Seit 1558

Friedrich-Schiller-Universität Jena

*Kompetenztest.de*

**Lehrstuhl für Methodenlehre und Evaluationsforschung (Prof. Rolf Steyer)**
Projektleiter: Dr. Christof Nachtigall

# Landesbericht

## Thüringer Kompetenztests 2010

**Dr. Christof Nachtigall (Hrsg.)**

Kompetenztest.de

# Inhaltsverzeichnis


**1 Die Kompetenztests 2010** **3**
1.1 Ziele der Kompetenztests . . . 3
1.2 Chancen und Risiken der Kompetenztests . . . 4

**2 Testentwicklung** **6**
2.1 Gütekriterien . . . 7

**3 Durchführung und Datenerfassung** **8**
3.1 Anmeldung zu den Kompetenztests 2010 und zum Schülerlängsschnitt . . . 8
3.2 Teilnahme an den Kompetenztests 2010 . . . 9
3.3 Webbasierte Datenerfassung . . . 10
3.4 Auswertung der Hotline . . . 13

**4 Diagnostische Kompetenz** **15**
4.1 Einleitung . . . 15
4.2 Auswertung der diagnostischen Kompetenz für Klassenstufe 3 . . . 17
4.3 Auswertung der diagnostischen Kompetenz für Klassenstufe 6 . . . 19
4.4 Auswertung der diagnostischen Kompetenz für Klassenstufe 8 . . . 23
4.5 Zusammenhänge der Ergebnisse der diagnostischen Kompetenz mit anderen Merkmalen . . . 26
4.6 Fazit . . . 27

**5 Rückmeldung der Testergebnisse** **29**
5.1 Formen der Rückmeldungen . . . 29
5.2 Korrigierter Landesmittelwert und faire Vergleiche . . . 34

**6 Testergebnisse auf Landesebene** **36**
6.1 Erreichte Gesamtpunktzahlen . . . 36
6.2 Erreichte Gesamtpunktzahlen getrennt nach Schulart . . . 42
6.3 Aufgabenschwierigkeiten . . . 47

**7 Analyse leistungsrelevanter Faktoren** **76**
7.1 Zusammenhang von Prozess- und Wirkungsqualität . . . 76
7.1.1 Allgemeines . . . 76
7.1.2 Daten und Stichprobe . . . 76
7.1.3 Ergebnisse und Schlussfolgerungen . . . 77
7.1.4 Diskussion & Anregungen . . . 78
7.2 Einflussfaktoren auf Schülerebene . . . 79
7.2.1 Mathematische und sprachliche Kompetenzen in Abhängigkeit vom Geschlecht . . . 81
7.2.2 Muttersprache . . . 85
7.2.3 Sozioökonomischer Status (SES) . . . 90
7.2.4 Jahre des Schulbesuchs . . . 94
7.2.5 Besondere Lernschwierigkeiten . . . 103
7.2.6 Vorwissen der Schüler . . . 106

7.2.7 Zusammenhang Halbjahresnote und Testergebnisse . . . . . 107
7.2.8 Zusammenfassende Betrachtung . . . . . 110

**8 Rezeption und Evaluation der Tests 113**
8.1 Befragung der Fachlehrer nach der Dateneingabe . . . . . 113
8.2 Rezeption der Testergebnisse an den Schulen . . . . . 118
8.3 Bewertung der Kompetenztests . . . . . 120
8.3.1 Evaluation der Kompetenztests 2009 der Thüringer Grundschulen . . . . 121
8.3.2 Evaluation der Kompetenztests 2010 . . . . . 122
8.3.3 Fazit . . . . . 123

**9 Schlussbemerkung 124**

**10 Literatur 126**

# 1 Die Kompetenztests 2010

Die Thüringer Kompetenztests fanden im Schuljahr 2009 / 2010 zum achten Mal statt. Die Testentwicklung, Durchführung und Rückmeldung der Ergebnisse erfolgte gemäß den in den vergangenen Jahren optimierten Verfahren, welche in diesem Bericht in ihren verschiedenen Schritten ausführlich beschrieben werden.

Die Tests werden jährlich in der Klassenstufe 3 in den Kernfächern Deutsch und Mathematik, in den Klassenstufen 6 und 8 in den Kernfächern Deutsch, Mathematik und Englisch durchgeführt, in Klassenstufe 8 kann statt Englisch auch Französisch gewählt werden. Während die Teilnahme in den Klassenstufen 3 und 8 an allen Tests verpflichtend ist, müssen die Schulen in Klassenstufe 6 nur an mindestens einem Test teilnehmen. Mehr als die Hälfte der Schulen nahm jedoch in allen drei Fächern teil.

Die Durchführung der Tests und die Erstellung und Rückmeldung der Ergebnisse 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 ihrem passwortgeschützten Schulportal zur Verfügung stellte.

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 als Chance und nicht nur 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 der Sicht der Schulleiter und der Fachlehrer gibt. Die Schulleiter stehen den Tests deutlich positiver gegenüber als die Fachlehrer. Es sind weiterhin Anstrengungen nötig, um die Fachlehrer mit den Potenzialen und Nutzungsmöglichkeiten der Kompetenztests vertraut zu machen.

## 1.1 Ziele der Kompetenztests

Mit den jährlichen Kompetenztests wird in einer landesweiten Erhebung der Leistungsstand der Thüringer Schüler in wichtigen Kompetenzbereichen gemessen. 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 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 sollen Anstoß für die Weiterentwicklung des eigenen Unterrichts sein.

## 1.2 Chancen und Risiken der Kompetenztests

Im mittlerweile 8. Jahr ihrer Durchführung stellen die Kompetenztests ein ausgereiftes, etabliertes und akzeptiertes Instrument dar, 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*). Insgesamt sind die Tests Teil eines Kulturwandels an den Schulen hin zu mehr Offenheit und Eigenverantwortlichkeit gegenüber den Ergebnissen der eigenen Arbeit. Sie sind jedoch ungeeignet für ein Ranking der besten Schulen, denn die Ergebnisse spiegeln neben den Leistungen von Schule und Unterricht auch eine Fülle von weiteren Einflüssen wieder, welche der Lehrer nicht oder nur wenig beeinflussen kann. Die Deutungshoheit für die Ergebnisse liegt daher bei den Experten an der Schule selbst, also den Lehrkräften in der Zusammenarbeit mit Schülern und Eltern. Aus diesem Grunde ist es ein wichtiges Prinzip, dass die Schulen selbst über ihre Daten bestimmen können. Durch die alleinige Bereitstellung der Schulergebnisse auf den passwortgeschützten Schulportalen konnte hier in den letzten Jahren viel Vertrauen aufgebaut werden, welches die Voraussetzung für realistische Ergebnisse und deren schulische Nutzung ist. Eine Vielzahl von Lehrern nutzt bereits die Testergebnisse in differenzierter Weise für den eigenen Unterricht. 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. An dieser Stelle 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.

Während das Potenzial der Tests zur Unterrichts- und Schulentwicklung noch lange nicht ausgeschöpft ist, besteht bei den meisten Lehrern ein starkes Zutrauen in die diagnostischen Möglichkeiten der Tests im Hinblick auf den einzelnen Schüler. Aus diesem Grunde wurde in diesem Jahr erstmals im Fach Deutsch in den Klassenstufen 6 und 8 ein Schülerbericht angefertigt und den Lehrkräften zur Verfügung gestellt. Dieser Bericht gibt für den einzelnen Schüler seine Ergebnisse in Gesamtergebnis und den Teilbereichen des Tests wieder. Zusätzlich werden als Vergleichswerte die Ergebnisse aller Thüringer Schüler der gleichen Schulart dargestellt. Nun muss allerdings einschränkend gesagt werden, dass die Tests die Breite des Curriculums hinreichend abbilden müssen, was zu einer weniger großen Messgenauigkeit in den einzelnen Teilbereichen führt. Aus diesem Grunde werden die Ergebnisse in den Schülerberichten mit so genannten Vertrauensintervallen angegeben. Diese Intervalle kennzeichnen einen Bereich, in dem die Schülerleistung unabhängig von Messungenauigkeiten liegt. Es bleibt aber zu beachten, dass die Ergebnisse auf Schülerebene niemals alleine, sondern immer mit anderen diagnostischen Informationen interpretiert werden sollten.

# 2 Testentwicklung

Bei der Entwicklung der Kompetenztests im Schuljahr 2009 / 2010 gab es im Vergleich zum Vorjahr bedeutsame Veränderungen. Nachdem bisher die Testentwicklung in Klassenstufe 3 vom Projekt VerA an der Universität Landau vorgenommen wurde, ging dieser Aufgabenbereich im Schuljahr 2009 / 2010 erstmals an das Institut zur Qualitätsentwicklung im Bildungswesen (IQB) in Berlin, welches seit dem Schuljahr 2008 / 2009 auch die Tests in Klassenstufe 8 erstellt. Die Tests für die Klassenstufe 6 werden im Verbund der Bundesländer Hamburg, Hessen, Mecklenburg-Vorpommern, Sachsen, Schleswig-Holstein und Thüringen entwickelt. Dabei sind 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 wie bei den Tests in den Klassenstufen 3 und 8 im Hinblick auf Vorgehensweise und methodische Richtlinien an den Arbeiten zur Normierung der Bildungsstandards, welche ebenfalls vom IQB in Berlin durchgeführt werden. Auf diese Weise wird sichergestellt, dass die Tests wissenschaftlichen Gütekriterien genügen. Konkret erfolgt das Erstellen der Aufgaben (Itemwriting) und eine erste Vortestung (Präpilotierung) durch speziell geschulte Fachlehrer aus allen beteiligten Bundesländern. Die entwickelten Aufgaben müssen sich dann mehreren Rückmeldungen unterziehen, in denen sie mit Hilfe von fachdidaktischer und psychometrischer Expertise auf ihre Tauglichkeit für die Kompetenztests geprüft und optimiert wurden. Anders als die früheren Kompetenztests basierte die finale Auswahl der Testitems auf der Item Response Theorie[1]. Dabei werden die Testaufgaben gemäß dem so genannten Raschmodell ausgewählt und geprüft. Dazu muss eine Vielzahl von Aufgaben an einer großen Stichprobe von Schülern unter realistischen Einsatzbedingungen getestet werden. Diese Form der Testentwicklung hat 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]

Die Tests erfassten in der Regel jährlich eine Auswahl an relevanten Kompetenzbereichen. So kann z. B. der Kompetenzbereich Sprechen / Mündlicher Sprachgebrauch in den sprachlichen Fächern aus verständlichen Gründen nicht im Klassenverband erfasst werden, und wird daher trotz unbestreitbarer Relevanz bisher nicht erfasst. Im Fach Englisch können die Schulen im Schuljahr 2010 / 2011 erstmals freiwillig an einem solchen Test teilnehmen, bei dem die Testung außerhalb des Klassenverbandes erfolgt. Andere Kompetenzbereiche werden in einem rotierenden Verfahren durch die Tests berücksichtigt. So wurden in diesem Jahr in Klassenstufe 3 im Fach Deutsch die Bereiche Lesen, Rechtschreibung und Sprachbetrachtung getestet, im Fach Mathematik nur die Leitideen Daten, Häufigkeit und Zufall sowie Zahlen und Operationen. Begründet wird dieses Vorgehen damit, dass nur durch die Konzentration auf wenige Teilbereiche die Messgenauigkeit in diesen Bereichen hinreichend hoch ist.

[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.

Anders als in Klassenstufe 3 streuen die Aufgaben in Klassenstufe 6 im Fach Mathematik über alle Leitideen, allgemeinen mathematischen Kompetenzen und Anforderungsbereiche. Da die Messgenauigkeit auf Schülerebene nicht der zentrale Fokus der Tests ist, wird der Nutzen für die Lehrkräfte durch die breitere Abdeckung des Stoffes höher eingeschätzt als der Schaden durch geringere Messgenauigkeit in den verschiedenen Teilbereichen des Faches.

Im Fach Deutsch wurden in Klassenstufe 6 die Bereiche Schreiben, Sprache / Sprachgebrauch und Lesen getestet, im Fach Englisch die Bereiche Lesen, Hören und Schreiben. Anders als die Fächer Deutsch und Mathematik verfügt das Fach Englisch mit dem Gemeinsamen Europäischen Referenzrahmen für Sprachen (GER) über ein Kompetenzstufenmodell, mit welchem die Fähigkeiten der Schüler und die Schwierigkeit von Aufgaben auf einer gemeinsamen Fähigkeits- / Schwierigkeitsskala verortet werden können. Durch diese Verortung kann präzise beschrieben werden, auf welcher Niveaustufe des Spracherwerbs sich Schüler zu einem bestimmten Zeitpunkt befinden. Für die Entwicklung des Englischtests Klassenstufe 6 wurden nur Aufgaben der Niveaustufe A1 und A2 des GER verwendet, da diese das Lernziel für die Sekundarstufe 1 bilden. Leistungsstärkere Schüler und Klassen, wie sie z. T. an Gymnasien vertreten sind, erreichen im Test nahezu maximale Punktzahl. Für diese Schülergruppe hat der Test weniger Differenzierungsvermögen. Auch unterscheiden sich die Ergebnisse aufgrund dieser Testkonstruktion systematisch von denen in den Fächern Deutsch und Mathematik, bei denen die Differenzierung auch der leistungsstärkeren Schüler vorgesehen ist. Vergleiche der Ergebnisse zwischen den Fächern sind daher nicht möglich.

In Klassenstufe 8 wurden im Fach Deutsch die Bereiche Leseverstehen und Zuhören getestet, im Fach Mathematik kamen Aufgaben aus allen Leitideen, allgemeinen mathematischen Kompetenzen und Anforderungsbereichen vor, während der Englischtest die Kompetenzbereiche Leseverstehen und das Schreiben untersuchte.

## 2.1 Gütekriterien

Die Testentwicklung aller Kompetenztests findet auf hohem methodischem Niveau statt. Alle Aufgaben durchlaufen einen mehrfachen Begutachtungsprozess und werden im Rahmen umfangreicher Pilotierungen auf ihre Tauglichkeit als Testaufgabe geprüft. Dadurch ist gewährleistet, dass die Tests in hohem Maße den wissenschaftlichen Gütekriterien der Objektivität, Reliabilität und Validität entsprechen. Konkrete Angaben zu den Gütekriterien finden sich in den Dokumentationen der Aufgabenentwicklergruppen. Eine Konsequenz dieses hohen methodischen Standards ist jedoch, dass die Aufgabenformate und Korrekturanweisungen für die Lehrkräfte oft ungewohnt und nicht unbedingt praxisnah erscheinen. Entsprechend wichtig ist es, den Unterschied von Testsaufgaben und Lernaufgaben an den Schulen verständlich zu machen, um unnötige Vorbehalte gegenüber den Tests zu vermeiden.

# 3 Durchführung und Datenerfassung

Die Kompetenztests 2010 wurden in der Klassenstufe 3 am 28. April (Deutsch Lesetest), am 4. Mai (Deutsch Rechtschreibtest) und am 6. Mai (Mathematik) durchgeführt. In der Klassenstufe 6 erfolgte die Testdurchführung am 3. März (Deutsch), 23. Februar (Mathematik) und 25. Februar (Englisch). In Klassenstufe 8 wurde der Kompetenztest Deutsch am 24. Februar, der Kompetenztest Mathematik am 4. März und der Kompetenztest Englisch am 2. März geschrieben. Der Kompetenztest Französisch wurde nur in ausgewählten Schulen ebenfalls am 2. März durchgeführt.

## 3.1 Anmeldung zu den Kompetenztests 2010 und zum Schülerlängsschnitt

Wie in den vergangenen zwei Jahren wurde auch dieses Jahr fast jeder Schüler, der an den Thüringer Kompetenztests teilnahm, ebenfalls am Thüringer Schülerlängsschnitt angemeldet. Die Anmeldung erfolgte mittels des speziell dafür entwickelten Erhebungsprogramms.

Mit diesem Programm wurden die Stammdaten der Schüler für den Thüringer Schülerlängsschnitt erfasst und an das Thüringer Landesrechenzentrum gesendet. Bei der Anmeldung der Schüler zu den Kompetenztests wurde für die Dateneingabe und Ergebnisrückmeldung für jeden Schüler ein Schülerkode erstellt. Auf diese Weise wurde die Anonymität der Schüler gewährleitstet und der Thüringer Schülerlängsschnitt realisiert.

Dieser Schülerlängsschnitt ist dadurch gekennzeichnet, dass weder die Universität Jena (Datenauswertung) noch das Thüringer Landesrechenzentrum (Treuhänder der Stammdaten aller Schüler) Leistungsdaten zusammenführen können, ohne dass dieses Vorgehen durch das Thüringer Kultusministerium autorisiert und eine Verknüpfung durch das Thüringer Landesrechenzentrum hergestellt 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.

Das Erhebungsprogramm wurde allen Thüringer Schulen online zur Verfügung gestellt und kann zu jedem Zeitpunkt unter: http://www.kompetenztest.net/ heruntergeladen werden. Ist das Erhebungsprogramm einmal installiert, ist keine erneute Installation notwendig. Die Aktualisierungen für die Schulen werden automatisch beim Starten des Erhebungsprogramms durch Updates aus dem Internet durchgeführt.

Durch die ständige Weiterentwicklung und Verbesserung der Handhabbarkeit des Erhebungsprogramms, konnte die Anzahl der Schulen, die das Erhebungsprogramm nicht benutzen konnten, verringert werden. Die Änderungen dieses Jahres am Erhebungsprogramm umfassten die Ermöglichung der Zwischenspeicherung der Daten, so dass die Dateneingabe unterbrochen und zu einem späteren Zeitpunkt fortgesetzt werden kann. Weiterhin wurden die Durchführungsschritte vier und fünf vertauscht. Infolgedessen müssen jetzt erst die Schülerkodelisten ausgedruckt werden und erst danach können die Daten zum Thüringer Landesrechenzentrum

hochgeladen werden. Dadurch soll vermieden werden, dass Schülerkodelisten nicht ausgedruckt werden. Auch zukünftig gilt es das Erhebungsprogramm leicht bedienbar und benutzerfreundlich zu gestalten.

## 3.2 Teilnahme an den Kompetenztests 2010

Es nahmen insgesamt 44 984 Schüler aus 834 Schulen an den Kompetenztests teil. In Tabelle 3.1 sind die genauen Teilnehmerzahlen, gegliedert nach Klassenstufe und Schulart, aufgeführt.

**Tabelle 3.1:** Teilnehmerzahlen an den Kompetenztests 2010

| Klassenstufe | Schulart | Anzahl Schulen | Anzahl Klassen | Anzahl Schüler |
|---|---|---|---|---|
| 3 | Förderschule | 23 | 28 | 221 |
| | Grundschule | 460 | 851 | 15 876 |
| 6 | Förderschule | 30 | 32 | 261 |
| | Gymnasium | 93 | 310 | 6 837 |
| | Regelschule | 246 | 484 | 8 529 |
| 8 | Förderschule | 31 | 39[1] | 265 |
| | Gymnasium | 93 | 257[1] | 5 582 |
| | Regelschule | 244 | 781[1] | 7 413 |

Verglichen mit 2009 sind die Teilnehmerzahlen in allen Klassenstufen geringfügig angestiegen. Eine Ausnahme bilden dabei die Förderschulen. Da die Teilnahme an den Kompetenztests verpflichtend ist, ist dies einerseits ein Hinweis darauf, dass die Schülerzahlen weiterhin leicht ansteigen, andererseits weist es darauf hin, dass weniger Schüler an Förderschulen angemeldet werden.

Wie im letzten Jahr konnten die Kompetenztests in Klassenstufe 6 wieder in den Fächern Deutsch, Mathematik und Englisch und in der Klassenstufe 8 zusätzlich dazu im Fach Französisch geschrieben werden. Dabei war die Teilnahme an mindestens einem Test in Klassenstufe 6 verpflichtend. Eine Grenze nach oben wurde nicht gesetzt, so dass die Schulen auch an zwei oder mehr Tests teilnehmen konnten. Um die Vergleichbarkeit zu sichern, galt dabei allerdings, dass alle Klassen einer Schule in den jeweils gleichen Fächern teilnehmen mussten. Tabelle 3.2 gibt einen detaillierteren Einblick der gewählten Teilnahme in Klassenstufe 6.

Insgesamt zeigt sich, dass mehr als die Hälfte der Schulen sich in Klassenstufe 6 für die Teilnahme an allen Tests entschieden hat. Dies weist darauf hin, dass die Kompetenztests an den Schulen als nützliches Instrument angesehen werden. Wurden nicht alle Tests gewählt, so wurde in den meisten Fällen entweder eine Kombinationen aus Mathematik und einem anderen Fach oder die ausschließliche Teilnahme an den Tests in Mathematik gewählt. Somit scheinen die Tests insbesondere hier als nützlich gesehen zu werden. Andererseits sind die Kompetenztests in Mathematik auch mit dem geringsten Korrekturaufwand verbunden, was einen anderen Grund für die Wahl der Tests in Mathematik darstellen kann.

[1] In Klassenstufe 8 ist hier die Zahl der Klassen und Kurse angegeben, diese ist nicht direkt vergleichbar mit den Klassenzahlen der anderen Klassenstufen

**Tabelle 3.2:** Teilnehmerzahlen an den verschiedenen Fächerkombinationen in Klassenstufe 6

| Testteilnahme an Fächern | Anzahl Schulen | Anzahl Klassen | Anzahl Schüler |
|---|---|---|---|
| Nur Deutsch | 31 | 65 | 1 181 |
| Nur Mathematik | 43 | 115 | 2 187 |
| Nur Englisch | 22 | 55 | 1 024 |
| Deutsch und Mathematik | 37 | 81 | 1 623 |
| Deutsch und Englisch | 10 | 25 | 522 |
| Mathematik und Englisch | 20 | 47 | 890 |
| Deutsch, Mathematik und Englisch | 206 | 438 | 8 200 |

**Das Wichtigste in Kürze**

- Die Mehrheit der Schulen nahm in Klassenstufe 6 freiwillig allen Tests teil. Dies zeigt, dass die Tests als nützlich angesehen werden.

## 3.3 Webbasierte Datenerfassung

Nach der Korrektur der Testhefte gemäß Auswertungsmanualen wurden die Ergebnisse im Schulportal auf *kompetenztest.de* für jeden Schüler itemweise durch die Fachlehrer eingegeben.

Seit einigen Jahren wird dazu ein hinsichtlich möglichst einfacher Bedienung sowie Verkürzung der Eingabezeit optimiertes AJAX-System[2] eingesetzt. Neben diesem dynamischen System werden alle Eingaben zusätzlich als Standard-Formular angeboten, damit die Daten auch auf Computern, die das neue Eingabesystem nicht unterstützen, eingegeben werden können. Mit der neuen Eingabeversion wurden 95 % der über 125 000 Datenzeilen eingegeben.

Aufgrund der guten Resonanz des Eingabesystems bei den Fachlehrern wurde dies fast unverändert auch in diesem Jahr eingesetzt. Es wurde lediglich eine Modifikation realisiert, die verhindert, dass die Eingabe aller Schülerkodes getrennt im Vorfeld der Eingabe der Daten erfolgen kann. Da die Dateneingabe ohne diese Modifikation später auch ohne Kodeliste z. B. von einer anderen Person durchgeführt werden könnte, wäre die korrekte Zuordnung der Schüler zu den Schülerkodes dann nicht mehr sicher gestellt.

Wie in den letzten Jahren auch erfolgte in beiden Eingabeversionen die Eingabe der Test- und soziodemographischen Daten der Schüler zusammen mit den Schülerkodes. Alle Lehrer mussten dazu die bei der Anmeldung der Klassen erzeugten Schülerkodelisten erhalten, die die Zuordnung der Kodes zu den realen Schülernamen enthalten. In jeder Klasse konnten mehrere nicht angemeldete Schüler durch Sonderkodes („999“, „888“, ...) eingegeben werden, wovon in diesem Jahr 830 Mal Gebrauch gemacht wurde. Das ist im Vergleich zum Vorjahr eine anteilmäßige Abnahme und spricht angesichts von fast 45 000 Schülern für die Güte des eingesetzten Kodierverfahrens.

[2] „AJAX“ bezeichnet ein Konzept zur asynchronen Datenübertragung, welches zur dynamischen Gestaltung von Webseiten eingesetzt werden kann. Weitere Informationen hierzu finden sich im Landesbericht Thüringer Kompetenztests 2007

Zeitgleich mit der Eingabe der Testdaten wurde eine im Vergleich zum letzten Jahr stark gekürzte Begleitbefragung der Fachlehrer durchgeführt. Die Items des Begleitfragebogens 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 Zeitenwerte nicht betrachtet. Tabelle 3.3 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, und die 25. und 75. Perzentile. Die Hälfte aller analysierten Zeileneingaben hatte eine Dauer im Bereich zwischen diesen beiden Werten. Ebenfalls in Tabelle 3.3 ersichtlich sind die Anzahlen der Items für jeden Test, also die Anzahl der Spalten, die für jeden Test online eingegeben werden mussten, inklusive der Kovariaten. In der letzten Zeile ist die Änderung der Anzahl der Items gegenüber dem Vorjahr für jeden Test angegeben. Da beide Teile des Tests in Deutsch, Klassenstufe 3 von einem Fachlehrer bearbeitet und eingegeben werden müssen, werden diese beiden Teile hier auch zusammen betrachtet.

**Tabelle 3.3:** Zusammenfassung der Eingabezeiten pro Zeile in Sekunden

| | **Klassenstufe 3** | | | **Klassenstufe 6** | | |
|---|---|---|---|---|---|---|
| | **DK3L** | **DK3R** | **MK3** | **DK6** | **MK6** | **EK6** |
| 25. Perzentil | 29 | 35 | 63 | 68 | 35 | 98 |
| **50. Perzentil** | **36** | **42** | **76** | **84** | **43** | **118** |
| 75. Perzentil | 44 | 52 | 94 | 106 | 54 | 145 |
| Anzahl Items | 23 | 37 | 59 | 47 | 31 | 94 |
| Zuwachs Items | +82 % | | +26 % | +7 % | +11 % | +15 % |

| | **Klassenstufe 8** | | | |
|---|---|---|---|---|
| | **DK8** | **MK8** | **EK8** | **FK8** |
| 25. Perzentil | 75 | 42 | 74 | 86 |
| **50. Perzentil** | **92** | **51** | **89** | **104** |
| 75. Perzentil | 113 | 63 | 109 | 130 |
| Anzahl Items | 71 | 40 | 72 | 76 |
| Zuwachs Items | +58 % | -15 % | +177 % | +52 % |

Nur der Test MK8 hat weniger Items als im letzten Jahr. Die Zuwächse bei der Itemzahl bei den anderen Tests sind zum Teil beträchtlich, nur die Tests DK6, MK6, EK6 und MK3 haben moderate Zuwächse, wobei Ek6 bereits im letzten Jahr mit 82 Items der mit Abstand längste Test war. In diesem Jahr gab es nur drei Tests, bei denen weniger als 50 Items zu bearbeiten waren (im Vorjahr waren dies sieben Tests).
Die Zunahme der Testlängen wirkt sich natürlich auf die Eingabezeiten aus. Allerdings sind diese Zuwächse nicht proportional zu den Zuwächsen der Zahl der Items. Anscheinend setzt sich die Eingabezeit zu einem sehr großen Teil aus der Zeit zusammen, die benötigt wird, die nächste Zeile auf dem Erhebungsbogen auszuwählen und ggf. zu markieren, sowie den

Schülerkode und die Kovariaten (meist per Maus) einzugeben und weniger aus der Zeit für die Eingabe der eigentlichen Werte der Testitems, die überwiegend nur aus „0“ oder „1“ bestehen. Die Eingabe der Testitems zu beschleunigen war eines der Hauptziele bei der Einführung des neuen AJAX-Eingabesystems.

Die Mathematiktests in Klassenstufen 6 und 8 wurden dieses Jahr am schnellsten eingegeben. Im letzten Jahr hatten sieben Tests eine mittlere Eingabezeit von etwas über einer Minute oder weniger pro Zeile, dieses Jahr erreichten nur noch zwei Tests dieses Kriterium. Erstmals gab es in diesem Jahr auch einen Test, dessen Eingabe im Mittel fast zwei Minuten pro Zeile dauerte: EK6, der mit 94 Items auch der bisher längste Test ist. Die Eingabe einer Klasse von 20 Schülern hat dieses Jahr bei 4 Tests durchschnittlich fast 30 Minuten oder mehr betragen (im Vorjahr war dies nur ein Test). Somit ergibt sich aufgrund der gestiegenen Zahl von Items allein bei der Onlineeingabe eine deutliche Mehrbelastung der beteiligten Fachlehrer, die bei Korrektur der Testhefte noch deutlicher zum Tragen kommen dürfte.

Die Analyse der Zeitpunkte der Dateneingaben zeigt wie schon in den letzten Jahren eine deutliche Häufung der Eingaben von Montag bis Donnerstag in der Zeit von 10:00 Uhr bis 14:00 Uhr. Die Zahl der Eingaben, die zu eher untypischen Arbeitszeiten stattfanden, war ebenfalls vergleichbar zum Vorjahr: Fast 27 % aller Eingaben erfolgten später als 19:00 Uhr oder am Wochenende. Unter diesen Eingaben waren die Tests in Klassenstufe 3 sowie die Tests in Mathematik Klassenstufe 6 und 8 am seltensten und Tests in Deutsch, Klassenstufe 6 und 8 am häufigsten vertreten. Die Eingabe für Französisch, Klassenstufe 8 fand mit Abstand am häufigsten am Wochenende statt, die Wochenendeingaben allgemein fanden deutlich häufiger sonntags statt.

Die Verteilung der Eingaben über den gesamten Eingabezeitraum zeigt einen relativ konstanten Anstieg mit einer moderaten Häufung innerhalb der jeweils ersten drei Wochen. Von den Lehrern wird jedoch wie im Vorjahr der gesamte Zeitraum der Eingabe (ca. 4-5 Wochen) genutzt. Die Eingaben der Klassenstufe 3 haben dabei den steilsten Anstieg, d. h. diese werden am schnellsten begonnen und sind zuerst fertig, haben aber auch einen kürzeren Eingabezeitraum.

**Tabelle 3.4:** Zusammenfassung der Bewertungen in Schulnoten 1–6 („Welche ’Note’ würden Sie der Dateneingabe im Internet geben?“) und die Anzahl der jeweils zu tätigenden Eingaben (Spalten)

| | **Klassenstufe 3** | | | **Klassenstufe 6** | | |
|---|---|---|---|---|---|---|
| | **DK3L** | **DK3R** | **MK3** | **DK6** | **MK6** | **EK6** |
| Mittelwert | 1,7 | 1,8 | 1,8 | 2,1 | 1,6 | 2,0 |
| (Streuung) | (0,7) | (0,8) | (0,8) | (0,8) | (0,7) | (0,9) |
| Anzahl Items | 23 | 37 | 59 | 47 | 31 | 94 |

| | **Klassenstufe 8** | | | |
|---|---|---|---|---|
| | **DK8** | **MK8** | **EK8** | **FK8** |
| Mittelwert | 2,0 | 1,8 | 2,0 | 1,8 |
| (Streuung) | (0,9) | (0,7) | (0,8) | (0,7) |
| Anzahl Items | 71 | 40 | 72 | 76 |

Die Lehrer hatten direkt im Anschluss an die Dateneingabe die Möglichkeit, die Eingabe im Internet zu bewerten. Tabelle 3.4 zeigt die Ergebnisse dieser Bewertung getrennt für die verschiedenen Tests. Die Bewertungen sind vergleichbar mit den Bewertungen des letzten Jahres. Es sind keine Trends über Klassenstufen oder Fächer erkennbar. Der Gesamtdurchschnitt von 1,9 über alle Tests spricht erneut für eine große Zustimmung zum verwendeten Online-System.

**Das Wichtigste in Kürze**

- Zeitaufwand für Onlineeingabe ist aufgrund zum Teil erheblich gestiegener Anzahl der einzugebenden Items stark angestiegen
- Bewertung der Onlineeingabe durch die Lehrer ist trotzdem vergleichbar gut wie in Vorjahren

## 3.4 Auswertung der Hotline

Zur Klärung von Problemen wurden auch in diesem Jahr wöchentliche Hotlinezeiten angeboten, um die Schulen bei der Durchführung und Auswertung der Kompetenztests zu unterstützen. Die telefonischen Sprechzeiten wurden in besonders problematischen Phasen spontan verlängert, um Probleme auch dann möglichst zeitnah klären zu können. Außerdem konnten sich die teilnehmenden Schulen weiterhin jederzeit per Fax oder E-Mail an das Projekt Kompetenztest wenden. Die angebotenen Möglichkeiten zur Kontaktaufnahme wurden auch in diesem Jahr von den Lehrern rege genutzt und dankbar angenommen, wobei Anfragen meist direkt und schnell geklärt werden konnten.

Die aufgetretenen Probleme in den teilnehmenden Schulen ähnelten den Anfragen des vergangenen Jahres. Aufgrund der Erfahrungen der letzten Jahre konnten unter anderem Passwortanfragen zügig bearbeitet werden und den Lehrern bei Fragen bezüglich der Ergebniseingabe im Internet oder Freischaltungen schnell geholfen werden. Es traten vereinzelt Unstimmigkeiten zwischen den Schülerkodelisten der Schulen und den im Projekt Kompetenztest vermerkten Listen auf, deren unterschiedliche Ursachen erfolgreich mit Hilfe der Lehrer identifiziert wurden. Vermehrt baten die teilnehmenden Schulen um Hinweise, wie man mit kranken oder neu hinzugekommenen Schülern, die noch nicht in den Schülerkodelisten vermerkt waren, verfahren solle. In diesen Fällen verwies die Hotline auf die Zusatzkodes für neue Schüler und verdeutlichte, dass auch am Testtag fehlende Schüler mit ihrem Schülerkode bei der Ergebniseingabe vermerkt sein müssen, damit diese abgeschlossen werden konnte.

Um die Hotline noch effizienter zu gestalten, sind wir auf die gute Zusammenarbeit mit den Schulen angewiesen. Ein gut funktionierender Informationsaustausch innerhalb der Schulen trägt dazu bei, dass die Fachlehrer über bereits zur Verfügung stehende Informationen Bescheid wissen. So gibt zum Beispiel das bereits am Testtag auf dem Schulportal vorhandene Durchführungsmanual alle benötigten Hinweise zum Verfahren mit z. B. kranken Schülern bei der Ergebniseingabe im Internet. Außerdem ist es von Vorteil, Schülerkodelisten und Passwörter an zentraler Stelle zu verwalten, so dass alle Fachlehrer jederzeit wissen, wohin sie sich wenden müssen. Wichtig ist auch, dass Veränderungen der Schülerdaten erneut zur Universität Jena hochgeladen werden, um Unstimmigkeiten vorzubeugen. Damit es keine Verzögerung bei der Beantwortung von E-Mail-Anfragen gibt, ist es wichtig, dass der Hotline immer die Schulnummer und eventuell die betreffende Klasse oder der Lehrer mitgeteilt wird.

Insgesamt hat sich die Hotline in diesem Jahr wieder als ein zentraler Bestandteil bei der Durchführung der Kompetenztests erwiesen. Sie fungiert als Schnittstelle zwischen der Durchführung an den Schulen und den organisatorischen bzw. technischen Abläufen und ist nicht nur eine Anlaufstelle bei Fragen für die teilnehmenden Lehrer. Durch den direkten Kontakt mit den Schulen konnten auch Hinweise und Verbesserungsmöglichkeiten aufgenommen und umgesetzt werden.

# 4 Diagnostische Kompetenz

## 4.1 Einleitung

Wie in den vergangenen Jahren gab es auch 2010 die Möglichkeit für die in die Thüringer Kompetenztests involvierten Fachlehrer, an der diagnostischen Kompetenz teilzunehmen. Hierzu können sie die zu erwartenden Schülerleistungen für ihre Klasse(n) bereits einen Tag vor den Kompetenztests einschätzen. Die Fachlehrer erhalten durch diese Erhebung eine wertvolle Rückmeldung darüber, wie genau die eigene Leistungseinschätzung der Klasse ist. Im folgenden Kapitel sollen die Ergebnisse dieser freiwilligen Einschätzung beschrieben werden.

Die diagnostische Kompetenz wurde auch dieses Jahr wieder für alle Kompetenztests erhoben. Das heißt, dass die Leistungseinschätzung in der Klassenstufe 3 in den Fächern Mathematik und Deutsch möglich war, während sie in den Klassenstufen 6 und 8 zusätzlich zu den Fächern Deutsch und Mathematik noch in Englisch angeboten wurde. In Klassenstufe 8 gab es die diagnostischen Kompetenz weiterhin im Fach Französisch, allerdings werden diese Daten hier aus Datenschutzgründen nicht veröffentlich, weil die geringe Teilnehmerzahl keine Anonymität gewährleistet.

Die teilnehmenden Lehrer hatten jeweils über die Schulportalseite die Möglichkeit, für durchschnittlich sechs ausgewählte Testaufgaben aus den jeweiligen Kompetenztests eine Einschätzung darüber abzugeben, wie viele der Schüler diese Aufgaben des Kompetenztests richtig und vollständig lösen werden. Nach dem Abschluss der Dateneingabe erfolgte dann ein Vergleich zwischen der Einschätzung der Fachlehrer und den tatsächlichen Ergebnissen der Klasse. Dieses Ergebnis erlaubt dem Fachlehrer, die Genauigkeit seines Urteils über die Leistungen der Schüler abzuschätzen. Der Lehrer erhält somit auch Rückmeldungen und Anregungen, sich intensiver mit dem Leistungsniveau seiner Klasse zu beschäftigen. Eine realistische Einschätzung der Schüler einer Klasse bezüglich ihrer fachspezifischen Kompetenzen ist eine notwendige Bedingung für die Förderung von leistungsstarken und leistungsschwachen Schülern im Unterricht. Bei vorhandenen Fehleinschätzungen kann es dazu kommen, dass Schüler unter- oder überfordert werden und Misserfolgserlebnisse könnten sich möglicherweise negativ auf die Schüler- oder Lehrermotivation auswirken. Es ist jedoch anzumerken, dass perfekte Leistungsaussagen über die eigenen Schüler kaum möglich sind, insbesondere, wenn es sich um eine zahlenmäßig sehr große Klassen handelt. Außerdem können die Leistungen in den Kompetenztests auch von unerwarteten Ereignissen beeinflusst werden, wie beispielsweise dem Fehlen einzelner Schüler oder der wechselnden Tagesform der Schüler am Testtag. Ziel der diagnostischen Kompetenz soll es demnach sein, den Lehrern eine Möglichkeit zu bieten, die eigenen Einschätzungen der Klasse zu überprüfen und die Genauigkeit bei der Kompetenzeinschätzung der Schüler auszubauen.

*Teilnehmerzahlen zur diagnostischen Kompetenz – global*:

Um die Teilnahme an der diagnostischen Kompetenz zu ermitteln, wurde in diesem Landesbericht die erste Frage der diagnostischen Kompetenz zugrunde gelegt. Hier wird die Lehrkraft gebeten, die erwartete Anzahl der an den Tests teilnehmenden Schüler zu nennen. Die Anzahl der Eintragungen bei dieser Frage wurde mit der Anzahl der Lehrkräfte des entsprechenden Faches in der jeweiligen Klassenstufe ins Verhältnis gesetzt. Betrachtet man die Anzahl der

freiwilligen Teilnehmer an der diagnostischen Kompetenz über alle Klassenstufen, Fächer und Schulformen hinweg, ist kaum Veränderung im Vergleich zum letzten Jahr festzustellen. Im Durchschnitt haben in diesem Jahr 35,2 % aller Lehrer an Thüringer Schulen die diagnostische Kompetenz genutzt. Eine über Fächer und Klassenstufen differenzierte Darstellung findet sich im jeweiligen Kapitel.

Die erweiterte Form der Datenerfassung der diagnostischen Kompetenz ermöglichte auch nachzuvollziehen wie viele Lehrer sich die Testaufgaben zwar online angeschaut, jedoch nur teilweise oder gar nicht bearbeitet hatten. Bei diesen Abbrecher-Analysen zeigt sich, dass speziell in den Deutschtests die Einschätzung der Lösungshäufigkeit der Testaufgaben vorzeitig beendet wurde (21,5 %). In Englisch liegt die relative Zahl der Abbrecher bei durchschnittlich 19,3 %, wobei vor allem die geringe Abbrecher-Anzahl bei den Aufgaben der Teilkompetenz Rechtschreibung in Klassenstufe 8 (15,9 %) die höheren Werte aus Klassenstufe 3 und 6 nach unten korrigiert. In Mathematik ist die relative Zahl der Abbrecher über alle Klassenstufen hinweg mit 15,4 % am geringsten.

In den nachfolgenden Unterkapiteln werden die auffälligen Ergebnisse der diagnostischen Kompetenz sortiert nach den einzelnen Tests für die verschiedenen Klassenstufen und in den weiterführenden Schulen auch getrennt nach Schularten berichtet. Dabei wurden nur vollständig bearbeitete Testaufgaben zur diagnostischen Kompetenz ausgewertet. Die Ergebnisse werden mithilfe von drei Maßen dargestellt, die eine unterschiedliche Beurteilung der Lehrereinschätzung zulassen. Bei dem ersten Maß handelt es sich um die mittlere Abweichung der von den Lehrern vorab geschätzten und der tatsächlichen Lösungshäufigkeit. Diese Angabe gibt Auskunft darüber, ob ein Fachlehrer seine Klasse eher unterschätzt oder überschätzt hat. Um einen Vergleich unabhängig von der Teilnehmerzahl zu ermöglichen, wird diese Differenz in Prozent angegeben, wobei ein positiver Prozentwert eine mittlere Überschätzung der tatsächlichen Lösungshäufigkeiten und ein negativer Wert eine mittlere Unterschätzung der realen Schülerleistungen bedeutet. Das zweite Maß ist die Standardabweichung der Differenzwerte, welche Schlüsse über die Breite der Verteilung zulässt und ein Maß für die Streuung um den Wert der mittleren Abweichung ist. Die Standardabweichung gibt Hinweise auf die Genauigkeit der Lehrereinschätzung und zeigt, wie ähnlich die verschiedenen Lehrer innerhalb eines Tests geschätzt haben. Dabei bedeuten hohe Werte, dass es innerhalb der Gruppe der Lehrer für einen Test große Abweichungen in der Einschätzung der Schülerleistungen gibt. Als drittes Maß werden Extremwerte berichtet, die angeben, wie viel Prozent der Lehrer bei einem Test die Schüler im Mittel über 25 %-Punkte über- oder unterschätzt haben. Diese Grenze ist durch die Erfahrungen mit der diagnostischen Kompetenz in den letzten fünf Jahren festgesetzt wurden. Zusätzlich werden die Ergebnisse durch Abbildungen veranschaulicht, in denen sowohl die mittlere Abweichung als auch die Standardabweichung über alle Schulformen und Aufgaben hinweg dargestellt sind. Aufgrund durchgängig kleiner Teilnehmerzahlen in den Förderschulen findet in der Klassenstufe 3 keine Schulartunterscheidung statt und in den weiterführenden Schulen wird nur zwischen Regelschulen und Gymnasien unterschieden. Andernfalls könnte keine Anonymität mehr gewährleistet werden. Für die betreffenden Fachlehrer der Förderschulklassen kann ein Vergleich ihrer Ergebnisse zur diagnostischen Kompetenz mit den Gesamtergebnissen zu jedem Test aber dennoch hilfreich sein und nützliche Hinweise zur Unterrichtsarbeit liefern.

Anschließend werden die Einschätzungsleistungen der Fachlehrer noch mit weiteren Merkmalen in Zusammenhang gebracht. Dafür werden speziell die Beziehungen zwischen diagnostischer Kompetenz und der tatsächlich erreichten Punktzahl der Klasse und den Ergebnissen der Verwendbarkeitsanalysen (Lernstandsdiagnostik, Lehrplanadäquatheit, Unterrichts- und Schulentwicklung) betrachtet. Zum Ende des Kapitels werden die zentralen Ergebnisse der diagnostischen Kompetenz noch in einem Fazit zusammengefasst.

## 4.2 Auswertung der diagnostischen Kompetenz für Klassenstufe 3

Im Fach Deutsch haben mit 31,1 % ca. 10% weniger Lehrer als im vergangenen Jahr an der diagnostischen Kompetenz teilgenommen. Im Gegensatz dazu haben im Fach Mathematik deutlich mehr Lehrer als in 2009 die diagnostische Kompetenz zur Einschätzung ihrer Bewertungsfähigkeiten genutzt. Erfreuliche 45,9 % machten in diesem Jahr eine Angabe zu der erwarteten Anzahl der Schüler bei den Kompetenztests. 2009 waren dies nur 27,4 %. Allerdings füllten 2010 nur 22,5 % alle Aufgaben aus, während 2009 fast alle Teilnehmer bis zu letzten Aufgabe Angaben machten. Trotzdem stellt eine Beteiligung von einem Drittel bzw. der Hälfte der Fachlehrer eine erfreulich hohe Rate dar, denn die Teilnahme an der diagnostischen Kompetenz war freiwillig. Offenbar wird die diagnostische Kompetenz von den Lehrern gut angenommen.

**Fach Deutsch**
Im Fach Deutsch unterscheiden sich die von den Lehrern vorhergesagte und die tatsächliche Anzahl von Schülern, die die ausgewählten Aufgaben vollständig gelöst haben, im Mittel um -11,1 %. Dieser durchschnittliche Abweichungswert ist relativ groß und deutlich ausgeprägter als die mittlere Abweichung von -1,3 % für die Kompetenztests in Deutsch in der Klassenstufe 3 im vergangenen Jahr. Inhaltlich bedeutet der Wert, dass die Schülerleistungen insgesamt von den Lehrern unterschätzt wurden. Anhand der Standardabweichung von 12,4 % für die mittlere Abweichung kann man erkennen, dass die Lehrereinschätzungen voneinander abweichen. Bei 19,2 % der Deutschlehrer, die an der diagnostischen Kompetenz für die Klassenstufe 3 teilgenommen haben, liegen die Unterschiede zwischen der geschätzten und der tatsächlichen Lösungshäufigkeit bei 25 %-Punkten im Sinne einer Unterschätzung der Schülerleistung. Im Vergleich zu 13,8 % im vergangenen Jahr ist dies ein Anstieg.

*Aufgabenebene*: Eine genauere Analyse der diagnostischen Kompetenz ist durch die Betrachtung der mittleren Abweichung für einzelne Aufgaben möglich. Dabei treten in diesem Jahr einige erhebliche Abweichungen auf. Im Testteil Lesen kommt es bei der Aufgabe 2 namens „Der aufmerksame Beobachter“ in der Teilaufgabe 2.3 mit einer mittleren Abweichung von -15,9 % zu einer Unterschätzung. In den Teilaufgaben 2.7 und 2.10 beträgt die mittlere Abweichung dagegen 21,1 % und 20,1 %. Die Leistung der Schüler wird also überschätzt. Die Schüler sollten bei der Aufgabe zunächst einen Text lesen und danach einige Fragen dazu beantworten. Die Aufgabe lässt sich dem Kompetenzbereich 3.1. „über Lesefähigkeiten verfügen“ zuordnen. Im Testteil Rechtschreibung zeigten sich bei einigen Teilaufgaben der Aufgabe 1 „Lückensätze“ auffällige mittlere Abweichungen. Bei diesen Aufgaben wurden den Schülern Sätze mit Lücken vorgelegt. Der vollständige Satz wurde dann vorgelesen und die Schüler sollten die fehlenden Wörter in ihren Heften ergänzen. In den Teilaufgaben 1, 2 und 3 kam es mit mittleren Abweichungen von -15,8 %, -24,7 % und -22,5 % zu deutlichen Unterschätzungen der Leistungen der Schüler. Die Aufgaben gehören zum Kompetenzbereich 5 „Rechtschreibung“. Ebenfalls zu Unterschätzungen der Schülerleistungen kam es bei den vier Lupenstellen in Aufgabe 2.2. Die Schüler hatten hier die Aufgabe, die Wörter durchzustreichen, die fehlerhaft waren, und darüber die korrekte Schreibweise einzutragen. Die mittleren Abweichungen schwanken dabei von -15,2 % bei der ersten Lupenstelle bis zur großen Abweichung von -34,5 % bei der dritten Lupenstelle. Die Schüler haben die Aufgabe offensichtlich sehr viel besser gemeistert, als ihre Lehrer dies erwartet haben.

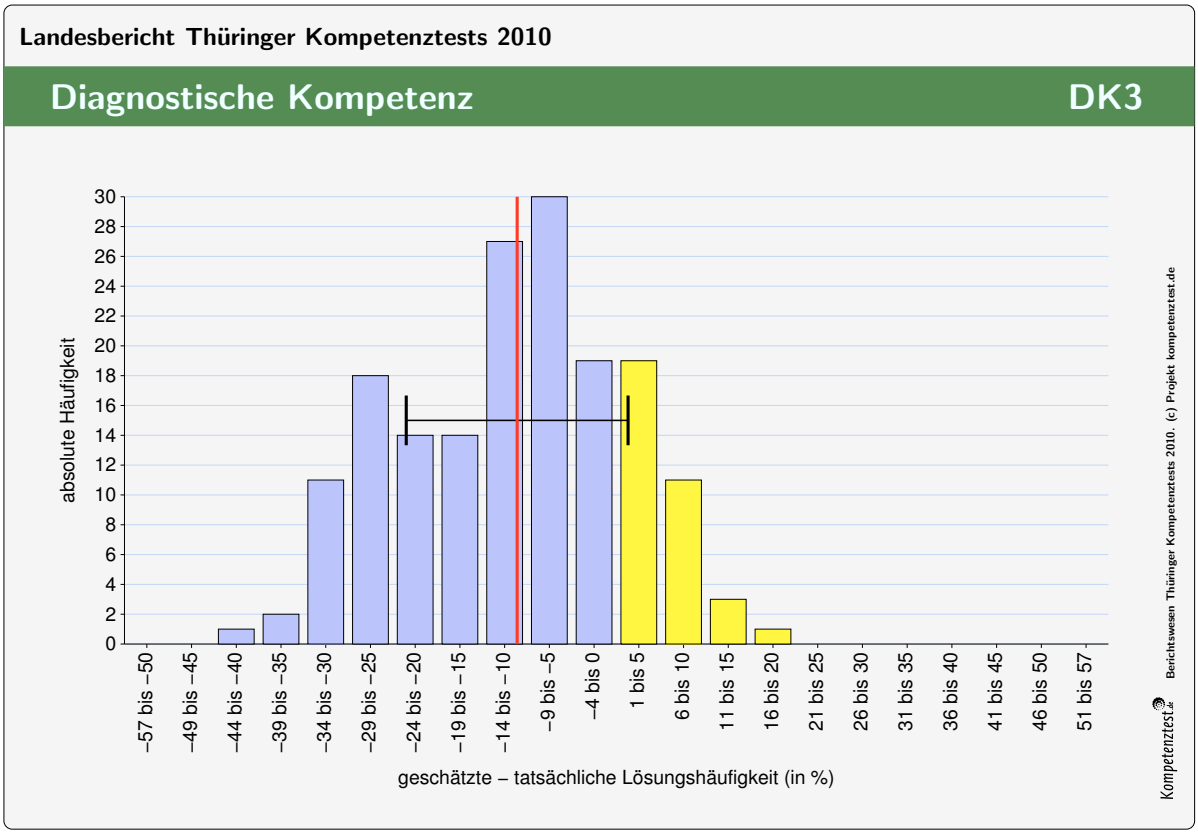


**Abbildung 4.1:** DK3: Abweichung der Lehrerschätzung von der tatsächlichen Lösungshäufigkeit mit mittlerer Abweichung (rot, senkrechter Strich) und Standardabweichung (schwarz, waagerechter Strich)

**Fach Mathematik**

Ähnlich wie im Fach Deutsch zeigt sich in Mathematik ebenfalls eine deutliche negative mittlere Abweichung. In diesem Jahr liegt hier die mittlere Abweichung zwischen Lehrereinschätzung und Schülerleistung bei -13,8 %. Die Standardabweichung beträgt 13,3 %. Diese Werte entsprechen in etwa denen aus dem vergangenen Jahr. Mit 19,2 % ist der Anteil der Lehrer, bei denen die Unterschiede zwischen geschätzter und tatsächlicher Lösungshäufigkeit größer als 25 %-Punkte ist, allerdings größer geworden. Dieser bezieht sich in diesem Jahr sowohl auf negative als auch auf positive Abweichungen.

*Aufgabenebene*: Deutliche Fehleinschätzungen sind bei der diesjährigen diagnostischen Kompetenz bei drei Aufgaben zu finden. Bei allen handelt es sich um negative mittlere Abweichungen zwischen den Einschätzungen der Lehrer und den tatsächlichen Lösungshäufigkeiten, d.h. es handelt sich um Unterschätzungen. Bei Aufgabe 2 a) unterschätzten die Lehrer die Leistung ihrer Schüler mit einer mittleren Abweichung von -16,7 %. Hier hatten die Schüler die Aufgabe nach Betrachten eines Histogramms mit der Anzahl der Geburtstage in jedem Monat festzustellen, in welchem Monat die meisten Kinder Geburtstag haben. Die Aufgabe ist der Leitidee 5 „Daten, Häufigkeit und Wahrscheinlichkeit“ zuzuordnen. Ebenfalls zu dieser Leitidee gehört die Aufgabe 5, bei der die mittlere Abweichung -25,0 % beträgt und damit eine deutliche Unterschätzung der Schülerleistung aufzeigt. Hier sollten die Schüler einschätzen, ob bestimmte Aussagen sicher, möglich oder unmöglich sind. Auch bei der Aufgabe 3 b) des zweiten Teststeils erwarteten die Lehrer schlechtere Leistungen. Hier beträgt die mittlere Abweichung -22,1 %. Die Aufgabe gehört zu der Leitidee 1 „Zahlen und Operationen“. Offensichtlich sind die Schüler in einigen Bereichen mit dem Lernstoff vertrauter als ihre Mathematik-Lehrer dies wahrnehmen.

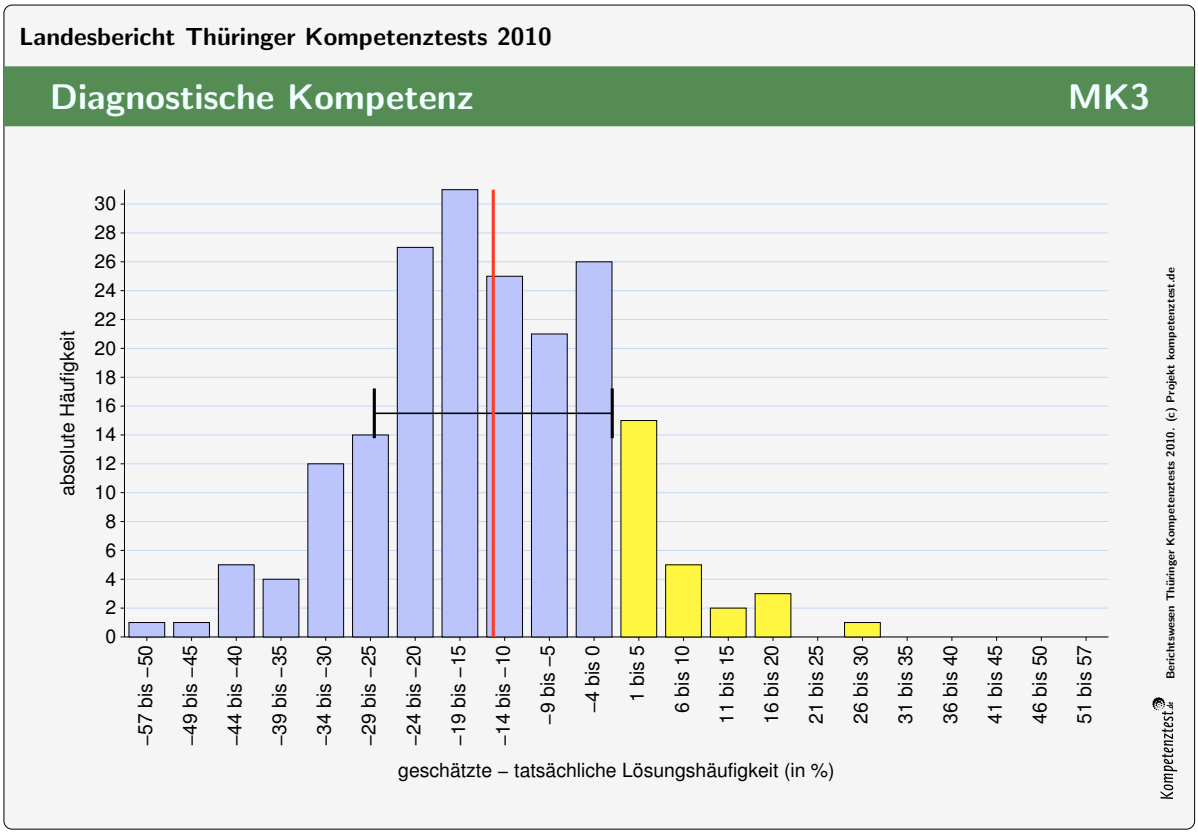


**Abbildung 4.2:** MK3: Abweichung der Lehrerschätzung von der tatsächlichen Lösungshäufigkeit mit mittlerer Abweichung (rot, senkrechter Strich) und Standardabweichung (schwarz, waagerechter Strich)

In der Abbildung 4.2 ist die Verteilung der Abweichungen der geschätzten von den tatsächlichen relativen Lösungshäufigkeit in Mathematik wie oben beschrieben dargestellt. Die im Text erwähnten negativen mittleren Differenzen lassen sich hier gut nachvollziehen und die Breite der Verteilung gibt Aufschluss darüber, wie genau geschätzt wurde. Dabei wird deutlich, dass im Vergleich zum Deutschtest (Abbildung 4.1) die Streuung etwas größer ausfällt, die Grundschullehrer im Fach Mathematik also etwas weniger genau schätzten.

## 4.3 Auswertung der diagnostischen Kompetenz für Klassenstufe 6

In diesem Jahr nahmen in Deutsch 31,4 % der Lehrer an der diagnostischen Kompetenz teil. Im Vergleich zu 2009 mit 39,7 % sind dies etwa 8 % weniger. Ähnlich wie in Klassenstufe 3 ist der Trend in Mathematik in umgekehrter Richtung. Mit 43,8 % nutzten etwa 5 % mehr Lehrer die diagnostische Kompetenz. Die Englischlehrkräfte nutzen im Fächervergleich die diagnostische Kompetenz mit 23,1 % am wenigsten. Diese Quote entspricht in etwa der Teilnahme im vergangenen Jahr. In allen drei Fächern beträgt die Differenz zwischen der Prozentzahl der Teilnahme an der ersten Aufgabe und der Prozentzahl der Teilnahme an der letzten Aufgabe etwa 10 %. Möglicherweise haben einige Lehrer die Bereitstellung einiger Aufgaben genutzt, um einen Einblick in die am nächsten Tag stattfinden Tests zu bekommen.

**Fach Deutsch**

*Schularten*: Im Fach Deutsch unterscheiden sich die im Mittel von den Lehrern vorhergesagte und die tatsächliche Anzahl von Schülern, die die ausgewählten Aufgaben vollständig gelöst haben, um 22,8 %. Das bedeutet, dass die Lehrer ihre Schüler stark überschätzen. Im Vergleich zum Vorjahr zeichnet sich ein negativer Trend ab, denn im letzen Jahr überschätzten die Lehrer die Leistungen lediglich leicht mit 4,5 %. Die Standardabweichung der mittleren Abweichung beträgt 15,7 %, was bedeutet, dass die Lehrereinschätzungen relativ stark streuen. Außerdem stieg der Anteil der Lehrer, bei denen die mittlere Differenz zwischen geschätzter und tatsächlicher Lösungshäufigkeit größer als 25 %-Punkte ist, von 7,7 % im letzten Jahr auf 45,5 % in diesem Jahr. Inhaltlich ausgedrückt heißt das, dass fast jeder zweite Lehrer seine Schüler stark unter- oder überschätzt, was einen sehr alarmierenden Anteil darstellt. In den Regelschulen unterscheiden sich die von den Lehrern vorhergesagte und die tatsächliche Lösungshäufigkeit um 23,9 %, während dieser Wert mit 20,2 % am Gymnasium etwas kleiner ist. Ebenso haben die verschiedenen Lehrer an den Gymnasien ähnlicher geschätzt als an den Regelschulen, was sich in einer Standardabweichung von 13,5 % beziehungsweise 16,3 % an den Regelschulen widerspiegelt.

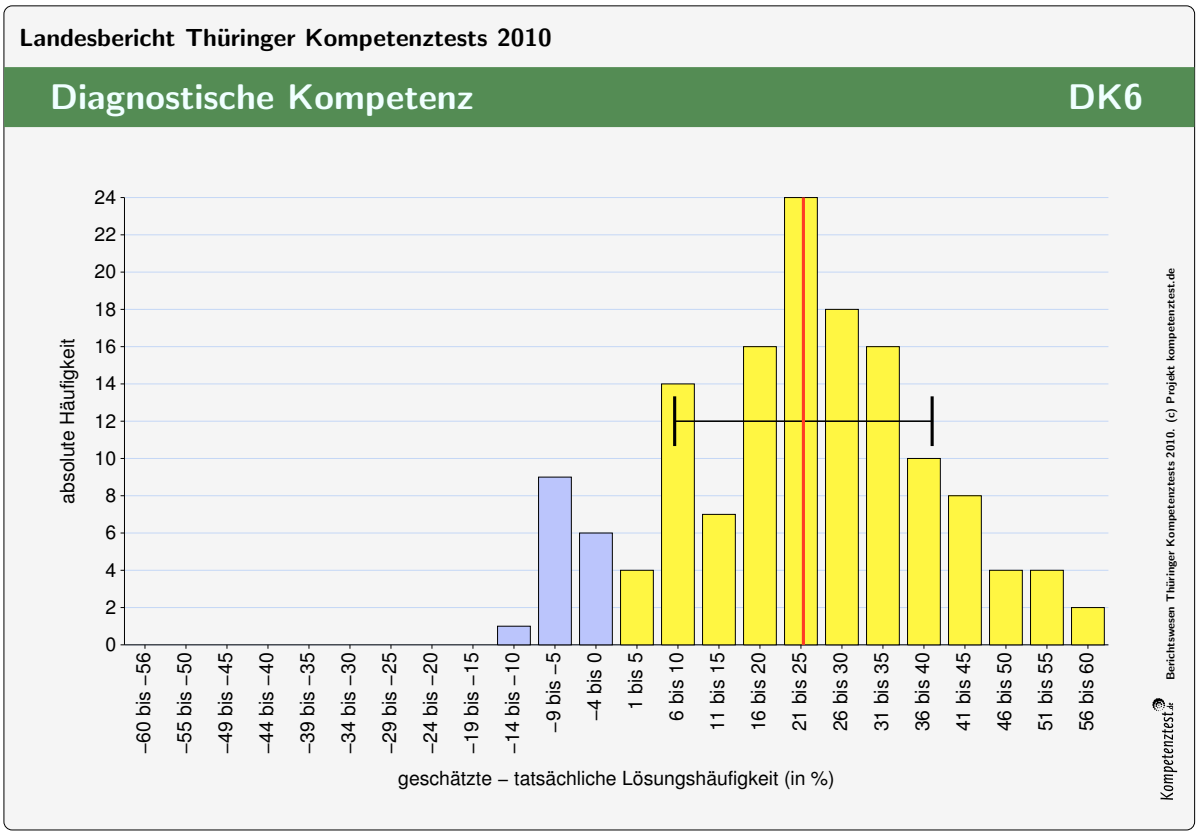


**Abbildung 4.3:** DK6: Abweichung der Lehrerschätzung von der tatsächlichen Lösungshäufigkeit mit mittlerer Abweichung (rot, senkrechter Strich) und Standardabweichung (schwarz, waagerechter Strich)

*Aufgabenebene*: Bei einer genaueren Betrachtung der mittleren Abweichung auf Aufgabenebene, sind zwei Aufgaben besonders auffällig. Im Kompetenzbereich Lesen kommt es bei der Aufgabe „Pünktlich wie das Ferkelkraut“ Unteraufgabe 3 zu einer mittleren Abweichung von 42,2 % und somit einer starken Überschätzung der Schülerleistungen durch den Lehrer. Von den Schülern wurde verlangt, zunächst einen Text zu lesen und anschließend mehrere Fragen zu diesem zu beantworten. Bei der Aufgabe 3 handelt es sich um den Anforderungsbereich 3, in dem die Schüler reflektieren und bewerten sollen. Außerdem überschätzten die Lehrer ihre

Schüler stark bei der Aufgabe „Präfixe“, in der die Schüler fehlende Vorsilben in kurzen Sätzen ergänzen sollten. Hier liegt eine mittlere Abweichung von 43,4 % vor. Diese Aufgabe gehört zum Kompetenzbereich Sprachgebrauch und stellt Anforderungsbereich 2 dar.

**Fach Mathematik**

*Schularten*: Im Fach Mathematik ergibt sich über alle Schulformen und Aufgaben hinweg eine mittlere Abweichung von -2,3 %. Die Lehrer schätzen die Leistungen der Schüler also im Durchschnitt relativ realistisch ein. Im Vergleich zum vorherigen Jahr (-1,5 %) ist dieser Wert ungefähr gleich geblieben, was positiv einzuschätzen ist. Die Standardabweichung beträgt 13,1 %, womit die Einschätzung der Mathematiklehrer moderat streut. Bei 7,5 % der Lehrer ist die mittlere Differenz zwischen geschätzter und tatsächlicher Lösungshäufigkeit größer als 25 %-Punkte. Vergleicht man diese Zahl mit dem Vorjahreswert wird deutlich, dass sich weniger Mathematiklehrer der Klassenstufe 6 extrem verschätzen. An den Regelschulen beträgt die mittlere Abweichung 0 %. Es liegt also im Durchschnitt eine exakte Einschätzung der Schülerleistungen vor, während die Gymnasiallehrer ihre Schüler mit einem Abweichungswert von 5 % leicht unterschätzen.

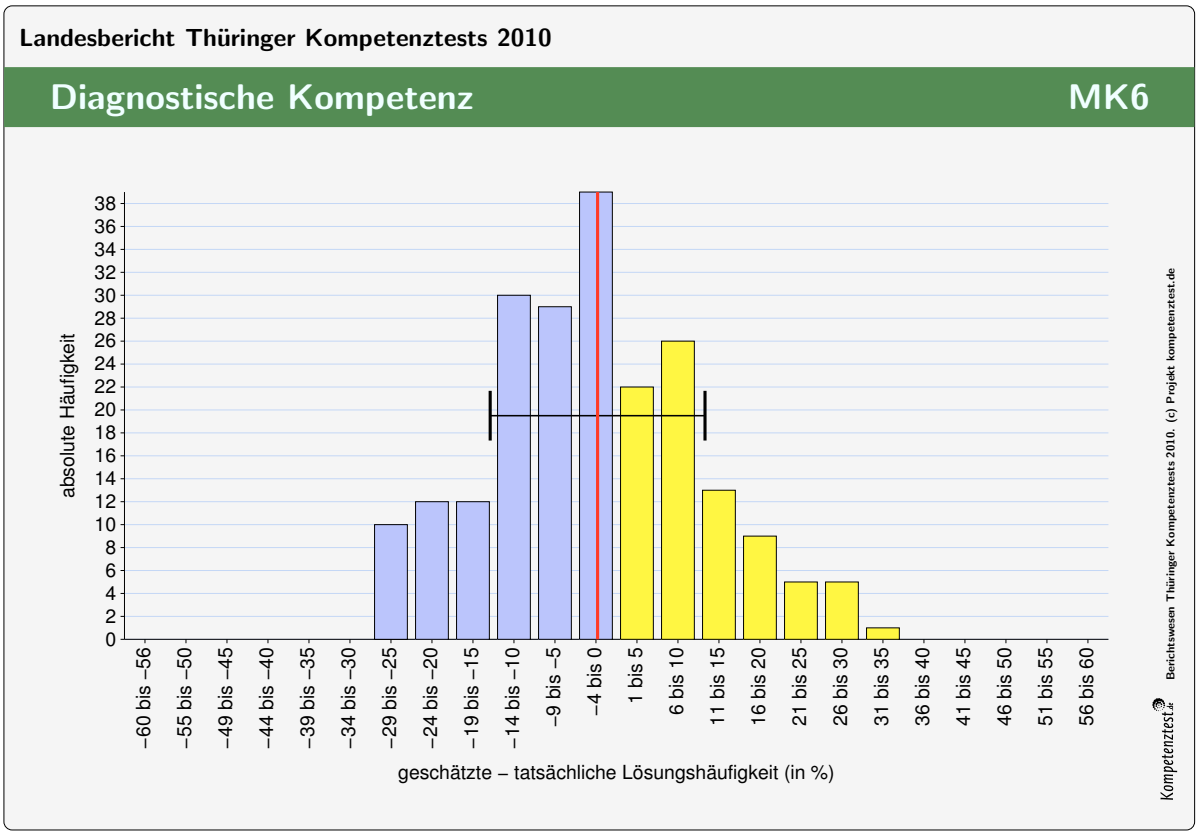


**Abbildung 4.4:** MK6: Abweichung der Lehrerschätzung von der tatsächlichen Lösungshäufigkeit mit mittlerer Abweichung (rot, senkrechter Strich) und Standardabweichung (schwarz, waagerechter Strich)

*Aufgabenebene*: Auch wenn die Mathematiklehrer die Schülerleistungen im Mittel relativ genau einschätzen, zeigt sich bei der Betrachtung einzelner Aufgaben, dass dieser Effekt größtenteils deswegen zustande kommt, weil sich die Über- und Unterschätzungen bei den einzelnen Aufgaben wieder ausgleichen. So findet sich bei der Aufgabe 5 „Größen und ihre Maßeinheiten“ eine starke Überschätzung von 32,7 %. Diese Aufgabe ist dem Kompetenzbereich 3 „Mathematisch modellieren“ und der Leitidee 2 „Messen“ zu zuordnen. Bei dieser Aufgabe sollten die Schüler sinnvolle Maßeinheiten für die aufgeführten Beispiele angeben. Die Schülerleistungen wurden bei den anderen Aufgaben unterschätzt, wobei die auffälligste mittlere Abweichung mit

-16,5 % bei der Aufgabe 22 „Glücksrad“ zu finden ist. Sie gehört zum Kompetenzbereich 4 „Mathematische Darstellungen verwenden“ und der Leitidee 5 „Daten und Zufall“. Die Schüler konnten offensichtlich besser Vergleiche zu Flächen durchführen, als die Mathematiklehrer dies erwartet haben.

**Fach Englisch**

Im Fach Englisch in der Klassenstufe 6 werden die Ergebnisse der diagnostischen Kompetenz nicht schulartspezifisch differenziert, weil keine ausreichende Datenbasis für eine inhaltliche Auswertung gegeben ist. In diesem Jahr ergibt sich eine mittlere Abweichung von -14,27 % über alle Schulformen und Aufgaben hinweg, während letztes Jahr ein Wert von 13,1 %-Punkten vorlag. Das bedeutet, dass die Leistungen der Schüler moderat schlechter eingeschätzt wurden, als sie tatsächlich sind, wobei im vorherigen Jahr noch eine Überschätzung gegeben war. Die Ergebnisse lassen sich folglich auch hier keinem eindeutigen Trend zuordnen. Betrachtet man die Standardabweichung für diese mittlere Abweichung von 13,6 %, so wird deutlich, dass die Lehrereinschätzungen auch moderat voneinander abweichen. Der Anteil der Lehrer, bei denen die mittlere Differenz zwischen geschätzter und tatsächlicher Lösungshäufigkeit größer als 25 %-Punkte ist, liegt bei 17,8 %. Dieser Wert ist positiv zu bewerten, denn er betrug im letzten Jahr noch 48,2 % und ist somit erheblich gesunken. Weniger Englischlehrer unter- oder überschätzen ihre Schüler folglich deutlich.

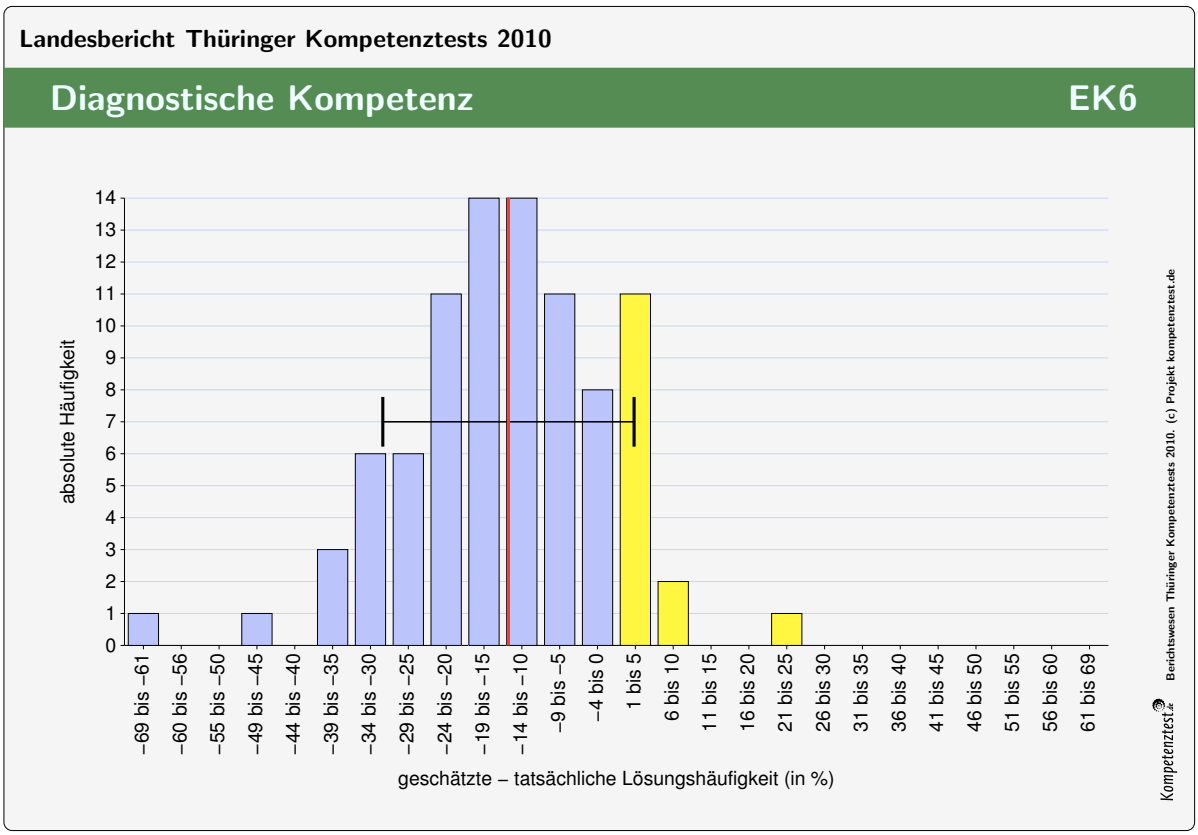


**Abbildung 4.5:** EK6: Abweichung der Lehrerschätzung von der tatsächlichen Lösungshäufigkeit mit mittlerer Abweichung (rot, senkrechter Strich) und Standardabweichung (schwarz, waagerechter Strich)

*Aufgabenebene*: Eine genauere Analyse auf Aufgabenebene zeigt, dass die Lehrer die Schülerleistungen bei allen Aufgaben unterschätzt haben. So liegt bei den Unteraufgaben 1, 2 und 3 von „How to surprise Mum on Mother’s Day?“ eine mittlere Abweichung von -22, %, -21,5 % und -19,2 % vor. Die Schüler haben bessere Leistungen darin erzielt, Bilder zu Instruktionen zu zuordnen, als die Lehrer es erwartet haben. Diese Aufgabe gehört zum Bereich Leseverstehen,

genau wie Aufgabe 9, in der die Schüler Bilder zu Sätzen zuweisen sollten. Auch hier unterschätzen die Englischlehrer ihre Schüler in den Unteraufgaben 1, 2 und 3, wobei die deutlichste negative Abweichung bei der Unteraufgabe 1 mit -32,3 % zu finden ist.

Die Abbildungen 4.3, 4.4 und 4.5 stellen die Verteilung der Abweichungen der geschätzten von der tatsächlichen Lösungshäufigkeit in den Fächern Deutsch, Mathematik und Englisch in der Klassenstufe 6 dar. Außerdem lässt sich hier gut die Breite der Verteilung beobachten, welche den angegebenen Standardabweichungen entspricht. Je weniger Lehrer sich in den beiden Extrembereichen (über 25 %-Punkte in positive und negative Richtung) verschätzen, umso geringer wird die Standardabweichung und desto genauer sind die Einschätzungen der Lehrer im Mittel. Die Streuungen sind jedoch immer im Zusammenhang mit dem Wert der mittleren Abweichung zu interpretieren.

## 4.4 Auswertung der diagnostischen Kompetenz für Klassenstufe 8

Wie in Klassenstufe 3 und 6 ist die Teilnahme an der diagnostischen Kompetenz im Fach Deutsch in Klassenstufe 8 gesunken. 2009 haben 40,9 % bei der ersten Aufgabe eine Angabe gemacht, während dies 2010 34 % der Deutschlehrkräfte waren. In Mathematik ist mit 35,9 % ein ein sehr ähnlicher Wert im Vergleich zu 2009 (34,3 %) zu verzeichnen. Die Englischlehrkräfte, denen das Angebot der Teilnahme an der diagnostischen Kompetenz in diesem Jahr zum zweiten Mal gemacht wurde, haben sich in diesem Jahr sehr viel reger beteiligt. So nahmen 2009 21,3 % der Lehrer teil, während 2010 31,5 % Angaben machten.

**Fach Deutsch**
*Schularten*: Über alle Schularten und Aufgaben hinweg unterscheiden sich die von den Lehrern vorgesagte und die tatsächliche Lösungshäufigkeit im Mittel um 20,9 %. Ähnlich wie im Fach Deutsch in der Klassenstufe 6 überschätzen auch die Deutschlehrer in der Klassenstufe 8 die Schülerleistungen stark, wobei auch die Streuung mit 18,5 % relativ hoch ist. Bei 41,1 % der Deutschlehrer, die an der diagnostischen Kompetenz teilgenommen haben, liegen die Unterschiede zwischen geschätzter und tatsächlicher Lösungshäufigkeit bei über 25 %-Punkten. Das bedeutet, dass fast jeder zweite Lehrer die Leistungen seiner Klassen sehr deutlich über- oder unterschätzt hat. Sehr interessant ist auch ein Vergleich der mittleren Abweichung nach Schularten. Während dieser Wert am Gymnasium 7,7 % beträgt, erreicht er an den Regelschule alarmierende 25,4 %. Im vorherigen Jahr lag der Wert bei den Regelschullehrer noch bei 7,3 % und ist somit stark angestiegen.

*Aufgabenebene*: Bei einer genaueren Analyse der mittleren Abweichung auf Aufgabenebene zeigt sich, dass die Lehrer ihre Schüler bei einigen Aufgaben stark überschätzen. Bei der Aufgabe 2.1 „Zeit für Freizeit“ beträgt der Wert der mittleren Abweichung 24 %. Sie ist dem Kompetenzbereich 3 „Lesen“ zu zuordnen, wobei die Schüler Fragen zu einer Grafik beantworten sollten. Das gelang ihnen offenbar nicht so gut, wie die Deutschlehrer es erwartet haben. Auch die Aufgabe 3.2 und 3.3 „Download-Fieber“ gehört zum Kompetenzbereich „Lesen“, wobei die Schüler zunächst einen Text lesen und anschließend Fragen dazu beantworten sollten. Hier liegt eine mittlere Abweichung von 30,5 % beziehungsweise 28,3 % vor, was zeigt, dass die Lehrer bessere Leistungen erwartet haben.

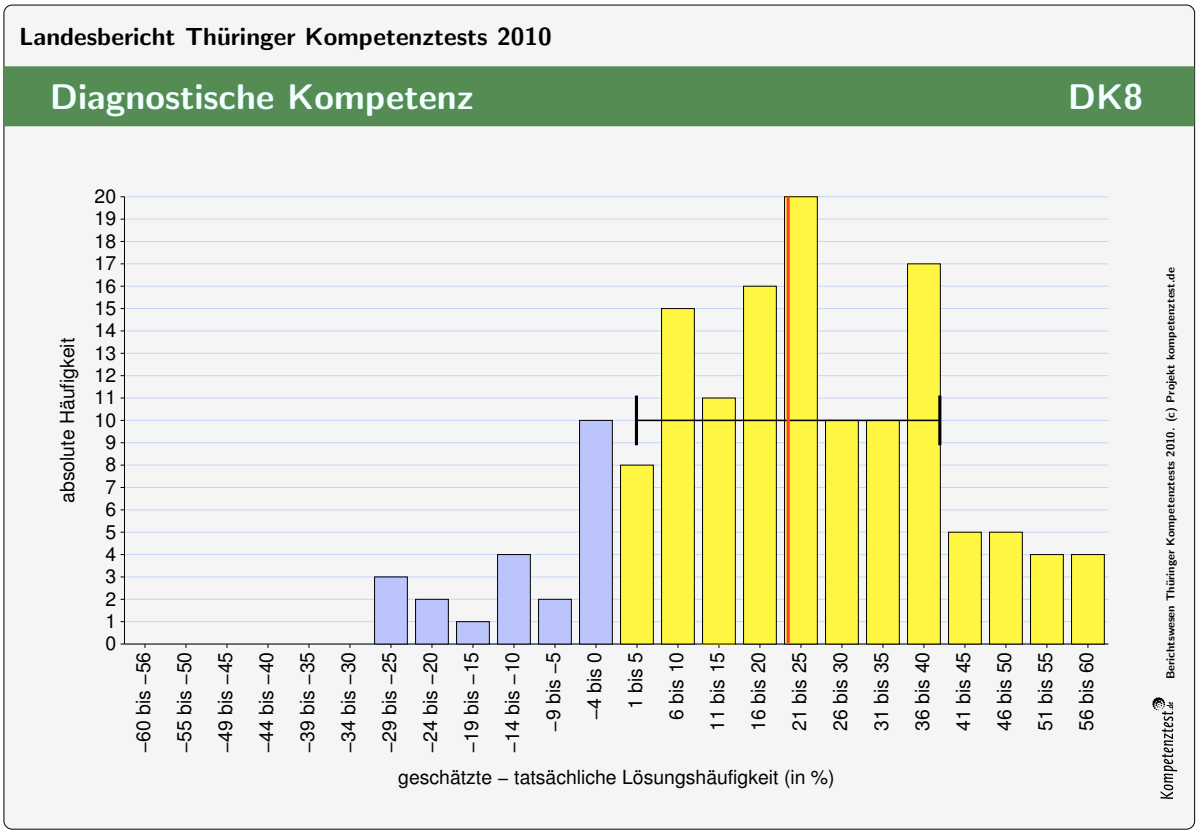


**Abbildung 4.6:** DK8: Abweichung der Lehrerschätzung von der tatsächlichen Lösungshäufigkeit mit mittlerer Abweichung (rot, senkrechter Strich) und Standardabweichung (schwarz, waagerechter Strich)

**Fach Mathematik**

*Schularten*: Auch die Ergebnisse im Fach Mathematik ähneln denen aus der Klassenstufe 6. Es zeigt sich hier eine sehr gute Einschätzung durch die Mathematiklehrer, denn die mittlere Abweichung beträgt 0,3 %-Punkte. Im letzten Jahr überschätzten die Lehrer ihre Schüler noch moderat, so dass dies eine positive Entwicklung ist. Die Einschätzung der Lehrer streut moderat bei einer Standardabweichung von 15,2 % und der Anteil der Lehrer, bei denen eine extreme Über- oder Unterschätzung vorliegt, beträgt 11,2 %. Im letzten Jahr waren es noch 15,5 %, so dass dieser Anteil leicht zurückgegangen ist. Sinnvoll ist auch eine Schulartunterscheidung bei der mittleren Abweichung, denn während in den Regelschulen die von den Lehrern vorhergesagte und die tatsächliche Lösungshäufigkeit sich im Mittel um 4 % unterscheidet, erreichen Gymnasiallehrer einen Wert von -7,2 %. Es zeigt sich ein entgegengesetzter Trend. Regelschullehrer neigen eher dazu ihre Klasse leicht zu überschätzen, während am Gymnasium die Leistungen unterschätzt werden.

*Aufgabenebene*: Die Einschätzung der Mathematiklehrer der Klassenstufe 8 ist nicht nur im Mittel sehr genau. Auch auf Aufgabenebene sind nur wenige Aufgaben auffällig. Ein relativ starke Unterschätzung findet sich bei Aufgabe 17 „Superman“. Hier beträgt die mittlere Abweichung -18,4 %-Punkte. Die Schüler sollten die Größe einer Superman-Statue schätzen und anschließend ihren Lösungsweg angeben. Diese Aufgabe gehört zum Anforderungsbereich 3 „Verallgemeinern und Reflektieren“ und der Leitidee 2 „Messen“. Außerdem lässt sie sich den Kompetenzbereichen 2-6 zuordnen und deckt somit viele mathematische Bereiche ab. Entgegen der Erwartung der Mathematiklehrer schneiden die Schüler hier relativ gut ab.

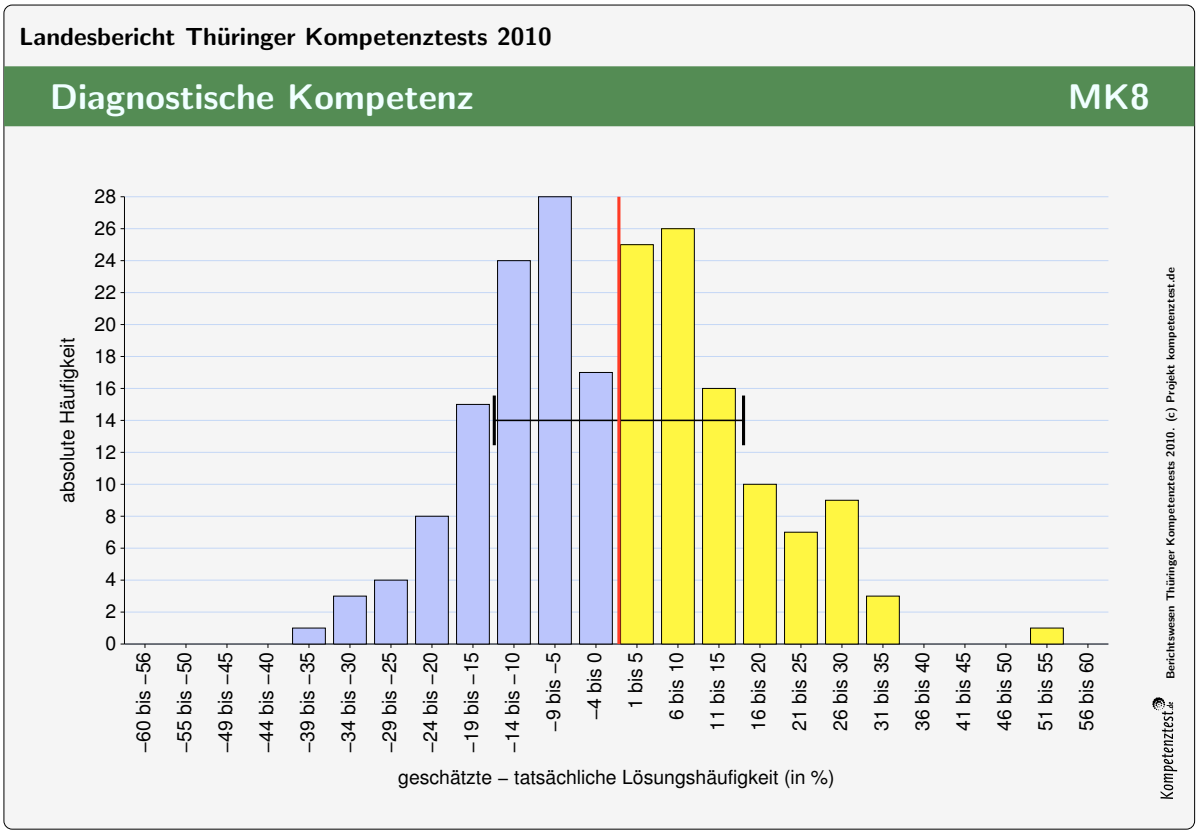


**Abbildung 4.7:** MK8: Abweichung der Lehrerschätzung von der tatsächlichen Lösungshäufigkeit mit mittlerer Abweichung (rot, senkrechter Strich) und Standardabweichung (schwarz, waagerechter Strich)

**Fach Englisch**

Wie in Klassenstufe 6 findet sich auch in Klassenstufe 8 im Fach Englisch ein starker Unterschied zum letzten Jahr. Während die mittlere Abweichung im vorherigen Jahr 13,1 % betrugt, liegt sie dieses Jahr bei -11,1 %. Die Englischlehrer neigen also über alle Aufgaben und Schulformen hinweg dazu ihre Schüler zu unterschätzen. Die Standardabweichung der mittleren Differenz beträgt 15,9 %, so dass die Einschätzung folglich moderat streut. Erfreulich ist, dass der Anteil der Englischlehrer, bei denen die Unterschiede zwischen geschätzter und tatsächlicher Lösungshäufigkeit bei über 25 %-Punkten liegen, von 30,7 % im letzten Jahr auf 18,4 % in diesem Jahr gesunken ist. Das bedeutet, dass nur noch knapp ein Fünftel der Englischlehrer sich deutlich verschätzt. Auch in Klassenstufe 8 wird von einer schulartspezifischen Unterscheidung im Fach Englisch aufgrund einer zu geringen Gesamtteilnehmerzahl abgesehen.

*Aufgabenebene*: Der deutlichste negative Abweichungswert für die Testaufgaben der diagnostischen Kompetenz zeigt sich bei den Aufgaben 2.1., 2.2. und 2.3. „Museum Guide“. Im Durchschnitt unterscheidet sich bei den 3 Aufgaben die von den Lehrern vorhergesagte und die tatsächliche Lösungshäufigkeit um -20,4 %. Bei dieser Aufgabe wurden die Teilnehmer des Kompetenztests dazu aufgefordert, bestimmte Antworten vorgegebenen Fragen zu zuordnen. Sie gehört zum Bereich Leseverstehen und entspricht Anforderungen auf Stufe A2 des Gemeinsamen Europäischen Referenzrahmens. Die Schüler haben hier bessere Leistungen erzielt, als es von den Lehrern erwartet wurde. Auffällig ist auch die Aufgabe 6 „Mini-Saga“, allerdings findet sich hier eine relativ starke Überschätzung der Schülerleistungen. Der Wert der mittleren Abweichung beträgt 21,2 %. Offenbar haben die Englischlehrer erwartet, dass es den Schülern besser gelingt, die Gefühle einer Person einzuschätzen, nachdem sie einen kurzen Text über sie gelesen haben. Die Aufgabe kann ebenfalls dem Leseverstehen zugeordnet werden.

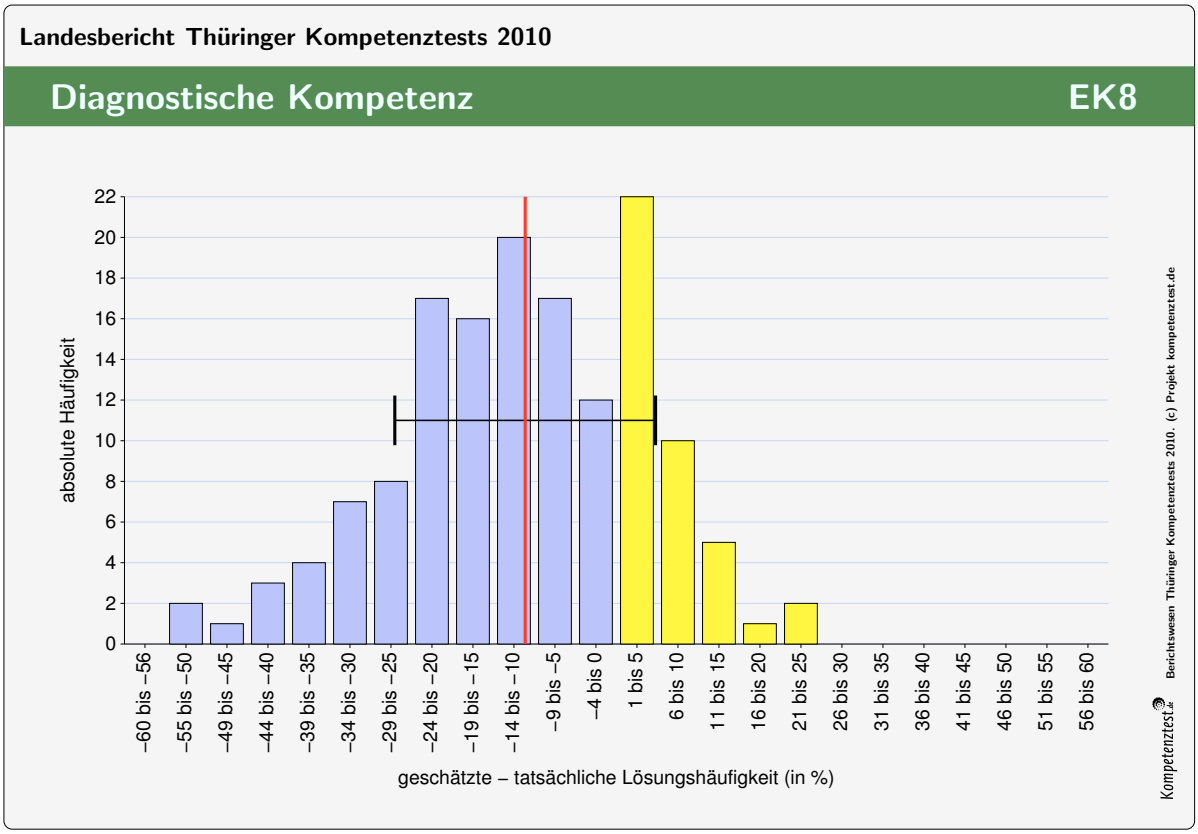


**Abbildung 4.8:** EK8: Abweichung der Lehrerschätzung von der tatsächlichen Lösungshäufigkeit mit mittlerer Abweichung (rot, senkrechter Strich) und Standardabweichung (schwarz, waagerechter Strich)

## 4.5 Zusammenhänge der Ergebnisse der diagnostischen Kompetenz mit anderen Merkmalen

Im Folgenden wird dargestellt, mit welchen anderen Merkmalen die diagnostische Kompetenz im Zusammenhang steht. Zuerst wird der Zusammenhang zwischen den tatsächlich erreichten Punktzahlen in den Kompetenztests und der diagnostischen Kompetenz untersucht. Danach wird die Beziehung zwischen den Ergebnissen der Verwendbarkeitsanalysen und der diagnostischen Kompetenz betrachtet. Bei den Verwendbarkeitsanalysen wurden die Lehrer gebeten, die Kompetenztests hinsichtlich der Nützlichkeit als Instrument zur Lernstandsdiagnostik, für die eigene Unterrichtsentwicklung und die Schulentwicklung zu bewerten und die Lehrplanadäquatheit einzuschätzen. Die berichteten Ergebnisse stellen lediglich eine Beschreibung der diesjährigen Ergebnisse dar. Eine Verallgemeinerung kann nur auf der Grundlage weiterer Untersuchungen stattfinden.

Die diagnostische Kompetenz und die tatsächlich erreichten Punktzahlen stehen in allen Fächern außer Englisch in einem statistisch bedeutsamen negativen Zusammenhang. Das heißt, dass Klassen, die von ihren Lehrern in der diagnostischen Kompetenz stärker überschätzt wurden, eher geringe Punktzahlen aufweisen. Diese Beziehung zeigt sich besonders für Klassenstufe 8. Bei der Betrachtung dieses Zusammenhanges muss jedoch beachtet werden, dass eine Überschätzung mit zunehmender Punktzahl immer unwahrscheinlicher wird, da es sich um den Randbereich der Skala handelt. Dies trifft ebenfalls für die Unterschätzung bei niedrigen Punktzahlen zu.

Außerdem wurde der Zusammenhang zwischen den Ergebnissen der Verwendbarkeitsanalysen und den Teilnehmern der diagnostischen Kompetenz betrachtet. In diese Analyse konnten nur die Lehrer einbezogen werden, die sowohl an der freiwilligen Nutzung der diagnostischen Kompetenz als auch an der Bewertung der Tests bezüglich ihrer Verwendbarkeit teilgenommen haben. In der Klassenstufe 3 zeichnen sich die Deutschlehrer, die zur Extremgruppe der Unter- oder Überschätzer gehören, dadurch aus, dass sie die Aufgaben des Kompetenztests als lehrplanadäquater beurteilen als ihre Kollegen. Ein weiterer Zusammenhang konnte im Fach Englisch der Klassenstufe 8 festgestellt werden. Die Lehrer, die die Schülerleistungen extrem unter- oder überschätzen, schätzen die Kompetenztests eher als ein nützliches Instrument für die eigene Unterrichtsentwicklung und für die Schulentwicklung ein.

## 4.6 Fazit

Im letzten Abschnitt sollen die Ergebnisse der diagnostischen Kompetenz für das Schuljahr 2009 / 2010 zusammenfassend betrachtet und einige auffällige Trends berichtet werden.

*Teilnehmerzahlen*:
Über alle Schulformen, Fächer und Klassenstufen hinweg betrachtet hat sich die Anzahl der Teilnehmer an der diagnostischen Kompetenz im Vergleich zum letzten Jahr kaum verändert. Etwas mehr als ein Drittel der Lehrer haben die diagnostische Kompetenz zur Überprüfung ihrer Leistungseinschätzung genutzt. Wenn man bedenkt, dass diese Nutzung freiwillig ist, ist diese Teilnehmerquote erfreulich hoch.

*Fächervergleiche*:
Die mittleren Abweichungen der Lehrereinschätzungen von den tatsächlichen Lösungshäufigkeiten für die einzelnen Fächer lassen sich in diesem Jahr nur schwer über Klassenstufen zusammenfassen. Während in der Klassenstufe 3 die Deutschlehrer im Durchschnitt die Schülerleistungen mittelstark unterschätzen, neigen diese in den weiterführenden Schulen zu einer starken Überschätzung. Auch im Fach Mathematik unterschätzen die Fachlehrer ihre Schüler in der Klassenstufe 3, während die Lehrer der Klassenstufe 6 und 8 die zu erwartenden Ergebnisse relativ genau einschätzen. Im Fach Englisch lässt sich in den Klassenstufen 6 und 8 ein Fächertrend feststellen. In beiden Fällen unterschätzen die Lehrer ihre Klasse moderat.

Auch im Vergleich mit den letzten Jahren lassen sich die mittleren Abweichungen der Lehrereinschätzungen von den tatsächlichen Lösungshäufigkeiten nicht eindeutig einem Trend zuordnen. Das bedeutet, dass die Ergebnisse der diagnostischen Kompetenz nicht über die Jahre hinweg stabil in eine Richtung einschlagen. Innerhalb der einzelnen Fächer konnten zum Teil deutliche Veränderungen in positive und negative Richtung beobachtet werden. Diese auffällige Heterogenität impliziert, dass eine testspezifische Interpretation notwendig ist. Weiterhin beeinflusst auch die Aufgabenwahl das Ergebnis der diagnostischen Kompetenz. Bei einem Vergleich mit dem Vorjahr fällt zum Beispiel auf, dass sich im Fach Deutsch die Abweichungen vergrößert haben. In der Klassenstufe 3 unterschätzen die Fachlehrer ihre Schüler im Mittel stärker und in den weiterführenden Schulen hat die Überschätzung stark zugenommen. Im Fach Mathematik weichen die Werte der mittleren Abweichungen in der Klassenstufe 3 und 6 nicht so stark von denen des Vorjahres ab. Als positiv zu bewerten ist, dass die Mathematiklehrer in der Klassenstufe 8 die Schülerleistungen genauer einschätzen als im letzten Jahr. Auffällig ist weiterhin, dass die Ergebnisse im Fach Englisch gegenläufig zu den Vorjahresergebnissen sind. Während im vorherigen Jahr noch eine Überschätzung der zu erwartenden Lösungshäufigkeit vorlag, zeigt sich in den aktuellen Ergebnissen in beiden Klassenstufen eine mittelstarke Unterschätzung.

In Bezug auf die Standardabweichungen lässt sich festhalten, dass diese in allen Fächern relativ hoch sind. Das bedeutet, dass die verschiedenen Lehrer innerhalb eines Test nicht sehr ähnlich geschätzt haben. Am auffälligsten erweist sich die Standardabweichung im Fach Deutsch der Klassenstufe 8 mit 18,5 %. Ein auffälliger Fächertrend kann jedoch nicht festgestellt werden.

Bei dem Vergleich der mittleren Differenz zwischen geschätzter und tatsächlicher Lösungshäufigkeit, die größer als 25 %-Punkte ist, ergibt sich, dass die Deutschlehrer ihre Schüler häufiger extrem unter- oder überschätzen als die Mathematik- oder Englischlehrer. Während der Anteil der Lehrer in Deutsch über alle Klassen hinweg 35,3 % beträgt, liegt er in Mathematik mit 12,6 % und im Fach Englisch mit 12,0 % auf einem ähnlichen relativ niedrigen Niveau. Im Vergleich zum Vorjahr, ist der Wert in Englisch als positiv zu bewerten, während er im Fach Deutsch im Gegensatz zu 11,0 % im letzten Jahr deutlich zugenommen hat.

In den letzen Jahren hat sich kaum eine Fächerkultur bezüglich der diagnostischen Kompetenz entwickelt. Das wurde auch mit den diesjährigen Analysen wieder bestätigt. Im Hinblick auf die Nutzung der Ergebnisse der diagnostischen Kompetenz für die Lehrkräfte, sind die Resultate auf Aufgabenebene entscheidend. Wie bereits ausführlich berichtet 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 ungenauen Einschätzungen des Wissens- und Leistungsstandes der Schüler 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.

**Das Wichtigste in Kürze**

- Die Teilnehmerzahlen entsprechen in etwa denen des letzten Jahres. Insgesamt haben ca. ein Drittel der Lehrer an der diagnostischen Kompetenz teilgenommen.
- Bezüglich der Ergebnisse der diagnostischen Kompetenz gibt es keine Trends über die Jahre, die Fächer und die Klassenstufen. Lediglich die Deutschlehrer haben die Leistung ihrer Schüler in diesem Jahr häufiger extrem unter- bzw. überschätzt.

# 5 Rückmeldung der Testergebnisse

Die Rückmeldung der Testergebnisse wurde in den vergangenen Jahren im Hinblick auf Verständlichkeit und Nützlichkeit optimiert. Die Ergebnisberichte liegen mittlerweile in ausgereiften und von den Lehrkräften wertgeschätzten Formaten vor, welche im Jahr 2010 weitgehend unverändert blieben. Sie liefern Informationen auf verschiedenen Ebenen. Im Hinblick auf den einzelnen Schüler zeigen die Testergebnisse die individuellen Lernstände und liefern Hinweise auf Förderbedarf. Im Jahr 2010 wurde auf Wunsch vieler Lehrkräfte erstmals im Fach Deutsch der Klassenstufe 6 und 8 ein spezieller Schülerbericht angeboten. Dieser gibt die Ergebnisse jedes einzelnen Schülers im Gesamtergebnis sowie in den verschiedenen Kompetenzbereichen in grafischer Form wieder. Zusätzlich wird die Verteilung der Ergebnisse aller Schüler der gleichen Schulart zur vergleichenden Einschätzung der individuellen Schülerleistung dargestellt. Wenn sich der Schülerbericht bewährt, wird er im kommenden Jahr auch in den anderen Fächern den Lehrkräften zur Verfügung gestellt werden. Sofern die Schüler in Klassenverbänden und nicht in Kursen unterrichtet werden, kann dieser Bericht auch fachübergreifend angeboten werden.

Zentrales Ziel der Kompetenztests ist jedoch nicht der Blick auf die Leistung des einzelnen Schülers, sondern auf die eigene Klasse und den eigenen Unterricht. In diesem Sinne liefern die zusammengefassten Ergebnisse auf Klassen- und Schulebene 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.

## 5.1 Formen der Rückmeldungen

Entsprechend diesen drei Ebenen erfolgen 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).

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 5.2 beschrieben. Zusätzlich wurden wie oben erwähnt in einem Pilotprojekt für das Fach Deutsch individuelle Schülerberichte erstellt. Abb. 5.5 zeigt Ausschnitte aus einem solchen Bericht.

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 5.6 zeigt Ausschnitte aus einem solchen Schulbericht.

Die dritte Rückmeldewelle umfasst diesen Landesbericht. Er liefert einen Überblick über die landesweiten Testergebnisse sowie eine Vielzahl von vertiefenden Analysen möglicher Einflussvariablen und Hintergrundbedingungen.

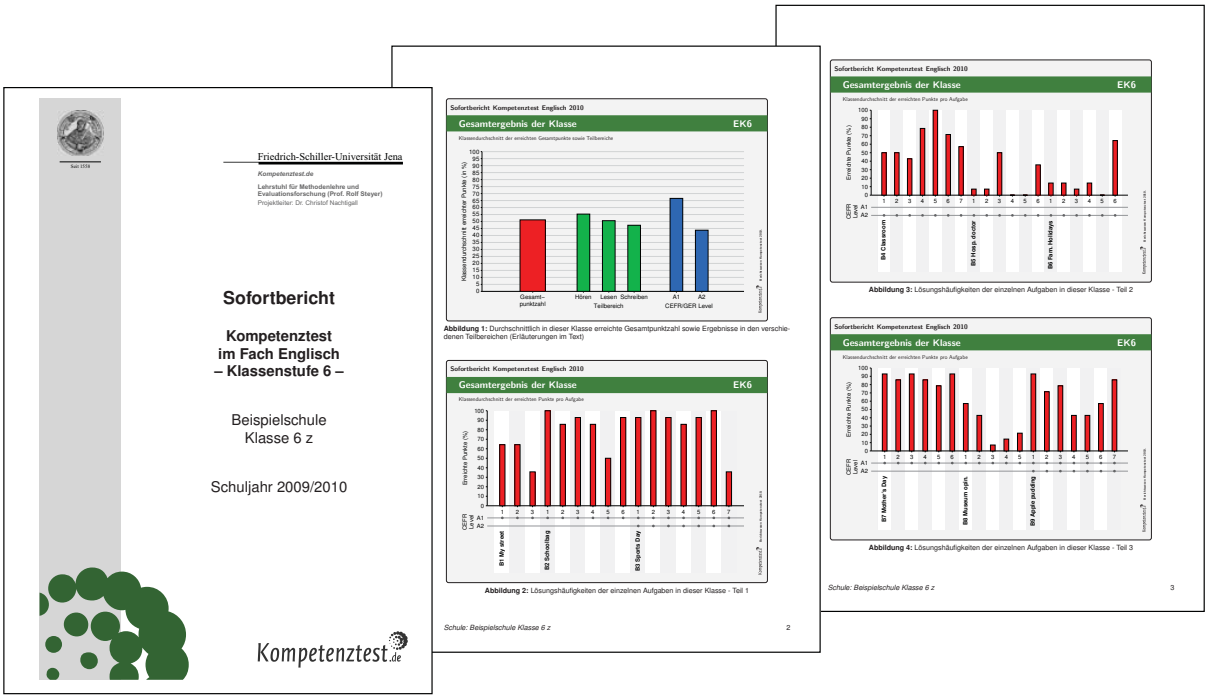


**Abbildung 5.1:** Beispielseiten eines Sofortberichts im Fach Englisch, Klassenstufe 6

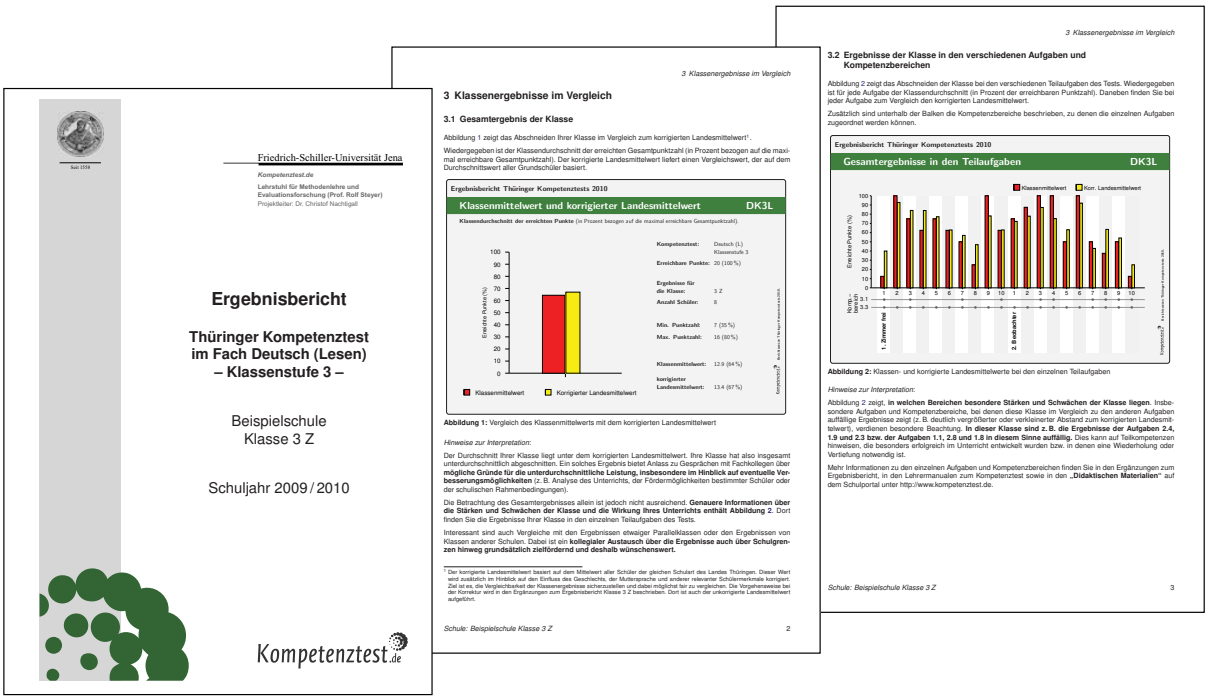


**Abbildung 5.2:** Beispielseiten eines Ergebnisberichts im Fach Deutsch, Klassenstufe 3

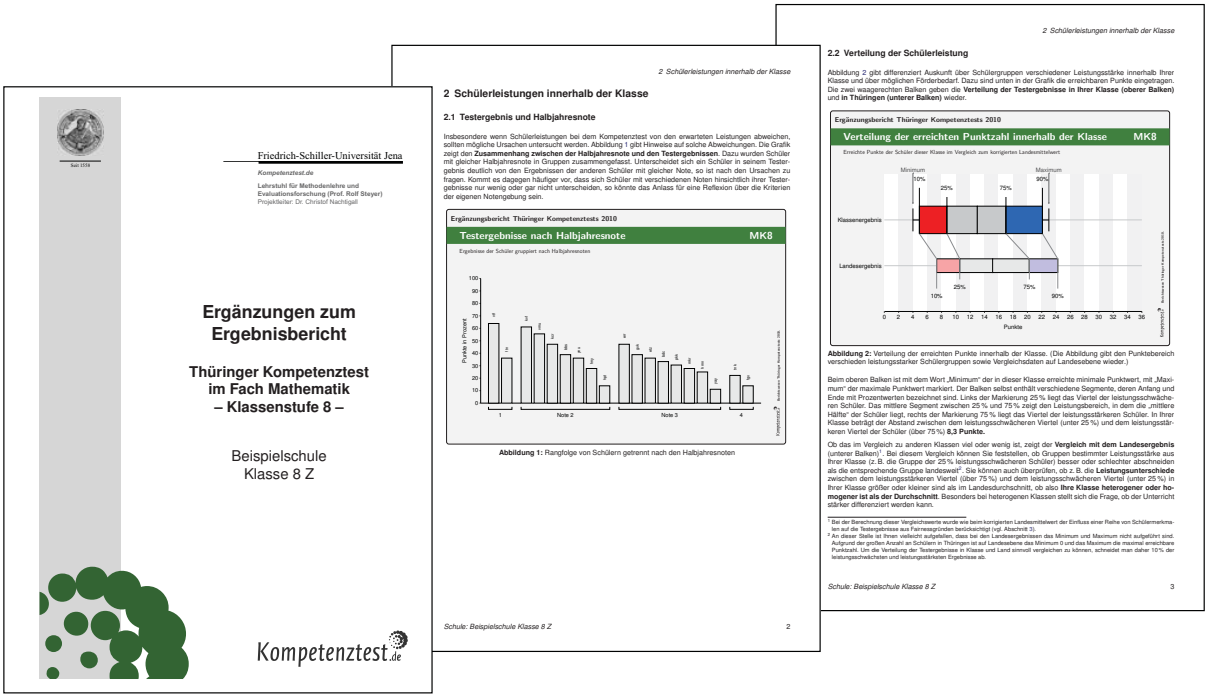


**Abbildung 5.3:** Beispielseiten der Ergänzungen zum Ergebnisbericht im Fach Mathematik, Klassenstufe 8 – Teil 1

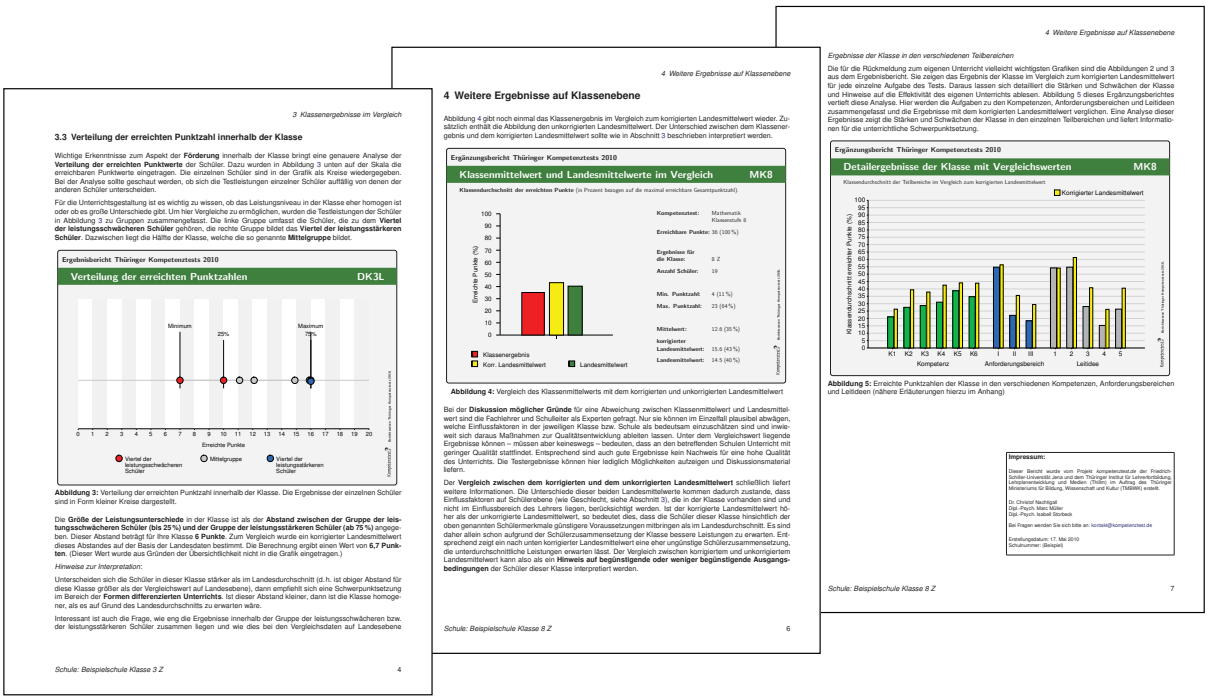


**Abbildung 5.4:** Beispielseiten der Ergänzungen zum Ergebnisbericht im Fach Mathematik, Klassenstufe 8 – Teil 2

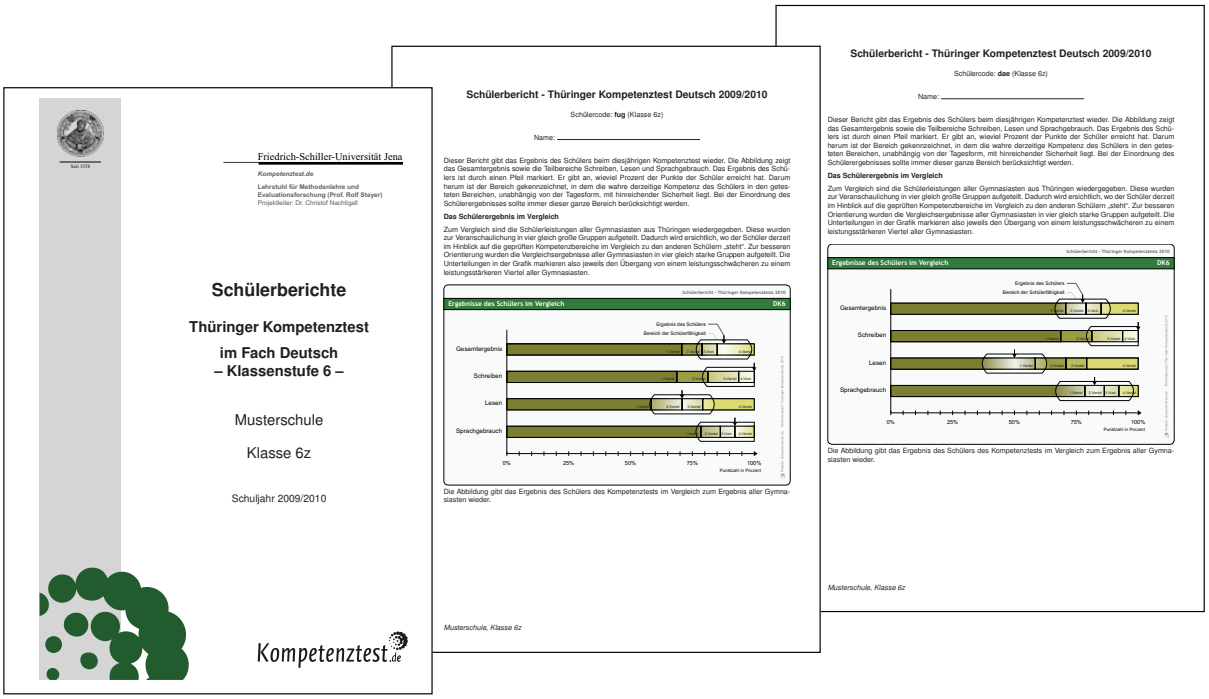


**Abbildung 5.5:** Beispielseiten aus Schülerberichten

**Abbildung 5.6:** Beispielseiten eines Schulberichts

## 5.2 Korrigierter Landesmittelwert und faire Vergleiche

Neben den Testergebnissen der eigenen Klasse bzw. Schule werden bei den Ergebnisrückmeldungen der zweiten und dritten Welle auch Vergleichswerte zurück gemeldet, die möglichst faire Vergleiche bieten. 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.

Insgesamt konnten im Jahr 2010 folgende Schülermerkmale berücksichtigt werden:

- Geschlecht
- Muttersprache
- Wiederholer einer Klassenstufe
- Diagnose besonderer Lernschwierigkeiten bzw. Sonderpädagogischer Förderbedarf
- Sozioökonomischer Status / Bildungsnähe[1]

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.

[1] Diese wichtigen Hintergrundmerkmale werden durch die so genannte „Bücherfrage“ erfasst (vgl. Abschnitt 7.2.3.)

Im Jahr 2009 konnte durch die erstmalige Hinzunahme des Vorwissens der Schüler deutlich fairere Vergleichswerte als in den früheren Jahren zur Verfügung gestellt werden. Bedauerlicher Weise ließ sich dies aus technischen Gründen im Jahr 2010 nicht realisieren, so dass die Qualität der Vergleichswerte in diesem Jahr auf dem Stand des Jahres 2008 zu sehen ist. Für die Kompetenztests 2011 ist die Berücksichtigung des Vorwissens erneut vorgesehen.

*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, Vorwissen 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.

# 6 Testergebnisse auf Landesebene

Dieser Abschnitt enthält die Ergebnisse der insgesamt neun verschiedenen Kompetenztests 2010 auf Landesebene. In diesem Jahr fanden die Tests im Bereich Deutsch und Mathematik in der Klassenstufe 3, Deutsch, Mathematik und Englisch in der Klassenstufe 6 und die Kompetenztests Deutsch, Mathematik, Englisch und Französisch 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 2010 entnommen werden.

## 6.1 Erreichte Gesamtpunktzahlen

In Tabelle 6.1 sind die Mittelwerte und Streuungen der erreichten Punktzahlen sowie die Teilnehmerzahlen der Kompetenztests 2010 wiedergegeben.

| Test | Mittelwert | (Streuung) | Schüleranzahl |
|---|---|---|---|
| **DK3L** | 12,14 | (4,41) | 15 530 |
| **DK3R** | 13,15 | (4,38) | 15 484 |
| **MK3** | 35,06 | (9,44) | 15 501 |
| **DK6** | 95,31 | (26,04) | 11 000 |
| **MK6** | 15,69 | (5,28) | 12 421 |
| **EK6** | 63,51 | (19,74) | 10 158 |
| **DK8** | 41,98 | (11,31) | 12 272 |
| **MK8** | 16,97 | (7,24) | 12 268 |
| **EK8** | 54,78 | (17,04) | 12 318 |
| **FK8** | 47,23 | (18,58) | 96 |

**Tabelle 6.1:** Ergebnisse der Kompetenztests 2010 für die Klassenstufen 3 Angegeben sind die Mittelwerte und Streuungen der erreichten Gesamtpunktzahlen aller Thüringer Schüler

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 beinhalten auch leichte Aufgaben (siehe Abschnitt 6.3), so dass die Wahrscheinlichkeit erhöht ist, dass schwächere Schüler ebenfalls zu Punkten kommen.

Auch für die leistungsstarken Schüler bieten die Tests genügend Differenzierungsmöglichkeiten. Die Ausnahme stellen diesbezüglich die Englischtests der Klassenstufen 6 und 8 dar: In diesen Tests wurde häufiger eine annähernd volle Punktzahl erreicht – vor allem in Thüringer Gymnasien. Dies ist in der Konstruktion der Tests begründet, die in Abschnitt 2 näher beschrieben sind. Dadurch werden systematisch bessere Ergebnisse erreicht, was jedoch nicht mit höherer Unterrichtsqualität gleichgesetzt werden kann - diese auch nicht ausschließt.

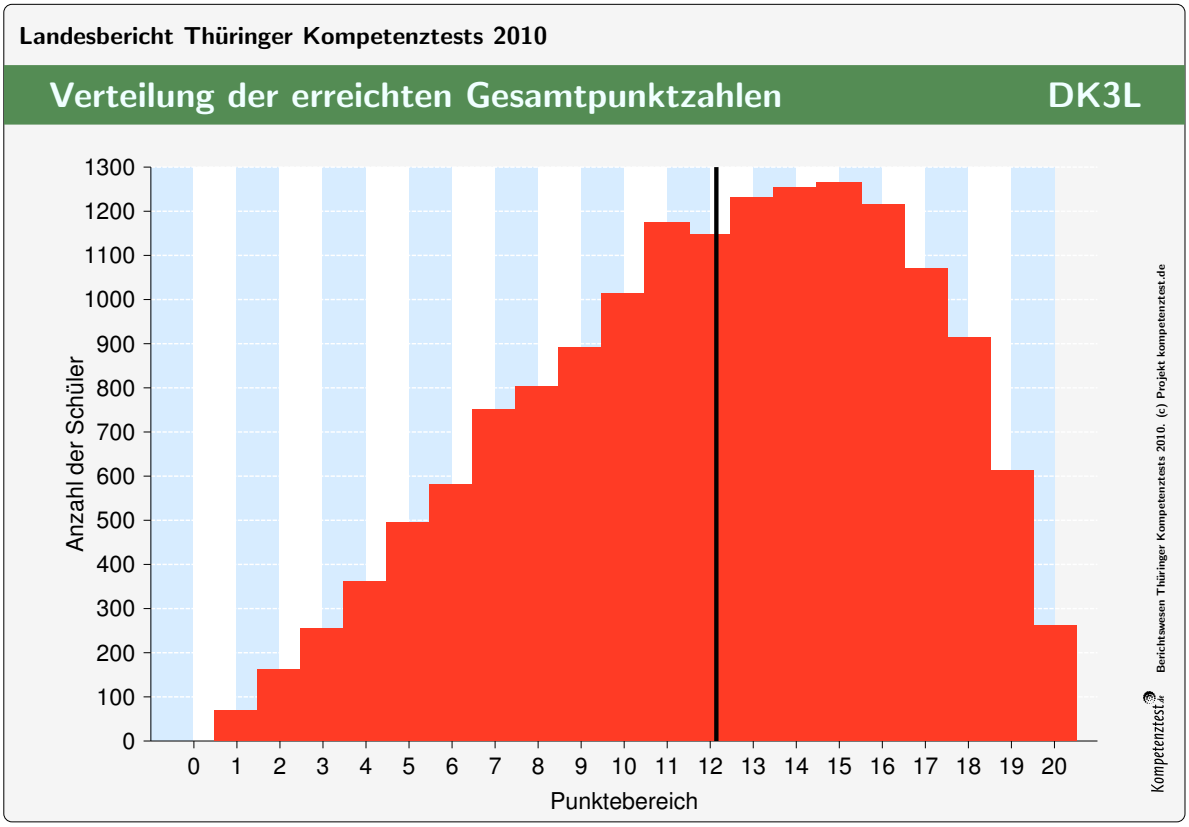


**Abbildung 6.1:** Verteilung der Gesamtpunktzahlen im Fach Deutsch Lesen, Klassenstufe 3

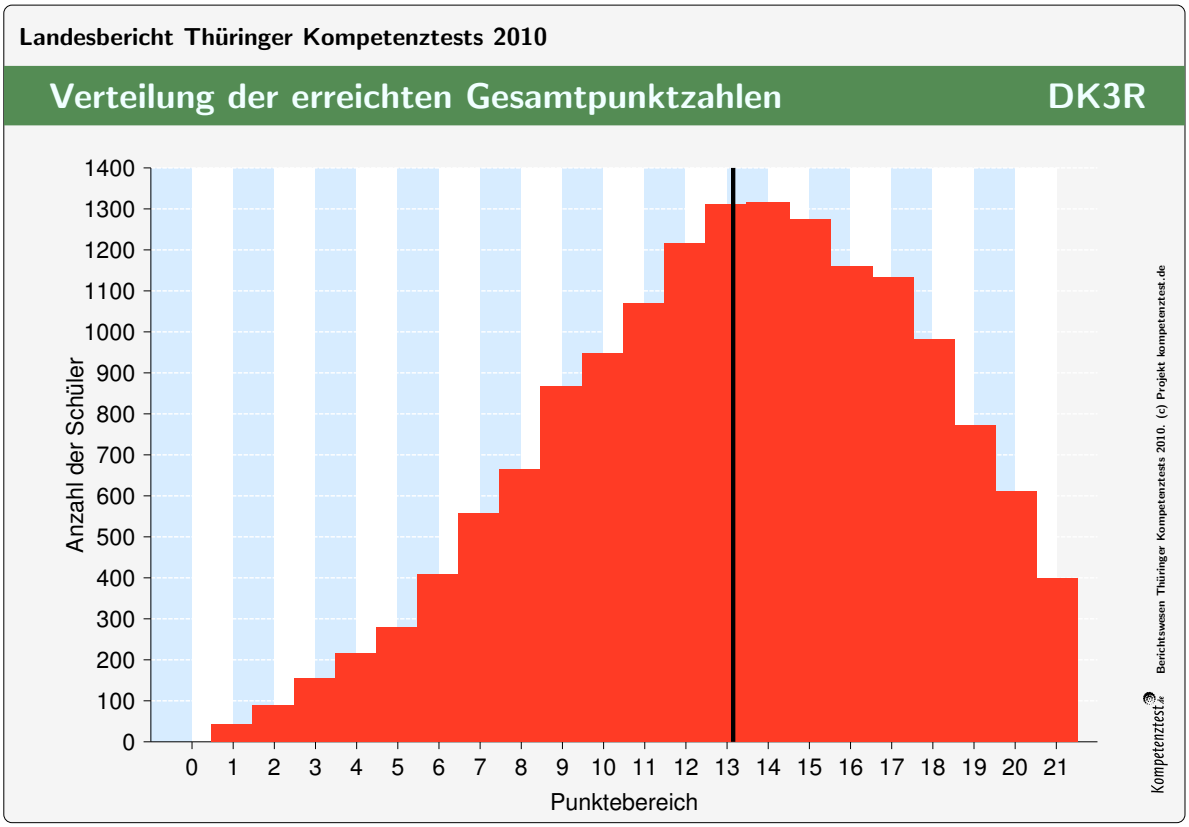


**Abbildung 6.2:** Verteilung der Gesamtpunktzahlen im Fach Deutsch Rechtschreibung, Klassenstufe 3

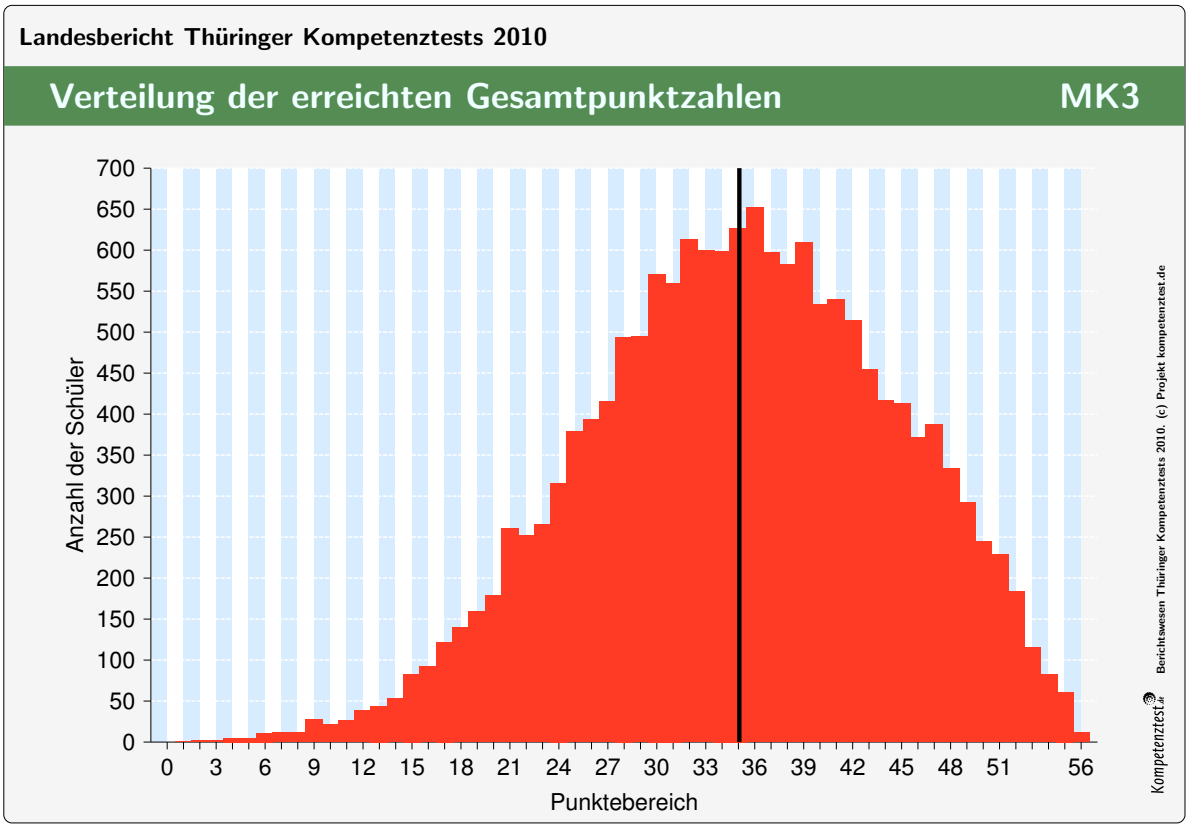


**Abbildung 6.3:** Verteilung der Gesamtpunktzahlen im Fach Mathematik, Klassenstufe 3

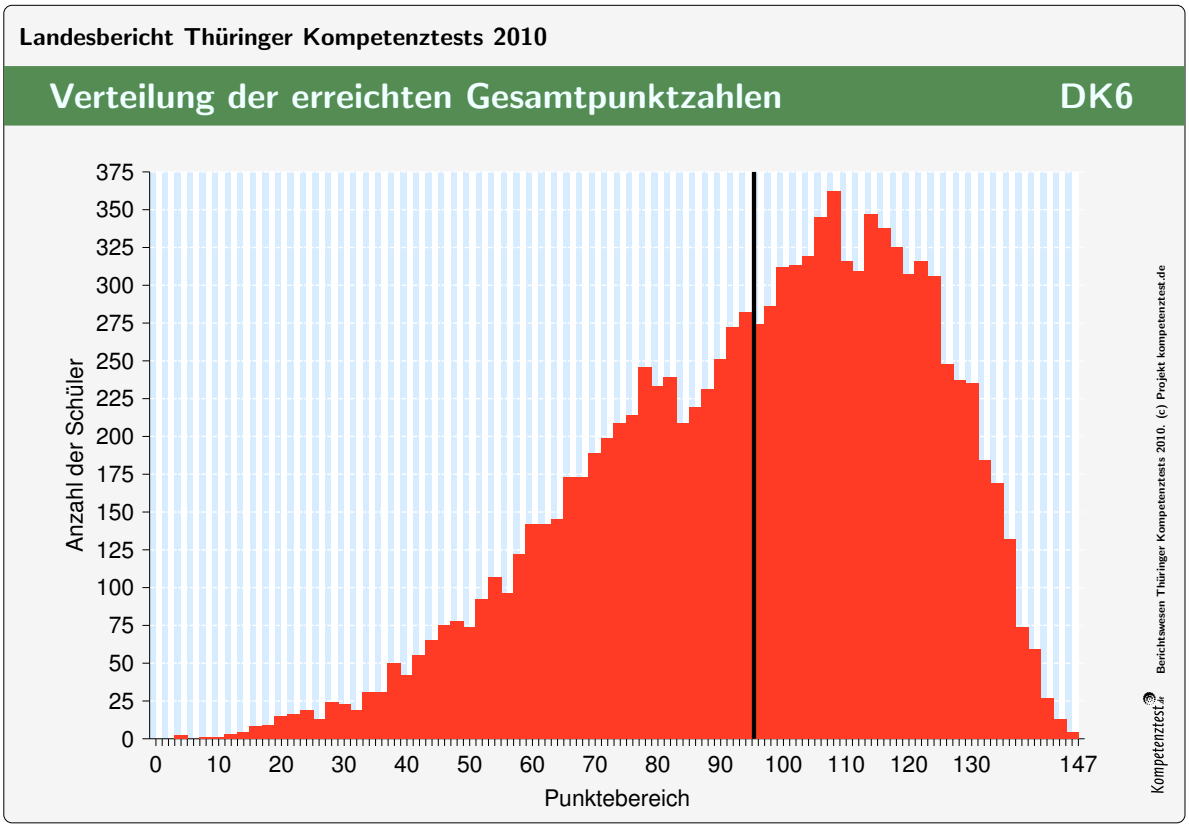


**Abbildung 6.4:** Verteilung der Gesamtpunktzahlen im Fach Deutsch, Klassenstufe 6

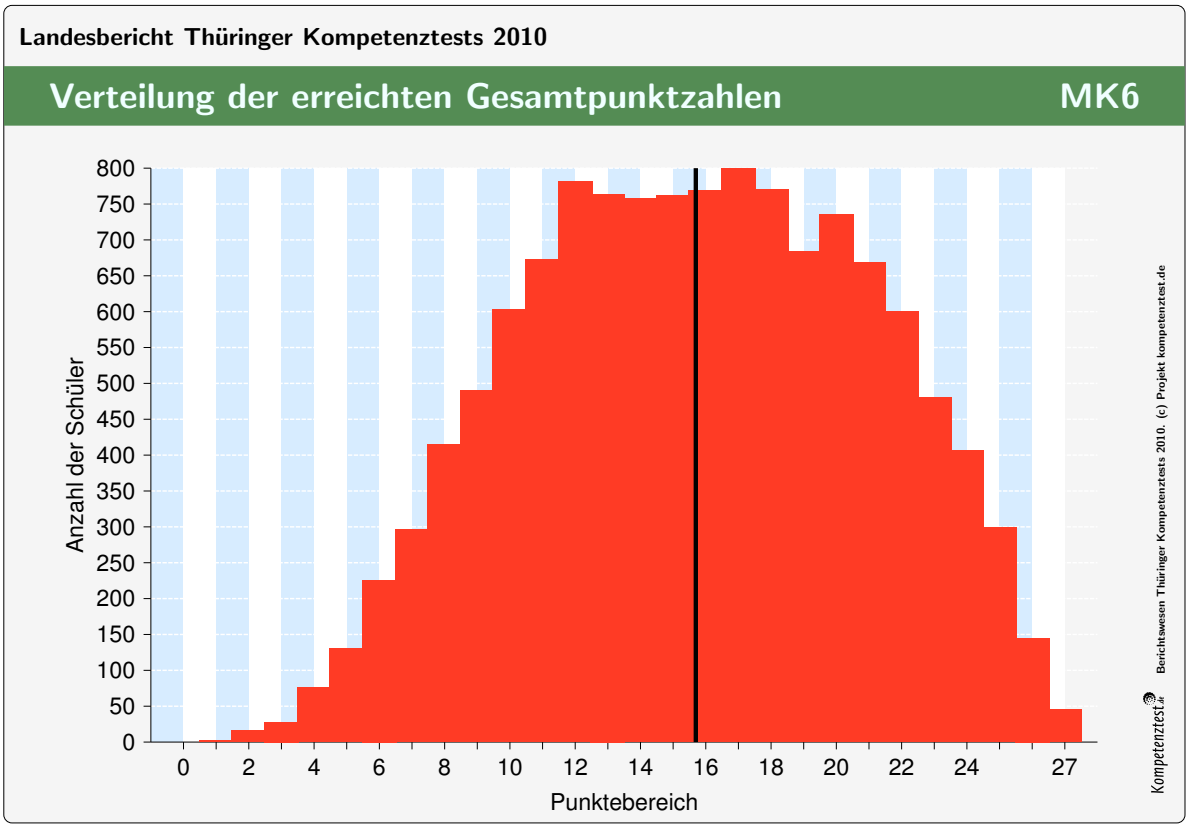


**Abbildung 6.5:** Verteilung der Gesamtpunktzahlen im Fach Mathematik, Klassenstufe 6

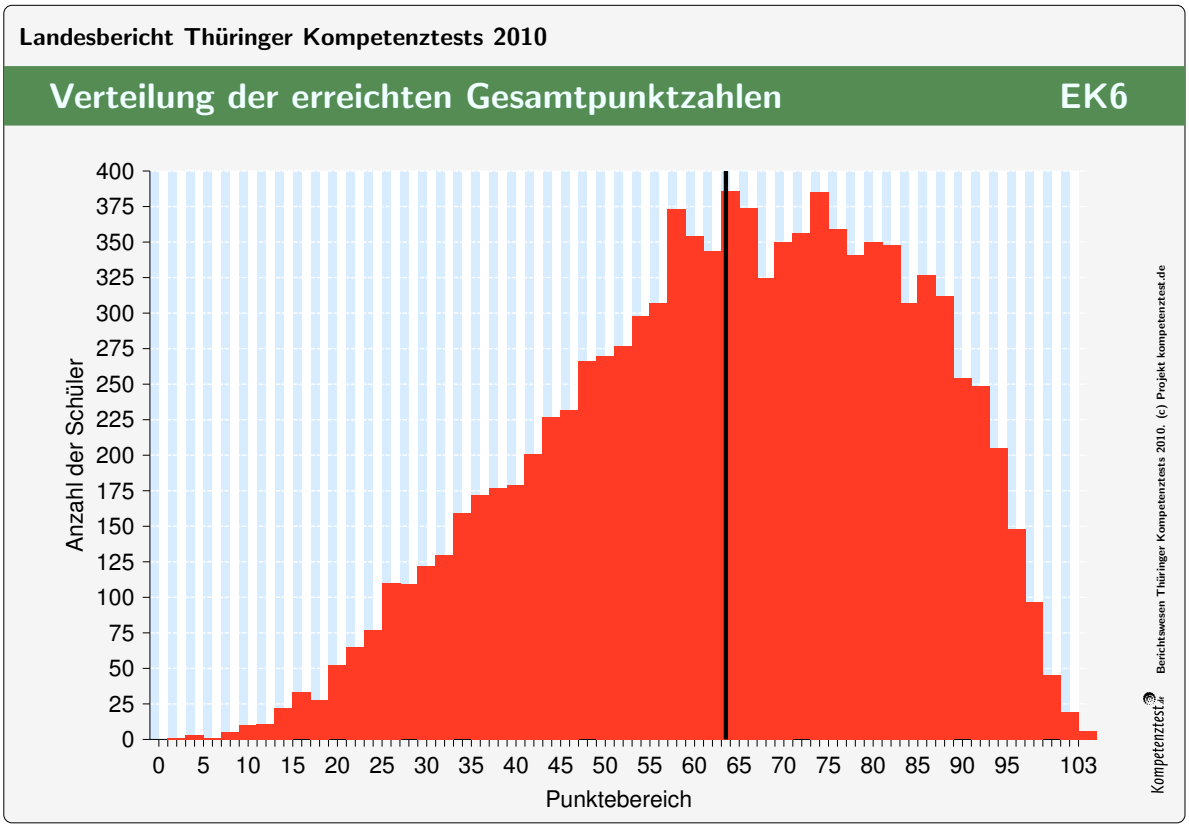


**Abbildung 6.6:** Verteilung der Gesamtpunktzahlen im Fach Englisch, Klassenstufe 6

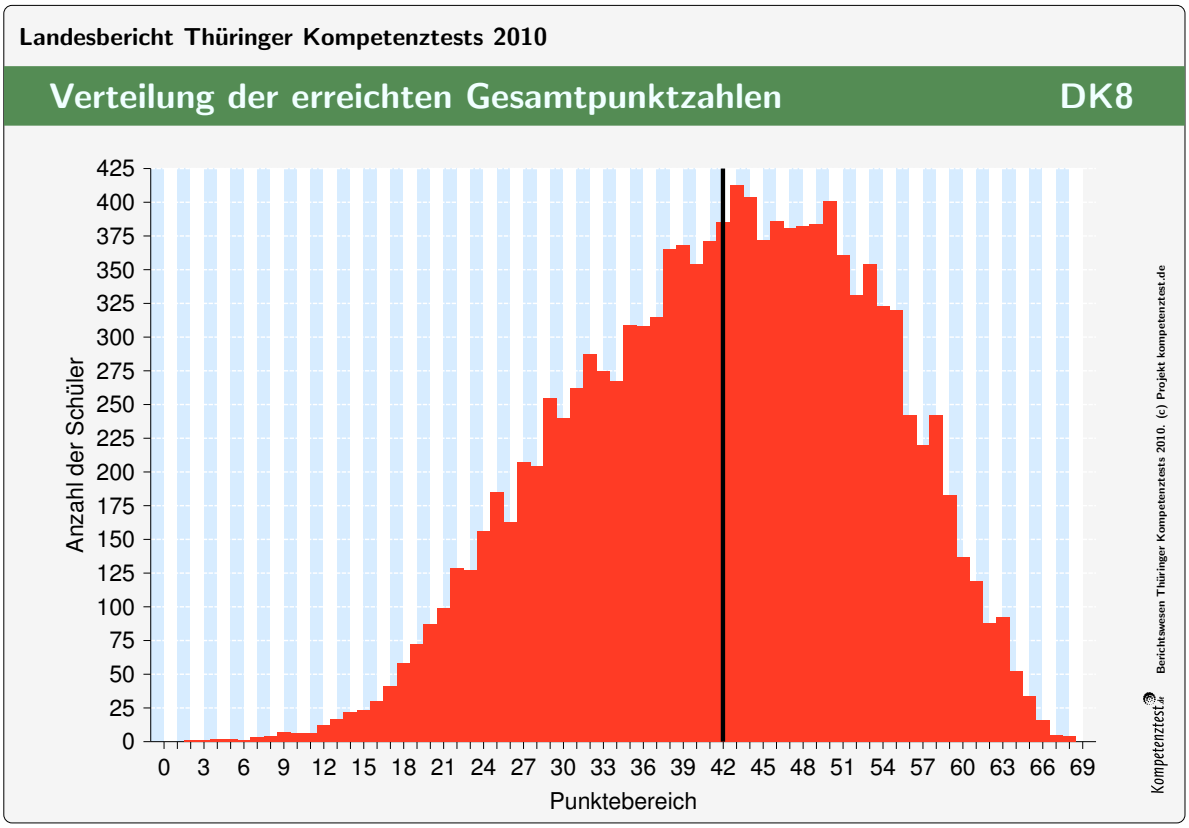


**Abbildung 6.7:** Verteilung der Gesamtpunktzahlen im Fach Deutsch, Klassenstufe 8

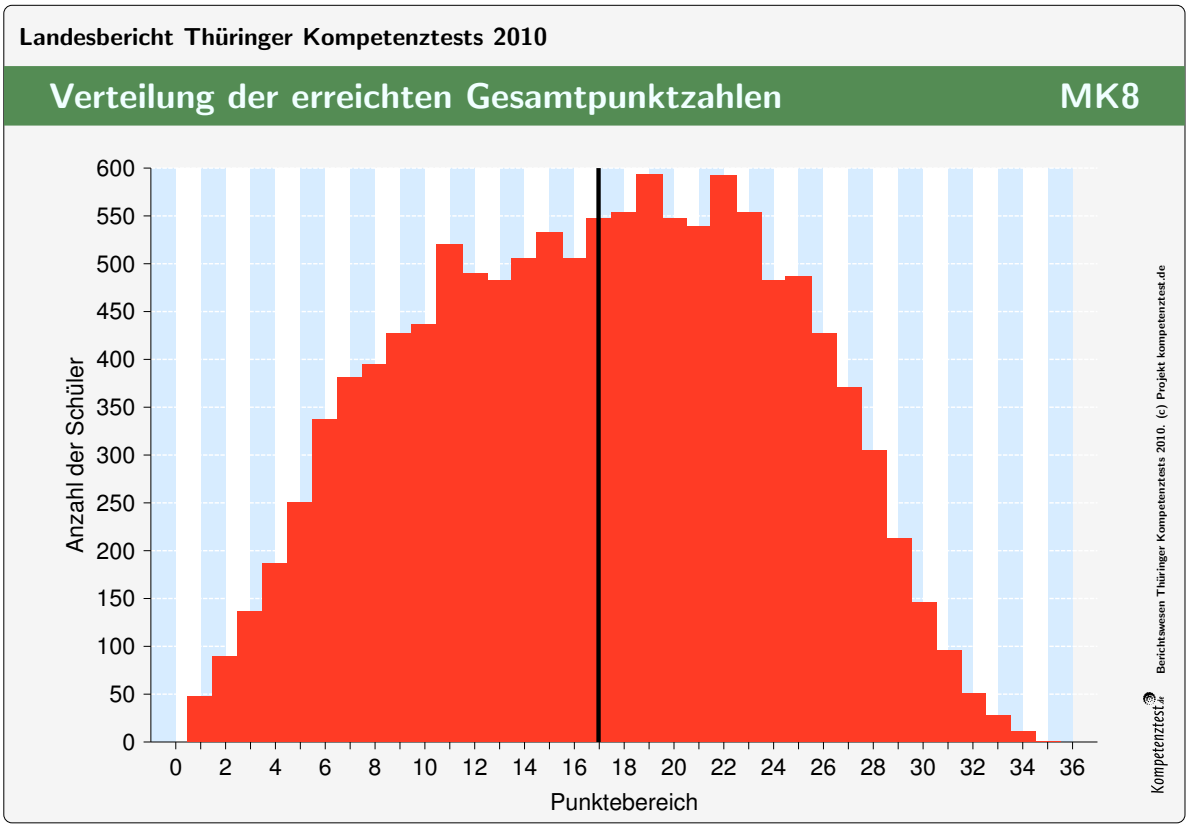


**Abbildung 6.8:** Verteilung der Gesamtpunktzahlen im Fach Mathematik, Klassenstufe 8

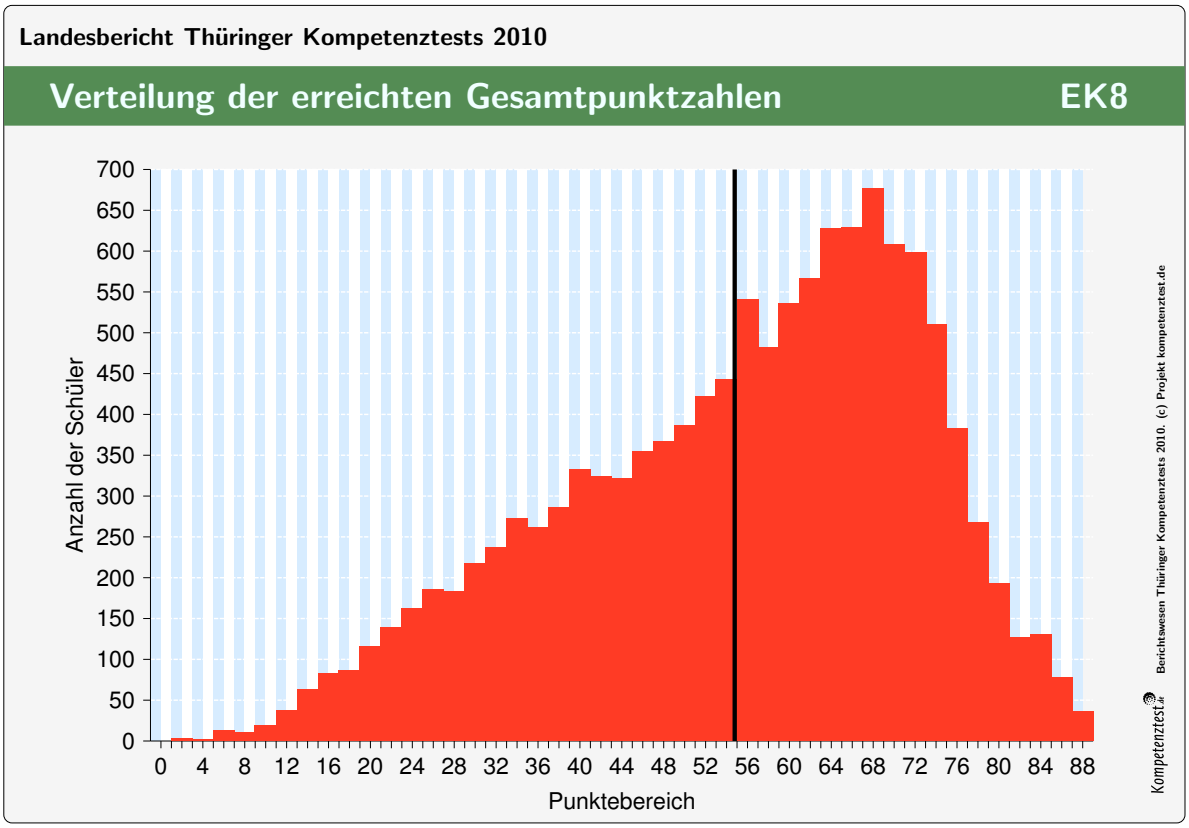


**Abbildung 6.9:** Verteilung der Gesamtpunktzahlen im Fach Englisch, Klassenstufe 8

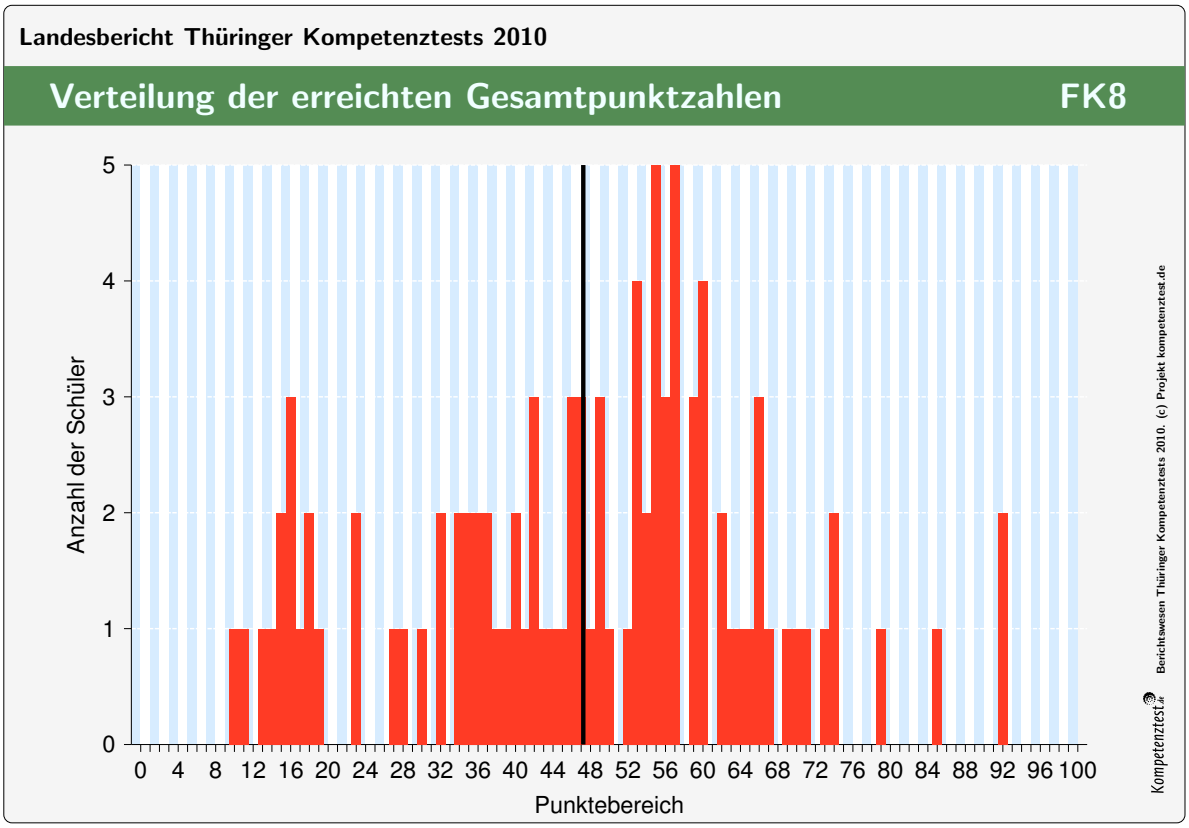


**Abbildung 6.10:** Verteilung der Gesamtpunktzahlen im Fach Französisch, Klassenstufe 8

## 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.

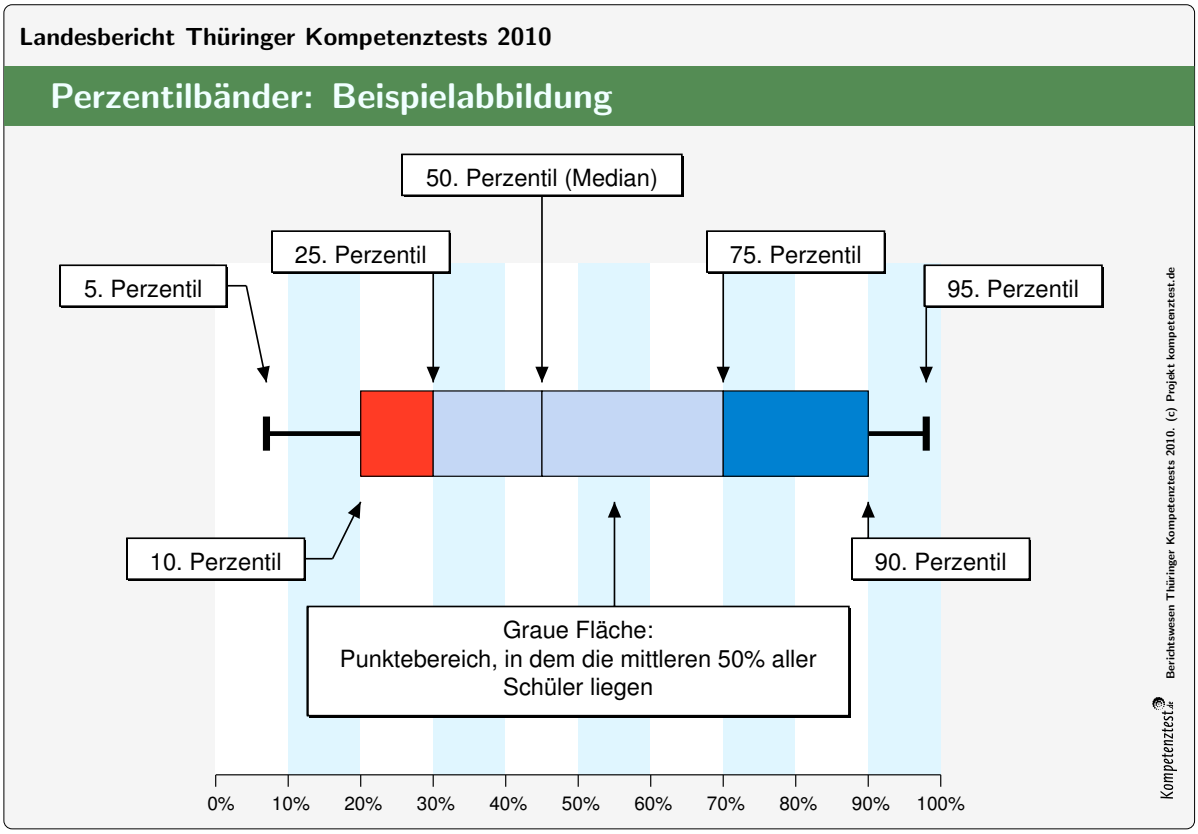


**Abbildung 6.11:** Beispiel für die Darstellung der erreichten Punktzahl in Prozent in Form eines Perzentilbandes

Abbildung 6.11 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 der Klassenstufe 8. Hier liegen die 25. Perzentile der Gymnasiasten mindestens auf Höhe des 75. Perzentils der Regelschüler. Das bedeutet, dass die Gymnasiasten im Mittel höhere Punktzahlen erreichen als die Regelschüler. Die Unterschiede des Englischtests, sind wie im Vorjahr zwischen Gymnasien und Regelschulen der 8. Klassenstufe noch deutlicher. Insgesamt kann festgestellt werden, dass durch diese Form der Darstellung die Ergebnisse differenziert nach Schularten durchweg gut veranschaulicht werden. Im Fach Französisch in der Klassenstufe 8 wurde aufgrund zu geringer Fallzahlen keine weitere Unterscheidung vorgenommen.

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

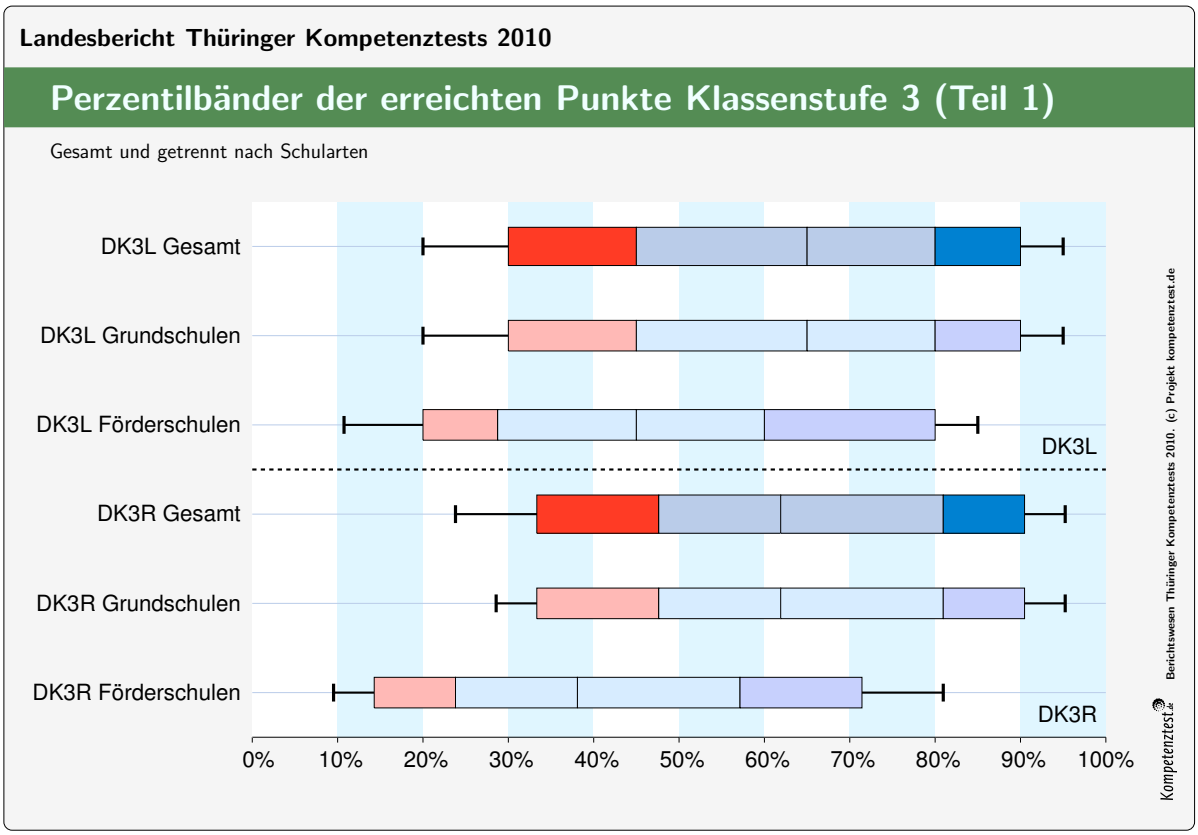


**Abbildung 6.12:** Ergebnisse bei den beiden Tests in Deutsch in der Klassenstufe 3 getrennt nach Schulart. Die Abbildung gibt die erreichte Punktzahl in Prozent wieder.

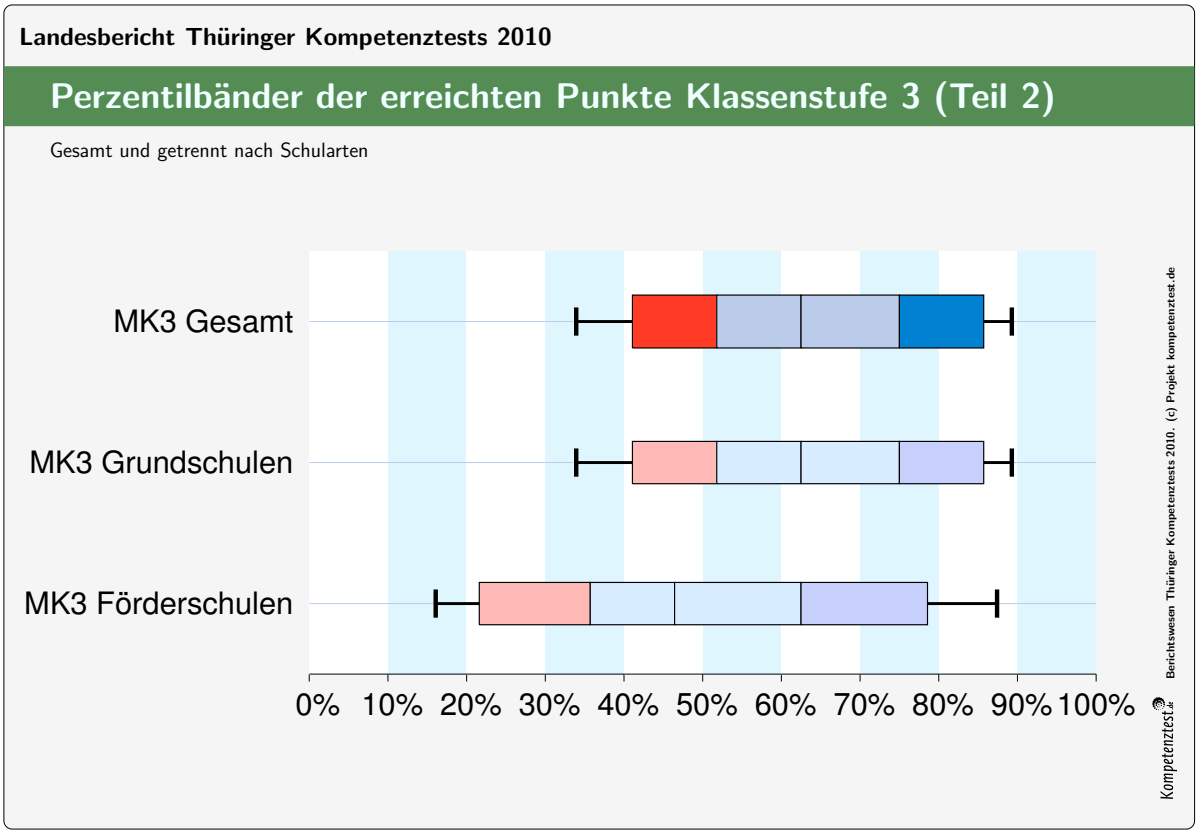


**Abbildung 6.13:** Ergebnisse beim Test Mathematik in der Klassenstufe 3 getrennt nach Schulart. Die Abbildung gibt die erreichte Punktzahl in Prozent wieder.

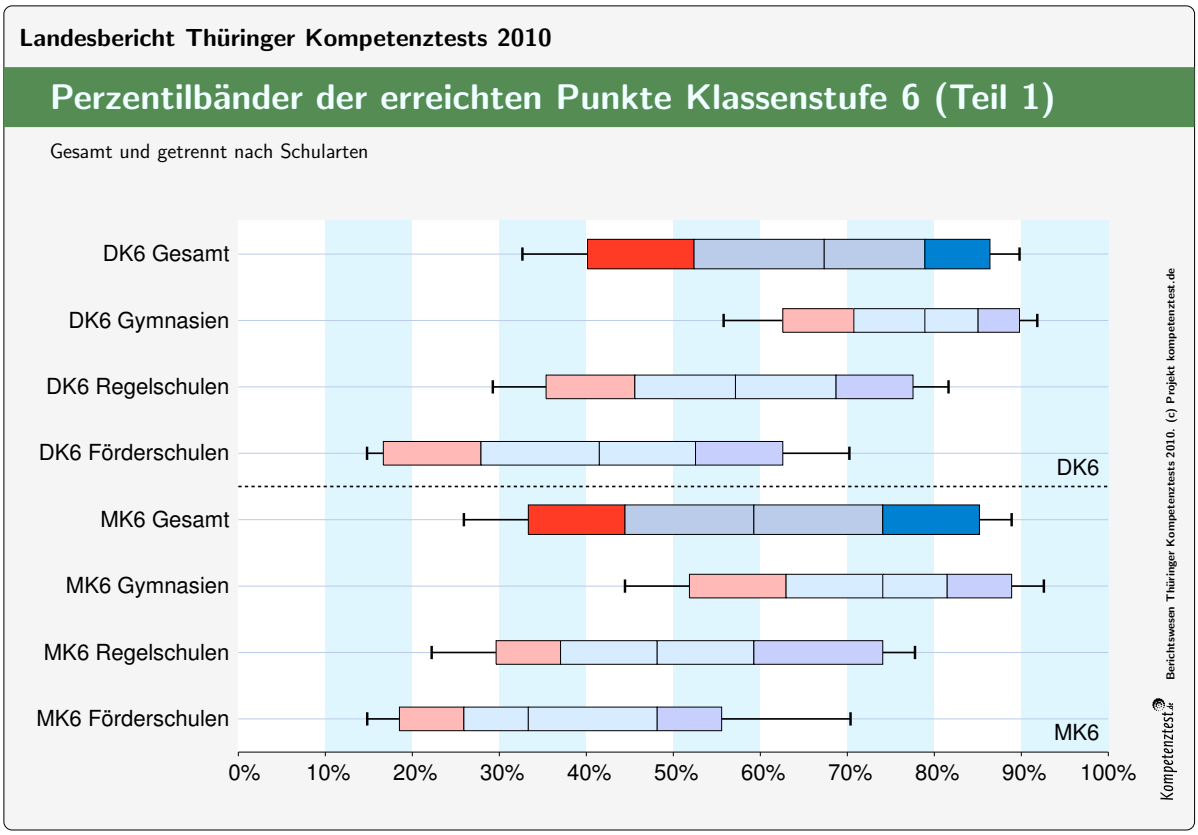


**Abbildung 6.14:** Ergebnisse bei den Tests Deutsch und Mathematik in der Klassenstufe 6 getrennt nach Schulart. Die Abbildung gibt die erreichte Punktzahl in Prozent wieder.

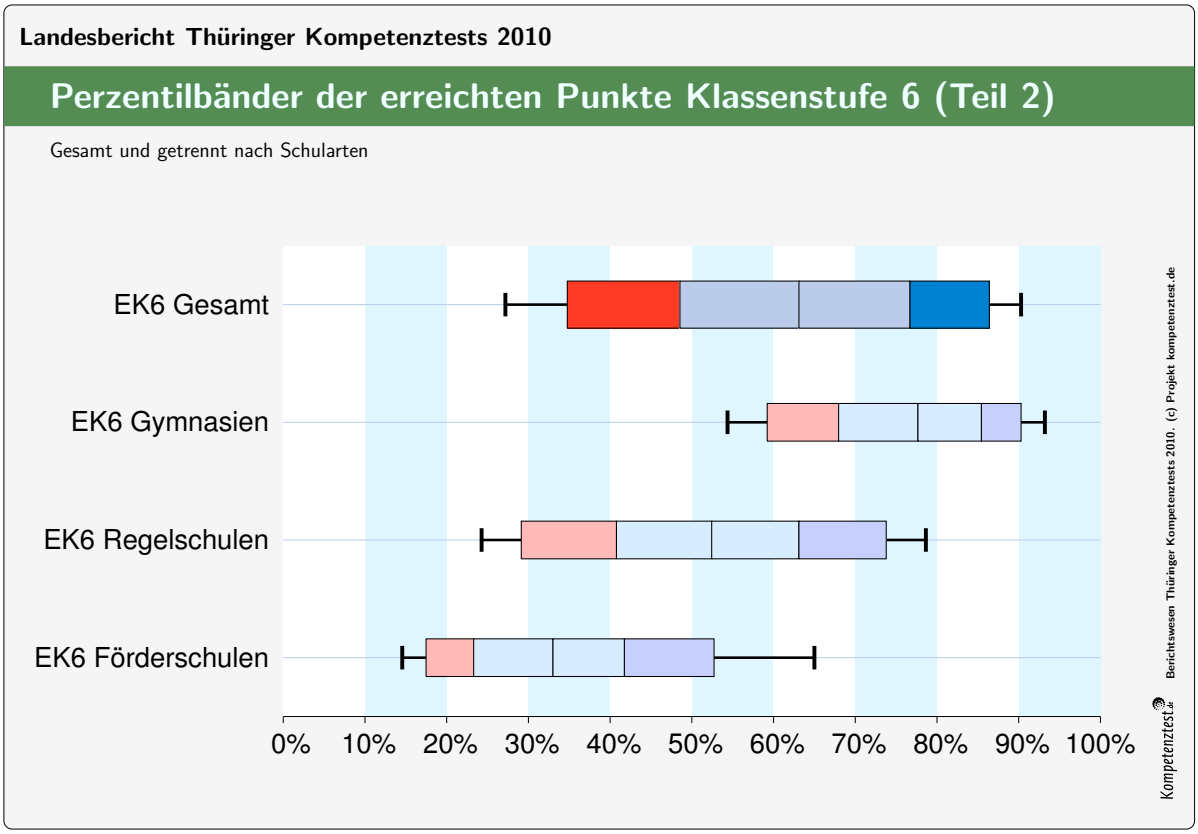


**Abbildung 6.15:** Ergebnisse beim Englisch-Test in der Klassenstufe 6 getrennt nach Schulart. Die Abbildung gibt die erreichte Punktzahl in Prozent wieder.

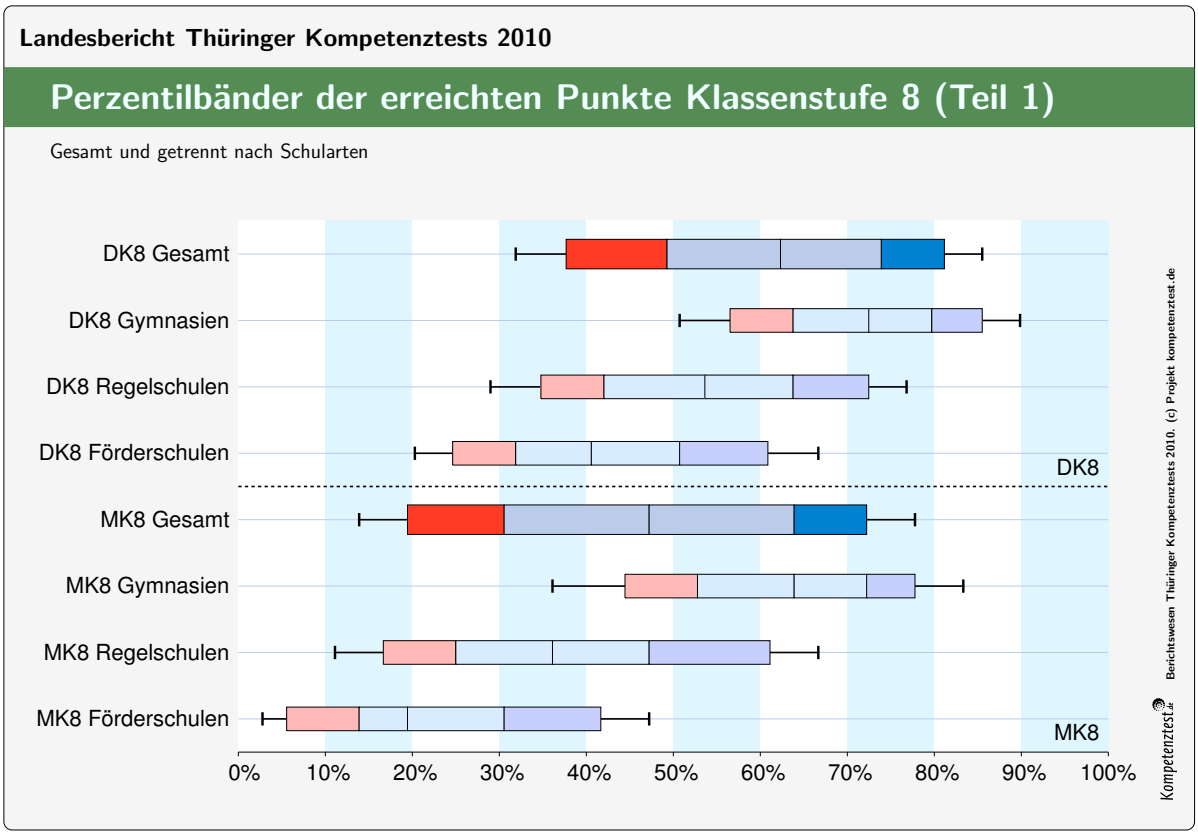


**Abbildung 6.16:** Ergebnisse bei den Tests Deutsch und Mathematik in der Klassenstufe 8 getrennt nach Schulart. Die Abbildung gibt die erreichte Punktzahl in Prozent wieder.

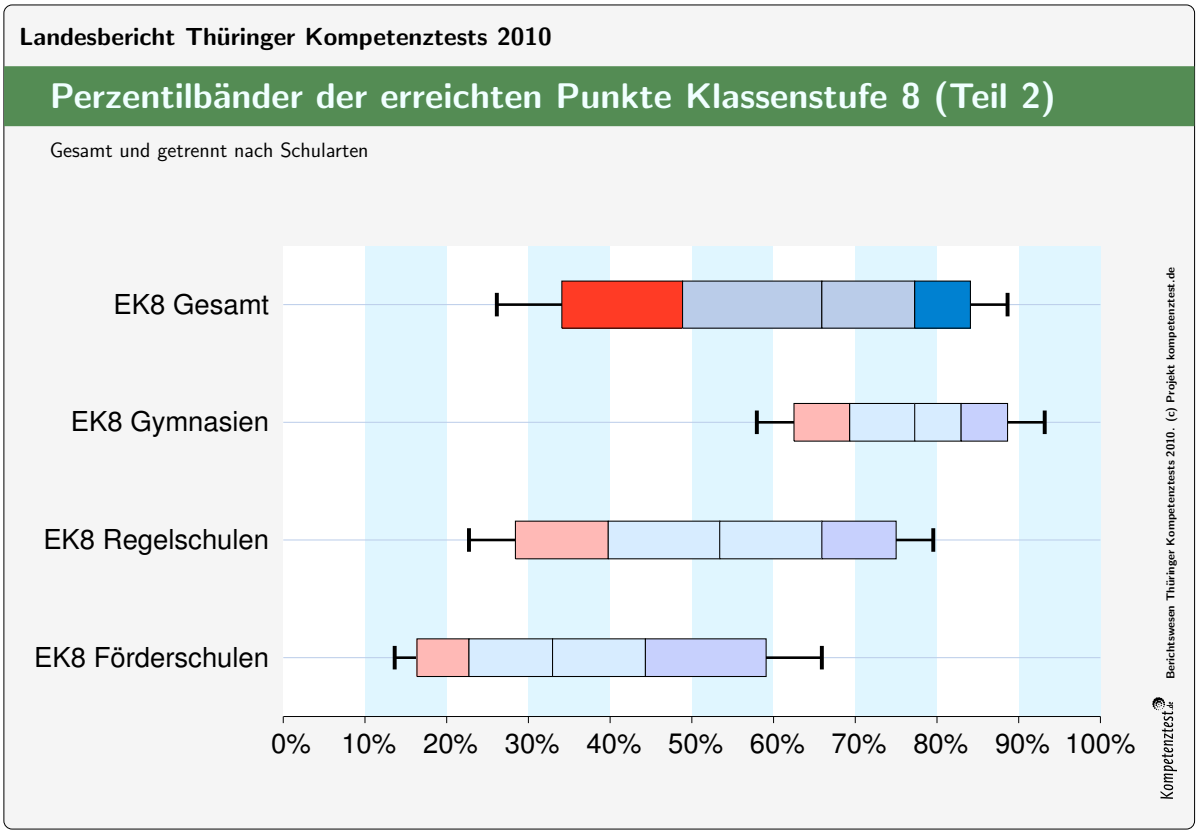


**Abbildung 6.17:** Ergebnisse beim Englisch-Test in der Klassenstufe 8 getrennt nach Schulart. Die Abbildung gibt die erreichte Punktzahl in Prozent wieder.

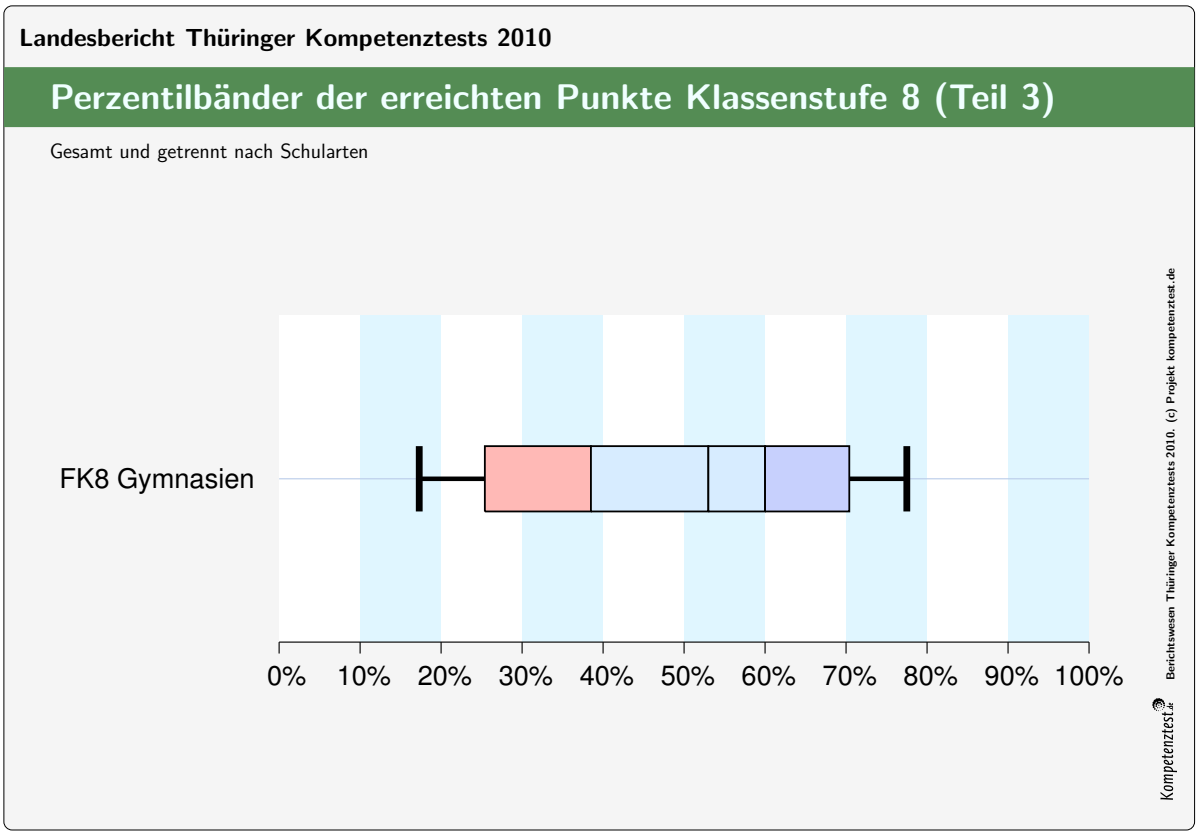


**Abbildung 6.18:** Ergebnisse beim Französisch-Test in der Klassenstufe 8. Die Abbildung gibt die erreichte Punktzahl in Prozent wieder.

## 6.3 Aufgabenschwierigkeiten

Neben der Analyse der Gesamtergebnisse sind die Ergebnisse 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.

In den Kompetenztests Englisch der Klassenstufe 6 und 8, sowie im Französischtest der Klassenstufe 8 wurde ein schriftlicher Teil durch die Lernenden bearbeitet. Die erreichten Punktzahlen lassen sich in diesen Teiltests nicht ohne Weiteres als Schwierigkeiten interpretieren. Aus diesem Grund wurden etwas veränderte Abbildungen (6.32 bis 6.49 ab Seite 61) an das Ende dieses Abschnittes angefügt. Diese geben die landesweiten Ergebnisse im Kompetenzbereich Schreiben wieder. Für die verschiedene Levels des CEFR / GER werden die Unterkriterien des Bereichs Schreiben dargestellt. Der linke Balken zeigt den Landesmittelwert aller teilnehmenden Schüler, welche die entsprechenden Kriterien gelöst haben. Der Mittlere und Rechte Balken gibt diesen Wert schulartspezifisch wieder.

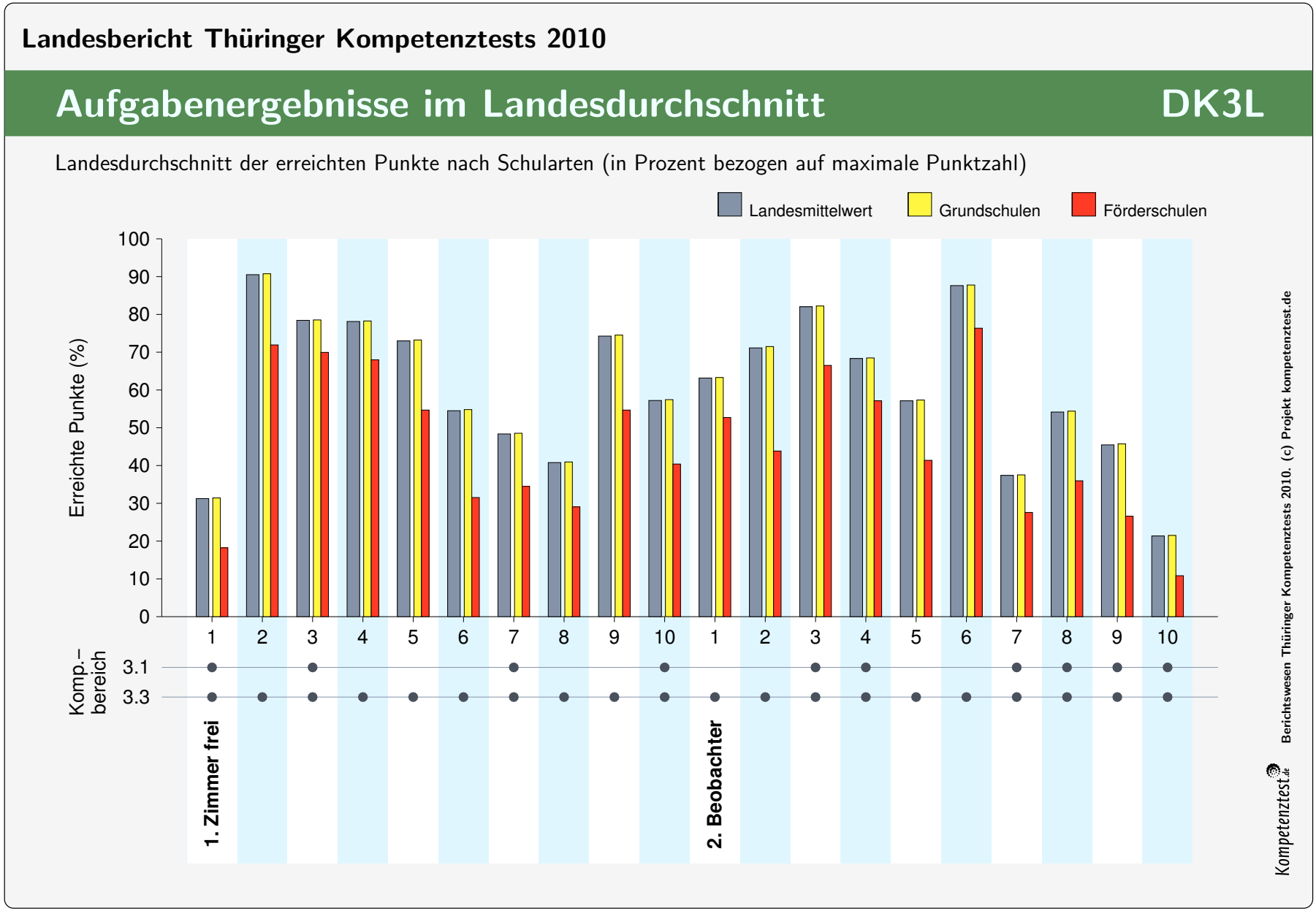


**Abbildung 6.19:** Aufgabenschwierigkeiten im Landesdurchschnitt sowie getrennt nach Schulart (Deutschtest Lesen, Klassenstufe 3, DK3L)

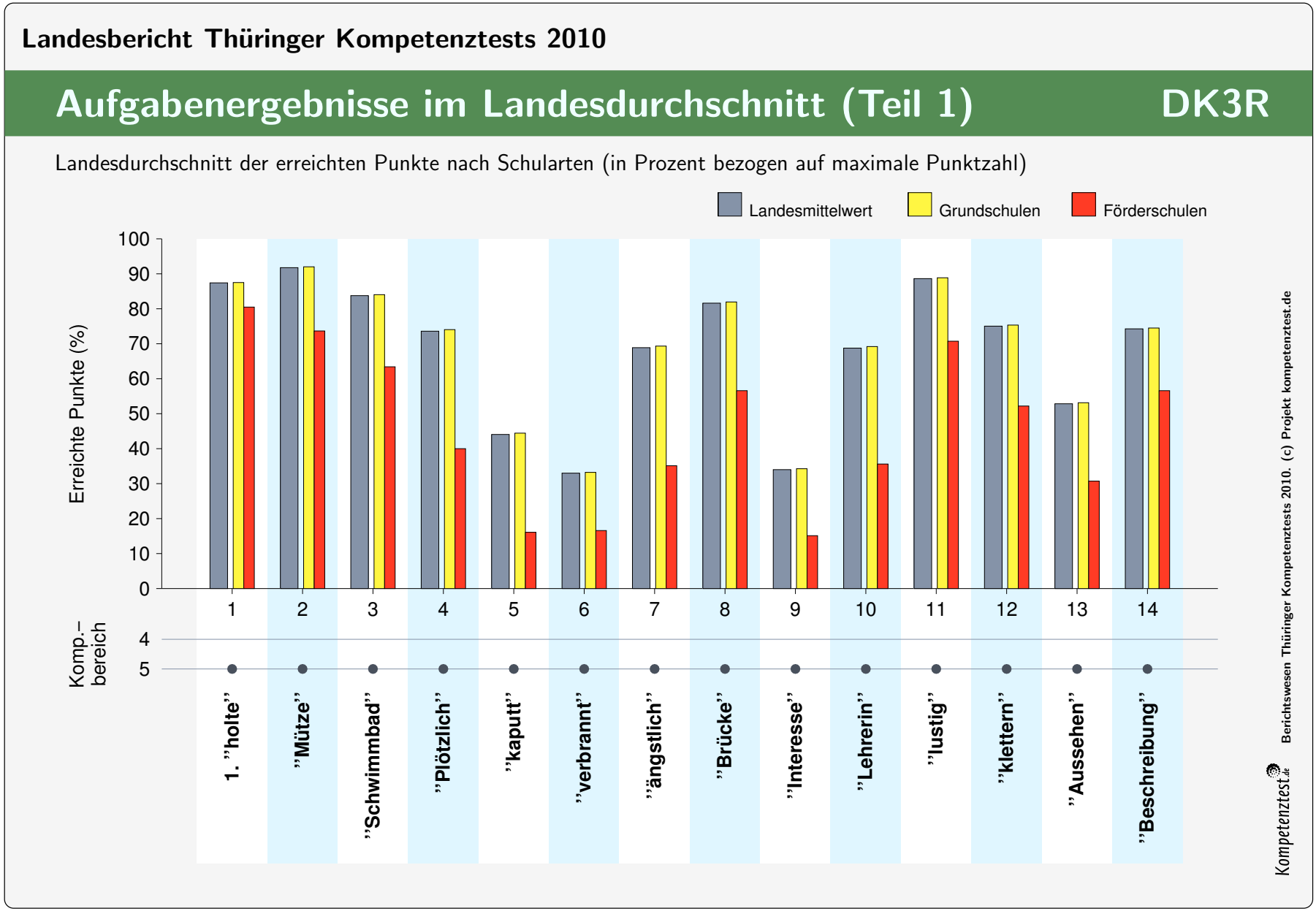


**Abbildung 6.20:** Aufgabenschwierigkeiten im Landesdurchschnitt sowie getrennt nach Schulart (Deutschtest Rechtschreibung, Klassenstufe 3, DK3R – Teil 1)

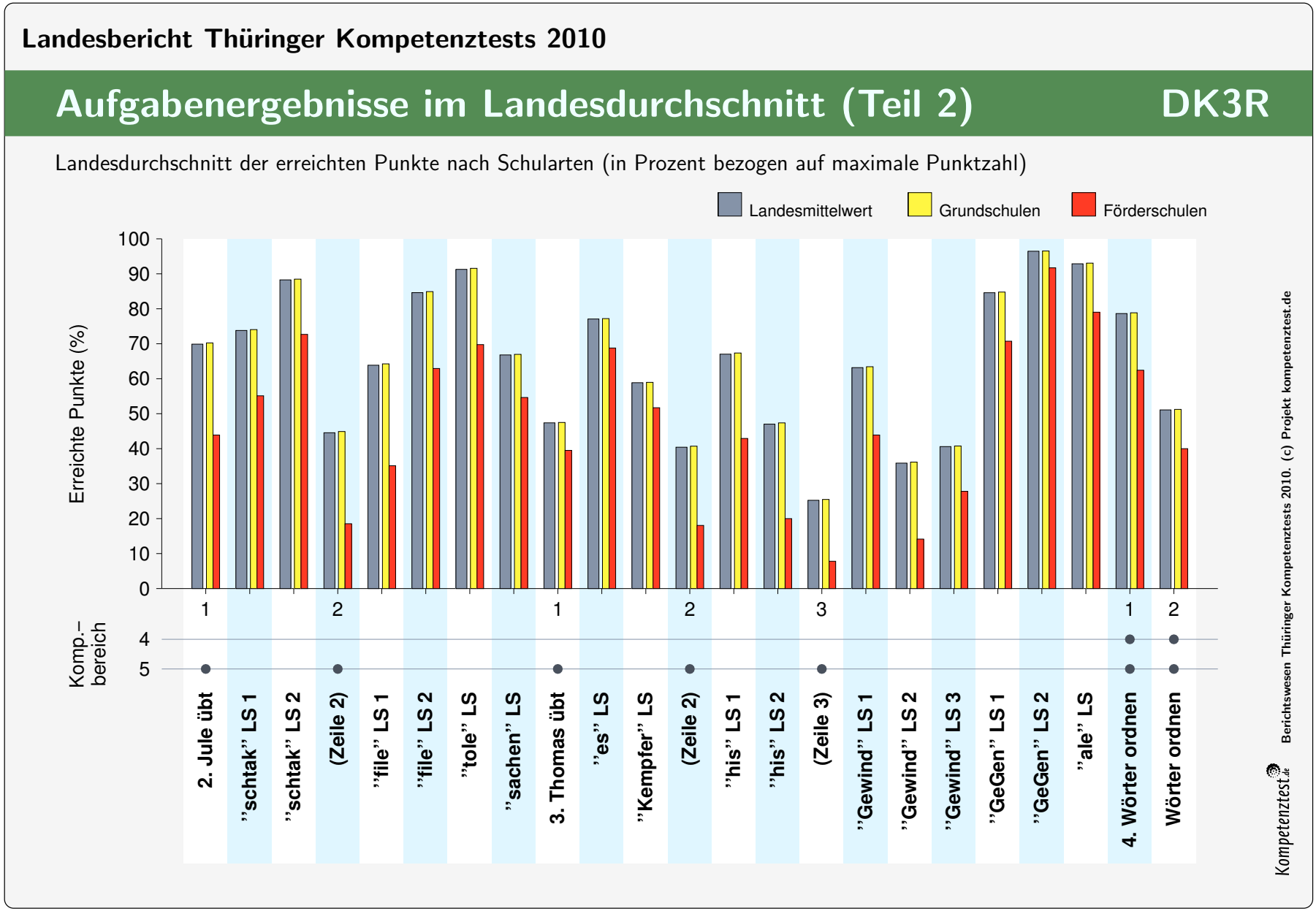


**Abbildung 6.21:** Aufgabenschwierigkeiten im Landesdurchschnitt sowie getrennt nach Schulart (Deutschtest Rechtschreibung, Klassenstufe 3, DK3R – Teil 2)

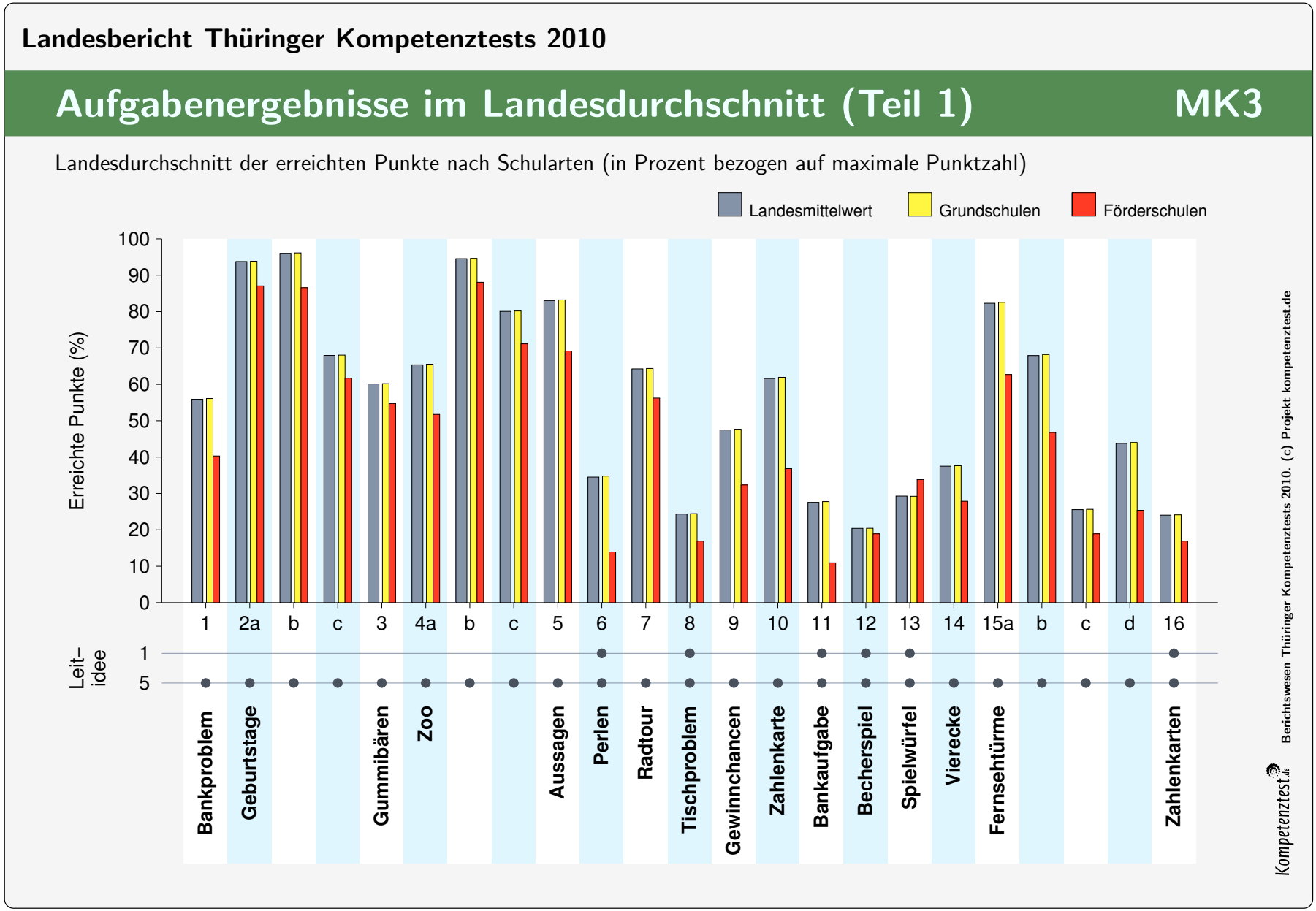


**Abbildung 6.22:** Aufgabenschwierigkeiten im Landesdurchschnitt sowie getrennt nach Schulart (Mathematiktest, Klassenstufe 3, MK3 – Teil 1)
Leitidee 1: Zahlen und Operationen; Leitidee 5: Daten, Häufigkeit und Wahrscheinlichkeit

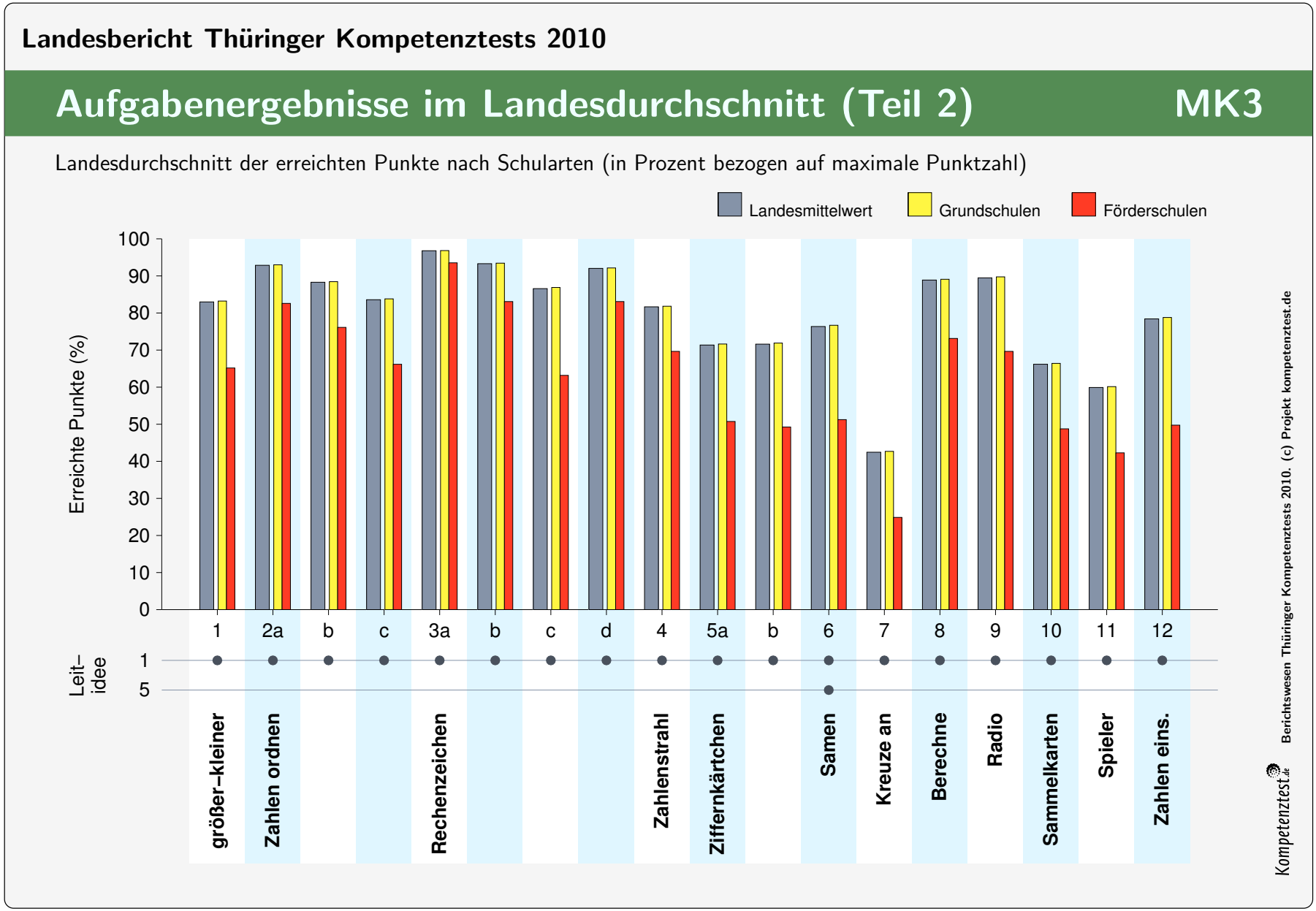


**Abbildung 6.23:** Aufgabenschwierigkeiten im Landesdurchschnitt sowie getrennt nach Schulart (Mathematiktest, Klassenstufe 3, MK3 – Teil 2)
Leitidee 1: Zahlen und Operationen; Leitidee 5: Daten, Häufigkeit und Wahrscheinlichkeit

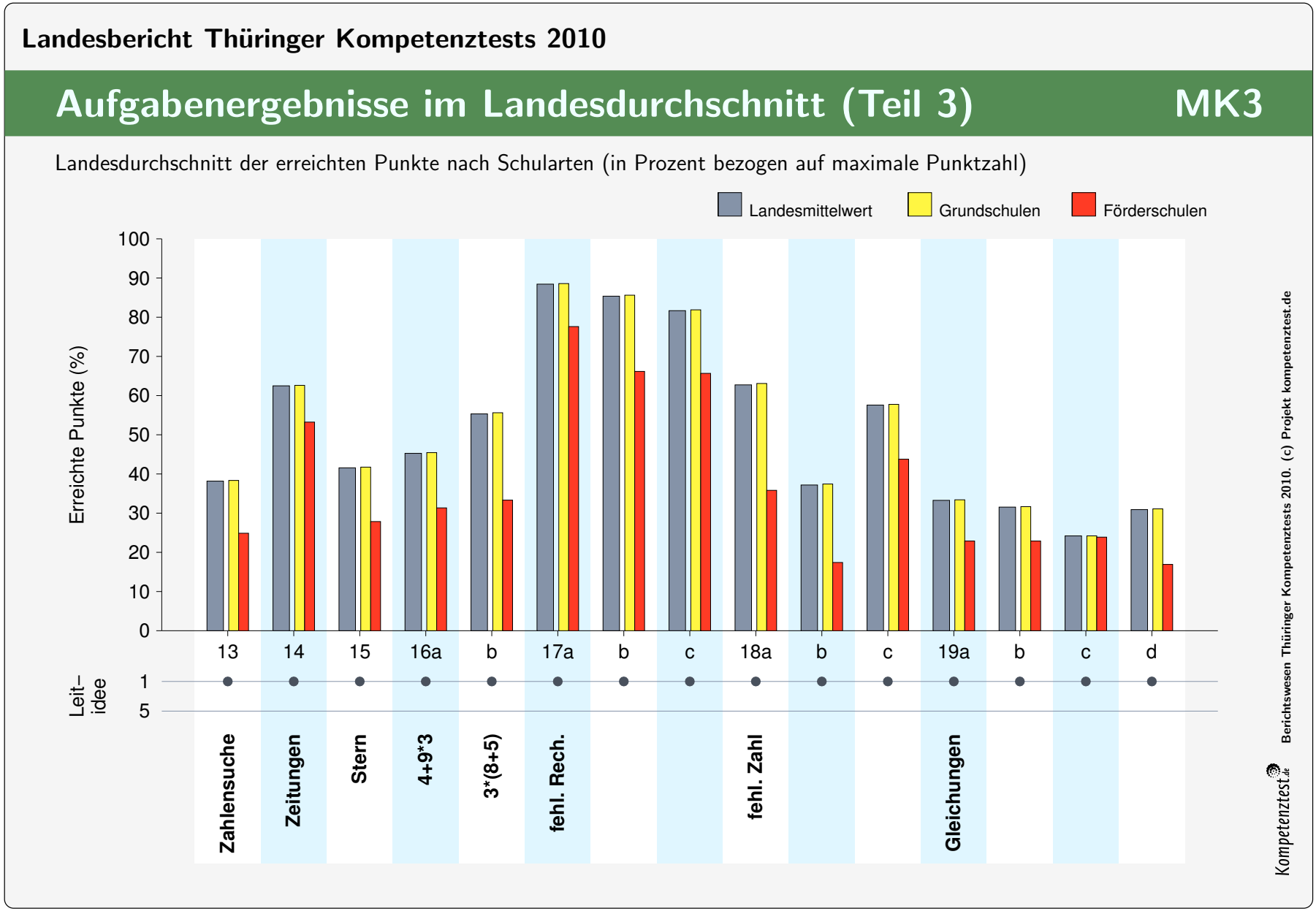


**Abbildung 6.24:** Aufgabenschwierigkeiten im Landesdurchschnitt sowie getrennt nach Schulart (Mathematiktest, Klassenstufe 3, MK3 – Teil 3)
Leitidee 1: Zahlen und Operationen; Leitidee 5: Daten, Häufigkeit und Wahrscheinlichkeit

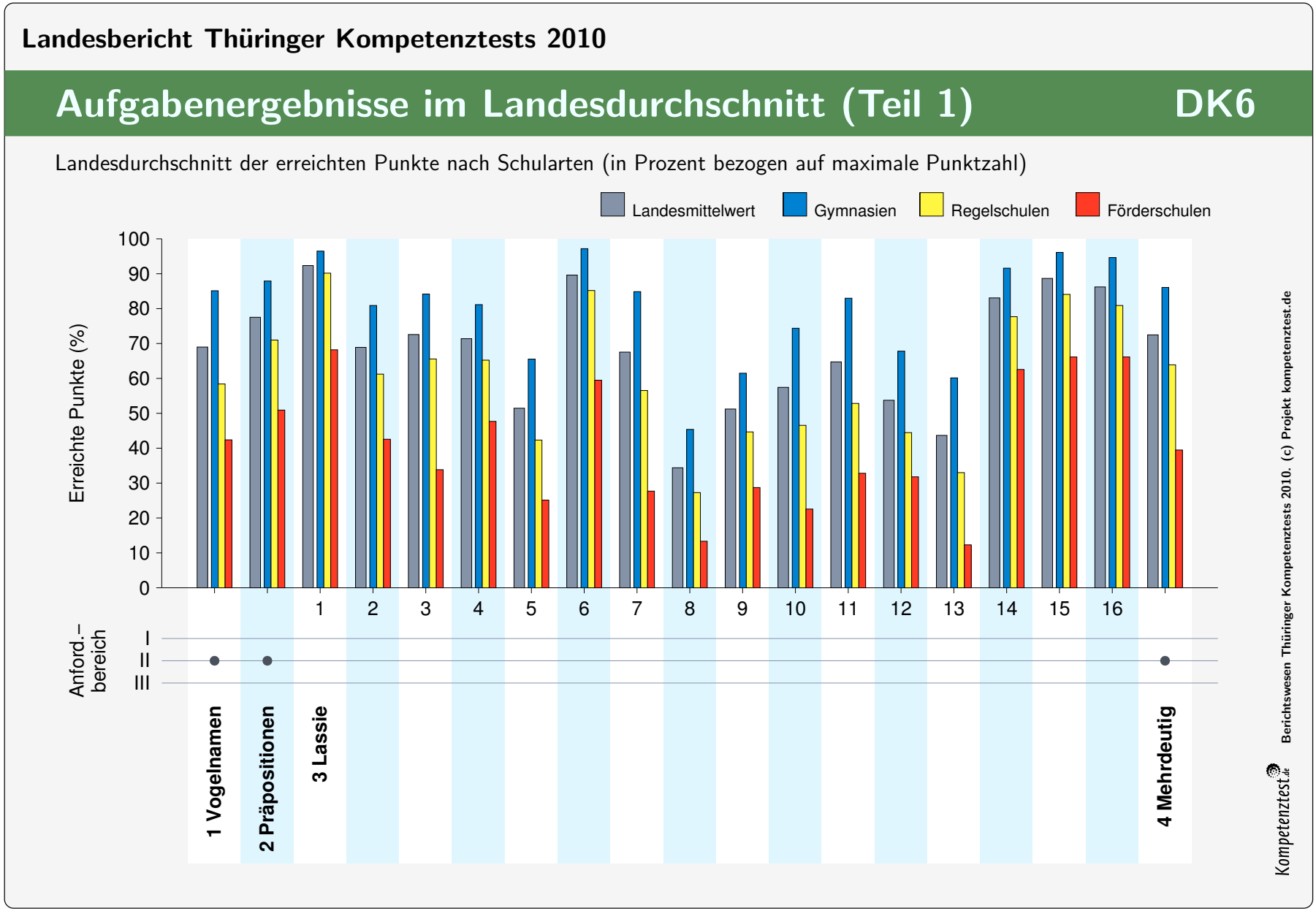


**Abbildung 6.25:** Aufgabenschwierigkeiten im Landesdurchschnitt sowie getrennt nach Schulart (Deutschtest, Klassenstufe 6, DK6 – Teil 1, Sprachgebrauch)

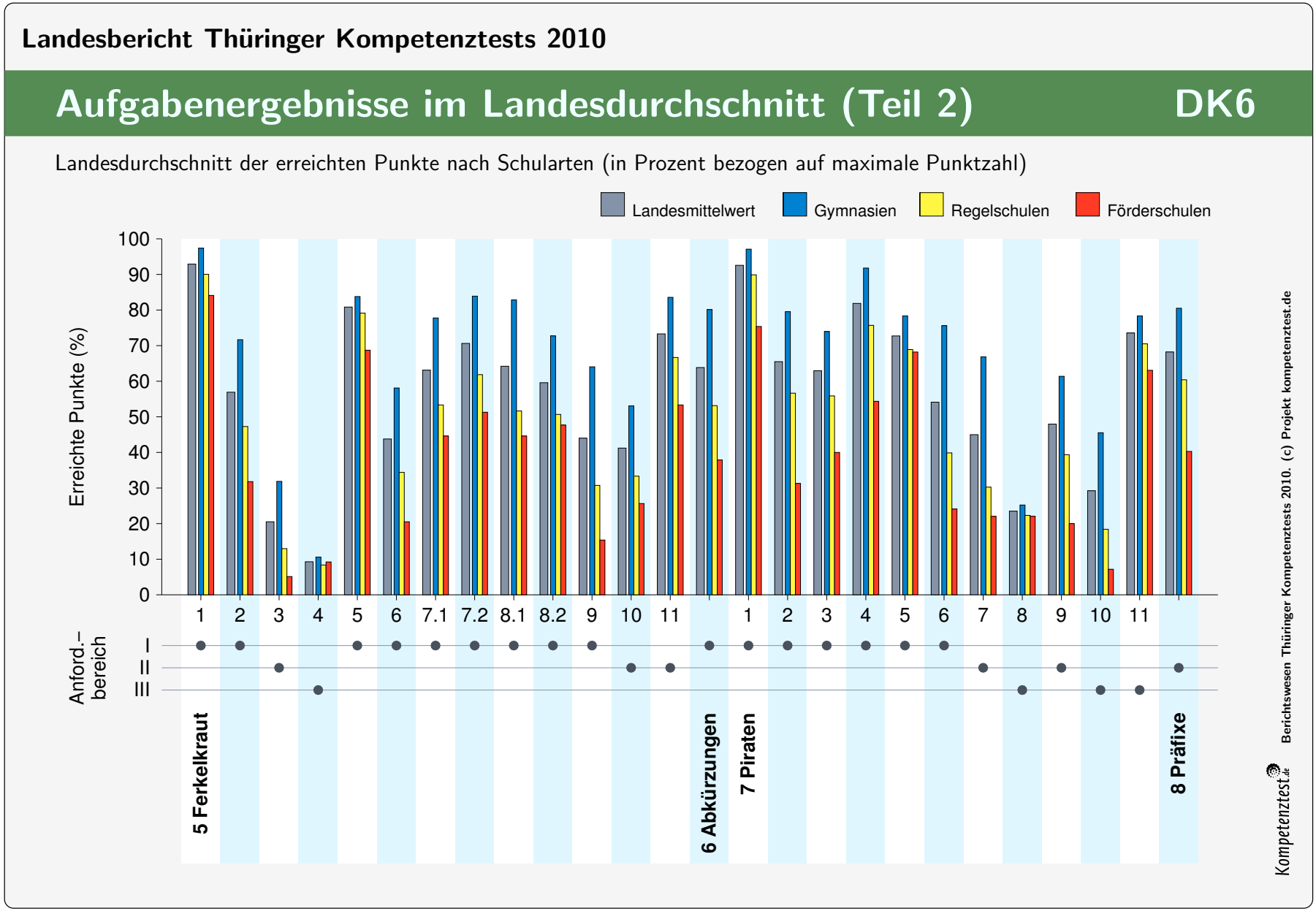


**Abbildung 6.26:** Aufgabenschwierigkeiten im Landesdurchschnitt sowie getrennt nach Schulart (Deutschtest, Klassenstufe 6, DK6 – Teil 2, Leseverstehen)

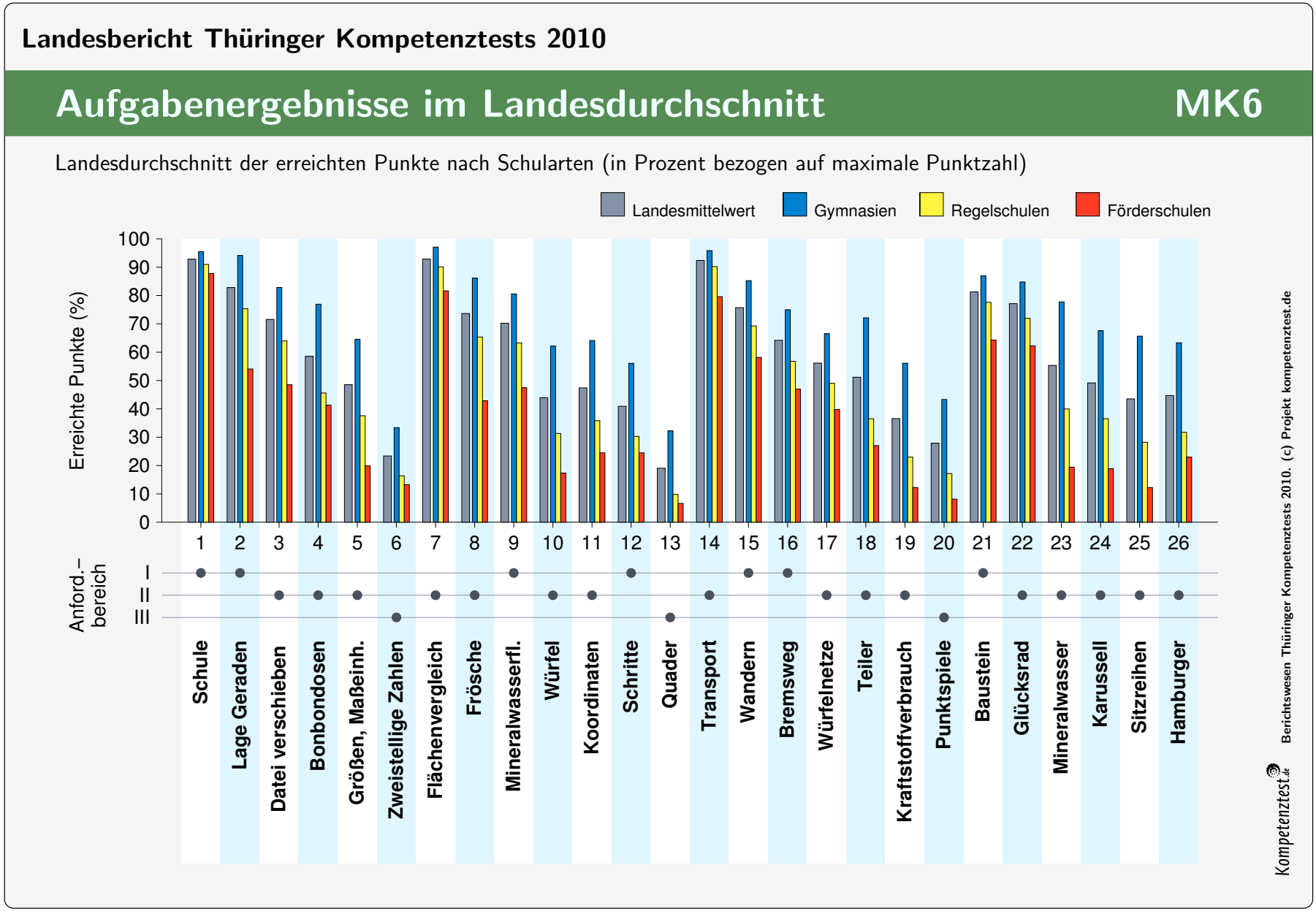


**Abbildung 6.27:** Aufgabenschwierigkeiten im Landesdurchschnitt sowie getrennt nach Schulart (Mathematiktest, Klassenstufe 6, MK6)

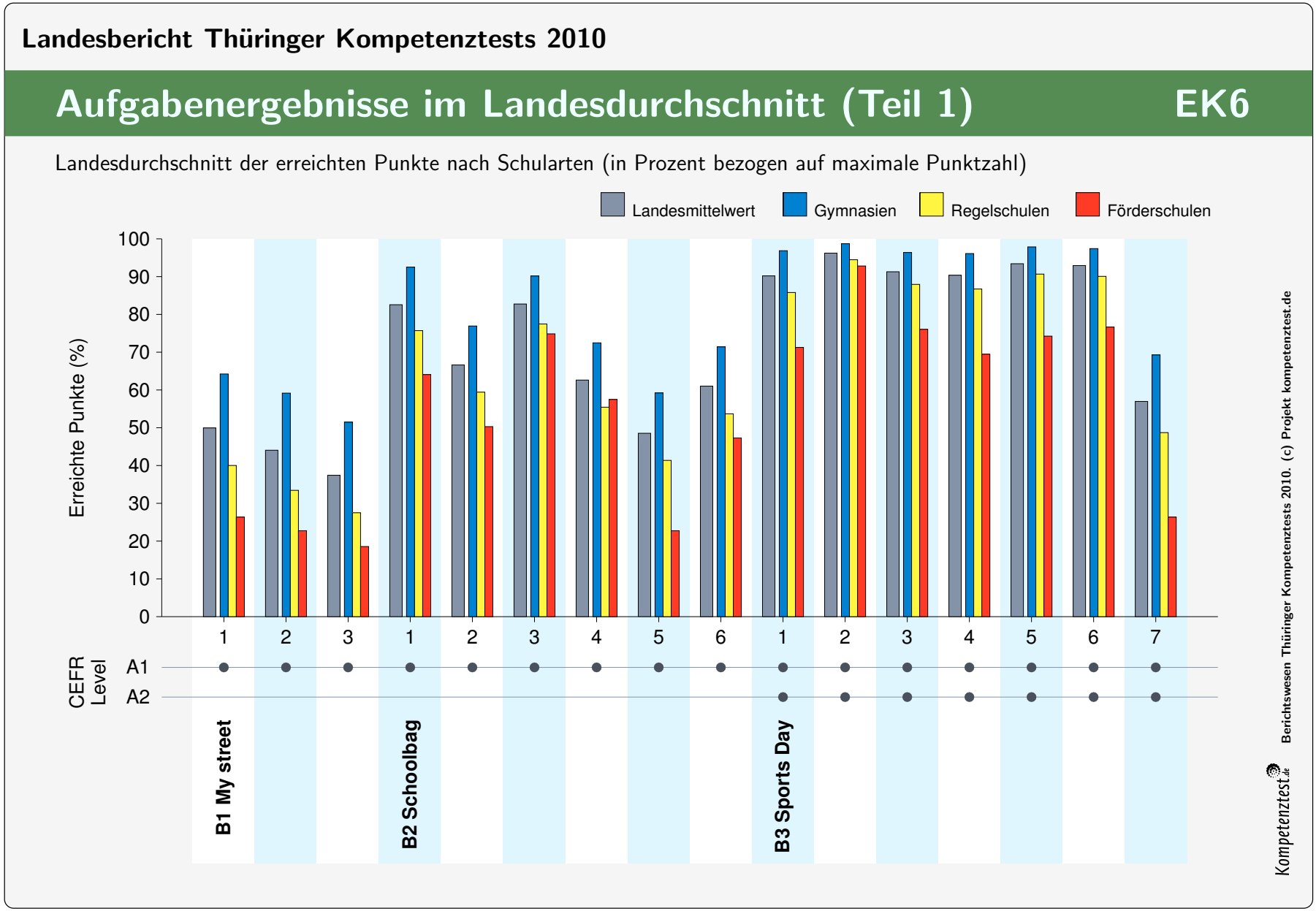


**Abbildung 6.28:** Aufgabenschwierigkeiten im Landesdurchschnitt sowie getrennt nach Schulart (Englischtest Hörverstehen, Klassenstufe 6, EK6 – Teil 1)

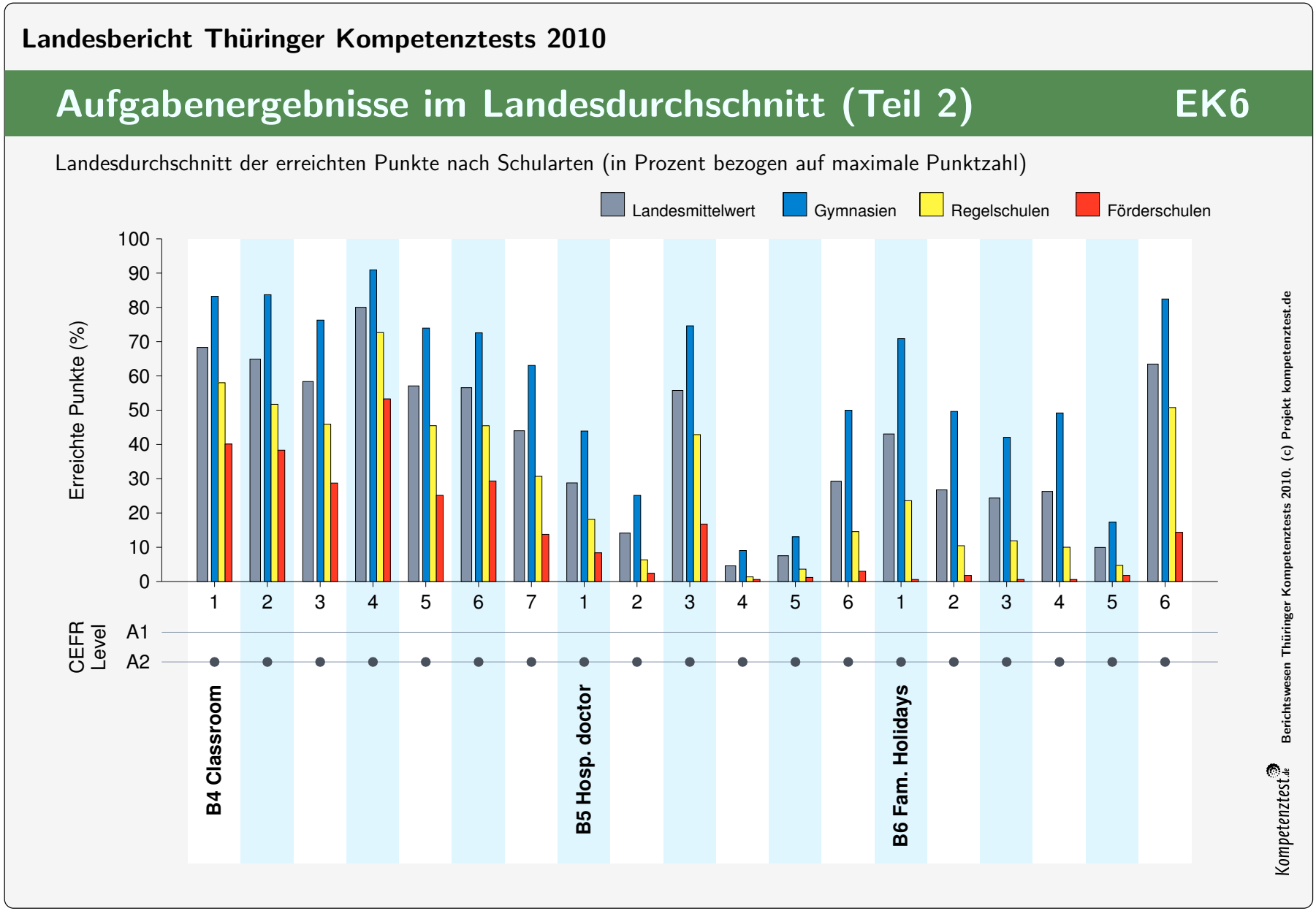


**Abbildung 6.29:** Aufgabenschwierigkeiten im Landesdurchschnitt sowie getrennt nach Schulart (Englischtest Hörverstehen, Klassenstufe 6, EK6 – Teil 2)

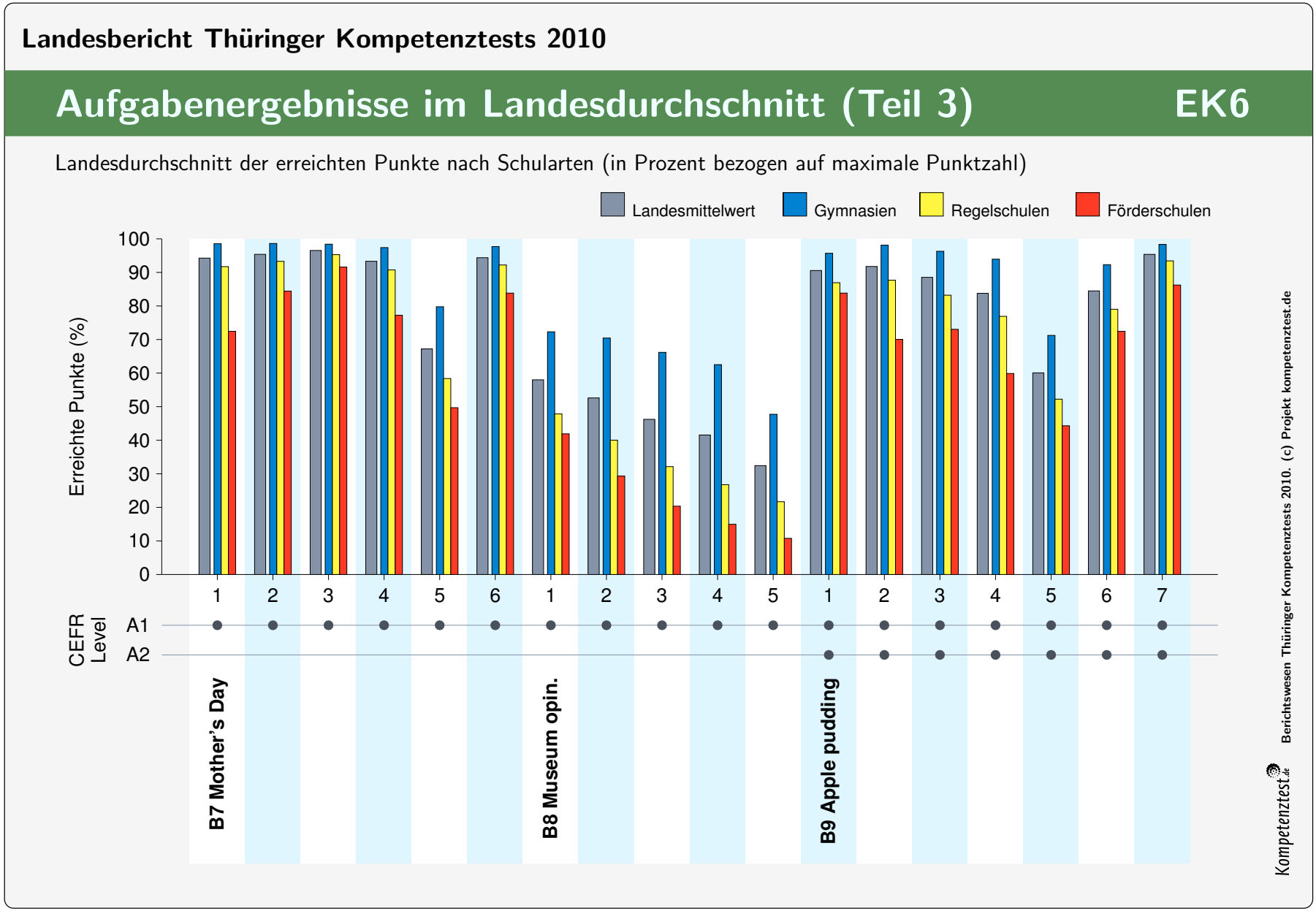


**Abbildung 6.30:** Aufgabenschwierigkeiten im Landesdurchschnitt sowie getrennt nach Schulart (Englischtest Leseverstehen, Klassenstufe 6, EK6 – Teil 3)

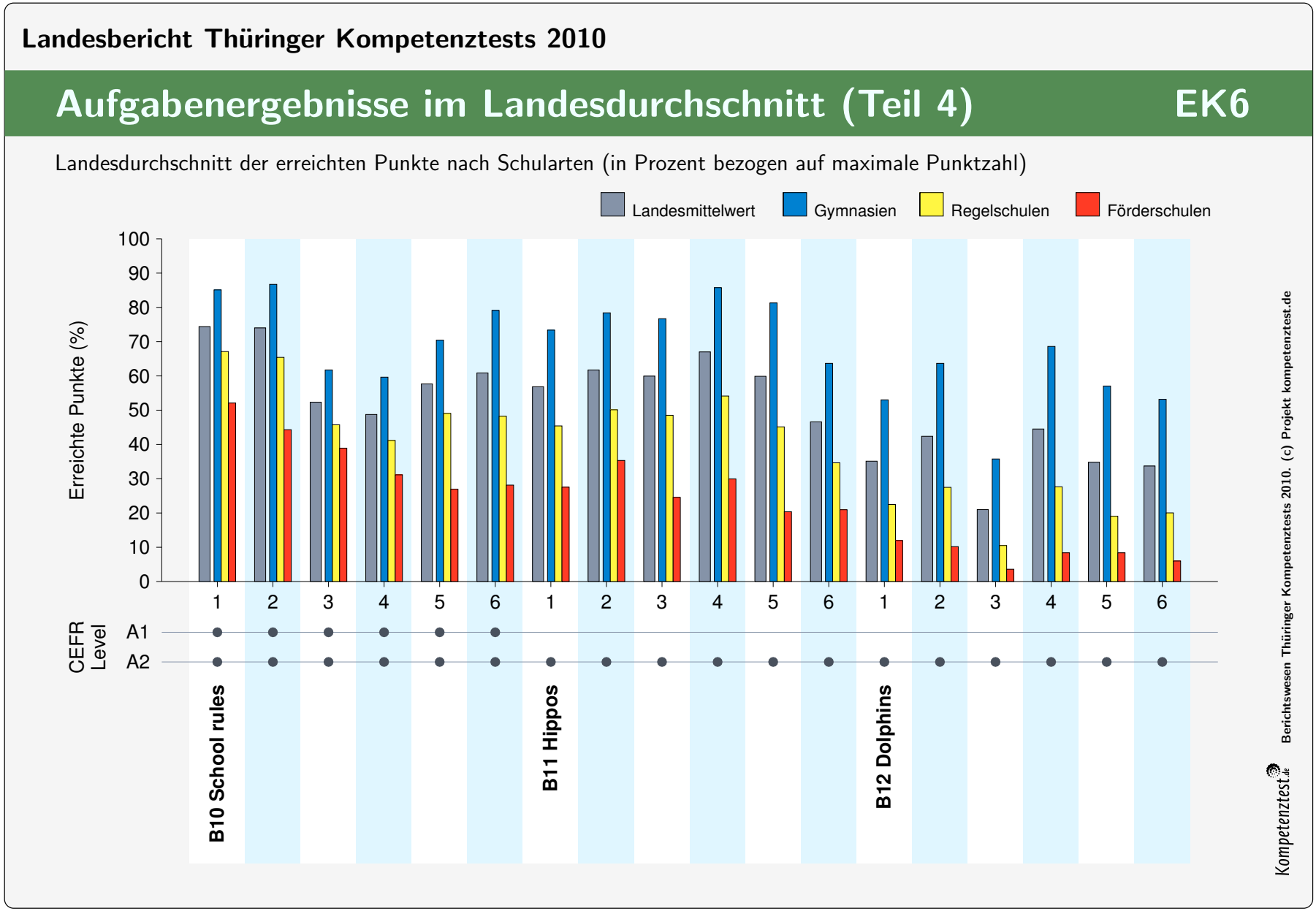


**Abbildung 6.31:** Aufgabenschwierigkeiten im Landesdurchschnitt sowie getrennt nach Schulart (Englischtest Leseverstehen, Klassenstufe 6, EK6 – Teil 4)

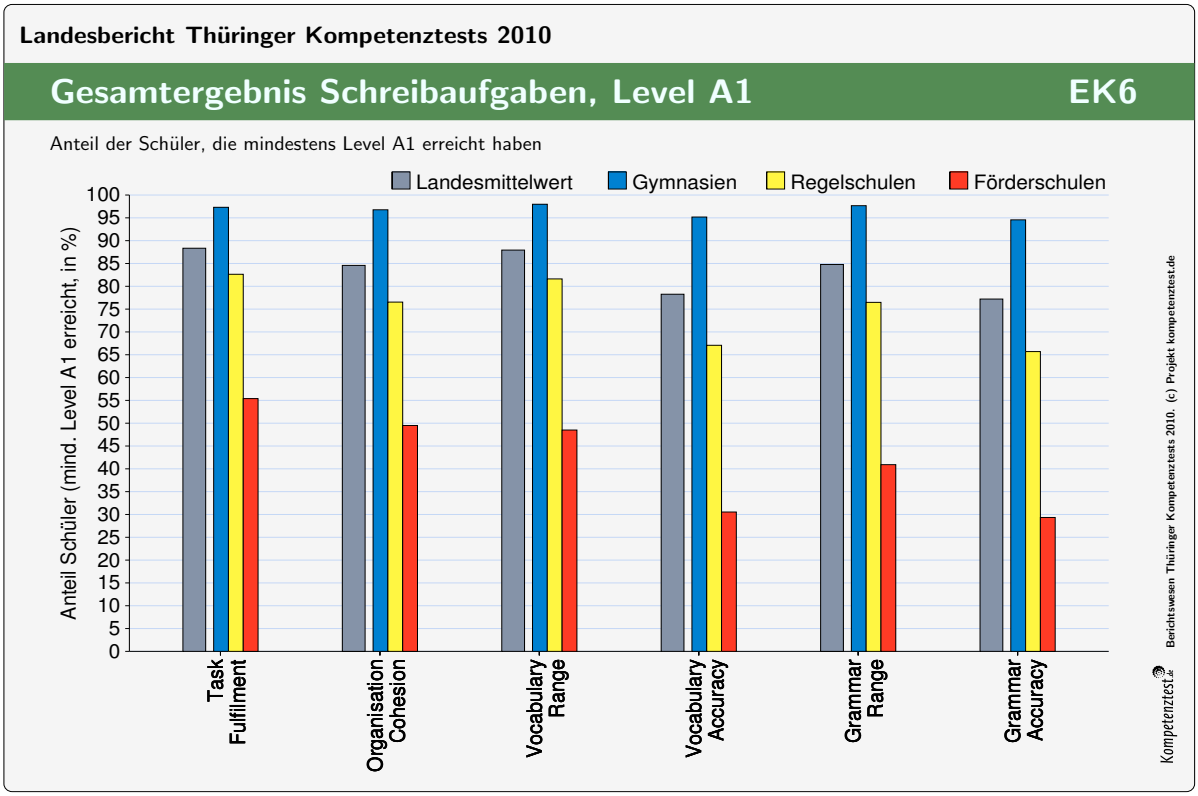


**Abbildung 6.32:** Ergebnisse der Schreibkriterien der Aufgaben 'All about me', 'Summer Holiday' und 'Room' (EK6, Level A1)

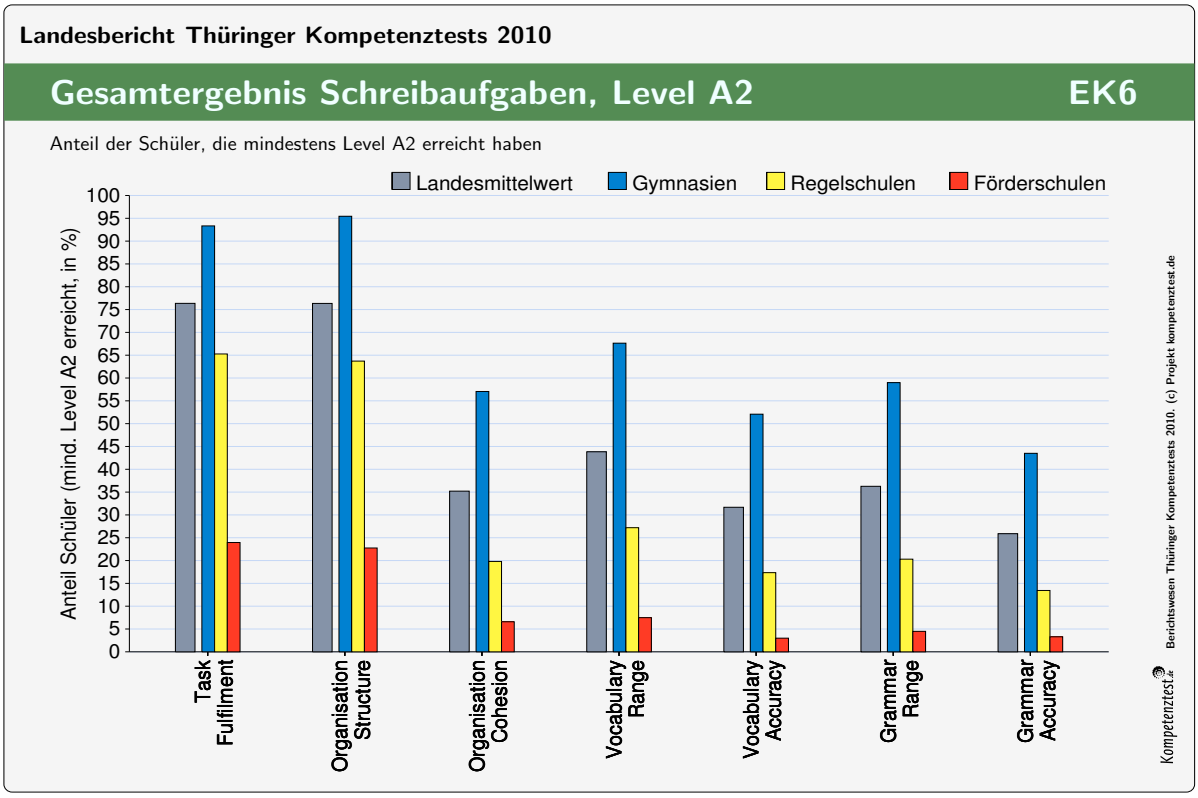


**Abbildung 6.33:** Ergebnisse der Schreibkriterien der Aufgaben 'Summer Holiday' und 'Room' (EK6, Level A2)

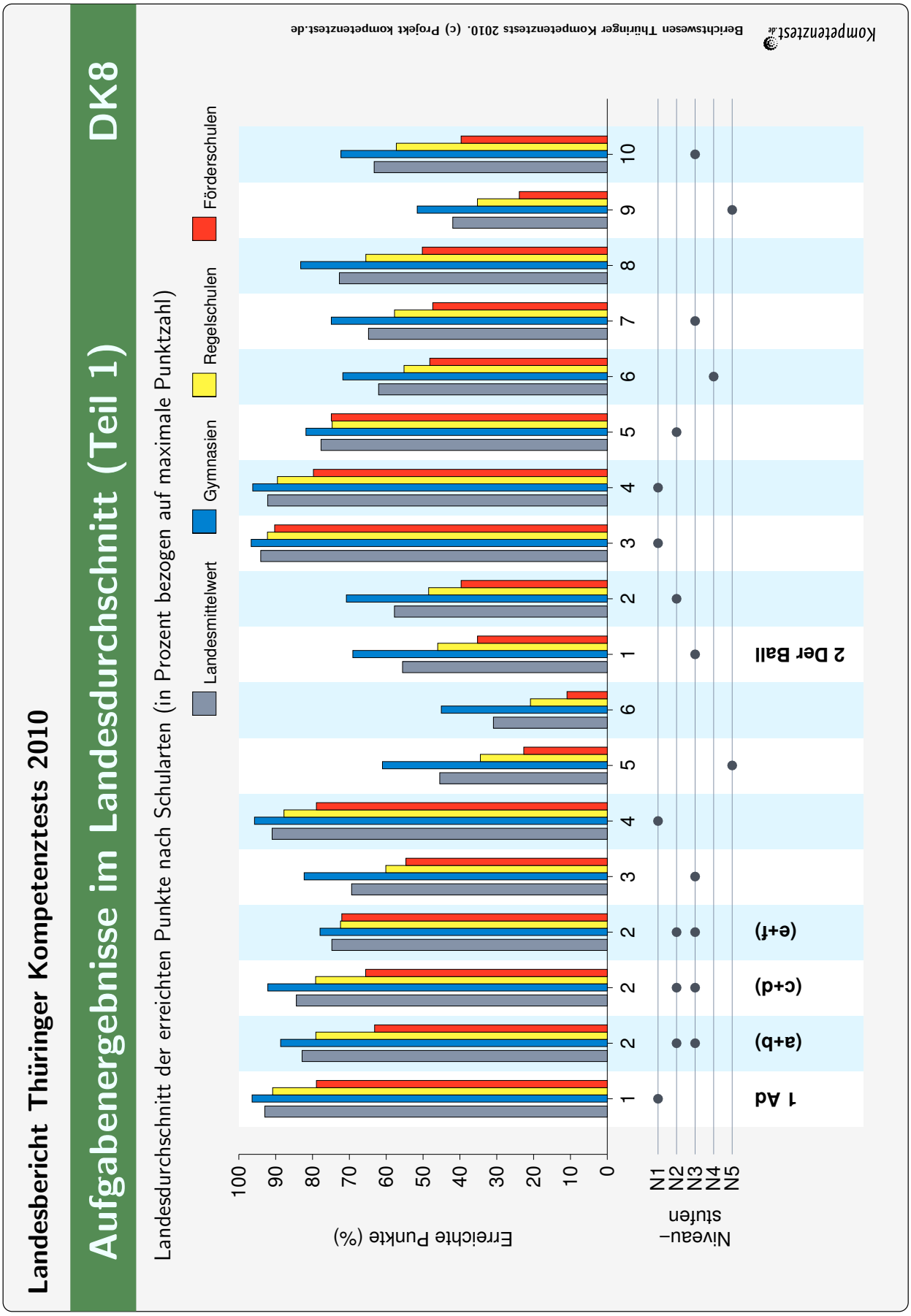


**Abbildung 6.34:** Aufgabenschwierigkeiten im Landesdurchschnitt sowie getrennt nach Schulart (Deutschtest Hörverstehen, Klassenstufe 8, DK8 – Teil 1)

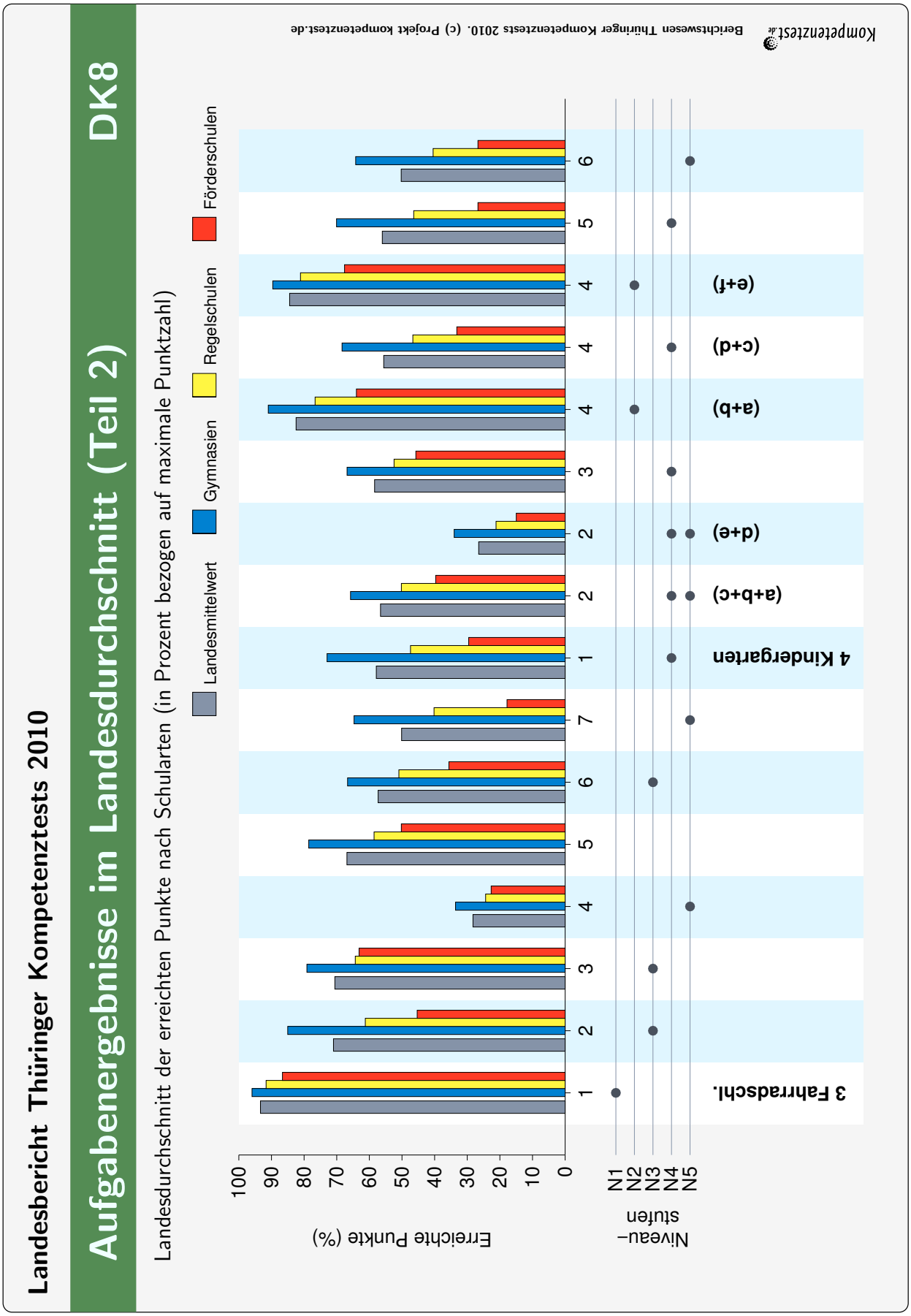


**Abbildung 6.35:** Aufgabenschwierigkeiten im Landesdurchschnitt sowie getrennt nach Schulart (Deutschtest Hörverstehen, Klassenstufe 8, DK8 – Teil 2)

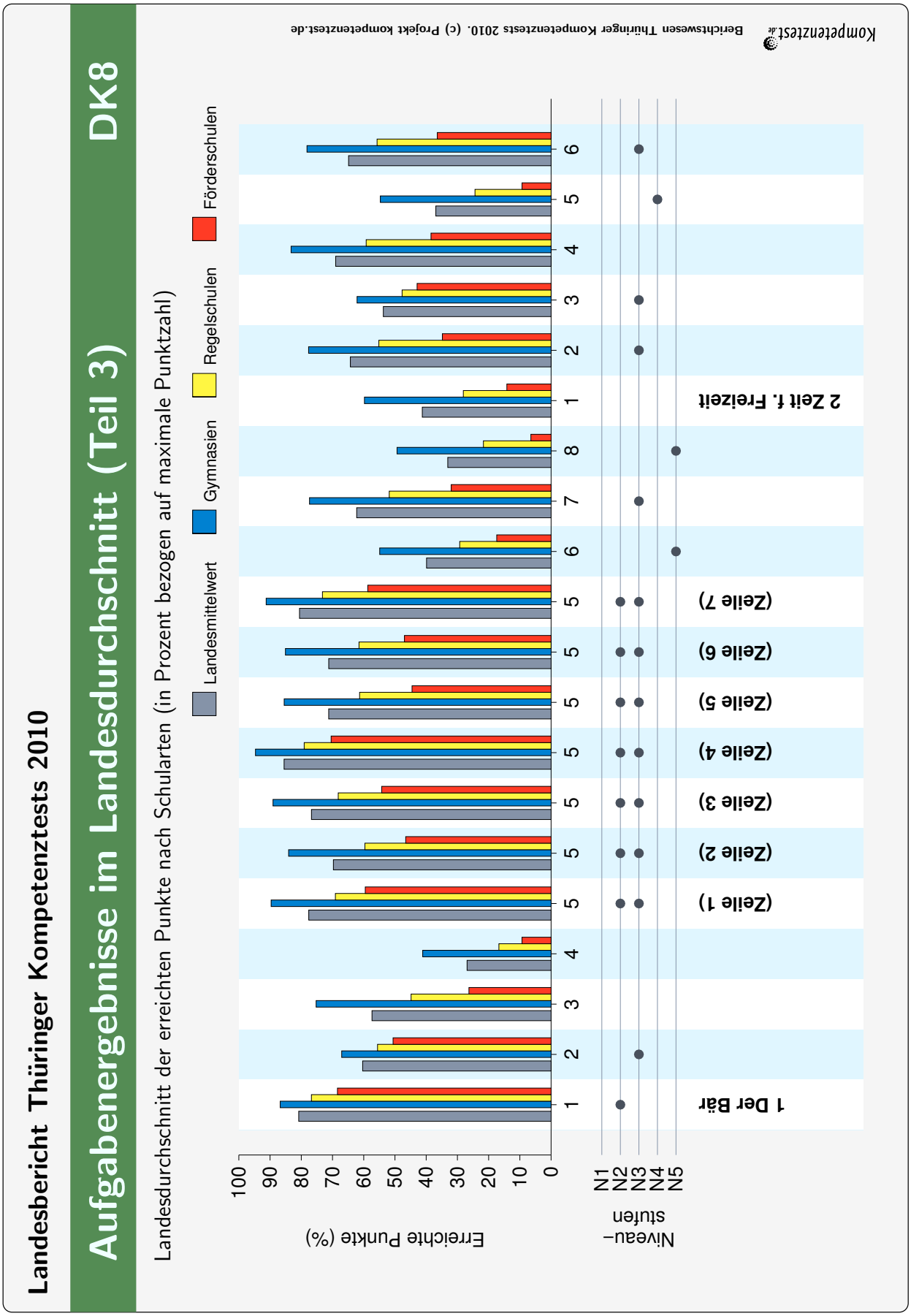


**Abbildung 6.36:** Aufgabenschwierigkeiten im Landesdurchschnitt sowie getrennt nach Schulart (Deutschtest Leseverstehen, Klassenstufe 8, DK8 – Teil 3)

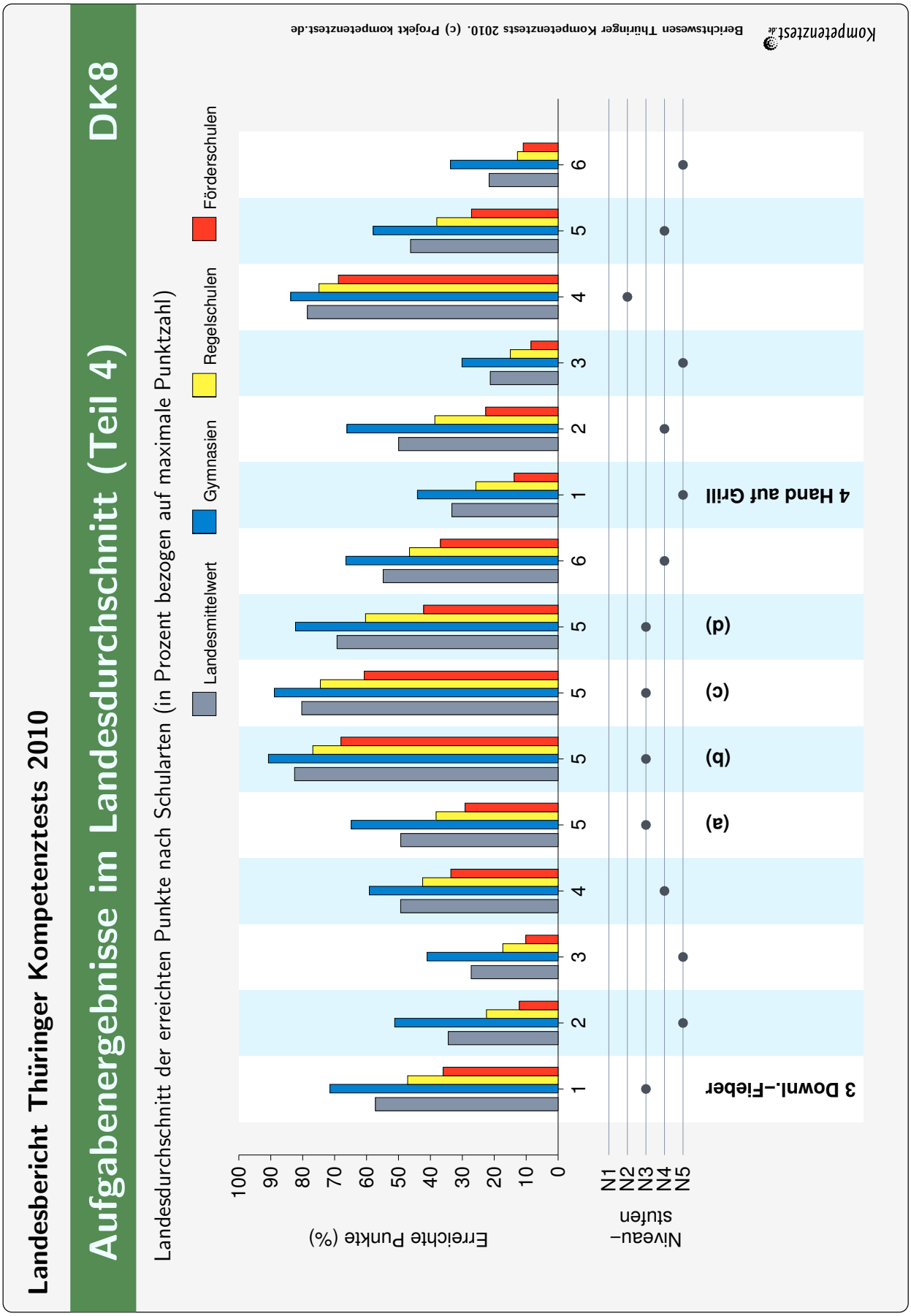


**Abbildung 6.37:** Aufgabenschwierigkeiten im Landesdurchschnitt sowie getrennt nach Schulart (Deutschtest Leseverstehen, Klassenstufe 8, DK8 – Teil 4)

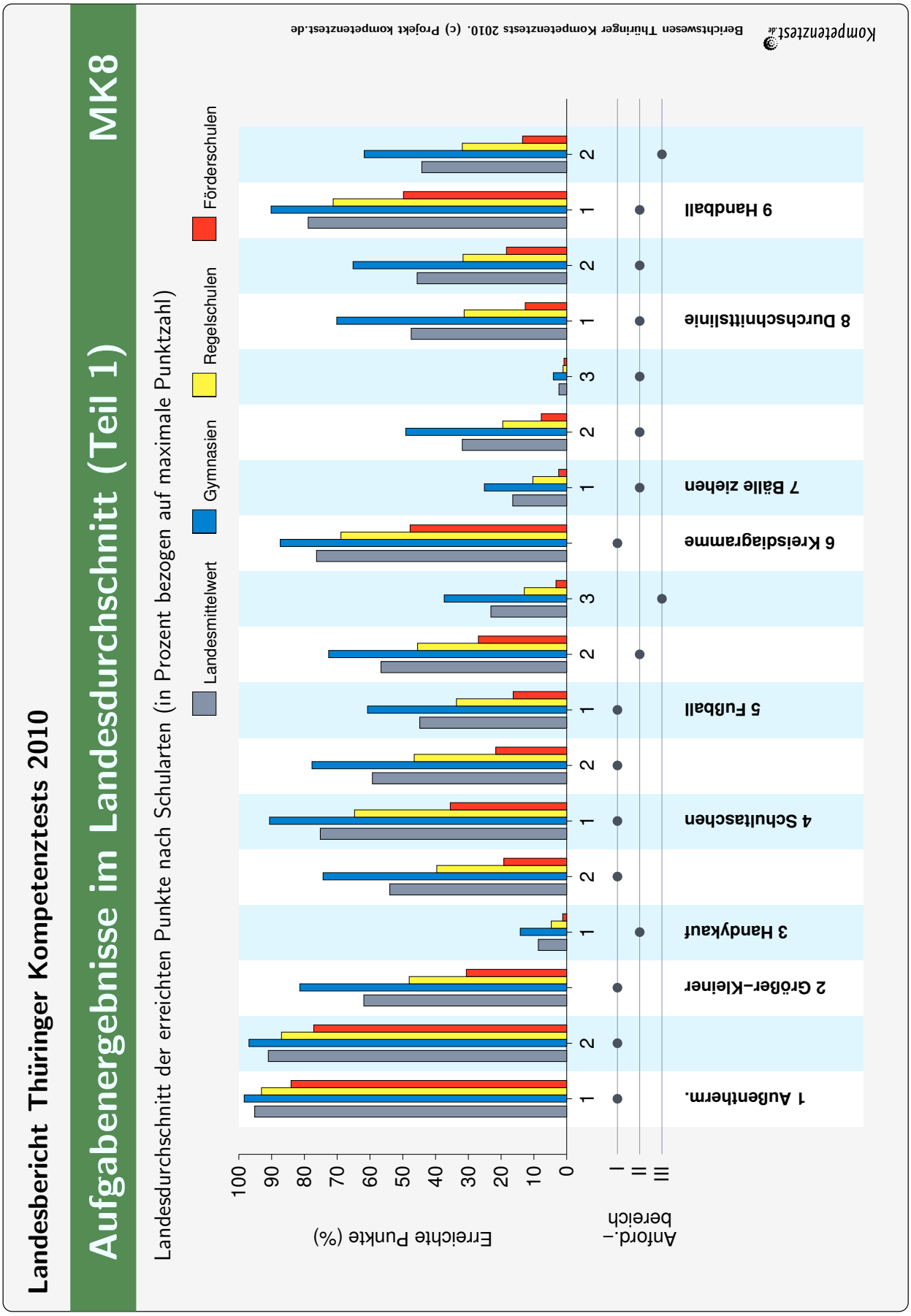


**Abbildung 6.38:** Aufgabenschwierigkeiten im Landesdurchschnitt sowie getrennt nach Schulart (Mathematiktest, Klassenstufe 8, MK8 – Teil 1)

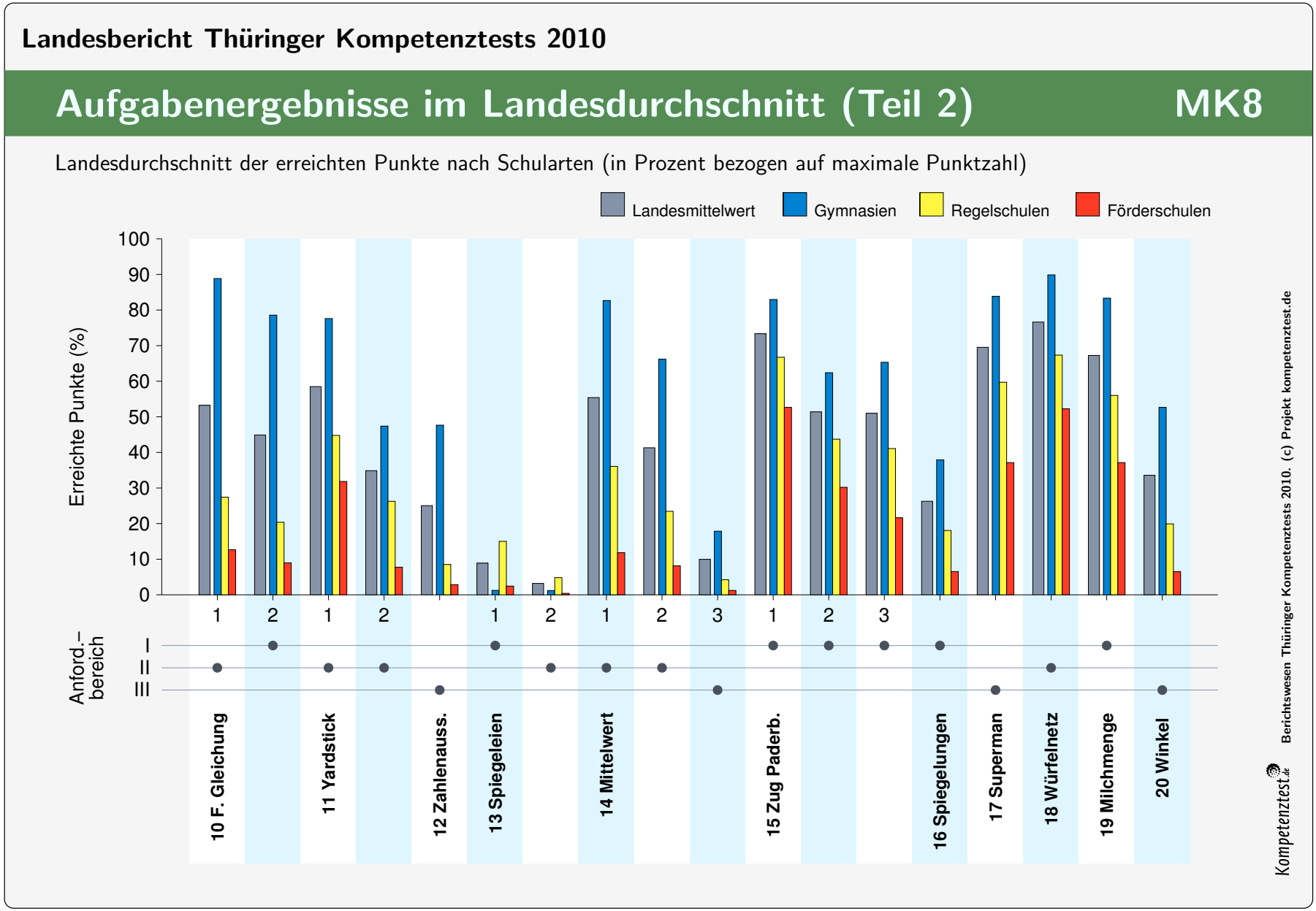


**Abbildung 6.39:** Aufgabenschwierigkeiten im Landesdurchschnitt sowie getrennt nach Schulart (Mathematiktest, Klassenstufe 8, MK8 – Teil 2)

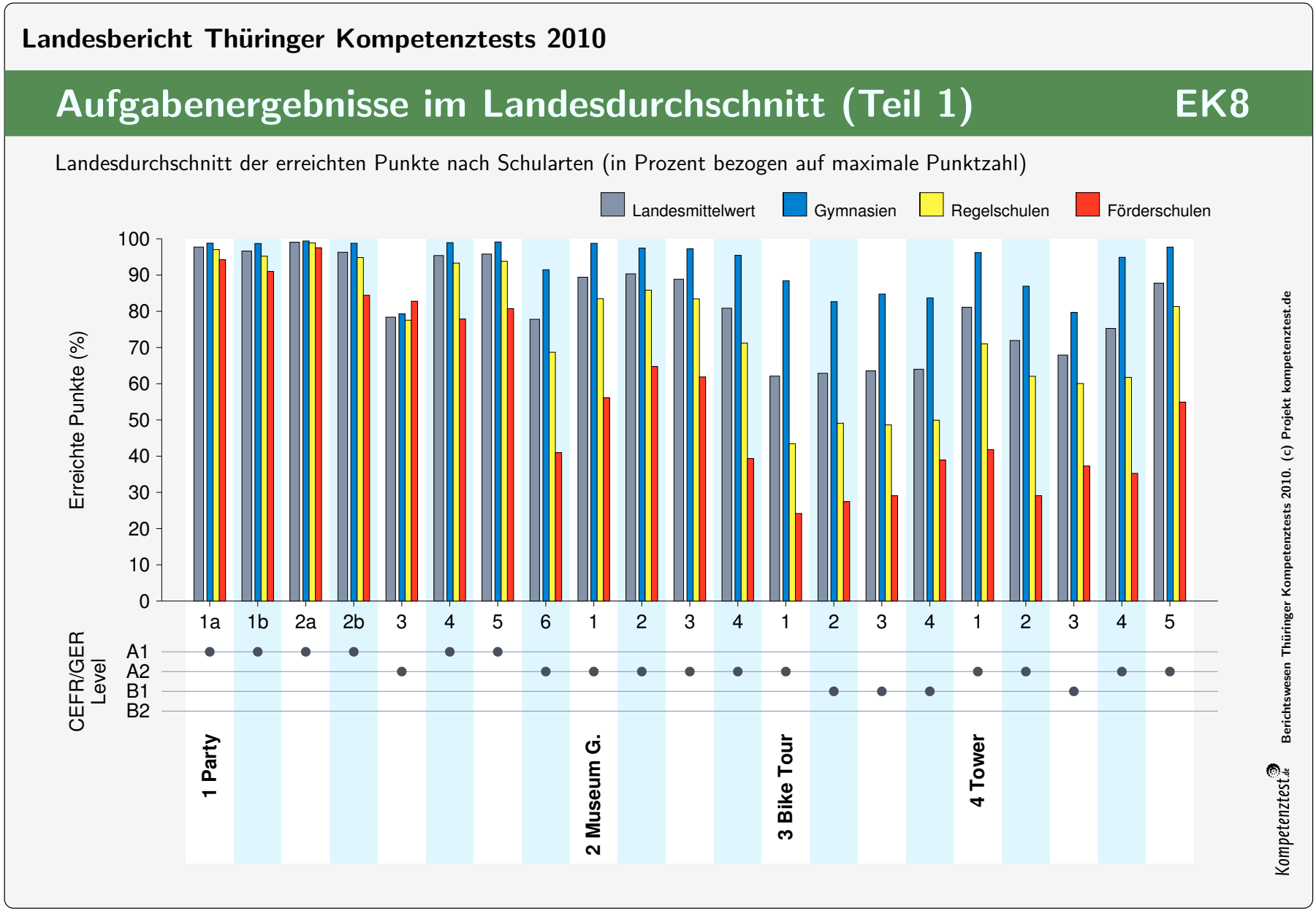


**Abbildung 6.40:** Aufgabenschwierigkeiten im Landesdurchschnitt sowie getrennt nach Schulart (Englischtest, Klassenstufe 8, EK8 – Teil 1)

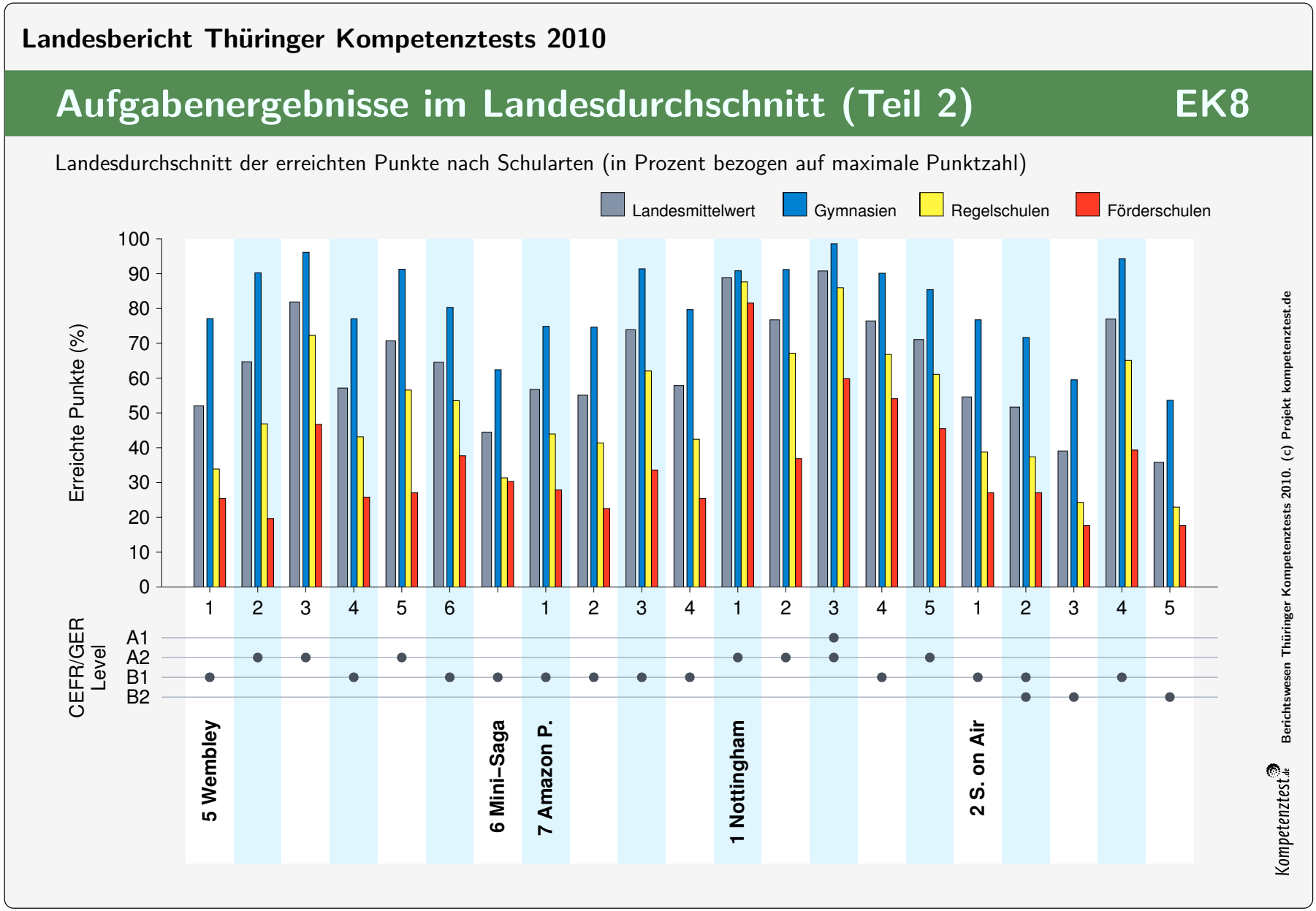


**Abbildung 6.41:** Aufgabenschwierigkeiten im Landesdurchschnitt sowie getrennt nach Schulart (Englischtest, Klassenstufe 8, EK8 – Teil 2)

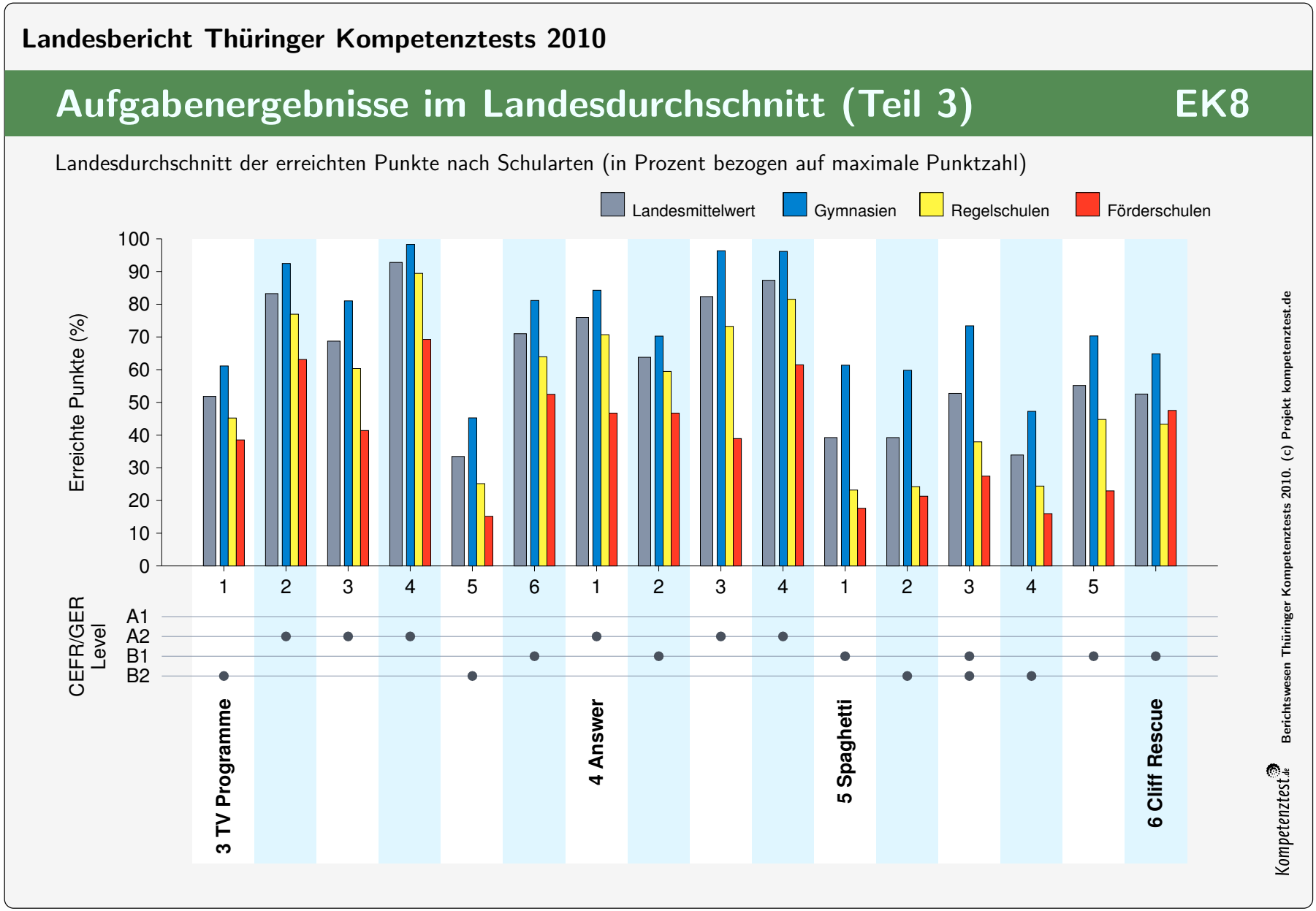


**Abbildung 6.42:** Aufgabenschwierigkeiten im Landesdurchschnitt sowie getrennt nach Schulart (Englischtest, Klassenstufe 8, EK8 – Teil 3)

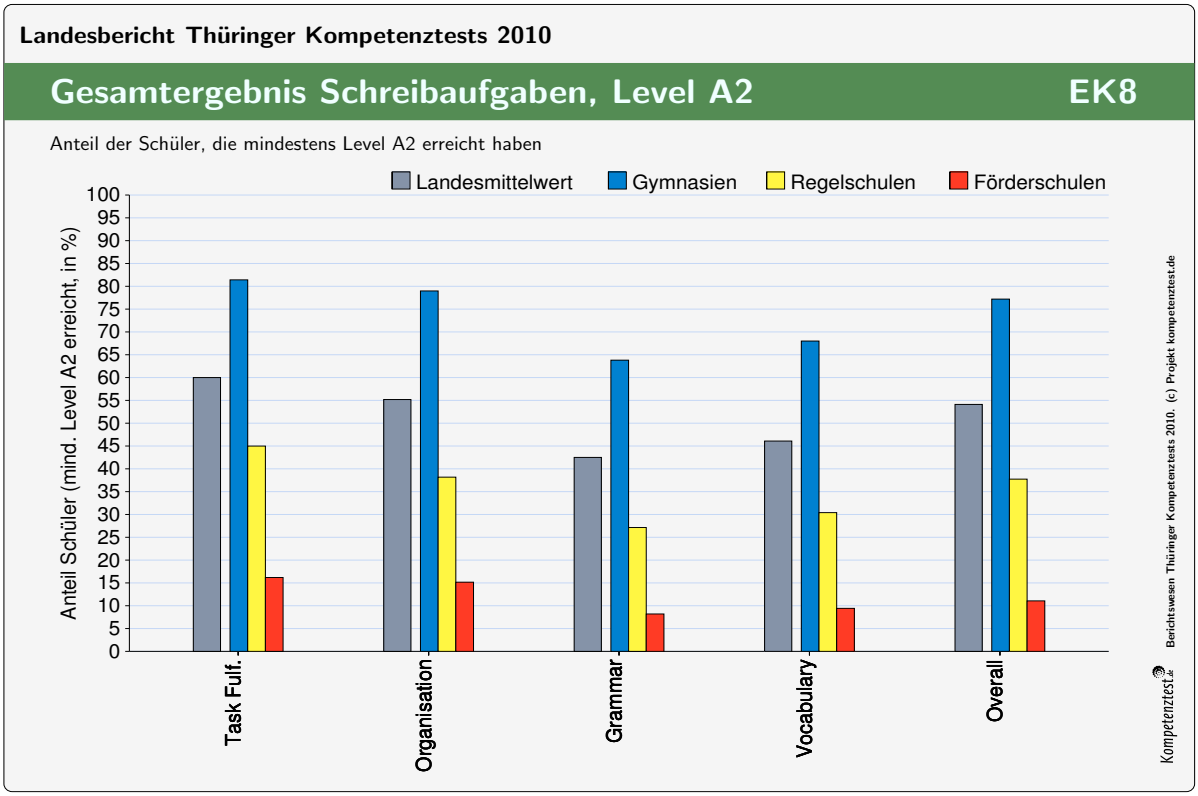


**Abbildung 6.43:** Ergebnisse der Schreibkriterien der Aufgaben 'Description of a Person' und 'Celebrations' (EK8, Level A2)

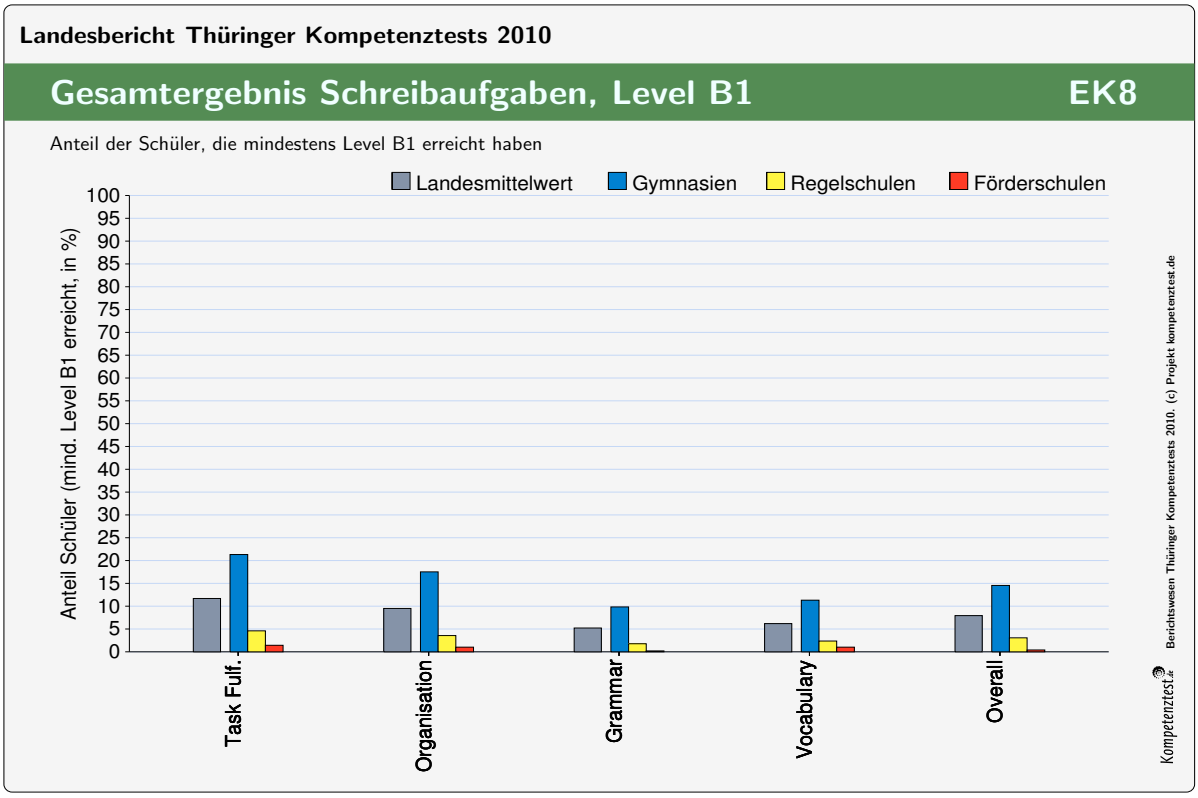


**Abbildung 6.44:** Ergebnisse der Schreibkriterien der Aufgaben 'Description of a Person' und 'Celebrations' (EK8, Level B1)

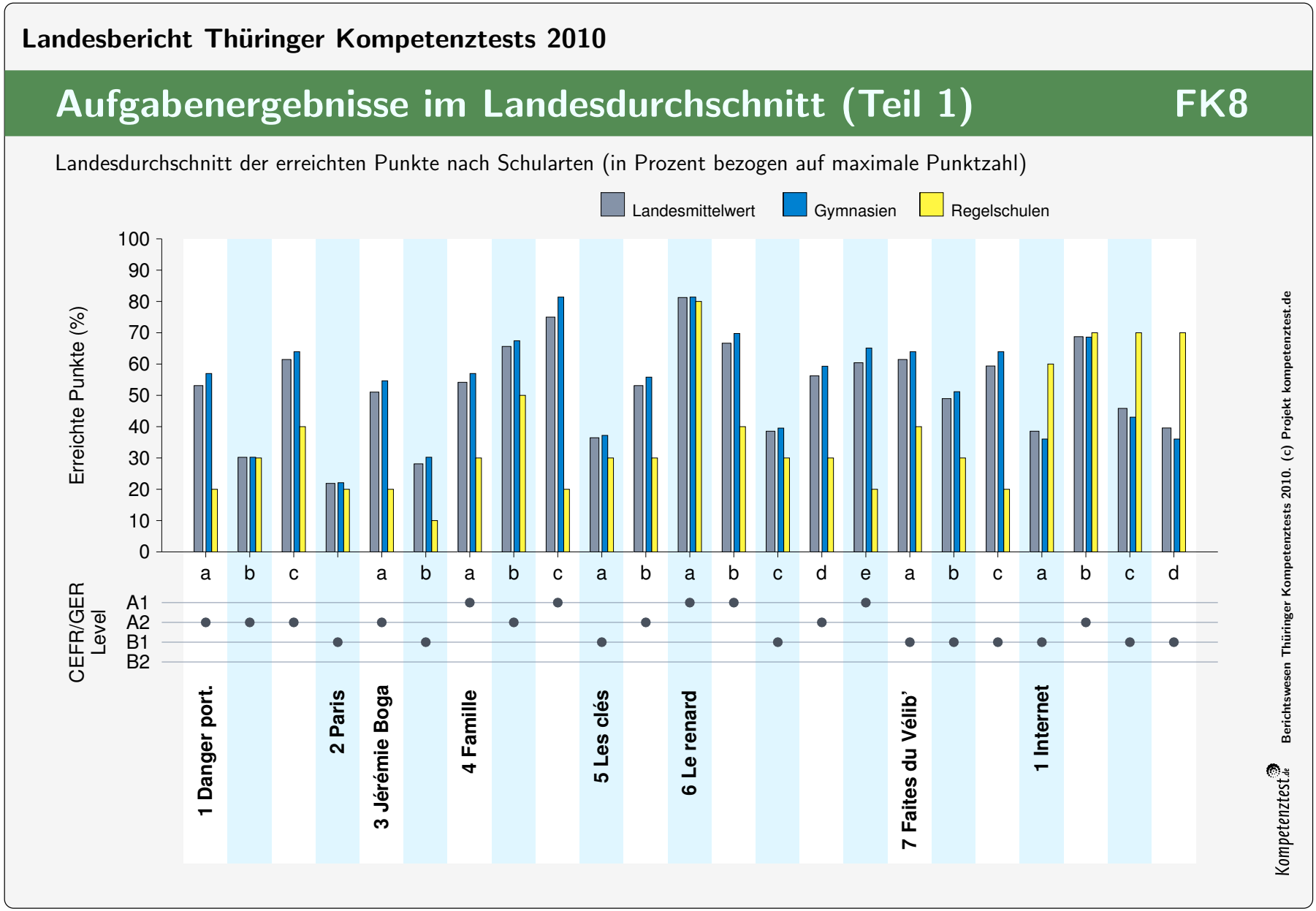


**Abbildung 6.45:** Aufgabenschwierigkeiten im Landesdurchschnitt sowie getrennt nach Schulart (Französischtest, Klassenstufe 8, FK8 – Teil 1)

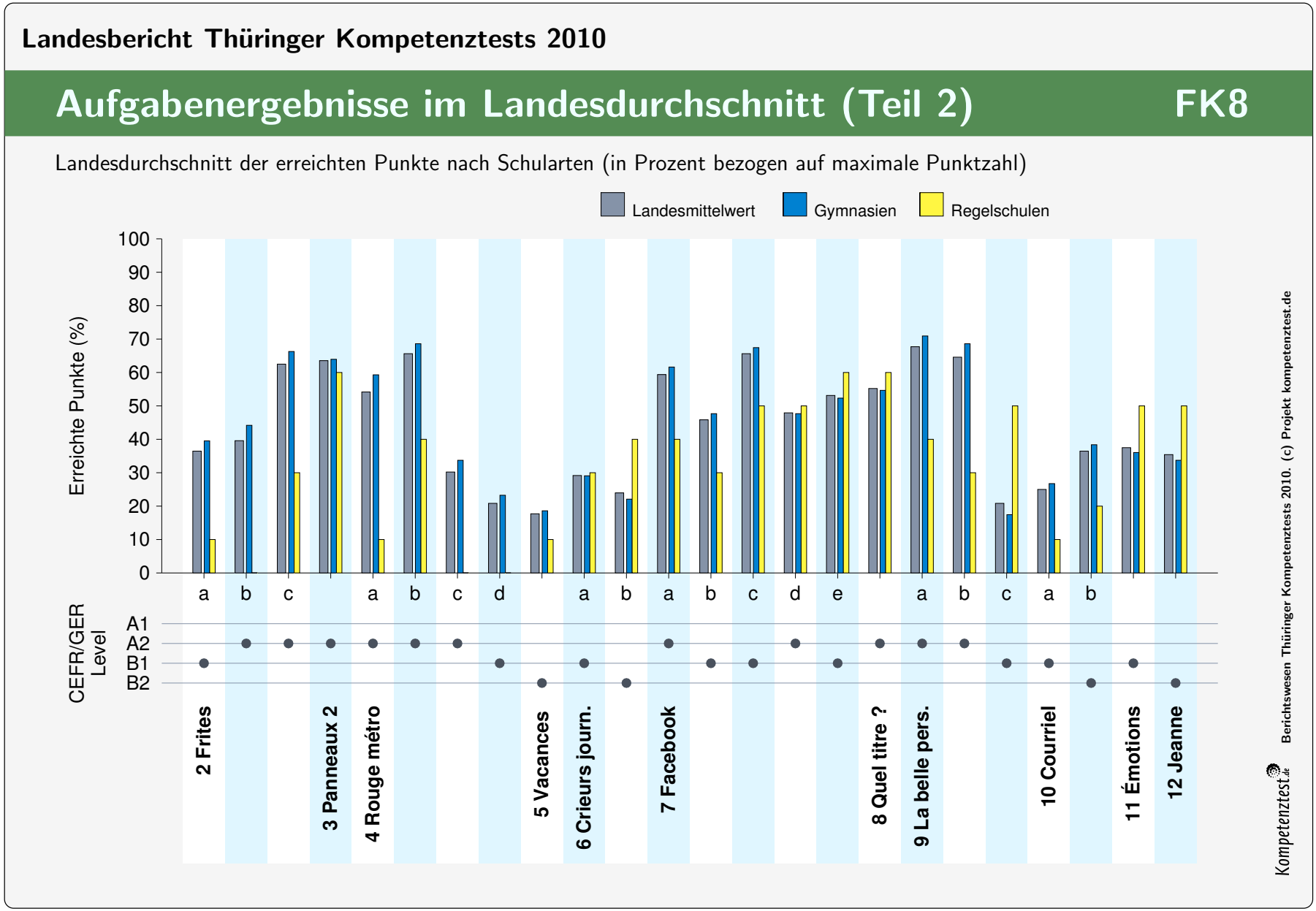


**Abbildung 6.46:** Aufgabenschwierigkeiten im Landesdurchschnitt sowie getrennt nach Schulart (Französischtest, Klassenstufe 8, FK8 – Teil 2)

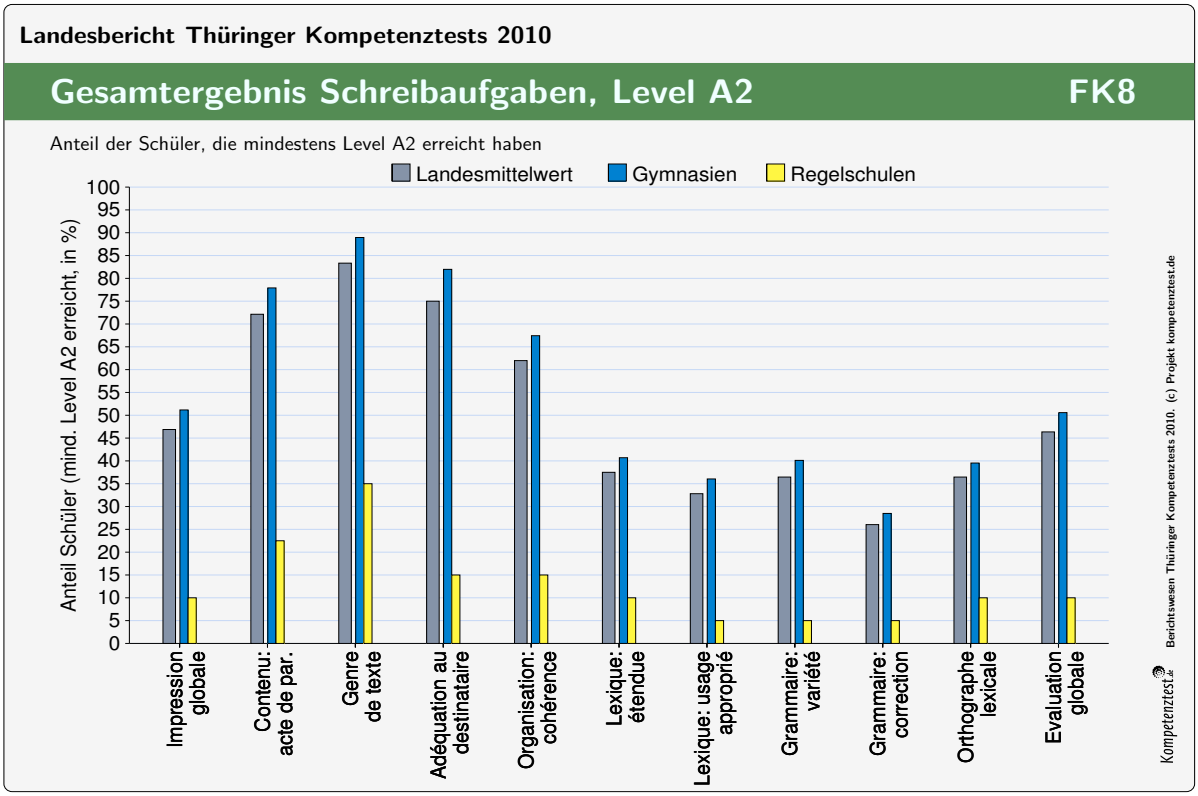


**Abbildung 6.47:** Ergebnisse der Schreibkriterien der Aufgaben 'Cadeau' und 'Film préféré' (FK8, Level A2)

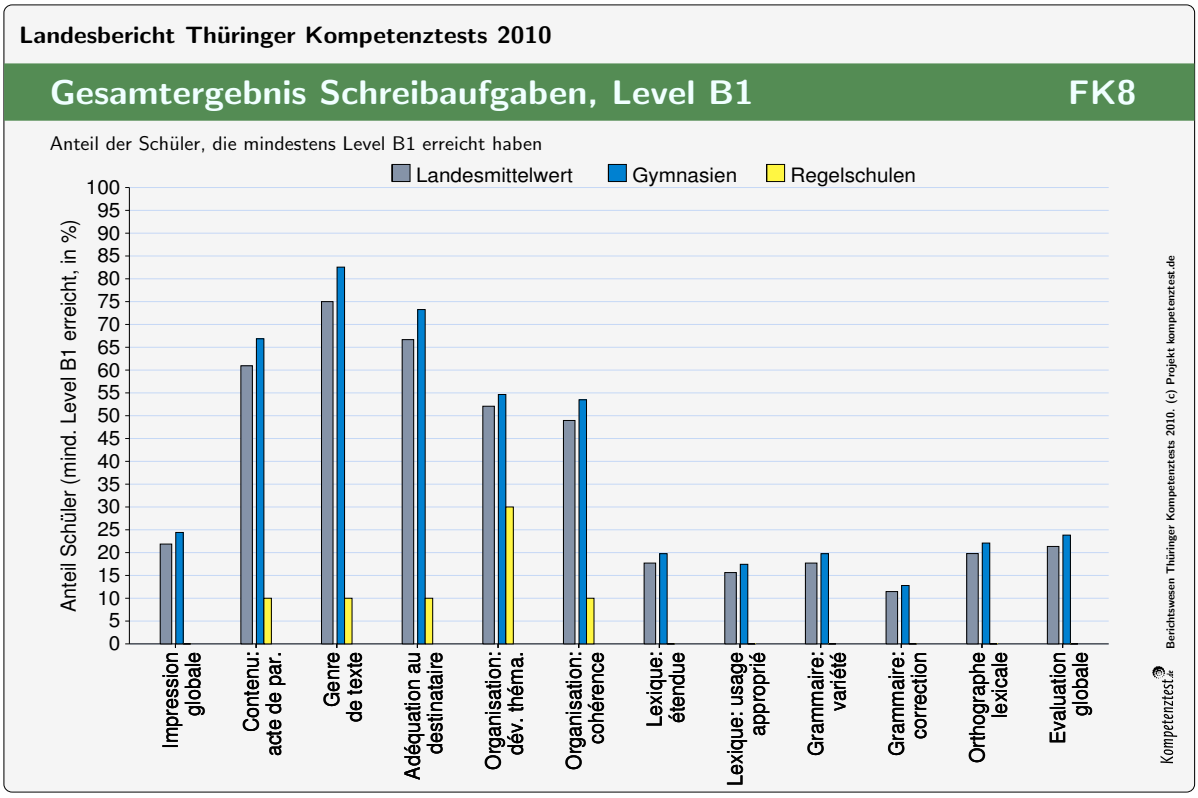


**Abbildung 6.48:** Ergebnisse der Schreibkriterien der Aufgaben 'Cadeau' und 'Film préféré' (FK8, Level B1)

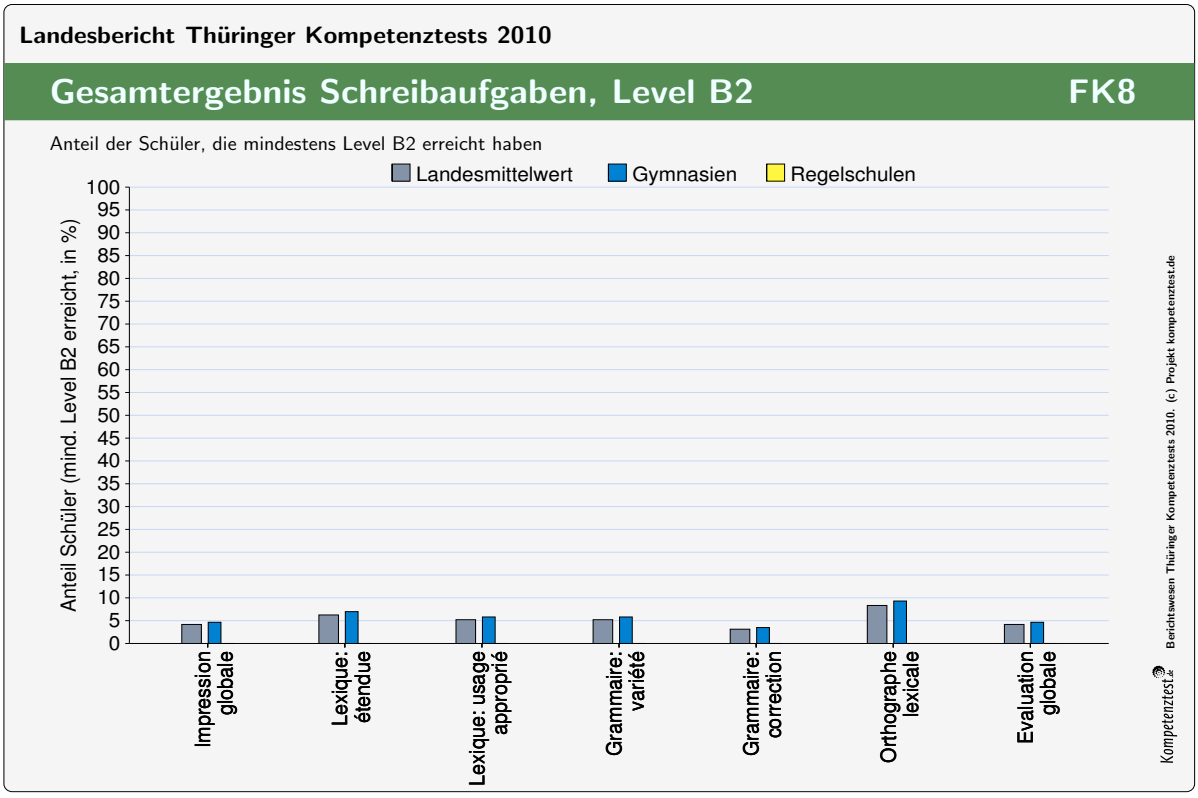


**Abbildung 6.49:** Ergebnisse der Schreibkriterien der Aufgabe 'Film préféré' (FK8, Level B2)

# 7 Analyse leistungsrelevanter Faktoren

## 7.1 Zusammenhang von Prozess- und Wirkungsqualität

### 7.1.1 Allgemeines

Mit dem Thüringer Qualitätsrahmen[1] wird den Schulen eine Orientierungshilfe zur schulischen Qualitätssicherung und Qualitätsentwicklung zur Verfügung gestellt. Dieser definiert drei zentrale Bereiche, mit deren Hilfe schulische Entwicklung beschreibbar und nachvollziehbar wird: Kontext-, Prozess- und Wirkungsqualitäten.

Die Thüringer Kompetenztests erfassen im Wesentlichen Methodenkompetenz und zeigen anhand einer Bestandsaufnahme die Wirkung schulischen Handelns (Wirkungsqualität). Als Instrument zur Verbesserung des Unterrichts werden die Tests in den einzelnen Unterrichtsfächern klassen- bzw. kursweise durchgeführt und die Berichte anschließend verfasst.

Das Selbstevaluationsinstrument *ThüNiS* hingegen erfasst schulische Prozesse, welche zu Kompetenzen der Lernenden führen[2]. Zudem werden Teile der Wirkungsqualitäten erfragt. Es werden dabei sowohl Schüler als auch Lehrer, Mitarbeiter, Ausbilder und Eltern befragt. Als Befragungsinstrument, das als Steuerungsinstrument auf Schulebene fungiert, werden die Berichte in zusammengefasster Form erstellt. Das bedeutet, dass die befragten Gruppen miteinander verglichen werden, nicht aber einzelne Klassen.

Da die an den Kompetenztests teilnehmenden Klassen und Kurse nicht denen an der ThüNiS-Befragung teilnehmenden entsprechen müssen, diese sogar meist andere sind, ist eine Auswertung auf Klassenebene nicht sinnvoll durchführbar. Eine solche Zuordnung würde zwar das Wissen über schulische Wirkung erhöhen, ist aber praktisch nicht möglich. Aus diesem Grund werden Schulmittelwerte für die vorliegenden Auswertungen herangezogen.

### 7.1.2 Daten und Stichprobe

Um die vorliegenden Analysen durchzuführen, müssen die Daten der Kompetenztests mit denen von ThüNiS in Zusammenhang gebracht werden. Es ist dabei eine essentielle Frage, welche Kennwerte als Hinweis eines „guten“ oder „schlechten“ Abschneidens bzw. „stärker ausgeprägter“ oder „geringerer“ Schulqualität anzusehen sind.

Bei den Klassenergebnissen der Kompetenztests wurde dazu die Abweichung vom korrigierten Landesmittelwert verwendet (vgl. Abschnitt *5.2 Korrigierter Landesmittelwert und faire Vergleiche* ab Seite 34). Hat eine Klasse einen negativen Differenzwert, so bedeutet das, dass diese schlechter abschneidet, als es unter den in den korrigierten Landesmittelwert einbezogenen Bedingungen zu erwarten gewesen wäre. Ist die Differenz positiv, so schneidet die Klasse entsprechend besser ab, als es unter den erfassten Rahmenbedingungen zu erwarten gewesen wäre.

[1] vgl. http://www.thueringen.de/de/publikationen/pic/pubdownload266.pdf
[2] vgl. http://www.kompetenztest.de/thuenis/thuenis1.html

Beim Evaluationsinstrument ThüNiS werden Zusammenfassungen der 24 erfassten Teilbereiche vorgenommen. Somit könnten Zusammenhänge nicht nur festgestellt, sondern möglicherweise inhaltlich interpretierbar sein. Bei der ThüNiS-Befragung ist große Zustimmung mit kleinen Werten versehen, Ablehnung wird durch große Werte repräsentiert. Zustimmung ist im Rahmen der ThüNiS-Befragung als positiv zu werten, da die Fragen positiv formuliert sind[3].

Um die Zusammenhangsanalyse durchzuführen, wurden alle Schulen welche bei ThüNiS 2010 teilgenommen haben betrachtet. Zu diesen Schulen wurden die Schulmittelwerte der oben beschriebenen Differenzen der einzelnen Kompetenztests diesen Jahres gebildet und mit den Einschätzungen der Schulqualität zusammengeführt. Durch dieses Verfahren entsteht eine Stichprobe von 120 Schulen mit dessen Mittelwerten der einzelnen Teilbereiche (ThüNiS) (getrennt nach Befragungsgruppen) und den Schulmittelwerten der Ergebnisse der Kompetenztests. Es wurden im Rahmen dieser Auswertung Einschätzungen von 10820 Schülerinnen und Schülern, 8690 Eltern und 1827 Lehrerinnen und Lehrern einbezogen.

An dieser Stelle sei nochmals darauf hingewiesen, dass die Kompetenztests in den Klassenstufen 3, 6 und 8 durchgeführt werden, wohingegen es bei der Befragung in ThüNiS keine Einschränkungen der Klassenstufe gibt. Es muss also davon ausgegangen werden, dass die Befragten bei ThüNiS nicht die sind, welche die Kompetenztests durchgeführt haben. Aus diesem Grund können Interpretationen auch nicht auf Individual- oder Klassenebene stattfinden, sondern müssen auf Schulebene erfolgen.

### 7.1.3 Ergebnisse und Schlussfolgerungen

Im Folgenden wird auf Zusammenhänge zwischen den prozessbezogenen Teilbereichen in ThüNiS und den wirkungsbezogenen Leistungsdaten der Kompetenztests eingegangen.

Wenn man vorerst lediglich die Zusammenhänge in ihrer Tendenz völlig ungeachtet der Signifikanz betrachtet, so ist festzustellen, dass diese fast ausschließlich positiv sind. Dass positive Einschätzungen der ThüNiS-Befragung mit höher als zu erwartenden Testergebnissen einhergehen bedeutet, dass in egal welcher Gruppe der Befragten eine Beziehung von schulischer Prozessqualität und Wirkungsqualität bestehen kann. Hier jedoch nochmals der Hinweis, dass, obwohl es sich hier um ein generelles Phänomen handelt, in vielen Fällen kein statistisch bedeutsamer Zusammenhang vorliegt. Eine Interpretation auf dieser Grundlage wäre höchst kritisch, wenngleich ausbleibende Signifikanz nicht bedeuten muss, dass kein tatsächlicher Zusammenhang vorliegt.

Betrachtet man zunächst nur die Ergebnisse *junger Schüler*[4], so ist im Wesentlichen festzustellen, dass keine Signifikanzen vorliegen, die über das Zufallsniveau[5] hinausgehen. Es ist jedoch eine Häufung signifikanter Zusammenhänge im Fach Deutsch der Klassenstufe 6 festzustellen. Auch bei *älteren Schülern*[6] und *Eltern* gibt es signifikante Zusammenhänge zwischen Merkmalen der Schulqualität und schulischer Leistung fast ausschließlich im Deutschtest der Klassenstufe 6. Anders ist das Zusammenhangsmuster bei Lehrkräften. Obwohl bei diesen ebenfalls eine stärkere Häufung bedeutsamer Zusammenhänge von eingeschätzter Schulqualität und Ergebnissen des Deutschtests der Klassenstufe 6 vorliegt, stehen außer dem Englischtest der Klassenstufe 6 und dem Deutsch-Rechtschreibtest der Klassenstufe 3 alle anderen Testergebnisse im Zusammenhang mit der Einschätzung der Schulqualität.

[3] bspw. Frage: „Ich achte genau auf das, was die Schülerinnen sagen.“
[4] junge Schüler sind Schüler der Klassenstufe 3 bis 6.
[5] Bei einem Signifikanzniveau von 5% ist damit zu rechnen, dass 5% der erhaltenen signifikanten Ergebnisse durch Zufall zustande gekommen sind.
[6] Ältere Schüler sind Schüler ab Klassenstufe 7

Unterschiede des Zusammenhangs in den unterschiedlichen Klassenstufen sind insofern festzustellen, als dass die Tests der Klassenstufe 3 weniger gut geeignet sind, um die eingeschätzte Schulqualität vorherzusagen und umgekehrt. Ein Unterschied zwischen den anderen beiden Klassenstufen ist nicht festzustellen, wenn man vom Fach Deutsch absieht.

Schwerpunkte der Schulqualität, bei denen ein Zusammenhang zu den Testergebnissen zu verzeichnen ist, sind Bereiche der Wirkungsqualität. Dies erscheint nachvollziehbar, da diese ebenfalls durch die Kompetenztests abgebildet werden. Hierbei zeigt sich innerhalb dieses Bereichs eine Häufung im Teilbereich Methodenkompetenz. Überraschend ist, dass gute Leistung insgesamt stärker mit schulischen Qualitätsaspekten zusammenhängt, die nicht direkt den Bereich „Lehren und Lernen“ betreffen.

Zusammenfassend kann man sagen, dass Zusammenhänge zwischen der eingeschätzten Schulqualität der Befragten und den Testergebnissen der Schulen beobachtbar sind. Kausale Schlüsse und Interpretationen können aus diesen Ergebnissen jedoch nicht gezogen werden. Ob eine gesteigerte Leistungsfähigkeit der Schüler für bessere Schulqualität verantwortlich ist, oder ob eine reine qualitative Intervention die Schülerschaft verändert, ist nicht zu schlussfolgern. Die vorliegenden Ergebnisse sind vorerst als rein deskriptiv zu betrachten und verlangen weitere Analysen, die jedoch unter der derzeit gegebenen Daten- und Datenschutzlage nicht zulässig sind.

### 7.1.4 Diskussion & Anregungen

Obgleich sich die Tendenz eines Zusammenhangs gesteigerter Qualitätskriterien und Leistungsvermögen der Schüler einer Schule zeigt und diese teils durch signifikante Ergebnisse gestützt wird, muss man sich die Grundfrage stellen, ob es einen solchen Zusammenhang gibt. Unter Betrachtung aller Ergebnisse wird deutlich, dass dieser positive Zusammenhang auftritt, jedoch nicht immer bedeutsam wird. Bedeutet das allerdings gleichzeitig, dass kein realer Zusammenhang vorliegt? Keinesfalls. Die statistische Bedeutsamkeit ist von mehreren Faktoren abhängig. So spielt zum Einen der reale Effekt bzw. Zusammenhang eine Rolle, ist aber neben Irrtumswahrscheinlichkeit und Stichprobengröße nur eine Voraussetzung für Signifikanz.

Zusätzlich zu den rein strukturell statistischen Merkmalen, sollte man jedoch auch über beobachtbare Problemfelder einer Analyse sprechen. Als diese kann hierbei die Zusammensetzung der Untersuchten Schulen und Klassen betrachtet werden. Im Rahmen der ThüNiS-Befragung liegt eine Stichprobe vor, welche durch freiwilligkeit gekennzeichnet ist. Es ist demnach nicht auszuschließen, muss sogar angenommen werden, dass die landesweit vorfindbare Diversität der Schulqualität hier nicht vorliegt. Es wäre denkbar, dass sich entweder Schulen durchschnittlicher Schulqualität oder mit besonderen Auffälligkeiten in schulischen Qualitätsdimensionen besonders bemühen, um die eigene Schule und somit auch die schulische Leistungsfähigkeit voranzubringen. Eine Kontrolle dessen wäre nur über eine Vollerhebung möglich, welche jedoch der Maßgabe der Eigeninitiierung schulischer Selbstevaluation widerspräche.

Bedingt durch die unterschiedlichen Bedingungen der Teilnahme bei ThüNiS und den Thüringer Kompetenztests, sind die Klassen, welche die Lernstandserhebungen durchführten, nicht zwangsläufig diese, welche zu schulischen Qualitätsmerkmalen befragt wurden. Die bietet auf der einen Seite eine Einschränkung, da man argumentieren könnte, dass damit die Ergebnisse nichts miteinander zu tun hätten, aber man könnte auch die Chancen hervorheben, welche sich deutlich auf Interpretationsebene Schule befinden. So könnte man annehmen, dass die eingeschätzte Schulqualität eine Art generalisiertes Verständnis für schulische Prozesse der Schule darstellt und gerade wegen der nicht zu verbindbaren Ergebnisse auf Klassenebene

die Qualitätsmerkmale nicht vom Leistungsstand der an den Kompetenztests teilnehmenden Schüler verfälscht sind. Eine objektive Erfassung der Zusammenhänge könnte somit sogar erreicht worden sein.

Aufgrund der zweifelsfreien Richtung des Zusammenhangs von schulischer Leistung mit Prozess- und Wirkungsqualität, gepaart mit bedeutsamen Ergebnissen, kann angenommen werden, dass dieser ein reales Phänomen darstellt. Da Prozesse in Schule und Unterricht der Weg zur Leistungsfähigkeit der Empfänger des Unterrichts - den Schülerinnen und Schülern - sind, wird nahegelegt, diese fortlaufend zu untersuchen. Eine anschließende Maßnahmenplanung schulischer Entwicklung auf Grundlage dieser fundierten Evaluation sollte ebenso erfolgen, wie dessen Umsetzung und Nachkontrolle. Dieser Kreislauf der Evaluation kann die schulischen Akteure unterstützen und fördern. Mit dem Instrument ThüNiS stellt das Projekt Kompetenztest.de ein umfangreiches Instrument zur Verfügung, welches interessierten Schulen in Thüringen kostenfrei zur Verfügung steht. Personal der Universität Jena, der Schulämter und des Thüringer Ministeriums für Bildung, Wissenschaft und Kultur steht in ihrer beratenden und unterstützenden Funktion zur Verfügung.

**Das Wichtigste in Kürze**

- Genereller Zusammenhang hoher Schulqualität und Leistung
- Bedeutsame Ergebnisse vor allem im Bereich Wirkungsqualität, Schulklima & Schulkultur und Kooperation & Kommunikation
- Zusammenhänge besonders deutlich bei Einschätzungen der Lehrkräfte
- Ergebnisse stützen die Notwendigkeit schulischer Selbstevaluation

## 7.2 Einflussfaktoren auf Schülerebene

In diesem Kapitel wird betrachtet, inwiefern sich die Thüringer Schüler hinsichtlich der gezeigten Leistungen in den Kompetenztests unterscheiden. Der hohe Zusammenhang zwischen den erfassten mathematischen und den sprachlichen Kompetenzen bietet viel Raum für Interpretationen (korrelativer Zusammenhang Mathematik und Sprache: $r \approx 0,65$; korrelativer Zusammenhang Deutsch und Englisch: $r \approx 0,75$). Vereinfacht könnte man sagen, dass Schüler, welche ein gutes Ergebnis im Kompetenztest Mathematik erreichten, tendenziell auch besser bei den sprachlichen Kompetenztests in Deutsch und Englisch abschnitten. Gibt es einfach leistungsstarke und leistungsschwache Schüler? Was sind mögliche Gründe für das gute oder weniger gute Abschneiden bei den Kompetenztests? In diesem Kapitel sollen Faktoren auf Schülerebene betrachtet werden, welche nicht im Einflussbereich der Lehrkraft liegen, aber als unterrichtsvorgeordnete Merkmale einen großen Einfluss auf das Abschneiden bei den Kompetenztests haben können. Es erfolgen detaillierte Analysen zu den Schülermerkmalen Geschlecht, Muttersprache, sozioökonomischer Status (aus Datenschutzgründen auf Klassenebene), Jahr des Schulbesuchs und besonderen Lernschwierigkeiten. Weiterhin wird untersucht, ob Halbjahresnoten und die Ergebnisse der Kompetenztests tendenziell in eine Richtung zeigen, d.h. ob Schüler mit guter Halbjahresnote auch ein gutes Testergebnis erzielen. Es wird der Frage nachgegangen, ob bestimmte Schülergruppen besser in ihrer Leistungsfähigkeit eingeschätzt werden können. In den einzelnen Unterkapiteln wird es eine Betrachtung der jeweiligen Variable über mehrere Jahre hinweg geben. Dabei steht die Frage im Vordergrund, ob Unterschiede zwischen Schülergruppen konsistent über die Zeit bestehen oder ob sich eine

positive bzw. negative Entwicklung abzeichnet. Am Ende dieses Kapitels wird der gemeinsame Einfluss der genannten Merkmale auf das Abschneiden in den Kompetenztests betrachtet (Regressionsanalyse).

***Statistische Auswertungen in diesem Kapitel und deren Interpretation***

Bei der Betrachtung des Zusammenhanges von Schülermerkmalen und den Kompetenztestleistungen werden folgende Analyseverfahren und Kennwerte verwendet und in diesem Kapitel berichtet: Mittelwerte (MW), Standardabweichungen (SD), die Fallzahl (N), Effektstärken ($d$), Korrelationen ($r$) und Regressionen. Um die Nachvollziehbarkeit der Analysen zu erhöhen, werden die verwendeten Methoden zunächst kurz erläutert und exemplarisch an Abbildungen veranschaulicht. Vertiefende Informationen sind beispielsweise bei Bortz (2002) oder Wirtz und Nachtigall (2004) nachlesbar. Bei der Interpretation aller Analysen ist zu beachten, dass die durchschnittliche Leistungsdifferenz nicht auf die Leistungsentwicklung einzelner Schüler übertragbar ist.

**Mittelwerte** geben an, wie gut die Schüler im Durchschnitt abgeschnitten haben. Sie werden berechnet, indem die Summe aller Werte durch die Anzahl aller Werte geteilt wird. Ergänzend zu diesem Wert wird die **Standardabweichung** angegeben. Die Standardabweichung weist darauf hin, wie stark sich die Testergebnisse der Schüler im Durchschnitt von dem Mittelwert unterscheiden. Sie gibt also Auskunft über die Unterschiedlichkeit der Testergebnisse. Um die Ergebnisse unterschiedlicher Tests mit unterschiedlichen Mittelwerten und Teilnehmern vergleichen zu können, kann man die Unterschiede in Prozentangaben betrachten (vgl. Abb. 7.1) oder die **Effektstärke** berechnen. Die Effektstärke gibt an, ob und wie stark sich die Schülergruppen in ihren Ergebnissen bei den Kompetenztests unterscheiden. Dabei wird eine Vergleichbarkeit der mittleren Ergebnisse unabhängig vom Test und der Art der Messung hergestellt. Dabei gilt nach Cohen (1988) eine Effektstärke ab $d = 0,2$ als geringer, ab $d = 0,5$ als mittelstarker und ab $d = 0,8$ als starker Effekt.

**Beispiel: Gruppenunterschiede mit Effektstärke angeben:** Aus den Mittelwerten und Standardabweichungen von zwei Schülergruppen (z.B. Mädchen und Jungen) errechnet sich eine Effektstärke $d = 0,287$. Diese zeigt zunächst an, dass sich die beiden Gruppen in ihren Leistungen bedeutend voneinander unterscheiden. Es liegt jedoch nur ein kleiner Effekt vor. Das Vorzeichen gibt an, dass die erste Gruppe (hier: Mädchen) im Durchschnitt bessere Leistungen erzielt. Ein negatives Vorzeichen würde anzeigen, dass die zweite Gruppe (hier: die Jungen) besser abschneidet.

**Beispiel: Grafische Veranschaulichung von Gruppenunterschieden:** Zur Veranschaulichung von Leistungsunterschieden wird es in diesem Kapitel vereinzelt Abbildungen geben. Exemplarisch werden diese mit Abbildung 7.1 veranschaulicht, in welcher die Leistungsdifferenz zwischen Jungen und Mädchen dargestellt ist. Mädchen lösen im Mittel 64,5% der Aufgaben eines Tests, Jungen dagegen 59,1%. Dies ist in der Abbildung durch die gestrichelten senkrechten Linien dargestellt. Die blaue (dunkle) und die grüne (helle) Kurve geben die Häufigkeiten an, mit der Schüler einen bestimmten Prozentsatz an Aufgaben lösten. Für beide Kurven gilt, dass sehr viele Schüler mittelmäßig abschnitten. Dies erkennt man daran, dass die Kurven im mittleren Bereich der Prozentskala ihren höchsten Punkt erreichen. Wenige Schüler erreichten sehr wenige oder sehr viele Punkte bei den Kompetenztests. Die Häufigkeitsverteilung der blauen, dunklen Kurve (Mädchen) liegt im Vergleich zur grünen, hellen Kurve (Jungen) etwas weiter rechts, d.h. mehr Mädchen lösten einen höheren Prozentsatz der Aufgaben.

**Korrelationen** geben die Enge des Zusammenhanges zwischen zwei Merkmalen an und können Werte zwischen $-1$ und 1 annehmen. Dabei ist nach Nachtigall und Wirtz (2004) der Betrag einer Korrelation bis 0.50 als gering, bis 0.70 als mittel und ab 0.70 als hoch einzustufen. Beispielsweise deutet die oben erwähnte Korrelation zwischen den mathematischen

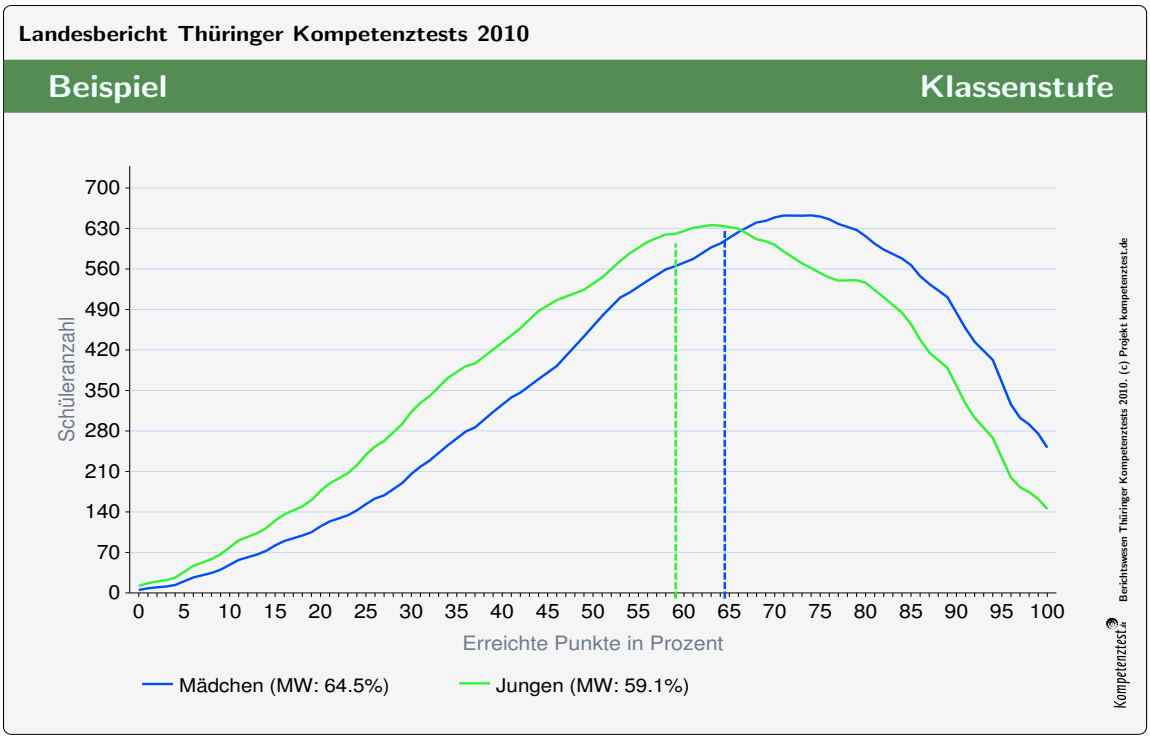


**Abbildung 7.1:** Vergleich der Testergebnisse zweier Schülergruppen

und den sprachlichen Kompetenztests von r$\approx 0,65$ auf eine mittlere Enge des Zusammenhanges hin. Liegt zwischen zwei Merkmalen ein bedeutender Zusammenhang vor, kann geprüft werden, ob eines der Merkmale zur Vorhersage des anderen eingesetzt werden kann. Diese Merkmalsvorhersage erfolgt in Form von **Regressionen**.

### 7.2.1 Mathematische und sprachliche Kompetenzen in Abhängigkeit vom Geschlecht

Schaut man sich die Halbjahresnoten der Thüringer Schüler und Schülerinnen an, so fällt auf, dass sich die Noten in Mathematik im Mittel nicht zwischen den Geschlechtern unterscheiden. Dagegen erhalten Mädchen in den sprachlichen Fächern Deutsch, Englisch und Französisch im Durchschnitt eine etwas bessere Note auf dem Halbjahreszeugnis als die Jungen. Da in die Halbjahresnote mehr Informationen eingehen als nur die Leistungsfähigkeit des Schülers (z.B. die Motivation des Schülers im Unterricht), wird es in diesem Abschnitt interessant zu schauen, inwiefern sich Jungen und Mädchen der verschiedenen Jahrgangsstufen in ihren Ergebnissen bei den Kompetenztests unterscheiden. Dazu werden zunächst die mathematischen Kompetenzen der Thüringer Schüler in diesem Jahr betrachtet. Weiterhin wird die Entwicklung der Kompetenzen in Mathematik eines bestimmten Jahrganges (Längsschnitt) und verschiedener Jahrgänge (Quasilängsschnitt) dargestellt. Die Betrachtung eines Jahrgangs hat den Vorteil, dass man nachvollziehen kann, ob sich Leistungsunterschiede in Abhängigkeit vom Geschlecht erst über die Zeit entwickeln bzw. ob sie stärker oder schwächer werden. Die Analyse der Kompetenzen verschiedener Jahrgänge gibt einen Hinweis darauf, ob Geschlechtsunterschiede typischerweise auftreten oder ob der Leistungsunterschied zwischen Jungen und Mädchen nur speziell für einen Jahrgang nachweisbar ist. Anschließend erfolgen die genannten Analysen auch für die sprachlichen Kompetenzen.

***Mathematische Kompetenzen im Jahr 2010***

Tabelle 7.1 zeigt das Abschneiden der Mädchen und Jungen in den Kompetenztests Mathematik der verschiedenen Jahrgangsstufen. Das negative Vorzeichen der angegebenen Effektstärken zeigt, dass Jungen tendenziell ein besseres Ergebnis erzielen. In Klassenstufe 3 der Thüringer Schulen ist diese Tendenz so gering, dass man davon ausgehen kann, dass Mädchen und Jungen in diesem Alter vergleichbare mathematische Kompetenzen besitzen. In Klassenstufe 6 und 8 der Regelschulen und Gymnasien lässt sich ein kleiner Effekt zugunsten einer besseren mathematischen Kompetenz der Jungen finden. Etwas stärker fällt dieser Geschlechtereffekt in den Förderschulen der Klassenstufe 8 aus (mittelstarke Effektstärke d = −0,703).

**Tabelle 7.1:** Mittelwertevergleich der mathematischen Kompetenzen zwischen Jungen und Mädchen

| | Mädchen | | | Jungen | | | Effektstärke |
|---|---|---|---|---|---|---|---|
| **Fach** | **MW** | **(SD)** | **(N)** | **MW** | **(SD)** | **(N)** | **(d)** |
| MK 3 | 34,75 | (9,29) | (7 460) | 35,42 | (9,56) | (7 719) | (−0,071) |
| MK 6 | 15,03 | (5,21) | (6 043) | 16,39 | (5,27) | (6 244) | (−0,260) |
| MK 8 | 16,33 | (7,13) | (5 876) | 17,72 | (7,28) | (6 094) | (−0,193) |

***Entwicklung der mathematischen Kompetenzen über die Jahre:***

Schaut man sich die Geschlechtereffekte in den Mathematik-Kompetenztests verschiedener Jahrgänge in Tabelle 7.2 an, so fällt anhand der negativen Vorzeichen sofort auf, dass Jungen verschiedener Jahrgänge immer etwas besser abschneiden als Mädchen. Jedoch erreicht die Stärke der Mittelwertsdifferenz in Form von Effektstärken selten ein bedeutendes Ausmaß (ab d = 0,2 spricht man von einem kleinen, aber bedeutenden Effekt). Betrachtet man die letzten beiden Jahre, lässt sich feststellen, dass sich besonders in den 3. Klassen kein Hinweis für einen Geschlechterunterschied hinsichtlich der mathematischen Kompetenz finden lässt. Dagegen zeigen die Effektstärken in den weiterführenden Schulen, dass Jungen im Mittel etwas bessere Leistungen zeigten als die Mädchen. Für Klassenstufe 6 lässt sich ein Trend zu zunehmenden Geschlechtsunterschieden in den verschiedenen Jahrgängen ablesen. Seit 2005 sind die Leistungsunterschiede stetig etwas größer geworden.

**Tabelle 7.2:** Entwicklung des Geschlechtereffekts in den Kompetenztests Mathematik

| **Jahr** | **Mathematik 3 Effektstärke d** | **Mathematik 6 Effektstärke d** | **Mathematik 8 Effektstärke d** |
|---|---|---|---|
| 2010 | −0,071 | − 0,260 | **−0,193** |
| 2009 | −0,044 | −0,232 | −0,189 |
| 2008 | −0,200 | **−0,193** | −0,176 |
| 2007 | −0,151 | − 0,197 | – |
| 2006 | −0,159 | −0,147 | – |
| 2005 | **−0,129** | −0,078 | – |

Tabelle 7.2 bietet weiterhin die Möglichkeit, die Entwicklung eines Jahrganges abzulesen. Die hervorgehobenen Effektstärken zeigen, wie sich die Geschlechtsunterschiede in den Leistungen für die diesjährige Klassenstufe 8 seit 2005 entwickelt haben. Zu allen drei Erhebungszeit-

punkten schnitten die Jungen tendenziell, aber statistisch nicht bedeutend, etwas besser ab als die Mädchen. Dabei ist festzustellen, dass die Tendenz in Klassenstufe noch etwas geringer ausgeprägt war als in den weiterführenden Schulen.

***Sprachliche Kompetenzen im Jahr 2010***

Tabelle 7.3 zeigt das Abschneiden der Thüringer Schüler und Schülerinnen bei den sprachbezogenen Kompetenztests in Deutsch und Englisch. Da nur sehr wenige Schüler am Kompetenztest in Französisch teilnahmen, wird an dieser Stelle nicht näher darauf eingegangen. Die Analysen dazu können aber im Anhang eingesehen werden. In allen drei Klassenstufen zeigt sich, dass Mädchen sowohl in Deutsch als auch in Englisch besser abschnitten. Es liegen kleine bis mittelstarke Effekte vor. Anzumerken ist, dass der Leistungsunterschied zugunsten der Mädchen in Abhängigkeit vom Geschlecht in den Förderschulen nicht so stark ausgeprägt ist wie in den anderen Schulformen.

**Tabelle 7.3:** Mittelwertevergleich der sprachlichen Kompetenzen

| | Weiblich | | | Männlich | | | Effektstärke |
|---|---|---|---|---|---|---|---|
| | **MW** | **(SD)** | **(N)** | **MW** | **(SD)** | **(N)** | **(d)** |
| **Deutsch** | | | | | | | |
| DK 3 | 26,53 | (7,55) | (7 233) | 24,34 | (7,70) | (7 430) | (0,287) |
| DK 6 | 102,2 | (23,65) | (5 348) | 88,9 | (26,47) | (5 542) | (0,529) |
| DK 8 | 43,12 | (11,09) | (5 906) | 40,99 | (11,42) | (6 116) | (0,189) |
| **Englisch** | | | | | | | |
| EK 6 | 67,05 | (18,76) | (4 995) | 60,03 | (19,974) | (5 027) | (0,362) |
| EK 8 | 57,66 | (16,31) | (5 922) | 52,15 | (17,27) | (6 106) | (0,328) |

***Entwicklung der sprachlichen Kompetenzen über die Jahre***

Der Vergleich des Abschneidens von Jungen und Mädchen verschiedener Jahrgänge bei den sprachbezogenen Kompetenztests Deutsch und Englisch zeigt einen durchgängigen Trend: Obwohl der Geschlechtereffekt in seiner Stärke von Jahr zu Jahr schwankt, kann festgehalten werden, dass Mädchen im Mittel in allen Jahrgängen und in beiden Sprachtests bedeutend besser abschnitten als die Jungen. Anzumerken ist, dass der Geschlechtereffekt im Englischkompetenztest der Klassenstufe 6 nicht in allen Jahrgängen von bedeutender Größe war. Es fällt auf, dass nur die Mädchen der letzten beiden Jahrgänge deutlich besser abschnitten. Damit könnten man vermuten, dass sich auch hier ein Trend zu zunehmenden Geschlechtsunterschieden abzeichnet.

Die hervorgehobenen Zahlen in Tabelle 7.4 zeigen wie sich die Geschlechtsunterschiede in einem Jahrgang in den letzten Jahren entwickelt haben. In Deutsch nimmt der Leistungsunterschied in Abhängigkeit vom Geschlecht deutlich ab. Während sich die Schüler und Schülerinnen in Klassenstufe 3 und 6 noch bei den Ergebnissen der Kompetenztests unterschieden, ist die Differenz in diesem Jahr (Klassenstufe 8) so gering, dass man nicht mehr von einem bedeutenden Kompetenzunterschied sprechen kann. In Englisch hingegen hat sich der Leistungsunterschied zwischen Jungen und Mädchen von Klassenstufe 6 bis Klassenstufe 8 weiter auseinander entwickelt.

**Tabelle 7.4:** Entwicklung des Geschlechtereffekts in den Kompetenztests Deutsch und Englisch

| Jahr | Deutsch 3 Effektstärke d | Deutsch 6 d | Deutsch 8 d | Englisch 6 d | Englisch 8 d |
|---|---|---|---|---|---|
| 2010 | 0,287 | 0,529 | **0,189** | 0,362 | **0,328** |
| 2009 | 0,378 | 0,245 | 0,255 | 0,308 | 0,219 |
| 2008 | 0,177 | **0,313** | – | **0,192** | – |
| 2007 | 0,342 | 0,280 | – | 0,157 | – |
| 2006 | 0,361 | 0,452 | – | 0,066 | – |
| 2005 | **0,364** | 0,361 | – | – | – |

### Fazit

Zusammenfassend lässt sich sagen, dass ein Geschlechtereffekt hinsichtlich der mathematischen und sprachlichen Kompetenzen relativ konsistent zu bestehen scheint. Es zeigt sich immer wieder, dass Mädchen in den sprachlichen Fächern anscheinend bessere Kompetenzen besitzen als die gleichaltrigen Jungen. Jungen erzielen dagegen meist in den Kompetenztests Mathematik eine leicht bessere Testleistung. Dabei fällt auf, dass die Unterschiede in den Testergebnissen für die sprachlichen Fächer etwas stärker ausgeprägt sind als für die mathematischen Kompetenztests. Dieses Ergebnis deckt sich mit den Befunden zur Halbjahresnote. Auch hier war festzustellen, dass sich Jungen und Mädchen in Mathematik nicht bedeutend voneinander unterschieden, in den sprachlichen Fächern dagegen schon. Die Betrachtung der Geschlechtereffekte über die Jahre hat gezeigt, dass es in Mathematik und Englisch einen Trend hin zu größeren Leistungsunterschieden in den verschiedenen Jahrgängen zu verzeichnen gibt (Quasilängsschnitt). Aus der Analyse ausgewählter Jahrgänge über die Zeit (Längsschnitt) geht hervor, dass sich die Geschlechtsunterschiede in Mathematik und Englisch eher verstärken, während sie in Deutsch schwächer werden. Es bleibt zu klären, warum der berichtete Geschlechtereffekt sowohl bei den Kompetenztests als auch bei der Halbjahresnote so konsistent zu finden ist. Unterscheiden sich Mädchen und Jungen tatsächlich in ihrer Leistungsfähigkeit oder werden Geschlechtsdifferenzen in Mathematik und Deutsch durch soziale Einflüsse seitens der Eltern und der Lehrer verstärkt? Diese Fragen können mit den vorliegenden Daten nicht geklärt werden, es gibt jedoch zahlreiche Publikationen, in welchen interessierte Leser nach möglichen Erklärungen recherchieren können. Als Beispiel seien hier der 30. Band zur Bildungsforschung des Bundesministeriums für Bildung und Forschung zum Thema Mathematik und Geschlecht oder eine Publikation von Budde et al. (2008) zu Chancen und Blockaden einer geschlechtergerechten Schule genannt.

**Das Wichtigste in Kürze**

- Jungen schneiden tendenziell etwas besser in den Kompetenztests Mathematik ab.
- Mädchen erzielen im Mittel bessere Ergebnisse in den sprachbezogenen Kompetenztests Deutsch und Englisch.

### 7.2.2 Muttersprache

Laut den PISA-Ergebnissen 2006 ist Deutschland einer der OECD-Mitgliedstaaten, in denen die Kompetenzunterschiede zwischen Jugendlichen mit und ohne Migrationshintergrund am größten sind [7]. Schüler, die in der zweiten Generation in Deutschland leben, deren Eltern beide im Ausland geboren wurden oder die selbst in einem anderen Land geboren sind, schneiden deutlich schlechter ab als die deutschsprachigen Mitschüler. In diesem Landesbericht soll nun untersucht werden, ob sich ähnliche Ergebnisse finden lassen. Ein möglicher Migrationshintergrund wurde erfasst, indem die Fachlehrer die Muttersprache ihrer Schüler angeben sollten. Sie konnten dabei wählen zwischen Deutsch, Englisch, Französisch oder einer anderen Muttersprache. Da es in den Gruppen Englisch bzw. Französisch als Muttersprache zu sehr kleinen Fallzahlen kommt, wurden die Schüler zu zwei Gruppen zusammengefasst: Schüler mit Deutsch als Muttersprache und Schüler mit einer anderen Muttersprache. Nach der Betrachtung des prozentualen Anteils von Schülern mit einer anderen Muttersprache wird der Frage nachgegangen, inwieweit sich Schüler deutscher und nicht-deutscher Muttersprache in ihren Testergebnissen unterscheiden. Es folgen Analysen zur Entwicklung des Muttersprachenefekts bestimmter Jahrgänge (jahrgangsspezifisch) und die Darstellung des Muttersprachenefekts verschiedener Jahrgänge in den Jahren 2005 bis 2010 (fachspezifisch).

**Häufigkeitsverteilung: Muttersprache Deutsch vs. Nicht-Deutsch**. Wie im letzten Jahr ist der Anteil der Schüler, die eine andere Muttersprache als Deutsch zu Hause sprechen in Thüringen relativ gering (vgl. Abbildung 7.2). Am geringsten ist der Prozentanteil in allen drei Klassenstufen in den Förderschulen mit 1,3-2,1 %. Da dies Fallzahlen von 3-7 Schülern entspricht, werden in diesem Kapitel keine weiteren Analysen mit Förderschulen berichtet. Analysen mit so wenigen Personen könnten zu einem verzerrten Bild führen und sind im Vergleich nicht aussagekräftig.

In Klassenstufe 3 ist aufgrund der anderen Muttersprache bei 3,4 % der Kinder ein Migrationshintergrund zu vermuten. Im Gymnasium sprechen in der Klassenstufe 6 1,9 % eine andere Muttersprache, während in Klassenstufe 8 der Anteil mit 2,5 % deutlich höher liegt. Der vergleichsweise höchste Anteil an Schülern mit einem möglichen Migrationshintergrund ist in den Regelschulen zu finden. In Klassenstufe 6 sprechen 2,2 % der Schüler eine andere Muttersprache und in Klassenstufe 8 3,1 %.

[7] vergleiche PISA-Konsortium Deutschland (2008)

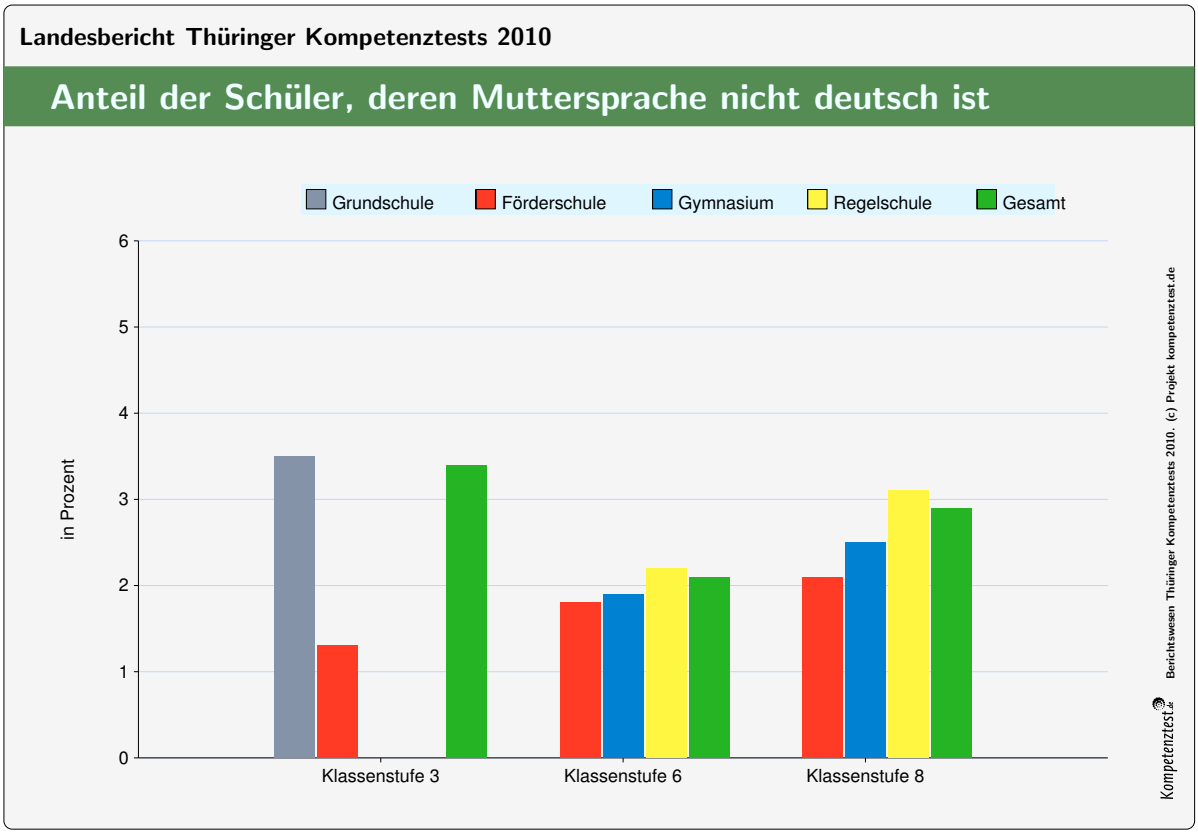


**Abbildung 7.2:** Prozentualer Anteil der Schüler, deren Muttersprache nicht deutsch ist, abhängig von der Schulform.

**Mittelwertevergleich der Testergebnisse**. Um herauszufinden, ob es einen Zusammenhang zwischen der Muttersprache und den Ergebnissen in den Kompetenztests gibt, kann man die Mittelwerte von Schülern deutscher Muttersprache mit den Mittelwerten der Schüler nichtdeutscher Muttersprache vergleichen. In die Analysen im Fach Englisch wurden die Schüler mit Englisch als Muttersprache nicht einbezogen, um die Ergebnisse nicht zu verfälschen. Das betrifft die Klassenstufe 8, in der bei acht Schülern angegeben wurde, dass sie Englisch als Muttersprache sprechen. Analysen mit diesen Schülern würden keine angemessenen Aussagen mehr darüber zulassen, inwiefern sich Schüler mit unterschiedlichen Muttersprachen beim Lernen der Fremdsprache Englisch unterscheiden. Die Ergebnisse zeigen, dass die Schüler mit einer anderen Muttersprache in allen Fächern in jeder Klassenstufe und jeder Schulform schlechter abschneiden als die Schüler mit Deutsch als Muttersprache. Im Detail zeigt sich, dass es sich in Klassenstufe 3 der Grundschulen sowohl in Deutsch als auch in Mathematik um einen kleinen Unterschied zugunsten der Deutsch-Muttersprachler handelt. In gleicher Richtung liegt in Klassenstufe 6 der Regelschulen ein mittelgroßer Effekt in Mathematik und Deutsch vor, während in Englisch ein kleiner Effekt zu finden ist. Sehr viel bessere Leistungen als Schüler mit einer anderen Muttersprache zeigen die Schüler deutscher Muttersprache in den Gymnasien der Klassenstufe 6 und 8 im Fach Deutsch. In Mathematik und Englisch ist der Unterschied in Klassenstufe 6 der Gymnasien in der gleichen Richtung klein, aber vorhanden. In Klassenstufe 8 in den Gymnasien erreichen Deutsch-Muttersprachler im Fach Mathematik mit einem mittelgroßen Effekt eine deutlich höhere Punktzahl als Schüler mit einer anderen Muttersprache. Dieselbe Tendenz ist in Englisch mit einem kleinen Effekt zu finden. Auch in Klassenstufe 8 der Regelschulen unterscheiden sich die Leistungen in Deutsch sehr stark im Sinne eines besseren Ergebnisses der Schüler mit Deutsch als Muttersprache. In Mathematik und Englisch erzielen Deutsch-Muttersprachler ein wenig bessere Leistungen als die Schüler mit einer anderen Muttersprache.

### Entwicklung des Muttersprachenefekts eines Jahrgangs

In den Kompetenztests 2010 zeigt sich, dass deutsche Muttersprachler besser abschneiden als Schüler mit einer anderen Muttersprache. Um den Zusammenhang zwischen Muttersprache und Leistung genauer zu erfassen, lohnt sich die Betrachtung der Entwicklung der Leistungsdifferenz über die Jahre hinweg (jahrgangsspezifische Betrachtung). Dazu wird im Folgenden der Muttersprachenefekt eines Jahrgangs mit den Effekten aus den vergangenen Jahren in Beziehung gesetzt. Es wird die Klassenstufe 6 aus 2008 und die Klassenstufe 8 aus 2010 betrachtet. In der Abbildung 7.5 bedeutet ein positiver Effekt, dass Schüler mit deutscher Muttersprache im Mittel besser abgeschnitten haben als Schüler mit einer anderen Muttersprache.

**Tabelle 7.5:** Leistungsunterschiede eines Jahrgangs in Abhängigkeit von der Muttersprache: 2008 und 2010

| | Muttersprache 2008 Effektstärke d | Muttersprache 2010 Effektstärke d |
|---|---|---|
| **Gymnasium** | | |
| Deutsch | 0,377 | 0,880 |
| Mathematik | 0,321 | 0,738 |
| Englisch | 0,245 | 0,231 |
| **Regelschule** | | |
| Deutsch | 0,657 | 0,897 |
| Mathematik | 0,429 | 0,379 |
| Englisch | 0,318 | 0,462 |

Die Tendenz, dass deutsche Muttersprachler besser abschneiden als Schüler mit einer anderen Muttersprache zeigt sich bei diesem Jahrgang bereits im Jahre 2008. Einige Veränderungen sind dennoch zu beobachten. Im Fach Deutsch vergrößert sich die Differenz zwischen den Leistungen von 2008 zu 2010 sowohl in den Regelschulen als auch in den Gymnasien. Dies trifft bei den Gymnasiasten auch im Fach Mathematik zu. In 2008 war ein kleiner Effekt zu beobachten, während in 2010 ein mittelgroßer Effekt einen größeren Unterschied zwischen den Leistungen anzeigt.

### Der Muttersprachenefekt in den Jahren 2005 bis 2010

In diesem Abschnitt wird der Muttersprachenefekt verschiedener Jahrgänge dargestellt. Die berichteten Effektstärken geben an, inwieweit sich Schüler mit Deutsch als Muttersprache und Schüler mit einer anderen Muttersprache in den verschiedenen Jahrgängen in ihrer Testleistung unterscheiden. Um die Darstellung übersichtlich zu halten, wird auf die Differenzierung nach Schulformen verzichtet. In Klassenstufe 3 wird in jedem Jahr ein bedeutsamer Unterschied in den Leistungen von deutschen Muttersprachlern und anderen Muttersprachlern berichtet. Die Stärke des Effekts schwankt zwischen klein und mittelgroß. In Mathematik zeigt sich in den meisten Jahren ein kleiner Unterschied zugunsten der Schüler mit Deutsch als Muttersprache. Lediglich in den Jahren 2007 und 2005 unterschieden sich die Schülergruppen nicht.

In der Klassenstufe 6 lässt sich für das Fach Deutsch feststellen, dass Schülern mit Deutsch als Muttersprache in jedem Jahr bessere Ergebnisse in den Tests erzielt haben als Schüler mit einer anderen Muttersprache. Die Stärke des Effekts schwankt stark zwischen den Jahren. In Mathematik sind die Unterschiede insgesamt kleiner. So tritt in keinem Jahr ein Unterschied mit einem großen Effekt auf. Dennoch lassen sich in den meisten Jahren Leistungsunterschiede zugunsten der Schüler mit Deutsch als Muttersprache finden. Im Fach Englisch finden sich in

**Tabelle 7.6:** Entwicklung des Muttersprachenefekts in den Kompetenztests der Klassenstufe 3: 2005 bis 2010

| Jahr | Deutsch 3 Effektstärke d | Mathematik 3 Effektstärke d |
|---|---|---|
| 2010 | 0,439 | 0,249 |
| 2009 | 0,510 | 0,318 |
| 2008 | 0,576 | 0,336 |
| 2007 | 0,386 | 0,157 |
| 2006 | 0,631 | 0,353 |
| 2005 | 0,409 | 0,162 |

den Jahren 2006 und 2007 gar keine Unterschiede zwischen den Schülergruppen mit unterschiedlichen Mutttersprachen, während in den anderen Jahren nur kleine Effektstärken zu beobachten sind. Auch diese deuten auf ein besseres Abschneiden der Deutsch-Muttersprachler hin.

**Tabelle 7.7:** Entwicklung des Muttersprachenefekts in den Kompetenztests der Klassenstufe 6: 2005 bis 2010

| Jahr | Deutsch 6 Effektstärke d | Mathematik 6 Effektstärke d | Englisch 6 Effektstärke d |
|---|---|---|---|
| 2010 | 0,629 | 0,443 | 0,314 |
| 2009 | 0,778 | 0,619 | 0,313 |
| 2008 | 0,433 | 0,312 | 0,227 |
| 2007 | 0,340 | 0,154 | 0,153 |
| 2006 | 0,804 | 0,379 | 0,125 |
| 2005 | 0,734 | 0,275 | – |

In Klassenstufe 8 werden die Kompetenztests in Deutsch und Englisch erst seit 2009 geschrieben. Deshalb beschreibt die folgende Tabelle nur in den Jahren die Tests dieser Fächer, während in Mathematik Daten bis zum Jahre 2007 vorliegen. In Deutsch zeigt sich erneut, dass Schüler mit Deutsch als Muttersprache bedeutsam besser abschneiden als Schüler mit einer anderen Muttersprache. In Mathematik und Englisch ist diese Tendenz ebenfalls zu finden, aber nicht in dem Ausmaß. Hier finden sich fast nur Unterschiede mit kleiner Effektstärke.

**Relative Wahrscheinlichkeit für den Gymnasialbesuch**

Die Gymnasialbeteiligung von Schülern mit verschiedenen Muttersprachen wird z.B. in den PISA-Studien als Indikator für die Fairness des Bildungssystems genutzt. Analog dazu soll in diesem Abschnitt der Frage nachgegangen werden, ob die bei der deskriptiven Darstellung des Anteils der Schüler mit einer anderen Muttersprache in den verschiedenen Schulformen gezeigten Unterschiede in der Gymnasialbeteiligung (vgl. Abbildung 7.2) bestehen bleiben, wenn die Leistungen in den Kompetenztests in die Analyse einbezogen werden. Es wird also das potenzielle Leistungsniveau der Schüler konstant gehalten, indem die Ergebnisse in den Mathematik-Tests in der Analyse berücksichtigt werden. Bei den Mathematik-Ergebnissen handelt es sich um relativ sprachfreie Testergebnisse, weshalb sie den Ergebnissen in den Deutsch-Tests vorzuziehen sind. Es zeigte sich, dass bei gleicher Leistung im Kompetenztest

**Tabelle 7.8:** Entwicklung des Muttersprachenefekts in den Kompetenztests der Klassenstufe 8: 2008 bis 2010

| Jahr | Deutsch 8 Effektstärke d | Mathematik 8 Effektstärke d | Englisch 8 Effektstärke d |
|---|---|---|---|
| 2010 | 0,759 | 0,434 | 0,311 |
| 2009 | 0,667 | 0,474 | 0,290 |
| 2008 | | 0,609 | |
| 2007 | | 0,365 | |

Mathematik die Chance für den Gymnasialbesuch für einen Schüler mit nicht-deutscher Muttersprache genauso hoch ist wie für einen Schüler mit deutscher Muttersprache. Das heißt, dass sich bei gleicher Leistung in Mathematik die Chancen für einen Gymnasialbesuch für Schüler beider Muttersprachen-Gruppen nicht bedeutsam unterscheiden. Dieses Ergebnis spricht für Fairness im Bildungssystem. Bei der Interpretation muss jedoch beachtet werden, dass in den oben dargestellten Analysen in jeder Klassenstufe Unterschiede in der Mathematik-Leistung von Schülern mit Deutsch als Muttersprache und Schülern mit einer anderen Muttersprache zu finden waren. Diese kann man nicht ohne weiteres auf Unterschiede in den kognitiven Fähigkeiten der Schüler zurückführen, denn auch in Mathematik ist der Lernerfolg eines Schülers nicht völlig unabhängig von der Sprache. Daher ist anzunehmen, dass ein Schüler mit einer anderen Muttersprache bei gleichen kognitiven Fähigkeiten schlechtere Leistungen erreicht, als ein Schüler mit deutscher Muttersprache. Seine Chancen auf den Besuch eines Gymnasiums fallen damit geringer aus.

**Fazit**

In allen Fächern aller Schulformen in jeder Klassenstufe schneiden Schüler mit einer anderen Muttersprache schlechter ab als ihre Mitschüler deutscher Muttersprache. Die Tatsache, dass sich dieser Effekt durch alle Klassen und Fächer durchzieht, ist bedenklich. Das Ergebnis spricht dafür, dass gezielte Sprachförderung dringend benötigt wird, damit die Schüler mit einer anderen Muttersprache die Möglichkeit erhalten, die gleichen Leistungen wie ihre Mitschüler zu erbringen. Besonders in der Grundschule sollte die schnelle Lernfähigkeit der Kinder genutzt werden, um spätere Sprachdefizite zu vermeiden. In Fach Deutsch sind diese Unterschiede quasi naturgemäß am größten. Interessant ist, dass im Fächervergleich tendenziell der Unterschied in Englisch am geringsten ist. Sprachkenntnisse scheinen auch in Mathematik von essentieller Bedeutung zu sein. Die durchgeführten Analysen bezüglich der relativen Wahrscheinlichkeit für den Gymnasialbesuch sprechen für Fairness im Bildungssystem.

**Das Wichtigste in Kürze**

- In allen Fächern aller Schulformen in jeder Klassenstufe schneiden Schüler mit einer anderen Muttersprache schlechter ab als ihre Mitschüler deutscher Muttersprache.
- Die durchgeführten Analysen bezüglich der relativen Wahrscheinlichkeit für den Gymnasialbesuch sprechen für Fairness im Bildungssystem.

### 7.2.3 Sozioökonomischer Status (SES)

Ein weiterer wichtiger Einflussfaktor, der regelmäßig im Rahmen der PISA-Studie betrachtet wird, ist der Zusammenhang zwischen sozialer Herkunft und Schülerkompetenz. Die PISA-Untersuchung von 2006[8] konnte erneut zeigen, dass es in allen OECD-Staaten einen bedeutsamen Zusammenhang zwischen dem sozioökonomischen Status (abgekürzt aus dem Englischen: SES) des Elternhauses der teilnehmenden Jugendlichen und den erreichten Kompetenzen gibt. In Bezug darauf weist Deutschland einen über dem Durchschnitt liegenden Zusammenhang auf, d. h. die Stärke und Steigung des sozioökonomischen Gradienten liegt in Deutschland über dem Durchschnitt der OECD-Länder[9]. Jedoch zeigt sich bei der Betrachtung der Veränderung in den sozialen Unterschieden seit PISA 2000 bis PISA 2006 am Beispiel der Lesekompetnz in Deutschland, dass sich der Zusammenhang zwischen sozialer Herkunft und Lesekompetenz für deutsche Schüler abgeschwächt hat[10]. Trotz der Tatsache, dass die verbleibenden Unterschiede weiter als hoch zu benennen sind, zeigt dieser positive Trend, dass es sich lohnt, auf Bildungsungerechtigkeiten aufmerksam zu machen, um einen Anstoß zu fördernden Maßnahmen zu liefern.

Innerhalb der PISA-Erhebung wird die soziale und kulturelle Herkunft der Schüler über eine Vielzahl von Indikatoren erfasst. Eine solche Handhabung ist aus forschungsmethodischer Sicht sehr wünschenswert. Aber sie ist auch mit einem großen Aufwand für Schüler und Lehrer verbunden. Im Zuge der Kompetenztests ist eine solche Vorgehensweise nicht realisierbar. Deshalb muss im Kontext der Kompetenztests auf sparsame Methoden zur Erfassung des sozioökonomischen Status als Maß für die soziale Herkunft der Schüler zurückgegriffen werden.

Im internationalen Grundschulvergleich IGLU[11] wurde der sozioökonomische Status u. a. mit Hilfe der so genannten „Bücherfrage“ erhoben. Die Bücherfrage misst auf einer fünfstufigen Skala die Anzahl der im Elternhaus vorhandenen Bücher. Sie hat sich als ein nützlicher Indikator für den SES erwiesen und ist somit ein gutes Forschungsinstrument, um den sozialen Hintergrund zu erfassen. Wie auch im vergangenen Jahr wurde die Bücherfrage im Kompetenztest anonymisiert erhoben. Daraus ergibt sich, dass für jede Klasse ein Mittelwert des sozioökonomischen Status vorliegt, welcher im Rahmen von Gruppenanalysen als Indikator für die soziale Herkunft genutzt wird. Dementsprechend sind sämtliche Ergebnisse dieses Kapitels auf Klassenebene zu interpretieren. Es ist wichtig zu beachten, dass es insbesondere in der Klassenstufe 3 zu methodischen Verzerrungen kommen kann, weil die Schüler die Bücheranzahl eventuell schlecht schätzen können. Deswegen kann die Bücherfrage nicht als genaues Abbild des sozialen Hintergrunds gesehen werden, sondern sie liefert lediglich Hinweise auf den sozioökonomischen Status der Schüler. Daraus ist zu folgern, dass die dargestellten Befunde ausschließlich als Gesamtergebnis interpretiert werden sollten und nicht auf Einzelfälle angewendet werden können. Vielmehr zeigen sie gewisse Trends auf und sollten von daher nicht überbewertet werden.

Zunächst soll untersucht werden, wie stark sich die verschiedenen Schulformen in der mittleren Bücheranzahl unterscheiden. Dazu erfolgt der Vergleich der mittleren Bücherzahl der verschiedenen Schulformen mittels Effektstärken. Eine detaillierte Übersicht zu den Effektstärken befindet sich in den Tabellen im Anhang zum Landesbericht unter der Überschrift „Vergleiche der Mittelwerte der Bücherfrage (SES) in Klassenstufen“. In der Klassenstufe 3 ist die Effektstärke d=1,08 ähnlich groß wie die im vergangenen Jahr, d. h. die Grundschulen und die Förderschulen unterscheiden sich im Mittel stark hinsichtlich der Angabe zur Bücherfrage.

[8] vergleiche PISA-Konsortium Deutschland (2007)
[9] vergleiche PISA 2006: Naturwissenschaftliche Kompetenzen für die Welt von morgen - Kurzzusammenfassung (2007)
[10] vergleiche PISA-Konsortium Deutschland (2007)
[11] vergleiche Bos (2003)

Die Grundschulklassen erreichen einen höheren Mittelwert in der Bücherfrage als die Förderschulklassen. Deswegen ist zu schlussfolgern, dass der SES in Grundschulen über dem in Förderschulen liegt. Die Ergebnisse in den Klassenstufen 6 und 8 zeigen ebenfalls große Effektstärken. Dies bedeutet, dass sich die Gymnasien, Regelschulen und Förderschulen stark in der Anzahl der Bücher im Elternhaus unterscheiden. Besonders die Gymnasialklassen stechen durch ihren hohen Mittelwert in der Bücherfrage hervor und liegen sowohl deutlich über den mittleren Werten der Regelschulklassen (d=1,71 in Kassenstufe 6 und d=2,12 in Klassenstufe 8) als auch über den Mittelwerten der Förderschulklassen (d=2,41 in Klassenstufe 6 und d=2,21 in Klassenstufe 8). Die beiden letztgenannten Schülergruppen unterscheiden sich dagegen kaum voneinander (d=0,17 in Klassenstufe 8). Aufgrund der aufgezeigten Differenzen kann davon ausgegangen werden, dass sich die Schularten in ihrem Klientel und deren sozialen Herkunft unterscheiden. Diese Auswertung kann als Hinweis dafür genommen werden, dass die soziale Herkunft eine große Rolle bei dem Besuch der Schulart spielt.

Nachdem gezeigt wurde, dass sich die verschiedenen Schularten hinsichtlich des sozioökonomischen Status ihrer Klassen stark unterscheiden, soll nun analysiert werden, inwieweit sich der SES auf die Leistung im Kompetenztest auswirkt. Hierfür wird der Zusammenhang zwischen SES und der Testleistung getrennt nach Schularten in Form von Korrelationen auf Klassenebene betrachtet. Bei Korrelationen handelt es sich um Zusammenhangsmaße. In den vorliegenden Berechnungen bedeuten positive Korrelationen, dass ein höherer SES mit einer höheren Testleistung einhergeht. Negative Korrelationen symbolisieren dagegen, dass ein höherer SES mit einer niedrigeren Testleistung einhergeht. An diesem Punkt sei darauf hingewiesen, dass die Neigung besteht Korrelationen in einen kausalen Zusammenhang zu bringen. Solche Ursache-Wirkungsbeziehungen lassen sich aber aus Korrelationen nicht ableiten, weil auch noch andere Faktoren wie beispielsweise die Muttersprache oder Stimmung die Testleistung beeinflussen können.

Die folgenden Tabellen geben die Korrelation r für die einzelnen Kompetenztests auf Klassenebene getrennt nach Schularten wieder. Die dargestellten Korrelationen sind in der Klassenstufe 3 ähnlich wie im vergangenen Jahr als klein zu bewerten. In den Klassenstufen 6 und 8 ist der Gesamtzusammenhang zwischen der Testleistung und dem SES äquivalent zum letzten Jahr als mittelgroß bis groß einzuschätzen. Wie zu erwarten, sind die einzelnen Korrelationen, getrennt nach Schularten betrachtet, wesentlich kleiner als die Gesamtkorrelation. Eine mögliche Schlussfolgerung ist, dass die Schularten die Unterschiede im SES widerspiegeln. Das bedeutet, dass die Selektion der Schüler in Regelschulen, Gymnasien und Förderschulen entsprechend des sozialen Hintergrunds erfolgt und demzufolge der Zusammenhang kleiner wird. Dieses Ergebnis könnte einen Anstoß für mehr Förderung geben, so dass auch die Schüler aus sozial schwachen Familien die Möglichkeit erhalten entsprechend ihrer Fähigkeiten einen höheren Bildungsgang in Anspruch zu nehmen.

**Tabelle 7.9:** Korrelation des SES mit den Testergebnissen (Klassenstufe 3)

| | **Grundschule** | **Förderschule** | **Gesamt** |
|---|---|---|---|
| $r_{SES-DK3}$ | 0,30 | −0,07 | 0,32 |
| $r_{SES-MK3}$ | 0,30 | −0,05 | 0,31 |

**Tabelle 7.10:** Korrelation des SES mit den Testergebnissen (Klassenstufe 6)

| | Gymnasium | Regelschule | Förderschule | Gesamt |
|---|---|---|---|---|
| $r_{SES-DK6}$ | 0,32 | 0,32 | 0,50 | 0,67 |
| $r_{SES-MK6}$ | 0,27 | 0,27 | 0,42 | 0,66 |
| $r_{SES-EK6}$ | 0,25 | 0,31 | −0,09 | 0,64 |

**Tabelle 7.11:** Korrelation des SES mit den Testergebnissen (Klassenstufe 8)

| | Gymnasium | Regelschule | Förderschule | Gesamt |
|---|---|---|---|---|
| $r_{SES-DK8}$ | 0,32 | 0,49 | 0,35 | 0,75 |
| $r_{SES-MK8}$ | 0,27 | 0,42 | 0,52 | 0,72 |
| $r_{SES-EK8}$ | 0,33 | 0,37 | 0,06 | 0,70 |

Aus den positiven Korrelationen geht hervor, dass die Klassen mit höherem SES unabhängig von der Klassenstufe und Schulart in allen Tests deutlich besser abschneiden. Da die Klassen mit einer geringeren Anzahl an Büchern im Elternhaus schwächere Leistungen zeigen, könnte man vermuten, dass diese Schüler weniger Unterstützung und Förderung aus dem Elternhaus erfahren. Deshalb ist es wichtig, solchen Klassen frühzeitig eine zusätzliche Unterstützung angeboten wird. Eine Möglichkeit wäre, ein spezielles Konzept für die entsprechenden Klassen in Zusammenarbeit mit dem Unterstützungssystem der Schulämter zu entwickeln, um die Lernmotivation und Kompetenzentwicklung zu fördern.

Die Abbildungen 7.3, 7.4 und 7.5 veranschaulichen die geographische Verteilung des 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 weisen höhere Werte auf einen höheren SES hin.[12] Dargestellt sind jeweils die aggregierten Werte auf Ebene der Schulamtsbereiche. Jedoch sind die dargebotenen Unterschiede zwischen den Regionen (Klassenstufe 3 zwischen 3,17 und 3,41; Klassenstufe 6 zwischen 3,16 und 3,5 und in der Klassenstufe 8 zwischen 3,15 und 3,49) als eher klein zu deuten. Bei der Betrachtung und Interpretation ist es ratsam darauf zu achten, welche größeren Städte sich in dem Schulamtsbezirk befinden. Städte wie zum Beispiel Jena nehmen einen nicht unentscheidenden Einfluss auf den Mittelwert den SES. Die Übersicht soll dabei helfen die Regionen zu ermitteln, in denen noch mehr Unterstützung und Förderbedarf bei den Schülern besteht.

**Das Wichtigste in Kürze**

- die Schularten unterscheiden sich deutlich hinsichtlich des sozioökonomischen Hintergrunds ihrer Schüler
- Klassen mit einem geringeren sozioökonomischen Status erreichen in den Kompetenztests schwächere Leistungen als Klassen mit einem hohen SES

[12] 1 = keine oder nur sehr wenige Bücher, 2 = genug Bücher, um ein Regalbrett zu füllen, 3 = genug Bücher, um ein Regal zu füllen, 4 = genug Bücher, um zwei Regale zu füllen und 5 = über 200 Bücher

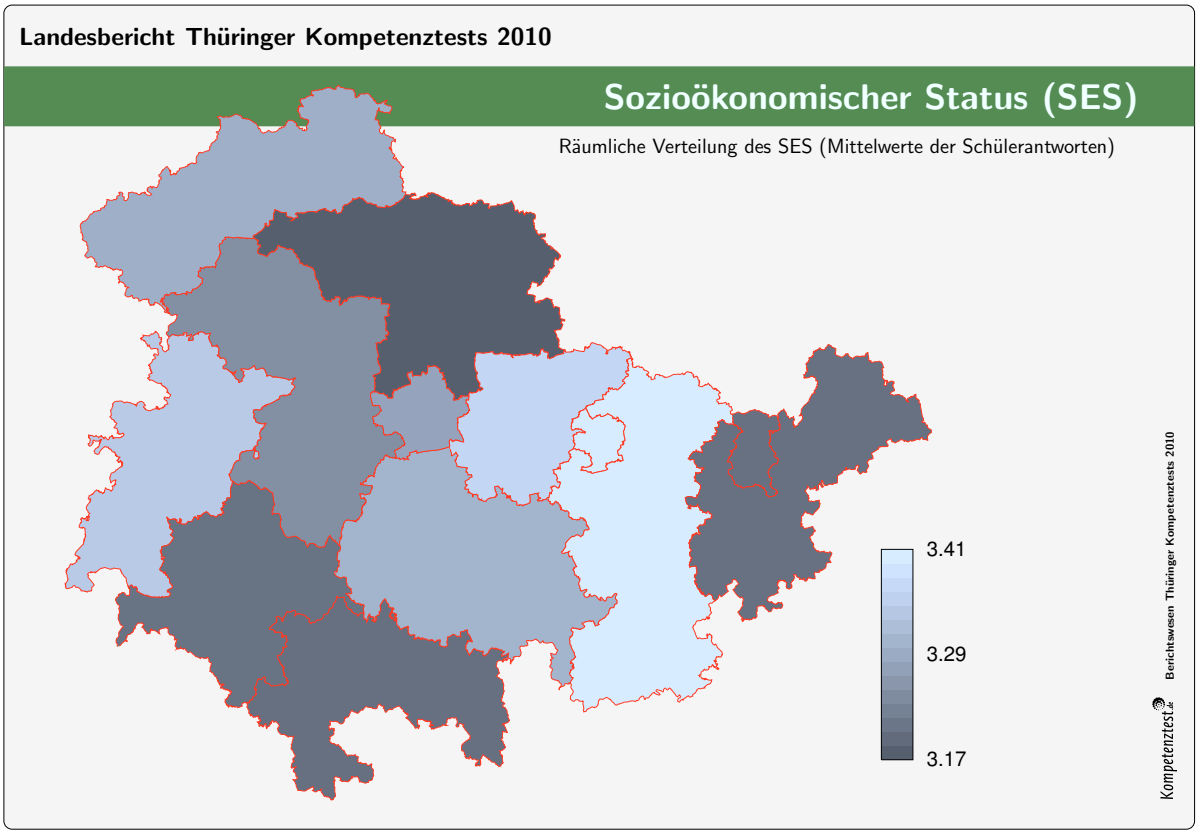


**Abbildung 7.3:** Räumliche Verteilung des SES bei den Schülern der Klassenstufe 3

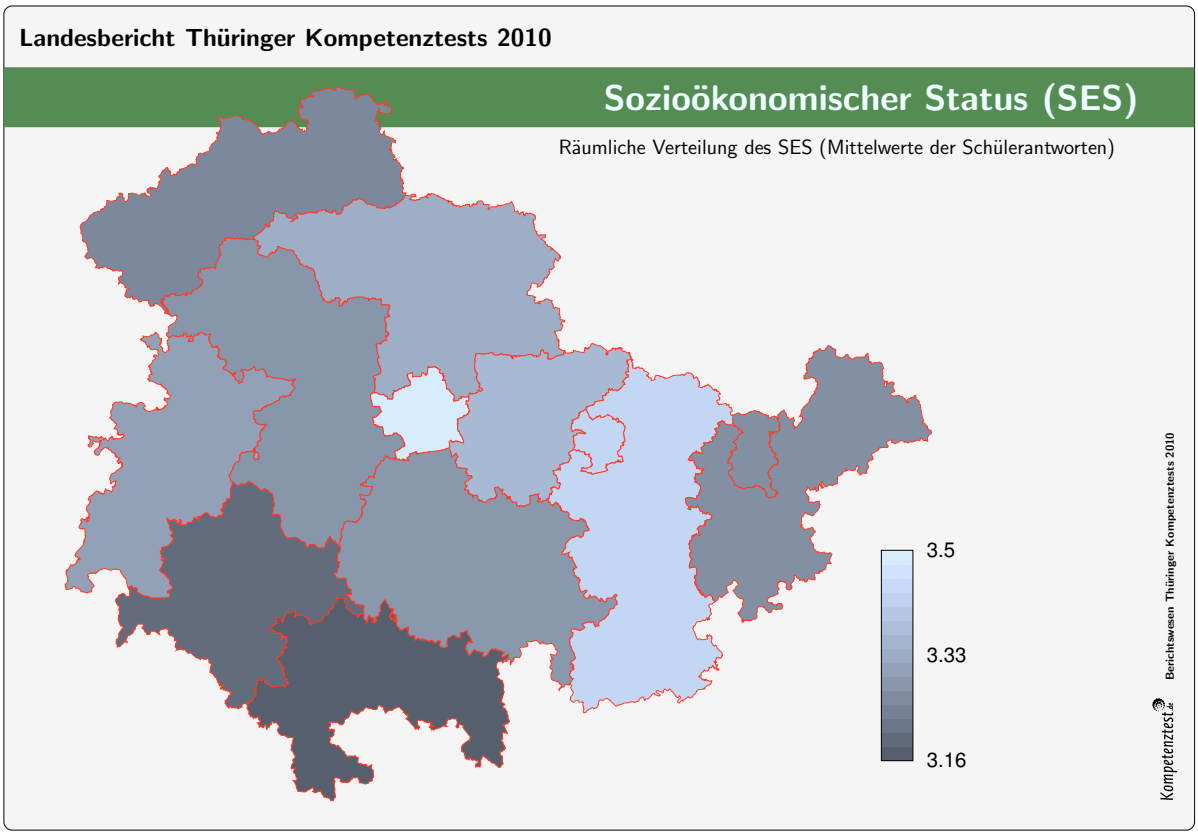


**Abbildung 7.4:** Räumliche Verteilung des SES bei den Schülern der Klassenstufe 6

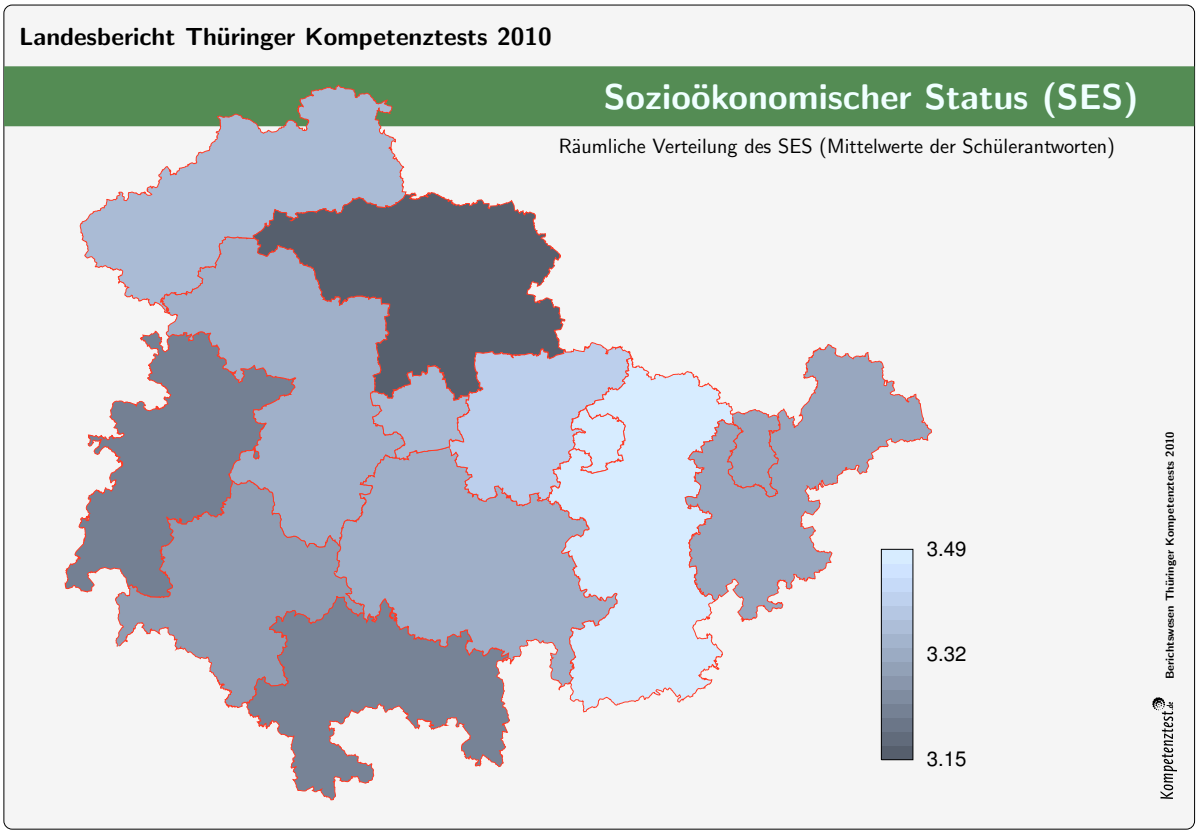


**Abbildung 7.5:** Räumliche Verteilung des SES bei den Schülern der Klassenstufe 8

### 7.2.4 Jahre des Schulbesuchs

Dieser Abschnitt untersucht, inwieweit die Anzahl der absolvierten Schuljahre einen Einfluss auf die Kompetenzen von Schülern hat. Dafür werden Schüler, die die Schuleingangsphase verkürzt haben und solche, die ein Jahr wiederholen mussten, genauer analysiert. Für die genauen Werte der Berechungen sei auf den Anhang zum Landesbericht verwiesen.

**Verkürzen der Schuleingangsphase**

Die hier vorgestellten Ergebnisse zum Verkürzen der Schuleingangsphase beziehen sich nur auf Schüler der Grundschulen. Grundsätzlich ist festzustellen, dass Schüler, die in ihrem zweiten Schuljahr bereits die Klassenstufe 3 besuchen, in den Kompetenztests bessere Leistungen zeigen als ihre Mitschüler. Besonders deutlich fällt dieser Leistungsunterschied im Fach Mathematik (d=0,51) aus. Dieser Befund deckt sich mit den Ergebnissen der letzten Jahre. Schüler, die ihre Schuleingangsphase verkürzen konnten, zeichnen sich durch überdurchschnittliche Leistungen im Unterricht aus. Somit ist es nicht verwunderlich, wenn dieses Schüler auch bei den Kompetenztests mehr Punkte erzielen als ihre Mitschüler.

**Entwicklung der Schüler mit verkürzter Schuleingangsphase**

Dank der eindeutigen Zuordnung der Schülercodes zu den Schülern über die Jahre hinweg ist es möglich, längsschnittliche Analysen durchzuführen. In diesem Abschnitt interessieren besonders jene Schüler, die in der Klassenstufe 3 im Jahr 2005 als Schüler mit zwei Jahren Schulbesuch angegeben waren. Für 22 dieser Schüler lassen sich die entsprechenden Kompetenztestergebnisse der Klassenstufen 6 zuordnen. In der Klassenstufe 8 sind noch 16 der Schüler mit verkürzter Schuleingangsphase registriert. Es zeigt sich, dass diese Schüler nach der Grundschule sehr häufig ans Gymnasium wechseln. Im Jahr 2008 waren 61 % der Schüler

mit verkürzter Schuleingangsphase an einem Gymnasium. Von den Schülern, die sich im Jahr 2005 regulär in ihrem dritten Schuljahr befanden, wechselten nur 39 % auf ein Gymnasium. Sogar zwei Jahre später finden sich diese Verteilungen noch (Klassenstufe 8 im Jahr 2010).

Natürlich stellt sich die Frage, ob die Schüler mit zwei Jahren Schulbesuch in der Klassenstufe 3 aus dem Jahr 2005 auch in den folgenden Kompetenztests noch leistungsstärker sind als ihre Mitschüler. Die Analysen zeigen, dass dieser Effekt tatsächlich über die Jahre bestehen bleibt. Die Schüler mit verkürzter Schuleingangsphase der Klassenstufe 3 sind bei den Kompentenztests Deutsch und Mathematik in der Klassenstufe 6 etwas leistungsstärker als ihre Mitschüler, die in der Grundschule die Schuleingangsphase nicht verkürzen konnten. Dieser Effekt verstärkt sich sogar noch in den Kompetenztests Deutsch und Mathematik der Klassenstufe 8 im Jahr 2010. Das Verkürzen der Schuleingangsphase scheint also ein guter Prädiktor für langfristig überdurchschnittlich gute Schulleistungen zu sein. Daraus lässt sich schlussfolgern, dass die Kriterien für die Auswahl von Schülern zur Verkürzung der Schuleingangsphase offensichtlich sinnvoll festgelegt sind.

**Wiederholung einer Klassenstufe**

Global betrachtet findet sich der Effekt, dass Schüler, die aufgrund von schwächeren Leistungen eine oder mehrere Klassenstufen wiederholen mussten, in den Kompetenztests leistungsschwächer sind als ihre Mitschüler. Dieser Befund trifft aber nicht auf alle Schularten, Klassenstufen und Fächer gleichermaßen zu und soll deshalb noch differenzierter betrachtet werden. Die Grafiken 7.6 , 7.7 und 7.8 stellen den grundsätzlichen Effekt nach Klassenstufen getrennt dar. Nachfolgend werden die Unterschiede zwischen den verschiedenen Schularten und Fächern berichtet.

Bei der Betrachtung der Schüler der Klassenstufe 3 fällt auf, dass sich Grundschüler deutlich von Förderschülern unterscheiden. Während nur etwa 8 % der Grundschüler vier oder mehr Jahre die Schule besucht haben, sind es bei den Förderschülern 87 %. Eine Gemeinsamkeit beider Schularten besteht darin, dass selten mehr als vier Jahre Schulbesuch angegeben wurden. Schüler mit vier oder mehr Jahren Schulbesuch in Grundschulen sind sowohl in Deutsch (d=−0,85) als auch in Mathematik (d=−0,91) leistungsschwächer als ihre Mitschüler. In den Förderschulen zeigt sich dagegen ein anderes Bild: Hier sind die Schüler mit vier oder mehr Jahren Schulbesuch nicht leistungsschwächer, sondern in Mathematik sogar etwas leistungsstärker als ihre Mitschüler (d=0,35). Diese Tatsache spricht dafür, dass leistungsschwache Schüler in Förderschulen besser in den Unterricht integriert und unterstützt werden als in Grundschulen. Dies ist möglich, weil die Klassenzusammensetzung in den Förderschulen eher homogen ist. In den Grundschulen dagegen bestehen die Klassen sowohl aus sehr guten Schülern, die später auf ein Gymnasium wechseln werden, als auch aus weniger leistungsstarken Schülern, die anschließend eine Regelschule besuchen werden. Aufgrund dieser heterogenen Klassenzusammensetzung ist es möglicherweise schwerer, auf leistungsschwächere Schüler ausreichend einzugehen.

Die Beobachtung, dass sich das Wiederholen eines Schuljahres in den Förderschulen kaum auf die Leistungen in den Kompetenztests auswirkt, zeigt sich auch in der Klassenstufe 6 und verliert sich dann in der Klassenstufe 8. In der Klassenstufe 6 sind die Förderschüler, die eine Klasse wiederholen mussten, in allen Fächern etwas leistungsstärker als ihre Mitschüler. In der Klassenstufe 8 finden sich keinerlei Unterschiede zwischen Wiederholern und Schülern mit 8 Jahren Schulbesuch. Ganz anders stellen sich die Effekte für die anderen Schularten dar. Regelschüler, die eine Klassenstufe wiederholen mussten, sind in den Kompetenztests in allen Fächern leistungsschwächer als ihre Mitschüler. Dieser typische Effekt findet sich sowohl in Klassenstufe 6 als auch in Klassenstufe 8. Bei den Gymnasien ist zu beachten, dass es kaum Wiederholer gibt. Unter 2 % aller Gymnasiasten mussten ein Schuljahr wiederholen. Diese geringe Zahl lässt sich damit erklären, dass leistungsschwächere Schüler eher vom Gymnasium

auf die Regelschule wechseln, statt eine Klassenstufe zu wiederholen. In der Klassenstufe 6 sind die Wiederholer besonders in Deutsch (d=−0,87) und Mathematik (d=−0,59) leistungsschwächer als ihre Mitschüler. In Englisch (d=−0,27) zeigen sich dagegen sehr viel geringere Unterschiede. Dieses Muster verliert sich in der Klassenstufe 8, denn dort erreichen die Gymnasiasten, die ein Schuljahr wiederholen mussten, in allen Fächern durchschnittlich geringere Punktzahlen als ihre Mitschüler.

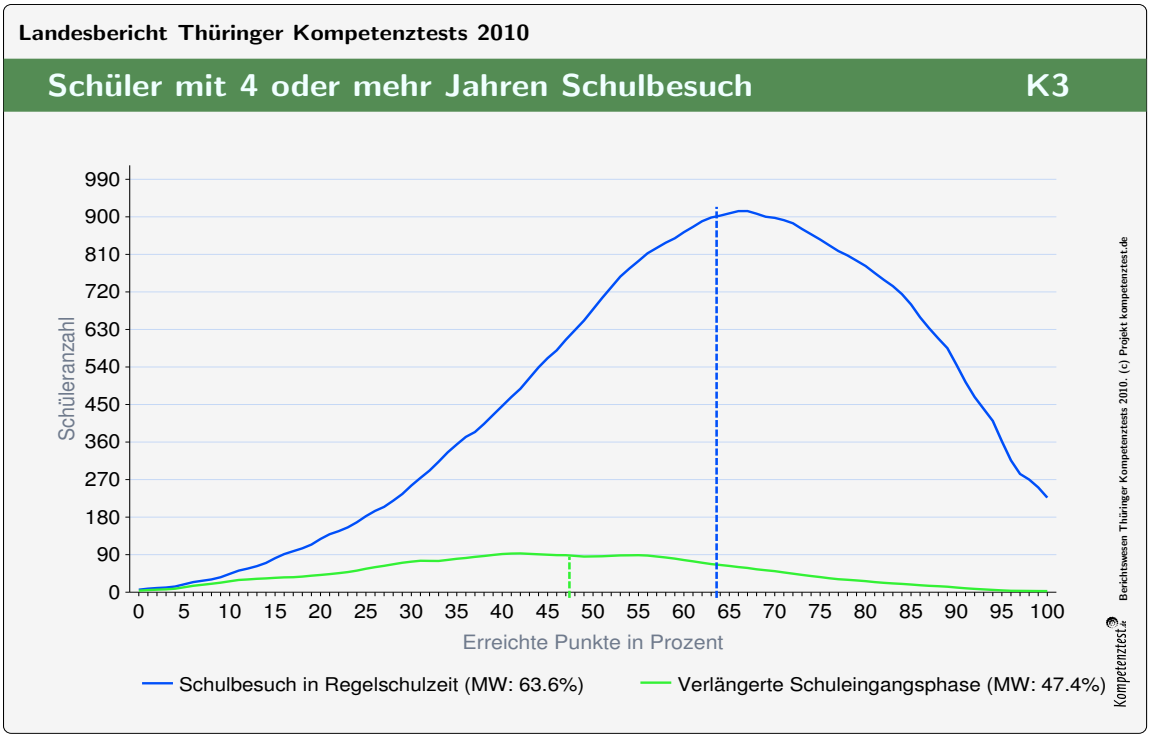


**Abbildung 7.6:** Vergleich der Testergebnisse Klassenstufe 3 zwischen Schülern mit 3 Jahren Schulbesuch und Schülern mit 4 oder mehr Jahren Schulbesuch

### Entwicklung der Schüler mit verlängerter Schuleingangsphase

Für rund 400 der Schüler, bei denen in der Klassenstufe 3 im Jahr 2005 angegeben wurde, dass sie vier oder mehr Jahre Schulbesuch haben, liegen die Kompetenztestergebnisse aus der Klassenstufe 6 vor. Für die Klassenstufe 8 lassen sich noch fast 300 Schüler mit ehemals verlängerter Schuleingangsphase zuordnen. Diese Schüler wechselten nach der Grundschule zum großen Teil auf eine Regelschule. In der Klassenstufe 6 im Jahr 2008 werden 85 % der Schüler mit vier oder mehr Jahren Schulbesuch von 2005 an Regelschulen unterrichtet. Von den Schüler, die im Jahr 2005 regulär drei Schuljahre absolviert hatten, sind im Jahr 2008 nur 58 % an einer Regelschule. Zwei Jahre später, also in der Klassenstufe 8, lernen sogar 88 % der Schüler mit ehemals verlängerter Schuleingangsphase an einer Regelschule. Die Zahl der Regelschüler, die in der Grundschule regulär drei Jahre Schulbesuch absolviert hatten, erhöht sich in der Klassenstufe 8 ebenfalls leicht und liegt bei 62 %.

Interessant ist selbstverständlich die Frage, ob die Schüler mit vier oder mehr Jahren Schulbesuch aus der Klassenstufe 3 vom Jahr 2005 auch in den folgenden Kompetenztests leistungsschwächer abschneiden als ihre Mitschüler. In den Analysen der Ergebnisse der Kompetenztests Deutsch und Mathematik der Klassenstufen 6 und 8 zeigen sich sehr starke Effekte, die dafür sprechen, dass dies der Fall ist. In allen vier untersuchten Kompetenztests erbringen die Schüler mit verlängerter Schuleingangsphase deutlich schwächere Leistungen als ihre Mitschüler. Es lässt sich schlussfolgern, dass sich die Tatsache, in der Klassenstufe 3 bereits vier oder mehr Jahre Schulbesuch absolviert zu haben, in den Schulleistungen der folgenden Jahre wiederspiegelt.

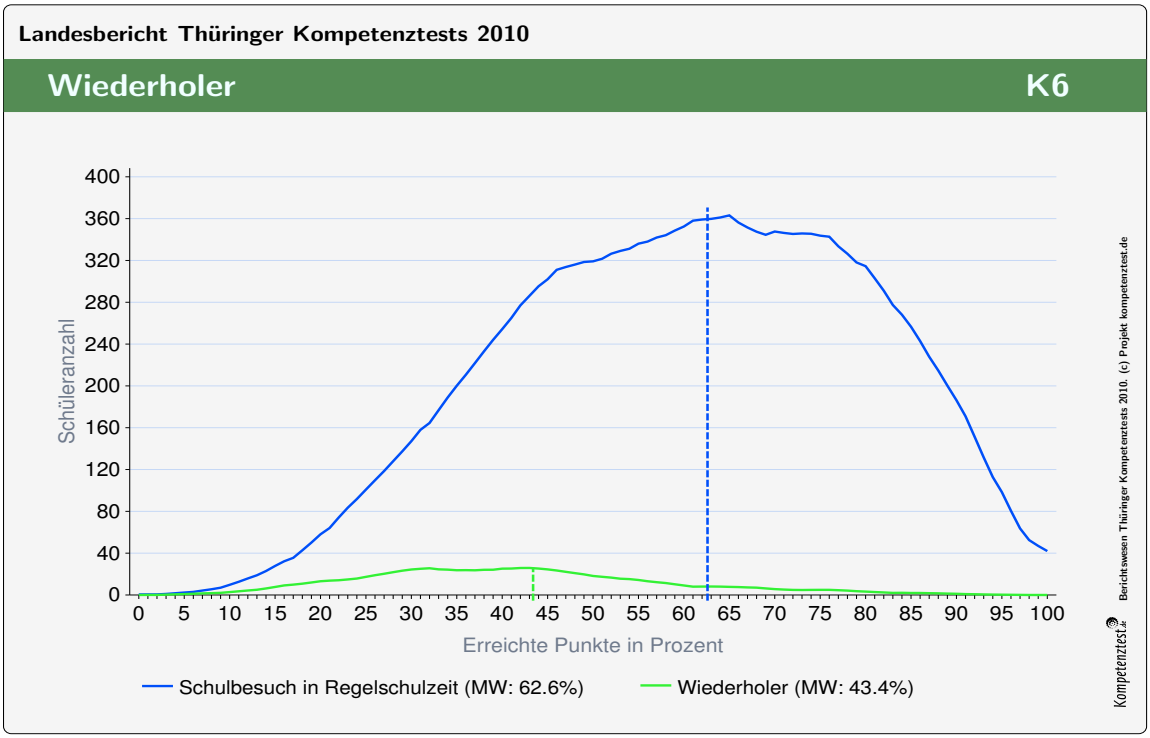


**Abbildung 7.7:** Vergleich der Testergebnisse Klassenstufe 6 zwischen Schülern mit und ohne Klassenwiederholung

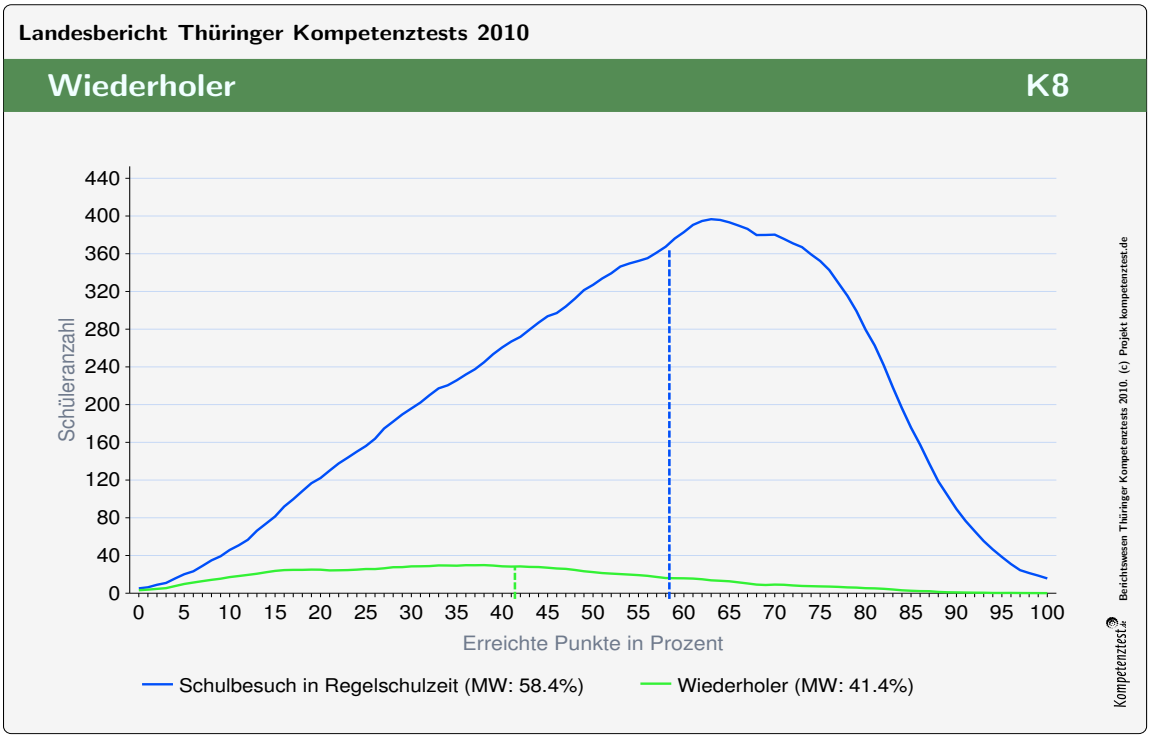


**Abbildung 7.8:** Vergleich der Testergebnisse Klassenstufe 8 zwischen Schülern mit und ohne Klassenwiederholung

### Räumliche Verteilung von Klassenwiederholern

In den Grafiken 7.9 bis 7.16 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. In der Klassenstufe 3 gibt es bei den Grundschulen kaum Unterschiede zwischen den Regionen, aber bei den Förderschulen zeigen sich deutliche regionale Differenzen. Eisenach, Artern und Weimar zählen zu den Regionen mit den wenigsten Schülern mit vier oder mehr Jahren Schulbesuch an Förderschulen. Rudolstadt, Schmalkalden und Bad Langensalza haben dagegen besonders viele Förderschüler, die ein ihre Schuleingangsphase verlängern mussten.

In der Klassenstufe 6 sind die regionalen Unterschiede bei den Gymnasien und Regelschulen weniger stark ausgeprägt. Viel mehr lohnt es sich, den Anteil an Wiederholern in den Förderschulen zu betrachten. Die Regionen Eisenach, Artern und Schmalkalden zeichnen sich durch eine besonders geringe Anzahl an Wiederholern aus. Auffällig viele Schüler, die eine Klassenstufe wiederholen mussten, finden sich in den Förderschulen der Regionen Jena/Stadtroda, Worbis und Weimar.

Auch in der Klassenstufe 8 fallen die regionalen Unterschiede bei den Gymnasien und Regelschulen gering aus. In den Förderschulen gibt es wieder ausgeprägte Ungleichheiten zwischen den Regionen. In Weimar und Rudolstadt ist der Anteil der Förderschüler, die eine Klasse wiederholen mussten, am geringsten. Besonders viele Wiederholer sind in den Förderschulen der Regionen Schmalkalden, Eisenach und Erfurt zu finden.

**Das Wichtigste in Kürze**

- Grundschüler, die die Schuleingangsphase verkürzen konnten, schneiden in den Kompetenztests leistungsstärker ab als ihre Mitschüler
- Grundschüler, Regelschüler und Gymnasiasten, die ein Schuljahr wiederholen mussten, sind in den Kompetenztests leistungsschwächer als ihre Klassenkammeraden

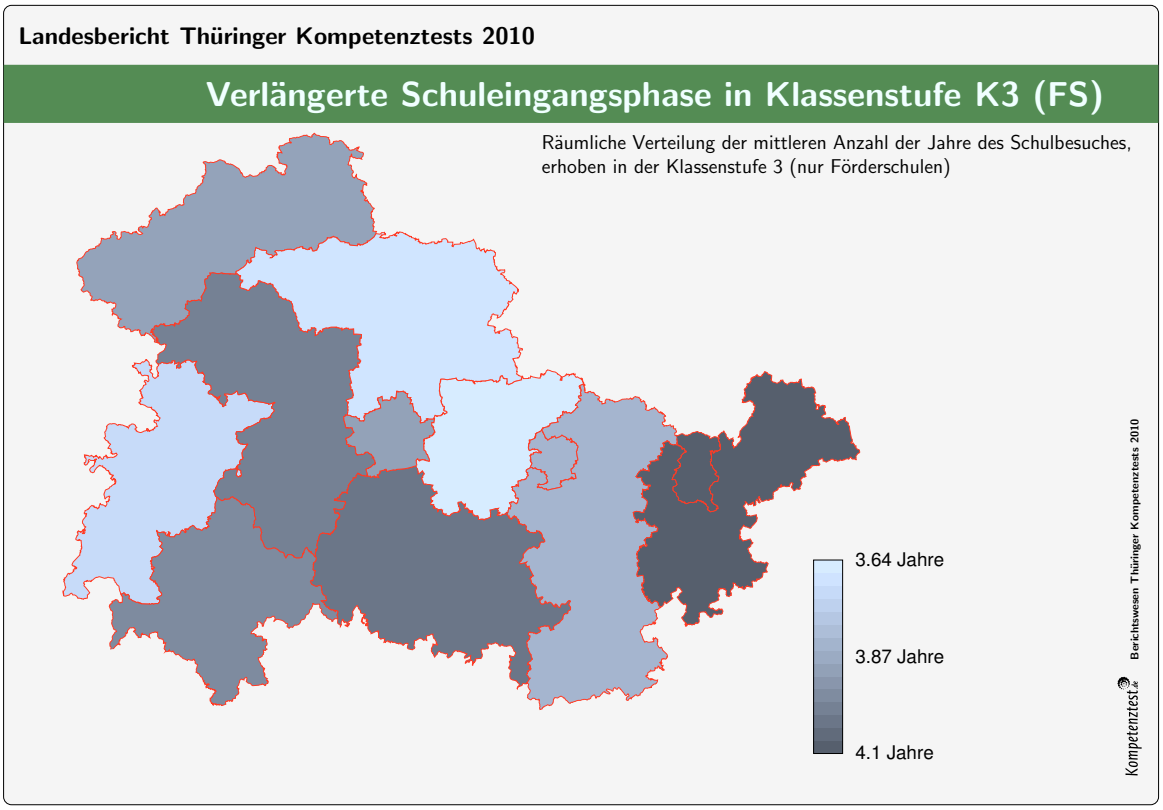


**Abbildung 7.9:** Räumliche Verteilung der mittleren Anzahl der Jahre des Schulbesuches, erhoben in der Klassenstufe 3 (nur Förderschulen). Für den Schulamtsbezirk Neuhaus können aufgrund der geringen Fallzahlen keine Angaben gemacht werden.

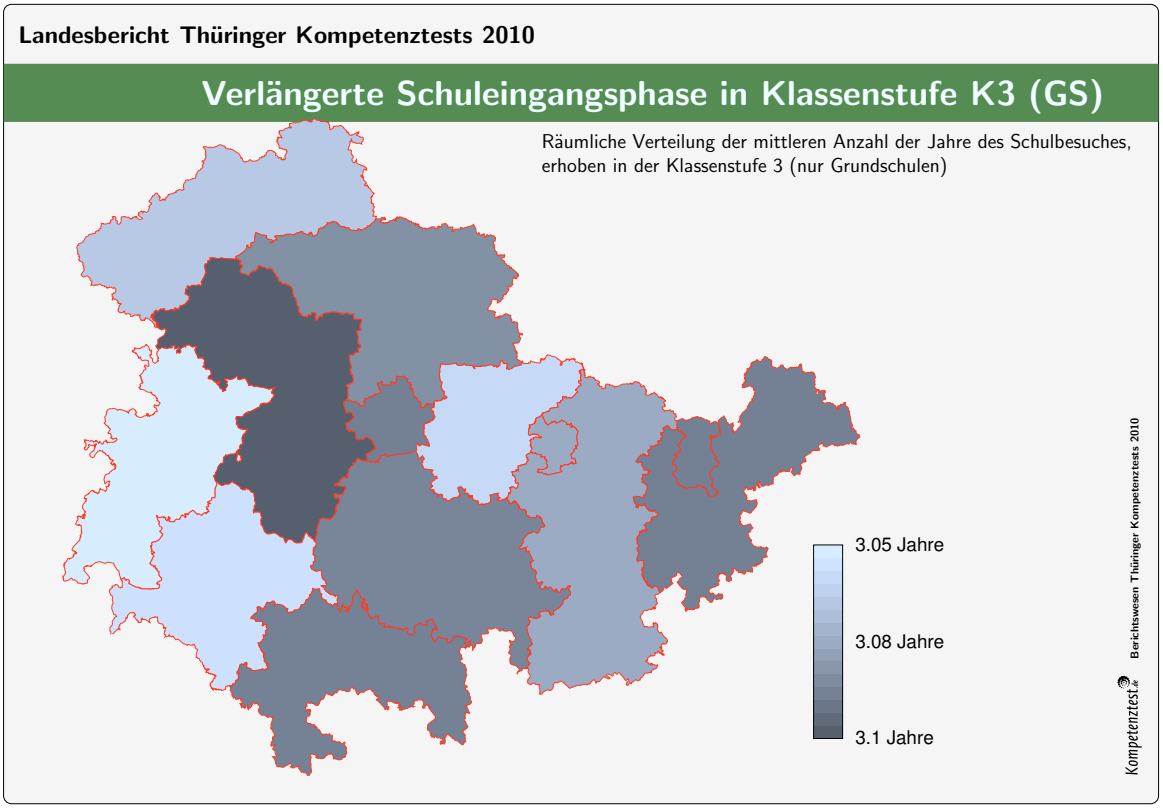


**Abbildung 7.10:** Räumliche Verteilung der mittleren Anzahl der Jahre des Schulbesuches, erhoben in der Klassenstufe 3 (nur Grundschulen)

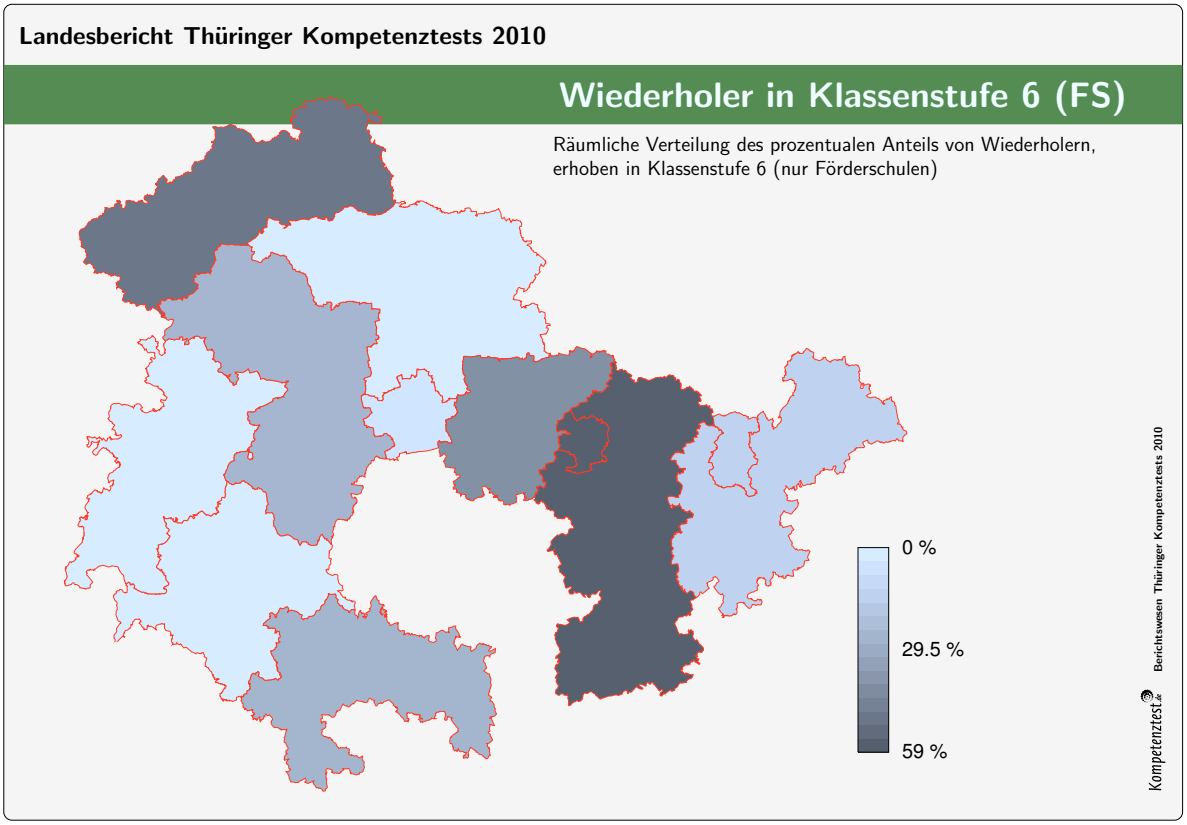


**Abbildung 7.11:** Räumliche Verteilung des prozentualen Anteils von Wiederholern, erhoben in der Klassenstufe 6 (nur Förderschulen)

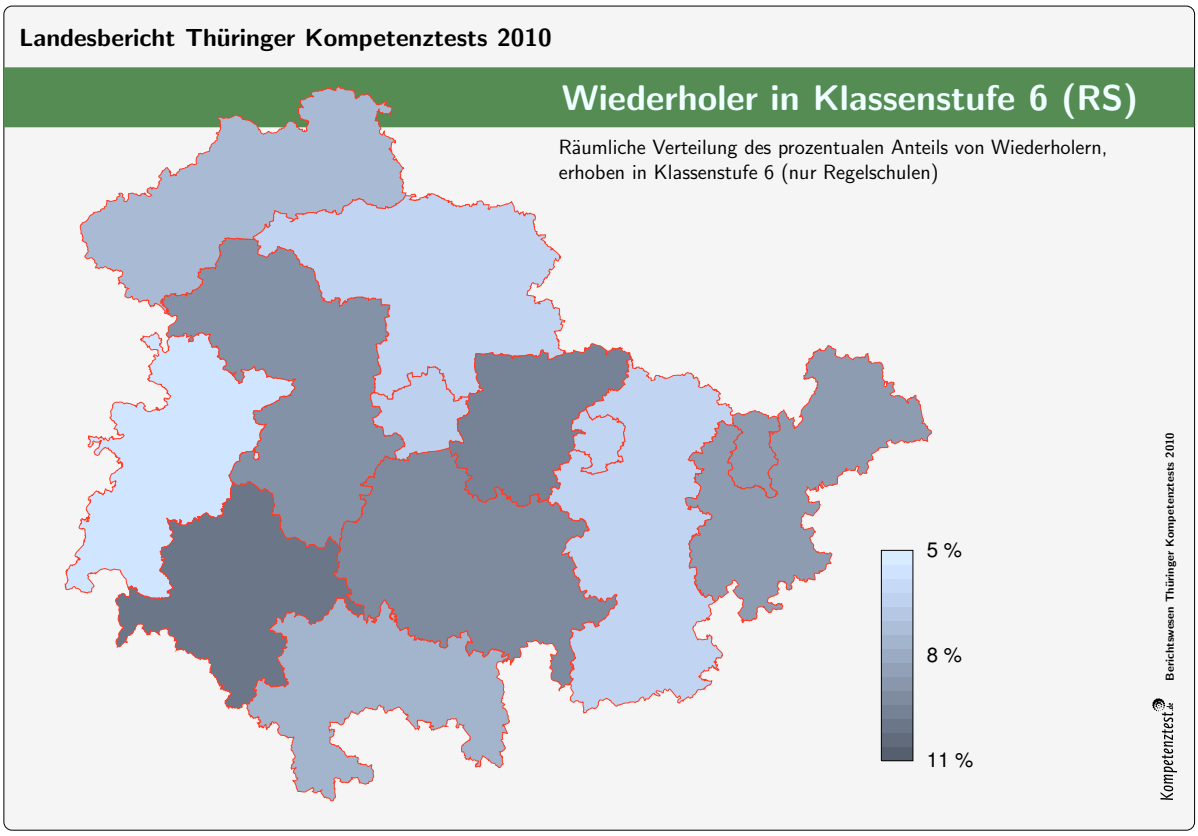


**Abbildung 7.12:** Räumliche Verteilung des prozentualen Anteils von Wiederholern, erhoben in der Klassenstufe 6 (nur Regelschulen)

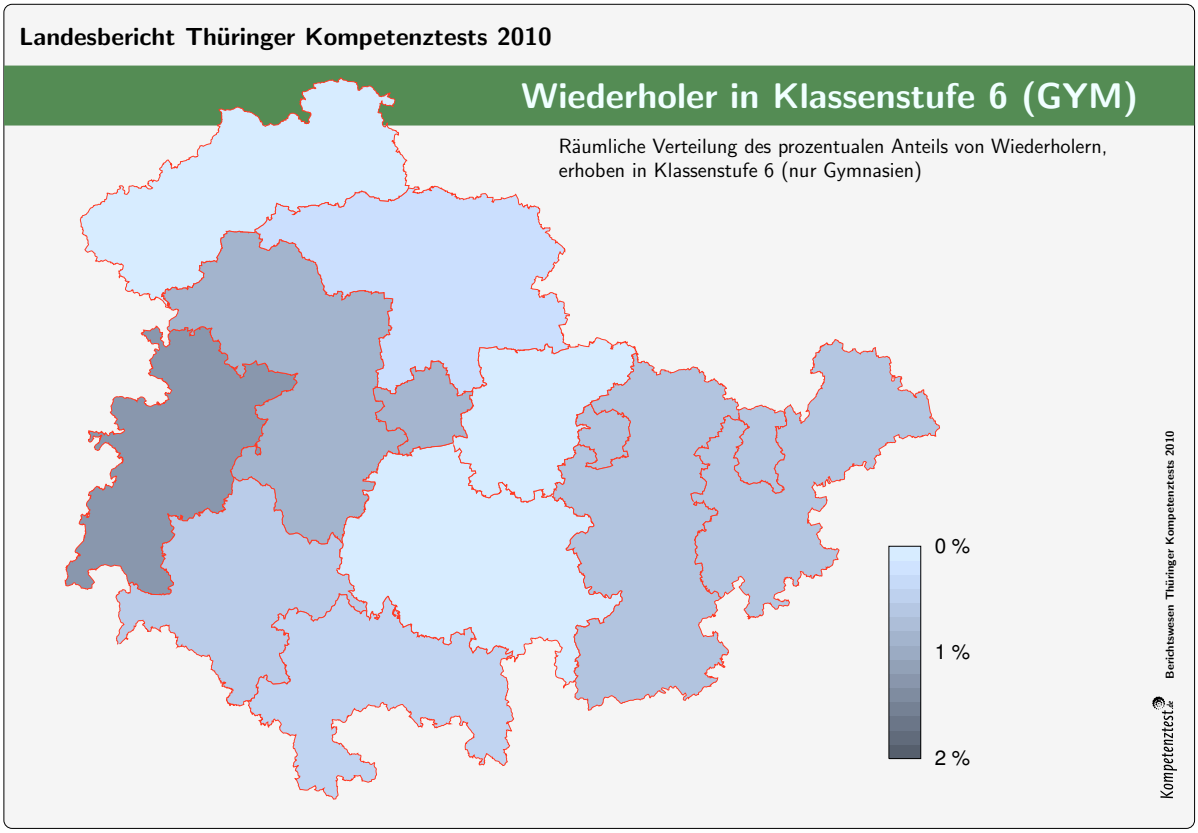


**Abbildung 7.13:** Räumliche Verteilung des prozentualen Anteils von Wiederholern, erhoben in der Klassenstufe 6 (nur Gymnasien)

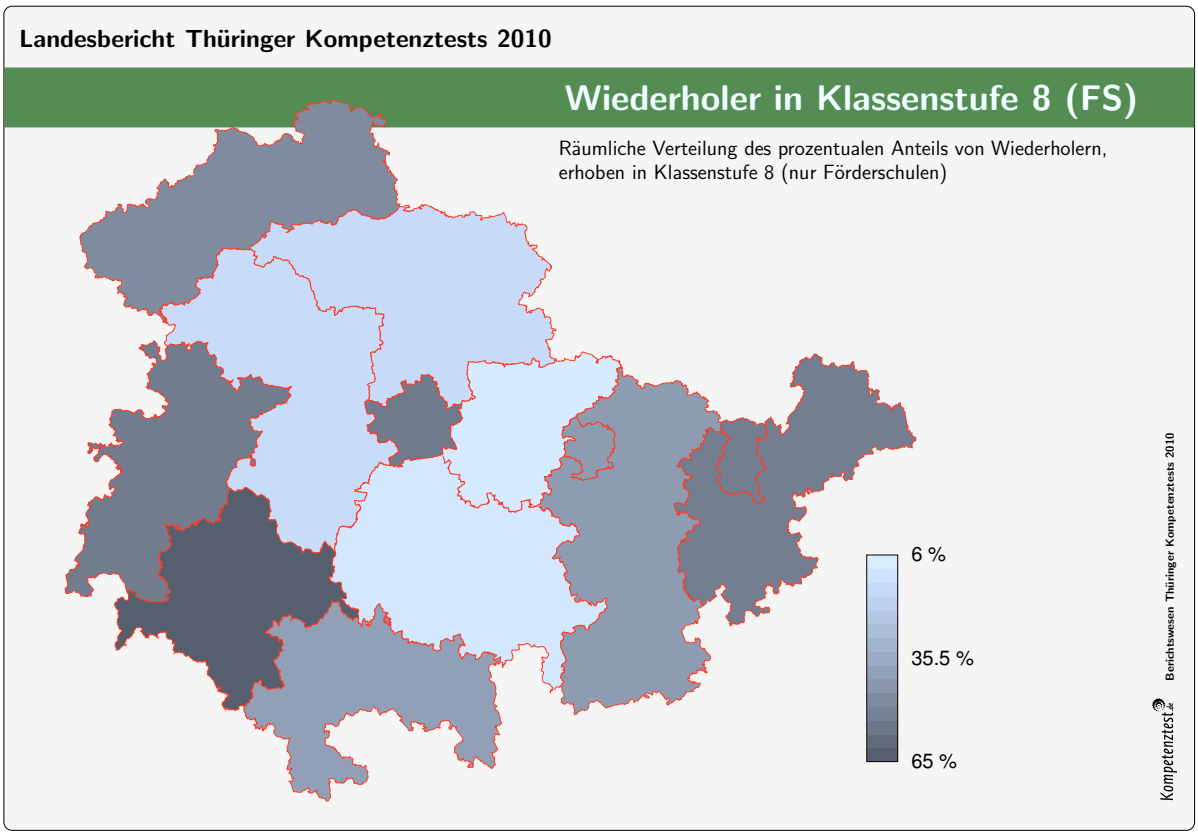


**Abbildung 7.14:** Räumliche Verteilung des prozentualen Anteils von Wiederholern, erhoben in der Klassenstufe 8 (nur Förderschulen)

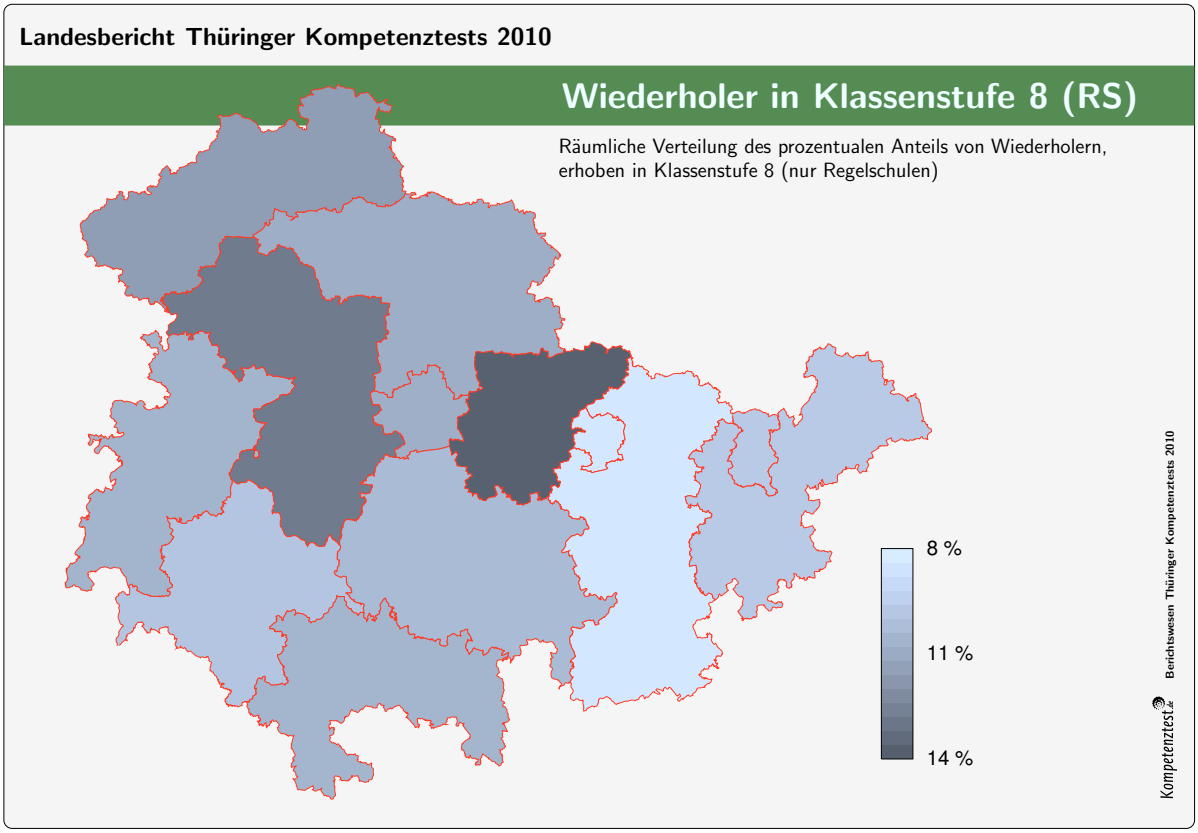


**Abbildung 7.15:** Räumliche Verteilung des prozentualen Anteils von Wiederholern, erhoben in der Klassenstufe 8 (nur Regelschulen)

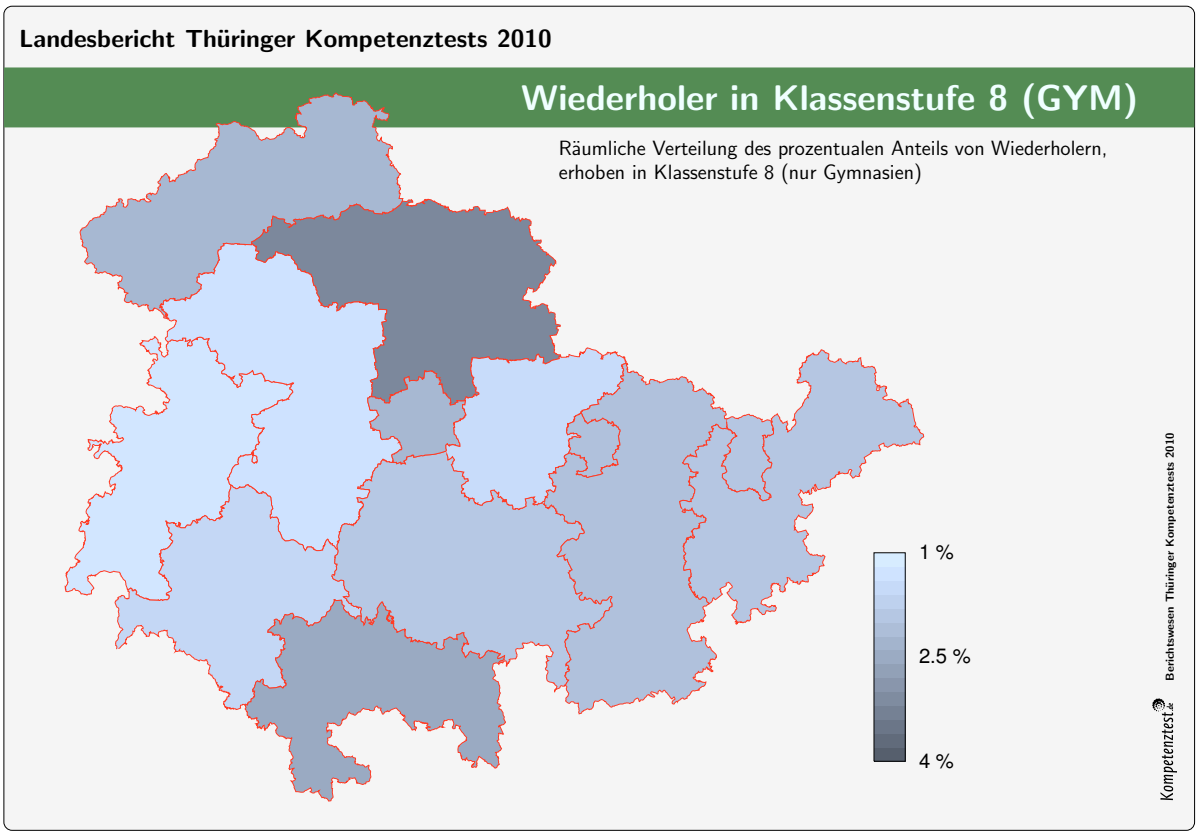


**Abbildung 7.16:** Räumliche Verteilung des prozentualen Anteils von Wiederholern, erhoben in der Klassenstufe 8 (von Gymnasien)

### 7.2.5 Besondere Lernschwierigkeiten

Wie bereits in den vergangenen Jahren wurden diagnostizierte Lernschwierigkeiten in die Analyse der Kompetenztests einbezogen. Es 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) und soziale Entwicklung (SF6).

Die Fachlehrer für Mathematik, Deutsch, Englisch und Französisch wurden gebeten bei der Dateneingabe anzugeben, ob bei einem Schüler eine besondere Lernschwierigkeit oder ein sonderpädagogischer Förderbedarf vorliegt. Stellt man die Angaben der Fachlehrer gegenüber, sind Abweichungen zu entdecken, die darauf hinweisen, dass ein noch intensiverer Informationsaustausch bezüglich der besonderen Lernschwierigkeiten bzw. dem sonderpädagogischen Förderbedarf vorteilhaft wäre. Im Folgenden sollen die typischen Auffälligkeiten berichtet werden, wobei im Anhang enthaltene ausführliche Kreuztabellen noch ergänzende Informationen liefern. In der Klassenstufe 3 fällt beim Vergleich der Angaben der Deutsch- und Mathematiklehrer auf, dass die Deutschlehrer häufiger über eine Lernschwierigkeit im Bereich Schriftspracherwerb informiert sind als die Mathematiklehrer. Insgesamt gaben die Deutschlehrer bei 7,0 % der Schüler eine Lernschwierigkeit in diesem Bereich an. Die Mathematiklehrer geben lediglich bei 38,6 % von diesen Schülern ebenfalls diese Lernschwierigkeiten an, während sie bei 31,4 % dieser Schüler keine Auffälligkeit diagnostiziert haben. Diese Abweichungen wurden beispielhaft nochmal in Abbildung 7.17 dargestellt. Unter der Voraussetzung, dass der Deutschlehrer eine Lernschwierigkeit im Bereich Schriftspracherwerb diagnostiziert hat, ist die Verteilung der Angaben des Mathematiklehrers zu sehen. Besonders relevant ist der Anteil der Schüler, bei dem gar keine Lernschwierigkeit vom Mathematiklehrer diagnostiziert wird. Wird eine andere Lernschwierigkeit angegeben (z.B. BLR) als vom Deutschlehrer, wird diese vermutlich als relevanter eingeschätzt.

Es zeigt sich, dass die Mathematiklehrer wiederum besser über eine Lernschwierigkeit im Bereich Rechnen informiert sind. Diese wurde bei 4,9 % der Schüler festgestellt, während die Deutschlehrer diese nur bei 1,3 % der Schüler diagnostizierten. Bei 38,8 % stimmen die Angaben beider Fachlehrer in diesem Bereich überein. Jedoch geben die Deutschlehrer bei 34,7 % der Schüler an, bei denen der Mathematiklehrer eine Lernschwierigkeit im Bereich Rechnen diagnostiziert, dass der Schüler keinerlei Beeinträchtigung beim Lernen hat. Diese relativ hohe Zahl kann auch dadurch nicht erklärt werden, dass die Deutschlehrer eine andere diagnostizierte Lernschwierigkeit als wichtiger einschätzen, denn in diesem Fall hätten Sie diese bei der Ergebniseingabe angeben können. Ähnliche Abweichungen zeigen sich auch bei den Deutsch- und Mathematiklehrern in Klassenstufe 6 und Klassenstufe 8. In der Klassenstufe 6 ist der Mathematiklehrer bei insgesamt 2 % der Schüler über eine besondere Lernschwierigkeit im Bereich Rechnen informiert, während der Deutschlehrer sogar bei 45 % von diesen Schülern nicht ausreichend darüber unterrichtet ist. Vergleicht man die Angaben der Deutsch- und Englischlehrer in Klassenstufe 6 zeigt sich, dass die Deutschlehrer bei 3,6 % eine Lernschwierigkeit im Bereich Schriftspracherwerb diagnostizieren, aber Englischlehrer bei 38,9 % von dieser Schülern nicht hinreichend informiert sind. Umgekehrt ist festzustellen, dass bei 29 % nur der Englischlehrer kundig ist. Immerhin stimmen die Angaben der Fachlehrer bei 52,8 % der Schüler überein. Das entspricht fast genau dem Wert von 52,9 % in Klassenstufe 8, in der es diesbezüglich keine abweichenden Auffälligkeiten zu berichten gibt. Die Angaben zum sonderpädagogischen Förderbedarf stimmen hingegen größtenteils gut überein, so dass diese hier nicht gesondert ausgewertet werden. Nur in Einzelfällen sind manche Fachlehrer nicht auf dem gleichen Kenntnisstand bezüglich der Problembereiche einzelner Schüler wie ihre Kollegen. Die Ergebnisse weisen darauf hin, dass ein intensiverer Informationsaustausch

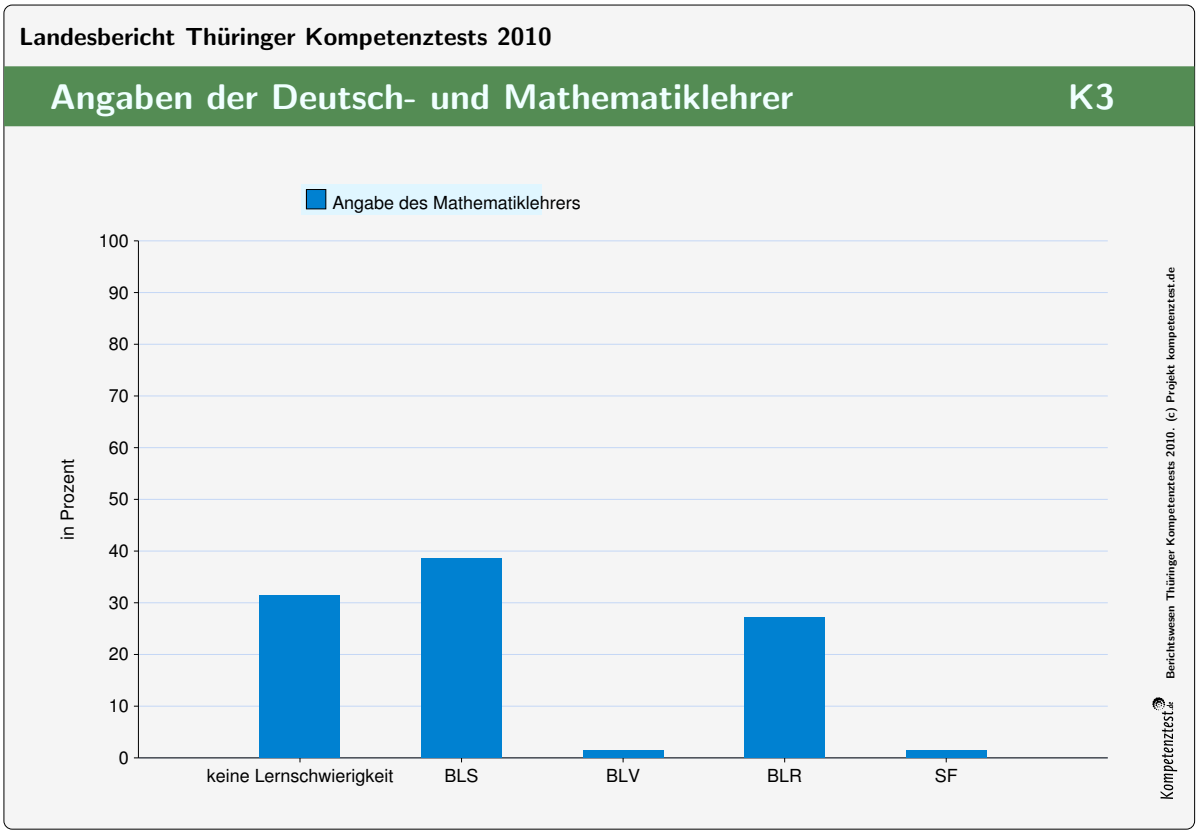


**Abbildung 7.17:** Angaben der Mathematiklehrer, wenn der Deutschlehrer BLS diagnostiziert

bezüglich diagnstizierten Lernschwierigkeiten zwischen den Fachlehrern wünschenswert wäre. Die Umstimmigkeiten sind vergleichbar mit den Befunden im letzten Jahr und weisen eher eine leicht vergrößerte Diskrepanz auf. Es ist deswegen von großer Bedeutung, dass die Kommunikation zwischen den Fachlehrern weiterhin gefördert und intensiviert wird, auch um im Unterricht angemessen auf die Bedürfnisse der Schüler eingehen zu können. Ein Schüler mit einer Lernschwierigkeit im Bereich Schriftspracherwerb wird beispielsweise auch bei Textaufgaben in Mathematik beeinträchtigt sein, weswegen es wichtig ist, dass auch der Mathematiklehrer darüber informiert ist. Ein geringer Teil der Abweichungen kann möglicherweise damit erklärt werden, dass die Lehrer nur eine Lernschwierigkeit bzw. einen sonderpädagogischen Förderbedarf bei der Ergebniseingabe im Internet angeben konnten. Schüler, die in mehreren Bereichen beeinträchtigt sind, wurden folglich nur mit der bedeutsamsten Lernschwierigkeit vermerkt.

Weiterhin wurden die Leistungen in den Kompetenztests von Schülern mit besonderen Lernschwierigkeiten oder sonderpädagogischen Förderbedarf verglichen mit den Leistungen von Schülern ohne diagnostizierte Lernschwierigkeit. Hierbei dient immer die Angabe des jeweiligen Fachlehrers als Differenzierungsgrundlage. So werden Schüler zum Beispiel bei den Kompetenztests im Fach Deutsch entsprechend den Angaben des Deutschlehrers den Gruppen zugeordnet. Aufgrund zu geringer Schülerzahlen können in einigen Fällen keine Unterschiede berichtet werden, weil die statistischen Analysen verzerrt sein könnten. In Folge dessen können zum Beispiel keine Ergebnisse für das Fach Französisch in Klassenstufe 8 angegeben werden. Im folgenden Abschnitt sollen jeweils die auffälligen und wichtigen Ergebnisse herausgegriffen werden. Die im Anhang zum Landesbericht vorhandene Tabellen können jedoch noch detailliertere Informationen für alle analysierten Ergebnisse liefern.

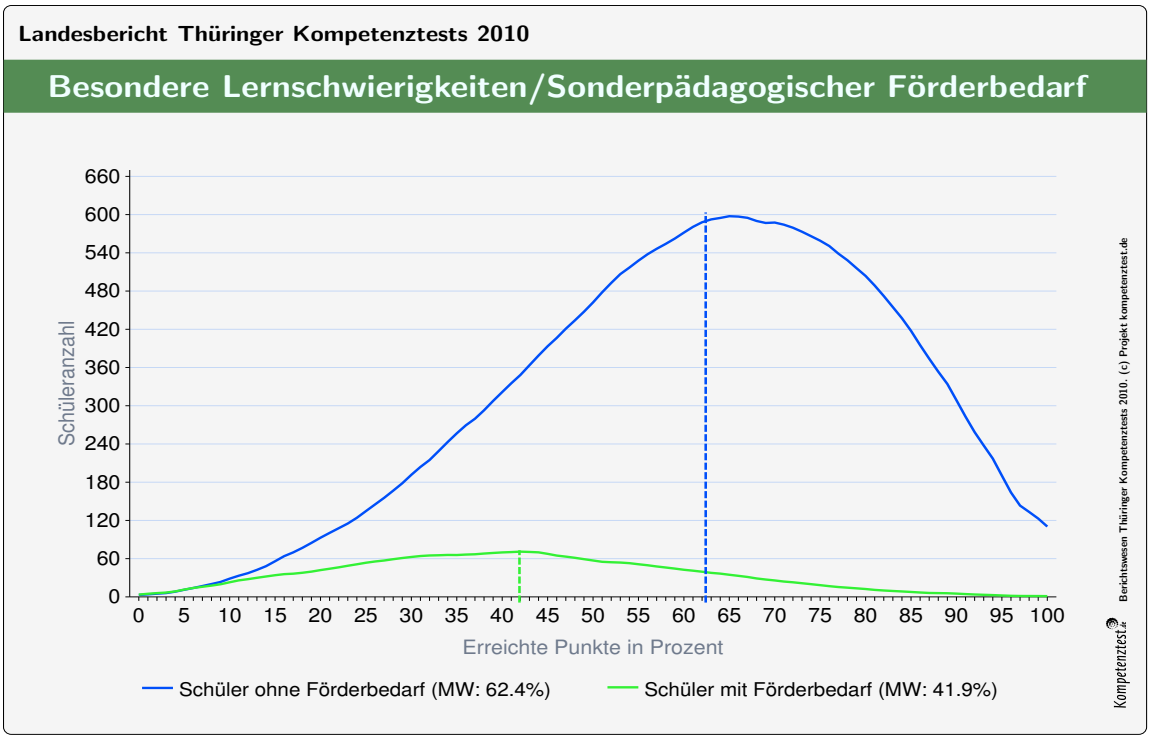


**Abbildung 7.18:** Vergleich der Testergebnisse von Schülern mit und ohne einer besonderen Lernschwierigkeit bzw. einem sonderpädagogischen Förderbedarf über alle Klassenstufen und Fächer hinweg

Es zeigt sich, dass Schüler mit einer besonderen Lernschwierigkeit oder sonderpädagogischen Förderbedarf deutlich schwächer in den Kompetenztests abschneiden als ihre Mitschüler. Bei einem Vergleich der Testergebnisse über alle Klassenstufen und Fächer hinweg liegt eine Effektstärke von d=1,14 vor, was einem großen Effekt entspricht [13]. Schüler mit einer Beeinträchtigung erreichen im Durchschnitt also nur 41,9 % der Gesamtpunktzahl, während diejenigen ohne Lernschwierigkeiten 62,4 % der Gesamtpunktzahl erzielen. In Abbildung 7.18 ist dieser bedeutsame Unterschied zwischen den Mittelwerten nochmal veranschaulicht. Die im Mittel erreichte Punktzahl ist in Prozent der Gesamtpunktzahl für die beiden Gruppen dargestellt [14].

Bei einer näheren Betrachtung der einzelnen Klassenstufen kann festgehalten werden, dass die Effekte in jeder Stufe annähernd gleich groß sind. Vergleicht man jedoch die Unterschiede in den Fächern Deutsch, Mathematik und Englisch fällt auf, dass Schüler mit einer besonderen Lernschwierigkeit vor allem im Fach Englisch unterlegen sind. Hier liegt eine Effektstärke von d=1,3 vor. In den Fächern Deutsch und Mathematik werden in der Klassenstufe 6 und 8 im Vergleich dazu kleinere, wenn auch immer noch bedenkliche Effekte vorgefunden. Eine Ausnahme im Fach Englisch stellt das Gymnasium in der Klassenstufe 6 dar. Hier schneiden die Schüler mit einer diagnostizierten Lernschwierigkeit nur geringfügig schwächer ab als ihre Mitschüler, was sich in einer kleinen Effektstärke von d=0,3 wiederspiegelt. Bedenkt man, dass letztes Jahr ein großer Effekt vorlag, ist das ein positiv zu bewertender Trend. Eine weitere Auffälligkeit findet sich in Klassenstufe 6 in den Förderschulen. Während im Gymnasium und der Regelschule mittlere positive Effekte im Fach Deutsch vorliegen, findet sich in den Förderschulen eine mittlere negative Effektstärke von d=–0,62. Das bedeutet, dass Schüler mit einer Lernschwierigkeit im Kompetenztest im Fach Deutsch im Mittel besser abschneiden als ihre Mitschüler ohne eine diagnostizierte Lernschwierigkeit. Bei diesem Effekt ist jedoch zu berücksichtigen, dass nur eine ziemlich geringe Anzahl von Schülern ohne besondere Lernschwierigkeit die Förderschulen

[13] siehe 7.2 für nähere Erläuterungen zur Effektstärke und deren Interpretation
[14] siehe Kapitel 7.2 für nähere Erläuterungen zur Darstellung der Effektstärke

der Klassenstufe 6 besucht und es dadurch womöglich zu einer kleinen Verzerrung kommen kann. Hervorzuheben ist außerdem, dass in Klassenstufe 8 am Gymnasium in den Fächern Deutsch und Mathematik nur kleine Effektstärken vorliegen und somit die Unterschiede zwischen den Testergebnissen der Schüler mit einer besonderen Lernschwierigkeit und deren ohne eine Beeinträchtigung nur gering sind.

Interessant ist auch die isolierte Betrachtung der Schüler, die über einen sonderpädagogischen Förderbedarf verfügen und derer, bei denen der Fachlehrer eine besondere Lernschwierigkeit angegeben hat. Dabei kann festgestellt werden, dass Schüler mit einem sonderpädagogischen Förderbedarf in Klassenstufe 6 im Fach Deutsch schwächer in den Kompetenztests abschneiden, als Schüler mit einer besonderen Lernschwierigkeit. Im Fach Mathematik und Englisch findet sich diese Auffälligkeit allerdings für die Klassenstufe 6 nicht. Auch in Klassenstufe 3 sind die Unterschiede zwischen diesen beiden Schülergruppen nicht so groß, allerdings finden sich wieder auffällige Abweichungen in Klassenstufe 8 in allen 3 Fächern.

Zusammenfassend ist über alle Klassenstufen und Fächer hinweg zu erkennen, dass sowohl besondere Lernschwierigkeiten als auch ein sonderpädagogischer Förderbedarf einen starken Einfluss auf die Ergebnisse der Kompetenztests haben. Die Schüler mit einer diagnostizierten Lernschwierigkeit erreichen im Durchschnitt eine geringere Punktzahl als ihre Mitschüler. Die Kompetenztests dienen hier also auch als ein Diagnoseinstrument. Da alle Schüler die Tests unter den gleichen Bedingungen schreiben, ist es zu erwarten, dass sich die Lernschwierigkeiten auch auf die Testergebnisse auswirken. Einige Lehrer hatten darauf hingewiesen, dass eine Verlängerung der Bearbeitungszeit für die Schüler mit einer Lernschwierigkeit sinnvoll wäre. Jedoch ist es gerade auch ein Anliegen der Kompetenztests, solche Lernschwierigkeiten zu diagnostizieren. Hingegen ist es im Schulalltag wichtig, dass gezielt auf die Schüler mit Beeinträchtigungen im Lernen eingegangen wird und ihnen angemessene Rahmenbedingungen geboten werden.

**Das Wichtigste in Kürze**

- Die Angaben der Fachlehrer zu besonderen Lernschwierigkeiten bzw. sonderpädagogischen Förderbedarf stimmen nicht immer überein, sodass ein intensiverer Informationsaustausch in den Schulen diesbezüglich wünschenswert wäre.
- Schüler mit einer diagnostizierten Lernschwierigkeit schneiden in den Kompetenztests im Mittel in allen Fächern und Klassenstufen schwächer ab als ihre Mitschüler.

### 7.2.6 Vorwissen der Schüler

Als weiteres Schülermerkmal konnte auch in diesem Jahr das Vorwissen der Schüler untersucht werden. Unter dem Merkmal Vorwissen wird dabei das in früheren Klassen erworbene Wissen in einem Fach verstanden. Erfasst wurde dieses frühere Wissen in Kompetenztests, an denen die Schüler bereits in niedrigeren Klassenstufen teilgenommen haben.

Die Vermutung ist, dass das Vorwissen in einem Fach Einfluss auf die aktuelle Testleistung in diesem Fach hat. Ob dem so ist, konnte mittels des Schülerlängsschnitts analysiert werden, sofern für die jeweiligen Schüler Testergebnisse aus früheren Klassenstufen vorlagen. So konnte untersucht werden, ob ein Schüler, der z.B. 2008 Klassenstufe 6 besuchte und in den damaligen Tests gute Leistungen erzielte auch in 2010 bei den Tests in Klassenstufe 8 noch gute

Leistungen erreichte und ob diese durch die guten Leistungen in Klassenstufe 6 erklärt werden können. Analysiert wurden einerseits Zusammenhänge zwischen den Schülerleistungen in Klassenstufe 6 und 8, andererseits war es in diesem Jahr erstmalig möglich auch Zusammenhänge der Schülerleistungen in Klassenstufe 6 mit denen in Klassenstufe 3 zu vergleichen.

Allerdings konnten in Deutsch nur bei 71,8 % der Schüler die Testergebnisse aus Klassenstufe 6 mit denen aus Klassenstufe 8 in Verbindung gebracht werden, in Klassenstufe 3 sogar nur bei 37,3 %. In Mathematik konnten sogar bei noch weniger Schülern auf Testergebnisse anderer Klassenstufen zürück gegriffen werden. Für 64,8 % der Schüler lagen Testergebnisse aus Klassenstufe 6 vor, aus Klassenstufe 3 für 33,8 %. In Englisch konnten bei 68,8 % der Schüler mögliche Zusammenhänge der Leistungen in Klassenstufe 6 und Klassenstufe 8 analysiert werden. Da für alle andereren Schülermerkmale prozentual gesehen mehr Daten vorlagen, wurde das Vorwissen einzeln betrachtet und nicht in die Analysen der folgenden Kapitel einbezogen (s. Abschnitt 7.2.7). Interessant ist das Merkmal Vorwissen, da es Rückschlüsse darauf zulässt, ob beispielsweise besondere Förderung einzelner Schüler sinnvoll ist und wie groß die Verbesserungsmöglichkeit eines Schülers ist, der in früheren Tests schlechte Leistungen erzielte.

Als Methode zur Untersuchung der Zusammenhänge von Vorwissen und aktuellen Testleistungen der Schüler wurde die Regressionsanalyse gewählt (s. Abschnitt 7.2). Mittels dieser ist es möglich zu überprüfen wieviel Varianz an der aktuellen Leistung durch das bereits bestehende Wissen eines Schülers erklärt werden kann. Weiterhin ermöglicht die Regressionsanalyse die Prüfung, ob und in wieweit sich aktuelle Testleistungen aus früheren Testleistungen vorhersagen lassen.

Die Ergebnisse dieser Analyse sind wie folgt: In Deutsch lassen sich 45,5 % der Testleistungen eines Schülers aus Klassenstufe 8 durch seine Testleistungen in Klassenstufe 6 erklären. Die Leistungen in Klassenstufe 3 erklären die aktuellen Testleistungen zu 35,5 %. Wie zu erwarten war, ist der Zusammenhang der Leistungen von Klassenstufe 6 und 8 größer als der von Klassenstufe 3 und 8. Gründe dafür sind einerseits der größere zeitliche Abstand zwischen Klassenstufe 3 und 8, andererseits der zwischenzeitlich erfolgte Schulwechsel, der von Klasse 6 zu 8 in den meisten Fällen ausbleibt.

Ein ähnliches Bild zeigt sich in Mathematik. Hier werden 60,8 % der aktuellen Testleistungen eines Schülers in Klassenstufe 8 durch seine Testleistungen in Klassenstufe 6 erklärt; die Leistungen aus Klassenstufe 3 ergeben eine Varianzaufklärung von 38,4 % der Leistugen aus Klassenstufe 8. Auch hier sind die Unterschiede der Zusammenhänge erwartungsgemäß. In Englisch wird die Varianz der erzielten Testleistungen in Klassenstufe 8 zu 65,6 % durch die Leistung in den Kompetenztests der Klasse 6 aufgeklärt.

Diese Ergebnisse weisen darauf hin, dass die Kompetenz eines Schülers über die Zeit hinweg realtiv stabil, jedoch nicht völlig unveränderlich ist. Da die Kompetenz einer Person besonders, allerdings nicht ausschließlich, von ihrer Intelligenz und Persönlichkeit beeinflusst wird, ist dieser Befund erwartungsgemäß und durchaus zu erklären.

### 7.2.7 Zusammenhang Halbjahresnote und Testergebnisse

Die Halbjahresnote eines Schülers bzw. einer Schülerin ist kein Schülermerkmal, das das Abschneiden bei den Kompetenztests per se beeinflusst. Sie kann als Ergebnis des Unterrichts gesehen werden, wobei hauptsächlich die eigentliche Leistungsfähigkeit eines Schülers ausschlaggebend für die Notengebung sein sollte. Sicherlich wirken aber auch Faktoren wie das Verhalten des Schülers im Unterricht, seine Motivation oder die Bewertungskriterien des Lehrers auf die Zensur am Ende eines Halbjahres. Es ist demnach anzunehmen, dass hohe Zusammenhänge zwischen den Halbjahresnoten und den Ergebnissen bei den Kompetenztests

bestehen. Schüler, die eine gute Halbjahresnote erhalten, sollten tendenziell ein leistungsstärkeres Ergebnis bei den Kompetenztests erzielen. Es wäre jedoch denkbar, dass der Zusammenhang zwischen Halbjahresnote und Testleistung für manche Schüler kleiner ausfällt als für andere. Beispielsweise für einen leistungsstarken Schüler, der ein gutes Ergebnis im Kompetenztest erreicht, aber aufgrund seines störenden Verhalten im Unterricht im Halbjahr schlechter abschneidet als es seine tatsächliche Leistungsfähigkeit hergeben würde. Andererseits könnte es in Einzelfällen auch sein, dass der Zusammenhang zwischen Halbjahresnote und Testleistung eher gering ausfällt, da ein Schüler aufgrund einer schlechten Tagesform nur ein mittelmäßiges Ergebnis beim Kompetenztest erzielt. In diesem Kapitel soll zuerst untersucht werden wie groß der Zusammenhang von Halbjahresnote und Testergebnissen ist. Anschließend wird analysiert, inwiefern sich dieser Zusammenhang für bestimmte Schülergruppen (z.B. Mädchen vs. Jungen) größer bzw. kleiner gestaltet. Damit sollen Hinweise gesucht werden, ob spezifische Schülergruppen hinsichtlich ihrer Leistung realistischer eingeschätzt werden, d.h. ob die erfassten Kompetenzen mit Notengebungskriterien hoch korrespondieren oder nicht.

**Tabelle 7.12:** Zusammenhänge zwischen Halbjahresnote und den Testergebnissen

| | **Deutsch Korrelation** | **Mathematik Korrelation** | **Englisch Korrelation** | **Französisch Korrelation** |
|---|---|---|---|---|
| **Klassenstufe** | | | | |
| 3 | −0,656 | −0,639 | – | – |
| 6 | −0,585 | −0,561 | −0,602 | – |
| 8 | −0,481 | −0,482 | −0,569 | −0,537 |

Tabelle 7.12 zeigt die Korrelationen r zwischen den Halbjahresnoten und den einzelnen Testergebnissen in Deutsch, Mathematik, Englisch und Französisch für alle Thüringer Schüler. Die Korrelation r gibt an dieser Stelle an, wie stark Note und Testergebnis voneinander abhängen. Nach Nachtigall und Wirtz (2002) sind die vorliegenden Korrelationen von r = −0,481 bis r = −0,656 als mittelstark einzuschätzen. Das bedeutet, dass Halbjahresnote und Kompetenztests miteinander einhergehen und die Leistungsfähigkeit eines Schülers für beide eine Rolle spielt. Es scheint aber noch andere Faktoren zu geben, die nur auf die Notengebung oder die Leistung im Kompetenztest einwirken. Denkbar wäre, dass die abgefragten Kompetenztestinhalte nicht notwendigerweise mit den Unterrichtsinhalten übereinstimmen. Die Halbjahresnote setzt sich aus mehreren Aspekten zusammen. Zum Einen gehen in die Leistungsbewertung mündliche Leistungskontrollen oder motiviertes Mitarbeiten im Unterricht ein. Zum Anderen könnten sich die Notengebungskriterien von Lehrern unterscheiden. Lehrer könnten in ihrer Bewertung eine soziale Bezugsnormorientierung bevorzugen, was bedeutet, dass der Leistungsvergleich innerhalb einer Klasse als Grundlage für die Notengebung dient. 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.

Interessant an Tabelle 7.12 ist weiterhin, dass der Zusammenhang zwischen Halbjahresnote und den Testergebnissen in Klassenstufe 3 etwas höher ausfällt als in den weiterführenden Schulen[15]. Eine Vermutung wäre, dass Grundschullehrer und -lehrerinnen insgesamt weniger Klassen unterrichten, ihre Schüler besser kennen und sie damit realistischer hinsichtlich der erworbenen Kompetenzen einschätzen. Es wäre aber auch möglich, dass die Halbjahresnote

[15] Um zu prüfen, ob die angegebenen Korrelationen sich bedeutend voneinander unterscheiden, wurden Fishers Z-Werte berechnet. Das Vorgehen ist bei Bortz (2002) näher beschrieben.

in den höheren Schuljahren stärker vom Verhalten der Schüler im Unterricht abhängt. Im Zuge der pubertären Entwicklung könnte es sein, dass Schüler ihr Kompetenzpotential, welches sie in den Kompetenztests zeigten, nicht konsequent über ein Halbjahr hinweg ausschöpfen.

**Zusammenhang zwischen Halbjahresnote und Testergebnissen in den verschiedenen Schulformen**

Um herauszufinden, ob es in den verschiedenen Schulformen Unterschiede hinsichtlich des Zusammenhanges von Halbjahresnote und Kompetenztestergebnis gibt, wurden die Korrelationen für die Klassenstufe 3, 6 und 8 der entsprechenden Schulformen miteinander verglichen (s. Anhang). Dabei fiel auf, dass Note und Testergebnis in den Förderschulen weniger stark miteinander einhergehen. Anscheinend lassen sich Förderschüler aufgrund ihres hohen Förderbedarfes im Unterricht schlechter in ihren mathematischen und sprachlichen Leistungen einschätzen. Vielleicht liegt es aber auch daran, dass die Leistungsanforderungen, die durch die Kompetenztests gestellt werden, weniger stark mit den Unterrichtsanforderungen in einer Förderschule korrespondieren als in einer Regelschule oder dem Gymnasium.

**Zusammenhang zwischen Halbjahresnote und Testergebnissen bei Mädchen und Jungen**

Bei den Kompetenztests der letzten Jahre fanden sich oft Geschlechtsunterschiede in den mathematischen und sprachbezogenen Kompetenzen: Mädchen schneiden meist in den Deutsch-Kompetenztests besser ab, Jungen hingegen in Mathematik (ausführliche Darstellung im Kapitel 7.2.1). In Verbindung mit der Halbjahresnote ist es interessant zu schauen, ob der Zusammenhang zwischen Note und Kompetenztestergebnissen für Mädchen oder Jungen höher ist. Wäre das z.B. in Deutsch für die Mädchen der Fall, dann könnte man schlussfolgern, dass die Halbjahresnote für Mädchen stärker die im Test gemessene Leistungskompetenz widerspiegelt. Bei Jungen hingegen würden andere Faktoren wie z.B. das Verhalten im Unterricht stärker auf die Notengebung einwirken und damit wäre der Bezug zu den Kompetenztestergebnissen geringer. In den vorliegenden Daten dieses Jahres lassen sich keine konsistenten Befunde erkennen. Für Mädchen ist der Zusammenhang zwischen Halbjahresnote und dem Ergebnis im Deutsch-Kompetenztest in Klassenstufe 3 und 8 höher als für Jungen, nicht aber in Klassenstufe 6. Für Jungen zeigt sich ein höherer Zusammenhang nur beim Mathematiktest der Klassenstufe 6. Es gibt folglich keinen konsistenten Hinweis darauf, dass Lehrer aufgrund des Geschlechts des Schülers unterschiedliche Bewertungskriterien anlegen.

**Zusammenhang zwischen Halbjahresnote und Testergebnissen in Abhängigkeit von der Muttersprache**

Betrachtet man, inwiefern Halbjahresnote und Testergebnisse für Schüler mit deutscher und nicht-deutscher Muttersprache einhergehen, lassen sich kaum Divergenzen feststellen. Die Einschätzung des Lehrers in Form der Halbjahresnote weist in ihrer Beziehung zu den Kompetenztestleistungen für Dritt- und Achtklässler keine Unterschiede auf. Deutsche und nicht-deutsche Muttersprachler scheinen gleich realistisch in ihrer Leistungsfähigkeit eingeschätzt zu werden. Ausnahme ist der Kompetenztest Englisch der Klassenstufe 6: Die Unterrichtsleistung der Schüler mit deutscher Muttersprache steht in engerem Zusammenhang mit dem Test als die der Schüler nicht-deutscher Muttersprache.

**Zusammenhang zwischen Halbjahresnote und Testergebnissen für Schüler mit und ohne besonderen Lernschwierigkeiten bzw. sonderpädagogischem Förderbedarf**

Für Schüler mit besonderer Lernschwierigkeit oder sonderpädagogischen Förderbedarf ist der Zusammenhang zwischen Halbjahresnote und den Kompetenztestergebnissen deutlich geringer als für Schüler ohne eine diagnostizierte Lernschwäche. Das bedeutet, dass die im Test gemessenen Kompetenzen und die Notengebungskriterien für diese Schüler weniger stark miteinander einhergehen. Das kann zum einen daran liegen, dass für Schülern mit diagnostizierter

Lernschwäche Faktoren wie die Anstrengungsbereitschaft im Unterricht eine stärkere Rolle für die Notengebung spielen als die reine Leistungsfähigkeit. Zum anderen kann es aber sein, dass der Zusammenhang von Note und Testergebnis geringer ausfällt, da die Schüler mit diagnostizierter Lernschwierigkeit bei den Kompetenztests nicht die Sonderbehandlung erfuhren wie es im Unterricht der Fall ist. Denkbar wäre beispielsweise, dass einem Schüler mit Lese-Rechtschreib-Schwäche mehr Bearbeitungszeit im Unterricht eingeräumt wird. In den Kompetenztests ist keine solche Sonderregelung vorgesehen, so dass Schüler ein Ergebnis unter ihrem Leistungsniveau erzielten.

**Zusammenhang zwischen Halbjahresnote und Testergebnissen für Klassen mit niedrigem vs. höherem sozioökonomischen Status**

Der Zusammenhang zwischen Halbjahresnote und Testergebnissen wurde in Bezug auf den sozioökonomischen Status auf Klassenebene betrachtet. Daraus folgt, dass sich die Analysen nicht auf den Einzelschüler übertragen lassen und nur begrenzte Aussagekraft besitzen. Für Deutsch und Englisch weisen die Daten darauf hin, dass sich die durchschnittliche Testleistung von Klassen mit einem höheren SES stärker in der mittleren Halbjahresnote widerspiegelt. Es liegt ein höherer Zusammenhang zwischen Halbjahresnote und Testergebnissen vor als für Klassen mit niedrigerem SES. In Mathematik hingegen finden sich inkonsistente Befunde. Der SES scheint sich nicht generell auf den Zusammenhang zwischen Halbjahresnote und Testleistung auszuwirken.

**Fazit**

Zusammenfassend lässt sich festhalten, dass es (wie in den Vorjahren auch) einen substantiellen Zusammenhang zwischen der Halbjahresnote und dem Abschneiden bei den Kompetenztests gibt. Jedoch scheint es neben der Leistungsfähigkeit des Schülers weitere Faktoren zu geben, die auf Note oder Testergebnis einwirken. Mögliche Erklärungsansätze wurden gegeben. Die Analysen unter Berücksichtigung von Schülermerkmalen haben gezeigt, dass Schüler mit einer besonderen Lernschwierigkeit oder sonderpädagogischem Förderbedarf schlechter in ihrer tatsächlichen Leistungsfähigkeit eingeschätzt werden können. Damit geht einher, dass der Zusammenhang zwischen Halbjahresnote und Testergebnissen für Förderschüler niedriger ist als für Schüler anderer Schulformen, da ein hoher Anteil an Förderschülern eine diagnostizierte Lernschwäche vorliegen hat. Bei der Betrachtung des Zusammenhanges zwischen Note und dem Abschneiden bei den Kompetenztests unter Berücksichtigung der Schülermerkmale Geschlecht, Muttersprache und sozioökonomischer Status findet sich kein Hinweis darauf, dass eine der Untergruppen weniger realistisch in der Leistungsfähigkeit eingeschätzt wird.

**Das Wichtigste in Kürze**

- Es gibt einen bedeutsamen Zusammenhang zwischen den Halbjahresnoten und den Kompetenztestergebnissen.
- Dieser Zusammenhang ist geringer für Förderschüler und Schüler mit diagnostizierter Lernschwierigkeit bzw. sonderpädagogischem Förderbedarf ausgeprägt. Diese Schüler scheinen schwerer in ihrer tatsächlichen Leistungsfähigkeit einschätzbar zu sein.

### 7.2.8 Zusammenfassende Betrachtung

In den vorangestellten Abschnitten wurde der Zusammenhang zwischen einzelnen Schülermerkmalen und den Ergebnissen in den Kompetenztests dargestellt. Abschließend wird in diesem Abschnitt betrachtet, wie groß der gemeinsame Einfluss von Geschlecht, Muttersprache,

den Jahren des Schulbesuchs und dem Vorliegen einer besonderen Lernschwierigkeit bzw. sonderpädagogischem Förderbedarf [16] auf das Abschneiden bei den Kompetenztests ist [17]. Es wurde gezeigt, dass sich Schüler im Durchschnitt aufgrund des Vorliegens eines bestimmten Schülermerkmals in ihren Testleistungen unterscheiden können. Beispielsweise schneiden Mädchen in den Deutschkompetenztests besser ab als die gleichaltrigen Jungen. Betrachtet man die Schülermerkmale Geschlecht, Muttersprache, das Jahr des Schulbesuchs und das Vorliegen einer besonderen Lernschwierigkeit gleichzeitig, kann man den Anteil der Kompetenztestleistung (in %) abschätzen, der auf die genannten Schülermerkmalen zurückzuführen ist. Tabelle 7.13 zeigt einen Überblick dieses gemeinsamen Einflusses der vier Schülermerkmale auf das Abschneiden bei den Kompetenztests. Die Prozentangabe im Kompetenztest Deutsch Klassenstufe 3 bedeutet beispielsweise, dass etwa 18% der Testleistung durch das Geschlecht, die Muttersprache, das Jahr des Schulbesuchs und das Vorliegen eines besonderen Lernschwierigkeit erklärt werden können.

Im Vergleich zu den sprachlichen Kompetenztests scheint der Einfluss von Schülermerkmalen auf die Kompetenztests in Mathematik geringer zu sein. Weiterhin zeigt sich der Trend, dass der Einfluss der genannten Schülermerkmale in den höheren Klassenstufen etwas abnimmt. Jedoch sind in Tabelle 7.13 verschiedene Jahrgänge dargestellt, so dass nicht auf eine Entwicklung über die Zeit geschlossen werden kann.

**Tabelle 7.13:** Anteil der Testleistung (in %), der auf Schülermerkmale zurückzuführen ist

| | **Deutschtest** | **Mathematiktest** | **Englischtest** |
|---|---|---|---|
| Klasse 3 | 18% | 13,2% | – |
| Klasse 6 | 16,6% | 10% | 14,6% |
| Klasse 8 | 10,2% | 10,4% | 14% |

Abschließend lässt sich sagen, dass alle erhobenen und in diesem Kapitel berichteten Schülermerkmale Geschlecht, Muttersprache, Sozioökonomischer Status, die Jahre des Schulbesuchs und Besondere Lernschwierigkeiten/sonderpädagogischer Förderbedarf in deutlichem Zusammenhang mit den Testergebnissen stehen. Die berichteten Ergebnisse sind mit Vorsicht zu interpretieren, da sie auf Mittelwertsanalysen beruhen und keinen Rückschluss auf das Individuum zulassen. Nur weil Mädchen im Durchschnitt etwas schlechter in Mathematik abschneiden, heißt das nicht, dass alle Mädchen schlechtere mathematische Kompetenzen besitzen. Es gibt eine sehr hohe Variabilität hinsichtlich der sprachlichen und mathematischen Kompetenzen zwischen den Schülern. Für eine faire Auswertung auf Schul- und Klassenebene ist die Berücksichtigung solcher Schülermerkmale unbedingt notwendig. Aus diesem Grund werden 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. Die Ergebnisse weisen erneut darauf hin, dass verschiedene Schülermerkmale die Kompetenztestleistung beeinflussen können. Faktoren wie beispielsweise die Motivation am Testtag oder die allgemeine Konzentrationsfähigkeit des Schülers wurden zwar im Zuge der Komptenztests nicht erhoben, könnten aber auch zu Leistungsunterschieden führen. Es ist von besonderer Bedeutung, möglichst viele relevante Schülermerkmale zu erfassen und in die

[16] Der sozioökonomische Status wird nicht in diese zusammenfassende Betrachtung der Merkmale auf Schülerebene einbezogen, da die Messung der Bücherfrage aus Datenschutzgründen nur auf Klassenebene erfolgte.

[17] Um den gemeinsamen Beitrag der einzelnen besprochenen Schülermerkmale auf die Testergebnisse abzuschätzen, wurden sie gemeinsam in ein Regressionsmodell zur Vorhersage der Testergebnisse eingebracht. Informationen zum Vorgehen einer Regressionsanalyse finden sich beispielsweise bei Bortz (2002). Die Regressionskoeffizienten können im Anhang eingesehen werden.

Auswertung der Kompetenztests einzubeziehen und somit Vergleiche zwischen Klassen und Schulen fairer zu gestalten. Die Berücksichtigung der potenziellen Einflussfaktoren bietet die Chance Leistungsdifferenzen zwischen Schülergruppen zu erkennen und liefert einen Hinweis für Interventionsbedarf. Bei leistungsschwächeren Schülern könnte eine gezielte Unterstützung zu einer optimaleren Ausschöpfung der Leistungsressourcen beitragen.

# 8 Rezeption und Evaluation der Tests

## 8.1 Befragung der Fachlehrer nach der Dateneingabe

Im Anschluss an die Dateneingabe der Kompetenztest-Ergebnisse hatten die Lehrer die Möglichkeit, die Verwendbarkeit der Tests zu beurteilen. Dabei wurden Fragen zur Nützlichkeit der Kompetenztests als Instrument zur Lernstandsdiagnostik für die eigene Unterrichtsentwicklung und die Schulentwicklung gestellt. Zusätzlich konnte die Lehrplanadäquatheit der Aufgaben eingeschätzt werden. Die vier genannten Aspekte wurden mit Hilfe einer fünfstufigen Skala ähnlich dem Schulnotensystem bewertet, wobei 1 = volle Zustimmung und 5 = gar keine Zustimmung bedeutet. Insgesamt haben in den verschiedenen Bereichen jeweils etwa 3 500 Lehrer eine solche Bewertung vorgenommen. Im Vergleich zum vergangenen Jahr, in dem etwa 4 200 Lehrer Bewertungen abgegeben haben, ist die Teilnehmerzahl damit gesunken. Neben den vorgegebenen Fragestellungen bestand auch die Möglichkeit, eine freiwillige Rückmeldung zur Dateneingabe zu geben. Teilweise wurde dies außerdem dazu genutzt, inhaltliche Anmerkungen zu den diesjährigen Kompetenztests zu machen oder Kritik an diesen zu üben.

Im Folgenden soll ein Überblick über die Einschätzungen der Fachlehrer zu den Kompetenztests gegeben werden. Dabei wird nicht speziell auf Förderschulen eingegangen. Auch wird auf die nähere Betrachtung der Bewertungen der teilnehmenden Französischlehrer verzichtet. In beiden Fällen wären aufgrund der für statistische Analysen zu kleinen Teilnehmerzahl keine verlässlichen Aussagen möglich. Abschließend werden Anregungen und Kommentare der Fachlehrer zu den einzelnen Tests dargestellt.

### Ergebnisse der Verwendbarkeitsanalysen

Auch in diesem Jahr lag die Einschätzung der Nützlichkeit der Kompetenztests der Lehrer über alle Kompetenztests und Schularten hinweg im mittleren Bereich. Die durchschnittliche Bewertung der Lernstandsdiagnostik, Lehrplanadäquatheit, Unterrichtsentwicklung und Schulentwicklung lag zwischen 2,50 und 2,99 (vgl. Abb. 8.1). Im Vergleich zur durchschnittlichen Bewertung im Jahr 2009 ist dabei kaum Veränderung festzustellen. Entgegen diesem allgemeinen Trend ist die Bewertung durch die Grundschullehrer in einigen Bereichen besser geworden. So schätzten sowohl die Mathematik- als auch die Deutschlehrer die Lehrplandäquatheit ein wenig höher ein ($d>0{,}2$ [1]). Zudem bewerteten die Deutschlehrer die Lernstandsdiagnostik positiver als im letzten Jahr ($d=-0{,}237$). Die Nützlichkeit für die Unterrichtsentwicklung und die Schulentwicklung wurden in beiden Fächern wie im vergangenen Jahr mit 2,6 (Unterrichtsentwicklung) und ca. 2,8 (Schulentwicklung) bewertet. Sowohl in Mathematik als auch in Deutsch korrelieren die Einschätzungen in der Lehrplanadäquatheit und der Lernstandsdiagnostik positiv miteinander. Das bedeutet, dass tendenziell eine positive Bewertung der Lehrplanadäquatheit mit einer positiven Bewertung der Lernstandsdiagnostik einher geht. Insgesamt erreichen die diesjährigen Einschätzungen in den Grundschulen dennoch nicht das hohe Bewertungsniveau von 2008.

[1] vergleiche die Erläuterungen zum Effektstärkemaß d in Kapitel 7

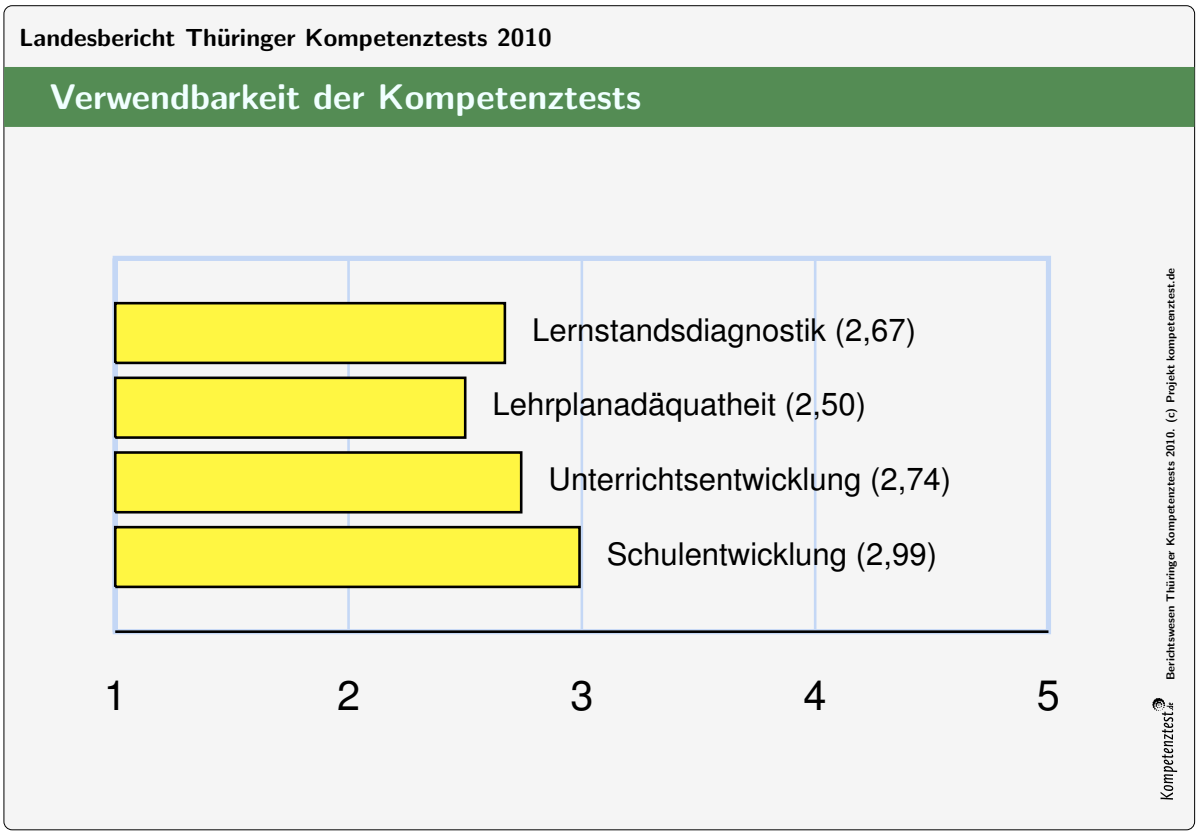


**Abbildung 8.1:** Mittlere Einschätzung der Verwendbarkeit der Kompetenztests durch die Fachlehrer über alle Klassenstufen und Schularten hinweg (gemessen auf einer Skala von 1 = volle Zustimmung bis 5 = gar keine Zustimmung).

Wie im vergangenen Jahr gaben die Mathematiklehrer der Klassenstufe 6 im Vergleich zu den anderen Fächern und Klassenstufen in allen Bereichen die besten Bewertungen ab. Insgesamt befinden sich die Bewertungen in Mathematik in den Klassenstufen 6 und 8 auf dem gleichen Niveau wie in 2009 (d<0,2). Im Vergleich zu Deutsch und Englisch geben die Mathematik-Lehrer in allen Bereichen bessere Noten. Im Fach Deutsch wurde in Klassenstufe 8 die Nützlichkeit der Kompetenztests für die Lernstandsdiagnostik, die Unterrichtsentwicklung und die Schulentwicklung etwas geringer eingeschätzt als im vergangenen Jahr (d>0,2). Als deutlich geringer wurde von diesen Lehrern die Lehrplanadäquatheit bewertet. Hier wurde zum ersten Mal in diesem Jahr der Kompetenzbereich “Sprechen und Hören“ mit Aufgaben zum Hörverstehen getestet. Möglicherweise wurde dieser Bereich von den Lehrern als weniger lehrplanadäquat eingeschätzt. Auch in Klassenstufe 6 neigten die Lehrer zu einer schlechteren Bewertung. Sowohl die Regelschullehrer (d=0,335) als auch die Gymnasiallehrer (d=0,221) schätzten die Lehrplanadäquatheit etwas geringer ein als im letzten Jahr. Zusätzlich bewerteten die Regelschullehrer die Lernstandsdiagnostik als weniger nützlich (d=0,292), während dies bei den Gymnasiallehrern bei der Schulentwicklung der Fall war (d=0,229). Ebenfalls Unterschiede zwischen den Regelschullehrern und den Gymnasiallehrern finden sich in Englisch in der Klassenstufe 6. Die Regelschullehrer vergaben in allen vier Bereichen schlechtere Noten als im vergangenen Jahr (d>0,2), während die Gymnasiallehrer alle Bereiche wie im letzten Jahr einschätzten (d<0,2). Möglicherweise passten die Tests in diesem Fach und dieser Klassenstufe besser zu den Anforderungen und Bedürfnissen der Gymnasien als der Regelschulen.

**Tabelle 8.1:** Einschätzung der Kompetenztests im Hinblick auf Nützlichkeit für Lernstandsdiagnostik, Lehrplanadäquatheit, Unterrichtsentwicklung und Schulentwicklung (Mittelwerte getrennt nach Test und Schularten, RS = Regelschule, Gym = Gymnasium).

| Kompetenztest | Schulart | Lernstandsdiagnostik | Lehrplanadäquatheit | Unterrichtsentwicklung | Schulentwicklung |
|---|---|---|---|---|---|
| DK3 | | 2,54 | 2,27 | 2,59 | 2,83 |
| MK3 | | 2,54 | 2,56 | 2,57 | 2,79 |
| DK6 | RS | 2,88 | 2,61 | 2,82 | 3,05 |
| DK6 | Gym | 2,67 | 2,51 | 2,82 | 3,14 |
| MK6 | RS | 2,42 | 2,20 | 2,51 | 2,76 |
| MK6 | Gym | 2,34 | 2,22 | 2,52 | 2,89 |
| EK6 | RS | 2,96 | 2,65 | 3,00 | 3,23 |
| EK6 | Gym | 2,69 | 2,35 | 2,85 | 3,15 |
| DK8 | RS | 2,91 | 2,74 | 2,97 | 3,18 |
| DK8 | Gym | 2,87 | 2,76 | 3,01 | 3,23 |
| MK8 | RS | 2,75 | 2,70 | 2,82 | 2,98 |
| MK8 | Gym | 2,62 | 2,95 | 2,77 | 3,04 |
| EK8 | RS | 2,89 | 2,65 | 2,96 | 3,23 |
| EK8 | Gym | 2,79 | 2,56 | 3,06 | 3,18 |

### Anmerkungen und Anregungen zu den Kompetenztests

Nach Beendigung der Dateneingabe bestand für die Fachlehrer die Möglichkeit, ihre Meinung, Wünsche und Kritik bezüglich der Dateneingabe zu äußern. Über 3 400 Mal wurde diese Gelegenheit wahrgenommen. Neben Anregungen, Hinweisen und Erwartungen für die Zukunft wurden auch einige Aspekte kritisiert. Die Kritik richtete sich dabei nicht nur gegen die Dateneingabe, sondern auch gegen verschiedene Aspekte der Kompetenztests. Im Folgenden soll für jedes Fach und jede Klassenstufe ein Überblick über die Rückmeldungen der Fachlehrer gegeben werden. Zu beachten ist, dass die hier aufgeführten Beispiele nicht als repräsentativ angesehen werden können. Sie dienen in erster Linie dazu, einen Eindruck von den unterschiedlichen Ansichten der Fachlehrer zu gewinnen. Natürlich können nicht alle Kommentare wiedergegeben werden, deswegen beschränkt sich die folgende Darstellung auf einige exemplarische Meinungen. Manche Anmerkungen wurden von mehreren Lehrern einheitlich über verschiedene Klassenstufen und Fächer geäußert: Viele Fachlehrer kritisieren die Kompetenztests als nicht lehrplanadäquat, sehen sie für bestimmte Schülergruppen (z. B. Schüler mit besonderer Lernschwierigkeit) als zu umfangreich an und stellen den Nutzen im Verhältnis zum Aufwand in Frage. Daran zeigt sich, dass die Fachlehrer nach wie vor nicht hinreichend über den Sinn der Kompetenztests informiert sind. Das Ziel besteht nicht darin, in allen Bereichen bestmögliche Leistungen zu erzielen, sondern die Ergebnisse der einzelnen Klassen

müssen vor ihrem individuellen Hintergrund interpretiert werden. Wenn manche Aufgaben von den Schülern nicht gelöst werden können, weil sie laut Lehrplan noch nicht behandelt wurden oder weil sie zu anspruchsvoll für bestimmte Schülergruppen sind, so ist dieser Umstand in die Interpretation der Ergebnisse mit einzubeziehen. Es ist nicht tragisch, wenn beispielsweise keiner der Schüler die maximale Punktzahl erreicht. Viel wichtiger ist es, die individuellen Stärken und Schwächen einer Klasse zu erkennen und dieses Wissen zur Unterrichtsentwicklung zu nutzen. So kann der Kompetenztest helfen, den eigenen Unterricht zu optimieren und damit langfristig zur Umsetzung von Verbesserungsmöglichkeiten beitragen.

**Anmerkungen zum Mathematiktest Klassenstufe 3:** Die häufigste Anmerkung der Mathematiklehrer der Klassenstufe 3 war, dass der Kompetenztest zu umfangreich sei, was dazu führe, dass die Schüler Zeitprobleme bekämen bzw. die Konzentration nachließe. Es wurde der Vorschlag geäußert, den Kompetenztest auf zwei Tage aufzuteilen (ähnlich dem Kompetenztest Deutsch in der Klassenstufe 3). Weiterhin kritisierten die Mathematiklehrer das hohe Anforderungsniveau der Aufgaben zu Daten, Häufigkeit und Wahrscheinlichkeit. Diese Testaufgaben wurden als zu schwierig angesehen und als nicht lehrplanadäquat empfunden. Außerdem merkten die Mathematiklehrer an, dass der Leseumfang für einen Mathematiktest zu groß gewesen sei. Zum Umfang und zur Schwierigkeit der Aufgaben ist zu sagen, dass diese Aspekte bei der Entwicklung der Kompetenztests Berücksichtigung finden. Wie weiter oben schon erwähnt ist es nicht der Anspruch der Kompetenztests, dass alle Schüler die Aufgaben in der vorgegebenen Zeit lösen können.

**Anmerkungen zum Mathematiktest Klassenstufe 6:** Wie auch im letzten Jahr, gab es beim Kompetenztest Mathematik der Klassenstufe 6 die wenigsten Kommentare. Bei den Anmerkungen der Mathematiklehrer handelt es sich um Einzelmeinungen, die nicht wiederholt angesprochen wurden. Daraus lässt sich schlussfolgern, dass der Großteil der Lehrer mit diesem Kompetenztest zufrieden ist und keine kritischen Punkte anzumerken hat.

**Anmerkungen zum Mathematiktest Klassenstufe 8:** Die häufigste Kritik der Mathematiklehrer der Klassenstufe 8 am diesjährigen Kompetenztest bezog sich auf die mangelnde Lehrplanadäquatheit. Besonders die Aufgaben zur Wahrscheinlichkeitsrechnung seien zu schwierig gewesen und die verwendete Fachsprache sei für die Schüler unverständlich gewesen. Auch hier ist wiederum darauf hinzuweisen, dass diese Umstände bei der Interpretation der Ergebnisse der eigenen Klasse berücksichtigt werden müssen. Bei den Kompetenztests geht es nicht darum, möglichst gut abzuschneiden, sondern darum den individuellen Leistungsstand der Schüler zu erkennen und dieses Wissen zur Optimierung des Unterrichts zu nutzen.

**Anmerkungen zum Deutschtest Lesen Klassenstufe 3:** Die Meinungen der Deutschlehrer, die die Möglichkeit wahrnahmen eine Anmerkung zu verfassen, waren in Bezug auf den Kompetenztest Lesen recht einheitlich. Es wurde bemängelt, dass der Test zu viele Aufgaben umfasste bzw. dass zu wenig Bearbeitungszeit zur Verfügung stand. Außerdem wurde die Bewertung der Multipe Choice Aufgaben als ungerecht und zu streng angesehen. Manche Deutschlehrer hielten die Texte des Kompetenztests für zu anspruchsvoll.

**Anmerkungen zum Deutschtest Rechtschreibung Klassenstufe 3:** Die häufigste Kritik bezüglich des Kompetenztests Rechtschreibung betraf das Verbot, Wörterbücher zu verwenden. Außerdem wurde es als ungünstig empfunden, dass den Schülern in einer Aufgabe falsch geschriebene Wörter präsentiert wurden. Die Deutschlehrer bedauerten es, dass sehr wenig Grammatikwissen abgefragt wurde und der Schwerpunkt der Aufgaben auf der Rechtschreibung lag. Hier äußern die Lehrer den Wunsch nach einer ausgeglicheneren Mischung. Es ist wichtig zu erwähnen, dass für eine hinreichend genaue Erfassung einer Kompetenz genügend Aufgaben verwendet werden müssen. Damit der Kompetenztest nicht zu umfangreich wird, können also nicht alle Lernbereiche gleichzeitig erhoben werden. Deswegen wechseln sich die Themenschwerpunkte von Jahr zu Jahr ab. Die neu eingeführten Lupenstellen führten zu unterschiedlichen Meinungen: Zum Teil wurde dieser Bewertungsmodus positiv aufgenommen, zum Teil wurde er aber auch als ungeeignet eingeschätzt.

**Anmerkungen zum Deutschtest Klassenstufe 6:** Der Hauptkritikpunkt bezüglich des Kompetenztests Deutsch der Klassenstufe 6 bezog sich auf den hohen Korrekturaufwand insbesondere bei der Aufgabe 3 “Lassie“. Viele Deutschlehrer empfanden den Kompetenztest als nicht lehrplanadäquat und bezweifelten den Sinn und Nutzen des Deutschtests. Weiterhin wurde die Beschwerde geäußert, dass keine Rechtschreibaufgaben enthalten seien. Die Erwiderung zu dieser Kritik deckt sich mit der beim Deutschtest Rechtschreibung Klassenstufe 3 aufgeführten Erklärung.

**Anmerkungen zum Deutschtest Klassenstufe 8:** Einige der Deutschlehrer der Klassenstufe 8 empfanden den diesjährigen Kompetenztest als zu einseitig. Sie kritisierten, dass zu wenig Aufgaben die Themen Rechtschreibung, Grammatik, Sprachbeherrschung und Ausdruck abdecken würden. Statt dessen umfasse der Deutschtest zu viele Texte und Höraufgaben. Außerdem sei die Bearbeitungszeit für die Leseaufgaben zu knapp bemessen. Wünschenswert wäre eine Pause zwischen den Testteilen gewesen. Weitere Anmerkungen bezogen sich auf die Korrektur: Der Korrekturaufwand wurde als zu hoch eingestuft und die Bewertung als ungerecht empfunden.

**Anmerkungen zum Englischtest Klassenstufe 6:** Ein Teil der Englischlehrer bemängelte, dass die Hörtexte zu schnell und zu undeutlich gesprochen seien. Es wurde die Empfehlung mitgeteilt, die Hörtexte bei zukünftigen Kompetenztests zwei Mal vorzuspielen. Die Bearbeitungszeit für die Leseaufgaben wurde als zu kurz bewertet, die des Schreibteils dagegen als zu lang eingestuft. Über die Schwierigkeit der Aufgaben gab es geteilte Meinungen: Manche hielten sie für zu schwierig, andere empfanden sie als zu anspruchslos. Hinsichtlich der Korrektur waren sich die Englischlehrer darin einig, dass die Korrekturanweisungen besonders bei den Schreibaufgaben zu umständlich waren und dass die Korrektur als sehr zeitaufwändig eingeschätzt wurde.

**Anmerkungen zum Englischtest Klassenstufe 8:** Manche Englischlehrer der Klassenstufe 8 fanden es bedauerlich, dass in diesem Jahr keine Hörkompetenz getestet wurde. Die für die Leseaufgaben zur Verfügung stehende Zeit wurde als zu kurz empfunden. Dagegen war die Zeit für die Schreibaufgaben zu lang bemessen und das Schwierigkeitsniveau im Schreibteil war zu niedrig. Beschwerden gab es darüber, dass das endgültige Korrekturmanual nicht termingerecht abrufbar und schlecht verständlich war. Darüber hinaus kritisierten die Englischlehrer die umständliche Bewertung der Schreibaufgabe.

**Fazit:**

Im Vergleich zum vergangenen Jahr lassen sich kaum Unterschiede in den Ergebnissen der Verwendbarkeitsanalysen feststellen. Die durchschnittliche Bewertung der Lernstandsdiagnostik, Lehrplanadäquatheit, Unterrichtsentwicklung und Schulentwicklung entspricht nahezu der Bewertung von 2009. Ein wenig verbessert hat sich die Einschätzung durch die Grundschullehrer. Dabei wird aber nicht das hohe Bewertungsniveau aus 2008 erreicht. In den Klassenstufen 6 und 8 ergibt sich ein recht gemischtes Bild. Tendenziell bewerten die Mathematiklehrer die Kompetenztests besser als ihre Kollegen in Deutsch und Englisch.

Auch in diesem Jahr nutzten viele Fachlehrer die Möglichkeit, ihre Meinung bezüglich der Kompetenztests mitzuteilen. Hauptsächlich wurden kritische Anmerkungen und Wünsche für die zukünftigen Kompetenztests geäußert. Alle Kommentare wurden an die entsprechenden Verantwortlichen weitergeleitet und sollten hier nur überblicksartig dargestellt werden. Die Anmerkungen der Fachlehrer werden dazu genutzt, die Kompetenztest zu optimieren. Doch sie machen auch deutlich, dass teilweise noch großer Informationsbedarf seitens der Lehrer besteht, insbesondere bezüglich des Nutzens der Kompetenztests für die Unterrichts- und Schulentwicklung.

**Das Wichtigste in Kürze**

- die durchschnittliche Bewertung der Lernstandsdiagnostik, Lehrplanadäquatheit, Unterrichtsentwicklung und Schulentwicklung entspricht in etwa der vom letzten Jahr
- viele Lehrer nutzten die Möglichkeit, Anmerkungen zur Dateneingabe zu machen und äußerten dabei hauptsächlich Kritik

## 8.2 Rezeption der Testergebnisse an den Schulen

Allein für die Kompetenztests wurden für Thüringer Klassen und Schüler im Jahr 2010 über 22 000 Berichte im PDF-Format erzeugt. Die Zahl der Abrufe von Berichten im geschützten Schulportal ist gegenüber dem letzten Jahr in jeder Kategorie moderat gesunken.

Die relative Gesamtzahl der Klassen mit Downloads von Rückmeldungen zeigt Abbildung 8.2 getrennt nach Schulart. Dargestellt ist jeweils in rot der Anteil an Klassen, die mindestens eine der Rückmeldungen abgerufen 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 relativ ähnlich und betragen durchschnittlich 95,4 % bei allen und 77,4 % bei den Ergebnisrückmeldungen mit Vergleichswerten. Wie auch in den vorangegangenen Jahren sind die Abrufquoten der Förderschulen deutlich geringer, 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 auch erwartet werden. Die Abrufquoten der jeweils ersten Rückmeldewelle sind mit denen des Vorjahres vergleichbar, bei den Berichten mit Vergleichswerten sind im Jahresvergleich die Quoten durchschnittlich um ca. 5 % gesunken.
Die Zahl der Schulen, die 10 Wochen nach Veröffentlichung der Berichte für Klassenstufe 3 keinerlei Rückmeldung vom Schulportal heruntergeladen hatten, ist im Vergleich zum Vorjahr von 8 auf 14 gestiegen. Dies betrifft 8 Grundschulen und 6 Förderschulen.

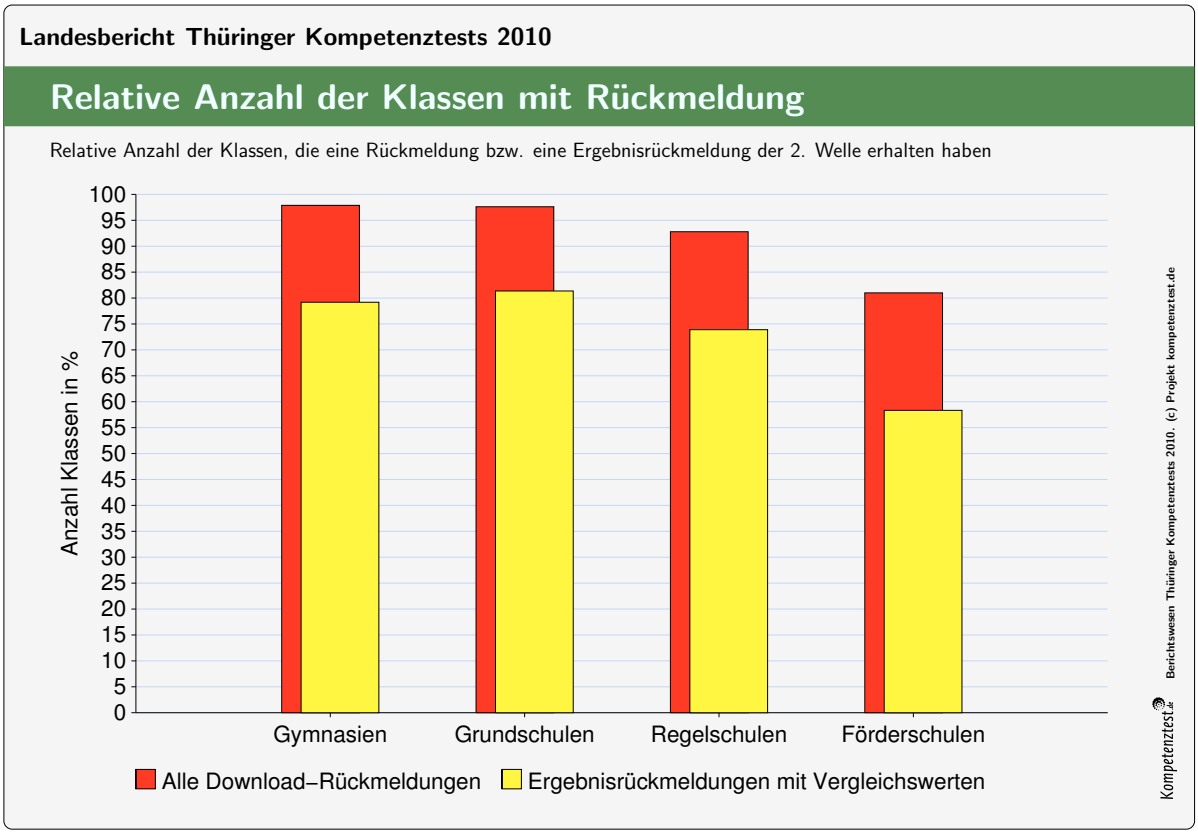


**Abbildung 8.2:** Relative Anzahl von Klassen mit Downloads mindestens einer Sofort-Rückmeldung vs. Ergebnisrückmeldung mit korrigierten Landesmittelwerten (Stand: 4. Oktober)

Die Abrufquoten der Klassenberichte mit Vergleichswerten getrennt nach Fach und Klassenstufe zeigt Abbildung 8.3 im zeitlichen Verlauf. Die Rückmeldungen zu den Tests der Klassenstufen 6 und 8 wurden bereits am 17. Mai veröffentlicht, also 10 Wochen vor den Tests in Klassenstufe 3. In den vorangegangenen Jahren zeigten die Abrufquoten der zuerst veröffentlichten Berichte einen nur geringen Anstieg auf, mit einem deutlich erkennbaren Schub zum Zeitpunkt der Veröffentlichung der restlichen Tests. Dieser Effekt ist in diesem Jahr nur schwach erkennbar. Die Abrufquoten in Klassenstufe 3 sind wie im letzten Jahr auch bereits nach wenigen Wochen am höchsten, d. h. in Klassenstufe 3 werden mehr Berichte heruntergeladen und dies auch in einem kürzeren Zeitraum nach der Veröffentlichung. Die Berichte für Klassenstufe 8 wurden etwas seltener als Berichte der Klassenstufe 6 abgerufen und das Interesse an Berichten für Mathematik scheint etwas höher als an den Berichten der Sprachenfächer zu sein. Schon im letzten Jahr wurden die Berichte des Englischtests Klassenstufe 8 mit etwas Abstand am seltensten abgerufen, in diesem Jahr sind die Abrufe von Deutsch Klassenstufe 8 ebenfalls vergleichbar selten. Im Jahresvergleich zeigt sich eine ähnliche Rangfolge der Abrufe wie im letzten Jahr, allerdings sind die Abrufquoten bei fast allen Fächern um ca. 5 % gesunken.

Für 819 Schulen konnten in diesem Jahr Schulberichte erzeugt werden, die gemittelte Vergleiche der Klassenergebnisse mit den jeweils korrigierten Landeswerten über mehrere Jahre hinweg enthalten. Von diesen Schulen haben 10 Wochen nach dem letzten Veröffentlichungstermin 35,2 % die Berichte noch nicht abgerufen. Unter den Gymnasien ist diese Quote mit 28,9 % am geringsten, Grund- und Regelschulen bilden mit 33,8 % bzw. 34,3 % das Mittelfeld,

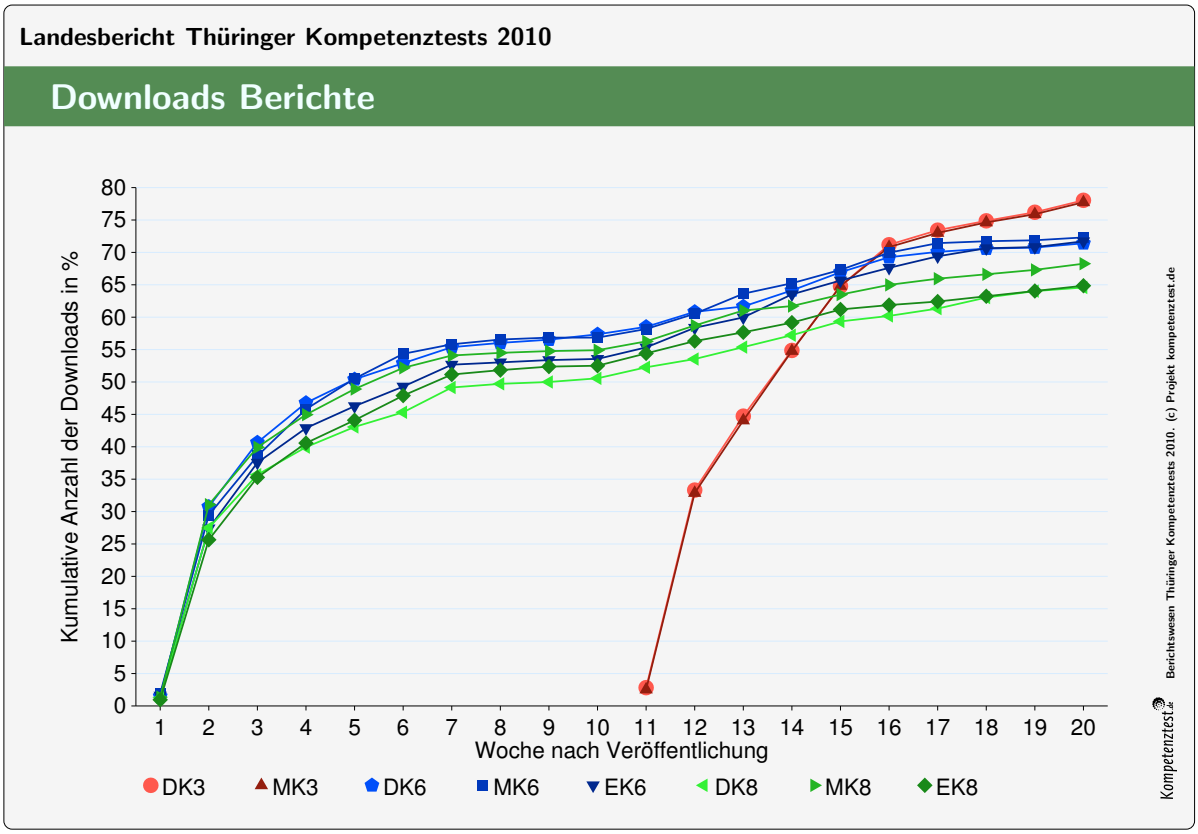


**Abbildung 8.3:** Downloads der Berichte mit Vergleichsdaten der verschiedenen Tests im zeitlichen Verlauf (pro Woche nach Veröffentlichung K6/K8, Stand: 4. Oktober)

die Förderschulen haben mit 73 % die höchste Quote von Schulen ohne Abruf.
Die Zahl der Schulen, die den Schulbericht nicht abgerufen haben, hat sich im Jahresvergleich erneut deutlich um fast 7 % erhöht.

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

**Das Wichtigste in Kürze**

- Abrufquoten der Berichte mit Vergleichswerten sind durchschnittlich um ca. 5 % gegenüber dem Vorjahr gesunken
- Abrufquoten der Schulberichte sind durchschnittlich um ca. 7 % gegenüber dem Vorjahr gesunken

## 8.3 Bewertung der Kompetenztests

Nach der Durchführung der jährlich stattfindenden Kompetenztests werden Schulleiter und Lehrer aller Schularten stichprobenartig um eine Einschätzung des Nutzens und der Auswirkungen der Kompetenztests gebeten. In diesem Abschnitt werden zunächst die Einschätzungen der Grundschulen vom Vorjahr berichtet. Diese konnten in den Landesbericht 2009 nicht einbezogen werden, da die Befragung der Lehrer und Schulleiter zu damaligen Zeitpunkt noch nicht abgeschlossen war. Im Anschluss daran werden die Ergebnisse der Schulleiter- und Lehrerbefragung des Jahres 2010 berichtet. Dabei werden nur ausgewählte Einschätzungen des

Lehrer- und Schulleiterfragebogens dargestellt. Es wurden Fragen zu folgenden Themenkomplexen zusammengefasst: Gesamtbewertung der Tests (Beispielfrage: Die landesweite Durchführung der Kompetenztests nehme ich als eine Chance wahr, über den eigenen Tellerrand zu blicken.), die Auswirkungen auf die Gestaltung des Unterrichts (z.B.: Impuls zum Einsatz innovativer Unterrichtsmethoden) und die Kommunikation mit externen Ansprechpartnern (z.B.: Auf- und Ausbau von Kontakten zu externen Schulberatern).[2]

### 8.3.1 Evaluation der Kompetenztests 2009 der Thüringer Grundschulen

Im Vorjahr nahmen 82 Lehrer und 43 Schulleiter der Grundschulen an der Nachbefragung zur Gesamtbewertung der Kompetenztests teil. Im Folgenden werden (1) die Einschätzungen der Fachlehrer mit den Einschätzungen der Schulleiter, (2) die Einschätzungen der Jahre 2008 und 2009 und (3) die Einschätzungen der Grundschulen mit denen der weiterführenden Schulen verglichen.

**(1) Vergleich der Einschätzungen von Schulleitern und Fachlehrern**

Insgesamt lässt sich festhalten, dass sowohl die Lehrer als auch die Schulleiter die Kompetenztests insgesamt als nützlich bewerten und angegeben wird, dass es positive Auswirkungen auf Unterrichtsebene gibt. Insbesondere gaben die Lehrer an, dass der Unterrichtsinhalt auf die Schüler angepasst wird und verstärkt das Prinzip der Binnendifferenzierung zur Anwendung kommt. Im Mittel werden die Lehrer aber nicht dazu angeregt ihren Unterricht in Folge der Kompetenztests stark umzustrukturieren oder innovative Unterrichtsmethoden anzuwenden. In der Regel werden die Ergebnisse der Kompetenztests in der Fachschaft und mit der Schulleitung besprochen. Weder die Schulleiter noch die Fachlehrer schätzten ein, dass infolge der Kompetenztests verstärkt Kontakt zu Fachlehrern anderer Schulen oder externen Schulberatern aufgenommen wurde. In den Grundschulen fällt auf, dass die Schulleiter sowohl den Nutzen der Kompetenztests, sowie die Auswirkungen auf Unterrichtsebene und die verstärkte Kommunikation mit Externen infolge der Kompetenztests tendenziell etwas besser einschätzen als die Fachlehrer.

**(2) Vergleich der Einschätzungen von 2008 und 2009**

Die Fachlehrer der Grundschulen unterscheiden sich hinsichtlich der Beurteilung des Gesamtnutzens und der Auswirkungen auf Unterrichtsebene infolge der Kompetenztests nicht zwischen 2008 und 2009. In beiden Jahren fällt die Einschätzung im Mittel positiv aus. Dies gilt nicht für die Kontaktaufnahme zu externen Fachlehrern oder Schulberatern. In beiden Jahren geben nur sehr wenige Lehrer an, den Austausch mit anderen Schulen oder externen Beratern aufzunehmen. Es ist trotz dieses niedrigen Niveaus ein leicht positiver Trend zu verzeichnen, da die Fachlehrer im Jahr 2009 etwas stärker einschätzten, dass infolge der Kompetenztests Kontakt zu externen Personen gesucht wird.

Die Einschätzungen der Schulleiter unterscheiden sich nicht bedeutend zwischen den beiden Jahren. Der Nutzen der Kompetenztest und die Auswirkungen auf Unterrichtsebene werden auch von den Schulleitern eher positiv eingeschätzt, während die Kontaktaufnahme mit Fachlehrern anderer Schulen oder externen Schulberatern auch aus dem Blickwinkel der Schulleiter heraus nur selten als Folge der Kompetenztests angegeben werden.

[2] Die analysierten Fragen, Skalen, sowie Mittelwerte und Standardabweichungen können im Anhang eingesehen werden.

**(3) Vergleich der Einschätzungen in Grundschulen und weiterführenden Schulen**

Vergleicht man die Beurteilungen hinsichtlich der Gesamtbewertung, der Auswirkungen auf Unterrichtsebene sowie die externe Kommunikation infolge der Kompetenztests, so fällt auf, dass sich die Schulleiter von Grundschulen und weiterführenden Schulen nicht bedeutend unterscheiden. Hingegen zeigt sich ein deutlicher Unterschied bei der Einschätzung der Lehrer: Grundschullehrer schätzen die drei Bereiche deutlich positiver ein als die Lehrer in weiterführenden Schulen.

**Das Wichtigste in Kürze**

- Schulleiter und Lehrer in den Grundschulen schätzten den Gesamtnutzen der Kompetenztests und die Auswirkungen auf Unterrichtsebene im Jahr 2009 insgesamt positiv ein.
- Die Beurteilungen der Schulleiter und Fachlehrer 2009 unterscheiden sich kaum zu den Einschätzungen des Vorjahres.
- Lehrer in den Grundschulen bewerteten die Kompetenztests etwas positiver als Lehrer aus weiterführenden Schulen.

### 8.3.2 Evaluation der Kompetenztests 2010

In diesem Jahr nahmen etwa 100 Schulleiter und 380 Lehrer aller Schulformen an der Nachbefragung der Kompetenztests teil. Im Folgenden werden (1) die Einschätzungen der Fachlehrer mit den Einschätzungen der Schulleiter, (2) die Einschätzungen der Jahre 2009 und 2010 und (3) die Einschätzungen der Grundschulen mit denen der weiterführenden Schulen verglichen.

**(1) Vergleich der Einschätzungen von Schulleitern und Fachlehrern**

Sowohl die Schulleiter als auch die Fachlehrer schätzten die Gesamtbewertung und die Auswirkungen auf Unterrichtsebene im Mittel eher positiv ein. Wie im Vorjahr scheinen sich die Auswirkungen auf Unterrichtsebene eher auf verstärkte Binnendifferenzierung und die Anpassung der Unterrichtsinhalten zu beziehen. Die Kompetenztests scheinen aber weniger zum Einsatz innovativer Unterrichtsmethoden anzuregen. Die Ergebnisse der Kompetenztests werden in den meisten Schulen wenigstens teilweise in der Fachschaft und mit der Schulleitung besprochen. Dagegen geben nur sehr wenige Schulen an, dass infolge der Kompetenztests intensiver mit Fachkollegen anderer Schulen oder externen Schulberatern kommuniziert wird. Die Bewertung des Gesamtnutzens der Kompetenztests sowie die Auswirkungen auf Unterrichtsebene werden von Schulleitern bedeutend positiver eingeschätzt als von Lehrern. Eine mögliche Erklärung für diesen Unterschied in der Bewertung könnte in dem mit den Kompetenztests verbundenen Arbeitsaufwand liegen, der vor allem bei den Fachlehrern liegt. An der Mehrzahl der Thüringer Schulen scheint dieser Arbeitsaufwand nicht ausgeglichen zu werden. 73% der Schulleiter geben an, dass die an den Kompetenztests beteiligten Lehrer keine Entlastung an anderer Stelle bekommen.

**(2) Vergleich der Einschätzungen von 2009 und 2010**

Im Vergleich zum Vorjahr unterscheiden sich die Einschätzungen der Lehrer und der Schulleiter nicht in Bezug auf die Auswirkungen auf Unterrichtsebene und die externe Kommunikation. Dagegen schätzten die Lehrer dieses Jahr den Nutzen der Kompetenztest etwas schlechter

ein als im Vorjahr. Sie empfinden insbesondere den mit den Kompetenztests verbundenen Arbeitsaufwand im Mittel als nicht gerechtfertigt. Die Schulleiter beider Jahrgänge unterscheiden sich nicht in ihren Einschätzungen der Kompetenztests.

**(3) Vergleich der Einschätzungen in Grundschulen und weiterführenden Schulen**

Der Gesamtnutzen der Kompetenztests, die Auswirkungen auf Unterrichtsebene und die Kommunikation mit Fachlehrern anderer Schulen oder externen Schulberatern werden in Grundschulen insgesamt positiver eingeschätzt. Dieser Unterschied in der Beurteilung gilt sowohl für die Lehrer als auch für die Schulleiter.

### 8.3.3 Fazit

Die hier berichteten Vergleiche zwischen Lehrern und Schulleitern, Grundschulen und weiterführenden Schulen und verschiedenen Jahrgängen können nur Unterschiede aufzeigen, jedoch nicht erklären. Gründe für die unterschiedliche Beurteilung der Kompetenztests können nur in Zusammenarbeit mit den Schulen analysiert werden. Um abzusichern, dass die Durchführung der Kompetenztests weiterhin als nützlich empfunden wird, sollten in direkter Zusammenarbeit mit den Lehrern und Schulleitern geklärt werden, was an der Durchführung und Rückmeldung der Kompetenztests verbessert werden kann. Es wäre beispielsweise denkbar, dass an Grundschulen bessere Bedingungen für die Durchführung der Kompetenztests herrschen, so dass der Nutzen und die Auswirkungen insgesamt positiver empfunden werden. Diese Bedingungen herauszufinden und auch an den weiterführenden Schulen herzustellen, wäre beispielsweise ein wünschenswertes Ziel, um die Akzeptanz und den Nutzen der Kompetenztests weiter auszubauen. Ein Hinweis für die Bedingungen der Durchführung bietet die Frage nach der Entlastung: Vielleicht empfinden die Lehrer den mit den Kompetenztests verbundenen Arbeitsaufwand als ungerechtfertigt, weil es von Seiten der Schule keinen Ausgleich für die mit den Kompetenztests verbunden Aufgaben gibt. Es wäre denkbar, dass die Akzeptanz der Kompetenztests bei den Lehrern (im Vergleich zu den Schulleitern) etwas höher wäre, wenn ein schulinterner Ausgleich bei der Aufgabenverteilung stattfinden würde.

**Das Wichtigste in Kürze**

- Schulleiter und Lehrer schätzten den Gesamtnutzen der Kompetenztests und die Auswirkungen auf Unterrichtsebene im Jahr 2010 insgesamt positiv ein.
- Schulleiter bewerten die Kompetenztests besser als die Fachlehrer.
- Im Vergleich zum Vorjahr bewerten die Lehrer den Gesamtnutzen der Kompetenztests etwas schlechter.
- An Grundschulen werden die Kompetenztests als nützlicher empfunden und es wird eine stärkere Auswirkung auf die eigene Unterrichtsgestaltung wahrgenommen als an weiterführenden Schulen.

# 9 Schlussbemerkung

Die Thüringer Kompetenztests werden seit mittlerweile 8 Jahren durchgeführt. Sie stellen ein bewährtes und etabliertes Instrument zur Kompetenzdiagnose und Unterrichtsentwicklung dar. Im Schuljahr 2009 / 2010 nahmen knapp 45 000 Schülerinnen und Schüler in den Klassenstufen 3, 6 und 8 an den Tests teil, mehr als 17 000 Ergebnisberichte wurden den Schulen auf ihrem Schulportal zur Verfügung gestellt. Um dem Wunsch vieler Lehrkräfte zu entsprechen und die positive Einschätzung der Tests als Diagnoseinstrument der Schülerleistung zu nutzen, wurde in diesem Schuljahr erstmals im Fach Deutsch in den Klassenstufen 6 und 8 spezielle Schülerberichte erstellt und den Lehrkräften zum Download angeboten. Diese Schülerberichte geben die Ergebnisse des einzelnen Schülers im gesamten Test sowie in den einzelnen Kompetenzbereichen grafisch wieder. Zusätzlich liefern sie Vergleichswerte, welche das Ergebnis in Relation zur Leistungsverteilung aller Schüler der gleichen Schulart setzen. Zu betonen ist allerdings, dass die Betrachtung des einzelnen Schülers nicht das zentrale Ziel der Tests ist. Ihr eigentliches Potenzial liegt in der Evaluation des eigenen Unterrichts, und sie dienen darauf aufbauend primär der Unterrichtsentwicklung.

Die Analyse der Downloadquoten und die Ergebnisse der Lehrerbefragungen zur Evaluation der Kompetenztests zeigen zum Teil unterschiedliche Ergebnisse. Dass die Downloadquoten der Ergebnisberichte leicht rückläufig sind, stimmt bedenklich. Es deutet auch darauf hin, dass die Auseinandersetzung mit den Ergebnissen nicht überall seitens Schulleitung und Aufsicht konsequent nachgefragt wird. Bei den Befragungsergebnissen zeigt sich, dass die durchschnittliche Bewertungen im positiven Bereich liegt und sich nach dem leichten Rückgang in den letzten Jahren stabilisiert hat. Dabei schätzt die Schulleitung die Tests und ihren Nutzen positiver ein als die Lehrkräfte. Von deren Seite wird immer wieder das Belastungsargument angeführt. Hier gehenin den letzten Jahren tatsächlich ein Rückgang der Bewertung mit einem wachsenden Korrektur- und Eingabeaufwand durch komplexer zu testende Kompetenzbereiche (Schreiben) sowie einer größeren Anzahl von Aufgaben miteinander einher. Aufgabe der Testentwickler muss es daher sein, einen gelungenen Kompromiss zwischen wissenschaftlicher Genauigkeit der Tests und einer einfachen Handhabung und Durchführung der Tests zu finden.

Auf der anderen Seite zeigen die Rückmeldungen durch die Lehrkräfte immer noch einen weiterhin bestehenden Informationsbedarf. Teile der Kompetenztest, die mit dem eigenen Unterricht korrespondieren, können durch die Lehrkräfte nach eigenem pädagogischen Ermessen benotet werden, was zu einer Entlastung der Lehrkräfte führt. Eine häufige Irritation folgt aus der Tatsache, dass sich die Testaufgaben in Bezug auf Format, Bepunktung und andere Aspekte von den im eigenen Unterricht verwendeten Lernaufgaben unterscheiden. Dies wird in manchen Fällen zum Anlass genommen, die Tests und ihre Ergebnisse nicht für die eigene Unterrichtsentwicklung zu nutzen. Testaufgaben dienen der möglichst genauen Messung relevanter Kompetenzen. Für die Unterrichtsentwicklung sind der Hintergrund der Aufgaben (Kompetenz- und Standardorientierung) sowie die Testergebnisse wichtig. Ein Üben von alten Testaufgaben, wie es leider immer wieder beobachtet werden kann, ist nicht sinnvoll. Vielmehr bieten die Tests die Möglichkeit, Stärken und Schwächen des eigenen Unterrichtes in relevanten Teilbereichen des Faches zu erkennen und daraus Entwicklungsziele abzuleiten und umzusetzen.

Es ist nach wie vor von großer Bedeutung, dass die Nutzung der Tests an den Schulen nicht als Selbstläufer betrachtet wird, sonder dass dieser Prozess weiterhin von allen Akteuren im Bildungssystem mit Aufmerksamkeit und Unterstützung bedacht wird. Maßnahmen wie die systematische Einbeziehung der Tests in die Fachschaftsarbeit (mit entsprechenden Schulungen) oder das Einbeziehen dieser Evaluationsinstrumente in die universitäre Lehramtsausbildung werden teilweise bereits umgesetzt oder auf dem Weg gebracht. Didaktische Materialien, welche bereits im Zuge der Testentwicklung jährlich im umfangreichen Maße zur Weiterarbeit nach den Tests entwickelt werden, können durch entsprechende Online-Lösungen leichter verfügbar und damit nutzbarer gemacht werden. Fachübergreifende Schülerberichte, wie sie für das Schuljahr 2010 / 2011 geplant sind, bieten verbesserte Diagnose- und Fördermöglichkeit auf Schülerebene. Ebenso bieten computergestützte adaptive Kompetenztests die Möglichkeit, besonders leistungsstarke oder leistungsschwache Schüler genauer einzuschätzen und darauf aufbauend die individuelle Förderung zu verbessern. Entsprechende Forschungs- und Entwicklungsprojekte werden derzeit an der Universität Jena geplant. Weiterführende Informationen wie die Angabe von erreichten Kompetenzstufen können konkrete Hilfen für differenzierten Unterricht und individuelle Förderung sein. All diese Maßnahmen sind auf dem Wege der Implementierung. Sie brauchen entsprechende Unterstützung und Nachhaltigkeit, um an den Schulen anzukommen und dort die Nutzung der Kompetenztests als wichtigen Baustein der Unterrichtsentwicklung zu etablieren.

# 10 Literatur


BmBF (2008). Mathematikunterricht und Geschlecht. Empirische Ergebnisse und pädagogische Ansätze. Band 30 Bildungsforschung.
http://www.bmbf.de/pub/band-dreissig-bildungsforschung.pdf. Zugriff am 16.10.2009

Bortz, J. (2005). Statistik für Human- und Sozialwissenschaftler. 6. Auflage, Springer, Berlin.

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

Budde, J., Scholand, B., Faulstich-Wieland, H. (2008). Geschlechtergerechtigkeit in der Schule: Eine Studie zu Chancen, Blockaden und Perspektiven einer gender-sensiblen Schulkultur. Weinheim: Juventa.

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.).

Nachtigall, C., Landmann, M. & Storbeck, I. (2009): Belastung oder Chance. Zur Nutzung von Vergleichsarbeiten, Lernstandserhebungen, Kompetenztests, Orientierungsarbeiten und Co. In: Schulleitung und Schulentwicklung (45). E 1.23 pp. 1-18.

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

Nachtigall, C. (Hrsg.) (2008). Landesbericht Thüringer Kompetenztests 2008.
http://www.kompetenztest.de

PISA-Konsortium Deutschland (Hrsg.) (2008): PISA 2006 in Deutschland: die Kompetenzen der Jugendlichen im dritten Ländervergleich. 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.  Nachtigall, C. (2004). Deskriptive Statistik. (3. Auflage), Weinheim: Juventa.

Nachtigall, C. & Wirtz, M. (2008): Statistik, Abschnitt II.C.2.3. Weinheim: Juventa.

**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 Ministeriums für Bildung, Wissenschaft und Kultur (TMBWK) erstellt.

Dr. Christof Nachtigall
Dipl.-Ing.(BA) Mario ten Venne
B.Sc. Torsten Schmidt

sowie die studentischen Hilfskräfte:
Simone Becker, Katrin Blarr, Judith Schäfer, Frauke Stoltefuß, Wiebke Wartenberg.

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

Erstellungsdatum: 17. November 2010