

VERA

**VERGLEICH SARBEITEN
JAHRGANGSSTUFE 4**

SCHULJAHR 2005/06

LÄNDERBERICHT BERLIN

Prof. Dr. A. Helmke, Jun.Prof. Dr. I. Hosenfeld,
J. Groß Ophoff, A. C. Halt, J. Hochweber,
K. Isaac, U. Koch und F. Scherthan

Universität Koblenz - Landau, Campus Landau

Inhalt

Zusammenfassung: Zentrale Befunde für Berlin.....	5
1 Einleitung und Überblick.....	7
1.1 Ziele	7
1.2 Organisation	8
2 Aufgabenentwicklung und Definition der Fähigkeitsniveaus	9
2.1 Deutsch	10
2.1.1 Aufgabenentwicklung	10
2.1.2 Skalierung.....	11
2.1.3 Fähigkeitsniveaus	12
2.2 Mathematik.....	15
2.2.1 Aufgabenentwicklung	15
2.2.2 Skalierung und Fähigkeitsniveaus des Mathematiktests	16
2.3 Gesamt-Ergebnisse für die Fähigkeitsniveaus.....	20
2.3.1 Gesamt-Verteilung.....	20
2.3.2 Zusammenhänge zwischen den Fähigkeitsniveaus.....	21
3 Berliner Ergebnisse	22
3.1 Fähigkeitsniveaus	22
3.1.1 Verteilung der Fähigkeitsniveaus in den Ländern	22
3.1.2 Veränderungstrends 2004–2005	23
3.1.3 Unterschiede innerhalb und zwischen den Klassen	24
3.1.4 Leistungen von Mädchen und Jungen	26
3.1.5 Migrationshintergrund	27
3.2 „Fairer Vergleich“	29
3.2.1 Beschreibung ausgewählter Kontextmerkmale	30
3.2.2 Bildung der Kontextgruppen	31
3.2.3 Verteilung der Fähigkeitsniveaus nach Kontextgruppen	35
3.2.4 Verteilung aller Schulen auf die Kontextgruppen	35
3.2.5 Vergleich tatsächliche vs. erwartete Leistung.....	36
3.3 Diagnosegenauigkeit	37

3.4	Lehrerfragebogen.....	40
3.4.1	<i>Kontinuität des Unterrichts in Mathematik und Deutsch und Fähigkeitsniveaus</i>	41
3.4.2	<i>Unterrichtserfahrung, grundständige Ausbildung und Fähigkeitsniveaus</i>	43
3.4.3	<i>Vorbereitung auf die Vergleichsarbeiten</i>	45
3.4.4	<i>Kooperation bei der Aufgabenauswahl und -auswertung</i>	46
4	Ausblick	47
5	Anhang: Lehrerfragebogen	49
6	Glossar	57
7	Literatur	63

Zusammenfassung: Zentrale Befunde für Berlin

Die Verteilung der Schüler/-innen auf die einzelnen Fähigkeitsniveaubereiche ist in 2005 verglichen mit der Verteilung 2004 in etwa gleich geblieben. In keinem der Inhaltsbereiche ergeben sich signifikante Unterschiede der Fähigkeitsniveauverteilung zwischen den beiden Jahren des Berichtszeitraums.

Im Fach **Mathematik** wird deutlich, dass in den Inhaltsbereichen Arithmetik und Geometrie der Großteil der Schüler/-innen erweiterte bis fortgeschrittene Fähigkeiten aufweist. Während sich für Arithmetik in etwa eine Gleichverteilung auf die Fähigkeitsniveaus zwei und drei zeigt, besetzen in Geometrie mehr Kinder den mittleren Niveaubereich. Im Sachrechnen erreichen verhältnismäßig wenige Schüler/-innen das dritte Fähigkeitsniveau und beinahe die Hälfte der Schüler/-innen ein Niveau, das höchstens das Beherrschen elementarer Aufgaben umfasst.

Im Inhaltsbereich **Leseverständnis** gibt es einen auffällig großen Anteil an Schülerinnen und Schülern, die höchstens das Fähigkeitsniveau eins erreichen.

Alles in allem fällt im Vergleich mit der länderübergreifenden Fähigkeitsniveauverteilung auf, dass in Berlin in allen erfassten Inhaltsbereichen weniger Kinder dem dritten Fähigkeitsniveau und damit mehr Kinder dem Fähigkeitsniveau eins bzw. nicht auswertbaren Leistungen zugeordnet werden. Im Leseverständnis und Sachrechnen scheint angesichts der häufigen Besetzung des unteren Fähigkeitsniveaus ein besonders großer Förderbedarf zu bestehen. Der hohe Anteil an Kindern mit nicht auswertbarer Leistung im Inhaltsbereich Leseverständnis steht vermutlich im Zusammenhang mit den Kontextbedingungen, vor allem mit der Sprachbeherrschung.

Es zeigen sich bedeutsame Unterschiede zwischen Kindern mit Deutsch als dominanter und Kindern mit Deutsch als nicht-dominanter Sprache sowohl in Mathematik als auch im Leseverständnis. So erreichen in Geometrie und Sachrechnen von den Schülerinnen und Schülern mit Deutsch als nicht-dominanter Sprache nur knapp zehn Prozent das höchste Fähigkeitsniveau (zum Vergleich Schüler/-innen mit Deutsch als dominanter Sprache 28,7 % bzw. 23,3 %), im Leseverständnis sind es unter neun Prozent (Schüler/-innen mit Deutsch als dominanter Sprache 28 %). Dabei fällt vor allem der hohe Anteil von Kindern auf, die in Sachrechnen und Leseverständnis ein Niveau erreichen, das höchstens das Beherrschen elementarer Aufgaben umfasst. Insbesondere beim Leseverständnis kann infolge des hohen Anteils von Kindern mit nicht auswertbarer Leistung (knapp ein Viertel) von einem substanziellen Unterschied zu Kindern, deren dominante Sprache Deutsch ist, gesprochen werden. Die Resultate bestätigen die Vermutung, dass Merkmale der Sprachherkunft Schüler/-innen bereits in der vierten Jahrgangsstufe mit erheblichen Leistungsunterschieden gekoppelt sind.

In Mathematik zeigt sich ein bedeutsamer Vorsprung der Jungen im Bereich Sachrechnen, dort unterscheiden sich Mädchen und Jungen. Im Leseverständnis ist ebenfalls ein bedeutsamer Effekt zu verzeichnen – hier schneiden Mädchen besser ab.

1 Einleitung und Überblick

Am 27. und 29. September 2005 schrieben im Rahmen des Projektes VERA mehr als 300 000 Schülerinnen und Schüler der vierten Klassenstufe in sieben deutschen Bundesländern¹ Vergleichsarbeiten in den Fächern Mathematik und Deutsch. Im Folgenden werden Hintergrund und Anlage des Projekts sowie einige empirische Ergebnisse deskriptiv dargestellt. Vertiefende Analysen müssen späteren Veröffentlichungen vorbehalten bleiben.

1.1 Ziele

Anders als beispielsweise TIMSS, PISA oder IGLU ist VERA kein bloßes „System Monitoring“ (dazu würde eine Stichprobe genügen), sondern umfasst neben der Bestandsaufnahme ausdrücklich auch die Schul- und Unterrichtsentwicklung:

So soll die aktive Einbeziehung der Lehrkräfte in den gesamten Prozess ein Anstoß für fachdidaktische Diskussion und Kooperation zwischen den Lehrkräften sein. Damit wird einem immer wieder geäußerten Ruf nach mehr schulinterner Kooperation und Teamarbeit Rechnung getragen. Die Rückmeldung des Leistungsstandes sowie von Informationen zur diagnostischen Kompetenz der Lehrkräfte und zu Fehlermustern der Schülerinnen und Schüler sollen pädagogische Impulse geben und damit die schulinterne Diskussion von (Bildungs-)Standards, der Schul- und Unterrichtsentwicklung oder der Beurteilungspraxis anregen. Darüber hinaus können die Informationen über die Fähigkeitsniveaus in den Fächern Deutsch und Mathematik als ergänzende Information zur Beratung der Eltern herangezogen werden.

Da die Aufgabenauswahl, die Auswertung und die Ergebnismeldung über das Internet erfolgen, trägt das Projekt VERA zu einem Aufschwung in der effizienten Nutzung des Internet für die schulische Qualitätssicherung bei und leistet somit auch einen Beitrag zur Förderung der Medienkompetenz.

Die Vergleichsarbeiten verfolgen also mehrere Ziele:

- Unterrichtsentwicklung: Nutzung pädagogischer und fachdidaktischer Impulse
- Schulentwicklung: Intensivierung schulinterner Kooperation und Teamarbeit
- Professionalisierung der Lehrkräfte im Hinblick auf diagnostische Kompetenzen
- Standardsicherung
- Ergänzende Information zur Beratung der Eltern
- Erleichterung und Beschleunigung der Umsetzung moderner Kerncurricula, Lehr- und Rahmenpläne
- Effizienzsteigerung bei der Nutzung des Internet für die schulische Qualitätssicherung

¹ Berlin, Brandenburg, Bremen, Mecklenburg-Vorpommern, Nordrhein-Westfalen, Rheinland-Pfalz, Schleswig-Holstein

1.2 Organisation

Durchgeführt wird das Projekt VERA von der Projektgruppe Empirische Bildungsforschung an der Universität in Landau (Leitung: Prof. Dr. Andreas Helmke und Jun.Prof. Dr. Ingmar Hosenfeld) in enger Zusammenarbeit mit den beteiligten Ministerien und Landesinstituten, Projekten und Institutionen der empirischen Bildungsforschung.

Im Herbst 2005 nahmen alle Grundschulen in den Bundesländern Berlin, Brandenburg, Bremen, Mecklenburg-Vorpommern, Nordrhein-Westfalen, Rheinland-Pfalz und Schleswig-Holstein verpflichtend an der Untersuchung teil. Darüber hinaus nahmen etliche deutsche Auslandsschulen mit 4. Klassen aus allen fünf Erdteilen das Angebot zur Teilnahme an VERA an. Abbildung 1 verdeutlicht die Vernetzung verschiedener Funktionen und Gruppen.

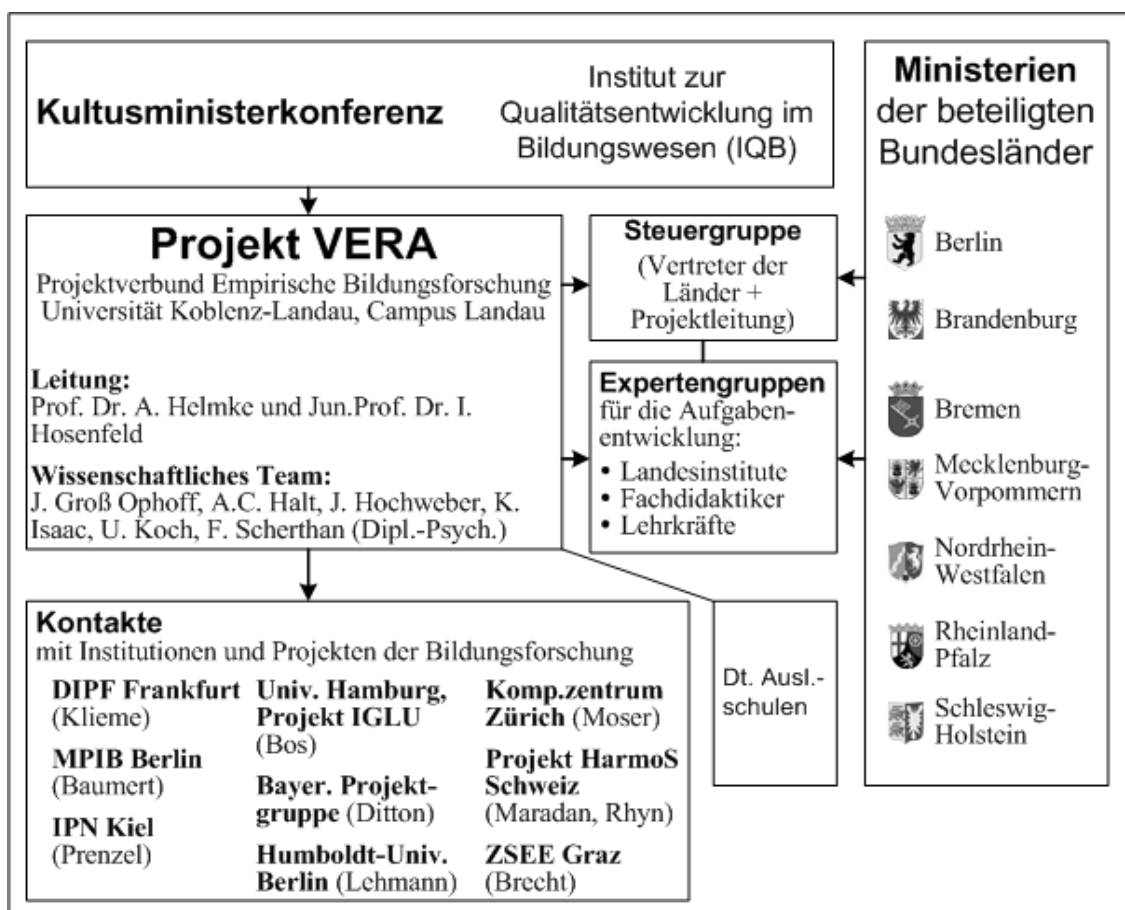


Abbildung 1: Bildungspolitische Kontext von VERA

Im Vergleich zum Jahr 2004 wurden die Durchführungsmodalitäten für die Vergleichsarbeiten verändert, insbesondere um den Aufwand bei der Durchführung und Auswertung der Vergleichsarbeiten zu reduzieren. So sollen z.B. die Fächer Mathematik und Deutsch im jährlichen Wechsel versetzt in unterschiedlichem Umfang erhoben werden. Für das Jahr 2005 bedeutete dies konkret, dass das Fach Deutsch nur im Bereich „Leseverständnis“ getestet wurde, wohingegen das Fach Mathematik

(Arithmetik, Geometrie, Sachrechnen) in vollem Umfang erhoben wurde. 2006 sollen im Fach Deutsch mehrere Inhaltsbereiche, dagegen Mathematik nur anteilig geprüft werden. Darüber hinaus wurde in den Vergleichsarbeiten im Jahr 2005 auf die Auswahl von Testaufgaben sowie die aufwändige Fehleranalyse im Fach Deutsch verzichtet.

2 Aufgabenentwicklung und Definition der Fähigkeitsniveaus

Die im Rahmen von VERA verwendeten Aufgaben wurden von zwei Expertengruppen (je eine für Mathematik und Deutsch) entwickelt und im Vorfeld der Vergleichsarbeiten im Jahr 2005 umfangreich empirisch untersucht. In einem ersten Schritt wurden ausgewählte Aufgaben durch die Referenzschulen beurteilt und zwar in Bezug auf Anwendbarkeit, typische Fehler sowie schulische Akzeptanz. Basierend auf den Rückmeldungen wurden erste Aufgabenentwürfe verbessert bzw. weitere Aufgaben entwickelt. Die eigentlichen empirischen Pilotierungs- und Normierungsstudien erfolgten im Mai 2005 (ca. 10 000 Schülerinnen und Schüler) und dienten der Realisierung folgender Ziele:

- Verbesserung des entwickelten Aufgabenmaterials und der Korrekturanweisungen (v.a. Pilotierung)
- Erprobung der von Didaktikern und Lehrkräften entwickelten Aufgaben (insbesondere in Hinsicht auf die Testsituation sowie Anregung zu Schul- und Unterrichtsentwicklung)
- Bestimmung der Aufgabenschwierigkeiten und Personenparameter als Basis für die Zuordnung zu den Fähigkeitsniveaus im Rahmen der Vergleichsarbeiten 2005
- Weiterentwicklung der Fähigkeitsniveau-Definitionen, darauf aufbauend Entwicklung der didaktischen Erläuterungen der Materialien
- Untersuchung der Art und Häufigkeit von Fehlern (Falschlösungen) der Aufgaben
- Gewinnung eines Referenzrahmens für die obligatorische Aufgabenauswahl durch die Kollegien
- Erhebung und Analyse von Kontextmerkmalen, die einen „fairen Vergleich“ ermöglichen.

Dazu wurden Zufallsstichproben von Schulen aus allen Bundesländern gezogen und jeweils mit einem Drittel dritten und zwei Dritteln vierten Klassen in die Untersuchung einbezogen. Innerhalb jeder Klasse kamen sowohl in der Pilotierung als auch in der Normierung je Fach zehn verschiedene Testhefte zum Einsatz. Jedes der Hefte enthielt eine Reihe von Aufgaben, die nur in dem jeweiligen Heft präsentiert wurden und weitere Aufgaben, die in mindestens einem weiteren Heft enthalten waren (Multi-Matrix-Sampling). So konnten in beiden Fächern jeweils mehr als 200 Aufgaben untersucht werden, während jedem einzelnen Kind nur eine überschaubare Anzahl von Aufgaben vorgelegt wurde. Zur Bearbeitung der Aufgaben standen in Deutsch in der Pilotierung 90 Minuten (mit zehn Minuten Pause) und in der Normierung 60 Minuten (mit fünf Minuten Pause) zur Verfügung. Im Fach Mathematik wurden 50 Minuten Bearbeitungszeit festgesetzt. Darüber hinaus wurden wichtige Hintergrundmerkmale der

Schülerinnen und Schüler (Alter, Geschlecht, zu Hause gesprochene Sprache etc.) im Rahmen eines kurzen Schülerfragebogens erhoben.

Im Folgenden wird der Prozess der Aufgabenentwicklung, getrennt für Mathematik und Deutsch, dargestellt.

2.1 Deutsch

2.1.1 Aufgabenentwicklung

Die Deutschaufgaben wurden von einer länderübergreifenden Expertengruppe entwickelt. Die Gruppe setzt sich zusammen aus erfahrenen Lehrkräften, Mitarbeitern der Landesinstitute, Herrn Prof. Dr. Albert Bremerich-Vos als fachdidaktischem Berater und Mitarbeitern des Landauer Projektteams als psychologische und psychometrische Experten. Die Entwicklung begann im Frühjahr 2003 und orientierte sich während des gesamten Entwicklungsprozesses auch an den im Oktober 2004 verabschiedeten Bildungsstandards für den Primarbereich.

Im Rahmen der Aufgabenentwicklung wurde ein besonderes Augenmerk darauf gelegt, die Entwicklung einer neuen Unterrichts- und Aufgabenkultur in den Schulen zu unterstützen. Das bedeutet konkret für die Aufgabenentwicklung, dass die generierten Items und die damit verbundenen Anforderungen nicht nur Testgütekriterien genügen müssen, sondern z.B. durch einen thematischen Rahmen und integrative Formate zu Entwicklungsprozessen anregen sollen.

Wie bereits erwähnt wurde in den Vergleichsarbeiten 2005 im Fach Deutsch nur der Bereich „Leseverständnis“ erfasst und dementsprechend in Vorüberhebungen normiert. Aus dem Anspruch des integrativen Deutschunterrichts heraus wurde für VERA 2005 der thematische Rahmen „Reisen“ bestimmt. Dabei wurde in curricularer Hinsicht darauf geachtet, dass die den Testaufgaben zugrunde liegenden Lesetexte das Spektrum von fiktionalen sowie von kontinuierlichen und diskontinuierlichen Sachtexten abdeckten.

Darüber hinaus wurden alle für die Vergleichsarbeiten ausgewählten Texte und Aufgaben auf schwierigkeitsbestimmende Merkmale hin überprüft. Kriterien hierbei waren in Bezug auf den Text u.a.

- Textlänge und –komplexität
- Gestaltung (Schriftauswahl, Zeilenabstand, Gliederung, Zeilennummern)
- Unterstützende Elemente (Bilder, Grafiken)
- Syntaktische Strukturen
- Vertrautheit des Wortschatzes (Text, Aufgabe)
- Vertrautheit und Interesse am Thema.

In Bezug auf die Aufgaben wurden des weiteren die folgenden Merkmale berücksichtigt:

- Streubreite der Aufgabenschwierigkeiten und Aufgabenformate (Multiple-Choice, offene Antworten unterschiedlicher Länge, Markierungen, Zuordnungen, Sortierungen)
- Vertrautheit mit dem Aufgabenformat
- Anordnung und Reihenfolge der Aufgaben
- Formulierung des Aufgabentextes (Eindeutigkeit).

Zunächst wurde eine umfangreiche Sammlung von Aufgaben erzeugt, die dann in sukzessiven Auswahlrunden reduziert und optimiert wurde. Dabei spielten neben klassischen Kriterien der Testentwicklung (Objektivität, Reliabilität, Validität) auch Überlegungen zur Durchführbarkeit, der erforderlichen Bearbeitungszeit und der Auswertungsökonomie sowie zur Relevanz für Schul- und Unterrichtsentwicklung eine Rolle. Insbesondere mit Blick auf den letztgenannten Gesichtspunkt, also die Anregung zu schulischen Entwicklungsprozessen, wurden die Hintergründe zu den VERA 2005-Aufgaben sowie Möglichkeiten des Umgangs mit den zurückgemeldeten Ergebnissen ausführlich in Form von „Didaktischen Erläuterungen“ (Projektgruppe VERA-Deutsch, 2005, siehe <http://www.uni-landau.de/vera/aufgaben.htm>) dargestellt.

2.1.2 Skalierung

Die Verwendung unterschiedlicher, über gemeinsame Aufgaben miteinander verbundener Testhefte erfordert den Einsatz probabilistischer Testmodelle, um Aufgaben sowie Schülerinnen und Schüler über die Testhefte hinweg auf einer gemeinsamen Dimension abzubilden. Hintergrund und Ziel dieser Methode können hier nicht ausführlich dargestellt werden (dazu s. Helmke & Hosenfeld, 2004). Nach der Erfassung, Aufbereitung und Qualitätskontrolle wurde 2004 für die vier Inhaltsbereiche jeweils ein ein-parametrisches Modell (Rasch-Modell; Rasch, 1960) angepasst. Als Ergebnis erhält man für jeden Inhaltsbereich eine Skala, auf der gleichzeitig die Aufgabenschwierigkeiten und die Personenfähigkeiten abgetragen werden können. Dabei gilt entsprechend des probabilistischen Charakters des Modells, dass sich aus der Differenz zwischen Personen(fähigkeit) und Aufgaben(schwierigkeit) angeben lässt, mit welcher Wahrscheinlichkeit eine Aufgabe gelöst wird. Hierbei wurde festgelegt (wie bei PISA 2000), dass eine 62,25-prozentige Lösungswahrscheinlichkeit resultiert, wenn Personenfähigkeit und Aufgabenschwierigkeit numerisch identisch sind, d.h. es ist gewährleistet, dass Personen entsprechende Aufgaben mit 'hinreichender' Sicherheit lösen können.

Dass Personenfähigkeiten und Aufgabenschwierigkeiten auf einer gemeinsamen Skala angeordnet werden, bietet einen großen Vorteil für die Interpretation der Ergebnisse: Die Fähigkeitsdimension lässt sich in Zonen (Fähigkeitsniveaus) einteilen, die sich anhand der ihnen zugeordneten Aufgaben kriterial beschreiben lassen. So kann verdeutlicht werden, welche Anforderungen von den zugeordneten Personen

typischerweise bewältigt werden. Natürlich wurden auch die im Jahr 2005 normierten Aufgaben auf der gleichen Fähigkeitsdimension verankert, die im Rahmen der Normierungsstudien von VERA 2004 ermittelt wurde. Das wiederum ermöglicht es, die Fähigkeitsniveau-Beschreibungen fortschreitend zu präzisieren.

2.1.3 Fähigkeitsniveaus

Was bedeuten die Fähigkeitsniveaus?

Ebenso wie in den großen internationalen Vergleichsstudien TIMSS, DESI, PISA oder PIRLS und IGLU ist es bei Lernstandserhebungen und anderen vergleichenden Leistungsmessungen im schulischen Kontext „state of the art“, inhaltlich definierte Kriterien zur Beschreibung und zum Vergleich des Leistungsstandes (Kompetenzen) heranzuziehen. In der Expertise zur „Entwicklung nationaler Bildungsstandards“ (Klieme et al., 2003) werden Kompetenzen als situationsangemessene bzw. domänen-spezifische Problemlösefähigkeiten beschrieben.

Üblicherweise wird die Ausprägung der Kompetenz in Anlehnung an das Anforderungsniveau der damit korrespondierenden Aufgaben in sog. Kompetenzstufen (PISA, z.B. Deutsches PISA - Konsortium, 2003) Kompetenzniveaus (Lernstand 9, z.B. Peek & Döbelstein, 2006) oder wie bei DESI und VERA in Fähigkeitsniveaus unterteilt. Das Kontinuum lässt sich von sehr gering ausgeprägten bis hin zu sehr entwickelten Fähigkeiten unterteilen. Dabei liegt die Vorstellung zugrunde, dass verschiedene Schwierigkeitsniveaus in einem hierarchischen Verhältnis zueinander stehen. Die Aussage „Schülerinnen und Schüler befinden sich in einem Inhaltsbereich X auf Niveau 2“ heißt soviel wie:

- Aufgaben mittleren Anforderungsniveaus (Fähigkeitsniveau 2) werden in diesem Inhaltsbereich mit hinreichender Sicherheit (Lösungswahrscheinlichkeit von 62,25 Prozent) gelöst.
- Aufgaben des Fähigkeitsniveaus 1, also einfachere Aufgaben mit grundlegenden Anforderungen, werden mit höherer Sicherheit gelöst.
- Anspruchsvollere Aufgaben, die sich auf dem Fähigkeitsniveau 3 befinden, werden dagegen mit geringerer Wahrscheinlichkeit gelöst.

Im Projekt VERA wurden die Fähigkeitsniveaus auf der Basis inhaltlicher und theoretischer Vorgaben in Kooperation mit Experten (Fachdidaktiker, Fachwissenschaftler, praktisch tätige Lehrkräfte und Psychologen) und ausführlicher empirischer Normierungsstudien im Rahmen der Vergleichsarbeiten 2004 entwickelt und basierend auf den Normierungsergebnissen im Jahr 2005 weiterentwickelt. Die Beschreibungen für den jeweiligen Inhaltsbereich in Deutsch und Mathematik wurden in Hinsicht auf die mit hinreichender Sicherheit zu bewältigenden Anforderungen konkretisiert. Der Entwicklungsprozess führt also empirisch erfasste Informationen und theoretische Überlegungen zusammen:

- Auf der einen Seite benötigt man die empirische Schwierigkeitsverteilung der Aufgaben, d.h. Informationen darüber, von wie vielen Schülerinnen und Schülern

die Aufgaben richtig gelöst wurden.

- Auf der anderen Seite sind Annahmen über Denkprozesse erforderlich, die der Aufgabebearbeitung zugrunde liegen, sowie Informationen darüber, was Aufgaben eigentlich unterschiedlich leicht oder schwer macht („schwierigkeitsbestimmende Merkmale“). Dies erfordert fachdidaktische und psychologische Expertise: Aufgrund welcher Merkmale, welcher Anforderungen an Denk- und Gedächtnisprozesse lassen sich Aufgaben charakterisieren?

Die Entwicklung der VERA-Fähigkeitsniveaus erfolgte in technischer Hinsicht in mehreren Schritten: Zunächst wurden die Aufgaben jedes Inhaltsgebietes hinsichtlich des geschätzten Schwierigkeitsparameters aufsteigend sortiert und vorläufig zu Gruppen zusammengefasst. Dann wurden die Aufgabengruppen von Experten in Hinsicht auf Aufgabeninhalte und Anforderungsbereiche (elementare, erweiterte und fortgeschrittene Fähigkeiten) analysiert. Dabei stand die Differenzierung gegenüber dem nächst niedrigeren Niveau im Fokus der Beschreibung. Es resultierte eine erste holistische Beschreibung der typischen Anforderungen jedes Niveaus. Diese Beschreibung wurde im zweiten Schritt verfeinert, indem zusätzlich Prozessüberlegungen (z.B. notwendige Problemlöseprozeduren) und schwierigkeitsbestimmende Merkmale der Aufgaben (z.B. Eindeutigkeit der Lösung, sprachliche Komplexität, Antwortformat) einbezogen wurden. Als Abschluss der Pilotierungs- und Normierungsstudien im Jahr 2005 wurden die aus den Vergleichsarbeiten 2004 hervorgegangenen Fähigkeitsniveau-Beschreibungen noch einmal mit den neu entwickelten Aufgaben abgeglichen und ggf. präzisiert.

In diesem Zusammenhang wird deutlich, dass die Beschreibungen von Fähigkeitsniveaus einen Beitrag zur Ausbildung von bereichsspezifischen Kompetenzmodellen darstellen, jedoch zur Zeit noch nicht von einer abgeschlossenen Theoriebildung gesprochen werden kann, wie es die Verwendung des „Kompetenz“-Begriffs nahe legt. Die bei VERA erhobenen Leistungen sind als fachlich definierte Fähigkeiten beschrieben, die sich auf das Beherrschen der gesamten Breite des jeweiligen Inhaltsgebietes (inkl. der notwendigen Vorkenntnisse) beziehen, also nicht nur – wie bei Klassenarbeiten – auf die Meisterung des vor kurzem durchgenommenen Unterrichtsstoffs.

Beschreibung der Fähigkeitsniveaus im Fach Deutsch

Für das Fach Deutsch wurden bei VERA Fähigkeitsniveaus in den vier Inhaltsbereichen Leseverständnis, Schreiben, Sprachbetrachtung und Orthografie entwickelt. Das aufgefundene Fähigkeitsspektrum wurde, wie auch in Mathematik, in drei Bereiche unterteilt. Für jede Schülerin bzw. jeden Schüler wird gemäß der erfassten Testleistung jeweils ein Fähigkeitsniveau in den vier Inhaltsbereichen ermittelt. Die Zuordnung besagt, dass die für dieses Niveau formulierten Anforderungen mit hinreichender Sicherheit bewältigt werden.

Die Fähigkeitsniveaus können bereichsübergreifend wie folgt beschrieben werden:

- **Fähigkeitsniveau 1: Elementare bzw. grundlegende Fähigkeiten**
Einfache Aufgaben mit grundlegenden Anforderungen werden hinreichend sicher gelöst.

- **Fähigkeitsniveau 2: Erweiterte Fähigkeiten**

Aufgaben mittleren Anforderungsniveaus werden hinreichend sicher gelöst.

- **Fähigkeitsniveau 3: Fortgeschrittene Fähigkeiten**

Es werden auch anspruchsvollere Aufgaben hinreichend sicher gelöst.

Liegen keine oder extrem unvollständige Daten vor, ist eine Zuordnung zu den beschriebenen Fähigkeitsniveaus nicht möglich. Dies wird bei VERA als „nicht auswertbare Leistung (n.a.L.)“ bezeichnet.

Da in den Vergleichsarbeiten 2005 im Fach Deutsch der Inhaltsbereich Leseverständnis fokussiert wurde, werden im Folgenden dafür die konkreten Anforderungen der Fähigkeitsniveaus dargestellt.

Leseverständnis

Fähigkeitsniveau 1: Elementare Fähigkeiten (Gewinnung von Einzelinformationen)

- Im Rahmen von Auswahlaufgaben werden einzelne Informationen im Lesetext gefunden, wenn sie in der Aufgabenstellung fast wortgleich oder in etwa bedeutungsgleich formuliert sind.
- Es kann angegeben werden, wie einzelne Stellen im Text unabhängig vom Kontext verstanden werden sollten.
- In Auswahlaufgaben können auf der Basis von elementarem Wissen einfache Schlüsse gezogen werden.
- Auf der Basis von elementarem Wissen und/oder Bildinformationen können einfache Schlüsse gezogen werden.

Fähigkeitsniveau 2: Erweiterte Fähigkeiten (Verknüpfung von mehreren Informationen)

- Über den Text verteilte Informationen können gefunden und miteinander verknüpft werden. Die zentrale Aussage eines Textes kann wiedergegeben werden.
- Im Rahmen von Auswahlaufgaben kann angegeben werden, wie einzelne Stellen im Text verstanden werden sollten – selbst dann, wenn dazu weitere Informationen im Text herangezogen werden müssen.
- Texte können ansatzweise plausibel beurteilt werden: Dabei wird vor allem auf inhaltliche Aspekte und persönliche Vorlieben Bezug genommen.
- Nahe liegende Schlüsse können unter Nutzung von einzelnen oder mehreren Informationen im Text und weit verbreitetem Wissen gezogen werden und/oder mit Textstellen belegt werden.

Fähigkeitsniveau 3: Fortgeschrittene Fähigkeiten (Komplexere Schlussfolgerungen)

- Über den Text verteilte Informationen können gefunden und miteinander verknüpft werden. Das gelingt auch dann, wenn diese Informationen in der Aufgabenstellung weder wortgleich noch sinngemäß vorkommen.
- Komplexere Schlussfolgerungen können gezogen und dargestellt werden, auch

wenn ein Kurztext zu schreiben ist. Dazu wird spezielleres Sach- und manchmal auch Sprachwissen benötigt. Texte können angemessen beurteilt werden. Dabei wird vor allem sowohl auf inhaltliche Aspekte und persönliche Vorlieben als auch auf formale Aspekte des Textes Bezug genommen.

2.2 Mathematik

2.2.1 Aufgabenentwicklung

Die Mathematikaufgaben waren ursprünglich von einer Gruppe erfahrener Lehrkräfte, maßgeblich beraten durch den Fachdidaktiker Prof. Dr. Jens-Holger Lorenz, auf der Basis des Grundschul-Rahmenplans des Landes Rheinland-Pfalz entwickelt worden. Dazu wurde ein zweidimensionales Raster zur Klassifikation der Aufgaben entworfen. Die beiden Dimensionen orientierten sich zum einen an den drei klassischen mathematischen *Inhaltsbereichen* in der Grundschule und zum anderen an verschiedenen *Tätigkeitsanforderungen*.

Für die Vergleichsarbeiten im Herbst 2005 erfolgte erstmalig eine explizite Orientierung der Aufgabenkonstruktion an den im Dezember 2004 von der KMK verabschiedeten Bildungsstandards². Alle Aufgaben wurden dafür entlang der fünf definierten **inhaltsbezogenen** mathematischen Kompetenzen (Zahlen und Operationen, Raum und Form, Muster und Strukturen, Größen und Messen, Daten, Häufigkeit und Wahrscheinlichkeit) und den definierten **allgemeinen** mathematischen Kompetenzen (Problemlösen, Kommunizieren, Argumentieren, Modellieren, Darstellen) entwickelt. Zusätzlich wurden alle Aufgaben wie im Jahr 2004 den „traditionellen“ Sachgebieten Arithmetik, Geometrie, Sachrechnen/Größen zugeordnet. Die genaue inhaltliche Klassifikation der Aufgaben gewährleistet eine umfassende Repräsentation der Bildungsstandards sowie aller Bereiche des Curriculums und der allgemeinen Anforderungen des Mathematikunterrichts.

Zunächst wurde eine umfangreiche Sammlung von Aufgaben (> 500) erzeugt, die dann in sukzessiven Auswahlrunden reduziert und optimiert wurde. Dabei spielten neben klassischen Kriterien der Testentwicklung (Objektivität, Reliabilität, Validität) auch Überlegungen zur Durchführbarkeit, der erforderlichen Bearbeitungszeit und der Auswertungsökonomie eine Rolle. So wurden z.B. alle Aufgaben, die besondere technische Hilfsmittel wie beispielsweise einen Zirkel erfordern, ausgeschlossen, da nicht zu gewährleisten war, dass jede Schülerin und jeder Schüler in der Vergleichsarbeit über die entsprechenden Hilfsmittel würde verfügen können.

² vgl. http://www.kmk.org/schul/Bildungsstandards/Grundschule_Mathematik_BS_307KMK.pdf

2.2.2 Skalierung und Fähigkeitsniveaus des Mathematiktests

Die Skalierung in Mathematik erfolgte analog zur oben geschilderten Skalierung in Deutsch, d.h. es wurde für jeden der drei Inhaltsbereiche ein ein-parametrisches Modell (Rasch-Modell) angepasst. Wie auch in Deutsch wird von einer hierarchischen Beziehung der Fähigkeitsniveaus ausgegangen. Die Aussage „Schülerinnen und Schüler befinden sich im Bereich Arithmetik auf Fähigkeitsniveau 2“ heißt soviel wie:

- Aufgaben mittleren Anforderungsniveaus (Fähigkeitsniveau 2) werden in diesem Inhaltsbereich mit hinreichender Sicherheit (Lösungswahrscheinlichkeit von 62,25 Prozent) gelöst.
- Aufgaben des Fähigkeitsniveaus 1, also Aufgaben mit grundlegenden Anforderungen, werden mit höherer Sicherheit gelöst.
- Anspruchsvollere Aufgaben, die sich auf dem Fähigkeitsniveau 3 befinden, werden dagegen mit geringerer Wahrscheinlichkeit gelöst.

Das Vorgehen bei der Definition der Fähigkeitsniveaus umfasste – genau wie bei Deutsch – im Wesentlichen zwei Schritte: Im ersten Schritt wurden die Aufgaben jedes Inhaltsgebietes hinsichtlich des geschätzten Schwierigkeitsparameters aufsteigend sortiert und vorläufig zu Gruppen zusammengefasst. Dann wurden die Aufgabengruppen über eine Analyse der Aufgabeninhalte und Anforderungen charakterisiert, wobei jeweils die Differenzierung gegenüber der nächst niedrigeren Stufe im Fokus der Beschreibung stand. Es resultierte eine holistische Beschreibung der typischen Anforderungen jeder Stufe.

Im Folgenden werden die Fähigkeitsniveaus getrennt für die einzelnen Bereiche Arithmetik, Geometrie und Sachrechnen/Größen dargelegt.

Arithmetik

Fähigkeitsniveau 1: Elementare Fähigkeiten (Grundlegende Kenntnisse arithmetischer Verfahren)

- Schriftliche Additionsaufgaben werden gelöst.
- Subtraktionsaufgaben ohne Übertrag werden gelöst und Ergänzungen zum nächsten Hunderter/Tausender gelingen.
- Einfache kombinatorische Aufgaben können gelöst werden (z.B. das Zusammensetzen von Zahlen aus drei Ziffern).
- Die Beziehung zwischen Zahlen und ihrer Darstellung in der Stellentafel kann hergestellt werden.
- Das Zahlbildungsprinzip wird beherrscht, innerhalb einfacher Aufgaben auch bei Zahlenfolgen.
- Die Aufgabenlösung gelingt bei sprachlich einfachen und kurzen Texten.
- In Gleichungen kann eine fehlende Rechenoperation eingesetzt werden.

Fähigkeitsniveau 2: Erweiterte Fähigkeiten (Umfassende Kenntnis der Addition und Subtraktion)

- Schriftliche Addition gelingt auch mit Überträgen in unüblichen Formaten (z.B. Lückenaufgaben).
- Schriftliche Subtraktion gelingt auch mit Überträgen oder in unüblichen Formaten (z.B. Lückenaufgaben).
- Kombinatorische Aufgaben, die zusätzlich einfache Rechnungen erfordern, können gelöst werden.
- Einsicht in das Stellenwertsystem liegt vor und der Stellenwert kann in unterschiedlichen Darstellungsformen erkannt werden.
- Das Vervollständigen von Zahlenreihen gelingt, wenn die zugrunde liegende Regel vorwiegend Strichrechnung erfordert.
- Aufgaben mit sprachlich komplexeren und längeren Texten werden gemeistert.
- Arithmetische Aufgaben mit ausgeprägten sprachlichen oder bildlichen Anteilen können gelöst werden.

Fähigkeitsniveau 3: Fortgeschrittene Fähigkeiten (Flexible Beherrschung der Grundrechenarten)

- Zahlen und Operationen können flexibel kombiniert werden. Hierbei werden mathematische Kenntnisse (z.B. Rechengesetze, Teilbarkeitsregeln) korrekt angewendet.
- Schriftliche Subtraktion gelingt auch mit Überträgen in unüblichen Formaten (z.B. Lückenaufgaben).
- Eine kombinatorische Problemstellung kann vollständig modelliert werden.
- Einsicht in das Stellenwertsystem liegt vor und Veränderungen können, z.B. durch geeignete arithmetische Operationen, vorgenommen werden.
- Das Erkennen und Benennen einer Zahlenfolge gelingt.
- Überschlagsrechnungen und Rundungen zum Tausender können vorgenommen werden.
- Mehrschrittige Rechnungen werden unter Berücksichtigung der Regel „Punktrechnung vor Strichrechnung“ bewältigt.
- Das Finden, Erklären und Korrigieren von Fehlern in schriftlichen Additionen oder Subtraktionen gelingt.
- Zahlen können durch geeignete Operationen zu einer Zielzahl kombiniert werden.

Geometrie

Fähigkeitsniveau 1: Elementare Fähigkeiten (Kenntnisse grundlegender geometrischer Formen und Abbildungen)

- Aufgaben zu Umfang, Flächeninhalt und Volumen, die sich durch einfache Operationen wie z.B. Abzählen lösen lassen, werden bewältigt.
- Ebene Figuren werden in Körpern wieder erkannt.
- Geometrische Körper werden in Alltagsgegenständen wieder erkannt.
- Ebene geometrische Formen können durch Verschiebung/Drehung zu anderen Formen kombiniert werden.
- Einfache geometrische Muster werden erkannt und können fortgeführt werden.
- Die Aufgabenlösung gelingt bei sprachlich einfachen und kurzen Texten.
- Das Ergänzen zu achsensymmetrischen Figuren gelingt.

Fähigkeitsniveau 2: Erweiterte Fähigkeiten (Erkennen und Zuordnen von ebenen und räumlichen Figuren)

- Flächeninhalte können verglichen werden, wenn verschiedene Antwortalternativen vorgegeben sind.
- Aufgaben zu Körpernetzen können gelöst werden.
- Geometrische Begriffe sind bekannt und können angewendet werden.
- Rechtecke können nach vorgegebenen Maßen konstruiert werden.
- Komplexere geometrische Muster können fortgesetzt werden.
- „Fehler“ in Mustern werden erkannt.
- Strecken können in ihrer Ausdehnung mit einer vorgegebenen Maßeinheit erfasst werden, auch wenn dabei halbe Maßeinheiten berücksichtigt werden müssen.
- Baupläne und Würfelbauten können einander zugeordnet werden.
- Über eine Raumvorstellung kann ein Perspektivenwechsel vollzogen werden, wenn die jeweiligen Ansichten keine Überdeckungen der einzelnen Körper aufweisen.
- Das Navigieren in Feldern bei vorgegebenen Koordinaten gelingt.

Fähigkeitsniveau 3: Fortgeschrittene Fähigkeiten (Konstruktionen in der Ebene und im Raum; Erkennen von Beziehungen zwischen geometrischen Begriffen)

- Flächeninhalte können in offenen Aufgabenstellungen bestimmt werden.
- Würfelnetze, Muster und räumliche Darstellungen einfacher Körper können vervollständigt werden.
- Körpereigenschaften können sachgerecht bezeichnet werden.
- Das Zerlegen einer Fläche in vorgegebene Figuren gelingt.

- Figuren in rechtwinkligen Rastern können in nicht-rechtwinklige Raster abgebildet (verzerrt) werden.
- Auch sprachlich komplexere, mehrschrittige Aufgaben werden gelöst.
- Es werden alle Spiegelachsen in ebenen Figuren erkannt.
- Das Erstellen von Bauplänen gelingt.
- Über eine Raumvorstellung kann ein Perspektivenwechsel vollzogen werden, auch wenn die jeweiligen Ansichten Überdeckungen der einzelnen Körper aufweisen.
- Punkte im Koordinatensystem werden gefunden.
- Die Relationen „senkrecht zueinander“ und „waagrecht zueinander“ werden beherrscht.
- Einfache Sachrechenaufgaben mit geometrischen Inhalten können gelöst werden.
- Seltener Körper (z.B. Pyramide, Kegel) sind bekannt.

Sachrechnen/Größen

Fähigkeitsniveau 1: Elementare Fähigkeiten (Grundlegende Fähigkeiten im Umgang mit Größen, Darstellungen und Wahrscheinlichkeit)

- Vertraute Maßeinheiten können bei einfachen Aufgaben geordnet, umgerechnet und verglichen und es kann mit ihnen gerechnet werden (Längen-, Zeit-, Gewichts- und Geldeinheiten).
- Die Anwendung einschränkter Operationen in authentischen Aufgaben gelingt. Die Anwendung von Addition (auch wiederholte Additionen) und Subtraktion in authentischen Aufgaben gelingt bei Aufgaben mit Auswahl aus vorgegebenen Lösungen.
- Daten können aus übersichtlich gestalteten Tabellen und Darstellungen entnommen und ggf. Folgerechnungen (einschrittige Addition/Subtraktion) vorgenommen werden.
- Wahrscheinlichkeitsaussagen zu den Begriffen sicher / unmöglich / möglich, aber nicht sicher werden korrekt kategorisiert.
- Einfache Größenvorstellungen sind vorhanden.
- Offensichtlich unlösbare Aufgaben werden erkannt.

Fähigkeitsniveau 2: Erweiterte Fähigkeiten (Entwickelte Fähigkeiten im Umgang mit Größen, Darstellungen und Wahrscheinlichkeit)

- Im Umgang mit vertrauten Maßeinheiten (Längen, Zeit-, Gewichts- und Geldeinheiten) können Aufgaben bis in den Tausender-Zahlenraum gelöst werden.
- Lösungen von authentischen Aufgaben, die Umrechnungen von Maßeinheiten erfordern, gelingen.
- Die Zuordnung arithmetischer Operationen/Relationen zu Sachsituationen gelingt.

- Daten können aus komplexeren Tabellen und Darstellungen entnommen und ggf. Folgerechnungen (Addition/Subtraktion) korrekt durchgeführt werden.
- Kenntnisse zur Wahrscheinlichkeit sind vorhanden und können in alltagsbezogenen Sachsituationen angewendet werden.
- Der Umgang mit elementaren Brüchen gelingt.
- Rundungen und Schätzungen gelingen bei Aufgaben mit vorgegebenen Lösungen.
- Verknüpfungen von Operationen werden bewältigt.
- Aufgaben mit mehreren zu verarbeitenden Größen werden gemeistert.

Fähigkeitsniveau 3: Fortgeschrittene Fähigkeiten (Eigenständige Problemlösungen)

- Funktionale Beziehungen zwischen Maßen können eigenständig hergestellt und verglichen werden.
- Aufgaben, die mehrere Teilschritte umfassen, werden beherrscht.
- Sprachlich formulierte Relationen können in arithmetische Terme übersetzt werden.
- Bei Aufgaben ohne vorgegebene Fragestellung kann eigenständig eine Aufgabe formuliert und bearbeitet werden.
- Unlösbare Aufgaben, die eine mentale Vorstellung des geschilderten Szenarios erfordern, werden erkannt.
- Die mathematische Modellierung problemhaltiger Sachsituationen gelingt.

Um den Zusammenhang zwischen den Aufgaben und der Zuordnung zu den Fähigkeitsniveaus insbesondere für die Lehrkräfte zu verdeutlichen und um Anreize zur Unterrichtsgestaltung zu geben, wurden sowohl für Mathematik als auch für Deutsch „Didaktische Erläuterungen“ (vgl. <http://www.uni-landau.de/vera/aufgaben2005.htm>) im Internet zur Verfügung gestellt.

2.3 Gesamt-Ergebnisse für die Fähigkeitsniveaus

2.3.1 Gesamt-Verteilung

Die zusammenfassende Darstellung der teilnehmenden Bundesländer ermöglicht einen allgemeinen Blick auf relative Stärken und Schwächen. Ein Vergleich zwischen den Inhaltsbereichen ist jedoch nur bedingt möglich, da es sich zwar um vergleichbare formale Strukturen (Niveaus), aber um verschiedene Inhalte handelt. In den Fähigkeitsniveaus wird in erster Linie beschrieben, was Schülerinnen und Schüler in der vierten Grundschulklasse können. Dementsprechend ist ein Vergleich der Inhaltsbereiche nur über die Fähigkeitsniveaueinteilungen sinnvoll.

Gesamtverteilung 2005

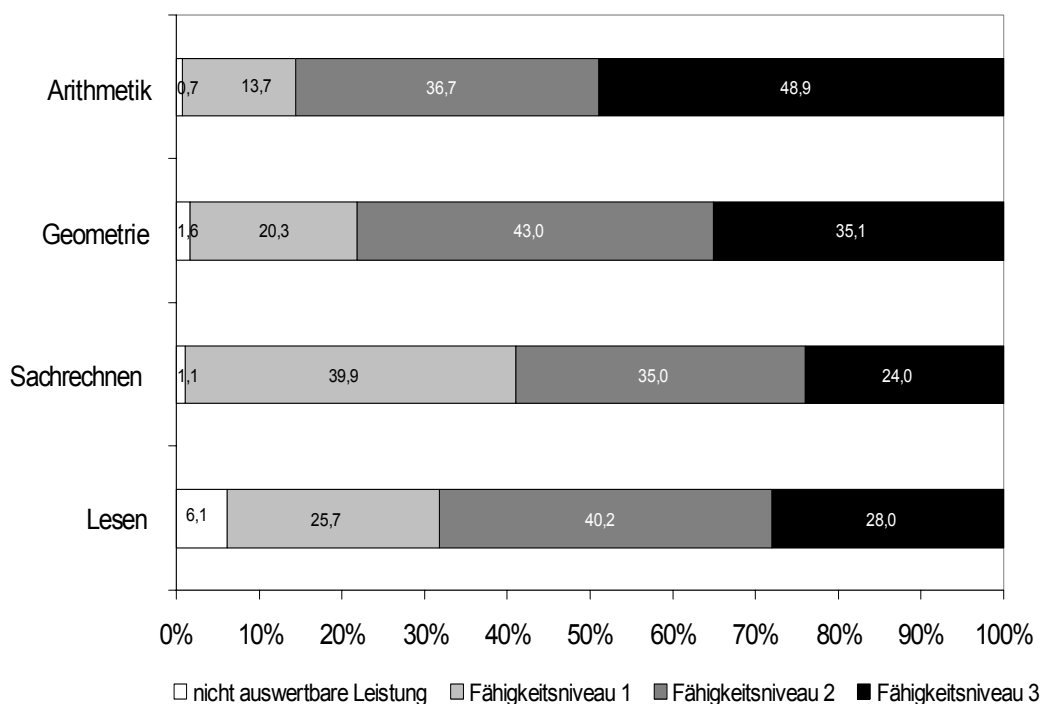


Abbildung 2: Gesamtverteilung der Fähigkeitsniveaus über alle Länder hinweg; Angaben in Prozent

Für das Fach *Mathematik* wird deutlich, dass in den Inhaltsbereichen Arithmetik und Geometrie der Großteil der Schülerinnen und Schüler erweiterte bis fortgeschrittene Fähigkeiten aufweist, also das Fähigkeitsniveau 2 und 3 erreicht (für Arithmetik deutlich mehr als 80 Prozent, für Geometrie knapp unter 80 Prozent). Der Anteil an Kindern auf dem dritten Fähigkeitsniveau ist für Arithmetik am höchsten (48,9 Prozent). Demgegenüber fällt auf, dass im Sachrechnen verhältnismäßig wenige Schülerinnen und Schüler dem dritten Fähigkeitsniveau (24 Prozent) und knapp 40 Prozent der Schülerinnen und Schüler einem Niveau, das höchstens das Beherrschen elementarer Aufgaben umfasst, zugeordnet werden können.

Für das Fach *Deutsch* im Inhaltsbereich Leseverständnis erreicht ein großer Anteil der Schülerinnen und Schüler höchstens ein Fähigkeitsniveau, das elementaren Fähigkeiten entspricht (31,8 Prozent).

Aus der Verteilung der Ergebnisse kann zusammenfassend geschlussfolgert werden, dass vor allem beim Leseverständnis und beim Sachrechnen ein Förderbedarf besteht.

2.3.2 Zusammenhänge zwischen den Fähigkeitsniveaus

Die folgende Tabelle 1 gibt den Zusammenhang (Interkorrelationen) zwischen den Leistungen in den verschiedenen Inhaltsbereichen wieder. Darin entsprechen die dunkelgrauen Felder den Interkorrelationen innerhalb des Faches Mathematik und die

hellgrauen Felder den Zusammenhängen zwischen den Inhaltsbereichen für Mathematik und dem Leseverständnis.

Die Korrelationen liegen in einem mittleren Wertebereich von $r = ,51$ (Leseverständnis/Geometrie) bis $r = ,59$ (Arithmetik/Sachrechnen) – dabei finden sich höhere Korrelationen innerhalb der Fachs Mathematik. Die Korrelationen zwischen dem Leseverständnis und den drei Mathematik-Inhaltsbereichen könnten als Hinweis für die fächerübergreifende Bedeutung des Leseverständnisses interpretiert werden.

Tabelle 1: Interkorrelationen zwischen den Leistungen in den Inhaltsbereichen (r)*2005

	Arithmetik	Geometrie	Sachrechnen
Lesen	0,53	0,51	0,52
Arithmetik		0,55	0,59
Geometrie			0,53

* minimale Stichprobengröße: $N = 280224$ (Kinder)

Alles in allem sprechen die mittleren, jedoch nicht perfekten Zusammenhänge zwischen den vier Inhaltsbereichen dafür, zwischen Arithmetik, Geometrie und Sachrechnen und dem Leseverständnis im Fach Deutsch zu differenzieren.

3 Berliner Ergebnisse

3.1 Fähigkeitsniveaus

Im Folgenden werden die Fähigkeitsniveaus unter verschiedenen Gesichtspunkten diskutiert: Zunächst wird allgemein die Verteilung der Schüler auf den einzelnen Niveaus dargestellt (vgl. 3.1.1, S. 20) sowie Veränderungstrends für die Vergleichsarbeiten in den Jahren 2004 und 2005 diskutiert (vgl. 3.1.2, S. 21). Im Anschluss werden Unterschiede zwischen (bzw. innerhalb) den untersuchten Klassen/Schulen (siehe 3.1.3, S. 22), Geschlechtsunterschiede (siehe 3.1.4, S. 24) sowie Unterschiede zwischen Schülerinnen und Schülern mit Deutsch als dominanter vs. nicht dominanter Sprache (vgl. 3.1.5, S. 25) genauer beleuchtet. Alle Angaben zu den Individualmerkmalen der Schülerinnen und Schüler stammen von den Lehrkräften und wurden vor der Durchführung der Vergleichsarbeiten online im „geschützten Bereich“ erfasst.

3.1.1 Verteilung der Fähigkeitsniveaus in den Ländern

Im Fach *Mathematik* wird deutlich, dass in den Inhaltsbereichen Arithmetik und Geometrie der Großteil der Schülerinnen und Schüler erweiterte bis fortgeschrittene Fähigkeiten aufweist (für Arithmetik 80,1 Prozent, für Geometrie 66,6 Prozent). Während sich für Arithmetik in etwa eine Gleichverteilung auf die Fähigkeitsniveaus zwei und drei zeigt, besetzen in Geometrie mehr Kinder den mittleren Niveaubereich. Im Sachrechnen erreichen verhältnismäßig wenige Schülerinnen und Schüler das dritte Fähig-

keitsniveau (20,3 Prozent) und 46,1 Prozent der Schülerinnen und Schüler ein Niveau, das höchstens das Beherrschen elementarer Aufgaben umfasst. Im Inhaltsbereich *Leseverständnis* gibt es einen auffällig großen Anteil an Schülerinnen und Schülern, die höchstens das Fähigkeitsniveau eins erreichen (39,8 Prozent, davon 10,7 Prozent nicht auswertbare Leistung).

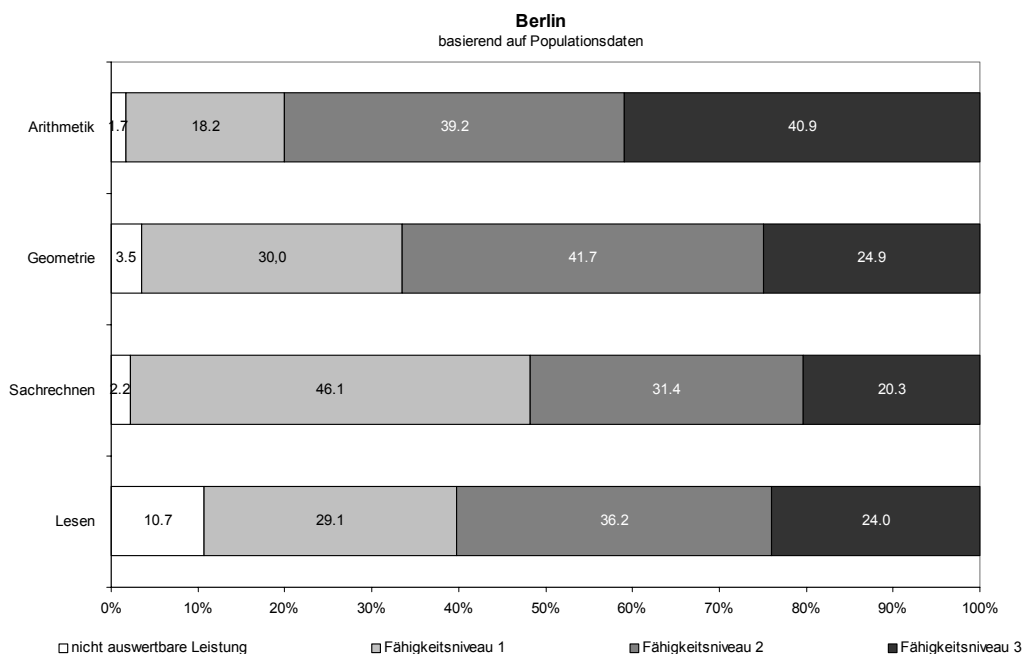


Abbildung 3: Gesamtverteilung der Fähigkeitsniveaus (2005); Angaben in Prozent

Alles in allem fällt im Vergleich mit der länderübergreifenden Fähigkeitsniveauverteilung auf, dass in Berlin in allen erfassten Inhaltsbereichen weniger Kinder dem dritten Fähigkeitsniveau und damit mehr Kinder dem Fähigkeitsniveau eins bzw. nicht auswertbaren Leistungen zugeordnet wurden: Im Leseverständnis und Sachrechnen scheint angesichts der häufigen Besetzung des unteren Fähigkeitsniveaus ein besonders großer Förderbedarf zu bestehen. Der hohe Anteil an Kindern mit nicht auswertbarer Leistung im Inhaltsbereich Leseverständnis steht vermutlich im Zusammenhang mit den Kontextbedingungen, v.a. mit der Sprachbeherrschung (siehe z.B. 3.1.5, S. 25).

3.1.2 Veränderungstrends 2004–2005

Die Verteilung der Schülerinnen und Schüler auf die einzelnen Fähigkeitsniveaubereiche ist in 2005 verglichen mit der Verteilung in 2004 (Abbildung 4) in etwa gleich geblieben. In keinem der Inhaltsbereiche ergeben sich signifikante Unterschiede der Fähigkeitsniveauverteilungen zwischen den Jahren.

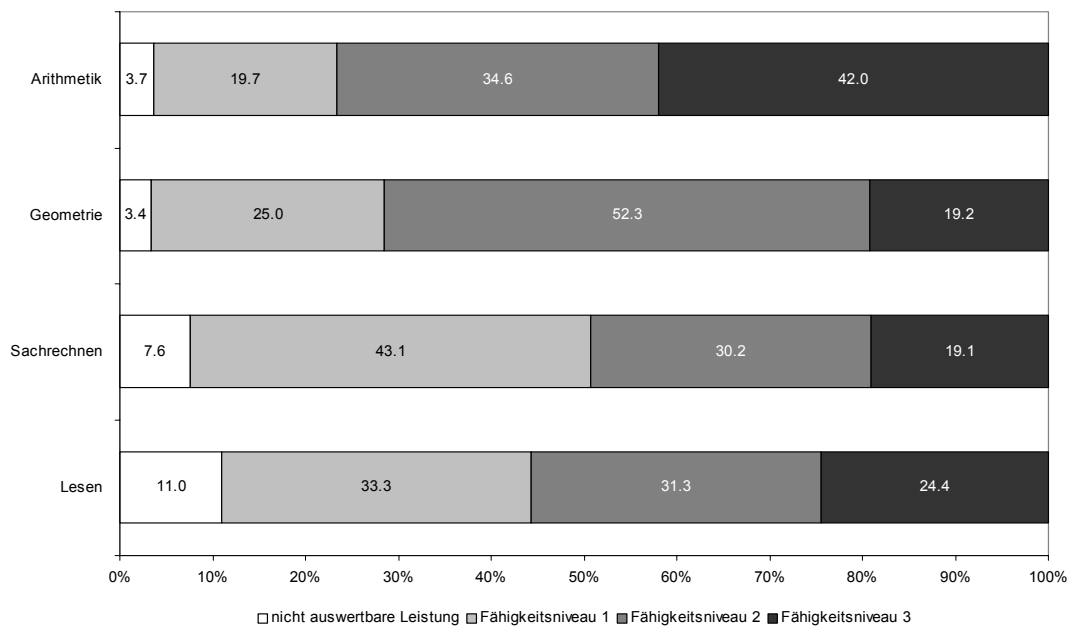


Abbildung 4: Gesamtverteilung der Fähigkeitsniveaus (2004); Angaben in Prozent

3.1.3 Unterschiede innerhalb und zwischen den Klassen

Leistungsunterschiede zwischen Schülern gehen nicht ausschließlich auf Merkmale zurück, welche mit dem Schüler als einzelner Person verknüpft sind (z.B. Geschlecht oder Erstsprache), sondern sind zu einem mehr oder minder großen Anteil auf die Zugehörigkeit zu einer bestimmten Schulklasse und einer bestimmten Schule zurückzuführen. Wissenschaftlich gesprochen, lässt sich die Leistungsvarianz also in Anteile der Individual-, der Klassen- und der Schulebene zerlegen. Ein immer wiederkehrender empirischer Befund aus Schulleistungsstudien ist, dass schulische Leistungsunterschiede zu einem überwiegenden Anteil auf interindividuelle Differenzen zurückgeführt werden können (Helmke & Weinert, 1997). Dies stellt eine teilweise, aber keineswegs weitgehende Relativierung der Bedeutung von Schule dar, da Unterricht zwar einen begrenzten, durch die Zusammenfassung in Klassen- und Schulverbänden jedoch weit streuenden Effekt hat: Von ungünstigen Unterrichts- und Kontextbedingungen ist jeweils nicht einer, sondern sind eine Reihe von Schülern betroffen.

Zerlegt man die (interindividuelle) Leistungsvarianz der Schüler, welche an den Vergleichsarbeiten teilgenommen haben, so resultieren die in Tabelle 2 dargestellten Prozentanteile.

Table 2: Zerlegung in die Varianz auf den drei Ebenen Schule, Klasse und Individuum;
Angaben in Prozent

Bereich	Schulebene	Klassenebene	Individualebene
Arithmetik	21,1	8,3	70,6
Geometrie	28,0	11,9	60,1
Sachrechnen	16,8	6,9	76,3
Lesen	24,1	7,8	68,1
N min	373	984	21132

Wie erwartet ist die Leistung in den Vergleichsarbeiten vorrangig mit Merkmalen der Individualebene verknüpft. Die Varianzanteile liegen zwischen 60,1 Prozent (Geometrie) und 76,3 Prozent (Sachrechnen). Ebenfalls erhebliche Anteile gehen auf schulische Merkmale zurück, sie bewegen sich zwischen 16,8 Prozent (Sachrechnen) und 28,0 Prozent (Geometrie). Den vergleichsweise geringsten Beitrag leistet die Zugehörigkeit zu einer Schulklasse (hinter der Unterschiede des Unterrichts und des Klassenkontextes stehen) mit 7,8 Prozent (Leseverständnis) bis 11,9 Prozent (Geometrie).

Nimmt man die Effekte der Schul- und Klassenebene zusammen, resultieren durchaus erhebliche Einflüsse schulischer Qualitätsmerkmale. Zudem können Effekte der Individualebene durch Merkmale der Klassen- bzw. Schulebene moderiert werden, der Einfluss des sozioökonomischen Hintergrunds kann beispielsweise von Schule zu Schule und von Klasse zu Klasse variieren. Andererseits sind Unterschiede zwischen Schulen und Klassen nicht etwa unabhängig von individuellen Faktoren, sondern zum Teil auf den Einfluss aggregierter Individualvariablen (z.B. den mittleren sozialen Status der Schülerschaft) zurückzuführen. Zusammenfassend darf der hohe Varianzanteil auf Individualebene nicht dazu verleiten, Unterricht und Schule für nebensächlich oder gar unbedeutend zu halten. Zum einen beeinflussen schulische Lernumgebungen nicht nur den einzelnen Schüler, sondern jeweils gesamte Schul- und Klassenverbände. Damit sind auch kleine Effekte bedeutsam, da sie immer eine größere Anzahl an Schülern betreffen. Zum anderen sollte die Wirkung von Schule nicht ausschließlich mit Blick auf Leistungsunterschiede beurteilt werden: Ohne Unterricht in Schulen erscheint der Aufbau persönlich und gesellschaftlich unentbehrlichen Wissens und vielfältiger kognitiver Fertigkeiten nahezu unmöglich (Helmke, Hosenfeld & Schrader, 2002, S. 420f.).

Der verhältnismäßig umfangreiche Varianzanteil der Schulebene, verglichen mit der Klasse, ist auf den ersten Blick überraschend und widerspricht den Ergebnissen anderer Studien, z.B. MARKUS (Hosenfeld, Helmke, Ridder & Schrader, 2001). Er erklärt sich vermutlich einerseits aus der Möglichkeit, einen Teil der zu bearbeitenden Aufgaben selbst auszuwählen. Die gemeinsame Auswahl der Aufgaben begünstigt infolge der notwendigen Abstimmung im Kollegium eine Leistungshomogenisierung *innerhalb* der Schulen, während Unterschiede *zwischen* den Schulen durch unterschiedliche Vorgehensweisen bei der Auswahl akzentuiert werden können.

Andererseits reflektiert der hohe Varianzanteil auf Schulebene auch die deutlichen Unterschiede in den Kontextbedingungen (insbes. Einzugsgebiet, soziotopisches Profil) der Schulen. So finden sich *innerhalb* der Schüler einer Schule oft keine allzu ausgeprägten Differenzen bezüglich Sozialschicht, Erwerbstätigkeit der Eltern usw., während diese *zwischen* den Schulen als Folge unterschiedlicher Einzugsgebiete erheblich variieren können.

Auf Klassenebene fällt der relativ große (und auf Individualebene entsprechend kleine) Varianzanteil im Bereich Geometrie ins Auge. Dieser Effekt der Klassenzugehörigkeit könnte die curriculare „Stiefkindrolle“ der Geometrie widerspiegeln (Blum et al., 2004, S. 66): Da dieser Bereich in den Lehrplänen in der Vergangenheit traditionell eine eher untergeordnete Rolle gespielt hat (im Gegensatz insbesondere zur Arithmetik), oblagen Entscheidungen zu Umfang und Art der Behandlung dieses Stoffgebiets verstärkt den Lehrkräften selbst. Diese relativ großen Handlungsspielräume könnten sich im Sinne einer Verstärkung von Leistungsunterschieden im Bereich Geometrie auswirken. Dass beträchtliche Unterschiede im Geometrieleistungsniveau auch auf *Schulebene* zu finden sind, könnte ein Indikator dafür sein, dass Entscheidungen zu Umsetzung und Ausgestaltung der Lehrpläne nicht nur Sache der einzelnen Lehrkraft sind, sondern auch im Rahmen der Organisationseinheit Schule getroffen werden.

3.1.4 Leistungen von Mädchen und Jungen

Das Geschlecht von Schülerinnen und Schülern ist ein weiterer schulleistungsrelevanter Bedingungsfaktor, welcher sich auf gut gesicherte Erkenntnisse über Unterschiede im kognitiven Bereich bezieht. So wäre etwa die Leistungsüberlegenheit von Jungen im räumlichen Denken und die von Mädchen im sprachlichen Bereich zu nennen.

Im Fach Mathematik wurden in der Regel etwas bessere Leistungen der Jungen nachgewiesen, während Mädchen in einschlägigen Studien im Leseverständnis bessere Werte aufweisen (Zimmer, Burba & Rost, 2004; Hosenfeld, Helmke, Ridder & Schrader, 2002). Obwohl diese Unterschiede in der Regel stabil sind, können sie dessen ungeachtet als marginal eingestuft werden.

In Tabelle 3 sind die Geschlechterunterschiede in den jeweiligen Inhaltsbereichen dargestellt. Als Maß für die Bedeutsamkeit eines Unterschieds gilt die Effektstärke d^* , bei der die Unterschiede zwischen den Gruppen auf die Streuung der Testwerte

* Als Faustregel gelten in der experimentellen Forschung Werte für d um 0,2 als kleine, um 0,5 als mittlere und um 0,8 als große Effektstärken. Im Kontext nicht-experimenteller pädagogisch-psychologischer Forschung sind auch kleinere Effekte beachtenswert und interpretationswürdig (vgl. Ditton, 1990). Da allerdings die jeweilige Forschungslage zu berücksichtigen ist, dürfen die angegebenen Werte nicht dogmatisch als absolute Grundlage der Bewertung aufgefasst werden. Effektstärkemaße werden unter anderem deshalb verwendet, weil Aussagen über die Signifikanz eines Effekts u.a. von der Stichprobengröße abhängen (bei großen Stichproben werden schon sehr kleine Effekte statistisch signifikant). Die Effektstärke ist dagegen weitgehend unabhängig von der Stichprobengröße.

standardisiert werden. Ein positiver d-Wert bedeutet eine Überlegenheit der Mädchen, ein negativer d-Wert umgekehrt eine Überlegenheit der Jungen.

Tabelle 3: Verteilung der Fähigkeitsniveaus, getrennt nach Geschlecht; Angaben in Prozent

		n.a.L.*	FN1	FN2	FN3	N (Kinder)	d**
Arithmetik	Mädchen	1,7	19,4	40,4	38,5	10721	-0,1
	Junge	1,7	17,0	37,9	43,4	10736	
Geometrie	Mädchen	3,1	31,5	41,5	23,8	10721	-0,06
	Junge	3,6	28,2	41,9	26,3	10736	
Sachrechnen	Mädchen	2,6	50,8	30,2	16,3	10721	-0,28
	Junge	1,7	39,8	33,7	24,7	10736	
Lesen	Mädchen	9,0	26,7	37,0	27,3	10548	0,20
	Junge	12,3	31,5	35,3	20,8	10582	

* nicht auswertbare Leistung

** Maß für die Effektstärke

Bei Betrachtung der geringen Effektstärken wird ersichtlich, dass die Leistungsunterschiede zwischen den bei VERA teilnehmenden Schülerinnen und Schülern erwartungsgemäß eher gering sind.

In Mathematik zeigt sich ein bedeutsamer Vorsprung ($d = -0,28$) der Jungen im Bereich Sachrechnen, dort unterscheiden sich Mädchen und Jungen im höchsten Fähigkeitsniveau um einen Anteil von 8,4 Prozentpunkten. In Arithmetik zeigen sich immerhin noch geringe Unterschiede in den Leistungsverteilungen der Jungen und Mädchen ($d = -0,10$), während für die Geometrie die Geschlechterzugehörigkeit kaum noch eine Rolle spielt.

Im Leseverständnis ist ebenfalls ein bedeutsamer Effekt ($d = 0,20$) zu verzeichnen - hier schneiden die Mädchen besser ab.

3.1.5 Migrationshintergrund

Die Sprachbeherrschung hängt vermutlich stärker mit der vorherrschenden Familiensprache zusammen als mit dem Geburtsort des jeweiligen „nicht-deutschen“ Elternteils. Anhand der Unterscheidung in „Deutsch dominant“ vs. „Deutsch nicht-dominant“ ($N_{\text{minimal}} = 16906$ bzw. 4224) wird bei VERA der Sprachherkunft Rechnung getragen. Dabei entspricht „Deutsch nicht-dominant“ zweisprachigen Schülerinnen und Schülern, bei denen – unabhängig von Nationalität und Geburtsort – Deutsch nicht die hauptsächlich gesprochene Sprache ist (Helmke & Reich, 2001). Mit dieser Unterscheidung soll dem Sachverhalt Rechnung getragen werden, dass ein Teil der Schülerschaft zwar in Deutschland geboren ist, aber nicht in erster Linie Deutsch spricht bzw. nicht in Deutschland geboren ist, jedoch hauptsächlich Deutsch spricht.

In Abbildung 5 und Abbildung 6 sind die prozentualen Schülerleistungen jeweils nach Deutsch als dominante und nicht-dominante Sprache dargestellt. Tabelle 4 zeigt die Effektstärken der Leistungsunterschiede.

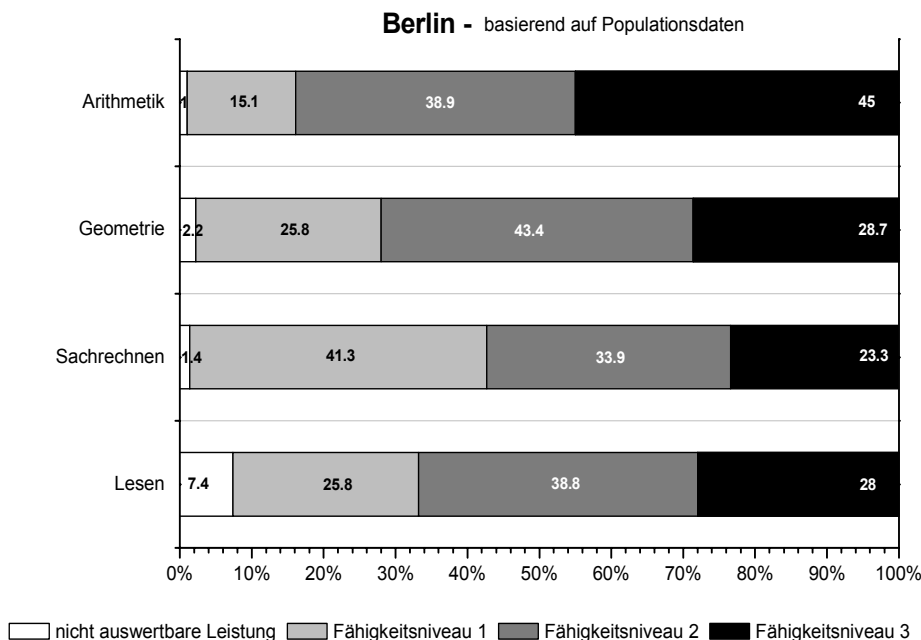


Abbildung 5: Gesamtverteilung der Fähigkeitsniveaus für Deutsch als dominante Sprache; Angaben in Prozent

Es zeigen sich bedeutsame Unterschiede zwischen Kindern mit Deutsch als dominanter und Kindern mit Deutsch als nicht-dominanter Sprache sowohl in Mathematik ($d = 0,57$ bis $0,65$) als auch im Leseverständnis ($d = 0,77$). So erreichen in Geometrie und Sachrechnen von den Schülerinnen und Schülern mit Deutsch als nicht-dominanter Sprache nur knapp zehn Prozent das höchste Fähigkeitsniveau, im Leseverständnis sind es unter neun Prozent. Dabei fällt vor allem der hohe Anteil von Kindern auf, die in Sachrechnen und Leseverständnis ein Niveau erreichen, das höchstens das Beherrschen elementarer Aufgaben umfasst (61,3 Prozent im Sachrechnen und 42,3 Prozent im Leseverständnis). Insbesondere beim Leseverständnis kann infolge des hohen Anteils von Kindern mit nicht auswertbarer Leistung (knapp 24 Prozent) von einem substanziellen Unterschied zu Kindern, deren dominante Sprache Deutsch ist, gesprochen werden.

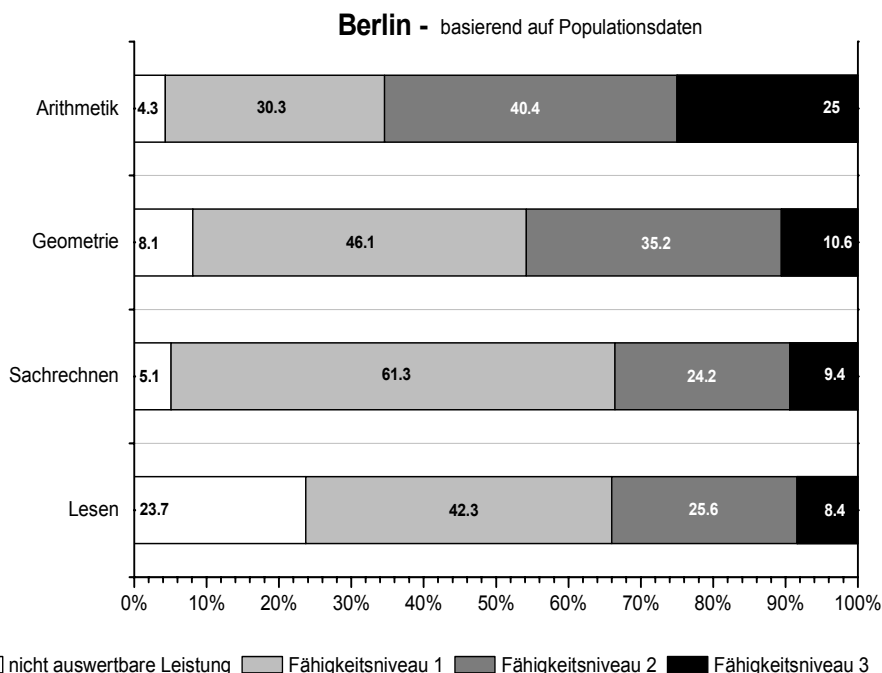


Abbildung 6: Gesamtverteilung der Fähigkeitsniveaus für Deutsch als nicht-dominante Sprache; Angaben in Prozent

Die Resultate bestätigen die Vermutung, dass Merkmale der Sprachherkunft für Schülerinnen und Schüler bereits in der Klassenstufe 4 mit erheblichen Leistungsunterschieden gekoppelt sind (Schwippert, Bos & Lankes, 2003).

Tabelle 4: Effektstärken der Leistungsunterschiede von Deutsch dominante vs. nicht-dominante Sprache

	Arithmetik	Geometrie	Sachrechnen	Lesen	N min (Kinder)
Effektstärke (d)	0,57	0,65	0,58	0,77	16906

3.2 „Fairer Vergleich“

Die großen Surveys der letzten Jahre, insbesondere PISA und IGLU/PIRLS, haben gezeigt, dass Merkmale des sozialen, ökonomischen und kulturellen Kapitals von Familien einen erheblichen Einfluss auf die Leistungsfähigkeit der Kinder ausüben. Auf der Ebene von Klassen und Schulen entspricht dies der wichtigen Rolle des Schuleinzugsgebietes und der Klassenzusammensetzung. Von Schulen „im sozialen Brennpunkt“ spricht man – obwohl es keine verbindlichen Definitionen gibt – gemeinhin dann, wenn verschiedene unterrichts- und lernerschwerende Faktoren in konzentrierter Form auftreten, etwa bei Schulen, deren Klientel durch stark überdurchschnittliche prozentuale Anteile mit Migrationshintergrund, geringer Bildungsnähe, soziale Unterschicht, Arbeitslosigkeit und Erhalt von Sozialhilfe gekennzeichnet ist.

Anders als Lernstandserhebungen und Forschungsprojekte vom Typ IGLU, MARKUS oder PISA wurden an dieser Stelle der VERA-Erhebung mit einem Lehrerfragebogen Angaben zur Klassenzusammensetzung und zum Einzugsgebiet der Schule in erster Linie zu dem Zweck erfasst, um für den „fairen Vergleich“ eine fundierte Datenbasis zu erzeugen. Alle Daten beruhen demnach auf Lehrerangaben und entsprechen nicht notwendigerweise den amtlichen Schulstatistiken. Aus diesem Grund sind die folgenden Ergebnisse keinesfalls als systematische Analyse kontextueller Bedingungen schulischer Leistungen zu verstehen.

3.2.1 Beschreibung ausgewählter Kontextmerkmale

Es werden zunächst die auf Klassenebene (Ausnahme: Sprachdominanz) erfragten Daten zum sozioökonomischen und sprachlichen Hintergrund berichtet (Tabelle 5). Anschließend folgt mit Tabelle 6 eine Darstellung der Zusammenhänge zwischen Kontextmerkmalen und Schülerleistungen. Tabelle 7 zeigt die Schülervariablen.

Tabelle 5: Kontextvariablen (Vergleich Berlin – Gesamtstichprobe)*; durchschnittliche Klassenanteile in Prozent

	Deutsch nicht- dominante Sprache ^{1,a)}	Schule liegt in sozialem Brennpunkt ²⁾	Familie gehört zur Grundschrift ^{2,b)}	Familie von Arbeitslosigkeit betroffen ²⁾	Familie bezieht Sozialhilfe ²⁾
Berlin	21,0	55,9	39,3	33,1	33,5
Gesamt	10,6	32,2	27,0	15,9	14,6

* Angaben beruhen auf Lehrerangaben und entsprechen nicht notwendigerweise den amtlichen Schulstatistiken

1) erfragt auf Individualebene in der Population

2) erfragt auf Klassenebene in der Zentralstichprobe, in Bremen in allen Klassen

a) Wortlaut: „Kinder, für die unabhängig von Nationalität und Geburtsort Deutsch die nicht-dominante Sprache darstellt.“

b) Wortlaut: „sog. Unterschicht oder Grundschrift: un- und angelernte Arbeiter, Landarbeiter, alle un- und angelernten Berufe aus dem manuellen Bereich sowie Dienstleistungstätigkeiten mit weitgehend manuellem Charakter und geringem Anforderungsniveau.“

Der Anteil von Schülerinnen und Schülern, für die unabhängig von Nationalität und Geburtsort Deutsch nicht die dominante Sprache darstellt, beträgt in Berlin 21 Prozent. Dabei ergibt sich eine Korrelation mit der Schülerleistung in Arithmetik von $r = -,42$ und im Leseverständnis von $r = -,48$. In diesem Zusammenhang ist auch die Abwesenheitsquote von knapp sieben Prozent in Deutsch erwähnenswert.

Über die Hälfte (55,9 Prozent) aller beteiligten Klassen liegen im so genannten „sozialen Brennpunkt“. Den Lehrerangaben zufolge, sind ein Drittel der Familien von Arbeitslosigkeit betroffen oder beziehen Sozialhilfe.

Tabelle 6: Zusammenhänge zwischen Kontextmerkmalen und Schülerleistungen
(Schulklassenebene)

	Arithmetik	Geometrie	Sachrechnen	Lesen	min. N
Prozentsatz Grundschicht	-0,58	-0,52	-0,64	-0,63	347
Prozentsatz Erhalt von Sozialhilfe	-0,581	-0,57	-0,69	-0,64	348
Prozentsatz Arbeitslosigkeit in der Familie	-0,53	-0,47	-0,63	-0,57	345
Prozentsatz deutsch nicht- dominant	-0,42	-0,41	-0,47	-0,48	985

Bei den Zusammenhängen zwischen Kontextmerkmalen und Schülerleistungen wurden in allen Inhaltsbereichen insgesamt mittlere Korrelationen gefunden (zwischen $r = -,41$ und $r = -,69$), wobei die höchsten in den Bereichen Sachrechnen und Leseverständnis auftreten.

Tabelle 7: Schülervariablen (Vergleich Berlin – Gesamtstichprobe)*; durchschnittliche Klassenanteile in Prozent

	ungenügende Sprachbe- herrschaft ^{1,a)}	nicht anwesend beim Deutshtest ¹⁾	nicht anwesend beim Mathetest ¹⁾	Klassen- wiederholer ¹⁾
Berlin	0,9	6,6	5,0	0,8
Gesamt	0,5	4,2	3,3	1,5

* Angaben beruhen auf Lehrerangaben und entsprechen nicht notwendigerweise den amtlichen Schulstatistiken

1) erfragt auf Individualebene in der Population

a) Wortlaut: „Kinder, die vor weniger als 12 Monaten nach Deutschland eingewandert sind und die deutsche Sprache noch nicht hinreichend beherrschen.“

3.2.2 Bildung der Kontextgruppen

Um den beteiligten Lehrkräften einen „fairen“ Vergleich anbieten zu können, indem Kontextunterschiede zwischen Klassen berücksichtigt werden, wurden drei landesspezifische Gruppen gebildet und mit Blick auf die Kontextmerkmale und die Leistungen beschrieben. Jede Lehrkraft hatte die Möglichkeit, auf der Grundlage einer Einschätzung der Zusammensetzung ihrer Klasse eine Zuordnung zu einer der drei definierten Kontextgruppen und somit einen „fairen“ Vergleichsmaßstab für die in der Klasse erzielten Leistungen zu erhalten. Die Bildung dieser Kontextgruppen erfolgte mit Hilfe eines regressionsanalytischen Mehrebenen-Ansatzes auf den Daten der Zentralstichprobe. Dieses Vorgehen stellt im Vergleich zum 2004 gewählten Ansatz eine Veränderung in drei Details dar: Berücksichtigung der hierarchischen Struktur der Daten, länderspezifische Analysen und Differenzierung in drei Kontextgruppen (je

Land), die so gewonnen Kontextgruppen lassen sich daher nicht auf die Daten des Jahres 2004 beziehen.

Konkret wurde zunächst ein Gesamtleistungswert ermittelt, in den die beiden Fächer Mathematik und Deutsch jeweils zu 50 Prozent eingingen, d.h. zunächst wurden die drei mathematischen Inhaltsbereiche gemittelt und dieser Wert wurde dann gleichgewichtig mit dem Leseverständnis (als einzigem verfügbaren Indikator im Fach Deutsch) zu einem Gesamtwert verrechnet. Im Anschluss daran wurden per Mehrebenenanalyse diejenigen Merkmale ermittelt, die mit diesem Gesamtleistungsindex im Zusammenhang stehen. In diese Analyse aufgenommen wurden auf Schülerebene die Merkmale Geschlecht, Deutsch als nicht-dominante Sprache, Wiederholung der vierten Klasse, Teilleistungsstörung Mathematik, Teilleistungsstörung Deutsch und kombinierte Teilleistungsstörungen in Mathematik und Deutsch. Auf der Ebene der Klassen wurden zwei verschiedene Blöcke von Informationen genutzt. Zum einen sind dies die auf Klassenebene aggregierten Schülermerkmale: die Geschlechterverteilung (Jungenanteil), der Anteil von Klassenwiederholern, der Anteil von Kindern mit Teilleistungsstörungen (getrennt für Mathematik und Deutsch), der Anteil von Kindern mit Deutsch als nicht-dominanter Sprache, der Anteil von Kindern mit sonderpädagogischem Förderbedarf⁴ und der Anteil von Kindern mit ungenügender Sprachbeherrschung⁵. Zum anderen sind dies die von den Lehrkräften im Lehrerfragebogen angegebenen Informationen zur Charakterisierung der Klasse: der Anteil der Grundschichtfamilien, der Anteil von Familien betroffen von Arbeitslosigkeit, der Anteil von Familien mit Bezug von Sozialleistungen.

In einem ersten Modell ohne Prädiktoren wurde die Verteilung der Varianz auf die beiden untersuchten Ebenen (Individuen; Klassen) ermittelt. Gut 61 Prozent der Leistungsvarianz liegt innerhalb der Klassen, etwa 39 Prozent sind mit der Klassenzugehörigkeit assoziiert. In der Folge wurde eine Serie von Analysen durchgeführt, um über schrittweise Erweiterungen des Modells die signifikanten Prädiktoren zu ermitteln.

Für Berlin erwiesen sich folgende Variablen als signifikante Prädiktoren: Deutsch als nicht-dominante Sprache, Teilleistungsstörung Mathematik, Teilleistungsstörung Deutsch, kombinierte Teilleistungsstörung Deutsch *und* Mathematik; auf Ebene der Klasse: Anteil der Grundschichtfamilien, Anteil der Familien mit Bezug von Sozialleistungen und der *Anteil* von Kindern mit Teilleistungsstörungen in Mathematik. Kinder mit Teilleistungsstörungen in Mathematik erzielen niedrigere Leistungen als Kinder ohne Teilleistungsstörungen. Gleiches gilt auch für Kinder mit Teilleistungsstörungen

⁴ Kinder mit sonderpädagogischem Förderbedarf nahmen nach Ermessen der unterrichtenden Lehrkraft an den Vergleichsarbeiten teil. Wenn entsprechende Kinder Aufgaben der Vergleichsarbeiten bearbeiteten, wurden auch Rückmeldungen für diese Kinder erstellt, ihre Leistungen gingen jedoch keinesfalls in die Leistungswerte der Klasse ein.

⁵ Unter die Definition „ungenügende Sprachbeherrschung“ fallen bei VERA nur solche Kinder, die weniger als sechs Monate in Deutschland zur Schule gehen und daher die deutsche Sprache noch nicht ausreichend beherrschen. Für die Rückmeldung und Wertung der Leistungen dieser Kinder gilt Analoges wie bei den Kindern mit sonderpädagogischem Förderbedarf.

in Deutsch. Kinder mit Teilleistungsstörungen in Mathematik *und* Deutsch sind ebenfalls leistungsschwächer als Kinder ohne Teilleistungsstörungen, jedoch nicht so leistungsschwach wie es sich aus der reinen Addition der Leistungsrückstände der beiden Formen von Teilleistungsstörungen ergeben würde. Ein ähnlicher Effekt ergibt sich auf der Ebene der Klasse. Der Einfluss des Klassenanteils mit Teilleistungsstörungen in Mathematik ist als Korrekturfaktor anzusehen: Klassen mit hohen Anteilen an Kindern mit Teilleistungsstörungen im Fach Mathematik schneiden *besser* ab als es auf Grund der reinen addierten Leistungsrückstände der einzelnen Schüler mit Teilleistungsstörung zu erwarten wäre.

Insgesamt können mit diesem Modell etwa acht Prozent der Leistungsvarianz auf Individualebene und gut 57 Prozent der Variation zwischen Klassen erklärt werden. Erwartungsgemäß fällt der Anteil der erklärten Varianz auf Ebene der Schülerinnen und Schüler gering aus, unerwartet hoch ist jedoch der erklärte Varianzanteil zwischen Klassen (in allen anderen Ländern liegt er unter 30 Prozent). Mit anderen Worten: In Berlin lassen sich die vorgefundenen Leistungsunterschiede besonders gut mit den zur Verfügung stehenden Informationen erklären. Insofern funktioniert hier der „faire Vergleich“ aus methodischer Sicht besonders gut.

Im nächsten Schritt wurden diese Informationen gemäß der ermittelten Regressionsgewichte zu einem Kontextwert kombiniert. Abbildung 7 zeigt die Verteilung dieser Werte der Berliner Zentralstichprobenklassen, für die vollständige Daten vorliegen und die daher berücksichtigt werden konnten (N = 162). Niedrige Werte stehen dabei für günstige, hohe Werte für ungünstige Kontextbedingungen.

Tabelle 8: Korrelation zwischen der mittleren Klassenleistung und dem Kontextwert in der Berliner Zentralstichprobe*

Mathematik	r	Deutsch	r
Arithmetik	-0,60	Leseverständnis	-0,71
Geometrie	-0,60		
Sachrechnen/Größen	-0,71		

* Stichprobengröße: N = 162 (Klassen mit vollständigen Daten)

Die Korrelationen dieses Kontextwertes mit den Leistungen in den verschiedenen Inhaltsbereichen sind in Tabelle 8 dargestellt. Erwartungsgemäß zeigen sich geringere Zusammenhänge bei den Leistungen im Bereich Arithmetik und Geometrie und engere Zusammenhänge mit den stärker sprachgebundenen Bereichen des Leseverständnisses und des Sachrechnens. Insgesamt fallen die Zusammenhänge vergleichsweise hoch aus, wie bereits der hohe Anteil aufgeklärter Varianz vermuten ließ.

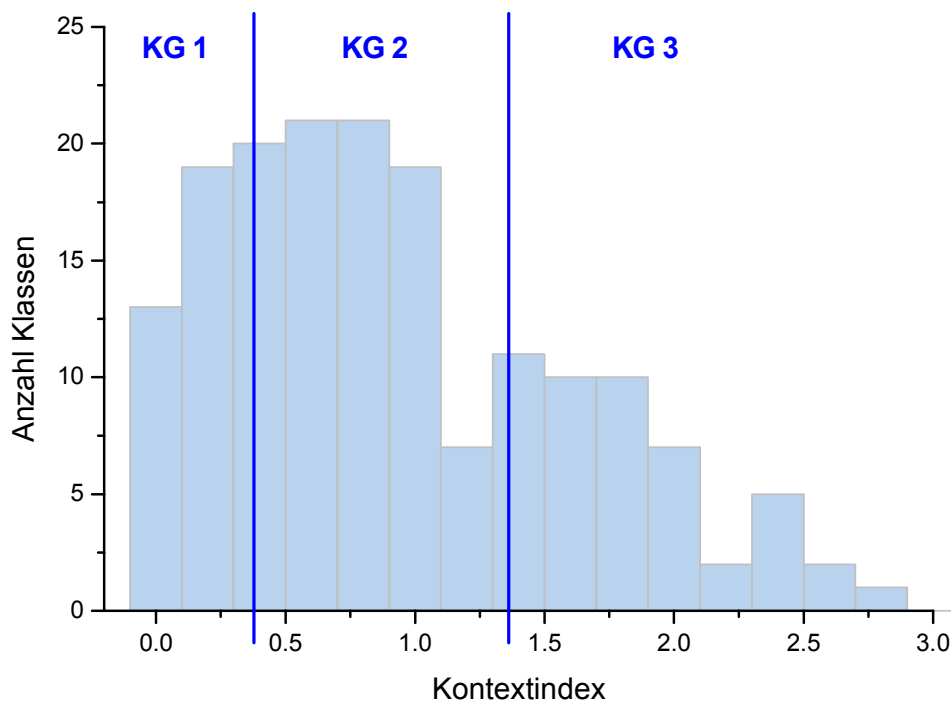


Abbildung 7: Verteilung der Kontextwerte inkl. der Grenzen zwischen den drei Kontextgruppen (KG 1 = günstige bis KG 3 = ungünstige Kontextbedingungen)

Anschließend wurden die Klassen in drei Kontextgruppen aufgeteilt, wobei die „günstigste“ Kontextgruppe 1 und die „ungünstigste“ Kontextgruppe 3 jeweils etwa 25 Prozent umfassen, während die mittlere Gruppe ca. 50 Prozent der Klassen umfasst.

Die Zuordnung aller „Nicht-Zentralstichproben-Klassen“ zu einer der drei Kontextgruppen erfolgt nach der Abfrage der relevanten Informationen im Internet. Durch die kompensatorische Art der Verknüpfung der Einzelinformationen können unterschiedliche Kontextkonfigurationen zur Zuweisung in die gleiche Kontextgruppe führen. Tabelle 9 gibt die Durchschnittswerte zentraler Kontextmerkmale der drei Gruppen wieder.

Tabelle 9: Beschreibung der drei Kontextgruppen anhand durchschnittlicher Merkmalsausprägungen; Angaben in Prozent

	Kontextgruppe		
	1	2	3
Anteil der Kinder mit Deutsch als nicht-dominanter Sprache	~2	~12	~50
Anteil der Kinder aus Familien der Grundschrift	~10	~38	~80
Anteil der Kinder aus Familien, die Sozialleistungen beziehen	~7	~32	~64

3.2.3 Verteilung der Fähigkeitsniveaus nach Kontextgruppen

Diese so gebildeten Kontextgruppen unterscheiden sich nicht nur mit Blick auf Merkmale der Klassenzusammensetzung, sondern auch hinsichtlich der erzielten Leistungen. In Tabelle 10 sind die Verteilungen der Zentralstichprobe auf die Fähigkeitsniveaus in Mathematik und Deutsch nach den drei Kontextgruppen differenziert dargestellt.

Tabelle 10: Verteilung der Fähigkeitsniveaus, aufgeschlüsselt nach Kontextgruppen; Angaben in Prozent

	Kontextgruppe	n.a.L.*	FN 1	FN 2	FN 3	N (Klassen)
Arithmetik	1	0,7	10,4	34,4	54,5	40
	2	0,7	19,2	39,6	40,5	82
	3	3,5	37,7	43,7	15,1	40
Geometrie	1	1,0	17,7	46,2	35,1	40
	2	2,7	27,0	43,2	27,0	82
	3	14,7	54,6	27,5	3,2	40
Sachrechnen	1	0,5	33,8	34,3	31,5	40
	2	2,3	45,3	34,2	18,2	82
	3	6,8	70,5	18,0	4,7	40
Lesen	1	5,7	18,4	43,1	32,8	40
	2	8,4	29,2	39,0	23,4	82
	3	29,2	45,4	20,4	5,0	40

* nicht auswertbare Leistung

Es wird deutlich, dass sich die Leistungen in den drei Kontextgruppen erwartungskonform unterscheiden. Der Anteil von Kindern mit fortgeschrittenen Fähigkeiten steigt mit zunehmender Günstigkeit des Kontextes stetig an, während der Anteil von Kindern mit nicht auswertbaren Leistungen nahezu stetig abnimmt. Angesichts der hohen Korrelationen zwischen dem gebildeten Kontextindex und den Schülerleistungen überrascht dies nicht.

3.2.4 Verteilung aller Schulen auf die Kontextgruppen

Wie oben dargestellt, erhielten alle beteiligten Klassen die Gelegenheit, sich nach der Eingabe weniger zusätzlicher Informationen automatisch einer der drei Kontextgruppen zuordnen zu lassen und so einen „fairen“ Vergleichsmaßstab zur Beurteilung der Leistungen der eigenen Klasse zu erhalten. Von diesem Angebot machten in Berlin 22,5 Prozent (N = 186) der Klassen Gebrauch⁶. Dabei ergaben sich hinsichtlich der Verteilung auf die drei Kontextgruppen nur geringe Abweichungen von der

⁶ Die oben bereits dargestellten 162 Klassen der Zentralstichprobe mit vollständigen Daten sind bei dieser Darstellung ausgeklammert.

Zentralstichprobe. Die mittlere Gruppe ist mit 43 Prozent der Klassen schwächer besetzt, während sowohl die günstige Gruppe 1 (29 Prozent) als auch die ungünstige Gruppe 3 (28 Prozent) geringfügig stärker vertreten sind.

Warum nur ein kleinerer Teil der Klassen dieses Angebot nutzte, lässt sich anhand der derzeit verfügbaren Daten nicht aufklären, es lässt sich jedoch festhalten, dass es sich hierbei nicht um ein Berlin-spezifisches Phänomen handelt. Lediglich in einem der anderen Bundesländer, in denen diese Option zur Verfügung stand, lag die Beteiligungsquote über 30 Prozent.

Ob sich diejenigen Klassen, die ihre Kontextgruppenzuordnung abgerufen haben, hinsichtlich der Leistung von denjenigen Klassen unterscheiden, welche dieses Angebot nicht genutzt haben, wurde anhand eines t-Tests geprüft. Es zeigt sich, dass kein statistisch bedeutsamer Unterschied in der gemittelten Leistung zwischen diesen beiden Gruppen existiert.

3.2.5 Vergleich tatsächliche vs. erwartete Leistung

Die folgende Gegenüberstellung der erwarteten Leistung und der tatsächlichen Leistung soll verdeutlichen, in welchem Ausmaß die gemittelte Leistung der Klassen über bzw. unter dem Erwartungswert liegt.

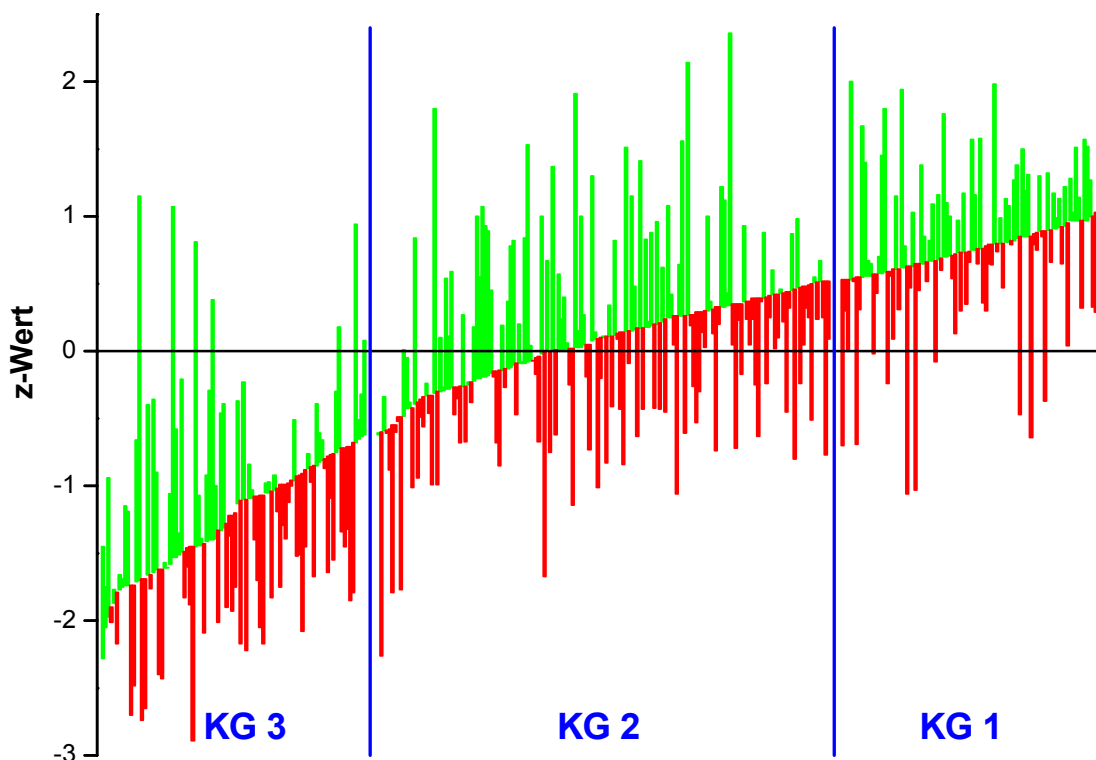


Abbildung 8: Differenzen zwischen erwarteter und tatsächlicher über beide Fächer gemittelter Leistung ($N = 348$ Klassen mit vollständigen Kontextdaten); KG = Kontextgruppe

Zu diesem Zweck wurden die Klassen, für die entsprechende Daten vorlagen ($N = 348$), nach dem Kontextindex sortiert, so dass die Klassen mit dem ungünstigsten Kontext ganz links und die Klassen mit dem günstigsten Kontext ganz rechts stehen. Entsprechend steigt der erwartete, gemittelte Leistungswert der Klassen von links nach rechts an. In der Abbildung dargestellt sind die Differenzen zwischen erwartetem und tatsächlichem Leistungswert. Jede Klasse wird durch einen senkrechten Balken dargestellt, je länger der Balken, desto größer die Differenz. Die Werte sind hierbei anhand der tatsächlichen Leistung aller 989 Berliner Klassen z-standardisiert. Liegt die tatsächliche Leistung über der erwarteten, so ist der Balken grün dargestellt und weist vom Erwartungswert aus nach oben, liegt die gezeigte Leistung unter dem Erwartungswert, so ist der Balken rot und geht nach unten. Die dicken blauen Linien kennzeichnen die Grenzen zwischen den drei Kontextgruppen (KG 1, KG 2, KG 3).

Auf den ersten Blick wird deutlich, dass es in allen Kontextbereichen deutliche Abweichungen vom Erwartungswert sowohl nach oben als auch nach unten gibt. Dabei liegen diese Abweichungen für etwa die Hälfte der Klassen in einer Zone von $-0,5$ bis $+0,5$ Standardabweichungen. Die schwächsten zehn Prozent unterschreiten ihren Erwartungswert jeweils um mindestens $0,90$ Standardabweichungen, die leistungsstärksten zehn Prozent erzielen gemittelte Leistungen, die jeweils mindestens $0,97$ Standardabweichungen über ihrem Erwartungswert liegen.

3.3 Diagnosegenauigkeit

Neben der Frage nach den Fähigkeitsniveaus und dem Leistungsstand der Viertklässler, aber aus pädagogischen Gründen gleichwohl bedeutsam, steht die Frage nach der Diagnosegenauigkeit von Lehrkräften – ein wichtiger Aspekt der diagnostischen Kompetenz. Wenn man die bei VERA vorgenommene Einschätzung der Aufgabenschwierigkeiten bei der Aufgabenwahl (in Mathematik) mit den tatsächlichen Aufgabenschwierigkeiten miteinander in Beziehung setzt, dann ergeben sich zwei interessante Kennwerte der Diagnosegenauigkeit: (a) das Ausmaß der *Unter- oder Überschätzung* der Aufgabenschwierigkeit, d.h. des Leistungsniveaus der Schulklasse und (b) die *Korrelation* zwischen beiden Rangreihen, also die Ähnlichkeit der Rangordnung geschätzter vs. realer Aufgabenlösungen.

Man muss diese beiden Aspekte der Diagnosegenauigkeit unbedingt unterscheiden. Eine Lehrerin kann z.B. alle Aufgaben konstant etwas über- oder unterschätzen, aber gleichwohl kann die Rangordnung der von ihr geschätzten Aufgabenschwierigkeit identisch mit der Rangordnung der gelösten Aufgabe sein. Eine hohe Korrelation der geschätzten mit der empirischen Aufgabenrangreihe sagt etwas über die *fachdidaktische* Kompetenz der Lehrkraft aus und ist am ehesten Ausdruck einer zutreffenden Orientiertheit über Schwierigkeitsunterschiede zwischen Aufgaben. Eine geringe Abweichung der durchschnittlichen geschätzten Aufgabenschwierigkeit von der empirischen Schwierigkeit sagt dagegen eher etwas über die *pädagogisch-psychologische* Diagnostik aus, d.h. wie gut die Lehrkraft über das durchschnittliche Leistungsniveau der Klasse im Bilde ist. Wir sprechen deshalb im Folgenden vereinfacht von der fachdi-

daktischen und der pädagogischen Komponente der Diagnosegenauigkeit. Die beiden Kennwerte sind nicht nur konzeptuell, sondern auch statistisch fast vollkommen unabhängig voneinander: Die Korrelation der beiden Kennwerte beträgt über alle Länder hinweg $r = ,08$. Die Frage der Diagnosegenauigkeit zu vertiefen, würde allerdings den Rahmen dieses Berichtes sprengen. Hierzu liegen nicht nur Handreichungen für die an VERA beteiligten Lehrkräfte, sondern auch Publikationen der VERA-Autoren vor (Helmke, Hosenfeld & Schrader, 2003, 2004).

a) Pädagogische Komponente: Unter- vs. Überschätzungstendenz

Wir berichten in der folgenden Abbildung das Ausmaß, in dem die Lehrkräfte die Leistungen ihrer Schülerinnen und Schüler unter- oder überschätzen. Hierzu liegen verwertbare Angaben von 327 Lehrkräften vor.

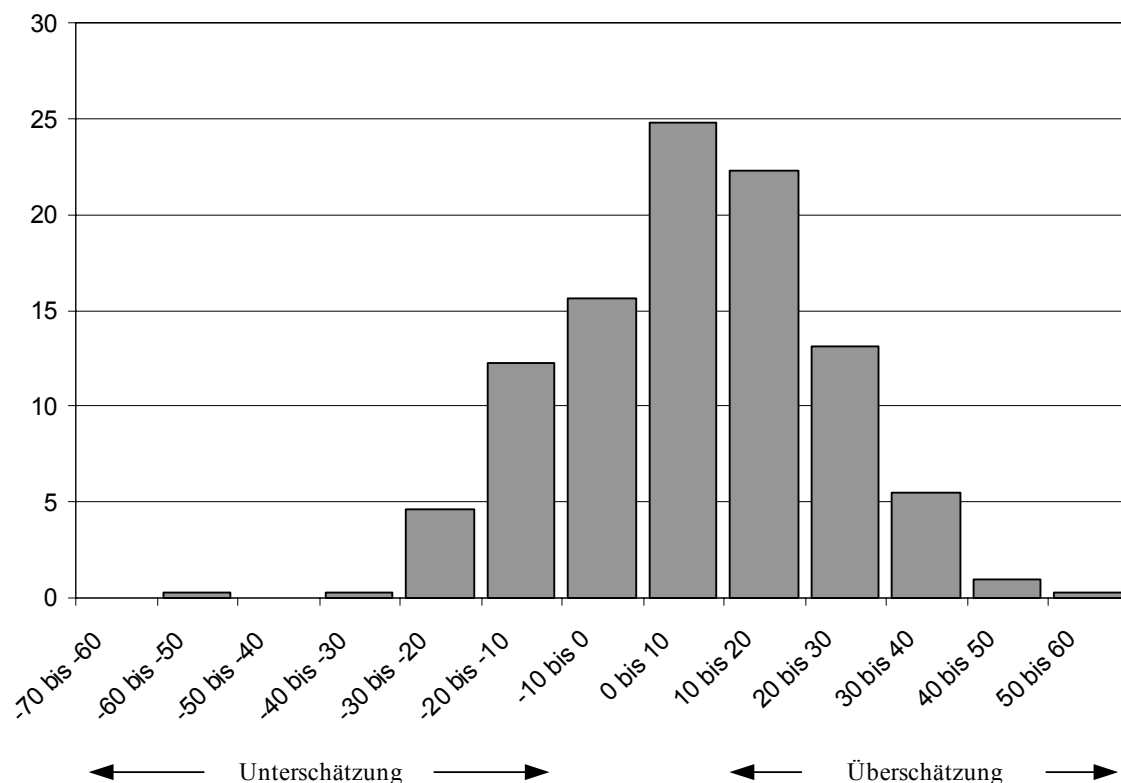


Abbildung 9: Prozentuale Unter- vs. Überschätzung des mathematischen Leistungsniveaus der Klasse (pädagogisch-psychologischer Aspekt der Diagnosegenauigkeit) durch die Lehrkräfte; Angaben in Prozent

Die Abbildung zeigt, dass die Lehrkräfte die Leistungen ihrer Klassen häufiger *überschätzen* als *unterschätzen*. Im Durchschnitt wird die tatsächliche Schülerleistung um 6,3 Prozent überschätzt.

Legt man das Ergebnis (die durchschnittliche Einschätzung) unter die Lupe und fragt, in welchem der drei Teilgebiete der Mathematik die geringsten und die größten Überschätzungen stattfinden, dann zeigt sich folgendes: Die Schülerleistungen im Bereich Sachrechnen/Größen werden um 1,0 Prozent *unterschätzt*, die in Arithmetik um 12,9 Prozent

und die im Bereich der Geometrie um 9,7 Prozent *überschätzt*. Die Lösungshäufigkeiten der Sachrechen-Aufgaben wurden damit 2005 realistischer eingeschätzt als jene der Arithmetik- und Geometrie-Aufgaben – eine Verkehrung gegenüber den Ergebnissen von 2004. Möglicherweise ist dieser Befund als Wirkung der Berichtlegung von 2004 zu interpretieren, in der auf die deutlichen Fehleinschätzungen im Bereich Sachrechnen aufmerksam gemacht wurde. An Plausibilität gewinnt dieser Schluss, wenn man die ausgeprägte Freiwilligkeit der Teilnahme in 2005 hinzunimmt – es dürfte sich bei den teilnehmenden Lehrkräften vor allem um die besonders interessierten und gut informierten gehandelt haben.

b) Fachdidaktische Komponente: Vergleich der Aufgaben-Rangordnungen

Im Land Berlin liegen verwertbare Angaben von 327 Lehrkräften zur Diagnosegenauigkeit vor. Die durchschnittliche Korrelation liegt bei $r = ,25$. Wenn man das Auflösungs-niveau erhöht, zeigt sich die folgende Verteilung.

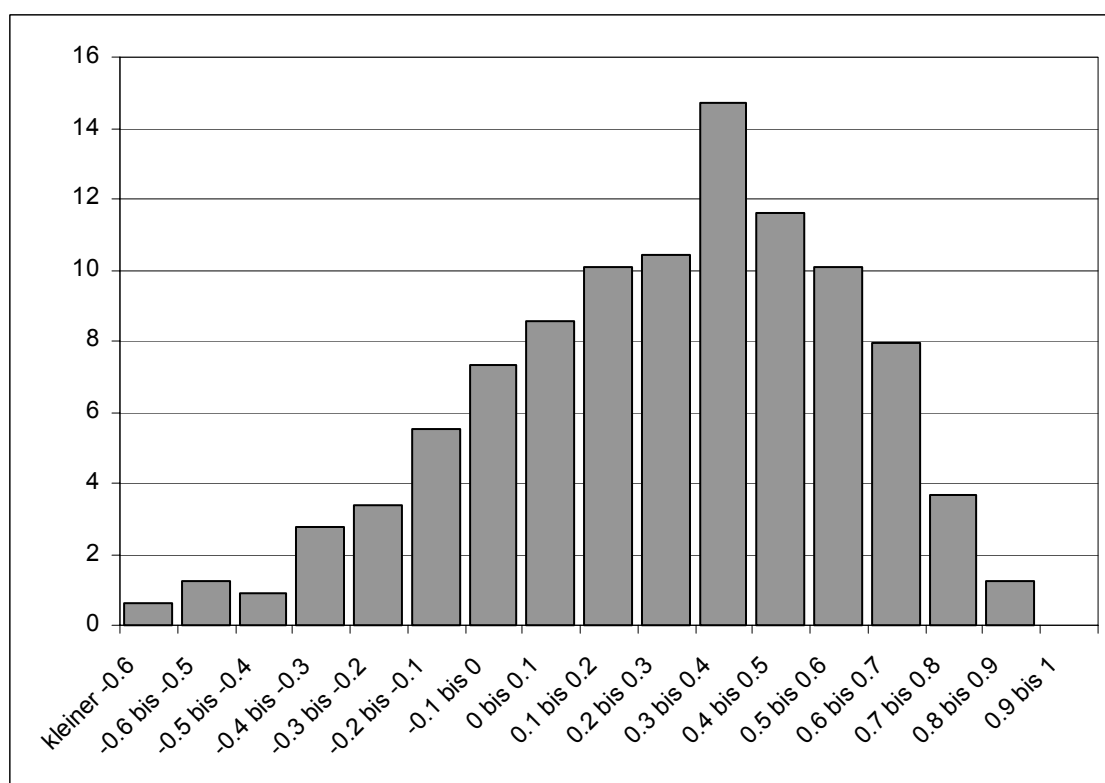


Abbildung 10: Verteilung der Korrelationskoeffizienten: Ähnlichkeit geschätzter und empirischer Aufgabenschwierigkeiten (fachdidaktischer Aspekt der Diagnosegenauigkeit); Angaben in Prozent

Es zeigt sich ein sehr großes Spektrum von Korrelationen um Null herum (dies ergibt sich, wenn man nach Zufall antwortet) und sogar einigen wenigen negativen Korrelationen, bis hin zu wenigen Klassen, deren Lehrkräfte eine sehr gute Diagnosegenauigkeit aufweisen (Korrelationen zwischen $r = ,80$ und $,90$). Korrelationen in Höhe von $r = ,50$ und höher würden wir als akzeptabel, Korrelationen in Höhe von $r = ,70$ und höher als gut bezeichnen, d.h. ca. fünf Prozent der Lehrkräfte verfügen über eine gute oder sehr gute Diagnosegenauigkeit, ca. 18 Prozent verfügen über eine akzeptable Diagnosege-

nauigkeit, aber ca. 77 Prozent der Lehrkräfte liegen mit ihrer Diagnosegenauigkeit z.T. deutlich unterhalb akzeptabler Werte.

Die 2005 gegebene Möglichkeit, auf Ebene der einzelnen Teilaufgaben Einschätzungen für die voraussichtliche Häufigkeit der richtigen Lösungen durch Schüler vorzunehmen, reduziert die Gefahr deutlicher Fehleinschätzungen, da auf der Ebene von Teilaufgaben ein feiner abgestimmtes Urteil möglich ist. Dies könnte als Ursache dafür angesehen werden, dass der Grad der Verschätzung (nach oben und nach unten) in diesem Jahr insgesamt geringer ausfällt. Andererseits wird es durch die größere Anzahl an (Teil-)Aufgaben schwieriger, die Häufigkeiten der richtigen Lösung in eine Rangreihe zu bringen. Dies dürfte zumindest zum Teil erklären, warum der Zusammenhang zwischen geschätzter und tatsächlicher Lösungshäufigkeit in diesem Jahr geringer ausfällt als 2004.

3.4 Lehrerfragebogen

Der Lehrerfragebogen sollte in Berlin für 187 Klassen (Zentralstichprobe) in 75 Schulen bearbeitet werden. An die Vollständigkeit des Rücklaufs des Lehrerfragebogens können unterschiedlich strenge Maßstäbe angelegt werden. An dieser Stelle sollen drei Kriterien berichtet werden:

- *Absolute Vollständigkeit:* Alle im Lehrerfragebogen erfragten numerischen Angaben wurden von der Lehrkraft gemacht. Lediglich die offenen Antwortfelder gehen nicht in die Wertung ein
- *Relative Vollständigkeit:* Fehlen höchstens zwei der numerischen Angaben, so liegen weitgehend vollständige Angaben vor, denn es werden hier beispielsweise Fälle mit aufgenommen, deren fehlende Werte aus versehentlichem Auslassen o. Ä. resultieren.
- *Vollständigkeit zentraler Angaben:* Die wichtigsten Angaben für die Kontextuierung einer Klasse sind solche zum
 - Anteil an Familien mit Grundschichtzugehörigkeit
 - Anteil an sozialhilfeempfangenden Familien
 - Anteil an Familien mit arbeitslosen Eltern(teilen)
 - Sozialen Brennpunkt.

In diesem Sinne besteht Vollständigkeit, wenn Angaben zu diesen vier Kriterien gemacht wurden.

Für die genannten Vollständigkeitsmaße resultieren die in Tabelle 11 dargestellten Werte.

Tabelle 11: Vollständigkeit des Rücklaufs des Lehrerfragebogens; Angaben in Prozent

	absolut vollständig	relativ vollständig	4 zentrale Angaben
komplett	73,8	82,9	84,0
unvollständig	26,2	17,1	16,0

Im Folgenden sind die Fähigkeitsniveauverteilungen in Bezug auf ausgewählte Variablen aus dem Lehrerfragebogen dargestellt. Zum einen werden bestimmte Lehrermerkmale (Kontinuität des Unterrichts, Unterrichtserfahrung, grundständiges Fach) diskutiert, die einen Einfluss auf die Schülerleistungen haben können. Zum anderen soll genauer beleuchtet werden, inwiefern VERA 2005 zumindest mittelfristig zu einer erhöhten Kooperation zwischen den beteiligten Lehrkräften geführt hat.

3.4.1 Kontinuität des Unterrichts in Mathematik und Deutsch und Fähigkeitsniveaus

Der Vergleich zwischen Klassen aus Schulen, in welchen *kein regulärer* Lehrerwechsel stattfindet, mit Klassen aus Schulen *mit regulärem* Lehrerwechsel nach der zweiten oder dritten Klasse, kann angesichts der teilweise geringen Gruppengrößen ($N < 10$) nicht zuverlässig interpretiert werden (s. Tabelle 12 und Tabelle 13).

Betrachtet man die Frage nach der Zahl der durch die einzelnen Lehrkräfte unterrichteten Halbjahre, so finden sich ebenfalls wieder teilweise geringe Gruppengrößen, weshalb auf eine weiterreichende Interpretation verzichtet werden sollte.

Tabelle 12: Lehrerwechsel während der ersten vier Grundschuljahre in den Fächern Mathematik und Deutsch; Angaben in Prozent

		n.a.L.*	FN 1	FN 2	FN 3	N (Klassen)
Arithmetik	keine Regel	2,2	21,3	40,0	36,6	58
	kein Lehrerwechsel	0,8	15,8	42,9	40,5	6
	Lehrerwechsel nach der 2. Klasse	0,0	13,1	32,0	54,9	10
	Lehrerwechsel nach der 3. Klasse	1,1	23,1	40,0	35,9	95
	eine andere Regel	2,3	16,5	43,5	37,7	18
Geometrie	keine Regel	5,1	33,2	39,4	22,2	58
	kein Lehrerwechsel	0,0	29,4	48,5	22,1	6
	Lehrerwechsel nach der 2. Klasse	1,9	16,7	45,9	35,5	10
	Lehrerwechsel nach der 3. Klasse	6,2	33,2	40,0	20,6	95
	eine andere Regel	3,4	34,4	37,5	24,7	18
Sachrechnen	keine Regel	4,1	50,1	29,3	16,5	58
	kein Lehrerwechsel	0,0	46,6	28,7	24,7	6
	Lehrerwechsel nach der 2. Klasse	0,0	29,6	36,7	33,7	10
	Lehrerwechsel nach der 3. Klasse	2,8	50,7	29,6	16,9	95
	eine andere Regel	2,2	50,2	30,4	17,2	18

		n.a.L.*	FN 1	FN 2	FN 3	N (Klassen)
Lesen	keine Regel	13,6	32,0	33,4	21,1	58
	kein Lehrerwechsel	11,0	27,9	32,5	28,6	6
	Lehrerwechsel nach der 2. Klasse	7,0	20,8	41,0	31,2	10
	Lehrerwechsel nach der 3. Klasse	13,8	31,4	36,4	18,4	95
	eine andere Regel	11,4	33,3	32,1	23,2	18

* nicht auswertbare Leistung

Table 13: Durch die Lehrkraft unterrichtete Halbjahre in den Fächern Mathematik und Deutsch, Angaben in Prozent

		n.a.L.*	FN 1	FN 2	FN 3	N (Klassen)
Arithmetik	Beginn 1. Klasse	1,4	18,8	36,7	43,2	33
	Mitte 1. Klasse	-	-	-	-	0
	Beginn 2. Klasse	0,0	17,5	34,7	47,9	6
	Mitte 2. Klasse	-	-	-	-	0
	Beginn 3. Klasse	0,5	14	32,9	52,5	24
	Mitte 3. Klasse	0,0	21,4	47,4	31,3	4
	Beginn 4. Klasse	1,9	23,5	42,4	32,2	112
Geometrie	Beginn 1. Klasse	3,4	26,4	39,1	31,1	33
	Mitte 1. Klasse	-	-	-	-	0
	Beginn 2. Klasse	2,3	22,7	43,9	31,7	6
	Mitte 2. Klasse	-	-	-	-	0
	Beginn 3. Klasse	1,7	18,1	43,2	37	24
	Mitte 3. Klasse	2,2	47,7	30,6	19,6	4
	Beginn 4. Klasse	6,7	37,2	39,2	16,9	112
Sachrechnen	Beginn 1. Klasse	2,7	45,9	29	22,3	33
	Mitte 1. Klasse	-	-	-	-	0
	Beginn 2. Klasse	3,4	34,5	40,2	22	6
	Mitte 2. Klasse	-	-	-	-	0
	Beginn 3. Klasse	0,8	35,1	36,5	27,6	24
	Mitte 3. Klasse	2,2	56	23,5	18,3	4
	Beginn 4. Klasse	3,3	54,1	28,4	14,2	112

		n.a.L.*	FN 1	FN 2	FN 3	N (Klassen)
Lesen	Beginn 1. Klasse	12,8	30,9	32,3	23,9	40
	Mitte 1. Klasse	-	-	-	-	0
	Beginn 2. Klasse	43,6	34,1	16,0	6,3	2
	Mitte 2. Klasse	26,3	42,1	31,6	0,0	1
	Beginn 3. Klasse	7,1	23,3	41,7	28,0	18
	Mitte 3. Klasse	4,7	29,5	39,2	26,6	3
	Beginn 4. Klasse	13,6	32,1	35,4	18,9	122

* nicht auswertbare Leistung

3.4.2 Unterrichtserfahrung, grundständige Ausbildung und Fähigkeitsniveaus

Aufgrund der verschiedenen in den Tabellen berücksichtigten Merkmalsebenen (z.B. Fach, Inhaltbereich und entsprechende Kategorien) und der vielfältigen damit zusammenhängenden Determinanten (z.B. bestimmte Kontextmerkmale wie Grundschichtzugehörigkeit) ist eine Interpretation komplex. Da sich darüber hinaus ähnlich wie bei MARKUS (Helmke et al., 2002) keine systematischen, insbesondere keine linearen (vom Typ „je...desto“) Unterschiede in Abhängigkeit von der Unterrichtserfahrung (Tabelle 14) sowie der grundständigen Ausbildung (Tabelle 15) ergeben, werden die entsprechenden Fähigkeitsniveaueverteilungen der Übersicht halber im Folgenden wiedergegeben, ohne sie im Detail zu diskutieren.

Tabelle 14: Durch die Lehrkraft unterrichtete Jahre in den Fächern Mathematik und Deutsch, Angaben in Prozent

		n.a.L.*	FN 1	FN 2	FN 3	N (Klassen)
Arithmetik	bis 2 Jahre	2,0	31,4	39,3	27,3	14
	2 bis 5 Jahre	3,4	27,1	41,5	28,0	5
	6 bis 10 Jahre	0,8	19,2	35,7	44,3	16
	11 bis 15 Jahre	1,9	19,3	41,8	37,0	26
	16 bis 20 Jahre	2,6	23,8	44,7	28,9	32
	21 bis 25 Jahre	1,1	21,8	42,9	34,3	27
	26 bis 30 Jahre	1,5	18,6	36,5	43,4	27
	31 bis 35 Jahre	0,5	18,0	35,6	45,9	21
	36 bis 40 Jahre	0,0	16,0	38,2	45,8	7
mehr als 40 Jahre	0,0	16,1	18,9	65,1	3	

		n.a.L.*	FN 1	FN 2	FN 3	N (Klassen)
Geometrie	bis 2 Jahre	8,3	34,1	46,5	11,0	14
	2 bis 5 Jahre	10,0	31,6	40,9	17,4	5
	6 bis 10 Jahre	4,1	30,6	32,4	32,9	16
	11 bis 15 Jahre	4,5	31,9	39,3	24,3	26
	16 bis 20 Jahre	6,5	37,6	41,5	14,4	32
	21 bis 25 Jahre	5,7	33,1	41,3	19,9	27
	26 bis 30 Jahre	4,3	29,5	40,8	25,4	27
	31 bis 35 Jahre	3,8	28,6	37,2	30,5	21
	36 bis 40 Jahre	3,2	36,9	35,3	24,6	7
	mehr als 40 Jahre	0,0	14,7	34,1	51,2	3
Sachrechnen	bis 2 Jahre	6,4	55,1	22,7	15,8	14
	2 bis 5 Jahre	0,8	52,3	26,0	20,9	5
	6 bis 10 Jahre	2,9	49,5	25,7	21,9	16
	11 bis 15 Jahre	3,6	47,2	30,7	18,5	26
	16 bis 20 Jahre	2,6	54,5	28,9	14,0	32
	21 bis 25 Jahre	2,7	47,5	33,3	16,5	27
	26 bis 30 Jahre	2,9	46,9	32,7	17,6	27
	31 bis 35 Jahre	1,9	48,4	29,4	20,4	21
	36 bis 40 Jahre	0,0	49,0	31,6	19,4	7
	mehr als 40 Jahre	1,5	24,7	31,5	42,3	3
Lesen	bis 2 Jahre	12,3	28,9	36,5	22,3	11
	2 bis 5 Jahre	14,7	32,6	22,0	30,8	6
	6 bis 10 Jahre	8,0	36,5	34,4	21,1	8
	11 bis 15 Jahre	10,1	29,5	41,0	19,4	32
	16 bis 20 Jahre	17,7	34,8	30,7	16,8	38
	21 bis 25 Jahre	14,3	32,3	34,3	19,2	19
	26 bis 30 Jahre	14,4	26,2	33,2	26,2	31
	31 bis 35 Jahre	10,8	30,3	41,2	17,7	29
	36 bis 40 Jahre	11,3	41,2	31,0	16,5	6
	mehr als 40 Jahre	3,1	13,1	36,8	47,0	4

* nicht auswertbare Leistung

Tabelle 15: Mathematik und Deutsch als grundständiges Fach, Angaben in Prozent

		n.a.L.*	FN 1	FN 2	FN 3	N (Klassen)
Arithmetik	Nein	1,8	19,8	39,8	38,6	101
	Ja	1,1	22,6	39,5	36,7	79
Geometrie	Nein	5,5	29,7	39,8	25,0	101
	Ja	4,8	35,3	40,0	19,9	79
Sachrechnen	Nein	2,7	48,1	29,6	19,6	101
	Ja	3,0	50,5	30,3	16,1	79
Lesen	Nein	12,9	30,5	36,1	20,5	89
	Ja	13,2	31,4	34,3	21,0	96

* nicht auswertbare Leistung

3.4.3 Vorbereitung auf die Vergleichsarbeiten

Mit Blick auf die im Projekt VERA formulierten Ziele beziehen sich gewünschte Konsequenzen im wesentlichen darauf, dass durch die zurückgemeldeten Ergebnisse eigene „blinde Flecke“ in Bezug auf die Klasse verdeutlicht und damit möglicherweise Maßnahmen zur Unterrichtsentwicklung angeregt werden. Dabei stellen die Vergleichsarbeiten als externe Evaluation sowohl für Lehrkräfte als auch für Schülerinnen und Schüler i.d.R. eine neuartige und aufregende Situation dar. Vor diesem Hintergrund ist es ein nachvollziehbares Bedürfnis, die eigene Klasse auf die Vergleichsarbeiten so vorzubereiten, dass das bisher Gelernte gezeigt werden kann und Ängste oder Belastungsreaktionen reduziert werden. In diesem Zusammenhang bieten sich unterschiedliche Zugangsweisen an wie z.B.

- *Vermittlung von Vertrautheit mit dem Testverfahren:* Vorbereitung auf den Ablauf, Besprechung von Aufgabenformaten und bei VERA eingesetzten Korrekturkriterien bzw. Anforderungen in den Korrekturanweisungen
- *Inhaltliche Vorbereitung:* Besprechung von VERA 2004-Aufgaben oder von typischen Inhalten der Vergleichsarbeiten
- *Vermittlung von Teststrategien*

Hier ist jedoch zu betonen, dass insbesondere mit der inhaltlichen Vorbereitung keinesfalls ein „teaching to the test“ von z.B. den herunter geladenen aktuellen Testaufgaben gemeint ist. Ein entsprechendes Vor-Üben der VERA-Aufgaben ist in keiner Hinsicht sinnvoll, insbesondere da die zurückgemeldeten Ergebnisse unter solchen Voraussetzungen allenfalls über die Reproduktionsleistung der Schülerinnen und Schüler Auskunft geben können.

Um die typischerweise eingesetzten Vorbereitungsformen zu erfassen, wurde eine entsprechende Frage in den Zentralstichprobenfragebogen aufgenommen. Die folgende

Tabelle 16 fasst die Häufigkeiten für die unterschiedlichen Formen der Vorbereitung zusammen.

Tabelle 16: Formen der Vorbereitung; Angaben in Prozent

		Häufigkeiten N = 180*	Prozent
Vertrautheit mit dem Testformat	Ablauf	144	80,0
	Aufgabenformate	40	22,2
	Korrekturkriterien	3	1,7
Inhaltliche Vorbereitung	Aufgaben 2004	35	19,4
	VERA-Inhalte	25	13,9
Teststrategien	Teststrategien	55	30,6

* Mehrfachantworten waren bei dieser Frage zugelassen.

Dabei scheint in Berlin der Fokus der Vorbereitung v.a. auf der Vertrautheit mit dem Testformat gelegen zu haben. So wurde am häufigsten angegeben, dass die Schülerinnen und Schüler auf den Ablauf der Vergleichsarbeiten vorbereitet wurden (80,0 Prozent). Dieses Ergebnis korrespondiert mit den Aufrufstatistiken für die Webseite mit den allgemeinen Materialien. Ungeklärt bleibt in diesem Zusammenhang, ob und aus welchem Grund in den anderen Fällen darauf verzichtet wurde. So ist es durchaus vorstellbar, dass die Lehrkräfte ihre Klassen auf diesem Wege das Entstehen von Angst vermeiden wollten. Ebenfalls häufig wurde angegeben, die Schülerinnen und Schüler in Bezug auf Teststrategien (30,6 Prozent) und Aufgabenformate bzw. VERA 2004-Aufgaben (etwa 20 Prozent) vorbereitet zu haben. Besonders selten berichteten die Lehrkräfte dagegen die Vorbereitung auf VERA-Korrekturkriterien: Dieses Vorgehen wurde nur in drei Fällen berichtet. Alles in allem zeigen die Ergebnisse jedoch deutlich, dass bei den Lehrkräften das Bedürfnis besteht, ihre Schülerinnen und Schüler auf die Vergleichsarbeiten vorzubereiten.

3.4.4 Kooperation bei der Aufgabenauswahl und -auswertung

Neben Informationen zum sozialen Kontext wurde im Lehrerfragebogen auch erfragt, inwiefern die Auswahl der Wahlaufgaben in Mathematik sowie die Auswertung der Vergleichsarbeiten in Kooperation mit Kolleg/innen erfolgte. Da die Kooperation (z.B. zur Besprechung der VERA-Ergebnisse) für die Tiefe der Verarbeitung wichtig ist und gerade die Auseinandersetzung mit den Inhalten und Konzepten der Vergleichsarbeiten den Blick auf Ansatzpunkte für Unterrichtsentwicklung eröffnet, wurde eine Kooperation durch das Prozedere der Aufgabenauswahl (pro Schule nur eine Aufgabekombination erlaubt) und der Vergleichsangebote in den Ergebnisrückmeldungen forciert. Die Antworten der Lehrkräfte zu diesen Fragen kann als Hinweis zu den schulischen Bedingungen, insbesondere dem Evaluations- und Kooperationsklima interpretiert werden. Tabelle 17 belegt, dass bei der *Aufgabenauswahl* in den meisten Fällen kooperiert wurde. Demgegenüber erfolgte die *schulinterne Auswertung* deutlich

seltener in Zusammenarbeit mit Kolleg/innen. Die von VERA intendierte Kooperation im Zusammenhang mit den Vergleichsarbeiten ist hier noch nicht überzeugend gelungen. Ein Grund könnte sein, dass die Lehrkräfte sich für die zeitintensivere Auswertung außerhalb des Unterrichts verabreden mussten. Dies lässt sich mit dem Arbeitsalltag von Lehrkräften u.U. schwer in Einklang bringen. Eine andere Erklärung wäre, dass immer noch sehr viel Angst davor besteht, die Klassenergebnisse und damit den eigenen Unterricht vor anderen offen zu legen. Bei der Auswertung in Mathematik ergibt sich erfreulicherweise eine Steigerung gegenüber dem Vorjahr von 15,9 Prozent auf 31,6 Prozent.

Tabelle 17: Kooperation im Rahmen der Vergleichsarbeiten; Angaben in Prozent

		ja	nein	N (Klassen)
Aufgabenauswahl im Team	Mathematik	79,2	20,8	187
Auswertung im Team	Deutsch	23,6	76,4	187
	Mathematik	31,6	68,4	187

4 Ausblick

Das Projekt VERA hat in vielfacher Hinsicht Neuland betreten, insbesondere mit dem auf Kommunikation und Kooperation zielenden Prinzip der schulinternen Auswahl von Aufgaben und mit der konsequenten Nutzung des Internet, die im letzten Jahr mit dem Ziel einer Verringerung der Arbeit für die beteiligten Lehrkräfte noch weiter ausgebaut und zugleich vereinfacht wurde. Die Vergleichsarbeiten der kommenden Jahren werden sich konsequent an den bundeslandübergreifenden Bildungsstandards orientieren.

An dieser Stelle möchten wir nochmals an die in Kapitel 1.1 skizzierten bildungspolitischen Ziele von Vergleichsarbeiten erinnern. Vergleichsarbeiten sind kein Selbstzweck, sondern ein Werkzeug zur Bestandsaufnahme, Sicherung und Verbesserung der Bildungsqualität. Der eigentliche Wert von VERA als einem bundeslandübergreifenden und flächendeckend angelegten Unternehmen, das im jährlichen Zyklus angelegt ist, wird sich danach bemessen, ob VERA zu einer Verbesserung insbesondere des Unterrichts (Nutzung pädagogischer und fachdidaktischer Impulse der Ergebnismeldungen, der Handreichungen und kommentierten Aufgabenbeispiele), der Schulentwicklung (Anstöße zur Verbesserung der Evaluations- und Kooperationskultur) sowie der Professionalisierung der Lehrerschaft im Bereich der pädagogischen Diagnostik (Erfassung und Verbesserung von Aspekten der Diagnosekompetenz) beiträgt. Für *System Monitoring* eignen sich Studien, die – wie TIMSS, PISA, IGLU oder DESI – anders angelegt sind als VERA.

Seit der „empirischen Wende“ ist der Grundsatz der Wirkungsorientierung im Bereich von Schule und Unterricht unbestritten: Das Ausmaß der Erreichung (oder Verfehlung) bildungspolitischer Ziele muss nachgewiesen, also empirisch überprüft werden. Einen wichtigen Schritt in diese Richtung haben wir in Gestalt internetbasierter Lehrerbefragungen zur Rezeption und zum Nutzen von VERA getan. Die Ergebnisse dieser

Befragungen, deren Ergebnisse nicht Teil dieses Berichtes sind, zeigen, welches Potenzial die Vergleichsarbeiten für die Reflektion und Veränderung des eigenen Unterrichts, die Professionalisierung und die Intensivierung der schulinternen Kooperation besitzen. Wenn dazu als flankierendes Maßnahmenbündel ein entfaltetes Unterstützungssystem kommt, dann wären dies günstige Bedingungen dafür, dass Evaluation nicht bei der Standortbestimmung stehen bleibt, sondern für Innovation genutzt wird. Ein zweiter Schritt in Richtung "Unterrichtsentwicklung" besteht in einer an VERA angedockten Studie des Grundschulunterrichts in den Fächern Deutsch und Mathematik, die im laufenden Schuljahr in den beiden Bundesländern Bremen und Rheinland-Pfalz durchgeführt wird. Diese als Zweipunktmessung angelegte Unterrichtsstudie, deren Kern eine Videographie des Unterrichts ist, nutzt die VERA-Erhebung vom Herbst 2005 als "base line" und führt am Ende des Schuljahres eine weitere Lernstandserhebung in Mathematik und Deutsch (Leseverständnis) durch. Die Nutzung der so gewonnenen Daten über unterrichtliche Bedingungen der Kompetenzentwicklung im Verlaufe der vierten Klassenstufe ist ein weiterer Mosaikstein, um die Vergleichsarbeiten stärker für die Verbesserung des Lehrens und Lernens zu nutzen.

5 Anhang: Lehrerfragebogen



UNIVERSITÄT
KOBLENZ · LANDAU

VERA

VERgleichsArbeiten
in 4. Grundschulklassen

Aktuell "Klassenbuch" Aufgaben Dateneingabe Ergebnisse Materialien Hilfe Kontakt

Klassenbuch
-> Lehrerfragebogen

Info
Zentral-
stichprobe

Angemeldet: B
DEMO004

[Seitenstruktur](#)



Abmelden

Information: Zentralstichprobe

Liebe Grundschule!

Sie sind Teil der Zentralstichprobe.
Sollten Sie noch nicht von Ihrem Ministerium Ihrer Senatsverwaltung benachrichtigt worden sein, werden Sie in den nächsten Tagen ein entsprechendes Schreiben erhalten.

Folgende Punkte sollten Sie bedenken:

1. Die Auswertung der Vergleichsarbeiten muss bis zum 4.11.2005 abgeschlossen, und die Daten müssen auf den VERA-Server hochgeladen sein.
2. Bitte beantworten Sie einen Fragebogen zu schulleistungsrelevanten Bedingungen Ihrer Klasse und Schule. Dies ist nötig, um die Ergebnisse der jährlich wechselnden, zufällig gezogenen Zentralstichprobe differenziert auswerten zu können. Die Zentralstichprobe dient der Analyse von Trends auf Landesebene und zur Ermittlung der Beziehung zwischen Fähigkeiten und sozialem Kontext (Einzugsgebiet, Klassenzusammensetzung).
Rankings von Schulen oder Klassen sind ausgeschlossen.

Für Ihre Unterstützung möchten wir uns im Voraus schon herzlich bedanken

Das VERA-Team

Wenn Sie auf "Weiter" klicken, gelangen Sie direkt zum Fragebogen für Lehrerinnen und Lehrer!



[Weiter](#)

1. Schule, Einzugsgebiet und Klassenzusammensetzung

1) Befindet sich Ihre Schule in einem Einzugsgebiet, das als sozialer Brennpunkt bezeichnet werden kann?

- nein ja

2) Wird in ihrer Schule eine Fremdsprache unterrichtet?

- nein ja, Englisch ja, Französisch Ja,

(sonstige: bitte eintragen)

Wichtige Anmerkung zu den folgenden Fragen:

Um bei VERA eine aussagekräftige schulübergreifende Einordnung der Schülerleistungen zu ermöglichen, werden die Ergebnisse differenziert nach sog. Kontextgruppen zurückgemeldet ("fairer Vergleich"). Dabei werden Unterschiede in den Standortvoraussetzungen der Schulen und der Zusammensetzung der Schülerschaft in den Klassen berücksichtigt. Wir bitten Sie zu diesem Zweck, die soziale Zusammensetzung Ihrer Klasse einzuschätzen. Diese Merkmale werden aus Datenschutzgründen in den Schulen nicht erfasst, damit die Schülerbeurteilung nicht von der sozialen Herkunft beeinflusst wird. Die Erfahrungen beim letzten Durchgang von VERA haben gezeigt, dass die Lehrkräfte dennoch ein weitgehend realistisches und treffsicheres Bild über die soziale Zusammensetzung ihrer Klassen haben und auf dieser Grundlage eine sinnvolle Zuordnung zu Kontextgruppen möglich ist. Zusätzliche Befragungen der Schülerinnen und Schüler zu ihrem sozialen Hintergrund sind auf diesem Hintergrund weder notwendig noch datenschutzrechtlich zulässig.

3) Bitte schätzen Sie die soziale Zusammensetzung Ihrer Klasse ein!

Wie viele Schülerinnen und Schüler gehören Familien an, die der sog. Unterschicht oder "Grundschrift" zuzuordnen sind?

(un- und angelernte Arbeiter, Landarbeiter, alle un- und angelernten Berufe aus dem manuellen Bereich sowie Dienstleistungstätigkeiten mit weitgehend manuellem Charakter und geringem Anforderungsniveau).

Schülerinnen und Schüler

4) Von Arbeitslosigkeit betroffene Familien

Für wie viele Schülerinnen und Schüler Ihrer Klasse gilt, dass ihre Familien (in Bezug auf das Haupteinkommen der Familie) von Arbeitslosigkeit betroffen sind?

Schülerinnen und Schüler

5) Bezug von Sozialleistungen

In den Haushalten wie vieler Schülerinnen und Schüler Ihrer Klasse wird der Lebensunterhalt überwiegend durch den Bezug von Sozialleistungen bestritten?

Schülerinnen und Schüler

6) Lehrkräfte Mathematik und Deutsch

Wie viele Lehrkräfte unterrichten diese Klasse in Mathematik und Deutsch?

- eine Lehrkraft unterrichtet beide Fächer
- Mathematik und Deutsch werden von unterschiedlichen Lehrkräften unterrichtet.
- Diese Klasse wird in den Fächern Deutsch und Mathematik von mehr als zwei Lehrkräften unterrichtet,

7) Gibt es in Ihrer Schule eine allgemeine Regel oder Gewohnheit hinsichtlich des Lehrerwechsels während der ersten vier Grundschuljahre?

- gar keine Regel
- kein Lehrerwechsel
- Lehrerwechsel nach der 2. Klasse
- Lehrerwechsel nach der 3. Klasse
- eine andere Regel

2. Fragen an die Deutsch-Lehrkraft

8) Seit wann unterrichten Sie diese Klasse in Deutsch? Seit...

- Beginn 1. Klasse Beginn 2. Klasse Beginn 3. Klasse Beginn 4. Klasse
 Mitte 1. Klasse Mitte 2. Klasse Mitte 3. Klasse

Wenn Sie die Klasse im Verlauf eines Halbjahres übernommen haben, wählen Sie die am ehesten zutreffende Antwort!

In der wissenschaftlichen Öffentlichkeit wird diskutiert, welche Formen der Vorbereitung die Ergebnisse der Schülerinnen und Schüler in Schulleistungstests beeinflussen. Wir bitten Sie daher, die folgende Frage zu beantworten.

9) Wie haben Sie Ihre Schülerinnen und Schüler auf die Vergleichsarbeiten vorbereitet?

- gar nicht
 Erläutern des Ablaufs der Vergleichsarbeiten anhand der Schülerinstruktionen
 Üben der veröffentlichten Aufgaben des VERA-Durchgangs 2004
 Üben von Aufgabenformaten, die typischerweise in Tests eingesetzt werden (z.B. Multiple-Choice-Format, etc.)
 Besprechen allgemeiner Testbearbeitungsstrategien
 Integration von VERA-Inhalten in den Unterricht
 Verwenden von an VERA angelehnten Korrekturkriterien in Klassenarbeiten

10) Wie leistungsstark ist Ihre Klasse im Leseverständnis?

Wenn Sie von der (geschätzten) Gesamt-Leistungsverteilung für das Leseverständnis in **Ihrem Bundesland** ausgehen, wie leistungsstark schätzen Sie Ihre Klasse im schulübergreifenden Durchschnitt ein?

Bitte kreuzen Sie auf der folgenden Skala, die vom untersten Fünftel (leistungsschwach) bis zum oberen Fünftel (leistungsstark) reicht, das zutreffende Feld an!

	leistungsschwach			leistungsstark	
	0-20	21-40	41-60	61-80	81-100
Leseverständnis	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

3. Fragen zu Ihnen selbst

11) Seit wie vielen Jahren unterrichten Sie Deutsch? Seit ...

Jahren

12) Haben Sie Deutsch als grundständiges Fach studiert?

nein ja

13) Haben Sie neben dem Grundschulstudium eine weitere Ausbildung mit Bezug auf den Deutschunterricht in der Grundschule?

14) Sind Sie an Ihrer Schule Fachkonferenzvorsitzende/r oder Fachleiter/in für das Fach Deutsch?

Fachkonferenzvorsitzende/r Fachleiter/in weder noch

4. Fragen an die Mathematik-Lehrkraft

15) Seit wann unterrichten Sie diese Klasse in Mathematik? Seit...

- Beginn 1. Klasse Beginn 2. Klasse Beginn 3. Klasse Beginn 4. Klasse
 Mitte 1. Klasse Mitte 2. Klasse Mitte 3. Klasse

Wenn Sie die Klasse im Verlauf eines Halbjahres übernommen haben, wählen Sie die am ehesten zutreffende Antwort!

In der wissenschaftlichen Öffentlichkeit wird diskutiert, welche Formen der Vorbereitung die Ergebnisse der Schülerinnen und Schüler in Schulleistungstests beeinflussen. Wir bitten Sie daher, die folgende Frage zu beantworten.

16) Wie haben Sie Ihre Schülerinnen und Schüler auf die Vergleichsarbeiten vorbereitet?

- gar nicht
 Erläutern des Ablaufs der Vergleichsarbeiten anhand der Schülerinstruktionen
 Üben der veröffentlichten Aufgaben des VERA-Durchgangs 2004
 Üben von Aufgabenformaten, die typischerweise in Tests eingesetzt werden (z.B. Multiple-Choice-Format, etc.)
 Besprechen allgemeiner Testbearbeitungsstrategien
 Integration von VERA-Inhalten in den Unterricht
 Verwenden von an VERA angelehnten Korrekturkriterien in Klassenarbeiten

17) Wie leistungsstark ist Ihre Klasse in Mathematik?

Wenn Sie von der (geschätzten) Gesamt-Leistungsverteilung für die drei Inhaltsbereiche Arithmetik, Geometrie und Sachrechnen in **Ihrem Bundesland** ausgehen, wie leistungsstark schätzen Sie Ihre Klasse im schulübergreifenden Durchschnitt ein?

Bitte kreuzen Sie auf der folgenden Skala, die vom untersten Fünftel (leistungsschwach) bis zum oberen Fünftel (leistungsstark) reicht, das zutreffende Feld an!

	leistungsschwach			leistungsstark	
	0-20	21-40	41-60	61-80	81-100
Arithmetik	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Geometrie	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sachrechnen	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

5. Fragen zu Ihnen selbst

18) Seit wie vielen Jahren unterrichten Sie Mathematik? Seit ...

Jahren

19) Haben Sie Mathematik als grundständiges Fach studiert?

nein ja

20) Haben Sie neben dem Grundschulstudium eine weitere Ausbildung mit Bezug auf den Mathematikunterricht in der Grundschule?

21) Sind Sie an Ihrer Schule Fachkonferenzvorsitzende/r oder Fachleiter/in für das Fach Mathematik?

Fachkonferenzvorsitzende/r Fachleiter/in weder noch

6. Durchführung der Vergleichsarbeiten

22) Haben Sie die Mathematik-Aufgabenauswahl im Team mit Kolleginnen / Kollegen durchgeführt?

- nein ja

23) Wie viel Zeit haben Sie in etwa für die Korrektur der Arbeiten benötigt (inkl. Dateneingabe)?

a) Für **Deutsch** insgesamt

	Stunden
	Stunden

b) Für **Mathematik** insgesamt

24) Haben Sie im Team ausgewertet?

- a) **Deutsch** nein ja
b) **Mathematik** nein ja

Herzlichen Dank für Ihre Mitarbeit!

6 Glossar

aggregieren, Aggregation, aggregierte Ebene, aggregierte Effekte, das Aggregieren bezeichnet einen datentechnischen Vorgang, bei dem mehrere Fälle einer Gruppe zu einem neuen Fall zusammengefasst („aggregiert“) werden. Beispielsweise lassen sich in der vorliegenden Untersuchung die Daten von allen Schülerinnen und Schülern einer Klasse zu \rightarrow arithmetischen Mitteln auf Klassenebene aggregieren. Neben der Aggregation von der Individualebene (Angaben einzelner Schüler) auf die Klassenebene sind auch Aggregationen auf die Ebene der Schule oder der Schulart denkbar.

arithmetisches Mittel, arithmetischer Mittelwert, Durchschnittswert

\rightarrow Mittelwert.

DESI, Deutsch Englisch Schülerleistungen International. Projekt der Kultusministerkonferenz, in dem es - als Komplement zu \rightarrow PISA - um die aktive Beherrschung der Muttersprache und des Englischen als Fremdsprache geht. Das Projekt DESI wird von einem Konsortium unter der Leitung des DIPF (Deutsches Institut für Internationale Pädagogische Forschung) Frankfurt durchgeführt.

Effektstärke, Maß für die Größe bzw. die praktische Bedeutsamkeit eines Effekts (d.h. eines Unterschieds zwischen Mittelwerten, Streuungen, Korrelationen usw.). Es gibt verschiedene Effektstärkemaße: Beim Vergleich der Mittelwerte zweier Gruppen (\rightarrow t-Test) kann das Maß d verwendet werden: Größe des Unterschieds zwischen beiden Gruppenmittelwerten, dividiert durch die gemittelte Streuung. Als Faustregel gelten in der experimentellen Forschung Werte für d um 0,2 als kleine, um 0,5 als mittlere und um 0,8 als große Effektstärken. Im Kontext nicht-experimenteller pädagogisch-psychologischer Forschung sind auch kleinere Effekte beachtenswert und interpretationswürdig. Da allerdings die jeweilige Forschungslage zu berücksichtigen ist, dürfen die angegebenen Werte nicht dogmatisch als absolute Grundlage der Bewertung aufgefasst werden. Effektstärkemaße werden unter anderem deshalb verwendet, weil Aussagen über die Signifikanz eines Effekts u.a. von der Stichprobengröße abhängen (bei großen Stichproben werden schon sehr kleine Effekte statistisch signifikant). Die Effektstärke ist dagegen weitgehend unabhängig von der Stichprobengröße.

erklärte Varianz, aufgeklärte Varianz, die erklärte Varianz ist derjenige prozentuale Anteil der \rightarrow Varianz der Werte einer Variablen x , der aufgrund der Werte einer anderen Variable y erklärbar ist. Bei einer Korrelationsrechnung wird die erklärte Varianz durch das Quadrat des \rightarrow Korrelationskoeffizienten bestimmt.

IEA, International Association for the Evaluation of Educational Achievement. Diese Organisation hat die weltweit meisten internationalen Vergleichsstudien, darunter →TIMSS, durchgeführt.

IGLU, Internationale Grundschul-Lese-Untersuchung. Deutsche Teilstudie der Studie PIRLS (Progress in International Reading Literacy Study) der →IEA, ergänzt um Mathematik und naturwissenschaftliche Teilkomponenten (IGLU-E). Die Hauptuntersuchung fand 2001 statt, die ersten Ergebnisse werden 2003 publiziert. Alle 16 Bundesländer haben sich an IGLU beteiligt, 13 an IGLU-E.

Intervallskala, intervallskalierte Variable, Skala, bei der gleich große Unterschiede zwischen den Skalenwerten gleich große Merkmalsunterschiede anzeigen (z.B. ist der Temperaturunterschied zwischen den Skalenwerten 16° und 18° genau so groß wie der zwischen 21° und 23°); bei einer Verhältnisskala (z.B. Länge, Gewicht) ist darüber hinaus der Nullpunkt eindeutig festgelegt.

Item, Bezeichnung für die Aufgaben eines Tests oder die Fragen/Feststellungen eines Fragebogens. Die Items werden häufig zu einer →Skala zusammengefasst.

Koeffizient, ein Koeffizient ist ein statistischer bzw. mathematischer Kennwert. Pearsons r ist z.B. ein →Korrelationskoeffizient, d. h. ein statistisches Zusammenhangsmaß.

Korrelation, korrelieren, Korrelationskoeffizient, eine Korrelation ist ein statistisches Maß für den Grad des linearen Zusammenhangs zwischen zwei →Variablen (Merkmalen) x und y . Für →intervallskalierte Daten ist das Korrelationsmaß der Pearsonsche Produkt-Moment-Korrelationskoeffizient r_{xy} (kurz „**Pearsons r** “ oder nur „ **r** “). r_{xy} hat einen Wertebereich von -1 bis $+1$. Ein hohes negatives r_{xy} besagt: Je höher das eine Merkmal ausgeprägt ist, desto niedriger ist das andere Merkmal, und je niedriger das eine Merkmal, desto höher das andere Merkmal. Ein hohes positives r_{xy} besagt sinngemäß entsprechend: Je höhere Werte das eine Merkmal annimmt, desto höhere hat auch das andere (bzw. je niedriger, desto niedriger). Ein r_{xy} von Null sagt aus, dass zwischen den beiden Merkmalen kein linearer Zusammenhang besteht. r_{xy}^2 ist ein Maß für die →erklärte Varianz.

Kriterium(s-Variable)

Ein anderer Begriff für →abhängige Variable, siehe auch →Regressionsanalyse.

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

M

Abkürzung für \rightarrow Mittelwert (engl. Mean).

$M \pm SD$, Wertebereich, der durch eine Streuungseinheit ($\rightarrow SD$) oberhalb und unterhalb des Mittelwerts (M) abgedeckt wird.

MARKUS, Mathematik-Gesamterhebung Rheinland-Pfalz: Kompetenzen, Unterrichtsmerkmale, Schulkontext. Gegenstand ist eine Vollerhebung der Mathematikleistungen der Schüler in der 8. Jahrgangsstufe sowie zu individuellen Lernvoraussetzungen, zum persönlichen Lernhintergrund und zu Unterrichtsmerkmalen (Mathematiktests, Schüler-, Lehrer- und Schulleiterfragebogen). Durchgeführt von einer Forschergruppe der Universität Landau.

Median, der Median beschreibt den Wert einer Verteilung, ober- und unterhalb dessen 50% aller Fälle oder Werte angesiedelt sind.

Mittelwert, Kurzbezeichnung für das arithmetische Mittel. Es entspricht der Summe der Einzelwerte aller Fälle dividiert durch die Zahl der Fälle. Der Mittelwert ist ein sinnvolles Maß, wenn mindestens \rightarrow intervallskalierte Daten vorliegen.

N, bezeichnet die Anzahl der untersuchten Personen.

Normalverteilung, auch Gaußsche Verteilung oder Glockenkurve genannt, ist eine symmetrische, glockenförmige Verteilung, die anhand von nur zwei Kennwerten vollständig beschrieben wird. Diese Kennwerte sind der \rightarrow Mittelwert und die \rightarrow Standardabweichung. Bei einer Normalverteilung entfallen 68 % aller Fälle auf das Intervall von einer Standardabweichung unterhalb bis einer Standardabweichung oberhalb des Mittelwertes.

Objektivität, Objektivität ist ein Gütekriterium für sozialwissenschaftliche Messungen. Sie besagt, dass die Ergebnisse der Messung unabhängig vom Untersucher sind.

Objektivität in Schulleistungsuntersuchungen ist gegeben, wenn für alle Schülerinnen und Schüler gleiche Aufgabenstellungen, Bearbeitungszeiten, Erläuterungen der Aufgaben, Arbeitsmaterialien u. ä. gelten und wenn Auswertung und Interpretation nach klaren Kriterien, die unabhängig von der Person des Auswerters sind, erfolgen.

PISA, Programme for International Student Assessment. Studie der OECD (1998 - 2007) zur Lesekompetenz, zur mathematisch-naturwissenschaftlichen Grundbildung und zu fächerübergreifenden Kompetenzen mit vielfältigen Indikatoren für Lernergebnisse bei 15jährigen Schülern. Federführend für den ersten Zyklus (PISA 2000) mit dem Schwerpunkt Leseverständnis: MPI für Bildungsforschung Berlin; für den zweiten Zyklus (PISA 2003) mit dem Schwerpunkt Mathematik: das IPN Kiel.

Populationsdaten, diese Daten repräsentieren die untersuchte Gesamtanzahl von Individuen (Grundgesamtheit oder auch Grundpopulation). Bei bestimmten Fragestellungen wird aus pragmatischen Erwägungen normalerweise nicht die Grundgesamtheit, sondern eine Stichprobe untersucht, die für die Grundgesamtheit repräsentativ ist.

Rasch-Modell, ein Messmodell im Rahmen der probabilistischen →Testtheorie, mit dessen Hilfe Personen unterschiedlicher Fähigkeit und Aufgaben unterschiedlicher Schwierigkeit auf einer gemeinsamen Skala bzw. Dimension abgebildet werden.

Regressionsanalyse, die (multiple) Regressionsanalyse ist ein Analyseverfahren, das den Zusammenhang zwischen einer →intervallskalierten abhängigen (zu erklärenden) Variable (dem so genannten Kriterium) und mehreren, ebenfalls intervallskalierten unabhängigen (erklärenden) Variablen (den so genannten Prädiktoren) aufdeckt. Bei der Berechnung der Regressionsgleichung werden die →Korrelationen der Prädiktoren untereinander berücksichtigt.

Reliabilität, Reliabilität ist ein Gütekriterium für sozialwissenschaftliche Messungen, das die Zuverlässigkeit einer Messung kennzeichnet. Reliabel ist ein Test oder eine Skala, wenn nur geringe Messfehler auftreten.

SD, Standard deviation: englisch für Streuung oder →Standardabweichung.

Signifikanz, signifikant, Signifikanzniveau, von einem signifikanten oder statistisch bedeutsamen Ergebnis spricht man im allgemeinen dann, wenn die Irrtumswahrscheinlichkeit sehr gering (in der Regel höchstens 5%) ist.

Skala, 1. Kurzbezeichnung für eine Einschätz- oder Beurteilungsskala (Ratingskala). So entsprechen z.B. die Antwortmöglichkeiten von 0 = „nie“ bis 4 = „sehr oft“ im Lehrerfragebogen zur Einschätzung der inneren Differenzierung einer fünfstufigen Skala. **2.** Inhaltlich zusammenpassende Einzelitems oder –fragen (→Items) können, z. B. durch Aufsummieren oder Mittelwertbildung, zu einer Skala zusammengefasst werden. Ein Beispiel ist die Skala „Innere Differenzierung“, bei der für jede Lehrkraft der Mittelwert ihrer Antworten auf 7 Fragen des Lehrerfragebogens berechnet wurde, um ein Maß für ihre Bereitschaft zu erhalten, Maßnahmen der inneren Differenzierung einzusetzen.

Standardabweichung, SD, Die Standardabweichung ist ein so genanntes Streuungsmaß, das für intervallskalierte Daten Auskunft darüber gibt, wie homogen oder heterogen eine Merkmalsverteilung ist. Je kleiner die Standardabweichung ist, desto enger gruppieren sich die Werte der einzelnen Fälle um den →Mittelwert; je größer sie ist, desto weiter streuen sie um den Mittelwert. Liegt eine →Normalverteilung vor, so lässt sich über die Verteilung folgendes sagen: Im Bereich Mittelwert \pm eine Standardabweichung liegen etwa 68 Prozent der Fälle; im Bereich Mittelwert \pm zwei Standardabweichungen liegen etwa 95 Prozent der Fälle.

Streuung

→Standardabweichung

t-Test, beim t-Test handelt es sich um ein statistisches Testverfahren, mit dessen Hilfe geprüft wird, ob sich die →Mittelwerte zweier Gruppen statistisch →signifikant voneinander unterscheiden. So könnte z.B. geprüft werden, ob sich die mittlere Testleistung der Mädchen statistisch signifikant von der der Jungen unterscheidet. Als Prüfgröße wird der t-Wert berechnet. Das analoge statistische Verfahren für den Vergleich der Mittelwerte mehrerer Gruppen ist die →Varianzanalyse.

t-Wert

Statistische Prüfgröße bei einem →t-Test.

Testtheorie, die der Konstruktion von psychologischen und pädagogischen Tests zugrundeliegende mathematisch-statistische Theorie. Die Testtheorie befasst sich u.a. mit der Frage, wie empirische Testwerte und die zu messenden Merkmalsausprägungen zusammenhängen. Aus den Annahmen einer Testtheorie können Gütekriterien wie →Reliabilität, →Validität und →Objektivität abgeleitet werden. Man kann z.B. mit Hilfe einer Testtheorie prüfen, ob eine →Skala statistisch akzeptiert werden kann.

TIMSS, Third International Mathematics and Science Study. Diese Studie setzt die Reihe der international vergleichenden Schulleistungsuntersuchungen fort, die seit 1959 von der →IEA durchgeführt werden. TIMSS umfasste drei Altersgruppen: Population I (Ende der Grundschule), II (Sekundarstufe I) und III (Ende der Pflichtschulzeit, Sekundarstufe III) und fokussierte auf naturwissenschaftliche und mathematische Leistungen. In Deutschland wurden nur die Populationen II und III untersucht.

Validität, Validität ist ein Gütekriterium für sozialwissenschaftliche Messungen. Validität gibt die Gültigkeit eines Messinstruments, z. B. eines Tests, an. Ein Test ist valide, wenn er tatsächlich das misst, was er zu messen beansprucht.

Varianz, die Varianz entspricht dem Quadrat der →Standardabweichung. Mathematisch ist die Varianz der Durchschnitt aus den quadrierten Abweichungen aller Einzelwerte vom Mittelwert.

Varianzanalyse, die Varianzanalyse (ANOVA, analysis of variance) ist ein Verfahren zur statistischen Überprüfung von Mittelwertsunterschieden zwischen verschiedenen Gruppen und stellt damit die Verallgemeinerung des →t-Tests auf mehr als 2 Gruppen dar. So könnte z.B. geprüft werden, ob sich die mittleren Testleistungen der Schülerinnen und Schüler aus den 4 Bildungsganggruppen Gymnasium, Realschule, Hauptschule A-Kurs und Hauptschule G-Kurs statistisch signifikant voneinander unterscheiden.

7 Literatur

- Blum, W., Neubrand, M., Ehmke, T., Senkbeil, M., Jordan, A., Ulfig, F. & Carstensen, C. H. (2004). Mathematische Kompetenz. In M. Prenzel, J. Baumert, W. Blum, R. Lehmann, D. Leutner, M. Neubrand, R. Pekrun, H.-G. Rolff, J. Rost & U. Schiefele (Hrsg.), *PISA 2003. Der Bildungsstand der Jugendlichen in Deutschland - Ergebnisse des zweiten internationalen Vergleichs* (S. 47-92). Münster: Waxmann.
- Deutsches PISA - Konsortium (Hrsg.). (2003). *PISA 2000 - Ein differenzierter Blick auf die Länder der Bundesrepublik Deutschland*. Opladen: Leske + Budrich.
- Helmke, A. & Hosenfeld, I. (2004). Vergleichsarbeiten - Kompetenzmodelle - Standards. In M. Wosnitza, A. Frey & R. S. Jäger (Hrsg.), *Lernprozesse, Lernumgebungen und Lerndiagnostik. Wissenschaftliche Beiträge zum Lernen im 21. Jahrhundert* (S. 56-75). Landau: Verlag Empirische Pädagogik.
- Helmke, A., Hosenfeld, I. & Schrader, F.-W. (2002). Unterricht, Mathematikleistung und Lernmotivation. In A. Helmke & R. S. Jäger (Hrsg.), *Die Studie MARKUS - Mathematik-Gesamterhebung Rheinland-Pfalz: Kompetenzen, Unterrichtsmerkmale, Schulkontext*. (S. 413-480). Landau: Verlag Empirische Pädagogik.
- Helmke, A., Hosenfeld, I. & Schrader, F.-W. (2003). Diagnosekompetenz in Ausbildung und Beruf entwickeln. *Karlsruher Pädagogische Beiträge* (55), 15-34.
- Helmke, A., Hosenfeld, I. & Schrader, F.-W. (2004). Vergleichsarbeiten als Instrument zur Verbesserung der Diagnosekompetenz von Lehrkräften. In R. Arnold & C. Griese (Hrsg.), *Schulleitung und Schulentwicklung* (S. 119-144). Hohengehren: Schneider-Verlag.
- Helmke, A. & Reich, H. H. (2001). Die Bedeutung der sprachlichen Herkunft für die Schulleistung. *Empirische Pädagogik*, 15 (4), 567-600.
- Helmke, A. & Weinert, F. E. (1997). Bedingungsfaktoren schulischer Leistungen. In F. E. Weinert (Hrsg.), *Psychologie des Unterrichts und der Schule* (Enzyklopädie der Psychologie, Pädagogische Psychologie, Vol. 3, S. 71-176). Göttingen: Hogrefe.
- Hosenfeld, I., Helmke, A., Ridder, A. & Schrader, F.-W. (2001). Eine mehrbenenanalytische Betrachtung von Schul- und Klasseneffekten. *Empirische Pädagogik*, 15 (4), 513-534.
- Hosenfeld, I., Helmke, A., Ridder, A. & Schrader, F.-W. (2002). Die Rolle des Kontextes. In A. Helmke & R. S. Jäger (Hrsg.), *Die Studie MARKUS - Mathematik-Gesamterhebung Rheinland-Pfalz: Kompetenzen, Unterrichtsmerkmale, Schulkontext*. (S. 155-256). Landau: Verlag Empirische Pädagogik.
- Klieme, E., Avenarius, H., Blum, W., Döbrich, P., Gruber, H., Prenzel, M., Reiss, K., Riquarts, K., Rost, J., Tenorth, H.-E. & Vollmer, H. J. (2003). *Zur Entwicklung nationaler Bildungsstandards. Eine Expertise*. Frankfurt a. M.: DIPF.

- Peek, R. & Dobbelstein, P. (2006). Benchmarks als Input für die Schulentwicklung. In H. Kuper & J. Schneewind (Hrsg.), *Rückmeldung und Rezeption von Forschungsergebnissen - Zur Verwendung wissenschaftlichen Wissens im Bildungssystem* (S. 41-58). Münster: Waxmann.
- Projektgruppe VERA-Deutsch. (2005). *Didaktische Erläuterungen Lesen*. Landau.
- Rasch, G. (1960). *Probabilistic Models for Some Intelligence and Attainment Tests* (Studies in Mathematical Psychology). Copenhagen: Nielsen & Lydiche.
- Schwippert, K., Bos, W. & Lankes, E. M. (2003). Heterogenität und Chancengleichheit am Ende der vierten Jahrgangsstufe im internationalen Vergleich. In W. Bos, E. M. Lankes, M. Prenzel, K. Schwippert, G. Walther & R. Valtin (Hrsg.), *Erste Ergebnisse aus IGLU* (S. 265-302). Münster: Waxmann.
- Zimmer, K., Burba, D. & Rost, J. (2004). Kompetenzen von Jungen und Mädchen. In M. Prenzel, J. Baumert, W. Blum, R. Lehmann, D. Leutner, M. Neubrand, R. Pekrun, H.-G. Rolff, J. Rost & U. Schiefele (Hrsg.), *PISA 2003. Der Bildungsstand der Jugendlichen in Deutschland - Ergebnisse des zweiten internationalen Vergleichs* (S. 211-223). Münster: Waxmann.