



**Ergebnisbericht
VERA 2005:
Nordrhein-Westfalen**

Andreas Helmke, Ingmar Hosenfeld, Jana Groß Ophoff, Ann Christin Halt,
Jan Hochweber, Kevin Isaac, Ursula Koch und Frank Scherthan
Universität Koblenz - Landau, Campus Landau

Stand: 31.03.2006

Inhalt

1 Einleitung und Überblick	4
1.1 Ziele	4
1.2 Organisation	5
2 Aufgabenentwicklung und Definition der Fähigkeitsniveaus	6
2.1 Deutsch	7
2.1.1 Aufgabenentwicklung.....	7
2.1.2 Skalierung.....	8
2.1.3 Fähigkeitsniveaus	9
2.2 Mathematik.....	12
2.2.1 Aufgabenentwicklung.....	12
2.2.2 Skalierung und Fähigkeitsniveaus des Mathematiktests	13
3 Vergleichsarbeiten im September 2005	17
3.1 Ablauf und Zentralstichprobe.....	17
3.2 Elemente der Ergebnisrückmeldungen.....	18
3.3 Ergebnisse zur Durchführung der Vergleichsarbeiten.....	19
3.3.1 Auswertung in den Schulen.....	20
3.3.2 Ergebnisse zur Nutzung des Internet.....	23
3.3.3 Ergebnisse zum Supportaufkommen.....	28
3.4 Gesamt-Ergebnisse für die Fähigkeitsniveaus.....	29
3.4.1 Gesamt-Verteilung.....	29
3.4.2 Zusammenhänge zwischen den Fähigkeitsniveaus.....	30
4 Landesspezifische Ergebnisse	30
4.1 Fähigkeitsniveaus	30
4.1.1 Verteilung der Fähigkeitsniveaus in den Ländern	31
4.1.2 Veränderungstrends 2004–2005.....	32
4.1.3 Unterschiede innerhalb und zwischen den Klassen	32
4.1.4 Leistungen von Mädchen und Jungen	34
4.1.5 Migrationshintergrund	35
4.2 „Fairer Vergleich“	37
4.2.1 Beschreibung ausgewählter Kontextgruppen.....	38
4.2.2 Bildung der Kontextgruppen	39

4.2.3	<i>Verteilung der Fähigkeitsniveaus nach Kontextgruppen</i>	42
4.2.4	<i>Verteilung aller Schulen auf die Kontextgruppen</i>	43
4.2.5	<i>Vergleich tatsächliche vs. erwartete Leistung</i>	44
4.3	<i>Lehrerfragebogen</i>	45
4.3.1	<i>Kontinuität des Unterrichts in Mathematik und Deutsch und Fähigkeitsniveaus.....</i>	46
4.3.2	<i>Unterrichtserfahrung, grundständige Ausbildung und Fähigkeitsniveaus.....</i>	48
4.3.3	<i>Vorbereitung auf die Vergleichsarbeiten</i>	50
4.3.4	<i>Kooperation bei der Aufgabenauswahl und -auswertung</i>	51
5	Ausblick	52

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 Bundesländern 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

1.2 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 vierten Klassen aus allen fünf Erdteilen das Angebot zur Teilnahme an VERA an. Abbildung 1 verdeutlicht die Vernetzung verschiedener Funktionen und Gruppen.

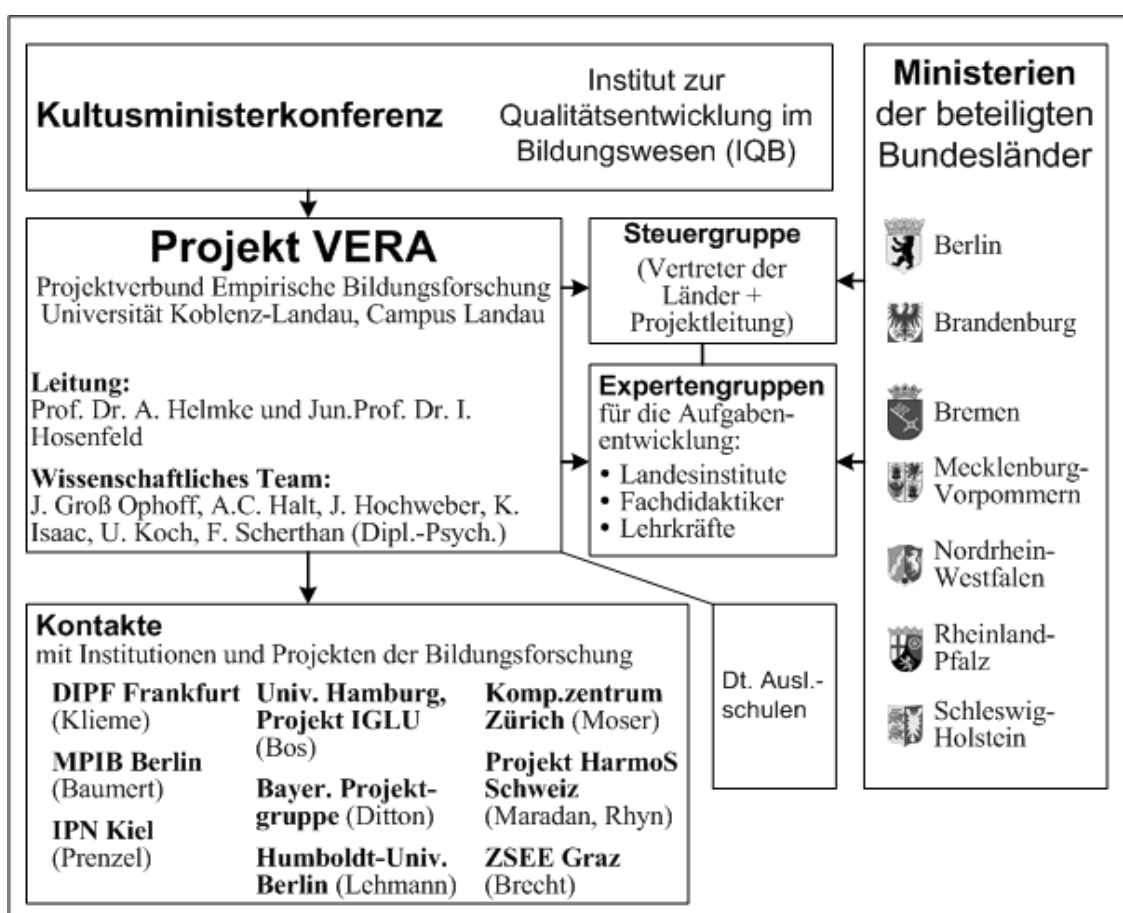


Abbildung 1: Bildungspolitische Kontext von VERA

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 Vorabhebungen 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 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“ (VERA-Projektgruppe 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 (s. dazu Helmke & Hosenfeld, 2004). Nach der Erfassung, Aufbereitung und Qualitätskontrolle wurde 2004 für die vier Inhaltsbereiche jeweils ein ein-parametrisches Modell (sog. 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. Baumert et al., 2002), Kompetenzniveaus (Lernstandserhebungen Klasse 9, NRW¹) 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

* Für jeden der Inhaltsbereiche der Fächer Deutsch und Mathematik wurden jeweils drei Fähigkeitsniveaus festgelegt.

¹ <http://www.learn-line.nrw.de/angebote/lernstand9/>; Zugriff am 29. März 2005

die Aufgaben richtig gelöst wurden.

- Auf der anderen Seite sind Annahmen über Denkprozesse erforderlich, die der Aufgabenbearbeitung 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

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

gewährleisten war, dass jede Schülerin und jeder Schüler in der Vergleichsarbeit über die entsprechenden Hilfsmittel würde verfügen können.

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 (einschränkte 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.

3 Vergleichsarbeiten im September 2005

3.1 Ablauf und Zentralstichprobe

Von Ende August bis Anfang September 2005 fanden in den sieben beteiligten Ländern Erprobungsphasen mit Fokus auf der Internetnutzung statt. Vom 12. September bis zum 16. September konnten die Schulen in Berlin, Brandenburg, Bremen, Mecklenburg-Vorpommern, Nordrhein-Westfalen und Rheinland-Pfalz internetbasiert und menügesteuert einen Teil der Mathematik-Aufgaben auswählen (zehn Aufgaben, die Hälfte der insgesamt erfassten Aufgaben). Vom 12. September bis zum 27. September hatten alle Lehrkräfte außer in Nordrhein-Westfalen außerdem die Möglichkeit, die Lösungshäufigkeiten für die gewählten Aufgaben einzuschätzen, um freiwillig die eigene Diagnosegenauigkeit überprüfen zu können. In Nordrhein-Westfalen und Bremen wurden die Deutsch-Testhefte zentral gedruckt und kurz vor den Vergleichsarbeiten an die Schulen gesendet. In allen anderen Ländern sowie im Fach Mathematik

konnten die Testhefte (Mathematik: Zentral- und Wahlaufgaben, Deutsch: nur Zentralaufgaben) von den Schulen vier Tage vor der Durchführung der Vergleichsarbeit in Mathematik heruntergeladen werden. Die Schulen in Schleswig-Holstein erhielten nur zentral vorgegebene Testaufgaben.

Zeitgleich am 27. September 2005 wurde in allen öffentlichen und einigen privaten Grundschulen der beteiligten Bundesländer die Vergleichsarbeit im Fach Mathematik geschrieben, am 29. September in Deutsch. Die Bearbeitungszeit betrug in beiden Fächern 50 Minuten. Nach der Durchführung der Vergleichsarbeiten konnten die Lehrkräfte neben diversen Informationsmaterialien (z.B. didaktische Erläuterungen) die Korrekturanweisungen für die zentral vorgegebenen und die von ihnen ausgewählten Mathematik-Aufgaben vom Landauer Universitätsserver herunterladen. Nach der Korrektur durch die Lehrkräfte konnten diese die Ergebnisse offline in ein Excel-Sheet oder online direkt im geschützten Internet-Bereich von VERA auf dem Universitätsserver eingeben. Die offline eingegebenen Werte wurden dann zum Server hochgeladen und konnten auf den VERA-Internetseiten eingesehen werden. Zwei Wochen nach der vollständigen Dateneingabe konnten die Lehrkräfte die ersten Ergebnisse ihrer Schülerinnen und Schüler abrufen. Für den Abschluss der Dateneingabe hatten die Schulen bis zum 18. November Zeit. Ausgenommen von dieser Regelung waren die Schulen der Zentralstichprobe, welche die Dateneingabe zwei Wochen früher, also bis zum 04. November, abgeschlossen haben mussten. Die Zentralstichprobe dient dem Zweck, für das jeweilige Bundesland repräsentative Aussagen zur Standortbestimmung zu gewinnen, mit deren Hilfe ein fairer Vergleich möglich werden soll. Hierfür wurde im Vorfeld der Vergleichsarbeiten eine Stichprobe gezogen. Die Ziehung der Stichprobe erfolgte länderweise unter Beachtung von drei Kategorien (einzügige, zwei- bis dreizügige, mehr als dreizügige Schulen). In die Zentralstichprobe gingen aus den Bundesländern Berlin, Brandenburg, Mecklenburg-Vorpommern, Nordrhein-Westfalen, Rheinland-Pfalz und Schleswig-Holstein 1064 Klassen in 545 Schulen ein, zusätzlich alle nicht-freiwilligen Schulen aus Bremen (252 Klassen in 91 Schulen). Um spezielle Vorbereitung und unerwünschtes „Training“ zu vermeiden, erhielten die Schulen erst nach der Durchführung der Vergleichsarbeiten beim Herunterladen der Korrekturanweisung die Mitteilung, dass sie Teil der diesjährigen Zentralstichprobe seien. Die Durchführung der Vergleichsarbeiten wie auch die schulinterne Auswertung waren somit in den Klassen der Zentralstichprobe identisch mit der Durchführung in allen anderen Klassen. Sowohl die Eingabe der Daten als auch die Ergebnisrückmeldung aus Landau waren ebenfalls für alle Schulen gleich, die Lehrkräfte der Zentralstichprobe beantworteten zusätzlich einen Kontextfragebogen, dessen Informationen für den „fairen Vergleich“ benötigt wurden.

3.2 Elemente der Ergebnisrückmeldungen

Die Ergebnisse wurden den Schulen in zwei Wellen über das Internet zurückgemeldet, ab dem 19. Dezember 2005 standen die Zusatzinformationen zur Verfügung. Die Lehrerinnen und Lehrer sollten anschließend die Informationen in geeigneter Weise an Eltern, Schulleitung oder Schulaufsicht weiterleiten und erläutern.

Zeitnah zur vollständigen Dateneingabe konnten die Lehrkräfte die **Basisinformationen** (erste Rückmeldewelle) einsehen:

- Fähigkeitsniveaus auf Ebene der einzelnen Schülerinnen und Schüler als Tabellenübersicht, auf Klassen- und auf Schulebene als Verteilungen (gestaffelte Balkendiagramme)

Nur im Fach Mathematik:

- Analysen zur Fehlerhäufigkeit: Die in der jeweiligen Klasse aufgetretenen Fehler wurden den Fehlerhäufigkeiten aus der Normierungsstichprobe gegenübergestellt, um so Informationen über Fehlerschwerpunkte und damit Ansatzpunkte für differenzierte Interventionsmaßnahmen in der Klasse zu gewinnen.
- Analysen zur diagnostischen Kompetenz: Die Einschätzungen der Lösungshäufigkeiten für die Wahlaufgaben wurden der tatsächlichen Lösungshäufigkeit gegenüber gestellt.

In der zweiten Rückmeldewelle wurden den Lehrkräften ab dem 19. Dezember 2005 weitere **Zusatzinformationen** zur Verfügung gestellt, und zwar:

- Fähigkeitsniveaus im Vergleich zum eigenen Bundesland: Die Gegenüberstellung der Verteilung der Fähigkeitsniveaus in der eigenen Klasse und Schule wurde kontrastiert mit der Darstellung der Ergebnisse des jeweiligen Bundeslandes.
- Fähigkeitsniveaus im „fairen Vergleich“: Zusätzlich erhielten die Lehrkräfte Ergebnisse zum Vergleich der eigenen Klasse mit einer länderunspezifischen kontextähnlichen Vergleichsgruppe.
- Lösungshäufigkeiten in den Zentralaufgaben bei VERA 2005: In dieser Darstellung wurden die richtigen Lösungen der jeweiligen Klasse im Vergleich zur Gesamtpopulation verortet.
- Analysen zur diagnostischen Kompetenz: Aus der Rangordnung der geschätzten und tatsächlichen Lösungshäufigkeiten in den Wahlaufgaben wurde ein Genauigkeitsindex für die eigene Einschätzung (Korrelation) abgeleitet. Dieser Wert wurde in einer Häufigkeitsverteilung der Population von VERA 2005 gegenüber gestellt.

Als Neuerung wurden den Schulen ab Mitte Januar zwei Gesamtdokumente mit allen Ergebnissen aus VERA 2005 zur Verfügung gestellt (Teil 1: Ergebnisse zu den Fähigkeitsniveaus; Teil 2: Aufgabenbezogene Ergebnisse). Diese können von den Schulen als schulinterne Dokumentation der Ergebnisse in den Vergleichsarbeiten 2005 oder als Basis für Berichte an z.B. die Schulaufsicht genutzt werden.

3.3 Ergebnisse zur Durchführung der Vergleichsarbeiten

Erstmalig wurden in Landau nach den Vergleichsarbeiten 2005 gezielt Informationen zur Nutzung des Internets sowie zum Supportaufkommen während der Durchführung der Vergleichsarbeiten ausgewertet. Außerdem wurde eine stichprobenartige Überprüfung der Ergebnis-Eingaben der Lehrkräfte vorgenommen, unter anderem mit dem Ziel, Erkenntnisse über die Verständlichkeit der Korrekturanweisungen zu erlangen.

In den folgenden Unterkapiteln soll etwas näher auf diese Ergebnisse eingegangen sowie einzelne Aspekte aus der Befragung der Zentralstichprobe dargestellt werden.

3.3.1 Auswertung in den Schulen

Um eine Einschätzung für die Genauigkeit der Ergebnis-Eingabe durch die Lehrkräfte sowie damit verbunden für die Verständlichkeit und Eindeutigkeit der in VERA verwendeten Korrekturanweisungen zu bekommen, wurde eine zufällige Stichprobe der Schulen gezogen, die ihr Test-Material zur Verfügung stellen sollten. Insgesamt wurden 50 Schulen von uns angeschrieben und darum gebeten, Kopien der Testhefte (Mathematik und Deutsch) von jeweils zwei per Zufall bestimmten Schülerinnen bzw. Schülern einzusenden. Dieser Aufforderung kamen 49 Schulen nach, eine Schule hatte die Testhefte bereits vernichtet.

Diese wurden in Landau erneut korrigiert und mit den entsprechenden Eingaben durch die Lehrkräfte abgeglichen.

Die von uns gefundenen Abweichungsarten sind in Tabelle 1 schematisch dargestellt.

Tabelle 1: Abweichungsarten

		Korrekte Lösung lt. Korrekturanweisung		
		richtig	falsch	nicht bearbeitet
Lehrerurteil	richtig	X	1. und 2.	kommt nicht vor
	falsch	4.	X	kommt nicht vor
	nicht bearbeitet	kommt nicht vor	3.	X

Deutsch

Bei allen Schulen lagen Abweichungen (Korrekturmängel) verschiedener Arten vor, die wie folgt unterschieden wurden:

Tabelle 2: Prozentuale Verteilung aller Abweichungen in Deutsch

Abweichungsarten	N	Prozent (vorliegende Aufgaben)
1. falsch gelöst und als „richtig“ bewertet	68	3,65
2. großzügige Bewertung	36	1,93
3. falsch gelöst und als „n.b.“ bewertet	16	0,86
4. richtig gelöst und als „falsch“ bewertet	10	0,54
Anzahl der Beanstandungen	130	6,98
Anzahl vorliegende Aufgaben	1862	100

Falsch gelöste Aufgaben, die als „richtig“ bewertet wurden:

Diese Art der Abweichung beschreibt Fälle, in denen als „falsch“ zu bewertende Bearbeitungen als „richtig“ bewertet wurden. Sie machen mit 68 Stück oder 3,65 Prozent den Löwenanteil aller Korrekturmängel aus. Diese Art der Abweichung kam bei 37 Schulen vor. Es kamen jedoch auch einzelne Schulen (N = 5) vor, bei denen zusätzlich ebenfalls Abweichungen der Art vorkamen, dass richtig gelöste Aufgaben als „falsch“ bewertet wurden. Diese Mischfälle könnte man als unaufmerksame Korrekturen auffassen. 19 Schulen zeigen neben der hier dargestellten Abweichung auch großzügige Bewertungen.

Großzügige Bewertung:

Hier handelt es sich um Grenzfälle, in denen die Bearbeitung streng genommen als „falsch“ hätte eingestuft werden müssen und in denen die Lehrkräfte die Korrekturanweisungen etwas liberaler ausgelegt haben. Diese Abweichungen können im Wohlwollen der Lehrkräfte oder in bestimmten Formulierungen in den Korrekturanweisungen (Interpretationsspielraum, der insbesondere für offene Deutsch-Aufgaben typisch ist) begründet sein. Diese Abweichungen entsprechen 1,93 Prozent (36 Fälle) und kamen in 29 Schulen vor.

Falsch gelöste Aufgaben, die als „nicht bearbeitet (n. b.)“ gewertet wurden:

In diesem Fall wurden Aufgaben, die aus unserer Sicht eindeutig als „falsch“ einzustufen waren, von den Lehrkräften als „nicht bearbeitet“ gewertet. Diese Abweichungen entsprechen 0,86 Prozent bzw. 16 Fällen und kamen in zehn Schulen vor.

Richtig gelöste Aufgaben, die als „falsch“ bewertet wurden:

Diese Art der Abweichung benachteiligt letztlich den Schüler, da hier eine richtige Antwort von der Lehrkraft als „falsch“ eingestuft wurde. Diese Form der Fehlzuordnung kam nur in 0,54 Prozent der Fälle vor (neun Schulen). In acht Schulen trat dieser Fehler nur einmal auf. Nur in einem Fall wurde dreimal eine richtige Antwort irrtümlicherweise als „falsch“ eingestuft.

Fazit:

In Bezug auf einzelne Schulen und Bundesländer sind keine auffälligen Fehlermuster zu berichten. Es scheint keine systematische Tendenz zur „besseren Darstellung“ der eigenen Schüler zu geben, selbst wenn es nachweislich einen geringen Prozentsatz an Aufgabenkorrekturen gibt, den man als absichtliche Verfälschung interpretieren könnte. Viele der Beanstandungshäufigkeiten gehen vermutlich auf Konzentrationsprobleme oder nicht auf als Täuschung zu interpretierendes Wohlwollen seitens der Lehrkräfte zurück. Darüber hinaus ist insbesondere der Bereich Deutsch mit der Tatsache konfrontiert, dass die Korrekturanweisungen bestimmte Interpretationsspielräume zulassen müssen (z.B. bei offenen Aufgaben), was sich in den am zweithäufigsten vertretenen Beanstandungen, nämlich den „großzügigen Bewertungen“ ausdrückt. Dass knapp 60

Prozent aller 130 Beanstandungen auf nur drei der 19 Teilaufgaben entfallen, kann als deutliches Indiz für diese These aufgefasst werden.

Mathematik

Für 43 der 49 Schulen wurden Abweichungen zwischen der Landauer Korrektur und der Ergebnisübermittlung durch die Lehrkraft gefunden. Unterscheiden lassen sich insgesamt drei verschiedene Abweichungsarten: 1. falsch gelöste Aufgaben wurden nicht dem richtigen Fehlertypen zugeordnet, 2. falsch gelöste Aufgaben wurden als „richtig“ bewertet, 3. richtig gelöste Aufgaben wurden als „falsch“ bewertet.

Die folgende Tabelle zeigt die prozentuale Verteilung aller Abweichungen (N = 183) auf die drei Abweichungstypen:

Tabelle 3: *Prozentuale Verteilung aller Abweichungen in Mathematik*

Abweichungsart	Anzahl	Prozent
falsch gelöst, falsche Fehlertypangabe (1)	110	ca. 60
falsch gelöst, als „richtig“ bewertet (2)	43	ca. 23
richtig gelöst, als „falsch“ bewertet (3)	30	ca. 16
gesamt	183	

Am häufigsten wurde der Fehlertyp der Falschlösungen nicht entsprechend der Vorgaben in den Korrekturanweisungen bestimmt. Insgesamt kam es bei dieser Art der Abweichung zu 110 falschen Fehlertypzuordnungen, die sich auf 39 der 43 überprüften Schulen verteilten. In 43 Fällen, verteilt auf 26 Schulen, wurden Falschlösungen als „richtig“ eingestuft, in 30 Fällen, verteilt auf 19 Schulen, die richtige Aufgabenlösung als „falsch“ gewertet. Mögliche Ursachen hierfür könnten in einer Fehlinterpretation der Korrekturanweisungen bzw. in empfundenen Widersprüchen in der Art der vorgegebenen Aufgabenkorrektur liegen.

Tabelle 4 gibt einen Überblick über die Häufigkeiten der Angabe falscher Aufgabenlösungen als „richtig“ bzw. umgekehrt.

Tabelle 4: *Auftretenshäufigkeit der Abweichungsarten (2) und (3)*

Abweichungsart	Anzahl
alleiniges Auftreten von Abweichtungstyp (2)	17
alleiniges Auftreten von Abweichtungstyp (3)	9
gleichzeitiges Auftreten von Abweichtungstyp (2) und (3)	19

Das gehäufte gleichzeitige Auftreten von Abweichtungstyp (2) und (3) in vielen der Schulen legt nahe, dass die Fehl kategorisierung von falsch gelösten Aufgaben als „richtig“ bzw. richtig gelöster Aufgaben als „falsch“ nicht Konsequenz absichtlicher Verfälschung ist, sondern aus unterschiedlich strengen Bewertungsmaßstäben und möglicherweise auftretenden Konzentrationsproblemen bei der Dateneingabe resultiert. Die meisten Fehlzuordnungen (auch bei der Falscheinordnung von Fehlertypen) kon-

zentrieren sich auf wenige, in der Auswertung aufwändige Aufgaben mit evtl. verbesserungswürdigen Korrekturanweisungen (z.B. FV093: Zuordnung von Alltagsgegenständen zu geometrischen Formen; FV103: Zerlegen einer Figur in Teilfiguren). Keine der Schulen und keines der Bundesländer fällt durch eine gehäufte Anzahl von Fehl kategorisierungen auf.

Setzt man die aufgetretenen Abweichungen ins Verhältnis zu der Gesamtzahl der durch die Lehrkräfte insgesamt im Rahmen der Vergleichsarbeiten für eine Klasse korrigierten (Teil-)Aufgaben, so wird deutlich, dass diese nur einen geringen Prozentsatz, nämlich deutlich unter zwei Prozent aller Beurteilungen, ausmachen. Daraus lässt sich folgern, dass die Materialien zur Korrektur der VERA-Mathematikaufgaben weitestgehend eindeutig und verständlich sind und von den Lehrkräften gewissenhaft angewendet werden.

3.3.2 Ergebnisse zur Nutzung des Internet

Grundlegend für die Durchführung von flächendeckenden und bundesländerübergreifenden Vergleichsstudien ist die Nutzung des Internet: So wählen die an VERA teilnehmenden Lehrkräfte über das Internet einen Teil der Testaufgaben aus, laden das dynamisch generierte Aufgabenheft herunter und geben im Anschluss die Resultate der Schülerinnen und Schüler online ein. Die aufbereiteten Ergebnisse und Informationsmaterialien werden ebenfalls über das Internet distribuiert. Ergänzend nehmen die Lehrkräfte an einer Online-Befragung zum sozialen Hintergrund ihrer Klassen ein und haben die Möglichkeit, in einer freiwilligen Befragung Lob und Kritik am Projekt zu äußern.

Mit der Nutzung des Internet als Instrument für Schulleistungsstudien wurde durch das Projekt VERA Neuland betreten: Obwohl inzwischen so gut wie alle Schulen über einen Zugang zum Internet verfügen, bedeutet die Nutzung des Internet für einige Schulen eine Umstellung. Sie soll jedoch zu einem Aufschwung in der effizienten Nutzung des Internet für die schulische Qualitätssicherung führen und damit auch einen Beitrag zur Förderung der Medienkompetenz leisten.

Soft- und Hardware-Voraussetzungen der VERA-Nutzer

Sowohl an die Hardware als auch an die Software gibt es grundlegende Anforderungen, damit der VERA-Prozess von den Schulen vollständig durchlaufen werden kann. Der Begriff Hardware wird dabei als Oberbegriff für die maschinentechnische Ausrüstung eines Computersystems verwendet. Von besonderer Bedeutung ist in diesem Zusammenhang die Bildschirmauflösung der für die Durchführung und Auswertung der Vergleichsarbeiten verwendeten Computer.

Entsprechende Informationen werden mit Hilfe von JavaScript auf der Startseite des geschützten Bereichs erfasst. Die in Tabelle 5 dargestellten Häufigkeiten beziehen sich demnach nur auf den Aufruf der Startseite von August 2005 bis Februar 2006. Alles in allem haben wir 53 unterschiedliche Bildschirmauflösungen registriert. Da die VERA-Seiten für 800 x 600 Pixel optimiert wurden, sind bei niedrigeren Auflösungen Teile der Seite wie z.B. Navigationselemente weder sicht- noch bedienbar. Entsprechend proble-

matische Bildschirmauflösungen wurden insgesamt 215-mal festgestellt, also in ca. 0,1 Prozent der Fälle.

Tabelle 5: registrierte Bildschirmauflösungen

	Häufigkeiten (Seitenaufrufe)		Prozent
Bildschirmauflösung	1024 x 768	136 778	67,4
	800 x 600	35 899	17,7
	1280 x 1024	18 418	9,1
	1152 x 864	5 032	2,5
	Gesamt	196 127	96,7

Gewissermaßen als Komplement zur Hardware stellt die Software die nichtphysischen Funktionsbestandteile eines Computers dar (Computerprogramme sowie die zur Verwendung mit Computerprogrammen bestimmten Daten). Dabei wird zwischen Systemsoftware und Anwendungssoftware unterschieden: Während Systemsoftware für das ordentliche Funktionieren des Computers erforderlich ist (v.a. Betriebssysteme), unterstützt Anwendungssoftware den Benutzer bei der Ausführung seiner Aufgaben (z.B. Internet-Browser).

Auch diese Informationen wurden von uns geloggt und zwar in Bezug auf den Aufruf der Seiten im geschützten Bereich von VERA. In Bezug auf die verwendeten Betriebssysteme konnten wir 18 unterschiedliche Programme identifizieren. Zu den ältesten Betriebssystemen gehören Windows 3.11 und OS/2. Da für diese Versionen keine aktuellen Browser verfügbar sind, kann das Webangebot nur sehr eingeschränkt benutzt werden. Diese Versionen wurden von uns jedoch nur 247-mal registriert.

Tabelle 6: Betriebssysteme

	Häufigkeiten (Seitenaufrufe)		Prozent
Betriebssysteme	Windows XP	9 711 158	51,0
	Windows 2000	4 179 498	21,9
	Windows 98	3 435 934	18,0
	Unix und Derivate	188 415	1,0
	MacIntosh	172 767	0,9
	Gesamt	17 687 772	92,8

In Bezug auf die Browser-Software haben wir 21 Programme unterschiedlicher Version registriert. Die häufigsten Browser können der Tabelle 7 entnommen werden. Für eine reibungslose internetgestützte Durchführung der Vergleichsarbeiten sind auch hier bestimmte Mindestanforderungen zu erfüllen. Ältere Browser wie z.B. Netscape 4.75 unterstützen aktuelle Webstandards nicht ausreichend und können demnach nur eine eingeschränkte Funktionalität des VERA-Webangebots gewährleisten. Entsprechende problematische Versionen lassen sich bei etwa 10,5 Prozent der Seitenaufrufe nach-

weisen. Mit anderen Worten bedeutet dies, dass über 89 Prozent der Seitenaufrufe mit einem ausreichend aktuellen Browser erfolgt sind.

Tabelle 7: Browser

	Häufigkeiten (Seitenaufrufe)	Prozent
Browser		
Internet Explorer	12 470 571	85,2
davon veraltete Versionen (3.0 bis 5.23)	1 509 956	10,4
Gecko: Netscape, Mozilla, Firefox	1 670 741	11,4
davon veraltete Versionen (Netscape 3.01 bis 4.79)	10 176	0,1
Sonstige	498 560	3,4

Aufrufe der Seiten im geschützten Bereich

Die Betrachtung der Nutzungsstatistiken unter zeitlicher Perspektive (in Abhängigkeit von Tageszeit und Monat) ist insbesondere aussagekräftig für den Unterstützungsbedarf im Rahmen der Durchführung der Vergleichsarbeiten. Die Abbildung 2 zeigt deutlich, dass praktisch den gesamten Tag und auch während der Nacht Aktivitäten im geschützten Bereich zu verzeichnen waren. Die Nutzung steigt gegen sieben Uhr morgens auf etwa fünf Prozent an und nimmt bis 13 Uhr zu (11,7 Prozent). Durch die Besetzung der Support-Hotline sowie des E-Mail-Supports von sieben Uhr bis 15 Uhr konnten etwa 68 Prozent der Aktivitäten bei Bedarf begleitet werden. Möglicherweise wäre es zu überlegen, den Support auf Beginn acht Uhr zu verlegen und bis 18 Uhr zu verlängern: So hätten etwa 87 Prozent der Aktivitäten abgedeckt werden können.

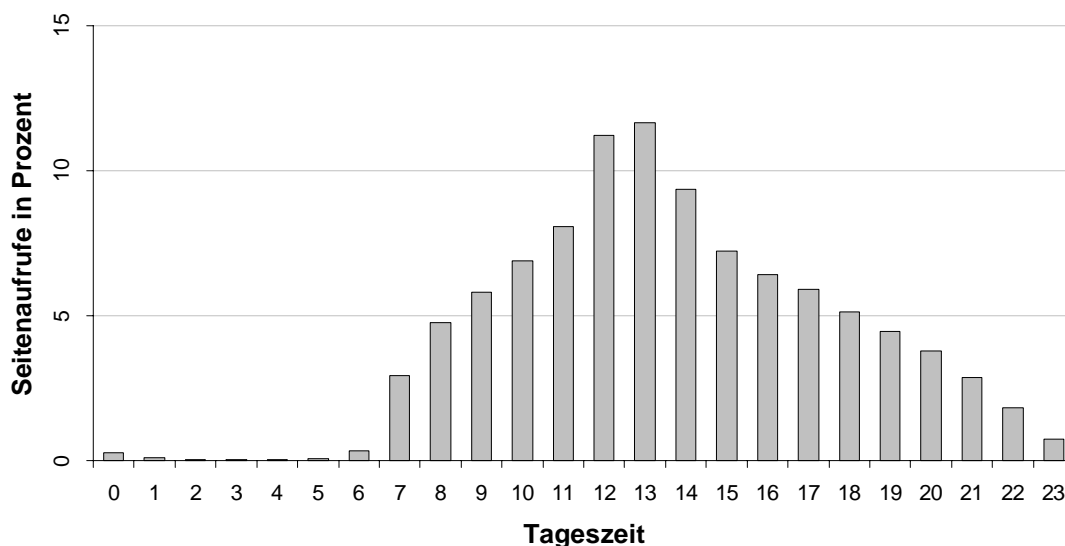


Abbildung 2: Häufigkeiten der Seitenaufrufe in Abhängigkeit von der Tageszeit

Ebenso aussagekräftig wie der Blick auf die Nutzung in Abhängigkeit von der Tageszeit sind die Nutzungsstatistiken im Verlauf der VERA-Durchführung. In Tabelle 8 zeigt

sich deutlich, dass sich die Aktivitäten im geschützten Bereich v. a. auf die Monate September, Oktober und November konzentrieren. Während dieser Zeit war der Telefon- und E-Mail-Support durch die Universität Landau voll besetzt und wurde erst im Lauf des Novembers allmählich reduziert.

Tabelle 8: Nutzungsstatistiken im VERA-Zyklus

	Aufgaben	Häufigkeiten (Seitenaufrufe)	Prozent
Monat August	<ul style="list-style-type: none"> • Anmeldung im geschützten Bereich • Probelauf in den Schulen 	211 120	3,2
September	<ul style="list-style-type: none"> • Eingabe der Schul- und Schülerdaten • Aufgabenauswahl Mathematik • Download der Testhefte und Korrekturanweisungen • Schreiben der Vergleichsarbeiten 	1 778 866	27,0
Oktober	<ul style="list-style-type: none"> • schulinterne Auswertung • Dateneingabe der Ergebnisse durch die Lehrkräfte • Klassen- und Schul-Ergebnisse 	1 871 574	28,4
November	<ul style="list-style-type: none"> • Befragung der Zentralstichprobe zum Kontext der Schule 	1 541 646	23,4
Dezember	<ul style="list-style-type: none"> • länderbezogene Ergebnisse und "fairer Vergleich" • Rückmeldung der Ergebnisse an die Eltern 	627 952	9,5
Januar	<ul style="list-style-type: none"> • innerschulische Auseinandersetzung mit den Ergebnissen 	432 205	6,5
Februar		136 576	2,1

Angesichts der Tatsache, dass 99 Prozent der teilnehmenden Schulen den VERA-Prozess vollständig durchlaufen haben, ist es über den zeitlichen Verlauf hinaus interessant, welche Seiten besonders häufig aufgerufen wurden. Denn für das Erreichen der dargestellten Ziele von VERA (vgl. 1.1, S. 4) sollen sich die Lehrkräfte und Schulen intensiv mit den Ergebnissen und zur Verfügung gestellten Informationsmaterialien auseinandersetzen. Im Folgenden werden daher für spezifische Seiten Nutzungsstatistiken berichtet.

Die folgende Tabelle 9 gibt den prozentualen Anteil der an VERA teilnehmenden Schulen wieder, die bestimmte Seiten im geschützten Bereich aufgerufen haben. Dazu wurden die Webseiten differenziert nach Inhaltsbereich (Allgemein, Mathematik, Deutsch), Rückmeldungszeitpunkt (Basis- vs. Zusatzinformationen) sowie nach spezifischem Inhalt. In den beiden letzten Tabellenspalten sind die landesspezifischen Nutzungsstatistiken den länderübergreifenden Prozentwerten gegenüber gestellt. Bei letzteren Werten ist zu berücksichtigen, dass es sich um einen gewichteten Wert handelt, d.h. dass die einzelnen Länderwerte gleich gewichtet eingehen. So wird ver-

mieden, dass Bundesländer mit größeren Schülerzahlen (wie z.B. Nordrhein-Westfalen) die länderübergreifenden Vergleichswerte dominieren. Gleichzeitig sind aufgefundene Unterschiede in den einzelnen Ländern eindeutiger zu interpretieren.

Zunächst einmal fällt auf, dass die Zahl der Schulen, die sich Ergebnisse ansehen, im zeitlichen Verlauf abnimmt. So werden die Basisinformationen von rund 80 bis 90 Prozent der Schulen aufgerufen – die Zusatzinformationen jedoch etwas seltener. Dabei ist wenig überraschend, dass die Seite mit den allgemeinen Materialien am häufigsten und mit einem Wert von rund 95 Prozent durch nahezu alle Schulen aufgerufen wurde. Auf dieser Seite finden sich u.a. Informationsmaterialien, die für den Ablauf der Vergleichsarbeiten grundlegend sind (Handreichung zur Durchführung der Vergleichsarbeiten, Erläuterungen zur Dateneingabe) sowie die Druckversion der Ergebnisse, Elternbrief und Elternrückmeldeformular.

Für die Schulen aus Nordrhein-Westfalen ist ein deutlicher Trend dahin gehend zu verzeichnen, dass praktisch alle Zusatzinformationen („fairer Vergleich“, Lösungshäufigkeiten, Diagnosegenauigkeit im Vergleich zum Land) sowie die fachspezifischen Informationsmaterialien deutlich häufiger als in der Gesamtstichprobe aufgerufen wurden. Dabei wurde insbesondere den Ergebnissen zum „faireren Vergleich“ (Mathematik: 84,9 Prozent; Deutsch: 76,8 Prozent) verstärkte Aufmerksamkeit gewidmet.

Alles in allem können also die Häufigkeiten für die Seitenaufrufe als erfreuliches Ergebnis gewertet werden. Denn neben der hohen Aktivität der nordrhein-westfälischen Schulen ist außerdem zu verzeichnen, dass insbesondere die klassenspezifischen Fähigkeitsniveauverteilungen, die wir als grundlegend für Schul- und Unterrichtsentwicklungsprozesse erachten, hier mit am häufigsten aufgerufen wurden.

Tabelle 9: Nutzungsstatistiken für den geschützten Bereich, Angaben in Prozent

			Nordrhein- Westfalen	Gesamt
			N = 3583	N = 6624
Allgemein		Informationsmaterialien	97,3	94,9
Mathematik	Basis	Fähigkeitsniveaus Klasse/Schule	96,3	91,8
		Fehleranalyse	88,6	83,8
		Zusatz	Fähigkeitsniveaus im Landesvergleich	85,6
		Fairer Vergleich	84,9	46,1
		Lösungshäufigkeiten	74,4	55,5
		Informationsmaterialien	86,9	77,2
Deutsch	Basis	Fähigkeitsniveaus Klasse/Schule	94,9	89,8
	Zusatz	Fähigkeitsniveaus im Landesvergleich	88,1	67,7
		Fairer Vergleich	76,8	36,8
		Lösungshäufigkeiten	47,8	36,9
		Informationsmaterialien	61,9	61,9

3.3.3 Ergebnisse zum Supportaufkommen

Zur Verfolgung der einzelnen Supportanfragen wurde im Jahr 2005 ein Ticketsystem eingesetzt. Die Wahl fiel hierbei auf OTRS (Open source Ticket Request System), das unter der GNU General Public License, GPL, vertrieben wird. Dieses System ist frei verfügbar und bietet durch das Webinterface die Möglichkeit, auch die bei den Ländersupports eingehenden Anfragen zentral zu verfolgen. Über entsprechende Konfigurationen konnte das System auf den hier vorliegenden Anwendungsfall angepasst werden.

Damit konnten sowohl E-Mail- als auch Telefonanfragen gemeinsam verwaltet werden. Die Abarbeitung der einzelnen Anfragen konnte durch die Zuordnung zu verschiedenen Warteschlangen prioritäts- und inhaltsbezogen gesteuert werden. Tabelle 10 gibt einerseits einen Überblick über das gesamte Supportaufkommen und andererseits über das aus Ihrem Bundesland.

Ein Ticket symbolisiert hierbei eine inhaltlich abgeschlossene Anfrage. Bei der Beantwortung und endgültigen Bearbeitung eines solchen Tickets kommt es teilweise zu mehreren Kontakten mit einer Schule. Eine Telefonanfrage entspricht dabei normalerweise einem Kontakt, Mailanfragen zwei Kontakten (Anfrage und Antwort). Vom Beginn des Probelaufs bis zum 15. März 2006 gingen insgesamt 3824 Anfragen (Tickets) ein. Die 1832 Anfragen aus Nordrhein-Westfalen kommen von 1161 verschiedenen Schulen. Der VERA-Support stand somit mit 32,8 Prozent der Schulen aus Nordrhein-Westfalen in Verbindung. Pro Anfrage ergaben sich dabei im Durchschnitt 1,8 Kontakte. 65,2 Prozent dieser Kontakte erfolgten per Mail, entsprechend 34,8 Prozent per Telefon. In Nordrhein-Westfalen ist der E-Mail-Anteil damit etwas höher als im Durchschnitt über die sieben Bundesländer (59,8 Prozent). Im Durchschnitt dauerte es 12,5 Stunden, bis eine Anfrage mit dem Status „erfolgreich beantwortet“ geschlossen werden konnte. Da in diese Berechnung auch die Wartezeiten von Tickets eingehen, die außerhalb der Supportzeiten eintrafen (z.B. abends oder am Wochenende), zeigt dies die schnelle Reaktion auf Anfragen aus den Schulen.

Bei der inhaltlichen Zuordnung ergibt sich folgendes Bild: 30 Prozent der Anfragen bezogen sich auf technische Fragestellungen, vier Prozent beinhalteten positives oder negatives Feedback und 66 Prozent waren allgemeiner Natur (Fragen zum Prozedere, zu den Aufgaben usw.).

Tabelle 10: Informationen über das Supportaufkommen

	Gesamt	Nordrhein-Westfalen
Anzahl der Tickets	3824	1832
Anzahl der Schulen mit Ticket %	35,4	32,8
Anzahl Kontakte	6631	3342
Kontakte pro Ticket	1,7	1,8
Mailkontakte %	59,8	65,2
Telefonkontakte %	40,2	34,8

3.4 Gesamt-Ergebnisse für die Fähigkeitsniveaus

3.4.1 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 Fähigkeitsniveaueverteilungen sinnvoll.

Gesamtverteilung 2005

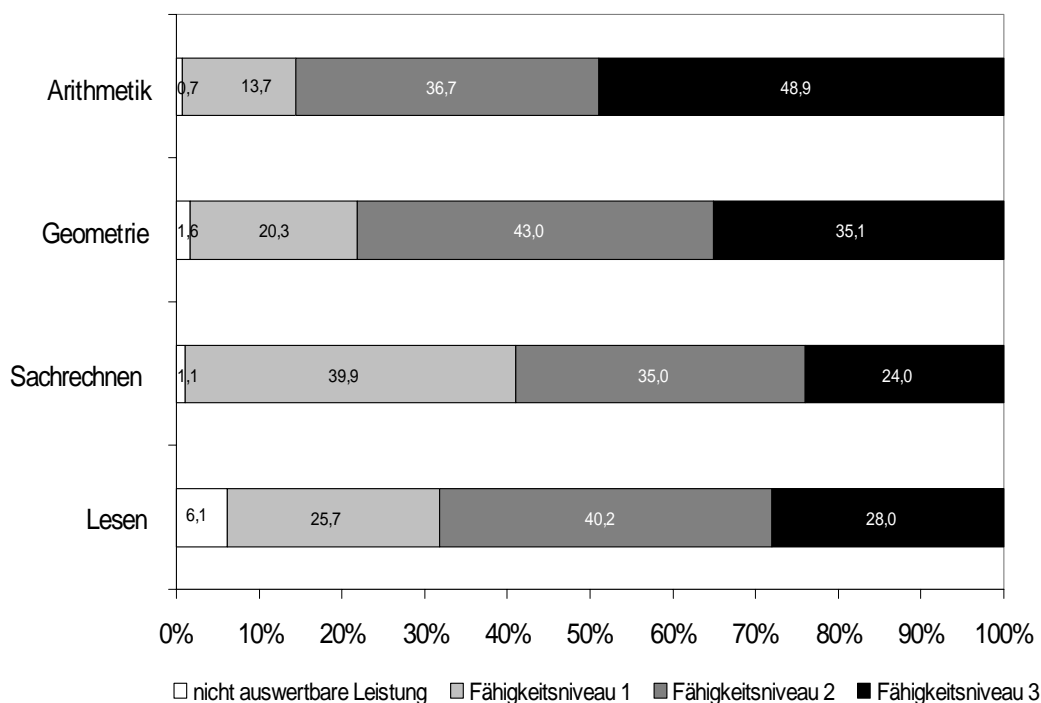


Abbildung 3: Gesamtverteilung der Fähigkeitsniveaus über alle Länder hinweg; Angaben in Prozent

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.

3.4.2 Zusammenhänge zwischen den Fähigkeitsniveaus

Die folgende Tabelle 11 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 Faches Mathematik. Die Korrelationen zwischen dem Leseverständnis und den drei Mathematik-Inhaltsbereichen könnten als Hinweis für die fächerübergreifende Bedeutung des Leseverständnisses interpretiert werden.

Tabelle 11: 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.

4 Landesspezifische Ergebnisse

4.1 Fähigkeitsniveaus

Im Folgenden werden die Fähigkeitsniveaus unter verschiedenen Gesichtspunkten diskutiert: Zunächst wird allgemein die Verteilung der Schülerinnen und Schüler auf den einzelnen Niveaus dargestellt (vgl. 4.1.1, S. 31) sowie Veränderungstrends für die Vergleichsarbeiten in den Jahren 2004 und 2005 diskutiert (vgl. 4.1.2, S. 32). Im Anschluss werden Unterschiede zwischen bzw. innerhalb den untersuchten Klassen/Schulen (siehe 4.1.3, S. 32), Geschlechtsunterschiede (siehe 4.1.4, S. 34) sowie Unterschiede zwischen Schülern mit Deutsch als dominanter vs. nicht dominanter Sprache (vgl. 4.1.5, S. 35) genauer beleuchtet.

4.1.1 Verteilung der Fähigkeitsniveaus in den Ländern

In den *Mathematik*-Inhaltsbereichen Arithmetik und Geometrie weist der Großteil der Schülerinnen und Schüler erweiterte bis fortgeschrittene Fähigkeiten auf (für Arithmetik 90,3 Prozent, für Geometrie 73,1 Prozent). Während in Arithmetik über die Hälfte dem Fähigkeitsniveau 3 zugeordnet wird, liegt in Geometrie der Anteil der Kinder im Bereich der fortgeschrittenen Fähigkeiten mit 30,2 Prozent deutlich niedriger. Im Sachrechnen erreichen verhältnismäßig weniger Kinder (61,2 Prozent) das zweite bzw. das dritte Fähigkeitsniveau und knapp unter 40 Prozent ein Niveau, das höchstens das Beherrschen elementarer Aufgaben umfasst (37,8 Prozent mit Fähigkeitsniveau 1; ein Prozent nicht auswertbare Leistung).

Im Vergleich zu der länderübergreifenden Verteilung fällt auf, dass in dem Inhaltsbereich Arithmetik vergleichsweise mehr Kinder das dritte Fähigkeitsniveau erreichen. Dem entgegengesetzt zeigen in Geometrie in Nordrhein-Westfalen weniger Kinder fortgeschrittene Fähigkeiten. Diesen Unterschieden liegen möglicherweise verschiedene landesspezifische Schwerpunktsetzungen im Grundschulunterricht zu Grunde.

Im Inhaltsbereich *Leseverständnis* zeigen 71,7 Prozent der Kinder erweiterte bis fortgeschrittene Fähigkeiten. 28,3 Prozent der Schülerinnen und Schülern erreichen höchstens das erste Fähigkeitsniveau (davon 4,6 Prozent nicht auswertbare Leistung).

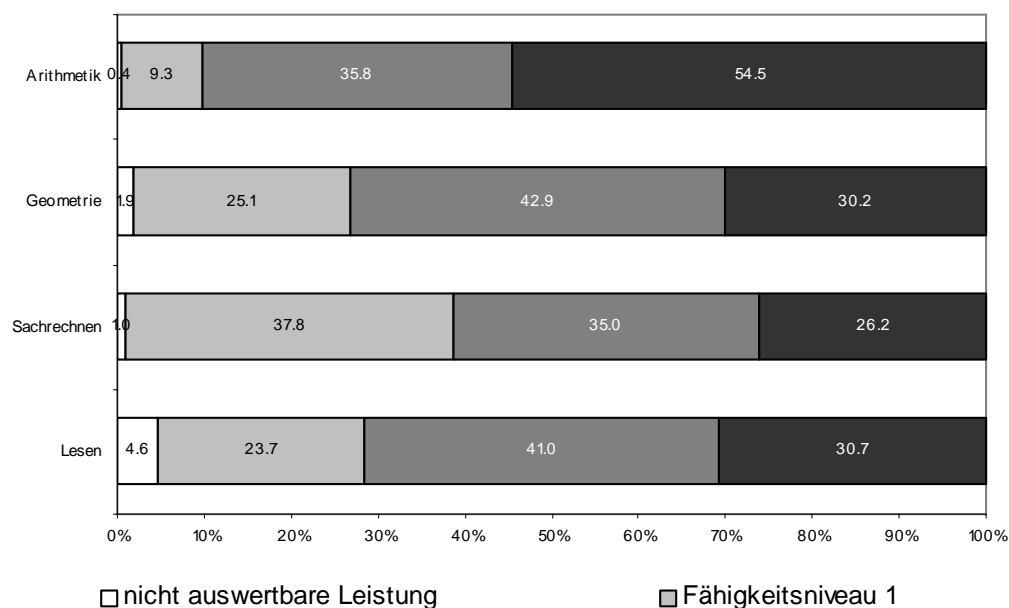


Abbildung 4: Gesamtverteilung der Fähigkeitsniveaus (2005); Angaben in Prozent

Im Vergleich zur länderübergreifenden Fähigkeitsniveaueverteilung fällt hier auf, dass in Nordrhein-Westfalen die Anzahl der nicht auswertbaren Leistungen etwas niedriger liegt.

Die häufige Besetzung des Fähigkeitsniveaus 1 in allen Inhaltsbereichen außer Arithmetik verdeutlicht, dass in diesen Gebieten nach wie vor Förderbedarf zu bestehen scheint.

4.1.2 Veränderungstrends 2004–2005

Die Verteilung der Schülerinnen und Schüler auf die einzelnen Fähigkeitsniveaubereiche ist in 2005 verglichen mit der Verteilung in 2004 (Abbildung 5) in allen vier Inhaltsbereichen in etwa gleich geblieben. Es ergeben sich in keinem der Inhaltsbereiche signifikante Unterschiede der Fähigkeitsniveaueverteilungen zwischen den Jahren.

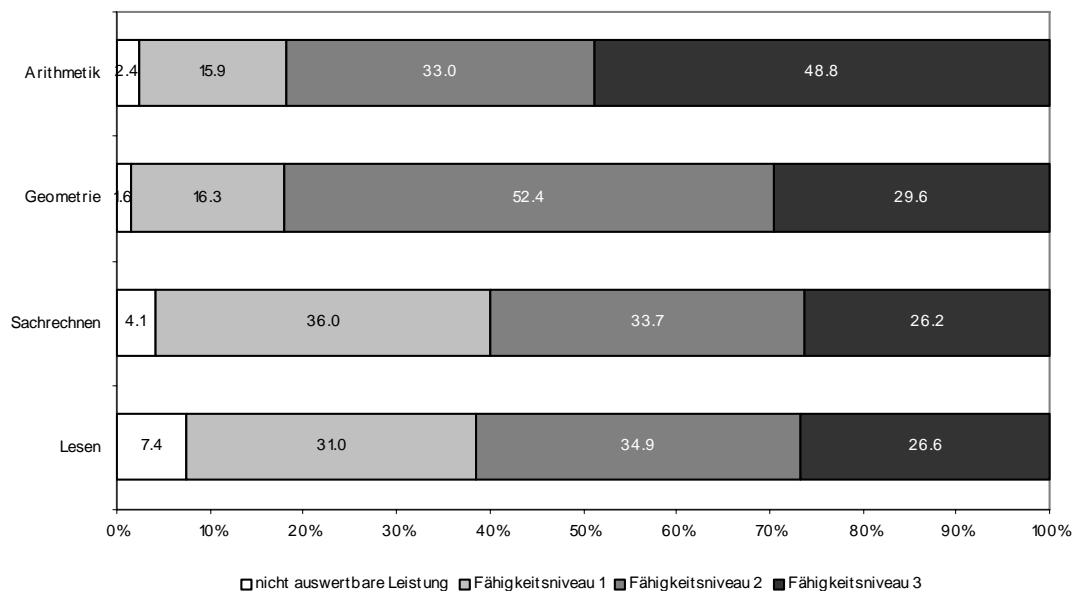


Abbildung 5: Gesamtverteilung der Fähigkeitsniveaus (2004); Angaben in Prozent

4.1.3 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 (vgl. 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 12 dargestellten Prozentanteile.

Table 12: Zerlegung in die Varianz auf den drei Ebenen Schule, Klasse und Individuum;
Angaben in Prozent

Bereich	Schulebene	Klassenebene	Individualebene
Arithmetik	16,6	7,1	76,3
Geometrie	25,3	9,8	64,9
Sachrechnen	13,4	5,5	81,1
Lesen	17,5	6,9	75,6
N min	3419	7926	172858

Wie erwartet ist die Leistung in den Vergleichsarbeiten vorrangig mit Merkmalen der Individualebene verknüpft. Die Varianzanteile liegen zwischen 64,9 Prozent (Geometrie) und 81,1 Prozent (Sachrechnen). Ebenfalls erhebliche Anteile gehen auf schulische Merkmale zurück, sie bewegen sich zwischen 13,4 Prozent (Sachrechnen) und 25,3 Prozent (Geometrie). Den vergleichsweise geringsten Beitrag leistet die Zugehörigkeit zu einer Schulklasse (hinter der Unterschiede des Unterrichts und des Klassenkontextes stehen) mit 5,5 Prozent (Sachrechnen) bis 9,8 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 sie sind 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 (vgl. Helmke, Hosenfeld & Schrader, 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 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 Unterschiede in den Kontextbedingungen (insbes. Einzugsgebiet, soziotopisches Profil) der Schulen. So

finden sich *innerhalb* der Schülerschaft 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 Klassen- und Schulebene fällt der relativ große Varianzanteil im Bereich Geometrie ins Auge. Dieser Effekt der Klassen- bzw. Schulzugehörigkeit könnte die curriculare „Stiefkindrolle“ der Geometrie widerspiegeln (vgl. 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 gerade 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.

4.1.4 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 (vgl. 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 13 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 standardisiert werden. Ein positiver d -Wert bedeutet eine Überlegenheit der Mädchen, ein negativer d -Wert umgekehrt eine Überlegenheit der Jungen.

* 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.

Table 13: Verteilung der Fähigkeitsniveaus, getrennt nach Geschlecht; Angaben in Prozent

		n.a.L.*	FN1	FN2	FN3	N (Kinder)	d**
Arithmetik	Mädchen	0,4	9,9	37,0	52,6	86086	-0,10
	Junge	0,3	8,6	34,6	56,4	87683	
Geometrie	Mädchen	1,9	26,8	43,2	28,1	86086	-0,11
	Junge	1,8	23,1	42,8	32,4	87683	
Sachrechnen	Mädchen	1,3	43,1	34,3	21,4	86086	-0,30
	Junge	0,7	31,4	36,8	31,1	87683	
Lesen	Mädchen	3,8	21,3	41,1	33,8	85666	0,17
	Junge	5,4	26,1	40,9	27,6	87152	

* 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 statistisch bedeutsamer Vorsprung ($d = -0,30$) der Jungen im Bereich Sachrechnen, - dort unterscheiden sich Mädchen und Jungen im höchsten Fähigkeitsniveau um einen Anteil von 9,7 Prozentpunkten. In Arithmetik und Geometrie zeigen sich immerhin noch geringe Unterschiede in den Leistungsverteilungen der Jungen und Mädchen ($d = -0,10$ bzw. $d = -0,11$).

Im Leseverständnis ist ebenfalls ein kleiner Unterschied ($d = 0,17$) zu verzeichnen, - hier schneiden die Mädchen besser ab.

4.1.5 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“ ($N_{\text{minimal}} = 152212$ bzw. 20606) 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 (vgl. 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.

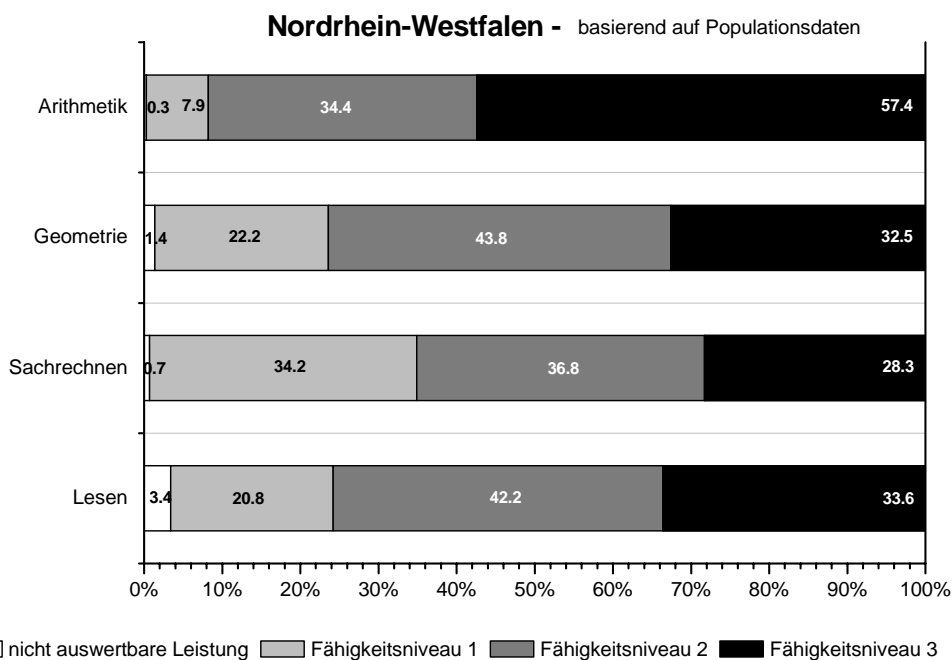


Abbildung 6: Gesamtverteilung der Fähigkeitsniveaus für Deutsch als dominante Sprache; Angaben in Prozent

In Abbildung 6 und Abbildung 7 sind die prozentualen Schülerleistungen jeweils nach Deutsch als dominante und nicht-dominante Sprache dargestellt. Es zeigen sich bedeutsame Unterschiede zwischen Kindern mit Deutsch als dominanter und Kindern mit Deutsch als nicht-dominanter Sprache sowohl in Mathematik ($d = 0,56$ bis $0,63$) als auch in Deutsch ($d = 0,85$). So sind Schülerinnen und Schüler mit Deutsch als nicht-dominanter Sprache in Mathematik vorwiegend im untersten und mittleren Fähigkeitsniveau zu finden (nicht auswertbare Leistung zwischen ein und fünf Prozent), im Leseverständnis zeigt sich ein ähnliches Bild. Ebenso wie im Sachrechnen erreicht auch hier über die Hälfte der Kinder nur ein Niveau, das höchstens das Beherrschen elementarer Aufgaben umfasst (für Sachrechnen 62,1 Prozent, davon 2,8 Prozent nicht auswertbare Leistung, für Leseverständnis 59,2 Prozent, davon 13,8 Prozent nicht auswertbare Leistung). Insbesondere im Leseverständnis kann infolge des hohen Anteils von Kindern mit nicht auswertbarer Leistung (knapp 14 Prozent) von einem substantiellen Unterschied zu Kindern, deren dominante Sprache Deutsch ist, gesprochen werden.

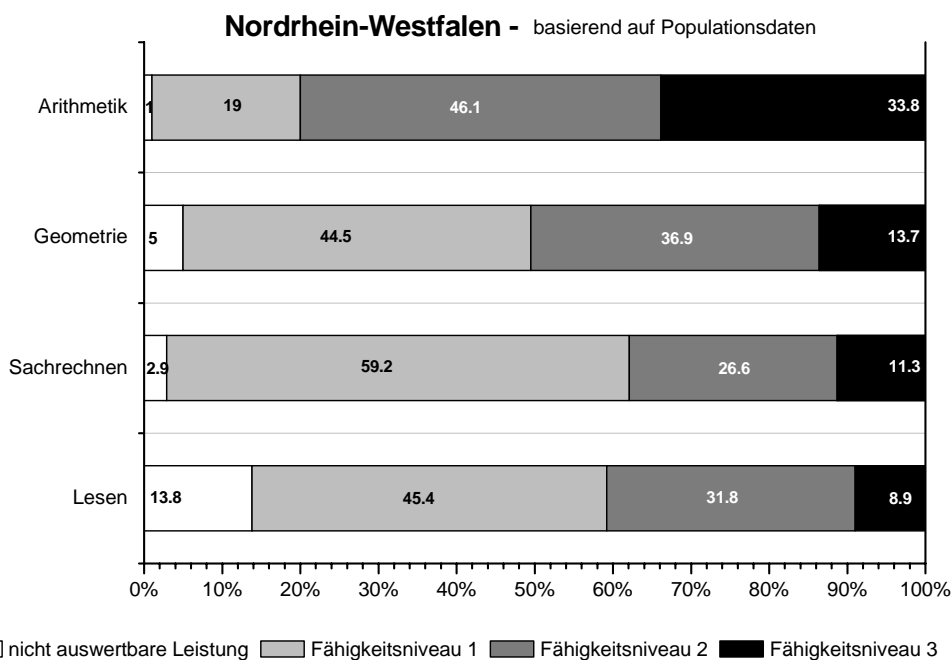


Abbildung 7: Gesamtverteilung der Fähigkeitsniveaus für Deutsch als nicht-dominante Sprache; Angaben in Prozent

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

Tabelle 14 zeigt die Effektstärken der Leistungsunterschiede.

Tabelle 14: Effektstärken der Leistungsunterschiede von Deutsch dominante vs. nicht-dominante Sprache

	Arithmetik	Geometrie	Sachrechnen	Lesen	N min (Kinder)
Effektstärke (d)	0,56	0,62	0,63	0,85	152212

4.2 „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.

4.2.1 Beschreibung ausgewählter Kontextgruppen

Es werden zunächst die auf Klassenebene (Ausnahme: Sprachdominanz) erfragten Daten zum sozioökonomischen und Sprachhintergrund berichtet (Tabelle 15). Anschließend folgt mit Tabelle 16 eine Darstellung der Zusammenhänge zwischen Kontextmerkmalen und Schülerleistungen. In Tabelle 17 sind die Schülervariablen dargestellt.

Tabelle 15: Kontextvariablen (Vergleich NRW – 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⁾
NRW	12,2	32,6	26,8	14,4	13,0
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 ist, beträgt in Nordrhein-Westfalen 12,2 Prozent. Dabei ergibt sich eine Korrelation mit der Schülerleistung im Leseverständnis von $r = -,35$.

Fast ein Drittel (32,6 Prozent) aller beteiligten Klassen liegen im so genannten „sozialen Brennpunkt“. Den Lehrerangaben zufolge sind 14,4 Prozent der Familien ihrer Schülerinnen und Schüler von Arbeitslosigkeit betroffen.

Table 16: Zusammenhänge zwischen Kontextmerkmalen und Schülerleistungen
(Schulklassenebene)

	Arithmetik	Geometrie	Sachrechnen	Lesen	min. N
Prozentsatz Grundschicht	-0,40	-0,37	-0,47	-0,47	4570
Prozentsatz Erhalt von Sozialhilfe	-0,34	-0,30	-0,40	-0,41	4572
Prozentsatz Arbeitslosigkeit in der Familie	-0,34	-0,31	-0,40	-0,40	4570
Prozentsatz deutsch nicht- dominant	-0,26	-0,23	-0,33	-0,35	7927

Bei den Zusammenhängen zwischen Kontextmerkmalen und Schülerleistungen wurden in allen Inhaltsbereichen insgesamt mittlere Korrelationen gefunden (zwischen $r = -,23$ und $r = -,47$).

Table 17: Schülervariablen (Vergleich NRW – Gesamtstichprobe)*; durchschnittliche Klassenanteile in Prozent

	ungenügende Sprachbe- herrschung ^{1,a)}	nicht an- wesend beim Deutschtest ¹⁾	nicht an- wesend beim Mathetest ¹⁾	Klassen- wiederholer ¹⁾
NRW	0,4	3,5	2,8	1,5
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.“

4.2.2 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 gewonnenen 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 verfügbarem 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 78 Prozent der Leistungsvarianz liegt innerhalb der Klassen, etwa 22 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 Nordrhein-Westfalen erwiesen sich folgende Variablen als signifikante Prädiktoren: Deutsch als nicht-dominante Sprache, Teilleistungsstörung Mathematik, Teilleistungsstörung Deutsch, Teilleistungsstörung Deutsch und Mathematik, Geschlecht, Anteil der Grundschichtfamilien, Schule im sozialen Brennpunkt und der Anteil von Kindern mit Teilleistungsstörungen in Mathematik. Dabei kehrt sich das Vorzeichen der kombinierten Teilleistungsstörung (Mathematik + Deutsch) um, d. h. Kindern mit Teilleistungsstörungen in Mathematik erzielen niedrigere Leistungen als Kinder ohne Teilleistungsstörungen. Gleiches gilt auch für Kinder mit Teilleistungsstörungen in Deutsch. Kinder mit kombinierten Teilleistungsstörungen sind zwar ebenfalls leistungsschwächer als

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.

Kinder ohne Teilleistungsstörungen, jedoch leistungstärker als sich aus der reinen Addition der Leistungsrückstände der Kinder mit einer Teilleistungsstörung ergeben würde. Ein ähnlicher Effekt ergibt sich auf der Ebene der Klasse: Für den Anteil der Kinder mit Teilleistungsstörungen im Fach Mathematik wird ein (schwach) positiver Koeffizient geschätzt, der inhaltlich jedoch nicht interpretiert werden darf, da er nicht auftritt, wenn das Individualmerkmal Teilleistungsstörung Mathematik nicht im Modell spezifiziert wird. Der geschätzte positive Koeffizient muss also als ein Korrekturfaktor verstanden werden, der verhindert, dass die erwartete Klassenleistung aufgrund der Aggregation von Leistungsrückständen von Kindern mit Teilleistungsstörungen unterschätzt wird. Hinsichtlich des Geschlechts ergibt sich, dass Jungen geringfügig höhere über die Fächer gemittelte Leistungen zeigen als Mädchen.

Insgesamt können mit diesem Modell gut zwölf Prozent der Leistungsvarianz auf Individualebene und knapp 26 Prozent der Variation zwischen Klassen erklärt werden. Erwartungsgemäß fällt der Anteil der erklärten Varianz auf Ebene der Schülerinnen und Schüler gering aus. Der erklärte Varianzanteil zwischen Klassen liegt auf dem Niveau wie in fast allen anderen Bundesländern und kann angesichts der verfügbaren Prädiktoren als zufrieden stellend gelten.

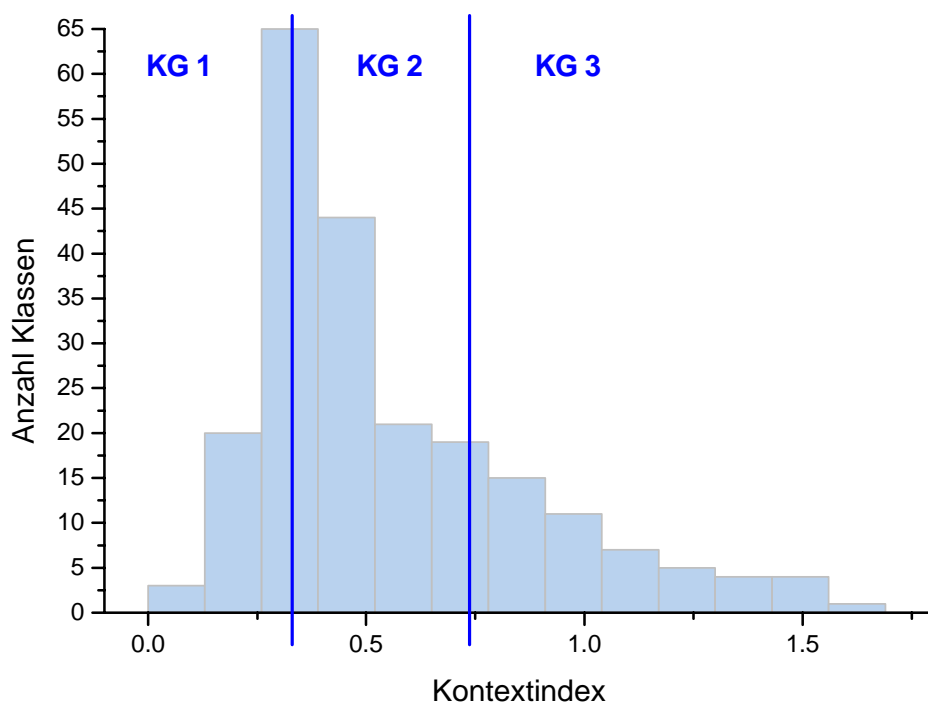


Abbildung 8: Verteilung der Kontextwerte inkl. der Grenzen zwischen den drei Kontextgruppen (KG 1 = günstige bis KG 3 = ungünstige Kontextbedingungen)

Im nächsten Schritt wurden diese Informationen gemäß der ermittelten Regressionsgewichte zu einem Kontextwert kombiniert. Abbildung 8 zeigt die Verteilung dieser Werte der nordrhein-westfälischen Zentralstichprobenklassen, für die vollständige Daten vorliegen und die daher berücksichtigt werden konnten (N = 206). Niedrige Werte stehen dabei für günstige, hohe Werte für ungünstige Kontextbedingungen.

Die Korrelationen dieses Kontextwertes mit den Leistungen in den verschiedenen Inhaltsbereichen sind in Tabelle 18 dargestellt. Erwartungsgemäß zeigen sich geringere Zusammenhänge zu den Leistungen im Bereich Arithmetik und Geometrie und engere Zusammenhänge mit den stärker sprachgebundenen Bereichen des Leseverständnisses und des Sachrechnens. Insgesamt fallen die Zusammenhänge zufrieden stellend aus.

Tabelle 18: Korrelation zwischen der mittleren Klassenleistung und dem Kontextwert in der nordrhein-westfälischen Zentralstichprobe*

Mathematik	r	Deutsch	r
Arithmetik	-0,32	Leseverständnis	-0,51
Geometrie	-0,32		
Sachrechnen/Größen	-0,45		

* Stichprobengröße: N = 206 (Klassen mit vollständigen Daten)

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 19 gibt die Durchschnittswerte zentraler Kontextmerkmale der drei Gruppen wieder.

Tabelle 19: Beschreibung der drei Kontextgruppen anhand durchschnittlicher Merkmalsausprägungen; Angaben in Prozent

	Kontextgruppe		
	1	2	3
Anteil der Kinder mit Deutsch als nicht-dominanter Sprache	~2	~7	~31
Anteil der Klassen im sozialen Brennpunkt	~0	~10	~90
Anteil der Kinder aus Familien der Grundschrift	~9	~26	~55

4.2.3 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 20 sind die Verteilungen der Zentralstichprobe auf die Fähigkeitsniveaus in Mathematik und Deutsch nach den drei Kontextgruppen differenziert dargestellt.

Table 20: Verteilung der Fähigkeitsniveaus, aufgeschlüsselt nach Kontextgruppen; Angaben in Prozent

	Kontextgruppe	n.a.L.*	FN 1	FN 2	FN 3	N (Klassen)
Arithmetik	1	0,1	9,9	32,9	57,2	51
	2	0,2	8,2	37,0	54,6	104
	3	0,2	17,1	38,3	44,3	51
Geometrie	1	1,5	20,4	42,8	35,3	51
	2	1,8	25,3	43,8	29,2	104
	3	5,3	34,3	39,4	20,9	51
Sachrechnen	1	0,4	31,8	36,5	31,3	51
	2	0,7	36,9	35,7	26,7	104
	3	1,6	48,0	32,1	18,3	51
Lesen	1	2,1	20,2	40,7	37,0	51
	2	4,2	22,4	43,7	29,7	104
	3	9,5	35,0	36,3	19,2	51

* 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.

4.2.4 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 Nordrhein-Westfalen bis Anfang März 56,3 Prozent (N = 4419) der Klassen Gebrauch⁶. Dabei ergaben sich hinsichtlich der Verteilung auf die drei Kontextgruppen nur geringe Abweichungen von der Zentralstichprobe. Die mittlere Gruppe ist mit 43,2 Prozent der Klassen schwächer besetzt, während sowohl die günstige Gruppe 1 (27,1 Prozent) als auch die ungünstige Gruppe 3 (29,7 Prozent) geringfügig stärker vertreten sind.

⁶ Die oben bereits dargestellten 206 Klassen der Zentralstichprobe mit vollständigen Daten sind bei dieser Darstellung ausgeklammert.

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 kein signifikanter Unterschied in der gemittelten Leistung zwischen diesen beiden Gruppen existiert.

4.2.5 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. Zu diesem Zweck wurden die Klassen, für die entsprechende Daten vorlagen ($N = 4625$), 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 8053 nordrhein-westfälischen 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).

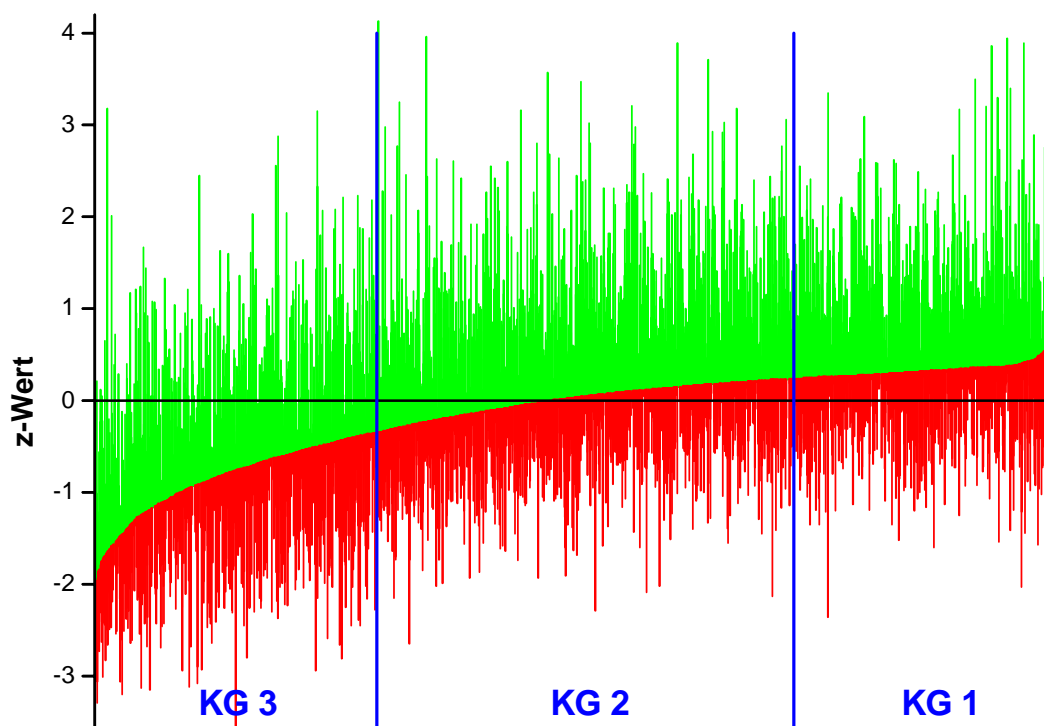


Abbildung 9: Differenzen zwischen erwarteter und tatsächlicher über beide Fächer gemittelter Leistung ($N = 4625$ Klassen mit vollständigen Kontextdaten); KG = Kontextgruppe

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 $-0,4$ bis $+0,7$ Standardabweichungen. Die schwächsten zehn Prozent unterschreiten ihren Erwartungswert jeweils um mindestens $0,9$ Standardabweichungen, die leistungsstärksten zehn Prozent erzielen gemittelte Leistungen, die jeweils mindestens $1,2$ Standardabweichungen über ihrem Erwartungswert liegen.

4.3 Lehrerfragebogen

Der Lehrerfragebogen sollte in Nordrhein-Westfalen für 222 Klassen in 116 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 numerische Werte, so sind die Angaben weitgehend vollständig, denn es werden 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 21 dargestellten Werte.

Tabelle 21: Vollständigkeit des Rücklaufs des Lehrerfragebogens; Angaben in Prozent

	absolut vollständig	relativ vollständig	4 zentrale Angaben
komplett	88,7	93,7	94,1
unvollständig	11,3	6,3	5,9

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.

4.3.1 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, zeigt über die verschiedenen Inhaltsbereiche kein gleichmäßiges Muster.

Betrachtet man die Frage nach der Zahl der durch die einzelnen Lehrkräfte unterrichteten Halbjahre, so können ebenfalls keine systematischen Zusammenhänge gefunden werden. Hier finden sich zusätzlich teilweise geringe Gruppengrößen, weshalb auf eine weiterreichende Interpretation verzichtet werden sollte.

Tabelle 22: Lehrerwechsel während der ersten vier Grundschuljahre in den Fächern Mathematik und Deutsch; Angaben in Prozent

Bereich		n.a.L.*	FN 1	FN 2	FN 3	N (Klassen)
Arithmetik	keine Regel	0,1	8,8	36,2	54,9	47
	kein Lehrerwechsel	0,2	11,4	34,9	53,6	109
	Lehrerwechsel nach der 2. Klasse	0,4	11,4	39,2	49,0	35
	Lehrerwechsel nach der 3. Klasse	-	-	-	-	0
	eine andere Regel	0,2	12,0	40,8	47,0	25
Geometrie	keine Regel	1,0	22,2	43,1	33,6	47
	kein Lehrerwechsel	3,3	27,0	40,9	28,9	109
	Lehrerwechsel nach der 2. Klasse	1,4	24,8	43,4	30,4	35
	Lehrerwechsel nach der 3. Klasse	-	-	-	-	0
	eine andere Regel	3,3	32,7	46,0	18,0	25
Sachrechnen	keine Regel	1,0	33,2	35,3	30,5	47
	kein Lehrerwechsel	0,7	40,9	33,4	24,9	109
	Lehrerwechsel nach der 2. Klasse	1,0	38,4	36,4	24,2	35
	Lehrerwechsel nach der 3. Klasse	-	-	-	-	0
	eine andere Regel	1,1	38,8	39,4	20,7	25
Lesen	keine Regel	3,3	20,5	43,4	32,8	47
	kein Lehrerwechsel	5,5	25,9	40,3	28,3	109
	Lehrerwechsel nach der 2. Klasse	4,5	23,2	40,9	31,4	35
	Lehrerwechsel nach der 3. Klasse	-	-	-	-	0
	eine andere Regel	6,8	29,6	41,6	22,0	25

* nicht auswertbare Leistung

Table 23: 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	0,1	10,1	34,5	55,3	97
	Mitte 1. Klasse	0,0	36,4	31,8	31,8	1
	Beginn 2. Klasse	0,0	7,2	37,8	55,0	12
	Mitte 2. Klasse	0,0	13,4	32,5	54,1	5
	Beginn 3. Klasse	0,3	10,7	38,6	50,4	56
	Mitte 3. Klasse	0,0	13,4	41,0	45,7	14
	Beginn 4. Klasse	0,4	13,9	38,4	47,3	29
Geometrie	Beginn 1. Klasse	2,7	25,2	40,8	31,2	97
	Mitte 1. Klasse	13,6	50,0	27,3	9,1	1
	Beginn 2. Klasse	2,9	27,8	38,5	30,8	12
	Mitte 2. Klasse	1,8	22,1	41,6	34,5	5
	Beginn 3. Klasse	2,1	24,7	44,5	28,7	56
	Mitte 3. Klasse	1,1	28,5	42,0	28,4	14
	Beginn 4. Klasse	3,1	31,4	45,2	20,3	29
Sachrechnen	Beginn 1. Klasse	0,7	38,7	33,0	27,6	97
	Mitte 1. Klasse	0,0	81,8	13,6	4,5	1
	Beginn 2. Klasse	1,3	33,1	32,9	32,7	12
	Mitte 2. Klasse	0,0	33,2	38,4	28,4	5
	Beginn 3. Klasse	1,0	36,2	37,3	25,5	56
	Mitte 3. Klasse	0,4	46,5	37,2	16,0	14
	Beginn 4. Klasse	1,4	41,6	36,1	20,8	29
Lesen	Beginn 1. Klasse	5,2	24,7	40,8	29,3	133
	Mitte 1. Klasse	7,1	36,9	34,5	21,4	2
	Beginn 2. Klasse	6,8	25,4	42,1	25,7	12
	Mitte 2. Klasse	2,2	14,4	34,1	49,3	2
	Beginn 3. Klasse	3,8	21,7	42,5	32,0	42
	Mitte 3. Klasse	6,6	32,5	44,4	16,5	12
	Beginn 4. Klasse	3,2	26,4	40,7	29,7	13

* nicht auswertbare Leistung

4.3.2 Unterrichtserfahrung, grundständige Ausbildung und Fähigkeitsniveaus

Aufgrund der verschiedenen in den Tabellen berücksichtigten Merkmalsebenen (z. B. Fach, Inhaltsbereich 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, Hosenfeld & Schrader, 2002) keine systematischen, insbesondere keine linearen (vom Typ „je...desto“) Unterschiede in Abhängigkeit von der Unterrichtserfahrung sowie der grundständigen Ausbildung ergeben, werden die entsprechenden Fähigkeitsniveaueverteilungen der Übersicht halber im Folgenden wiedergegeben, ohne sie im Detail zu diskutieren.

Table 24: Durch die Lehrkraft unterrichtete Jahre in den Fächern Mathematik und Deutsch; Angaben in Prozent

Bereich		n.a.L.*	FN 1	FN 2	FN 3	N (Klassen)
Arithmetik	bis 2 Jahre	0,0	14,3	41,7	44,0	10
	2 bis 5 Jahre	0,9	11,4	37,7	50,0	15
	6 bis 10 Jahre	0,5	13,8	37,3	48,3	32
	11 bis 15 Jahre	0,0	8,9	32,3	58,9	29
	16 bis 20 Jahre	0,0	7,5	39,6	52,9	13
	21 bis 25 Jahre	0,0	15,6	35,7	48,7	17
	26 bis 30 Jahre	0,2	10,7	36,6	52,5	56
	31 bis 35 Jahre	0,1	8,7	36,9	54,3	29
	36 bis 40 Jahre	0,0	8,9	36,5	54,6	13
	mehr als 40 Jahre	0,0	8,7	43,5	47,8	1
Geometrie	bis 2 Jahre	1,5	21,1	50,8	26,6	10
	2 bis 5 Jahre	1,2	29,1	35,7	33,9	15
	6 bis 10 Jahre	2,7	31,0	44,1	22,2	32
	11 bis 15 Jahre	0,7	21,2	42,9	35,2	29
	16 bis 20 Jahre	1,8	23,0	43,3	31,9	13
	21 bis 25 Jahre	4,4	29,7	42,9	23,1	17
	26 bis 30 Jahre	4,2	28,0	40,7	27,1	56
	31 bis 35 Jahre	1,2	20,6	42,9	35,3	29
	36 bis 40 Jahre	2,1	32,3	43,4	22,3	13
	mehr als 40 Jahre	0,0	21,7	30,4	47,8	1

Bereich		n.a.L.*	FN 1	FN 2	FN 3	N (Klassen)
Sachrechnen	bis 2 Jahre	1,4	42,8	35,8	20,0	10
	2 bis 5 Jahre	0,7	37,5	33,1	28,8	15
	6 bis 10 Jahre	1,7	40,0	36,9	21,4	32
	11 bis 15 Jahre	1,0	33,6	34,6	30,9	29
	16 bis 20 Jahre	0,7	45,6	33,0	20,6	13
	21 bis 25 Jahre	0,6	43,2	31,5	24,7	17
	26 bis 30 Jahre	0,5	37,5	35,2	26,7	56
	31 bis 35 Jahre	0,7	37,3	35,8	26,2	29
	36 bis 40 Jahre	0,7	41,1	35,9	22,4	13
	mehr als 40 Jahre	0,0	26,1	43,5	30,4	1
Lesen	bis 2 Jahre	4,0	32,2	42,7	21,1	7
	2 bis 5 Jahre	5,2	30,6	41,6	22,7	15
	6 bis 10 Jahre	3,7	24,0	41,3	31,1	30
	11 bis 15 Jahre	5,7	21,4	41,9	31,1	32
	16 bis 20 Jahre	5,4	25,5	45,3	23,8	16
	21 bis 25 Jahre	3,7	22,3	42,3	31,7	19
	26 bis 30 Jahre	6,3	27,5	39,7	26,5	48
	31 bis 35 Jahre	4,4	21,2	41,5	32,9	33
	36 bis 40 Jahre	5,5	26,3	38,6	29,5	13
	mehr als 40 Jahre	1,5	20,0	34,8	43,7	3

* nicht auswertbare Leistung

Tabelle 25: Mathematik und Deutsch als grundständiges Fach; Angaben in Prozent

Bereich		n.a.L.*	FN 1	FN 2	FN 3	N (Klassen)
Arithmetik	nein	0,1	11,5	36,5	52,0	69
	ja	0,3	10,7	36,7	52,3	146
Geometrie	nein	3,4	26,4	40,9	29,4	69
	ja	2,1	26,3	43,0	28,5	146
Sachrechnen	nein	0,9	40,7	34,1	24,4	69
	ja	0,9	37,7	35,4	26,0	146
Lesen	nein	5,0	23,0	43,4	28,6	60
	ja	5,0	25,4	40,4	29,2	156

* nicht auswertbare Leistung

4.3.3 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 häufig eingesetzte Vorbereitungsformen zu erfassen, wurde eine entsprechende Frage in den Zentralstichprobenfragebogen aufgenommen. Die folgende Tabelle 26 stellt die Häufigkeiten unterschiedlicher Formen der Vorbereitung dar.

Tabelle 26: Formen der Vorbereitung; Angaben in Prozent

		Häufigkeiten N = 216*	Prozent
Vertrautheit mit dem Testformat	Ablauf	182	84,3
	Aufgabenformate	41	19,0
	Korrekturkriterien	3	1,4
Inhaltliche Vorbereitung	Aufgaben 2004	22	10,2
	VERA-Inhalte	25	11,6
Teststrategien	Teststrategien	64	29,6

*Mehrfachantworten waren bei dieser Frage zugelassen.

Dabei scheint in Nordrhein-Westfalen 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 (84,3 Prozent). Dieses Ergebnis korrespondiert mit den Aufrufstatistiken für die

Webseite mit den allgemeinen Materialien (z.B. Handreichung zur Durchführung der Vergleichsarbeiten, Erläuterungen zur Dateneingabe, vgl. 3.3.2, S. 23). 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 ihren Klassen auf diesem Wege das Entstehen von Angst ersparen wollten.

Ebenfalls häufig wurde angegeben, die Schülerinnen und Schüler in Bezug auf Teststrategien (29,6 Prozent) und Aufgabenformate (19,0 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.

4.3.4 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 Kolleginnen und Kollegen 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 richtet, wurde eine Kooperation durch das Prozedere der Aufgabenauswahl (pro Schule nur eine Aufgabenkombination erlaubt) und der Vergleichsangebote in den Ergebnisrückmeldungen forciert. Die Antworten der Lehrkräfte zu diesen Fragen können als Hinweis zu den schulischen Bedingungen, insbesondere dem Evaluations- und Kooperationsklima interpretiert werden. Tabelle 27 belegt, dass bei der *Aufgabenauswahl* relativ oft kooperiert wurde. Demgegenüber erfolgte die *schulinterne Auswertung* deutlich seltener in Zusammenarbeit mit Kolleg/innen.

Tabelle 27: Kooperation im Rahmen der Vergleichsarbeiten; Angaben in Prozent

		ja	nein	N (Klassen)
Aufgabenauswahl im Team	Mathematik	76,7	23,3	222
Auswertung im Team	Deutsch	24,9	75,1	222
	Mathematik	28,0	72,0	222

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 22,0 Prozent auf 28,0 Prozent, in Deutsch reduziert sich die Zahl der Kooperationen leicht von 27,4 auf 24,9 Prozent.

5 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 Ergebnisrückmeldungen, 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 zwei Bundesländern 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.