klassenstufe 6, EK6H

Landesbericht Thüringer Kompetenztests 2008

Verteilung der erreichten Punkte

EK6L

Punkteverteilung pro Halbjahresnoten

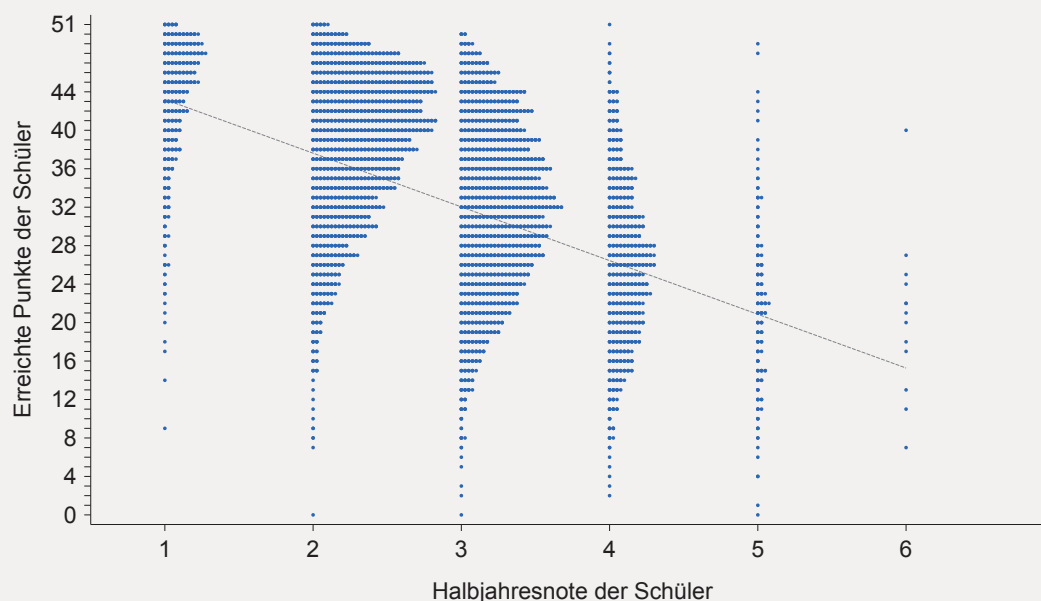
Kompetenztest.de
Berichtswesen Thüringer Kompetenztests 2008. (c) Projekt kompetenztest.de

Abbildung 7.39: Zusammenhang zwischen Halbjahresnote und Testergebnis im Fach Englisch Leseverstehen, Klassenstufe 6, EK6L

Schaut man sich die Abbildungen 7.33 bis 7.39 einmal genauer an, so fallen jene Schüler besonders auf, welche eine Halbjahresnote 5 oder 6 bekamen, aber bei den Kompetenztests eine hohe Punktzahl erreichten. Es ist schwierig zu erklären, warum Schüler mit einer recht ausgeprägten Kompetenz derart schlechte Schulleistungen zeigen. Vielleicht liegt es daran, dass dies Schüler sind, welche besonders unmotiviert sind. Damit wäre aber nicht beantwortet, warum sich die Schüler beim Kompetenztest anstrebten und eine gute Leistung zeigten. Es wäre auch vorstellbar, dass bei jenen Schülern eine sehr schlechte Beziehung zum Lehrer besteht. Die Frage ist, ob sich eine schlechte Beziehung derart in der Halbjahresnote wiederfinden sollte. Es bleibt also offen zu klären, woran es liegt, dass sich Halbjahresnote und Testleistung so stark voneinander unterscheiden. Bei Schülern, welche eine recht gute Halbjahresnote haben, aber eine schlechte Testleistung zeigten, liegt zwar auch eine große Differenz zwischen den Leistungsmaßen vor, diese ist aber durch Tagesform oder Motivationszustand eher erklärbar.

7.2.7 Vorhersage der Testergebnisse durch die Schülermerkmale

In den vorangestellten Abschnitten wurde der Zusammenhang zwischen einzelnen Schülermerkmalen und den Ergebnissen in den Kompetenztests dargestellt. Um den gemeinsamen Beitrag der einzelnen besprochenen Schülermerkmale auf die Testergebnisse abzuschätzen, wurden sie gemeinsam in ein Regressionsmodell zur Vorhersage der Testergebnisse eingebracht.

Bei der Regressionsanalyse wird versucht, eine Funktion zu finden, die die Schülerleistung mithilfe bestimmter Prädiktoren (hier die einzelnen Schülermerkmale) bestmöglich beschreibt. In den folgenden Analysen wurden die in den Kompetenztests erhobenen Schülermerkmale Geschlecht, Muttersprache, SES, Jahr des Schulbesuchs und das Vorliegen einer besonderen Lernschwierigkeit bzw. von sonderpädagogischem Förderbedarf berücksichtigt. Mit der Schulart als Prädiktor könnte man sicherlich einen großen Teil an der Kompetenztestleistung erklären. Jedoch soll sich bei der angeführten Regressionsanalyse auf eine einfache Regression beschränkt werden, welche nur Prädiktoren auf Schülerebene umfasst. Die Güte der Vorhersage wird gemessen am Anteil der Varianz der Schülerleistung, der durch die Prädiktoren erklärt wird. D. h. es wird geschaut, wie viel der Varianz in der Schülerleistung durch Unterschiede in den Schülermerkmalen erklärt werden kann. Nachfolgend wird für alle Kompetenztests diese Varianzaufklärung dargestellt. Weiterführende Informationen zu den einzelnen Regressionen, zum Beispiel die Regressionskoeffizienten, befinden sich im Anhang.

Klassenstufe 3

Im Kompetenztest Deutsch wurde durch die berücksichtigten Schülermerkmale (Geschlecht, Muttersprache, SES, Jahr des Schulbesuchs und besondere Lernschwierigkeiten) 15,4 % Varianz der Testleistungen aufgeklärt. Im Kompetenztest Mathematik ergab sich eine Varianzaufklärung in Höhe von 14,9 %.

Klassenstufe 6

Am Kompetenztest Deutsch der Klassenstufe sechs konnten durch die Schülermerkmale Geschlecht, Muttersprache, SES, Jahr des Schulbesuchs und besondere Lernschwierigkeiten 19,2 % der Varianz der Testergebnisse aufgeklärt werden. Für das Testergebnis in Mathematik und Englisch spielte die Muttersprache keine prädiktive Rolle. Die restlichen Schülermerkmale sorgten für eine Varianzaufklärung von 20,1 % für die Leistung im Kompetenztest Mathematik und für eine aufgeklärte Varianz von 16,9 % für Englisch-Hören und 19,9 % für Englisch-Lesen. Auffällig ist, dass Muttersprache lediglich für das Abschneiden im Kompetenztest Deutsch prädiktiv für die Vorhersage des Testergebnisses war. Erklärbar ist dieses Ergebnis dadurch, dass

das Beherrschen der deutschen Sprache zwar für das Lernen in anderen Fächern eine Rolle spielen sollte. Jedoch beginnen beim Lernen der Fremdsprache Englisch (fast) alle Schüler am gleichen Ausgangspunkt. Gute Deutsch Kenntnisse sollten dabei nicht unbedingt vorteilhaft sein, sondern eher kontinuierliches Lernen und Üben der neuen Sprache. Im Fach Mathematik sollte vor allem Zahlenverständnis und logisches Denken im Vordergrund stehen, Sprachverständnis sollte eine eher untergeordnete Rolle spielen.

Klassenstufe 8

Das Abschneiden im Kompetenztest Mathematik der Klassenstufe 8 kann zu 23,3 % durch die erfassten Schülermerkmale erklärt werden. Im Gegensatz zur Klassenstufe 6 leistet hier auch die Muttersprache einen Beitrag zur Aufklärung der Varianz der Testleistung. Das könnte daher kommen, dass in höheren Klassenstufen komplexere mathematische Themen eine Rolle spielen, die sprachliches Verständnis unbedingt erfordern.

Als einer der stärksten Prädiktoren kann über die Klassenstufen hinweg der sozioökonomische Status angesehen werden. Während die Bücherfrage in der der Klassenstufe 3 eher wenig Varianz an der Testleistung erklären kann, gewinnt sie in der sechsten und Klassenstufe 8 an Bedeutung und zeigt eine durchaus respektable Varianzaufklärung. In der Klassenstufe acht können immerhin 15,7 % der Varianz auf den SES zurückgeführt werden. Das zeigt erneut, dass das Abschneiden im Kompetenztest in bedeutsamer Weise vom sozialen Hintergrund der Schüler abhängt. Die durch die Schülermerkmale aufgeklärte Varianz ist in Anbetracht der Tatsache, dass nur einige wenige Merkmale berücksichtigt werden konnten, recht groß. Es sind viele weitere Einflussfaktoren auf Schülerebene (z. B. kognitive Basisfähigkeiten, Motivation, Interesse, Tagesform), aber auch auf Klassen- und Schulebene vorhanden, die Auswirkungen auf die Leistungen der Schüler haben können.

7.2.8 Entwicklung über die letzten Jahre

Die Kompetenztests werden seit 2004 an Thüringer Schulen durchgeführt. In diesem Jahr soll es auch eine Betrachtung der Ergebnisse über die Zeit hinweg geben. Aufgrund des Thüringer Schülerlängsschnitts ist es möglich dieselben Schüler in verschiedenen Klassenstufen zu betrachten. Deshalb erfolgt zunächst eine längsschnittliche Betrachtung der einzelnen Schülermerkmale, d. h. für einen Jahrgang wird geschaut, inwieweit sich die Effekte hinsichtlich Geschlecht, Muttersprache und SES über die Jahre hinweg entwickeln. Dies ist möglich für den Kompetenztest Mathematik im diesjährigen Jahrgang der Klassenstufe 8. Die Effekte bezüglich der Schülermerkmale werden verglichen mit den Effekten desselben Jahrgangs vor zwei Jahren, also der Klassenstufe 6 von 2006. Anschließend folgt eine Betrachtung verschiedener Jahrgänge über die Zeit.

Schülerlängsschnitt

Geschlecht

Tabelle 7.28 zeigt den Vergleich des Geschlechtereffekts von 2006 und 2008 beim Kompetenztest Mathematik. Die Effekte wurden aus den jeweiligen Mittelwerten und Standardabweichungen von Jungen und Mädchen berechnet, diese sind aber aus Gründen der Übersichtlichkeit nicht noch einmal angeführt. Ein positiver Wert d bedeutet in diesem Fall, dass Jungen durchschnittlich eine bessere Testleistung erzielten als Mädchen. Es zeigt sich, dass Jungen in Gymnasien und Regelschulen im Mittel ihren Leistungsvorsprung beim Kompetenztest Mathematik halten können. Sowohl vor zwei Jahren, als auch in diesem Jahr liegt ein kleiner Effekt

zugunsten der Testleistung der Jungen vor. In der Förderschule ist der durchschnittliche Geschlechtereffekt seit 2006 sogar stärker geworden. Mädchen schneiden im Mittel schwächer ab als die gleichaltrigen Jungen der Förderschule. Während man 2006 noch von einem kleinen Effekt sprechen konnte, zeigte sich dieses Jahr immerhin ein mittlerer Effekt.

Tabelle 7.28: Entwicklung des Geschlechtereffekts im Kompetenztest Mathematik: 2006 - 2008

	Geschlecht Klassenstufe 6	Geschlecht Klassenstufe 8
	(d)	(d)
Gymnasium	0,308	0,325
Regelschule	0,319	0,350
Förderschule	0,286	0,558

Muttersprache

Bei der Analyse des Einflusses der Muttersprache auf die Testleistung wurden Schüler deutscher Muttersprache verglichen mit Schülern nicht-deutscher Muttersprache. Im Gymnasium bleibt der Vorteil deutscher Muttersprachler beim Abschneiden im Kompetenztest Mathematik über die zwei Jahre stabil. Schüler deutscher Muttersprache erzielten ein leicht besseres Testergebnis als Schüler nicht-deutscher Muttersprache, es liegt jeweils ein kleiner Effekt vor. In der Regelschule unterschied sich die Testleistung der beiden Gruppen 2006 nur leicht voneinander. Schüler deutscher Muttersprache schnitten im Durchschnitt etwas besser ab als nicht-deutsche Muttersprachler. Dieser Leistungsunterschied ist in diesem Jahr größer geworden, es liegt mit $d = 0,54$ ein mittlerer Effekt vor. Das könnte bedeuten, dass das Nicht-Beherrschen der Unterrichtssprache ein kumulatives Defizit nach sich zieht: Vielleicht werden aufgrund sprachlicher Probleme komplexe Sachverhalte nicht wirklich verstanden, was wiederum dazu führen kann, dass die Schüler mit nachfolgenden Aufgaben und Anforderungen nicht entsprechend mithalten können. Dieses Ergebnis liefert einen weiteren Hinweis darauf, wie wichtig die frühzeitige sprachliche Förderung und Integration von Schülern mit Migrationshintergrund ist. Die Entwicklung im Gymnasium zeigt, dass der Abstand zwischen deutschen Muttersprachlern und nicht-deutschen Muttersprachlern nicht zwangsläufig vergrößern muss. Daher ist die Entwicklung in den Regelschulen als bedenklich zu betrachten.

Tabelle 7.29: Mittelwertevergleich bzgl. Muttersprache im Kompetenztest Mathematik: 2006 - 2008

	Muttersprache Klassenstufe 6	Muttersprache Klassenstufe 8
	(d)	(d)
Gymnasium	0,344	0,331
Regelschule	0,382	0,540

Sozioökonomischer Status

In Tabelle 7.30 sind die Korrelationen des sozioökonomischen Status mit dem Testergebnis im Kompetenztest Mathematik dargestellt. Es zeigt sich, dass der Zusammenhang von 2006 bis 2008 soweit stabil geblieben ist. Das bedeutet, eine gute Testleistung im Kompetenztest Mathematik hängt in gewisser Weise über die Jahre hinweg vom SES ab. Vielleicht sorgt eine kontinuierliche Unterstützung aus dem Elternhaus für höhere Leistungsmotivation bei den Kindern, die auch in der Klassenstufe 8 zu einer besseren Testleistung führt. Das könnte für den

schulischen Alltag bedeuten, dass Lehrer und Eltern eng zusammenarbeiten müssen. Eltern sollten schon früh darüber informiert werden, wie wichtig ihre Mitwirkung und Unterstützung für Erfolg in der Schule ist.

Tabelle 7.30: Korrelationen des SES mit den Testergebnissen in Mathematik: 2006 - 2008

	SES Klassenstufe 6	SES Klassenstufe 8
	(d)	(d)
Gymnasium	0,19	0,21
Regelschule	0,21	0,21
Förderschule	0,13	0,16
Gesamt	0,36	0,40

Zeitliche Betrachtung verschiedener Jahrgänge

Ausgewählte Analysen der Testergebnisse über die letzten Jahre haben gezeigt:

1. Die Effekte zwischen den Schularten sind seit 2005 stabil geblieben. Natürlicherweise unterscheiden sich die Testleistungen von Gymnasiasten, Regelschülern und Förderschülern. Es ist jedoch eine positive Entwicklung, wenn die Leistungsspanne nicht wesentlich zusammen oder auseinander geht.
2. Die Durchlässigkeit zwischen den Schularten hat sich seit 2005 nicht wesentlich verändert. Das bedeutet, es gibt zwar immer Förder- und Regelschüler, welche das Potential haben einen anderen Bildungsweg einzuschlagen. Jedoch zeigt die Entwicklung über die Jahre, dass es prozentual nicht mehr Schüler werden. Das könnte bedeuten, dass wiederholt ein großer Teil der Schüler den Bildungsweg einschlägt, der auch dem tatsächlichen Kompetenzniveau entspricht. Die gleich bleibende Durchlässigkeit kann aber auch ein Zeichen dafür sein, dass es eine nicht zu unterschätzende Zahl von Schülern gibt, die mit etwas mehr Anregung und Förderung einen anderen Bildungsabschluss schaffen würden.
3. Der Einfluss des SES auf die Testleistung ist seit 2005 relativ stabil. In der Klassenstufe 3 findet sich eine Korrelation von etwa $r = 0,20$ und in der Klassenstufe 6 von ca. $r = 0,35$ über die Jahre hinweg. Das bedeutet, im Mittel kann der sozioökonomische Status als stabiles Korrelat der Testleistung angesehen werden. Die PISA-Ergebnisse⁵ von 2006 haben gezeigt, dass es durchaus möglich ist, den Einfluss des sozioökonomischen Status über die Jahre zu reduzieren. Dazu ist es insbesondere notwendig, dass Kinder aus anregungsärmeren Elternhäusern schon frühzeitig entsprechende Unterstützung erfahren.

7.2.9 Zusammenfassung

Dieser Abschnitt des Landesberichts hat sich mit dem Einfluss von Schülermerkmalen auf das Abschneiden im Kompetenztest beschäftigt. Zusammenfassend lässt sich festhalten, dass sich der bekannte Geschlechtereffekt auch in den Deutsch- und Mathematikergebnissen diesen Jahres wiederfinden ließ. Hingegen im Kompetenztest Englisch unterschieden sich die Thüringer Mädchen und Jungen nicht voneinander. Weiterhin zeigte sich, dass Schüler nicht-deutscher Muttersprache in allen Kompetenztests durchschnittlich schwächer abschnitten als die Schüler deutscher Muttersprache. Dieses Ergebnis liefert einen Hinweis darauf, wie wichtig die frühzeitige Sprachförderung für Schüler mit Migrationshintergrund ist. Eltern sollten deshalb unbedingt auf die Bedeutsamkeit der sprachlichen Integration ihrer Kinder aufmerksam gemacht werden. Der sozioökonomische Status wies auch dieses Jahr einen Zusammenhang

⁵ vergleiche PISA-Konsortium Deutschland (2007)

mit dem Abschneiden in den Kompetenztests auf. Die Regressionsanalysen der vergangenen Jahre stellte außerdem heraus, dass der SES als stabiler Einflussfaktor auf das Abschneiden im Kompetenztest angesehen werden kann. Bei der Analyse der Schüler, welche schon mindestens eine Klassenstufe wiederholt haben, fiel auf, dass jene Schüler auch nach der Wiederholung schwächer abschneiden als der Rest der Klassenstufe. Eine Ausnahme bildet dabei die Förderschule. Hier schneiden die Wiederholer genauso gut ab wie ihre Mitschüler. Anscheinend werden Wiederholer in Förderschulen spezieller unterstützt und werden nach ihrer Wiederholung besser integriert. Beim Vergleich der Leistung von Schülern verschiedener Schularten (siehe Analyse zur Durchlässigkeit der Schularten) zeigte sich, dass es einen nicht geringen Anteil von Regelschülern gibt, welche bei den Tests besser abschnitten als das schlechteste Drittel der Gymnasiasten bzw. Förderschüler, welche besser abschnitten als das schlechteste Drittel der Regelschüler. Dieses Ergebnis lässt vermuten, dass einige Schüler durchaus das Potential hätten, einen höheren Bildungsabschluss zu erlangen. Im Rahmen der Analyse zu Schülern mit besonderer Leistungsschwierigkeit oder sonderpädagogischem Förderbedarf konnte wie auch in den letzten Jahren festgestellt werden, dass die Absprache unter den Fachlehrern verbesserungswürdig ist.

Abschließend lässt sich sagen, dass die erhobenen Schülermerkmale Geschlecht, Muttersprache, Sozioökonomischer Status, Jahr des Schulbesuchs und besondere Lernschwierigkeiten /Sonderpädagogischer Förderbedarf in deutlichem Zusammenhang mit den Testergebnissen stehen. Die Berücksichtigung solcher Schülermerkmale ist demzufolge für eine faire Auswertung der Schul- und Klassenergebnisse unbedingt notwendig. Aus diesem die genannten Merkmale im Rahmen der klassenspezifischen Rückmeldung bei der Berechnung des korrigierten Landesmittels einbezogen, so dass die Rückmeldung in Form von Vergleichsdaten für die einzelnen Klassen nicht nur fairer, sondern auch aussagekräftiger gestaltet wird. Es sollte an dieser Stelle aber unterstrichen werden, dass mit dem Kompetenztest nicht alle relevanten Einflussvariablen auf Schülerebene erfasst werden. Für die kommenden Jahre besteht also weiterhin Entwicklungsbedarf. Beispielsweise könnte man schwerer erfassbare Merkmale wie die kognitiven Fähigkeiten und die Motivation der Schüler berücksichtigen.

8 Rezeption und Evaluation der Tests

8.1 Kurzbefragung nach der Dateneingabe

Die Erhebung der Verwendbarkeitsitems fand nach Abschluss der Eingabe der Schülerdaten statt. Dabei konnten die Lehrer Aussagen zur Nützlichkeit der Kompetenztests als Instrument zur Lernstandsdiagnostik, für die eigene Unterrichtsentwicklung und die Schulentwicklung treffen. Zusätzlich konnte die Lehrplanadäquatheit der Aufgaben beurteilt werden. Die vier genannten Aspekte wurden mit Hilfe verschiedener Aussagen auf einer fünfstufigen Skala ähnlich dem Schulnotensystem bewertet, wobei 1 = volle Zustimmung und 5 = gar keine Zustimmung bedeutet. Neben den vorgegebenen Bereichen bestand auch die Möglichkeit, ein freiwilliges Feedback zur Dateneingabe zu geben. Teilweise wurde dies außerdem dazu genutzt, inhaltliche und kritische Anmerkungen zu den diesjährigen Kompetenztests zu machen.

Im Folgenden soll ein Überblick über die abgegebenen Einschätzungen gegeben werden. Dabei wird nicht speziell auf Förderschulen eingegangen. Aufgrund der für statistische Analysen zu kleinen Teilnehmerzahl wären keine verlässlichen Aussagen möglich. Abschließend werden Anregungen und Kommentare der Fachlehrer zu den einzelnen Tests dargestellt.

8.2 Ergebnisse der Verwendbarkeitsanalysen

Insgesamt lässt sich eine positive Bewertung der Nützlichkeit der Kompetenztests der Lehrer über alle Schularten hinweg feststellen. Die mittleren Einschätzungen für die Lernstandsdiagnostik, Lehrplanadäquatheit, Unterrichtsentwicklung und Schulentwicklung lagen alle zwischen 2,3 und 2,8 (vgl. Abb. 8.1).

Besonders positiv wurden die jeweiligen Aspekte von den Grundschullehrern eingeschätzt. Deren Bewertungen lagen in fast allen Bereichen deutlich unter dem jeweiligen Durchschnitt. Auch der Mathematiktest der Klassenstufe 6 wurde im Vergleich zu den anderen an weiterführenden Schulen geschriebenen Tests (DK6, EK6 T1 und T2, MK8) besser eingestuft. 63,5 % der Gymnasial- und 58,6 % der Regelschullehrer stimmten voll oder überwiegend zu, dass der Test ein nützliches Instrument zur Lernstandsdiagnostik ist. Als voll oder überwiegend lehrplanadäquat stufen den Mathematiktest Klassenstufe 6 sogar 81,9 % der Gymnasial- und 77,8 % der Regelschullehrer ein. Alle anderen Tests (DK6, EK6 T1 und T2, MK8) befinden sich bezüglich der Einschätzungen auf ähnlichem Niveau. Auch sind hier keine systematischen Unterschiede zwischen den Schularten festzustellen. Der Aspekt der Schulentwicklung, dem die Kompetenztests dienen sollen, wurde generell weniger positiv eingestuft. Dies spiegelt wider, dass die Tests auf einer anderen Ebene, nämlich der der Klassen bzw. Klassenstufen, eine weitaus größere Bedeutung haben. Da außerdem nicht in jedem an einer Schule unterrichteten Fach ein solcher Kompetenztest durchgeführt wird, ist der Einfluss und die Nützlichkeit für die gesamte Schule von vornherein begrenzt. Dass die allgemeine Einschätzung dennoch im positiven Bereich liegt, ist besonders erfreulich. Eine genaue Übersicht über die Ergebnisse der Erhebung der Verwendbarkeit der verschiedenen Kompetenztests aufgegliedert nach Schulart findet sich in Tab. 8.1.

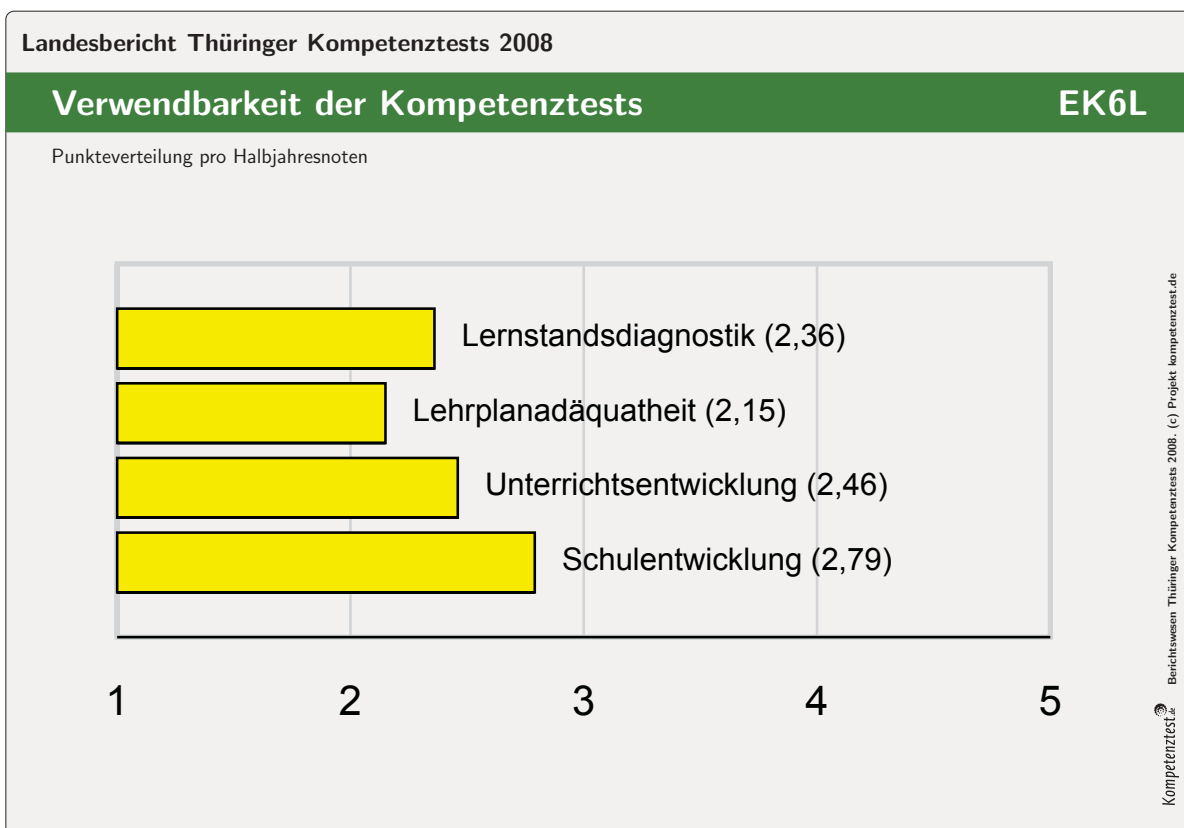


Abbildung 8.1: Mittlere Einschätzung der Verwendbarkeit der Kompetenztests durch die Fachlehrer über alle Klassenstufen und Schularten hinweg (gemessen auf einer Skala von 1 = volle Zustimmung bis 5 = gar keine Zustimmung).

8.2.1 Anmerkungen und Anregungen zu den Kompetenztests

Im Anschluss an die Dateneingabe hatten die Fachlehrer die Möglichkeit, Anregungen, Wünschen oder Kritik bezüglich der Dateneingabe zu äußern. Diese Gelegenheit wurde auch teilweise dazu genutzt, um Kritik an den Kompetenztests an sich zu üben. Diese Rückmeldungen sind nicht repräsentativ, sollten aber dennoch Beachtung finden und werden hier auszugsweise wiedergegeben, um das Bild zu vervollständigen. Beim Lesen sollte beachtet werden, dass diese Aussagen nur die Meinung eines kleinen Teils der Lehrerschaft widerspiegeln.

Anmerkungen zum Mathematiktest Klassenstufe 3: Insgesamt gab es bezüglich des Mathematiktests in der Klassenstufe 3 nur wenige Anmerkungen. Die Aussage eines Lehrers fasst die Äußerungen zusammen: „Die Aufgabenformate sind spitze, aber die Auswertung mangelhaft.“ Vor allem wurde bei der Auswertung die ungünstige Punktevergabe kritisiert, bei der Teilleistungen unberücksichtigt bleiben mussten. Ein weiterer Kritikpunkt war die zur Bearbeitung zur Verfügung stehende Zeit, die teilweise etwas knapp eingeschätzt wurde. Außerdem wurde darauf hingewiesen, dass es aufgrund der Flexibilität bei der Vermittlung der Lehrplaninhalte einige Stoffgebiete, die der Kompetenztest bereits abverlangte, noch nicht im Unterricht behandelt worden seien. Dies könne bei einigen Schülern zu Verunsicherung führen.

Tabelle 8.1: Übersicht der mittleren Einschätzungen von Lernstandsdiagnostik, Lehrplanadäquatheit, Unterrichtsentwicklung und Schulentwicklung nach Kompetenztests und Schularten (Gym = Gymnasium, RS = Regelschule) getrennt.

Kompetenztest	Schulart	Lernstandsdiagnostik	Lehrplanadäquatheit	Unterrichtsentwicklung	Schulentwicklung
DK3		2,25	1,88	2,24	2,59
MK3		2,17	2,05	2,17	2,46
DK6	Gym	2,51	2,11	2,68	3,03
DK6	RS	2,48	2,16	2,64	3,02
MK6	Gym	2,24	1,98	2,36	2,80
MK6	RS	2,36	2,04	2,47	2,79
EK6 T1	Gym	2,48	2,24	2,65	2,99
EK6 T1	RS	2,46	2,42	2,64	2,96
EK6 T2	Gym	2,34	2,27	2,66	2,96
EK6 T2	RS	2,29	2,22	2,42	2,79
MK8	Gym	2,35	2,39	2,65	3,10
MK8	RS	2,46	2,33	2,61	2,82

Diese Rückmeldungen zeigen, dass zum Teil immer noch Informationsdefizite über die Ziele der Tests und die Hintergründe der Testformate herrschen. Ähnliche Anmerkungen gab es auch zu den anderen Tests.

Anmerkungen zum Mathematiktest Klassenstufe 6: Zum Mathematiktest der Klassenstufe 6 wurden in diesem Jahr besonders Hinweise zur Punktevergabe gegeben. So äußerten einige Lehrer den Wunsch nach einer „Differenzierung der Punktevergabe nach Art des Fehlers“, so dass eine Unterscheidung zwischen falsch und gar nicht gelöst möglich wäre. Darüber hinaus wurde der Test von manchen Lehrern als zu „sprachlastig“ eingestuft und der Zeitpunkt, zu dem der Kompetenztest geschrieben werden musste, wurde als „zu zeitig im Schuljahr“ bemängelt.

Anmerkungen zum Mathematiktest Klassenstufe 8: „Die Aufgaben an sich haben mir gefallen und waren ein guter Mix aus Grundlagen und auch anspruchsvolleren Problemstellungen, sie haben die Schüler gefordert.“ Neben dieser positiven Einschätzung des Mathematiktest für die Klassenstufe 8 wurde wie auch in anderen Test die Punktevergabe ohne die Berücksichtigung von Folgefehlern bemängelt. Außerdem schien für einige Schüler die Zeitvorgabe zu knapp gewesen zu sein. Leider hat das Lehrermanual einen Fehler enthalten, was es in Zukunft zu vermeiden gilt. Insgesamt aber gab es nur wenig globale Kritik an den Mathematiktests.

Anmerkungen zu den Deutschtests Klassenstufe 3: Neben sehr positiven Einschätzungen zu den Aufgaben, die „sehr gut gestaltet und auch verständlich geschrieben“ waren, gab es auch in diesem Jahr einige kritische Anmerkungen zu dem Deutschtests Klassenstufe 3. Zum einen gab es den Wunsch, auch Teilergebnisse bzw. Folgefehler bei der Bewertung zu berücksichtigen und damit die als „ungünstig“ wahrgenommene Punktevergabe zu verändern. Zum anderen wurde des Öfteren bemängelt, dass die zu Verfügung stehende Zeit für die „umfangreichen“ Aufgaben zu knapp bemessen sei. Als Idee dazu wurde die Aufteilung des Tests auf zwei Tage vorgeschlagen. Einige Lehrer haben angeraten, ein Wörterbuch bei der Bearbeitung zuzulassen und größeren Wert auf die Rechtschreibung zu legen, indem diese in die Bewertung einbezogen werden könnte.

Anmerkungen zum Deutschtest Klassenstufe 6: Auch zu dem Deutschtest für die Klassenstufe 6 gab es generell positive Einschätzungen, die „den Test als sehr gut“ bewerteten. Außerdem wurde die Möglichkeit der Vorbereitung der Klasse positiv hervorgehoben, da auf zum Download zur Verfügung stehende Materialien aus den letzten Jahren zurückgegriffen werden konnte. Doch auch kritischere Anmerkungen wurden laut. Besonders die Rechtschreibfehler in Schülerbögen und Lehrermanualen wurden bemängelt. Des Weiteren wurden Zeitmangel bei der Briefaufgabe bzw. zu viel Bearbeitungszeit für die zweite Testhälfte negativ hervorgehoben. Unklarheiten traten bezüglich des Wörterzählens bei den von den Schülern verfassten Briefen auf. Ebenso sorgte die Witzaufgabe für einigen Unmut, da hier besonders die unterschiedliche Auslegung von Humor in dieser Altersstufe bemängelt wurde und zu „Verständnisschwierigkeiten“ geführt habe. Auch in diesem Test wurde der Wunsch nach der Berücksichtigung der Rechtschreibleistung geäußert. Außerdem wurde kritisiert, dass „die Kontrolle der Schülerarbeiten (...) für den unterrichtenden Fachlehrer eine fast unzumutbare Mehrarbeit“ darstelle.

Anmerkungen zum Englischtest Klassenstufe 6 Leseverstehen/Hörverstehen: „Der Test war schwer, zeigt mir aber, was erwartet wird“, so die Einschätzung eines Lehrers zum Englischtest Klassenstufe 6. Andere haben gegenteilige Erfahrungen gemacht: „Der Test selbst ist (...) für einen Großteil unserer Schülerinnen und Schüler zu einfach.“ Leider gab es „bei Aufgabe B5_6 ein[en] offensichtliche[n] Fehler“ im Lehrermanual, was von einigen Lehrern zu Recht kritisch angemerkt wurde. Es wurde der Hinweis gegeben, dass „[a]lle Aufgaben (...) zweimal gehört werden [könnten].“ Außerdem wurde der Wunsch nach einer einfacheren Formulierung wie „Umblättern“ geäußert, da bei dem Hinweis des teilweise „beidseitig bedruckt[en]“ Hefts einige Schüler die Bearbeitung aller Aufgaben versäumt hätten. Des Weiteren wurde die Druckqualität bemängelt.

Fazit: Im Vergleich zum letzten Jahr lässt sich für die Kompetenztests 2008 eine geringe positive Veränderung in der Bewertung feststellen - der Mittelwert sank sowohl bei der Lernstandsdiagnostik (2008: 2,36, 2007: 2,51) als auch bei der Einschätzung der Lehrplanadäquatheit (2008: 2,15, 2007: 2,18). Dies zeigt, dass die Tests inhaltlich durchaus am Lehrplan orientiert und geeignet sind, den Leistungsstand verschiedener Schüler zu erfassen. Der Beitrag der Kompetenztests zur Unterrichts- und Schulentwicklung wird im Vergleich dazu geringer eingeschätzt (Unterrichtsentwicklung: 2,46, Schulentwicklung: 2,79), befindet sich aber erfreulicherweise im positiven Bereich. Neben positivem Feedback gab es auch einige kritische Anmerkungen zu den diesjährigen Tests u. a. bezüglich einiger Aufgaben, der Punkteverteilung und Zeitvorgabe. Dies macht deutlich, dass es weiterhin Bedarf gibt, die Hintergründe der Verfahren transparent zu vermitteln. Einige der durchaus berechtigten Kritikpunkte sollten für die Weiterentwicklung der Kompetenztests berücksichtigt werden.

8.3 Rezeption der Testergebnisse an den Schulen

Für Thüringer Klassen und Schüler wurden in diesem Jahr allein für die Kompetenztests fast 16 500 Berichte im PDF-Format erzeugt. Die Analyse der Downloads der Klassenberichte vom geschützten Schulportal ergibt gegenüber dem letzten Jahr geringfügig gesunkene Abrufzahlen.

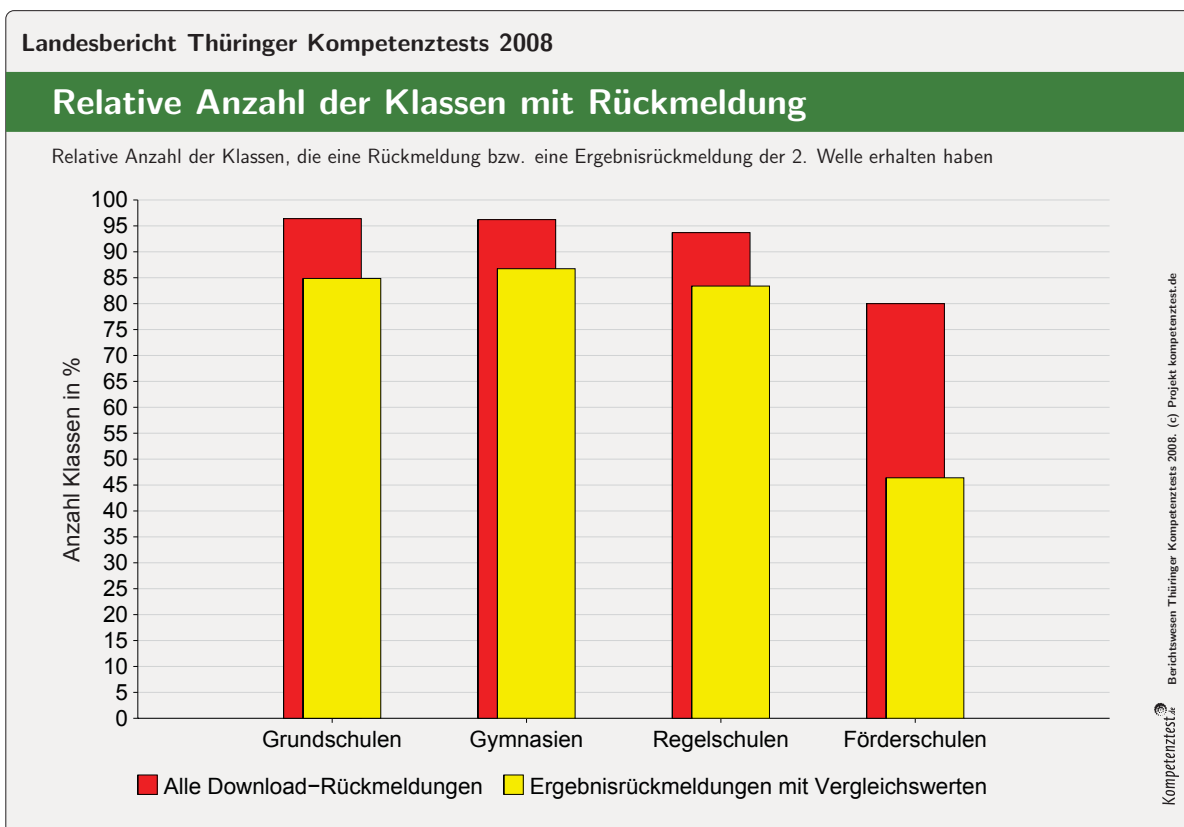


Abbildung 8.2: Relative Anzahl von Klassen mit Downloads mindestens einer Sofort-Rückmeldung vs. Ergebnissrückmeldung mit korrigierten Landesmittelwerten

Abbildung 8.2 stellt getrennt nach Schulart die relative Gesamtzahl der Klassen dar, die Rückmeldungen durch einen Download im Schulportal erhalten haben. Dargestellt ist jeweils in rot der Anteil an Klassen, die mindestens eine der Rückmeldungen erhalten haben. Dazu ist jeweils in gelb der Anteil der Klassen dargestellt, die in der zweiten Rückmeldewelle mindestens einen Bericht mit korrigierten Landesvergleichsdaten heruntergeladen haben.

Die Abrufquoten der Grundschulen, Gymnasien und Regelschulen sind einander sehr ähnlich und betragen durchschnittlich 95 % bei allen und 85 % bei den Ergebnissrückmeldungen mit Vergleichswerten. Wie auch im letzten Jahr ist eine geringere Abrufquote der Förderschulen deutlich erkennbar, insbesondere bei den finalen Ergebnisberichten mit Landesvergleichswerten. Da auch für Förderschulen die Teilnahme verpflichtend war, die Vergleichbarkeit mit anderen Schulen jedoch aufgrund der Heterogenität der Schülerschaft oft nur eingeschränkt sinnvoll ist, kann dieser Befund jedoch auch erwartet werden. Während im Vergleich zum letzten Jahr die Abrufquoten aller Berichte vergleichbar sind, sank die Zahl der Abrufe der Berichte mit Vergleichswerten um ca. 5 %.

Die Zahl der Schulen, die 10 Wochen nach Veröffentlichung keinerlei Rückmeldung vom Schulportal heruntergeladen hatten, ist im Vergleich zum Vorjahr von 9 auf 20 gestiegen. Dies betrifft 14 Grundschulen, 4 Förderschulen und 2 Regelschulen.

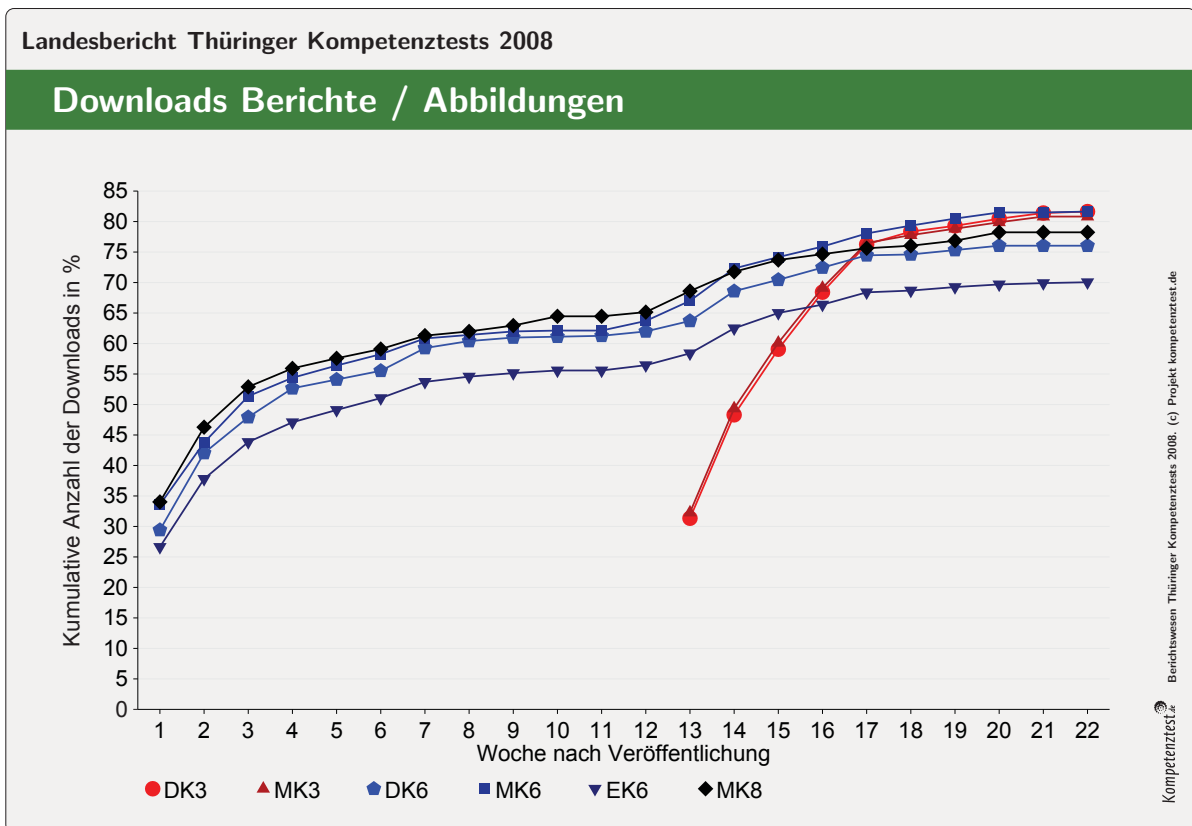


Abbildung 8.3: Downloads der Berichte mit Vergleichsdaten der verschiedenen Tests im zeitlichen Verlauf (pro Woche, Stand: 20. Oktober)

Abbildung 8.3 zeigt die Abrufquoten der Klassenberichte mit Vergleichswerten getrennt nach Fach und Klassenstufe im zeitlichen Verlauf. Die Rückmeldungen zu den Tests der Klassenstufen 6 und 8 wurden bereits am 19. Mai veröffentlicht, also 12 Wochen vor den Tests Klassenstufe 3. Wie auch im letzten Jahr zeigen die Abrufquoten der zuerst veröffentlichten Berichte einen nur geringen Anstieg auf, mit einem Knick zum Zeitpunkt der Veröffentlichung der restlichen Tests. Anscheinend waren auch in diesem Jahr viele Lehrer nicht über den ersten Rückmeldezeitpunkt informiert.

Die Abrufquoten sind in Klassenstufe 3 am höchsten, wie im letzten Jahr sind die Quoten in Deutsch und Englisch in Klassenstufe 6 am geringsten, wobei die Quoten im Fach Englisch einen deutlich sichtbaren Abstand haben.

Schulberichte, die gemittelte Vergleiche der Klassenergebnisse mit den jeweils korrigierten Landeswerten über mehrere Jahre hinweg enthalten, konnten in diesem Jahr für 826 Schulen erstellt werden. Von diesen Schulen haben ca. 22 % diesen Bericht bis zum 20. Oktober noch nicht abgerufen. Unter den Gymnasien ist diese Quote mit ca. 16 % am geringsten, die Förderschulen haben mit über 60 % die höchste Quote von Schulen ohne Abruf. Wie bereits dargestellt ist jedoch der Vergleich der Klassenergebnisse mit den korrigierten Landesmitteln – und somit auch der Schulbericht – für Förderschulen oft nur eingeschränkt sinnvoll.

Die Zahl der Schulen, die den Schulbericht nicht abgerufen haben, hat sich im Jahresvergleich halbiert. Letztes Jahr wurden die Schulberichte jedoch verspätet veröffentlicht, was sicherlich einen negativen Einfluss auf die Abrufquote hatte.

Die räumlichen Verteilungen der Abrufquoten der Ergebnis- und Schulberichte in den verschiedenen Schulämtern sind im separaten „Anhang zum Landesbericht“ dargestellt.

8.4 Evaluationen zu den Kompetenztests

Im Zuge der Kompetenztests werden die Schulleiter und Lehrer der verschiedenen Schularten stichprobenweise zur Einschätzung der Kompetenztests befragt. Diese Erhebung findet jährlich mit hinreichendem zeitlichen Anstand zu den Ergebnismeldungen statt, um die Nutzung der Ergebnisberichte zu erfassen. Die folgenden Ergebnisse beziehen sich daher auf die Bewertung und Nutzung der Kompetenztests 2006/2007. Die Befragung zu den Kompetenztests 2008 findet im November und Dezember 2008 statt, ihre Resultate werden daher erst im Landesbericht 2009 veröffentlicht.

Im Dezember 2007 wurden 185 Schulleiter und 429 an den Kompetenztests beteiligte Fachlehrer aller Schularten via Webfragebogen zur Einschätzung des Nutzens der Tests, zum Umgang mit den Ergebnissen und zu den Folgeprozessen befragt. 86 Schulleiter und 271 Lehrer gaben eine Rückmeldung zur Befragung. Die dabei verwendete sechsstufige Skala der vorgegebenen Antwortmöglichkeiten geht von 1 „stimme voll zu“ bis 6 „stimme gar nicht zu“ beziehungsweise bei der Durchführung von Maßnahmen von 1 „stark“ bis 6 „kaum“. Dabei bildet der Wert 3,5 einen neutralen Mittelwert, der weder zustimmend noch ablehnend ist. Ein Mittelwert kleiner als 3,5 bedeutet demzufolge, dass die zustimmenden Antworten überwiegen und der Sachverhalt von den Befragten überwiegend positiv gewertet wird. Ein Mittelwert größer als 3,5 zeigt ein Überwiegen der Ablehnungen. Die folgenden Abschnitte geben die Ergebnisse dieser Befragung wieder. Dabei werden die Resultate jeweils mit denen des Vorjahres 2005/2006 verglichen. Zunächst erfolgt eine gemeinsame Auswertung der Daten von Schulleitern und Fachlehrern, in den Folgeabschnitten wird nach Berufsgruppen und Schulart differenziert.

8.4.1 Gesamtergebnis

Abbildung 8.4 zeigt die Auswertung der Befragungsergebnisse aller 357 Schulleiter und Lehrer, die an der Befragung 2007 teilgenommen haben. Die Abbildung gibt die Einschätzung der Befragten zu den Tests im Vergleich zum Vorjahr wieder. Diese Einschätzung weist insgesamt leicht rückläufige Werte auf, die sich aber in den meisten Fällen als statistisch nicht bedeutsam herausstellen. Lediglich bei den Fragen danach, ob die Tests einen Baustein zur Schulentwicklung darstellen und die Chance zum Blick über den Tellerrand bieten, kann der Unterschied als signifikant (*), das heißt als nicht zufällig und als für die Darstellung relevant bezeichnet werden¹.

Betrachtet man hierbei die Antworten zu den Auswirkungen der Tests im Vergleich zum Vorjahr ist ebenfalls bei einigen Fragen ein statistisch nicht bedeutsamer leicht negativer Wert zu verzeichnen. Relevant beziehungsweise signifikant (*) sind dabei lediglich die Fragen nach dem Impuls für innovative Methoden, die Kontakte zu externen Schulberatern und der Bedarf an Fortbildung, wobei die letzteren beiden weniger bedeutend für den Schulalltag zu sein scheinen, da die Mittelwerte zwischen 3,7 und 4,4 (also schlechter als 3,5) auf ein Überwiegen der Ablehnungen in den Antworten hinweisen. Andererseits zeigen die Antworten bezüglich der Kommunikation mit dem Schüler, den Eltern und den Fachkollegen im Vergleich zum Vorjahr leicht positive Werte (siehe Abbildung 8.5).

¹ Das hierfür verwendete statistische Verfahren ist der Signifikanztest; dabei war das Signifikanzniveau bei $\alpha = 0.05$; nähere Informationen zu dieser Methode finden Sie bei Nachtigall & Wirtz (2004).

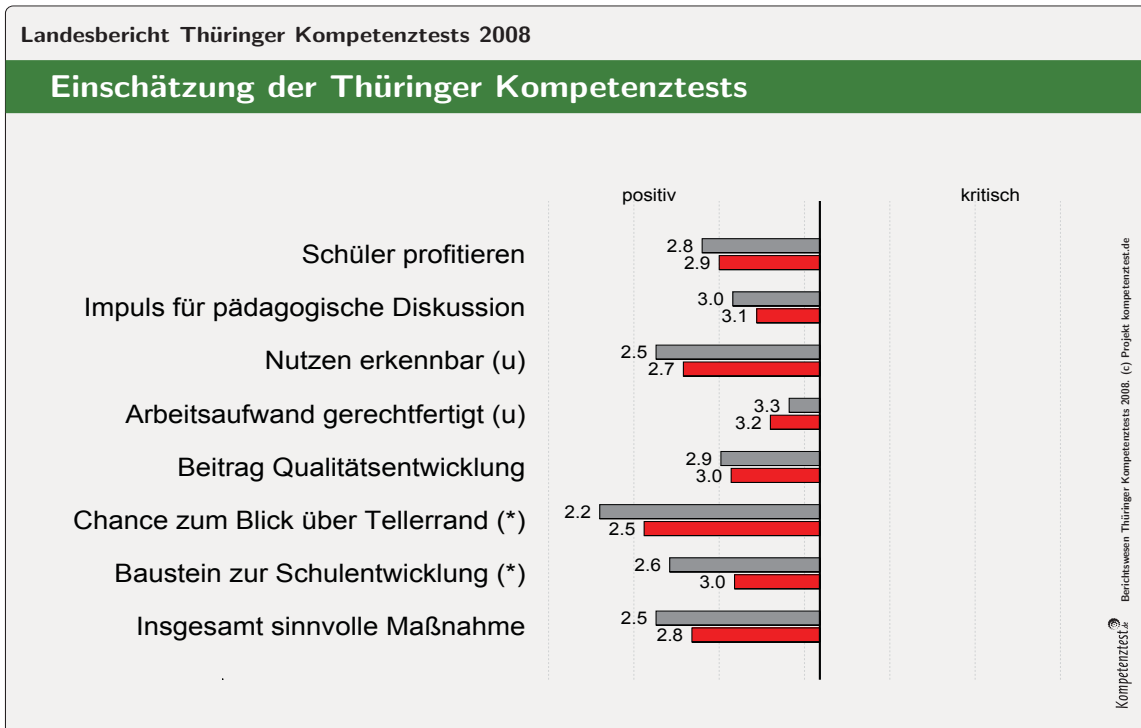


Abbildung 8.4: Schulleiter und Lehrer aller Schularten zur Einschätzung der Thüringer Kompetenztests 2005 / 06 und 2006 / 07 im Vergleich

Mittelwerte 2005 / 2006 (N>258; grau) und 2006 / 2007 (N>357; rot); Skalenmittelpunkt=3,5; (u) umgepolt; (*) Vergleich signifikant

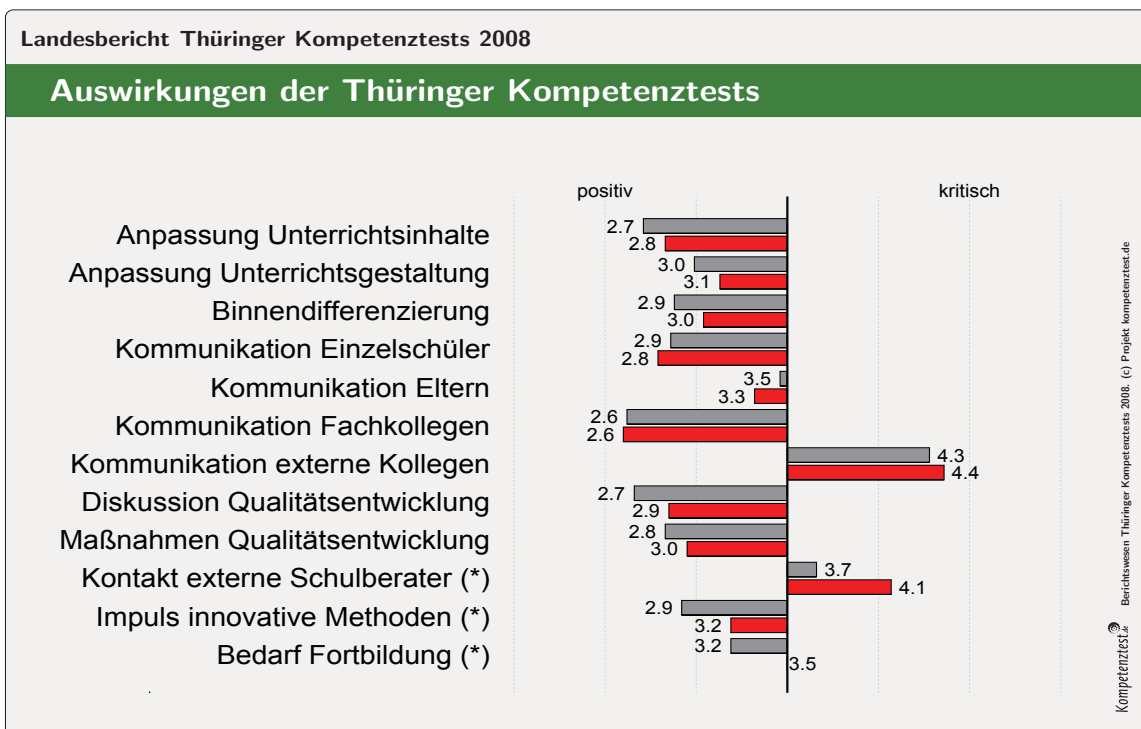


Abbildung 8.5: Schulleiter und Lehrer aller Schularten zu den Auswirkungen der Thüringer Kompetenztests 2005 / 06 und 2006 / 07 im Vergleich

Mittelwerte 2005 / 2006 (N>258; grau) und 2006 / 2007 (N>357; rot); Skalenmittelpunkt=3,5; (u) umgepolt; (*) Vergleich signifikant

Zusätzlich zu den geschlossenen Antworten hatten die Schulleiter und Lehrer die Möglichkeit, freie Kommentare abzugeben. Trotz der geäußerten Kritik in diesen freien Antworten zum Beispiel hinsichtlich des Zeitpunkts der Tests, des Schwierigkeitsgrades und der Bewertung der Aufgaben sowie des Arbeitsmehraufwandes für die Lehrer überwiegen die positiven Anmerkungen, die die Vorteile und Stärken der Tests hervorheben. Diese beziehen sich unter anderem auf die Nutzung der Tests als Vergleichsmöglichkeit auf verschiedenen Ebenen, als Instrument zur Analyse beispielsweise des Leistungsstandes, der Kompetenzentwicklung oder der Unterrichtsqualität sowie auf daraus abgeleitete Konsequenzen für den eigenen Unterricht und für die Schule.

Zusammenfassend kann man sagen, dass bei den meisten geschlossenen Fragen geringfügig schlechtere Werte im Vergleich zum Vorjahr auftauchen, die Veränderung jedoch in den meisten Fällen statistisch nicht signifikant ist. Die aufgrund des größeren Unterschieds signifikanten Merkmale sollten dagegen Anlass zum Handeln sein (siehe Abbildung 8.5).

Im Folgenden soll nun auf Unterschiede bei der Einschätzung und den Auswirkungen der Tests in den Antworten der Schulleiter und Lehrer einerseits und der verschiedenen Schularten andererseits eingegangen werden.

8.4.2 Vergleich der Schulleiter und Lehrer aller Schularten

Die Auswertung der Antworten von insgesamt 86 Schulleitern bezüglich der **Einschätzung** der Kompetenztests ergab, dass die Mehrzahl der Schulleiter die Tests insgesamt als eine sinnvolle Maßnahme einschätzen. Bei der Befragung gab es jeweils sechs Antwortmöglichkeiten von „stimme voll zu“ und „stimme fast voll zu“ bis zu „stimme nicht zu“ bis „stimme gar nicht zu“. Bei der Frage nach der Sinnhaftigkeit der Maßnahme gaben 69,7 % der befragten Schulleiter an, dass sie voll oder fast voll zustimmen, während nur 2,1 % meinten, die Tests sind nicht oder gar nicht sinnvoll. Ähnlich verhielt es sich mit der Frage danach, ob die Schüler von der Durchführung der Tests profitieren. Nur 2,2 % waren der Meinung, dass die Schüler nicht oder gar nicht profitieren. Befragt nach dem Nutzen der Tests für sich selbst und die Schule gaben sogar über die Hälfte der Schulleiter, nämlich 62,1 % an, dass die Tests einen sehr großen oder großen Nutzen haben, weitere 25,3 % sahen den Nutzen für ihre Schule teilweise. Diese überaus positive Tendenz der Annahme der Tests durch die Schulleiter ist besonders in Bezug auf die Schulentwicklung von maßgeblicher Bedeutung, da die Schulleiter die Folgeprozesse anstoßen und steuern.

Die Rückmeldung der Antworten der stichprobenweise befragten 258 Lehrer fiel im Vergleich zu der der Schulleiter nicht so positiv aus. Hier empfanden nur 45,4 % der befragten Lehrer die Tests insgesamt als eine sinnvolle oder sehr sinnvolle Maßnahme, während 18,7 % die Tests als nicht oder gar nicht sinnvoll ansahen. Ebenso meinte circa ein Fünftel der Lehrer, dass die Schüler von der Durchführung der Tests nicht oder gar nicht profitieren. Bei der Frage nach einem erkennbaren Nutzen der Tests für sich selbst und die Schule waren es auch 19 %, die sich ablehnend äußerten. Sogar 28,7 % meinten, der persönliche Arbeitsaufwand würde den Nutzen der Tests nicht rechtfertigen.

Die Unterschiede in den Aussagen der Schulleiter und Lehrer werden in der Abbildung 8.6 nochmals anhand von Mittelwerten verdeutlicht.

Als überaus positive Tendenz fällt auf, dass im Rahmen der **Auswirkungen** der Kompetenztests ausnahmslos alle Schulleiter von einer Zusammenarbeit in den Fachschaften berichten. Dennoch stimmen nur 94,8 % der Lehrer dem völlig oder fast zu, die restlichen 5,2 % geben an, dass die Auswertung der Tests in den Fachschaften kaum oder nicht stattfindet. Daraus lässt sich schlussfolgern, dass die Zusammenarbeit der Lehrer in den Fachschaften an allen

Schulen eine feste Arbeitsform darstellt, aber offensichtlich einige Lehrer nicht mit der Qualität der Auswertung der Kompetenztests zufrieden sind oder teilweise nicht hinreichend intensiv an den Auswirkungen der Tests gemeinschaftlich gearbeitet wird.

Ähnliches lässt sich auch für die Kommunikation mit den Eltern feststellen: während nur 4,5 % der Schulleiter sagen, die Kommunikation findet kaum oder nicht statt, behaupten das immerhin 19,7 % der Lehrer. Möglicherweise sind die Auswertungsgespräche mit den Eltern vorgesehen, finden aber noch nicht konsequent statt. Im Schulalltag zeigen die Erfahrungen, dass Elterngespräche oft erst nach aufgetretenen Lernschwierigkeiten oder anderen Problemen tatsächlich durchgeführt werden, teilweise nicht in allen Schulen feste, über das gesamte Schuljahr verteilte Termine für Lehrersprechstage oder Sprechstunden geplant sind oder die Abstände zwischen den Elternabenden relativ groß sind, so dass im Schuljahr mitunter nur ein Elternabend stattfindet. Betrachtet man hingegen den Fakt, dass noch vor zwei Jahren im Landesbericht der Thüringer Kompetenztests 2006 nur 39,5 % der Befragten in den Kompetenztests eine Möglichkeit der verstärkten Zusammenarbeit mit den Eltern sahen, sind es jetzt doppelt so viele Lehrer, die nach eigenen Angaben die Kompetenztests in Einzelgesprächen mit den Eltern auswerten, um Schwächen und Stärken des Schülers zu analysieren, gemeinsam Maßnahmen für die Förderung des Schülers festzulegen und ihn so in seiner Lernentwicklung voranzubringen. Abbildung 8.7 verdeutlicht dies wieder anhand von Mittelwerten.

Wie in den Abbildungen 8.6 und 8.7 zu sehen ist, überwiegen insgesamt die positiven Antworten der Schulleiter im Vergleich zu denen der Lehrer. Maßgeblich hierfür könnte sein, dass die Schulleiter die Tests vordergründig aus ihrer Leitungsposition heraus als ein aktivierendes Instrument im Schulentwicklungsprozess betrachten, welches vielschichtige Prozesse in Bezug auf die Unterrichtsqualität beziehungsweise die Zusammenarbeit mit den Schülern und Eltern und auch der Kollegen in den Fachschaften initiiert. Umfang und Kontrolle der Durchführung dieser Maßnahmen und die Ergebnisse dieser Prozesse sind an dieser Stelle nicht berück-

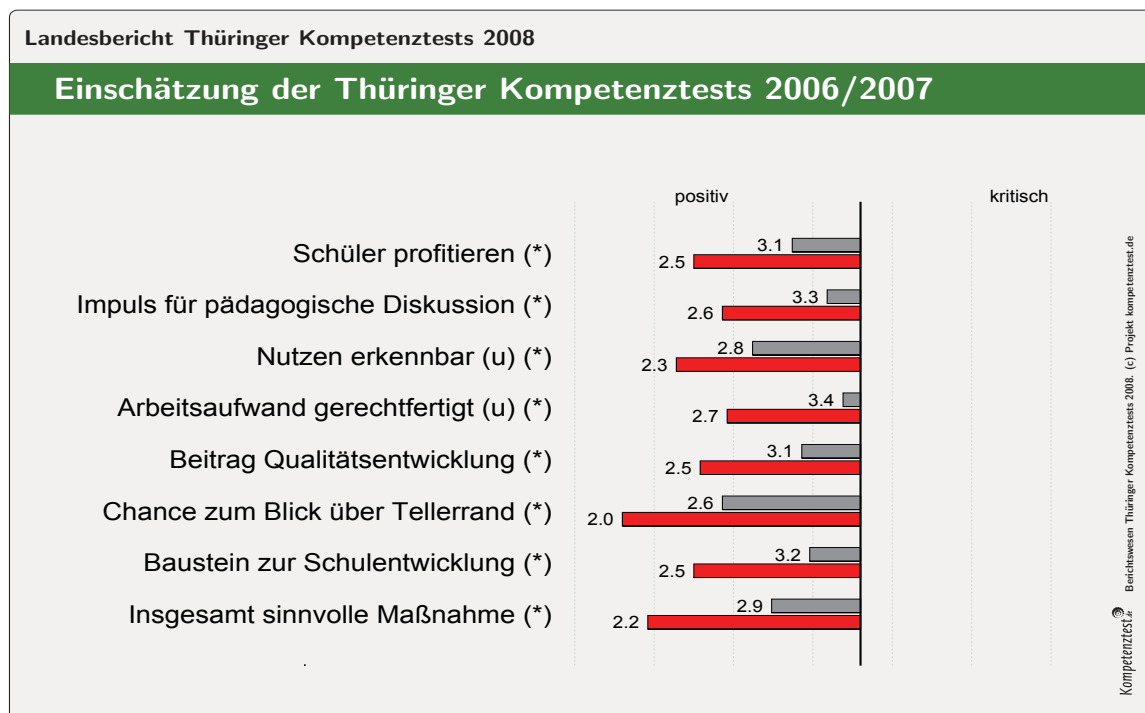


Abbildung 8.6: Einschätzung der Thüringer Kompetenztests 2006/2007 durch Lehrer und Schulleiter aller Schularten im Vergleich

Mittelwerte Lehrer (N>271; grau) und Schulleiter (N>86; rot); Skalenmittelpunkt=3,5; (u) umgepol; (*) Vergleich signifikant

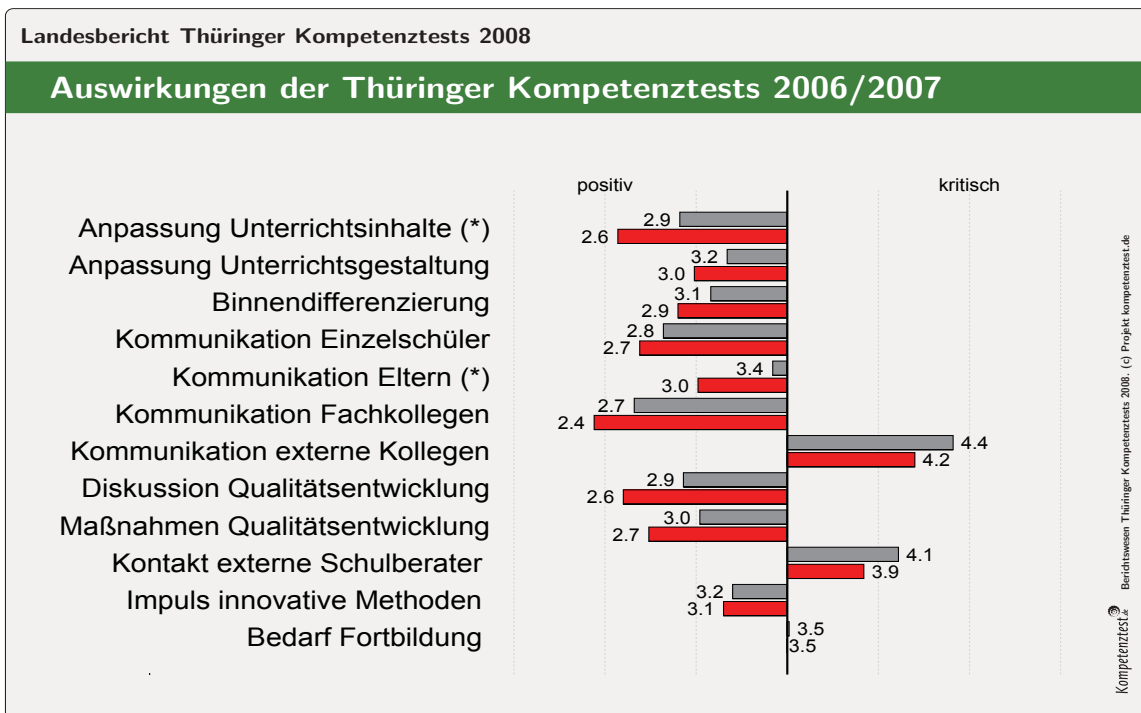


Abbildung 8.7: Auswirkungen der Thüringer Kompetenztests 2006/2007 durch die Lehrer und Schulleiter aller Schularten im Vergleich

Mittelwerte Lehrer (N>271; grau) und Schulleiter (N>86; rot); Skalenmittelpunkt=3,5; (u) umgepolt; (*) Vergleich signifikant

sichtigt. Es bleibt zu untersuchen, inwieweit die Schulleiter ihre Aussagen tatsächlich aufgrund von Hospitationen und den Jahresplanungen der Lehrer treffen und sie so die eingeleiteten Maßnahmen realistisch einschätzen können, diese fortlaufend begleiten und evaluieren. Ein kritischer Blick der Schulleitung auf die konkrete Art und Weise des Umgangs mit den Berichten und der eingeleiteten Maßnahmen ist in jedem Fall dringend erforderlich.

8.4.3 Schularttypische Besonderheiten

Schulartübergreifend verfolgen die Kompetenztests Ziele auf mehreren Ebenen: der Ebene der Schülerleistung, des Unterrichts, der Schulentwicklung sowie die Schulamts- und Landesebene. Exemplarisch sei die Ebene der Schülerleistung herausgegriffen, da hier eine Übereinstimmung der Meinungen festzustellen ist. So sind sich die Schulleiter und Lehrer aller Schularten in den freien Antworten zu den Kompetenztests darin einig, dass die Tests eine sehr gute Möglichkeit der objektiven Leistungsmessung innerhalb der Klasse, der Klassenstufe einer Schule und auch im Vergleich mit anderen Schulen darstellen. Dennoch gibt es schularttypische Unterschiede und Tendenzen hinsichtlich der Annahme der Tests, auf die nun eingegangen werden soll. Die Auswertung bezieht sich dabei immer auf die Aussagen der Schulleiter und Lehrer der Grundschulen, der Regelschulen und Gymnasien. Dabei werden aufgrund der geringen Anzahl der an den Kompetenztests beteiligten Förderschulen für diese Befragungsgruppe keine eigenen Ergebnisse vorgelegt.

Einschätzung der Tests

Bei den im Abschnitt 8.4.2 bereits genannten Kriterien der Einschätzung der Tests als eine sinnvolle Maßnahme und dem Profitieren der Schüler von den Tests liegt der Mittelwert der Grundschullehrer zwischen 2 bis 2,2, bei den Regelschulen und Gymnasien zwischen 3 bis

3,3 (siehe Erklärung Mittelwert zu Beginn des Kapitels 8.4). Hervorzuheben sind weitere positive Aspekte, die die Grundschullehrer darin sehen, dass die Tests die Chance zum Blick über den Tellerrand bieten (Mittelwert 1,9), Impulse für die pädagogische Diskussion liefern (Mittelwert 2,3) und einen Beitrag zur Qualitätsentwicklung an der Schule leisten (Mittelwert 2,2). Im Gegensatz zu diesen vorwiegend zustimmenden Antworten der Grundschulen sind es mehrheitlich die Regelschulen und Gymnasien, die den Arbeitsaufwand durch den resultierenden Nutzen nicht gerechtfertigt sehen (Regelschule / Gymnasium: Mittelwert 3,5, Grundschule: Mittelwert 2,6). Bei der Frage nach einer insgesamt sinnvollen Maßnahme liegt der Mittelwert nur bei 3,5 (Grundschule: 2,0), ob die Schüler profitieren bei 3,3 (Grundschule: 2,2) und beim Impuls für pädagogische Diskussionen nur bei 3,6 (Grundschule: 2,3), um nur einige gravierende Abweichungen zu nennen.

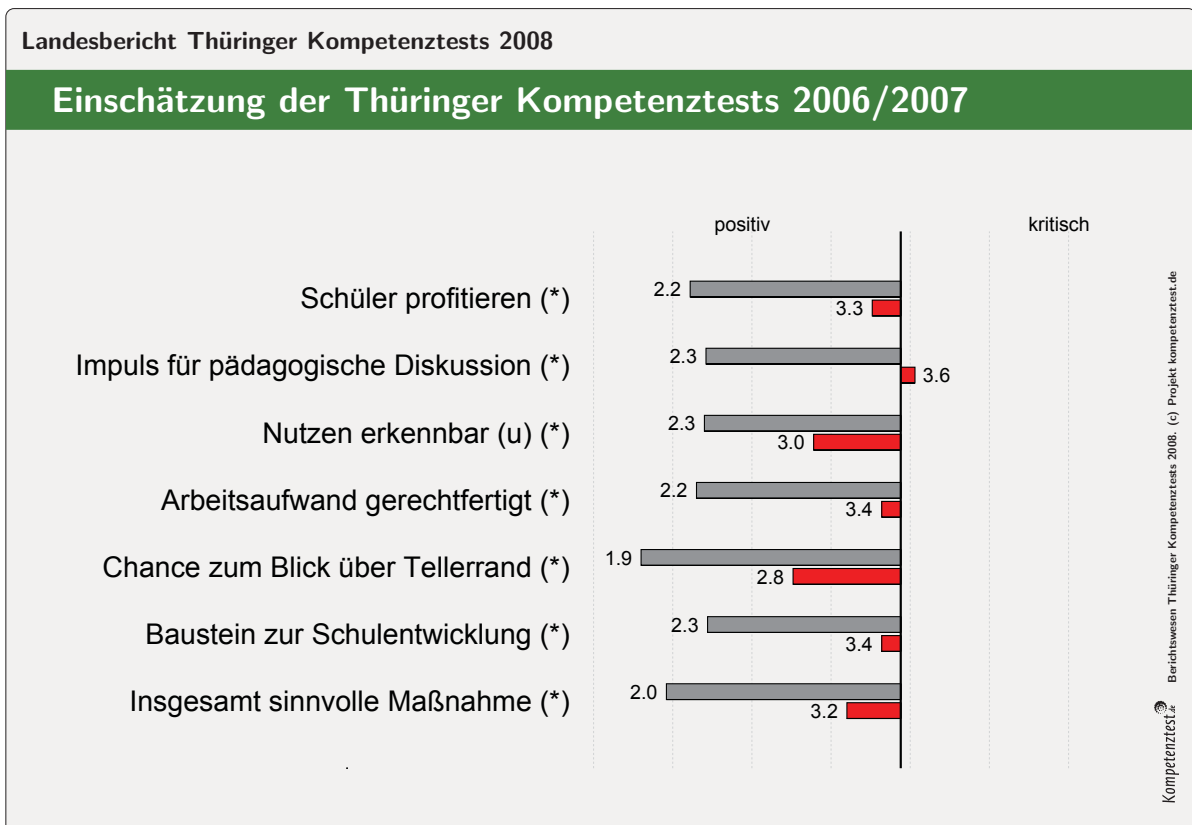


Abbildung 8.8: Einschätzung der Thüringer Kompetenztests 2006 / 2007 durch alle Lehrer und Schulleiter der Grundschulen im Vergleich zu den Regelschulen und Gymnasien
Mittelwerte Grundschulen (N>127; grau) und Regelschulen / Gymnasien (N>214; rot); Skalennittelpunkt=3,5; (u) umgepolt; (*) Vergleich signifikant

Bemerkenswert ist, dass die erheblich besseren Werte der Grundschulen (vergleiche Abbildung 8.8) nicht mit der von ihnen in den freien Antworten geäußerten Kritik an den Tests konform gehen. So bemängelten die Grundschullehrer zum Beispiel, dass die Kinder eine derartige Prüfungssituation nicht gewöhnt sind und der enorme Druck die realen Leistungen verzerren würde. Außerdem wurden immer wieder Probleme mit dem Termin der Tests und mit dem Stand der Lehrplanerfüllung genannt, da dieser auf zwei Schuljahre konzipiert ist.

Auswirkungen der Tests

Bezüglich der Auswirkungen der Tests ist der Durchschnitt der Antworten der weiterführenden Schulen im Jahr 2006 / 07 im Vergleich zu denen der Grundschulen ebenfalls erheblich negativer. Wie in Abbildung 8.9 zu sehen ist, liegen signifikante Unterschiede in der Kommunikation

mit den Eltern, der Kommunikation mit dem Einzelschüler, in den Maßnahmen und der Diskussion zur Qualitätsentwicklung sowie den Impulsen für innovative Methoden. Auch ist aufgrund der Antworten der Bedarf der Regelschul- und Gymnasiallehrer an einer Fortbildung zu den Tests erheblich geringer als der der Grundschullehrer. Als einzige übereinstimmende Auswirkung scheint bei den Regelschul-, Gymnasial- und Grundschullehrern die Kommunikation mit externen Lehrern und Kontakte zu externen Schulberatern nur eine untergeordnete Rolle zu spielen.

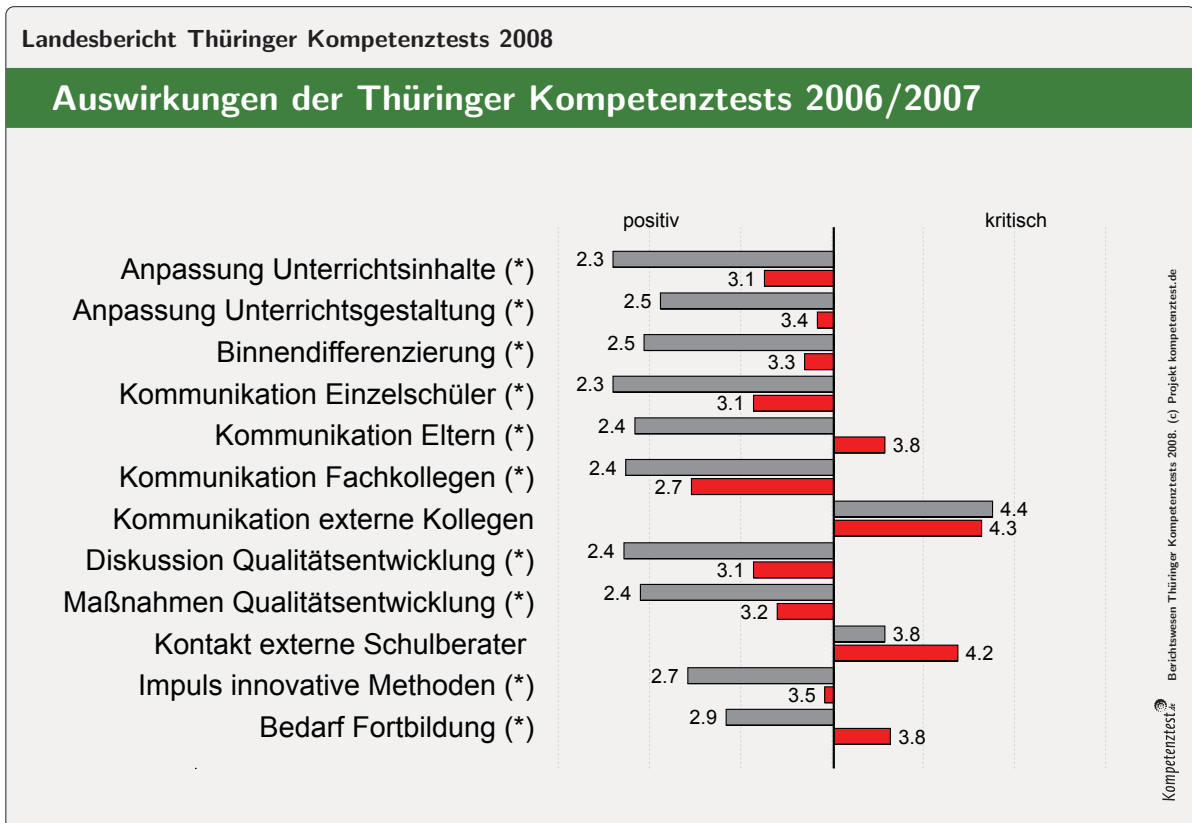


Abbildung 8.9: Auswirkungen der Thüringer Kompetenztests 2006/2007, eingeschätzt durch alle Lehrer und Schulleiter der Grundschulen im Vergleich zu den Regelschulen und Gymnasien

Mittelwerte Grundschulen (N>127; grau) und Regelschulen / Gymnasien (N>214; rot); Skalenmittelpunkt=3,5; (u) umgepolt; (*) Vergleich signifikant

Annahme der Tests

An dieser Stelle werden exemplarisch einige Erklärungsversuche gegeben, die möglicherweise für die unterschiedliche Akzeptanz der Tests an den verschiedenen Schularten mitverantwortlich sind.

Insgesamt kann aufgrund der Einschätzung und der Auswirkungen der Tests von einer positiven Annahme der Tests an den **Grundschulen** gesprochen werden. Mögliche Gründe dieser guten Annahme wurden in Expertengesprächen über die Erfahrungen im Umgang mit den Tests und aus dem Lehreralltag an den Grundschulen deutlich:

die Mehrzahl der Lehrer schon vertraut mit den Tests ist

Durch den 4- oder 2-Jahresrhythmus des Lehrereinsatzes und aufgrund der Tatsache, dass die meisten Grundschullehrer Deutsch und Mathematik unterrichten, ist es wahrscheinlich, dass die Lehrer die Tests wiederholt durchführen. So sind ihnen Ablauf und Auswertung bekannt, es könnte sich lediglich eine sogenannte Testmüdigkeit eingestellt haben.

- wahrscheinlich eine intensivere Zusammenarbeit in den Grundschulkollegien stattfindet

Begründet werden könnte dies zum Beispiel mit den zahlenmäßig oft relativ kleinen Kollegien, dem geringen Fächerspektrum (fast jeder Grundschullehrer unterrichtet selbst in den Testfächern) und der relativ eingegrenzten Zeitspanne von der ersten bis sechste Stunde, in der der Unterricht stattfindet. All dies erleichtert die Kommunikation der Kollegen untereinander und somit die Modalitäten der Auswertung und Nutzung der Tests.

- der Klassenleiter zum Teil beziehungsweise meist selbst der Deutsch- und / oder der Mathematiklehrer ist

Somit hat der Klassenleiter die Testergebnisse meist selbst vorliegen und kann die Tests in den Elterngesprächen fachgerecht auswerten, ohne erst Informationen von anderen Kollegen einholen zu müssen.

- meist von Klassenstufe 3 zu Klassenstufe 4 kein Lehrerwechsel stattfindet

Der Fachlehrer kann somit auch im folgenden Schuljahr nach Durchführung der Tests kontinuierlich an den Stärken und Schwächen der Schüler in seinem eigenen Unterricht arbeiten.

- die Tests als ein ergänzendes Mittel der objektiven Bewertung der Schülerleistung angesehen werden

Die Ergebnisse der Kompetenztests in Klassenstufe 3 liefern dem Lehrer für sich selbst und im Gespräch mit den Eltern ein zusätzliches Mittel der Bestätigung der Notengebung, da diese erst

ab dem zweiten Halbjahr der Klassenstufe 2 stattfindet.

- die Grundschullehrer möglicherweise in ihrem Arbeitsalltag mehr Zeit haben

Zwar unterrichten die Grundschullehrer mit 27 Wochenstunden eine Wochenstunde mehr als ihre Kollegen an den weiterführenden Schulen, doch sind die Korrekturen in den Klassenstufen 1 bis 4 weniger aufwendig als an den weiterführenden Schulen. Das benötigte Fachwissen der Lehrer ist begrenzt auf vier Schuljahre und nicht, wie in den weiterführenden Schulen, auf sechs bis acht oder neun Jahre, und somit wiederholen sich die Unterrichtsvorbereitungen ebenfalls in diesem Rhythmus. Prüfungen gibt es an der Grundschule keine.

Die Annahme der Tests an den **Regelschulen und Gymnasien** fällt deutlich negativer aus als an den Grundschulen. Bei Experteninterviews ergaben sich folgende, mögliche Ursachen aus dem Arbeitsalltag der Lehrer:

- relativ viele Neuerungen im Schulalltag in letzter Zeit

Beispielsweise werden im Lehreralltag neben dem normalen Unterricht die Einarbeitung in neue Lehrpläne, die Umsetzung neuer Prüfungsanforderungen sowie die Vorbereitung auf die Umstrukturierung der gymnasialen Oberstufe und der Sekundarstufe I erwartet. Deshalb fühlen sich einige Lehrer regelrecht überfordert.

- der den Test durchführende Lehrer oft nur ein Fachlehrer in der Klasse ist

Um den Zielen der Tests auf der Schülerebene gerecht zu werden, müssten die Ergebnisse der verschiedenen Fächer beim Klassenleiter zusammengeführt werden. Das bedeutet, dass der Klassenleiter gemeinsam mit den betreffenden Fachlehrern die Ergebnisse der Tests (im Schuljahr 2007/08 in Klassenstufe 6 drei, in Klassenstufe 8 zwei) individuell für jeden Schüler auswerten müsste, um daraus Schlussfolgerungen für die Arbeit mit den Eltern und die konkrete Förderung des Schülers ableiten zu können. Hier könnte die Schulleitung die Zusammenarbeit der Kollegen durch eine Terminvorgabe steuern.

- teilweise Fachlehrerwechsel zum nächsten Schuljahr stattfindet

Das erschwert die langfristige Arbeit mit den Ergebnisberichten. Es ist anzunehmen, dass der betreffende Fachlehrer die Ergebnisse der Tests mit dem neuen Fachlehrer auswertet, die Relevanz für die Veränderung des eigenen Unterrichts konkret in dieser Klasse aber geht verloren.

- die Lehrer eventuell weniger Interesse an den Tests haben

Die Fachlehrer unterrichten in den Klassenstufen 5-10/12, einige vorrangig in der Oberstufe, und sind so seltener in die Tests involviert. Das könnte zur Folge haben, dass sie sich noch nicht so intensiv mit den Tests beschäftigt haben, diese womöglich zum ersten Mal durchführen und sie mehr als „verordnete Maßnahme von Oben“ ansehen und dabei den Nutzen für sich und den eigenen Unterricht noch nicht in vollem Umfang erkennen.

8.4.4 Fazit

Zusammenfassend lässt sich aus den Auswertungen feststellen, dass sich die Kompetenztests an den Schulen gut etabliert haben. Allerdings beschreiben die vorliegenden Daten die Tests aus dem Schuljahr 2006/2007. Die Evaluation der Kompetenztests 2008 wird erst im November und Dezember 2008 durchgeführt. Die Ergebnisse werden im Landesbericht 2009 veröffentlicht.

Insgesamt schätzen die Lehrer und Schulleiter aller Schularten die Tests des Schuljahrs 2006/2007 ähnlich wie im Vorjahr ein, wobei einige wenige Kriterien relevante Verschlechterungen darstellen. So scheinen die Tests zum Beispiel weniger bedeutend für die Schulentwicklung und den Blick über den Tellerrand zu sein. Daraus ergeben sich Reserven für die Arbeit mit den Testergebnissen auf der Ebene der Schulentwicklung. Die Kompetenztests stellen ein Instrument der individuellen Förderung der Schüler dar und sollten im schulinternen Konzept besondere Beachtung finden.

Hinsichtlich der Akzeptanz der Tests sehen die Schulleiter aller Schularten die Tests insgesamt positiver als die Lehrer. Da die Leitungsebene maßgeblich für Schulentwicklungsprozesse verantwortlich ist, sollten die positiven Impulse der Schulleiter intensiver dazu genutzt werden, an den Schulen eine entsprechende Beratungskultur und einen offenen Erfahrungsaustausch unter den Kollegen, aber auch mit den Schülern und Eltern weiterzuentwickeln beziehungsweise im Jahresplan fest zu etablieren. Durch organisierte, kontinuierliche und verbindliche Teamarbeit kann den Kollegen der Umgang mit den Tests erleichtert werden, so dass die Tests tatsächlich eine Grundlage für Prozesse der Schulentwicklung und für die Veränderung des Unterrichts darstellen.

Die Betrachtung der Annahme der Tests durch die einzelnen Schularten ergab, dass diese bei den Grundschulen im Vergleich zum Vorjahr gleich geblieben oder sogar ein wenig positiver geworden ist, sich die der Regelschul- und Gymnasialkollegen aber negativer als im Vorjahr darstellt, wodurch sich der zahlenmäßige Abstand im Vergleich der Mittelwerte zur Grundschule vergrößerte. Die gegebenen Erklärungsversuche für diesen Unterschied legen den Schluss nahe, dass aufgrund einer Reihe von organisatorischen und zeitlichen Problemen die vielseitigen Potentiale der Tests besonders an den weiterführenden Schulen noch nicht ausreichend genutzt werden. Möglicherweise würde sich das Erkennen eines Nutzens der Testergebnisse für die eigene pädagogische Arbeit und die konkrete Gestaltung des eigenen Unterrichts in Zukunft auch auf die motivationale Ebene im Umgang mit den Tests positiv auswirken. Zur Zeit zeigt sich die teilweise Ablehnung oder Nichtakzeptanz einiger Lehrer gegenüber den Tests beispielsweise noch darin, dass sie den Arbeitsmehraufwand über den Nutzen der Tests für die gesamte Lehr- und Lernarbeit stellen. Des Weiteren scheint es so, dass der Schwerpunkt

der Auswertung der Tests auf der personellen Ebene, das heißt in der Kommunikation mit dem Schüler, den Eltern und den Kollegen liegt (siehe Abbildung 8.5, leichte Verbesserung dieser Werte im Vergleich zum Vorjahr), während die Evaluation des eigenen Unterrichts noch zu wenig fokussiert wird. Wahrscheinlich werden die verschiedenen Angaben der Ergebnisberichte noch nicht von allen Kollegen als Instrument für die Veränderung des eigenen Unterrichts erkannt, und somit auch nicht genutzt. Hier könnten schulinterne Fortbildungen und die vernetzte Arbeit in den Fachschaften besonders an den weiterführenden Schulen einen wesentlichen Beitrag dazu leisten, die vielseitigen Potentiale der Tests im Rahmen der Schulentwicklung weiter auszuschöpfen.

9 Zusammenfassung und Ausblick

9.1 Entwicklung und Durchführung der Kompetenztests 2008

Während die Durchführung der Kompetenztests im Schuljahr 2007 / 2008 weitgehend unverändert blieb, kam es im Bereich der Testentwicklung zu einer ganzen Reihe von Neuerungen. An erster Stelle ist die gemeinsame Entwicklung und Prüfung der Testaufgaben im Rahmen von Kooperationen verschiedener Bundesländer zu nennen. Die Tests in Klassenstufe 3 wurden vom Projekt VerA an der Universität Landau im Auftrag aller 16 Bundesländer entwickelt. In Klassenstufe 6 hat ein Verbundprojekt von 6 Bundesländern diese Aufgabe übernommen. Hier arbeiten die Länder Hamburg, Hessen, Mecklenburg-Vorpommern, Schleswig-Holstein, Sachsen und Thüringen erstmals gemeinsam an einer solchen Aufgabe. Dabei wurde neben den Tests in den Bereichen Deutsch, Mathematik und Englisch-Hörverstehen auch ein Test im Bereich Englisch-Leseverstehen entwickelt, der dieses Jahr erstmals zum Einsatz kam. Die Testentwicklung in Klassenstufe 8 erfolgt durch das Institut zur Qualitätsentwicklung im Bildungswesen (IQB) in Berlin. Durch diese Kooperationen konnte das inhaltliche und methodische Niveau der Tests weiter erhöht und vereinheitlicht werden. Die Tests werden nun alle, ähnlich wie die großen internationalen Schulleistungsstudien, auf der Basis der Item Response Theorie entwickelt. Dies erfordert einerseits eine aufwändigere Aufgabenentwicklung und umfangreiche Vortests, schafft aber andererseits die Voraussetzung für weitergehende Maßnahmen im Hinblick auf Unterrichtsentwicklung, wie z. B. die Entwicklung von Kompetenzstufenmodellen oder den Einsatz verschiedener Testhefte unterschiedlicher Schwierigkeit, deren Ergebnisse trotzdem auf einer gemeinsamen Skala verglichen werden können.

Die Durchführung der Tests mit der Anmeldung der Schulen, der Erfassung der Ergebnisse und deren Rückmeldung an die Schulen blieb weiterhin den einzelnen Ländern überlassen. In Thüringen wurde diese Aufgabe wie in den letzten Jahren vom Projekt *kompetenztest.de* der Universität Jena übernommen. Aufgrund der guten Erfahrungen der vergangenen Jahre hatten sich auch andere Länder wie Mecklenburg-Vorpommern und Sachsen entschlossen, die Universität Jena mit dieser Aufgabe zu betrauen. Die einzelnen Arbeitsschritte der Testdurchführung erwiesen sich aufgrund des in den letzten Jahren ausgereiften Systems als unproblematisch. Auch die im Vorjahr noch vermehrt aufgetretenen Probleme bei der Anmeldung zum Thüringer Schülerlängsschnitt waren nach der Überarbeitung der Hochladeprozedur vernachlässigbar.

Für die Thüringer Lehrer machten sich diese Änderungen bei der Testentwicklung in zum Teil neuen und ungewohnten Aufgaben- und Korrekturformaten bemerkbar. Bei allen Tests (mit Ausnahme des Mathematiktests der Klassenstufe 6) hatte sich die Anzahl der bei der Dateneingabe einzugebenden (Teil-)Aufgaben im Vergleich zum Vorjahr erhöht, zum Teil mussten mehr als doppelt so viele Items eingegeben werden. Entsprechend erhöhte sich der zeitliche Aufwand der Lehrkräfte für die Dateneingabe von durchschnittlich unter 20 Minuten auf etwa 30 Minuten pro Klasse und Fach, was sich in einer schlechteren Bewertung durch die Fachlehrer widerspiegelte. Die Dateneingabe wurde in diesem Jahr mit der Durchschnittsnote 1,82 (im Vorjahr: 1,74) bewertet, was allerdings immer noch als sehr zufrieden stellend gelten kann. Problematischer als der Mehraufwand bei der Dateneingabe erwiesen sich vereinzelte Rechtschreibfehler in den Testheften und ein falscher Korrekturhinweis bei den Tests in Klassenstufe

6. Insbesondere die Rechtschreibfehler im Fach Deutsch sorgten für nahe liegende und z. T. sehr ironische Kommentare durch Fachlehrer. Generell ist jedoch anzumerken, dass die unter veränderten Bedingungen entwickelten Tests von den Fachlehrern in der Regel gut angenommen wurden. Als problematisch wird der auf den Beginn des zweiten Halbjahres vorverlegte Testtermin in Klassenstufe 6 und 8 von den Lehrern eingeschätzt. Dies resultiert vor allem aus der Tatsache, dass noch immer viele Lehrer eine möglichst vollständige Lösung möglichst aller Aufgaben durch ihre Schüler als wichtigstes Ziel bei der Testdurchführung sehen, was durch den frühen Testzeitpunkt erschwert wird. Hier herrschen offenbar noch immer Informationsdefizite über Ziel und mögliche Nutzung der Tests.

Erfreuliches gibt es hinsichtlich der Erfassung der diagnostischen Kompetenz der Fachlehrer zu berichten. Hier können die Fachlehrer kurz vor Testdurchführung einen Teil der Testaufgaben einsehen und die erwartete Lösungshäufigkeit für ihre Klasse einschätzen. Der Vergleich zwischen der Einschätzung und den tatsächlichen Lösungshäufigkeiten gibt dem Lehrer Rückmeldung über die eigene diagnostische Kompetenz. Die Teilnahmezahlen bei diesem freiwilligen und anonymen Verfahren liegen im Durchschnitt bei 27,9 %, was einen leichten Anstieg im Vergleich zum Vorjahr bedeutet. Lediglich im Fach Mathematik Klasse 8 ist ein Rückgang zu verzeichnen. Einer der Gründe dürfte sein, dass dieser Test im Jahr 2007 erstmals durchgeführt wurde und viele Lehrer die Möglichkeit zur Erfassung der diagnostischen Kompetenz nutzten, um diesen für sie bis dahin unbekannten Test noch vor ihrer Durchführung kennen zu lernen.

Im Hinblick auf die Zweitkorrekturen wurde das sehr gute Ergebnis der letzten Jahre erneut bestätigt. Nur in sehr wenigen Einzelfällen kam es zu bedeutsamen Abweichungen zwischen Erst- und Zweitkorrektur. Am ehesten problematisch war wie in den letzten Jahren die Situation beim Test im Fach Deutsch, Klassenstufe 6. Die Aufgaben zum schriftlichen Sprachgebrauch sind traditionell schwierig im Hinblick auf eindeutige Bepunktung.

9.2 Nutzen und Nutzung der Kompetenztests

Die Kompetenztests stellen den Schulen eine Vielzahl von Informationen zur Verfügung, die auf verschiedenen Ebenen genutzt werden können: Für verbesserte Einschätzung des einzelnen Schülers, für die Evaluation und Weiterentwicklung des eigenen Unterrichts und für die Schulentwicklung. Die Ergebnisse sind hilfreich für Diagnose und Entwicklung, sie erlauben aber keine unmittelbaren Rückschlüsse auf Unterrichts- und Schulqualität, sondern müssen mit dem Wissen um die konkreten Bedingungen an der einzelnen Schule interpretiert werden. Daher sind sie z. B. nicht für ein Ranking geeignet. Konsequenter Weise wird in Thüringen auf Eigenverantwortung der Schulen im Umgang mit den Testresultaten gesetzt. Entscheidend ist, dass mit den Ergebnisberichten sinnvoll gearbeitet wird. Das bedeutet eine genaue Analyse der Ergebnisse, das Ableiten von Maßnahmen, deren Durchführung und Überprüfung. Diese Vorgehensweise ist für viele Lehrkräfte und Schulleiter noch ungewohnt. Ein solcher Wandel zu einer schulischen Evaluationskultur braucht Zeit und Unterstützung. Anders als im Jahr 2007 wurden diesbezüglich jedoch nur wenige Fortbildungen oder andere Unterstützungsmaßnahmen mit dem spezifischen Ziel der verbesserten Nutzung der Tests angeboten.

Hinweise auf die Nutzung der Tests an den Schulen geben z. B. die Abrufzahlen der Ergebnisberichte sowie die jährlichen Befragungen der Lehrkräfte. Bei den Abrufzahlen ist im Vergleich zum Vorjahr ein geringfügiger Rückgang zu verzeichnen. So wurden in immerhin 15 % aller Klassen keine Berichte mit Vergleichsdaten abgerufen. Bei den Schulberichten betrug diese Quote 22 %. Dies bedeutet allerdings bei diesen Berichten einen deutlichen Anstieg der Abrufzahlen im Vergleich zum Vorjahr. Die Schulberichte werden erst seit 2006 angeboten und sind an den Schulen noch deutlich weniger bekannt als die Klassenberichte.

Die jährliche Befragung von Schulleitern und Fachlehrer zur Nutzung der Kompetenztests findet jeweils im Dezember statt. Entsprechend gibt dieser Landesbericht die Ergebnisse zu den Tests 2007 wieder. Hier zeigt sich bei der Bewertung und Nutzung der Tests eine Stagnation, teilweise ist sogar ein Rückgang im Vergleich zu den Vorjahren zu verzeichnen. Diese bedenkliche Tendenz betrifft vornehmlich die weiterführenden Schulen. In den Grundschulen fallen die Bewertungen deutlich besser aus. In den Abrufzahlen sind diese Unterschiede hingegen nicht wieder zu finden. Auffällig sind die Unterschiede bei den Bewertungen durch Schulleiter und Fachlehrer. Erstere schätzen den Nutzen der Tests deutlich höher ein als die Fachkollegen.

Kritische Punkte

Im Hinblick auf weitere Verbesserungen der Tests hat das Schuljahr 2007/2008 eine Reihe von kritischen Punkten aufgezeigt. Manche dieser Punkte sind seit mehreren Jahren bekannt und nicht so leicht abzustellen, andere traten zum ersten Mal auf.

Im Hinblick auf die Aufgabenentwicklung muss kritisch angemerkt werden, dass die Tests im Fach Englisch in Klassenstufe 6 deutlich leichter waren als die übrigen Tests. In so mancher Gymnasialklasse wurden durchschnittlich mehr als 90 % aller Aufgaben gelöst, wohingegen die gleichen Klassen in den Fächern Deutsch und Mathematik deutlich schlechter abschnitten. Die (nahe liegende) Interpretation einer höheren Kompetenz der Schüler im Fach Englisch ist aber nicht gerechtfertigt. Die Unterschiede resultieren aus unterschiedlichen Formen der Testkonzeption. Im Fach Englisch steht mit dem Europäischen Referenzrahmen ein Bezugsrahmen zur Verfügung, mit Hilfe dessen die Schwierigkeiten der in Klassenstufe 6 zu lösenden Aufgaben festgelegt sind. Konkret sind dies die Stufen A1 und A2. Bei der Testkonstruktion wurden ausschließlich solche Aufgaben verwendet. Dies führt dazu, dass die Englischtests im oberen Schwierigkeitsbereich nicht differenzieren und insgesamt leichter ausfallen. So wertvoll der kriteriale Bezug der Englischaufgaben ist, so wichtig wäre diese Differenzierung im oberen Bereich sowie die Vergleichbarkeit der Gesamtschwierigkeit zwischen den Fächern. Andernfalls ist der Nutzen des Tests für den Fachlehrer deutlich reduziert, um die Ergebnisse im Fächervergleich laden förmlich zu falschen Interpretationen ein.

Auch im Jahr 2008 ist die Tendenz an den Schulen zu beobachten, die Schüler durch das Üben alter Testaufgaben auf die Tests vorzubereiten. Hier ist klar zu sagen: Das Ziel muss der Erwerb von Kompetenzen sein, die Testaufgaben sind lediglich Indikatoren für diese Kompetenzen. Durch Training bestimmter Aufgaben können zwar Testergebnisse verbessert werden, doch ohne eine tiefer gehende Erweiterung der zugrunde liegenden Kompetenzen ist dieses Vorgehen nicht nur unnötig sondern kontraproduktiv, die dafür verwendete Zeit kann nutzbringender investiert werden.

Die Ergebnisse der Analysen der Lehrefragebögen sowie der Schülerdaten weisen auf eine Reihe von kritischen Punkten hin. So zeigt sich ein bedeutsamer Zusammenhang zwischen dem Migrationshintergrund bzw. der Bildungsnähe des Elternhauses und den Testergebnissen der Schüler. Die Stärke des Zusammenhanges ist in den letzten Jahren im Wesentlichen stabil geblieben. Erneut zeigen sich dieses Jahr auch beachtliche Divergenzen der unterschiedlichen Fachlehrer hinsichtlich der Einschätzung, ob bei ihren Schülern besondere Lernschwierigkeiten (BL) vorliegen. Es kann leider nicht wie selbstverständlich davon ausgegangen werden, dass das Vorliegen einer BL bei einem Schüler den verschiedenen Fachlehrern bekannt ist. Mehr professioneller Austausch mit den Kollegen, etwa in Klassenkonferenzen, wäre hier notwendig, um die Förderung der Schüler zu verbessern.

9.3 Perspektiven

Die Ergebnisse der Befragung zur Bewertung und Nutzung der Kompetenztests durch die Fachlehrer zeigen, dass diese Nutzung kein Selbstläufer ist. Dies wird durch stagnierende und teilweise rückläufige Bewertungen an den weiterführenden Schulen genauso wie durch die immer noch steigerungsfähigen Zahlen beim Herunterladen von Ergebnisberichten mit Vergleichsdaten verdeutlicht. Nach der erfolgreichen Einführung und Etablierung des Evaluationsinstrumentes „Kompetenztests“ bedarf es andauernder und nachhaltiger Maßnahmen, damit das gesamte Potenzial der Tests von den Lehrern genutzt wird. Ein tiefer gehender Blick auf die Befragungsdaten zeigt durchaus unterschiedliche Aspekte bei der Nutzung der Tests. Einerseits werden die Tests von weiten Teilen der Lehrerschaft durchaus als Chance gesehen, mehr über ihre Schüler zu erfahren. Hingegen gibt es verbreitete Widerstände, aus den Testergebnissen auf Stärken und Schwächen des eigenen Unterrichts zu schließen und daraus Konsequenzen für die Weiterentwicklung der eigenen pädagogisch-professionellen Kompetenzen zu ziehen. Die Tests besitzen durchaus diesbezügliches Potenzial, jedoch ist ein solches Vorgehen ungleich riskanter und schwieriger, streift es doch so sensible Bereiche wie das professionelle Selbstwertgefühl. Bisher wird eine offene Auseinandersetzung mit eigenen Stärken und Schwächen in der Schule nur selten als Professionalität erkannt und wertgeschätzt, und dies kennzeichnet keineswegs nur den Lehrerberuf. Daher ist es unrealistisch zu erwarten, dass die entsprechenden Potenziale der Tests in kurzer Zeit von der Mehrheit der Lehrkräfte intensiver genutzt werden. Für einen derartigen Wandel hin zu einer Evaluationskultur an den Schulen ist ein langer Atem mit konsistenter und nachhaltiger Unterstützung notwendig. Eine Schlüsselrolle spielt hier sicherlich die Lehrerbildung, in der die kommende Generation von Lehrkräften mit dieser Kultur vertraut gemacht werden und durch welche diese Kultur am ehesten in den Schulen Fußfassen könnte.

Das entscheidende Kriterium für den Erfolg der Kompetenztests bleibt, dass die Tests für die konkrete Arbeit des Lehrers Nutzen zu bringen vermögen. Bisher enthalten die Ergebnisse vor allem soziale Vergleichswerte, welche die Ergebnisse in ähnlichen Klassen oder Schulen zeigen. Hinzu kommen in rudimentärer Form auch kriteriale Informationen. Der Fachlehrer kann z. B. die Lösungshäufigkeiten einzelner Aufgaben oder Aufgabengruppen mit selbstgesteckten Kriterien vergleichen und daraus Schlüsse für den eigenen Unterricht ziehen. Im Zuge der konkreten Festschreibung der Bildungsstandards werden in dieser Hinsicht auch Kriterien für die Erreichung von z. B. minimalen Bildungszielen beschrieben. Zukünftig sollen die Kompetenztests auch diese kriterialen Informationen stärker berücksichtigen. Dazu werden im Jahr 2009 die Testergebnisse in Klassenstufe 8 erstmals neben den Vergleichsdaten auch Angaben zu Kompetenzstufen enthalten. Zudem bietet die jetzige, auf der Item Response Theorie basierende Testentwicklung die Möglichkeit, Schülern gemäß ihrer Kompetenzstufe spezifische Unterstützungsmöglichkeiten in Form von passenden Übungsaufgaben anzubieten. Dies kann z. B. im Rahmen der didaktischen Materialien geschehen, welche den Lehrern zusammen mit den Tests im Hinblick auf die Weiterarbeit im Unterricht angeboten werden.

Perspektivisch bieten die bei der Testentwicklung erstellten Aufgaben die Möglichkeit, mit Hilfe von sogenannten adaptiven Tests auch die Diagnosegenauigkeit auf der Ebene des Einzelschülers deutlich zu erhöhen. Bisher dürfen Testergebnisse auf Schülerebene aufgrund der Messfehlerproblematik nur mit großer Vorsicht interpretiert werden, wohingehend die aggregierten Ergebnisse auf Klassenebene in der Regel eine große Genauigkeit aufweisen. Adaptive Tests könnten gerade für leistungsschwächere Schüler, deren individuelles Testergebnis durch die im Klassenverband einheitlichen Testhefte auf mittlerem Schwierigkeitsniveau eher ungenau ausfallen, wertvolle Information im Hinblick auf eine genaue Diagnose der vorliegenden Kompetenzstufe und darauf basierender optimaler Förderung erbringen.

10 Literatur

Bos, W., u. a. (Hrsg.) (2003): Erste Ergebnisse aus IGLU. Schülerleistungen am Ende der vierten Jahrgangsstufe im internationalen Vergleich. Waxmann: Münster.

Cohen, J. (1988): Statistical Power Analysis for the Behavioral Sciences. Hillsdale: Erlbaum.

Klieme, E., Avenarius, H., Blum, W., et al. (2003). Expertise zur Entwicklung nationaler Bildungsstandards. Bundesministerium für Bildung und Forschung (Hrsg.).

Landis, J. R. & Koch, G. G. (1977): The measurement of observer agreement for categorical data. In: Biometrics (33), pp. 159-174.

Nachtigall, C., Steyer, R. & Wüthrich-Martone, O. (2001): Causal Effects in Empirical Research. In M. May & U. Ostermeier (Eds.), Interdisciplinary perspectives on causality. Bern Studies in the History and Philosophy of Science. Norderstedt: Libri Books on Demand. pp. 81-100.

Nachtigall, C. & Wirtz, M. (2004): Wahrscheinlichkeitsrechnung und Inferenzstatistik. Weinheim: Juventa.

Nachtigall, C. (Hrsg.) (2007). Landesbericht Thüringer Kompetenztests 2007.

<http://www.kompetenztest.de>

PISA-Konsortium Deutschland (Hrsg.) (2007): PISA 2006. Die Ergebnisse der dritten internationalen Vergleichsstudie. Münster: Waxmann-Verlag.

Rost, J. (1996). Lehrbuch Testtheorie Testkonstruktion. (2. Aufl. 2004) Bern: Huber.

ThILLM (2005). Kompetenztests als Beitrag zur Schul- und Unterrichtsentwicklung. Heft Nr. 115 der ThILLM-Reihe „Materialien“.

Wirtz, M. & Caspar, F. (2002): Beurteilerübereinstimmung und Beurteilerreliabilität: Methoden zur Bestimmung und Verbesserung der Zuverlässigkeit von Einschätzungen mittels Kategoriensystemen und Ratingskalen. Göttingen, Bern [u.a.]; Hogrefe, Verlag für Psychologie, S.25.

Impressum:

Dieser Bericht wurde vom Projekt *kompetenztest.de* der Friedrich-Schiller-Universität Jena in Kooperation mit dem Thüringer Institut für Lehrerfortbildung, Lehrplanentwicklung und Medien (Thillm) im Auftrag des Thüringer Kultusministeriums (TKM) erstellt.

Dr. Christof Nachtigall
Dipl.-Psych. Marc Müller
Dipl.-Psych. Isabell Storbeck
B.Sc. Torsten Schmidt

sowie die studentischen Hilfskräfte:
Simone Becker, Linda Gräfe
Dorit Grundmann und Mario ten Venne.

Bei Fragen wenden Sie sich bitte an: kontakt@kompetenztest.de

Erstellungsdatum: 11. November 2008