



**Ergebnisbericht  
VERA 2004:**

**Nordrhein-Westfalen**

Andreas Helmke, Ingmar Hosenfeld, Jana Groß Ophoff, Jan Hochweber, Kevin Isaac,  
Ursula Koch, Frank Scherthan und Sonja Wagner

*Universität Koblenz Landau, Campus Landau*

Stand: 03.06.2005

# Inhalt

<b>1 Einleitung und Überblick</b> .....	<b>4</b>
1.1 Ziele .....	4
1.2 Organisation .....	4
<b>2 Aufgabenentwicklung und Definition der Fähigkeitsniveaus</b> .....	<b>5</b>
2.1 Deutsch .....	6
2.1.1 Aufgabenentwicklung .....	6
2.1.2 Skalierung .....	7
2.1.3 Fähigkeitsniveaus .....	8
2.2 Mathematik .....	13
2.2.1 Aufgabenentwicklung .....	13
2.2.2 Skalierung und Fähigkeitsniveaus des Mathematiktests .....	15
<b>3 Vergleichsarbeiten im September 2004</b> .....	<b>18</b>
3.1 Ablauf und Stichproben .....	18
3.2 Elemente der Ergebnismeldungen .....	19
3.2.1 Welle 1 .....	19
3.2.2 Welle 2 .....	20
3.3 Bildung von Kontextgruppen .....	20
3.4 Gesamt-Ergebnisse .....	23
3.4.1 Verteilung der Fähigkeitsniveaus .....	23
3.4.2 Fähigkeitsniveaus nach Kontextgruppen .....	24
3.4.3 Zusammenhänge zwischen den Fähigkeitsbereichen .....	26
<b>4 Landesspezifische Ergebnisse</b> .....	<b>27</b>
4.1 Beschreibung der Leistungen .....	27
4.1.1 Verteilung der Fähigkeitsniveaus .....	27
4.1.2 Gegenüberstellung: Wahl- vs. Zentralteil .....	28
4.1.3 Unterschiede innerhalb und zwischen Klassen bzw. Schulen .....	30
4.1.4 Geschlechterunterschiede .....	31
4.1.5 Unterschiede zwischen Deutsch als dominante vs. nicht-dominante Sprache .....	33
4.1.6 Diagnosegenauigkeit im Fach Mathematik .....	35

---

4.2	Rahmenbedingungen.....	38
4.2.1	<i>Verteilung ausgewählter Kontextmerkmale.....</i>	38
4.2.2	<i>Verteilung der Zentralstichprobe auf die Kontextgruppen.....</i>	40
4.3	Lehrerfragebogen (ausgewählte Ergebnisse).....	40
4.3.1	<i>Kontinuität des Unterrichts in Mathematik und Deutsch und Fähigkeitsniveaus ...</i>	40
4.3.2	<i>Unterrichtserfahrung, grundständige Ausbildung und Fähigkeitsniveaus.....</i>	43
4.3.3	<i>Kooperation bei der Aufgabenauswahl und -auswertung.....</i>	45
<b>5</b>	<b>Ausblick.....</b>	<b>46</b>
<b>6</b>	<b>Literatur.....</b>	<b>47</b>
<b>7</b>	<b>Anhang.....</b>	<b>49</b>

# Länderbericht VERA 2004

## 1 Einleitung und Überblick

Am 28. und 30. September 2004 schrieben im Rahmen des Projektes VERA etwa 300 000 Schülerinnen und Schüler der 4. 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 Beteiligung der Schulen an der Auswahl und Auswertung der Aufgaben 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 Ergebnismrückmeldung 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 2004 nahmen in Berlin, Bremen, Mecklenburg-Vorpommern, Nordrhein-Westfalen, Rheinland-Pfalz und Schleswig-Holstein alle Grundschulen verpflichtend an der Untersuchung teil. In Brandenburg nahmen insgesamt 188 Schulen am Projekt teil, wobei 74 zufällig gezogen und zur Teilnahme verpflichtet wurden, während sich die restlichen Schulen freiwillig beteiligten.

Darüber hinaus nahmen etliche deutsche Auslandsschulen mit 4. Klassen aus allen fünf Erdteilen das Angebot zur Teilnahme an VERA an. Abbildung 1 verdeutlicht die Vernetzung verschiedener Funktionen und Gruppen.

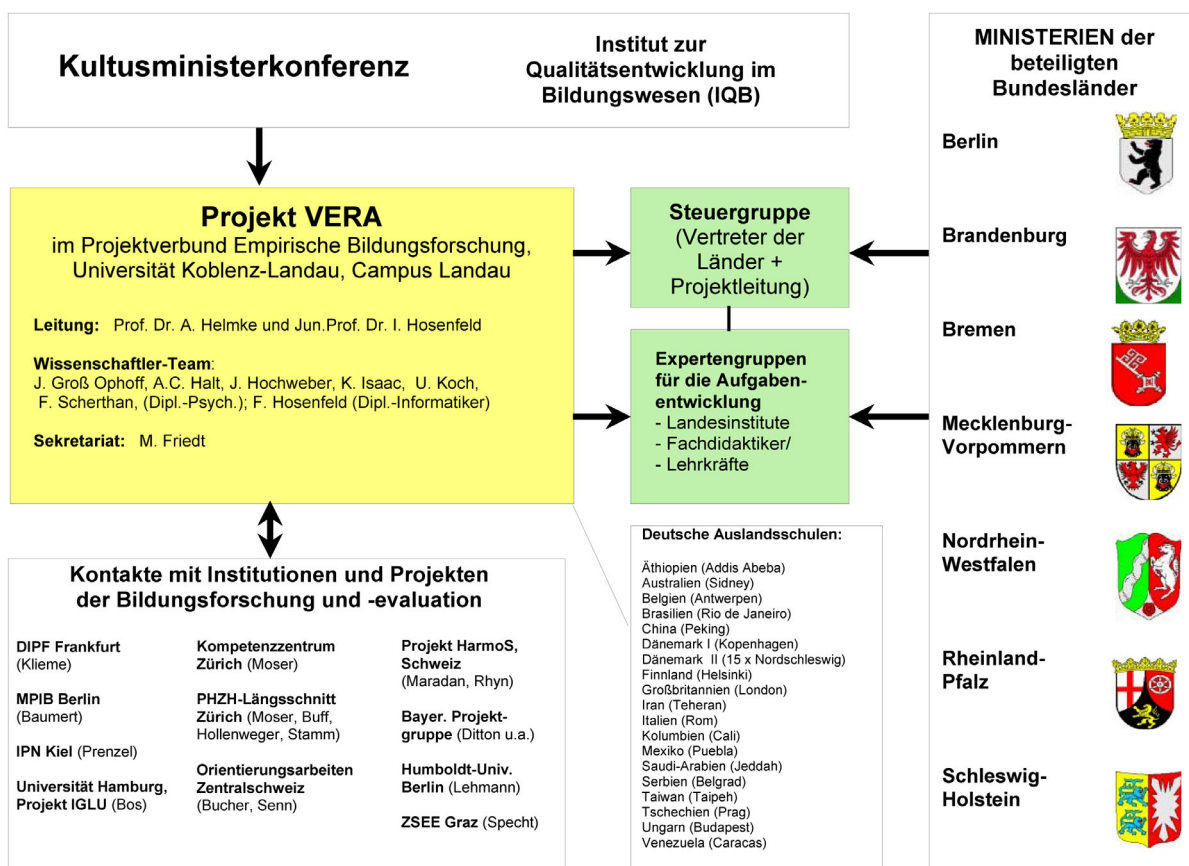


Abbildung 1: Bildungspolitischer Kontext von VERA

## 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 umfangreich empirisch untersucht. Dabei konnte 2004 für Mathematik auf die bereits zur Vergleichsarbeit 2003 geleistete Entwicklungsarbeit zurückgegriffen werden, während das Fach Deutsch 2004 erstmalig im Rahmen von VERA untersucht wurde. Die empirischen Vorerhebungen Ende 2003 (in Deutsch ca. 13 000 Schülerinnen und Schüler, in Mathematik ca. 6 000 Schülerinnen und Schüler) dienen der Realisierung folgender Ziele:

- Erprobung der von Didaktikern und Lehrkräften entwickelten Aufgaben
- Untersuchung der Art und Häufigkeit von Fehlern (Falschlösungen) der Aufgaben
- Entwicklung von Fähigkeitsniveau-Definitionen

- 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 fünf (Deutsch) bzw. zwei (Mathematik) Bundesländern gezogen und jeweils mit allen vierten Klassen in die Untersuchung einbezogen. Innerhalb jeder Klasse kamen je Fach zehn verschiedene Testhefte zum Einsatz. Jedes der zehn 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 100 Aufgaben untersucht werden, wobei jedem einzelnen Kind jedoch nur eine überschaubare Anzahl von Aufgaben vorgelegt wurde. Zur Bearbeitung der Aufgaben standen in Deutsch 90 Minuten (unterbrochen von 10 Minuten Pause) und Mathematik 50 Minuten Bearbeitungszeit zur Verfügung, alle Aufgabenbearbeitungen erfolgten direkt im Testheft. 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 psychologischen und psychometrischen Experten. Die Entwicklung begann im Frühjahr 2003 und orientierte sich während des gesamten Entwicklungsprozesses auch an der jeweils aktuellen Version Bildungsstandards. In diesem Zusammenhang sei vermerkt, dass einige der Aufgabenentwickler/innen auch in den Arbeitsgruppen der KMK vertreten waren. Da die technischen Voraussetzungen zur Erfassung des Hörverständnisses im flächendeckenden Einsatz noch nicht gegeben sind, erfolgte eine Beschränkung der Aufgabenentwicklung auf vier Inhaltsbereiche. Tabelle 1 gibt die Grob- und Fein-Klassifikation der Inhaltsbereiche wieder.

Die Zugrundelegung dieses Rasters gewährleistete, dass die Bereiche des Curriculums in hinreichender Breite im Aufgabenpool repräsentiert sind. Darüber hinaus wurde eine Vielfalt der *Aufgabenformate* (Multiple-Choice, offene Antworten unterschiedlicher Länge, Markierungen, Zuordnungen, Sortierungen) und eine Variation der *Schwierigkeiten* angestrebt. Dazu wurden alle Aufgaben hinsichtlich einer Reihe von bereichsspezifischen schwierigkeitsbestimmenden Merkmalen beurteilt, z.B. Textlänge, Vertrautheit der verwendeten Wörter etc. bei Aufgaben zum Leseverständnis.

Zunächst wurde eine umfangreiche Sammlung von Aufgaben (> 300) 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. auch alle Aufgaben, die besondere Hilfsmittel wie beispielsweise ein Wörterbuch erfordern, ausgeschlos-

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

**Tabelle 1:** Grob- und Feinklassifikation der bei der Aufgabenentwicklung berücksichtigten Inhaltsbereiche im Fach Deutsch

Grobklassifikation	Feinklassifikation
Leseverständnis	fiktionale Texte kontinuierliche Sachtexte diskontinuierliche Sachtexte
Schreiben	Intentionen (appellativ, expressiv, informativ) Darstellungsarten (Beschreiben/Berichten, Erzählen, Erörtern)
Sprachbetrachtung	Wortschatz: lexikalische Bedeutungen Wortschatz: Bedeutungsbeziehungen Syntax: Wortarten (inkl. Tempus) Syntax: Satzglieder Syntax: Satzbau (Zeichensetzung)
Orthographie	Korrektur vorgegebener Texte (Fehlschreibungen) Diktat (Satzdiktat, Wort-Diktat) Alphabetische Ordnungsaufgaben*

\* Da nicht sichergestellt werden kann, dass alle Kinder ein Wörterbuch zur Verfügung haben, entfallen Aufgaben, die unmittelbar auf die Verwendung des Wörterbuchs zielen. Aufgaben, in denen Wörter alphabetisch sortiert werden sollen, erfassen allerdings eine für die Nutzung des Wörterbuchs notwendige Vorläuferfähigkeit.

### 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 für die 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 dem probabilistischen Charakter des Modells, dass sich aus der Differenz zwischen Personen(fähigkeit) und Aufgaben(schwierigkeit) angeben lässt, mit welcher Wahrscheinlichkeit eine Aufgabe gelöst wird. Dabei 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. Im Bereich Orthographie wurde eine 77,7-prozentige Lösungswahrscheinlichkeit festgelegt.

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.

### 2.1.3 Fähigkeitsniveaus

#### *Was bedeuten die Fähigkeitsniveaus<sup>1</sup>?*

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. domänenspezifische 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, NRW<sup>2</sup>) 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 zwei“ heißt soviel wie:

- Aufgaben mittleren Anforderungsniveaus (Fähigkeitsniveau zwei) werden in diesem Inhaltsbereich mit hinreichender Sicherheit (Lösungswahrscheinlichkeit von 62,25 Prozent) gelöst.
- Aufgaben des Niveaus 1, also einfachere Aufgaben mit grundlegenden Anforderungen, werden mit höherer Sicherheit gelöst.
- Anspruchsvollere Aufgaben, die sich auf dem Niveau 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 entwickelt. Die Beschreibungen für den jeweiligen Inhaltsbereich in Deutsch und Mathematik wurden in Hinsicht auf die mit hinreichender Sicherheit zu bewältigenden Anforderungen konkretisiert. Der Entwicklungsprozess führt also empirisch erfasste Informationen und theoretische Überlegungen zusammen:

- Auf der einen Seite benötigt man die empirische Schwierigkeitsverteilung der Aufgaben, d.h. Informationen darüber, von wie vielen Schülerinnen und Schülern die Aufgaben richtig gelöst wurden.
- Auf der anderen Seite sind Annahmen über Denkprozesse erforderlich, die der 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?

<sup>1</sup> Für jeden der Inhaltsbereiche der Fächer Deutsch und Mathematik wurden jeweils 3 Fähigkeitsniveaus festgelegt.

<sup>2</sup> <http://www.learn-line.nrw.de/angebote/lernstand9/> ; Zugriff am 29. März 2005



In technischer Hinsicht umfasste das Vorgehen zur Bestimmung der VERA- Fähigkeitsniveaus damit korrespondierend zwei Schritte:

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.

Die Beschreibungen der Fähigkeitsniveaus stellen einen Beitrag zur Ausbildung von bereichsspezifischen Kompetenzmodellen dar. Es kann zur Zeit jedoch noch nicht von einer abgeschlossenen Theoriebildung gesprochen werden, 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 Lesen, Schreiben, Sprachbetrachtung und Orthographie 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.

Die konkreten Anforderungen der einzelnen Fähigkeitsniveaus werden im Folgenden in Bezug auf die einzelnen Inhaltsbereiche 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.

*Fähigkeitsniveau 2: Erweiterte Fähigkeiten (Verknüpfung von mehreren Informationen)*

- Über den Text verteilte Informationen können gefunden und miteinander verknüpft 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.

*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.
- Auch dann, wenn ein Kurztext geschrieben werden muss, können komplexere Schlüsse gezogen werden. Dazu werden 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.

**Schreiben:***Fähigkeitsniveau 1: Elementare Fähigkeiten*

- Das Schriftbild ist im Wesentlichen leserlich.
- Die Textmenge ist der Aufgabenstellung noch angemessen.
- Formale Standardelemente wie Anrede und Unterschrift werden verwendet, wenn die Aufgabenstellung ein Muster enthält.
- Auftretende Grammatikfehler sind überwiegend nicht schwerwiegend.
- Die Übernahme der Perspektive einer in der Aufgabenstellung genannten Person gelingt.
- Es werden einfache, häufig vorkommende Wörter verwendet.

- Eine Kurzmitteilung enthält einige der wichtigsten in der Aufgabenstellung verlangten Informationen.
- Das Schreiben einer Geschichte nach Reizwörtern gelingt ansatzweise. Die Handlung wird einfach, linear dargestellt und enthält keine Interaktionen zwischen Personen.

### *Fähigkeitsniveau 2: Erweiterte Fähigkeiten*

- Formale Standardelemente wie die Unterschrift werden selbständig verwendet und korrekt platziert.
- Syntaktische Strukturen werden flexibel und der Textsorte angemessen genutzt. So kommen bei der Kurzmitteilung Satzfragmente und in Briefen Satzgefüge vor.
- Mit Blick auf sprachliche Kriterien (Wortschatz, Grammatik) ergeben sich keine nennenswerten Unterschiede zum Fähigkeitsniveau 1.
- Beim Schreiben eines informierenden Textes werden alle wesentlichen Informationen wiedergegeben.
- In einem appellativen Text werden mehrere sachgerechte Argumente vorgebracht. Die Wortwahl entspricht dem appellativen Charakter des Schreibenanlasses.
- Beim freien Schreiben einer Geschichte nach Reizwörtern gelingt die Darstellung einer einfachen Handlung mit Interaktionen zwischen Personen.

### *Fähigkeitsniveau 3: Fortgeschrittene Fähigkeiten*

- Formale Anforderungen werden vollständig erfüllt, z.B. wird die Unterschrift mit einer Grußformel versehen.
- Sätze mit komplexer Struktur werden auch in informierenden oder appellativen Texten verwendet.
- Der verwendete Wortschatz ist erweitert, d.h. er enthält auch ungewöhnliche Wörter oder Fachausdrücke.
- Beim Schreiben eines informierenden Textes werden alle verfügbaren Informationen genutzt, d.h. es werden auch Details mitgeteilt. Dabei können ebenso Fachbegriffe verwendet werden.
- In einem appellativen Text wird originell und wirkungsvoll argumentiert.
- Beim freien Schreiben einer Geschichte nach Reizwörtern gelingt die Darstellung einer komplexen Handlung, d.h. es können mehrere Handlungsstränge dargestellt werden.

## **Sprachbetrachtung:**

### *Fähigkeitsniveau 1: Elementare Fähigkeiten*

- Wenn mehrere Wörter unterschiedlicher Wortartzugehörigkeit zur Wahl stehen, können Verben und Adjektive in der Regel richtig zugeordnet werden.
- Im Rahmen einer Auswahlaufgabe können häufig vorkommende Wortverbindungen (Kollokationen), die sich auf einfache Handlungen beziehen, meistens richtig bestimmt werden.
- Die Bedeutung von häufigen und eindeutigen Redewendungen kann richtig zugeordnet werden.

### *Fähigkeitsniveau 2: Erweiterte Fähigkeiten*

- Im Rahmen einer Auswahlaufgabe können häufig vorkommende Wortverbindungen (Kollokationen) auch dann richtig bestimmt werden, wenn sie sich auf komplexere Handlungen beziehen.
- Im Rahmen des Vergleichs zweier einfacher Sätze können die Wörter bzw. Wortgruppen mit gleicher Satzgliedfunktion ermittelt werden, wenn die Aufgabe mit Hilfe von Umstellproben gelöst werden kann.
- Gibt es mehrere Versionen von Satzgliedgrenzen in kurzen, einfachen Sätzen, kann die richtige Version erfasst werden.
- Die Bedeutung von Redewendungen kann fast durchgängig richtig angegeben werden.

### *Fähigkeitsniveau 3: Fortgeschrittene Fähigkeiten*

- Auch bei umfangreicheren einfachen Sätzen können die Satzgliedgrenzen korrekt zugeordnet werden.
- Wortarten werden auch dann identifiziert, wenn die Aufgabenstellung durch mehrdeutige Alternativen erschwert ist.
- Im Rahmen des Vergleichs zweier einfacher Sätze können die Wörter bzw. Wortgruppen mit gleicher Satzgliedfunktion ermittelt werden, selbst wenn dies eine Analyse der syntaktischen Struktur erfordert.
- Mehrdeutige Redewendungen, die neben der übertragenen auch eine wörtliche Bedeutung besitzen, werden korrekt zugeordnet und können frei erläutert werden.

## **Rechtschreibung<sup>3</sup>:**

### *Fähigkeitsniveau 1: Grundlegende Fähigkeiten*

- Die Laut-Buchstaben Zuordnung gelingt meistens.
- Häufig vorkommende einfache Wörter werden größtenteils richtig geschrieben.
- Substantive und Verben werden größtenteils groß- bzw. kleingeschrieben.

### *Fähigkeitsniveau 2: Erweiterte Fähigkeiten*

- Die Laut-Buchstaben-Zuordnung gelingt fast immer.
- Einfache Wörter werden richtig geschrieben.
- Substantive werden richtig groß- und Verben richtig kleingeschrieben.
- Die Kennzeichnung der Länge bzw. Kürze der Vokale gelingt in vielen Fällen.
- Die Auslautverhärtung stellt kein Problem mehr dar.

---

<sup>3)</sup> Für den Bereich „Rechtschreibung“ wird ein höherer Automatisierungsgrad vorausgesetzt: Das bedeutet, dass die niveauspezifischen Anforderungen mit einer Wahrscheinlichkeit von 77,7 Prozent bewältigt werden können.

### *Fähigkeitsniveau 3: Fähigkeit, weitgehend fehlerfrei zu schreiben*

- Die Schreibung von Länge bzw. Kürze wird fast durchgehend beherrscht.
- Auch die Großschreibung von substantivierten Verben gelingt.
- Über die Auslautverhärtung hinaus bereiten auch andere Aspekte der Morphemkonstanz (Vorsilben, Vokale im Wortstamm) keine Schwierigkeiten mehr.
- Falsche Getrenntschreibungen werden vermieden.

## **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 orientieren sich zum einen an den drei klassischen mathematischen *Inhaltsbereichen* in der Grundschule und zum anderen an verschiedenen *Tätigkeitsanforderungen*. Letzteres entspricht von der Struktur her einer Aufgliederung in „mathematische Prozesse“, wie sie seitens der NCTM-Standards im Jahre 2000 vorgelegt wurden (National Council of Teachers of Mathematics); für eine ausführliche Diskussion der verschiedenen Konzepte, die sich für eine Strukturierung von Aufgaben eignen (inhaltliche Themenbereiche, Prozesse, mathematische Leitideen, Outcomes) vgl. Kapitel 2 der Standards-expertise (Klieme, Avenarius, Blum, Döbrich, Gruber, Prenzel, Reiss, Riquarts, Rost, Tenorth & Vollmer, 2003). Obgleich das skizzierte konzeptuelle Raster nicht auf die Formulierungen der Bildungsstandards Bezug nimmt (die erst später folgten), gibt es bereits eine klare Orientierung an wesentlichen Inhalten dieser Standards. Eine explizite Orientierung an den im Dezember 2004 von der KMK verabschiedeten Standards wird in der nächsten Vergleichsarbeit (Herbst 2005) erfolgen.

Die Zugrundelegung dieses Rasters gewährleistete, dass die Bereiche des Curriculums in hinreichender Breite und in Kombination mit verschiedenen Anforderungen im Aufgabenpool repräsentiert sind. Obwohl die im Jahr 2004 verwendeten Aufgaben zunächst auf Grund des Rahmenplans in Rheinland-Pfalz entwickelt wurden, sind die Aufgaben durch die Aufgabenentwickler/-innen auf Kompatibilität mit den in den anderen Bundesländern gültigen Lehr- bzw. Rahmenplänen überprüft worden.

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

Die folgenden beiden Tabellen enthalten die Aufteilung der Inhaltsbereiche (Tabelle 2, S.14), sowie die Unterteilung der Tätigkeitsanforderungen (Tabelle 3, S.15).

**Tabelle 2:** *Grob- und Feinklassifikation der Inhaltsbereiche*

<b>Grobklassifikation</b>	<b>Feinklassifikation</b>
Arithmetik	Stellenwertsystem Zahleigenschaften Zahlbeziehungen Zahlenfolgen Gesetzmäßigkeiten Einpluseins Einmaleins (besondere Rolle der Null) Rechnen mit größeren Zahlen, mündlich und halbschriftlich Schriftliche Addition, Subtraktion Rechengesetze
Geometrie	Lage von Objekten im Raum; die Ebene mit Richtungen Entfernungen und Koordinaten Geometrische Formen Muster, Gesetzmäßigkeiten Inhalt und Umfang von Flächen; Volumen - Oberfläche Modelle und Netze Achsen-, Dreh- und Schubsymmetrie Abbildungen: z.B. ebene Figuren vergrößern/verkleinern (Ähnlichkeit) Perspektivisches Zeichnen
Sachrechnen/Größen	Stückzahlen, Geldwerte, Uhrzeiten, Zeitmaße, -dauer, Gewichte, Entfernungen, Längen, Geschwindigkeiten Maßbeziehungen und Umrechnungen Bruchteile von Größen, Kommazahlen Zuordnungen Terme Zufallsversuche, Wahrscheinlichkeit *

\* Der Bereich Zufallsversuche und Wahrscheinlichkeit ist als Bereich des Rahmenplans Mathematik in Rheinland-Pfalz enthalten, aber nicht eigens ausgewiesen.

**Tabelle 3:** Grob- und Feinklassifikation der Tätigkeitsaspekte

Grobklassifikation	Feinklassifikation
Standardverfahren und -techniken	Fachbegriffe kennen und anwenden Rechnen Schätzen Skizzieren Technische Hilfsmittel nutzen
Modellieren und Problemlösen	Modellieren Rückübersetzen Formulieren Argumentieren (Lösungen begründen, Lösungswege bewerten) Sich vorstellen (Vorstellungsvermögen) Günstige Strategien nutzen
Umgang mit Daten	Daten gewinnen/Daten darstellen/Daten interpretieren Textverständnis
	Kreativ sein *

\* Die Forderung nach kreativem Umgang mit Mathematik ist explizit im Rahmenplan enthalten, Kreativität lässt sich aber keiner anderen Klassifikation zuordnen.

## 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 Geometrie 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 Niveaus 1, also Aufgaben mit grundlegenden Anforderungen, werden mit höherer Sicherheit gelöst.
- Anspruchsvollere Aufgaben, die sich auf dem Niveau 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 erste holistische Beschreibung der typischen Anforderungen jeder Stufe. Diese Beschreibung wurde im zweiten Schritt verfeinert, indem zusätzlich schwerigkeitsbestimmende Merkmale (z.B. Textlänge, Antwortformat, sprachliche Komplexität etc.) der Aufgaben einbezogen wurden.

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

### **Arithmetik:**

#### *Stufe 1: Elementare Fähigkeiten (grundlegende Kenntnisse arithmetischer Verfahren)*

- In Gleichungen kann eine fehlende Rechenoperation eingesetzt werden.
- Schriftliche Additionsaufgaben werden gelöst.
- Schriftliche Subtraktionsaufgaben ohne Übertrag werden gelöst.
- Die Aufgabenlösung gelingt bei sehr kurzen und sprachlich einfachen Texten in der Aufgabenstellung.

#### *Stufe 2: Erweiterte Fähigkeiten (Umfassende Kenntnis der Addition und Subtraktion)*

- Einsicht in das Stellenwertsystem liegt vor und es kann angewendet werden.
- Schriftliche Subtraktion gelingt auch mit Überträgen oder in unüblichen Formaten (Lückenaufgaben).
- Das Vervollständigen von Zahlenreihen gelingt, wenn die zugrunde liegende Regel Strichrechnung erfordert.
- Sprachlich formulierte Aufgaben zu den Grundrechenarten werden bewältigt, im Falle von Strichrechnung auch mehrschrittig.
- Aufgaben mit sprachlich komplexeren und längeren Texten werden gemeistert.

#### *Stufe 3: Fortgeschrittene Fähigkeiten (Flexible Beherrschung der Grundrechenarten)*

- 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.
- Schriftliche Subtraktion gelingt auch mit Überträgen in unüblichen Formaten (Lückenaufgaben).
- Das Erkennen und Benennen einer zweischrittigen Regel für eine Zahlenfolge gelingt.
- Überschlagsrechnungen können vorgenommen werden.

### **Geometrie:**

#### *Stufe 1: Elementare Fähigkeiten (Kenntnisse grundlegender geometrischer Formen und Abbildungen)*

- Einfache geometrische Muster können fortgesetzt werden.
- Aufgaben zu Umfang, Fläche oder Volumen, die sich durch Abzählen lösen lassen, werden bewältigt.



- Ebene geometrische Formen können durch Verschiebung/Drehung zu anderen Formen kombiniert werden.
- Das Ergänzen zu achsensymmetrischen Figuren gelingt.
- Die Aufgabenlösung gelingt bei sprachlich einfachen und kurzen Texten.

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

- Körpernetze können zugeordnet werden.
- Aufgaben zu Körpernetzen können in der gedanklichen Vorstellung gelöst werden.
- Flächen können verglichen werden.
- „Fehler“ in Mustern werden erkannt.
- Komplexere Muster können fortgesetzt werden.
- 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.
- Rechtecke können nach vorgegebenen Maßen konstruiert werden.
- Das Navigieren in Felder bei vorgegebenen Koordinaten gelingt.

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

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

**Sachrechnen/Größen:**

*Stufe 1: Elementare Fähigkeiten*

- Die Anwendung von Addition (auch wiederholte Additionen) und Subtraktion in authentischen Aufgaben gelingt bei Aufgaben mit Auswahl aus vorgegebenen Lösungen.
- Grundlegende Kenntnis von vertrauten Maßeinheiten (Längen-, Zeit-, Gewichts- und Geldeinheiten).
- Offensichtlich unlösbare Aufgaben werden erkannt.

### *Stufe 2: Erweiterte Fähigkeiten (Entwickelte Fähigkeiten im Umgang und Rechnen mit Größen)*

- 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.
- Rundungen und Schätzungen gelingen bei Aufgaben mit vorgegebenen Lösungen.
- Verknüpfungen von Operationen werden bewältigt.
- Der Umgang mit elementaren Brüchen gelingt.
- Aufgaben mit mehreren zu verarbeitenden Größen werden gemeistert.

### *Stufe 3: Fortgeschrittene Fähigkeiten (Eigenständige Problemlösungen)*

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

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/aufgaben.htm>) im Internet zur Verfügung gestellt.

## **3 Vergleichsarbeiten im September 2004**

### **3.1 Ablauf und Stichproben**

Von Mitte Mai bis Ende August 2004 fanden in den sieben beteiligten Ländern jeweils zweiwöchige Erprobungsphasen mit Fokus auf der Internetnutzung statt. Vom 1. September bis zum 17. September konnten die Schulen und Kooperationsgruppen internetbasiert und menügesteuert einen Teil der Aufgaben auswählen. In Mathematik sollten hier zehn Aufgaben, in Deutsch eine Lese- und eine Schreibaufgabe ausgesucht werden. Dies entspricht in beiden Inhaltsbereichen etwa der Hälfte der Aufgaben. In Mathematik konnten die Lehrkräfte zusätzlich Lösungshäufigkeiten für die gewählten Aufgaben einschätzen. Die restlichen (zentralen) Aufgaben erhielten alle Schulen erst drei bis vier Tage vor der Durchführung der Vergleichsarbeit.

Zeitgleich am 28. September 2004 wurde in allen öffentlichen und einigen privaten Grundschulen der beteiligten Bundesländer sowie in zahlreichen deutschen Auslandsschulen (siehe Abbildung 1, S. 5) die Vergleichsarbeit in Mathematik geschrieben, am 30. September in Deutsch. Die Bearbeitungszeit betrug in Mathematik 50 Minuten, in Deutsch 90 Minuten plus einer Pause von 10

Minuten. Nach der Durchführung der Vergleichsarbeiten konnten die Lehrkräfte die Korrekturanweisungen für die zentral vorgegebenen und die von ihnen ausgewählten Aufgaben vom Landauer Universitätsserver herunterladen. Nach der Korrektur durch die Lehrkräfte konnten diese die Ergebnisse offline in ein Excel-Sheet oder online direkt auf dem Universitätsserver eingeben. Die offline eingegebenen Werte wurden dann zum Server hochgeladen. Ein bis zwei Wochen nach der vollständigen Dateneingabe konnten die Lehrkräfte die Ergebnisse abrufen. Diese wurden ihnen in zwei Wellen über das Internet zurückgemeldet (vgl. 3.2). Zeitnah zur vollständigen Dateneingabe konnten die Lehrkräfte die Basisinformationen einsehen, und ab dem 19. Januar 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.

Die *Zentralstichprobe* dient dem Zweck, für das 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, Nordrhein-Westfalen, Mecklenburg-Pommern, Rheinland-Pfalz und Schleswig-Holstein 983 Klassen in 527 Schulen ein, zusätzlich alle Schulen aus Bremen (274 Klassen in 99 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 Vergleichsarbeit wie auch die schulinterne Auswertung waren somit in den Klassen der Zentralstichprobe identisch mit der Durchführung in allen anderen Klassen. Die Eingabe der Daten als auch die Ergebnismeldung aus Landau war ebenfalls für alle Schulen gleich, die Lehrkräfte der Zentralstichprobe beantworteten zusätzlich einen Kontextfragebogen, dessen Informationen für den „fairen Vergleich“ benötigt wurden (siehe auch 3.3, S. 20).

## 3.2 Elemente der Ergebnismeldungen

### 3.2.1 Welle 1

In der ersten Rückmeldewelle von VERA wurden die so genannten **Basisinformationen** zurückgemeldet:

- Fähigkeitsniveaus auf Ebene der einzelnen Schülerinnen und Schüler als Zahlenwerte, auf Klassen- und auf Schulebene als Verteilungen
- 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 bei VERA 2004 nur im Bereich Mathematik erhoben. Zusätzlich zu der Gegenüberstellung der eingeschätzten und der tatsächlichen Lösungshäufigkeit wurde ein Genauigkeitsindex (Korrelationen) für die eigene Einschätzung berichtet und im Vergleich zur Häufigkeitsverteilung in der Population von VERA 2004 dargestellt.

### 3.2.2 Welle 2

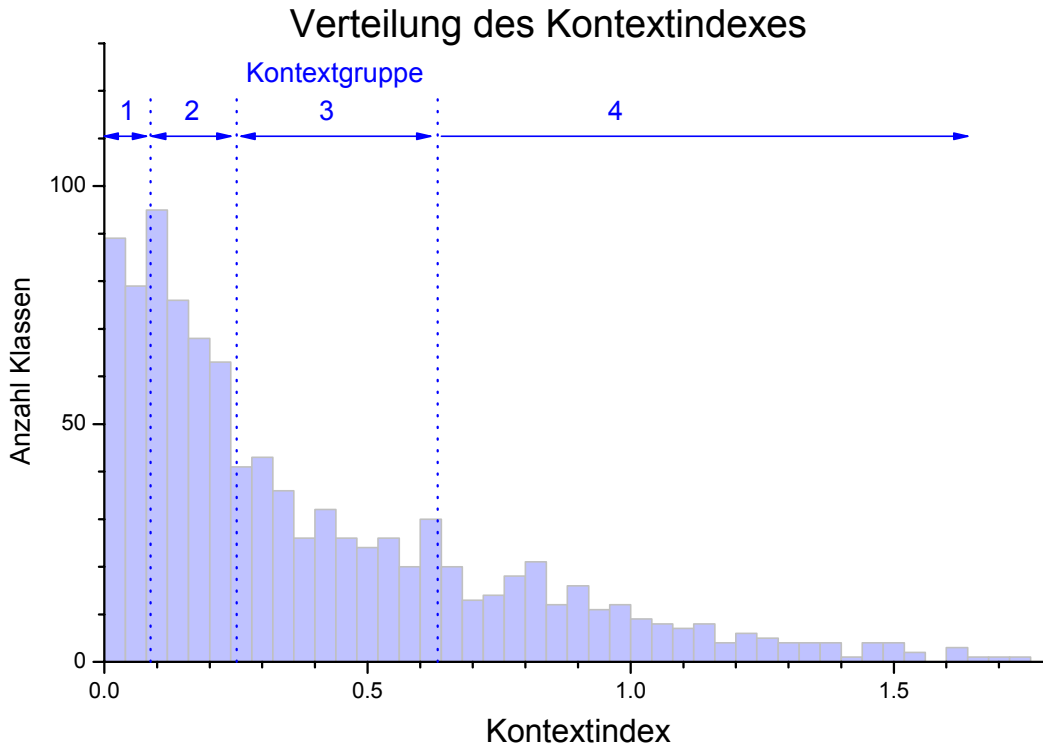
In der zweiten Rückmeldewelle wurden den Lehrkräften ab Januar 2005 weitere **Zusatzinformationen** zur Verfügung gestellt, und zwar:

- 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 2004: In dieser Darstellung wurden die richtigen Lösungen der jeweiligen Klasse im Vergleich zur Gesamtpopulation verortet.

### 3.3 Bildung von Kontextgruppen

Um den beteiligten Lehrkräften einen „fairen“ Vergleich anbieten zu können, indem Kontextunterschiede zwischen Klassen berücksichtigt werden, wurden vier Gruppen gebildet und mit Blick auf die Kontextmerkmale und die Leistungen beschrieben. Jede Lehrkraft konnte ihre Klasse auf der Grundlage von Einschätzungen der Klassenzusammensetzung einer von vier Kontextgruppen zuordnen und die in der Klasse erzielten Leistungen mit denen der „passenden“ Kontextgruppe vergleichen. Die Bildung dieser Kontextgruppen erfolgte mit Hilfe eines regressionsanalytischen Ansatzes auf der Ebene der Klassen der Zentralstichprobe. Da die Kontextgruppen länderübergreifend definiert werden sollten, musste zunächst sicher gestellt werden, dass die beteiligten Bundesländer gleichmäßig berücksichtigt werden, d.h. der Einfluss jedes Bundeslandes sollte unabhängig von der Größe der Zentralstichprobe ausfallen. Entsprechend wurden die Klassen gewichtet. Im Anschluss daran wurden per Regressionsanalyse diejenigen Merkmale ermittelt, die mit der (über alle sieben Inhaltsbereiche gemittelten) Leistung im Zusammenhang stehen. In diese Analyse aufgenommen wurden zum einen auf Klassenebene aggregierte Schülerdaten (Geschlecht, Deutsch als nicht dominante Sprache, Wiederholung der vierten Klasse), zum anderen die von den Lehrkräften im Lehrerfragebogen angegebenen Informationen zur Charakterisierung der Klasse (Sozialer Brennpunkt, Anteil Kinder aus Unterschichtfamilien, Anteil Kinder aus Familien betroffen von Arbeitslosigkeit, Anteil Kinder aus Sozialhilfe-Empfänger-Familien). Es zeigte sich, dass die folgenden vier Variablen die stärksten Prädiktoren darstellen: „Prozentualer Anteil von Kindern mit Deutsch als nicht dominanter Sprache“, „Prozentualer Anteil von Kindern aus Familien der Unterschicht“, „Prozentualer Anteil von Kindern aus Sozialhilfe beziehenden Familien“ und die Frage, ob sich die Schule in einem sozialen Brennpunkt befindet (Anteil insgesamt aufgeklärter Varianz: 25 Prozent). Im nächsten Schritt wurden diese vier Informationen gemäß der ermittelten Regressionsgewichte zu einem Kontextwert kombiniert. Abbildung 2 zeigt die Verteilung dieser Werte der 992 Klassen, für die vollständige Daten vorliegen und die deshalb berücksichtigt werden konnten.

Die Korrelationen dieses Kontextwertes mit den Leistungen in den verschiedenen Inhaltsbereichen sind in Tabelle 4 dargestellt. Erwartungsgemäß zeigt sich ein deutlich engerer Zusammenhang zu den Leistungen im Fach Deutsch, innerhalb dessen die engste Beziehung zum Leseverständnis besteht. Für Mathematik ergibt sich, aufgrund der höheren Leseverständnisanforderungen ebenfalls erwartungsgemäß, der engste Zusammenhang mit dem Bereich Sachrechnen/Größen.



**Abbildung 2:** Verteilung der Kontextwerte (Bereich 1 = günstige bis Bereich 4 = ungünstige Kontextbedingungen)

**Tabelle 4:** Korrelation zwischen der mittleren Klassenleistung und dem Kontextwert (r)\*

<b>Mathematik</b>	<b>r</b>	<b>Deutsch</b>	<b>r</b>
Arithmetik	-0,32	Leseverständnis	-0,48
Geometrie	-0,33	Schreiben	-0,41
Sachrechnen/Größen	-0,41	Sprachbetrachtung	-0,46
		Orthographie	-0,44

\* minimale Stichprobengröße: N = 262247 (Kinder)

Anschließend wurden die Klassen in vier Kontextgruppen aufgeteilt, wobei die „günstigste“ Kontextgruppe eins und die „ungünstigste“ Kontextgruppe vier jeweils etwa 20 Prozent umfassen, während die beiden mittleren Gruppen mit je ca. 30 Prozent vertreten sind.

Die Zuordnung aller „Nicht-Zentralstichproben-Klassen“ zu einer der vier Gruppen erfolgt nach der Abfrage der relevanten Informationen im Internet. Durch die kompensatorische Art der Verknüpfung der Einzelinformationen können jedoch sehr unterschiedliche Kontextkonfigurationen zur Zuweisung in die gleiche Kontextgruppe führen. Dies wird besonders im Unterschied zwischen den Ländern deutlich: Während Klassen der ungünstigsten Kontextgruppe in den alten Bundesländern vorwiegend von Kindern mit Deutsch als nicht dominanter Sprache besucht werden, trifft dies in den neuen Bundesländern nicht zu. Entsprechend erfolgte die Beschreibung der gebildeten Kontextgruppen für folgende drei Ländergruppen: Stadtstaaten (Berlin, Bremen), alte Flächenländer (Nordrhein-Westfalen, Rheinland-Pfalz, Schleswig-Holstein), neue Flächenländer (Brandenburg,

Mecklenburg-Vorpommern). Tabelle 5 gibt die durchschnittlichen Eigenschaften der jeweiligen Gruppen wieder.

**Tabelle 5:** Beschreibung der vier Kontextgruppen in den drei Ländergruppen anhand der durchschnittlichen Merkmalsausprägungen, Angaben in Prozent

	Kontextgruppe			
	1	2	3	4
<b>Stadtstaaten</b>				
Anteil der Schülerinnen und Schüler mit Deutsch als nicht dominanter Sprache	~1	~3	~13	~40
Anteil von Klassen im sozialen Brennpunkt	0	0	~ 42	~ 92
Anteil der Schülerinnen und Schüler aus Familien der Unterschicht	~3	~14	~28	>50
Anteil der Schülerinnen und Schüler aus Familien, die Sozialhilfe beziehen	~3	~11	~17	~38
<b>Alte Flächenländer</b>				
Anteil der Schülerinnen und Schüler mit Deutsch als nicht dominanter Sprache	<1	~3	~10	~35
Anteil von Klassen im sozialen Brennpunkt	0	< 1	~ 47	~ 98
Anteil der Schülerinnen und Schüler aus Familien der Unterschicht	~6	~18	~33	>50
Anteil der Schülerinnen und Schüler aus Familien, die Sozialhilfe beziehen	~1	~6	~10	~20
<b>Neue Flächenländer</b>				
Anteil der Schülerinnen und Schüler mit Deutsch als nicht dominanter Sprache	<1	<1	~2	~10
Anteil von Klassen im sozialen Brennpunkt	0	0	~ 47	100
Anteil der Schülerinnen und Schüler aus Familien der Unterschicht	~4	~18	~39	>50
Anteil der Schülerinnen und Schüler aus Familien, die Sozialhilfe beziehen	~1	~9	~17	~34

Erwartungsgemäß ergibt sich über die Länder hinweg eine ungleiche Verteilung der Klassen auf die Kontextgruppen. Abbildung 3 zeigt die resultierenden relativen Verteilungen der Zentralstichproben der Länder.

Es bestätigt sich die Erwartung, dass insbesondere die beiden Stadtstaaten Bremen und Berlin mit jeweils 10 Prozent ausgesprochen niedrige Anteile der günstigsten Kontextgruppe eins und mit mindestens 42 Prozent sehr hohe Anteile der ungünstigsten Kontextgruppe vier aufweisen. Die Unterschiede zwischen den Flächenländern fallen demgegenüber eher gering aus.

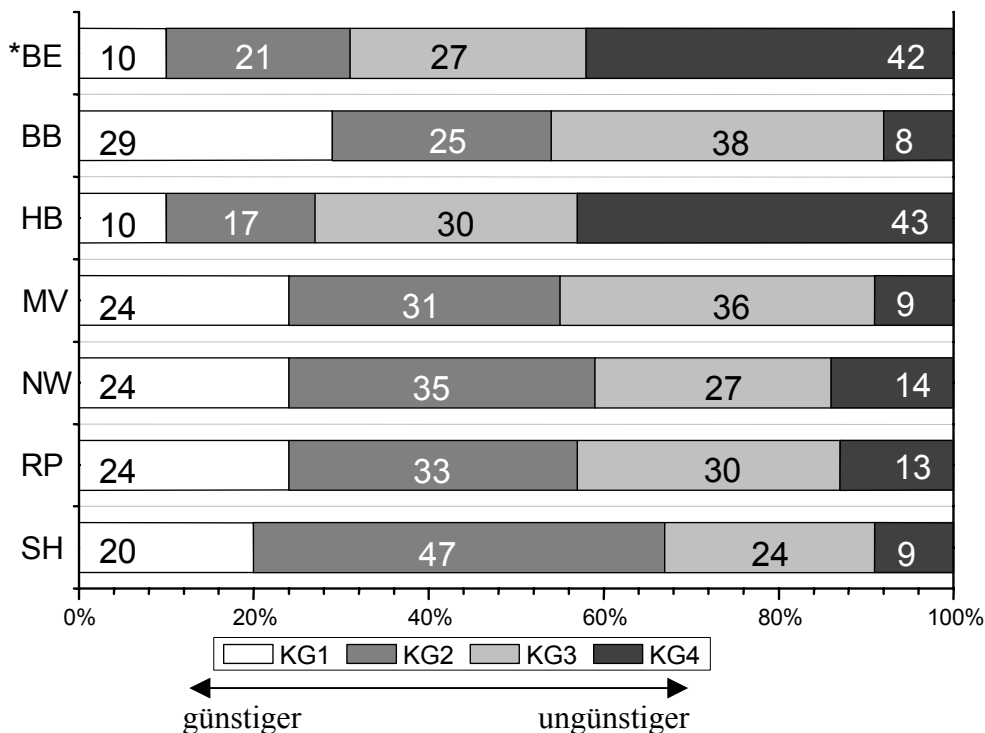


Abbildung 3: Verteilung der Klassen auf die vier Kontextgruppen (KG) nach Land\*; Angaben in Prozent

### 3.4 Gesamt-Ergebnisse

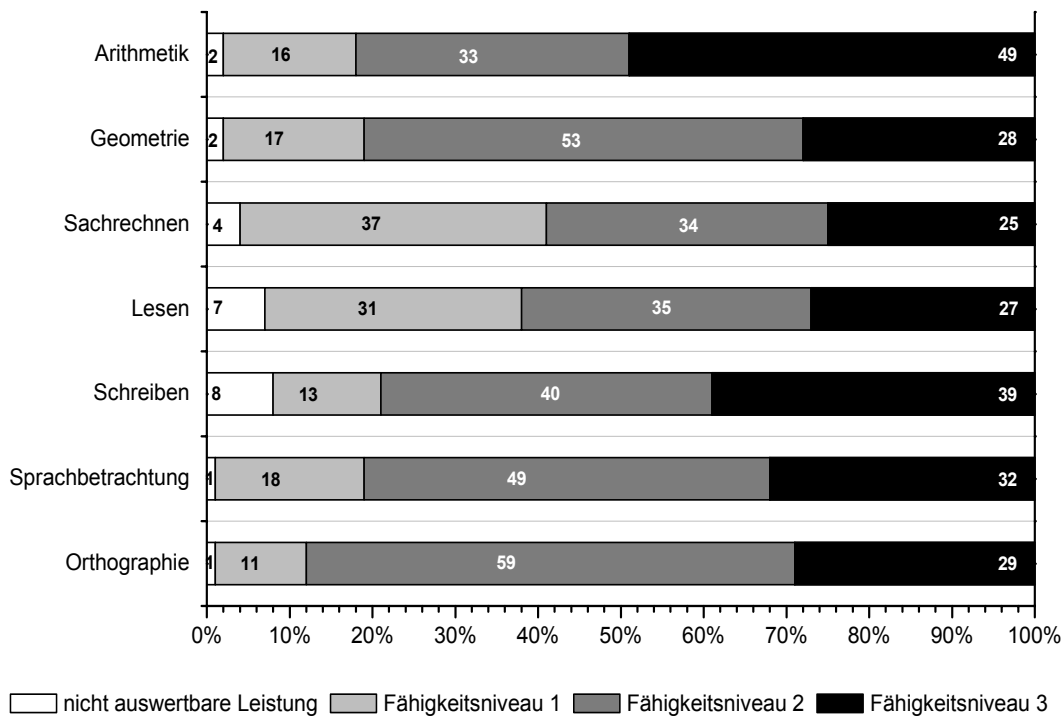
#### 3.4.1 Verteilung der Fähigkeitsniveaus

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 Schüler und Schülerinnen in der vierten Grundschulklasse können. Ein Vergleich der Inhaltsbereiche z.B. zwischen Orthographie und dem Schreiben ist nur sinnvoll über die Fähigkeitsniveauverteilungen: Die Schülerinnen und Schüler erreichen z.B. im Schreiben zu 39 Prozent das höchste Fähigkeitsniveau (vgl. Abbildung 4, S. 24), in Orthographie dagegen nur mit einem niedrigeren Prozentsatz (29 Prozent). Dementsprechend wird die folgende Abbildung in Bezug auf die Verteilungen in den sieben geprüften Inhaltsbereichen genauer besprochen.

Für das Fach *Mathematik* wird deutlich, dass in den Inhaltsbereichen Arithmetik und Geometrie der Großteil der Schülerinnen und Schüler (mehr als 80 Prozent) erweiterte bis fortgeschrittene Fähigkeiten aufweisen, also das Niveau 2 und 3 erreichen. Der Anteil an Kindern auf dem dritten Fähigkeitsniveau ist für Arithmetik am höchsten (49 Prozent). Demgegenüber fällt auf, dass im Sachrechnen verhältnismäßig wenige Schülerinnen und Schüler dem dritten Fähigkeitsniveau und

\* BE = Berlin, BB = Brandenburg, HB = Bremen, MV = Mecklenburg-Vorpommern, NW = Nordrhein-Westfalen, RP = Rheinland-Pfalz, SH = Schleswig-Holstein

über 40 Prozent der Schülerinnen und Schüler einem Niveau, das höchstens das Beherrschen elementarer Aufgaben umfasst, zugeordnet werden.



**Abbildung 4:** Gesamtverteilung der Fähigkeitsniveaus über alle Länder hinweg; Angaben in Prozent

In *Deutsch* sind die Fähigkeitsverteilungen für die Inhaltsbereiche Sprachbetrachtung und insbesondere Orthographie relativ günstig: Der Anteil an Kindern mit nicht auswertbaren Leistungen ist besonders gering und die meisten Kinder, nämlich jeweils über 80 Prozent, weisen Fähigkeiten auf dem zweiten und dritten Niveau auf. Die übrigen Ergebnisse aus Abbildung 4 sind selbsterklärend und werden daher nicht weiter kommentiert.

Aus der Verteilung der Ergebnisse kann zusammenfassend geschlussfolgert werden, dass vor allem beim Leseverständnis und beim Sachrechnen, vermutlich aber auch beim Schreiben ein Förderbedarf besteht. Demgegenüber zeigen in Arithmetik und in Orthografie ein Großteil der Kinder erweiterte bis fortgeschrittene Fähigkeiten.

### 3.4.2 Fähigkeitsniveaus nach Kontextgruppen

In Abschnitt 3.3 wurde dargestellt, wie die vier unterschiedlichen Kontextgruppen gebildet wurden. Es stellt sich natürlich die Frage, inwieweit sich diese Kontextgruppen im Niveau der Fähigkeiten voneinander unterscheiden bzw. in welchen Inhaltsbereichen Unterschiede besonders markant oder besonders gering ausfallen. In Tabelle 6 sind die Verteilungen der Zentralstichprobe auf die Fähigkeitsniveaus für die drei mathematischen Inhaltsbereiche nach den vier Kontextgruppen differenziert dargestellt. Tabelle 7 zeigt die entsprechenden Informationen für die vier Inhaltsbereiche im Fach Deutsch.



**Tabelle 6:** Verteilung der Fähigkeitsniveaus, aufgeschlüsselt nach Kontextgruppen (Mathematik); Angaben in Prozent

	Kontextgruppe	n.a.L.*	FN 1	FN 2	FN 3	N (Klassen)
Arithmetik	1	2,9	10,8	26,4	59,9	190
	2	3,6	12,3	30,0	54,1	285
	3	3,8	15,1	31,6	49,5	291
	4	6,0	21,4	31,9	40,7	220
Geometrie	1	2,4	11,5	51,8	34,3	190
	2	2,4	13,3	50,0	34,3	285
	3	3,3	18,8	50,5	27,4	291
	4	5,2	24,7	50,0	20,0	220
Sachrechnen	1	4,1	28,5	36	31,5	190
	2	5,0	33,6	33,7	27,7	285
	3	6,1	38,7	32,2	23,0	291
	4	11,0	43,2	28,8	17,1	220

\* nicht auswertbare Leistung

Zunächst wird deutlich, dass sich die vier Kontextgruppen insofern erwartungskonform unterscheiden, als der Anteil der Kinder mit fortgeschrittenen Fähigkeiten mit zunehmender Günstigkeit des Kontextes nahezu stetig ansteigt, während der Anteil von Kindern mit nicht auswertbaren Leistungen stetig abnimmt. Zugleich ist jedoch auch erkennbar, dass der deutlichste Unterschied zwischen den vier Gruppen in allen drei Inhaltsgebieten jeweils zwischen der dritten und vierten Kontextgruppe liegt. Offenbar wirken sich die ungünstigen Kontextbedingungen in der in Kontextgruppe vier vorhandenen Kumulation stärker aus.

**Tabelle 7:** Verteilung der Fähigkeitsniveaus, aufgeschlüsselt nach Kontextgruppen (Deutsch); Angaben in Prozent

	Kontextgruppe	n.a.L.*	FN 1	FN 2	FN 3	N (Klassen)
Lesen	1	5,0	20,4	37,2	37,4	189
	2	6,2	25,8	36,7	31,4	289
	3	9,2	31,6	33,8	25,4	289
	4	16,9	39,6	27,8	15,7	221
Schreiben	1	6,8	8,1	38,4	46,7	189
	2	7,4	11,0	37,9	43,7	289
	3	9,7	12,8	39,3	38,2	289
	4	15,6	18,8	37,3	28,3	221
Sprachbetrachtung	1	2,5	11,7	47,2	38,7	189
	2	2,5	14,0	47,4	36,2	289
	3	3,6	20,7	47,5	28,2	289
	4	6,2	30,7	47,0	16,2	221

	Kontextgruppe	n.a.L.*	FN 1	FN 2	FN 3	N (Klassen)
Orthographie	1	0,7	6,0	54,0	39,3	189
	2	1,1	8,7	57,1	33,1	289
	3	1,8	13,0	58,3	27,0	289
	4	4,5	20,0	58,8	16,7	221

\* nicht auswertbare Leistung

Das Ergebnis für Deutsch ist ähnlich: In allen vier Inhaltsbereichen liegt der größte Unterschied jeweils zwischen den Kontextgruppen drei und vier. Im Vergleich zu den Unterschieden im Fach Mathematik fallen die Unterschiede zwischen den Kontextgruppen für die vier Inhaltsbereiche des Faches Deutsch insgesamt größer aus. Dies wird insbesondere beim Vergleich der beiden günstigsten Kontextgruppen eins und zwei deutlich, z.B. im Vergleich der Anteile von Kindern unter Fähigkeitsniveau zwei. Darüber hinaus wird deutlich, dass im Bereich des Leseverständnisses und des Schreibens die größten Unterschiede zwischen den Kontextgruppen zu konstatieren sind. Dies kann im Sinne der elementaren Bedeutung des Leseverständnisses interpretiert werden. Im Einklang damit erweist sich im Bereich der Mathematik mit dem Sachrechnen derjenige Bereich als am kontextsensitivsten, der auch die größten Leseverständnisanforderungen umfasst.

### 3.4.3 Zusammenhänge zwischen den Fähigkeitsbereichen

Die folgende Tabelle 8 gibt den Zusammenhang (Interkorrelationen) zwischen den Leistungen in den verschiedenen Inhaltsbereichen wieder. Darin entsprechen die dunkelgrauen Felder den Interkorrelationen innerhalb eines Faches und die hellgrauen Felder den Zusammenhängen zwischen den Inhaltsbereichen für Mathematik und Deutsch. Die Korrelationen decken das Spektrum von niedrig ( $r = ,25$ , Geometrie/Schreiben) bis mittel ( $r = ,59$ , Lesen/Sprachbetrachtung) ab.

Zunächst fällt auf, dass die höchsten Korrelationen innerhalb der Fächer Mathematik und Deutsch (dunkelgraue Felder) auftreten. So kovariieren z.B. im Fach Mathematik die Leistungen in Arithmetik, Geometrie und Sachrechnen lediglich mittelhoch ( $r = ,50$  bis  $r = ,57$ ). Es hat sich aber als sinnvoll herausgestellt, nach Fähigkeitsbereichen innerhalb von Deutsch und Mathematik zu differenzieren. Ein ähnliches Bild zeigt sich für Deutsch: Auch hier bestehen zwischen den Inhaltsbereichen die mitunter höchsten Korrelationen wie z.B. zwischen Lesen und Sprachbetrachtung von  $r = ,59$ . Dass auch hier Teilbereiche bzw. -fähigkeiten im Fach Deutsch angenommen werden können, belegen die mittleren, also nicht perfekten Zusammenhänge zwischen den einzelnen Inhaltsbereichen.

**Tabelle 8:** Interkorrelationen zwischen den Leistungen in den Inhaltsbereichen ( $r$ )\*

	Arithmetik	Geometrie	Sachrechnen	Lesen	Schreiben	Sprachbetrachtung	Orthographie
Arithmetik		,53	,57	,43	,27	,43	,38
Geometrie			,50	,39	,25	,42	,30
Sachrechnen				,45	,26	,44	,36
Lesen					,41	,59	,52
Schreiben						,36	,36
Sprachbetrachtung							,48

\* minimale Stichprobengröße:  $N = 262247$  (Kinder)

## 4 Landesspezifische Ergebnisse

### 4.1 Beschreibung der Leistungen

Im Folgenden werden die Fähigkeitsniveaus unter verschiedenen Gesichtspunkten diskutiert: Zunächst wird allgemein die Verteilung der Schüler auf den einzelnen Niveaus dargestellt (vgl. 4.1.1, S. 27) und der Wahl- und Zentralteil miteinander verglichen (vgl. 4.1.2, S. 28). Im Anschluss werden Unterschiede zwischen (bzw. innerhalb) den untersuchten Klassen/Schulen (siehe 4.1.3, S. 30), Geschlechtsunterschiede (siehe 4.1.4, S. 31), sowie Unterschiede zwischen Schülern mit Deutsch als dominanter vs. nicht dominanter Sprache (vgl. 4.1.5, S. 33) genauer beleuchtet.

Abschließend wird unter 4.1.6 die diagnostische Kompetenz der Lehrkräfte anhand der empirischen Schülerleistungen bezüglich ihrer fachdidaktischen und pädagogischen Komponenten dargestellt.

#### 4.1.1 Verteilung der Fähigkeitsniveaus

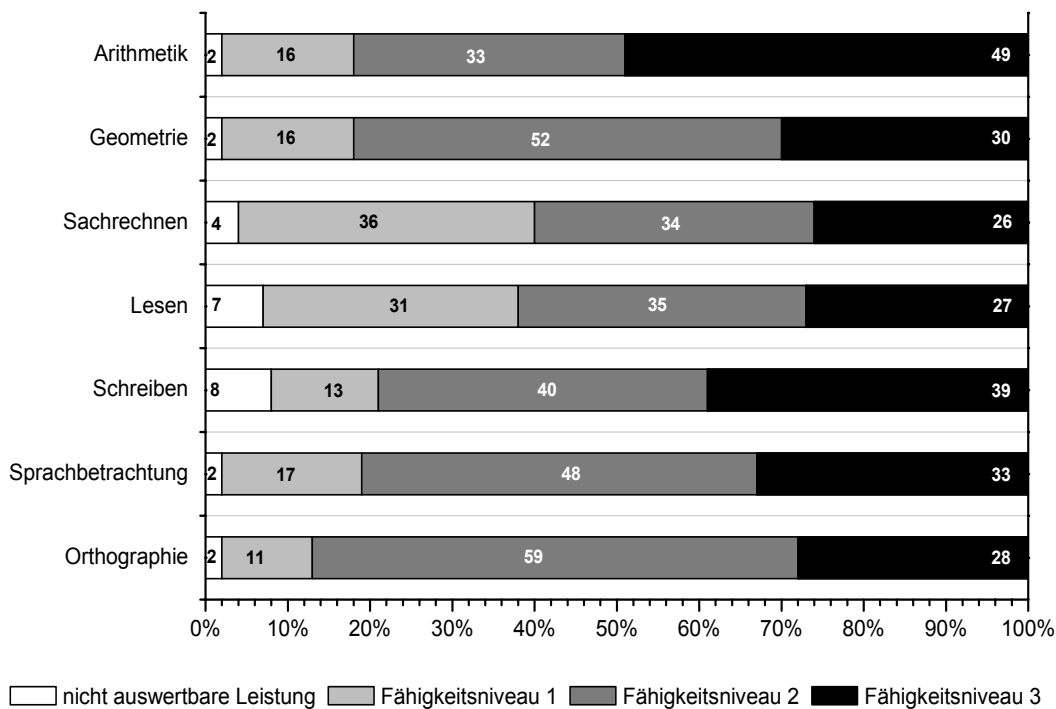


Abbildung 5: Gesamtverteilung der Fähigkeitsniveaus für NRW (Angaben in Prozent)

In den Mathematik- Inhaltsbereichen Arithmetik und Geometrie weist der Großteil der Schülerinnen und Schüler (jeweils über 80 Prozent) erweiterte bis fortgeschrittene Fähigkeiten auf. Im Sachrechnen erreichen verhältnismäßig weniger, nämlich 60 Prozent maximal das dritte Fähigkeitsniveau und 40 Prozent ein Niveau, das höchstens das Beherrschen elementarer Aufgaben umfasst. Im Fach Deutsch gibt es in den Inhaltsbereichen Sprachbetrachtung und Orthographie vergleichsweise wenig Kinder mit nicht auswertbaren Leistungen (2 Prozent) und viele Kinder auf dem zweiten und dritten Fähigkeitsniveau (81 Prozent bzw. 87 Prozent). Ein besonders großer Anteil an Schülerinnen und Schülern, die höchstens das Fähigkeitsniveau eins erreichen, findet sich in den Inhaltsbereichen Lesen (38 Prozent) und Schreiben (21 Prozent, davon 8 Prozent nicht auswertbare Leistungen).

Die obige Abbildung verdeutlicht, dass auch die landesspezifische Verteilung der Fähigkeitsniveaus ein vergleichbares Bild wiedergibt, wie es im Abschnitt 3.4.1 im Zusammenhang mit der länderübergreifenden Verteilung der Fähigkeitsniveaus dargestellt wurde: Sowohl beim Leseverständnis, Sachrechnen als auch beim Schreiben scheint Förderbedarf zu bestehen. Die nur sehr geringen Abweichungen von der länderübergreifenden Verteilung lassen sich dadurch erklären, dass das Land Nordrhein-Westfalen bei VERA den „Löwenanteil“ (62,8 Prozent) an Schülerinnen und Schülern gestellt hat und somit im Vergleich zu den anderen teilnehmenden Bundesländern maßgeblich die Verteilung beeinflusst.

#### 4.1.2 Gegenüberstellung: Wahl- vs. Zentralteil

Im Folgenden werden die Fähigkeitsniveaus der Kinder in Bezug auf die Zentral- und die Wahlaufgaben gegenüber gestellt. Dazu wurden die Fähigkeitsniveaus jeweils getrennt berechnet. Der Zentral- und der Wahlteil umfassen ca. die Hälfte der Aufgaben. Als Konsequenz daraus beruht die Schätzung der Fähigkeiten auf einer geringeren Anzahl von Aufgaben und ist mit einer größeren Unsicherheit verbunden. Die Zugehörigkeit einer Aufgabe zu Zentral- oder Wahlteil war dabei für die Kinder nicht zu erkennen.

Ist mit besseren Leistungen bei den Zentralaufgaben oder bei den Wahlaufgaben zu rechnen? Für beide Annahmen sprechen gute Gründe:

Einerseits standen die Aufgaben des Zentralteils am Anfang der Testhefte, die des Wahlteils dahinter. Es liegt die Vermutung nahe, dass die Kinder die Aufgaben der Reihe nach, also mit dem Zentralteil beginnend bearbeitet haben. Vor allem schwächere Schülerinnen und Schüler dürften in der Testzeit nicht die Aufgaben am Ende des Testheftes erreicht haben. Ebenso ist es möglich, dass im Verlauf der Bearbeitung die Motivation nachgelassen hat, so dass die Aufgaben des Zentralteils besser bearbeitet wurden als die des Wahlteils. Einschränkend muss erwähnt werden, dass in der Schülerinstruktion explizit darauf hingewiesen wurde, dass es sinnvoll ist, zuerst die Aufgaben zu bearbeiten, die man als leichter empfindet – die Reihenfolge der Aufgabenbearbeitung war den Kindern also freigestellt.

Andererseits ermöglicht das Konzept der Aufgabenauswahl, dass die Lehrkräfte die Aufgaben aussuchen konnten, die gut zu ihrem Unterricht passen bzw. die sie als angemessen für den Leistungsstand ihrer Schülerinnen und Schüler erachten. Da die Wahlaufgaben aus organisatorischen Gründen früher heruntergeladen werden konnten als die Zentralaufgaben, bestand für die ausgewählten Aufgaben außerdem die Möglichkeit, diese zu üben. Dies kann dazu führen, dass die Aufgaben des Wahlteils besser beherrscht wurden.

Das bedeutet also, dass in den Zentral- und in den Wahlaufgaben möglicherweise zwei gegenläufige Effekte wirken. Dementsprechend komplex ist die Interpretation von Unterschieden in den Leistungsverteilungen. Zur Interpretation der Unterschiede zwischen den Zentral- und Wahlaufgaben wird daher im Folgenden die *Effektstärke*<sup>4</sup> „d“ als zugrunde liegende Metrik herangezogen:

---

<sup>4</sup> 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

Dieses Maß gibt Hinweise auf die Bedeutsamkeit eines Unterschieds, indem die Unterschiede zwischen den Gruppen auf die Streuung der Testwerte standardisiert werden. Für die Betrachtung des  $d$ -Maßes spricht auch, dass die Berechnung der Effektstärke auf dem Fähigkeitskontinuum beruht und nicht durch die künstliche Unterteilung in Fähigkeitsniveaus in seiner Aussagekraft eingeschränkt wird.

**Tabelle 9:** Verteilung der Fähigkeitsniveaus Zentral- vs. Wahlteil, Angaben in Prozent

		n.a.L*	FN 1	FN 2	FN 3	N (Kinder)	d**
<b>Mathematik</b>							
Arithmetik	Zentral	0	21,8	30	48,2	171606	
	Wahl	5,5	17,4	27,9	49,2	171864	-0,13
Geometrie	Zentral	1,9	14	46,8	37,3	171605	
	Wahl	4,4	21,4	42,9	31,3	171867	-0,10
Sachrechnen	Zentral	7,1	31,5	28,9	32,5	171605	
	Wahl	0,5	48,5	16,5	34,5	171864	0,01
<b>Deutsch</b>							
Lesen	Zentral	8,1	25,5	30,7	35,7	170346	
	Wahl	13,9	25,5	40,2	20,4	170294	-0,29

\* nicht auswertbare Leistung

\*\* Maß für die Effektstärke

- negativer Wert = bessere Leistung bei den Zentralaufgaben
- positiver Wert = bessere Leistung bei den Wahlaufgaben

Die Gegenüberstellung in Tabelle 9 zeigt, dass z.B. im Sachrechnen der Unterschied zwischen den beiden Teilen, ausgedrückt in  $d = 0,01$ , unbedeutend ist. Beim Lesen ( $d = 0,29$ ), in der Arithmetik ( $d = 0,13$ ) und in der Geometrie ( $d = 0,10$ ) gibt es aber einen leichten Vorteil für die Zentralaufgaben. Beim Lesen bestand die Zentralaufgabe darin, eine Wetterkarte zu lesen und anschließend Fragen dazu zu bearbeiten. Dieser Text ist vergleichsweise schnell gelesen und hat einen hohen Lebensbezug für die Kinder (vgl. „Didaktische Erläuterungen“<sup>5</sup>). Die Annahme ist nicht unplausibel, dass wegen der Kürze des Textes die Aufgabe direkt begonnen und wenn möglich zu Ende geführt wird.

In der Arithmetik gibt es bei vielen Kindern erweiterte und fortgeschrittene Leistungen. Es liegt die Vermutung nahe, dass die Fähigkeiten in Arithmetik so umfassend sind, dass der Impuls, Aufgaben zu überspringen, nicht so hoch ist. Die weiter hinten stehenden Wahlaufgaben werden aus Zeit- oder Motivationsgründen möglicherweise nicht mehr so gründlich bearbeitet. In der Geometrie ist dieser Effekt etwas geringer. Beim Sachrechnen finden die Kinder dagegen eher Aufgaben, bei denen möglicherweise generell grundlegende Bearbeitungsstrategien (wie Modellierungsprozesse)

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.

<sup>5</sup> <http://www.uni-landau.de/vera/aufgaben.htm>

fehlen. Hier findet sich somit keine bessere Bearbeitung des Zentralteils. Dies sind allerdings nur Spekulationen.

Das Muster der Leistungsunterschiede in den Zentral- und Wahlaufgaben legt nahe, dass die Bearbeitung der Reihe nach erfolgte (mit entsprechenden Konsequenzen für die weiter hinten stehenden Aufgaben). Die Bearbeitungsreihenfolge war zwar freigestellt: Als Vorbereitung auf die Vergleichsarbeiten sollten vermutlich entsprechende Test-Bearbeitungsstrategien sowie situationsangemessenes Problemlösen durch die Lehrkräfte stärker vermittelt werden.

#### 4.1.3 Unterschiede innerhalb und zwischen Klassen bzw. Schulen

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 10 dargestellten Prozentanteile.

**Tabelle 10:** Zerlegung in die Varianz auf den drei Ebenen Schule, Klasse und Individuum; Angaben in Prozent

Bereich	Schulebene	Klassenebene	Individualebene
Arithmetik	21,9	8,7	69,4
Geometrie	27,8	10,5	61,7
Sachrechnen	16,3	7,1	76,6
Lesen	19,0	6,0	75,0
Schreiben	14,0	7,7	78,3
Sprachbetrachtung	18,6	7,8	73,6
Orthographie	14,2	6,7	79,1
N	3411	7881	170773

Wie erwartet ist die Leistung in den Vergleichsarbeiten vorrangig mit Merkmalen der Individualebene verknüpft. Die Varianzanteile liegen zwischen 61,7 Prozent (Geometrie) und 79,1 Prozent (Orthographie). Ebenfalls erhebliche Anteile gehen auf schulische Merkmale zurück, sie bewegen sich zwischen 14,0 Prozent (Schreiben) und 27,8 Prozent (Geometrie). Den vergleichsweise geringsten Beitrag leistet die Zugehörigkeit zu einer Schulklasse (hinter der Unterschiede des Unterrichts und des Klassenkontextes stehen) mit 6,0 Prozent (Lesen) bis 10,5 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 Sozial-schicht, 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 Geschlechterunterschiede**

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.

**Tabelle 11:** Verteilung der Fähigkeitsniveaus Mathematik, getrennt nach Geschlecht, Angaben in Prozent

		n.a.L.*	FN 1	FN 2	FN 3	N (Kinder)	d**
<b>Arithmetik</b>	Mädchen	2,6	16,8	33,9	46,8	85006	
	Jungen	2,2	14,9	32,1	50,8	86634	-0,09
<b>Geometrie</b>	Mädchen	1,6	16,4	53,4	28,6	85006	
	Jungen	1,6	16,2	51,5	30,7	86633	-0,04
<b>Sachrechnen</b>	Mädchen	5,0	41,1	33,3	20,6	85006	
	Jungen	3,2	30,9	34,2	31,7	86633	-0,29

\* nicht auswertbare Leistung

\*\* Maß für die Effektstärke

**Tabelle 12:** Verteilung der Fähigkeitsniveaus Deutsch, getrennt nach Geschlecht, Angaben in Prozent

		n.a.L.*	FN 1	FN 2	FN 3	N (Kinder)	d**
<b>Lesen</b>	Mädchen	7,0	29,2	35,2	28,6	84446	
	Jungen	7,9	32,8	34,6	24,7	85947	0,10
<b>Schreiben</b>	Mädchen	6,8	10,0	38,4	44,7	84214	
	Jungen	9,7	15,6	41,8	33,0	85647	0,24
<b>Sprachbetrachtung</b>	Mädchen	1,5	17,2	47,7	33,7	84437	
	Jungen	1,6	17,6	49,1	31,8	85930	0,04
<b>Orthographie</b>	Mädchen	1,0	8,9	57,6	32,5	84408	
	Jungen	2,2	13,7	60,1	24,0	85903	0,25

\* nicht auswertbare Leistung

\*\* Maß für die Effektstärke

In Tabelle 11 und Tabelle 12 sind die Geschlechterunterschiede in den jeweiligen Inhaltsbereichen dargestellt. Wie bereits erwähnt, gilt als Maß für die Bedeutsamkeit eines Unterschieds 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.

Bei Betrachtung der geringen Effektstärken wird ersichtlich, dass die Leistungsunterschiede der bei VERA teilnehmenden Schülerinnen und Schüler erwartungsgemäß minimal sind.

In Mathematik zeigt sich ein statistisch bedeutsamer Vorsprung ( $d = -0,29$ ) der Jungen im Bereich Sachrechnen, - dort unterscheiden sich Mädchen und Jungen im höchsten Fähigkeitsniveau um



einen Anteil von 11,1 Prozentpunkten. In Arithmetik und Geometrie ist die Geschlechterverteilung dagegen unerheblich.

In Deutsch ist ebenfalls ein statistisch bedeutsamer Effekt ( $d = 0,24$  und  $0,25$ ) in den Bereichen Schreiben und Orthographie zu verzeichnen – Mädchen schneiden hier geringfügig besser ab. In den Bereichen Lesen und Sprachbetrachtung zeigen sich hingegen keine bedeutsamen Unterschiede.

#### 4.1.5 Unterschiede zwischen Deutsch als dominante vs. nicht-dominante Sprache

An früherer Stelle wurde bereits erwähnt, dass der Anteil an Kindern mit nicht auswertbarer Leistung möglicherweise im Zusammenhang mit der Sprachbeherrschung steht (siehe 3.4.1, S.23). Die Sprachbeherrschung hängt vermutlich stärker mit der vorherrschende Familiensprache zusammen als mit dem Geburtsort des jeweiligen „nicht-deutschen“ Elternteils. Anhand der Unterscheidung in „Deutsch dominant“ vs. „Deutsch nicht-dominant“ ( $N_{\text{minimal}} = 148730$  bzw.  $21269$ ) 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 dominante 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 hauptsächlich Deutsch spricht bzw. nicht in Deutschland geboren ist, jedoch hauptsächlich Deutsch spricht.

In Abbildung 6 und Abbildung 7 sind die prozentualen Schülerleistungen jeweils nach Deutsch als dominante und nicht-dominante Sprache dargestellt. Tabelle 13 zeigt die Effektstärken der Leistungsunterschiede.

**Tabelle 13:** Effektstärken der Leistungsunterschiede von Deutsch dominante vs. nicht-dominante Sprache

	Arithmetik	Geometrie	Sach-rechnen	Lesen	Schreiben	Sprach-betrachtung	Ortho-graphie	N (Kinder)
<b>Effektstärke (d)</b>	0,39	0,51	0,49	0,84	0,59	0,84	0,53	169999

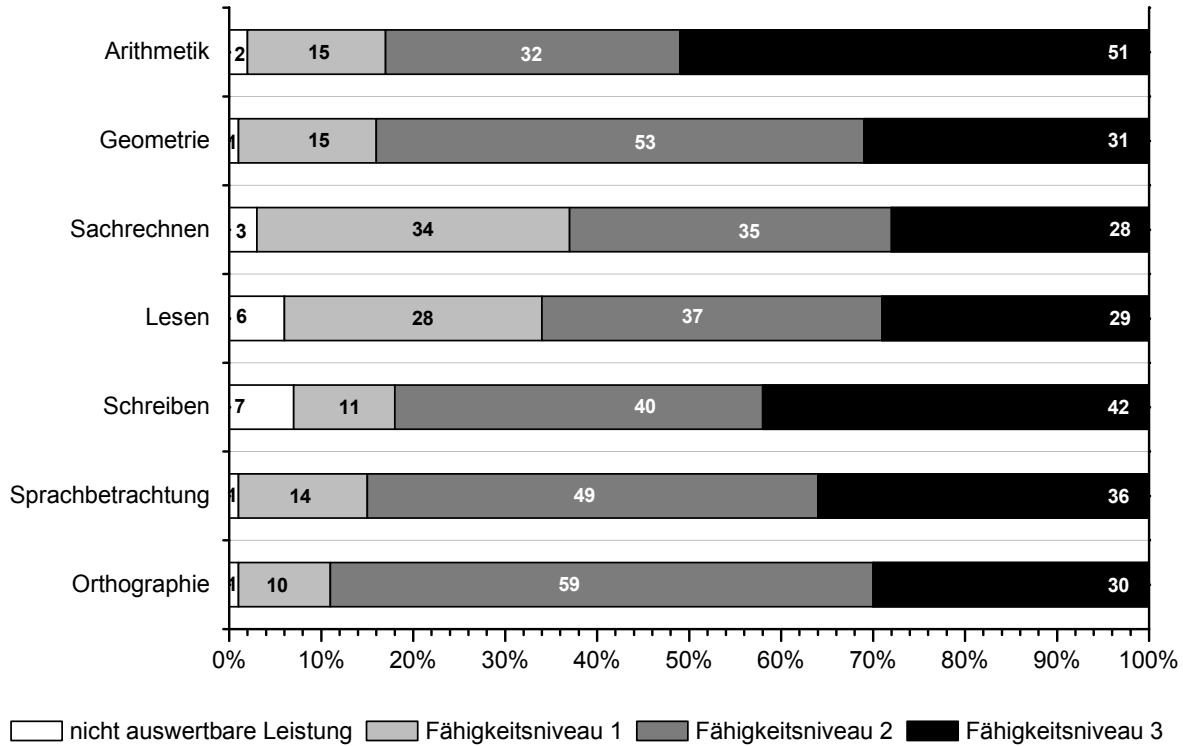


Abbildung 6: Gesamtverteilung der Fähigkeitsniveaus für Deutsch als dominante Sprache

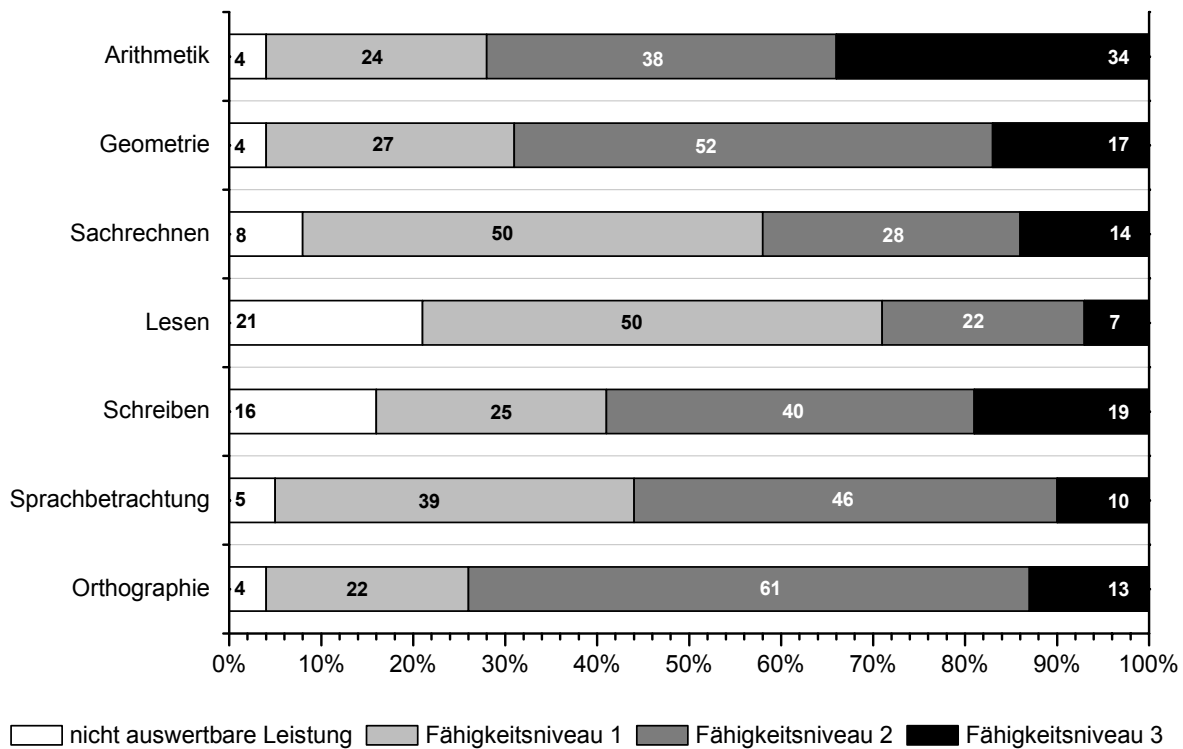


Abbildung 7: Gesamtverteilung der Fähigkeitsniveaus für Deutsch als nicht-dominante Sprache

Es zeigen sich bedeutsame Unterschiede sowohl in Mathematik ( $d = 0,39$  bis  $0,51$ ) als auch in Deutsch ( $d = 0,53$  bis  $0,84$ ). Diese großen Effektstärken repräsentieren die hohen Leistungsunterschiede. 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 4 und 8 Prozent), in Deutsch zeigt sich ein ähnliches Bild. Dabei fällt vor allem der hohe Anteil an Schülerinnen und Schülern (71 Prozent) auf, die in Lesen ein Niveau erreichen, das höchstens das Beherrschen elementarer Aufgaben umfasst. Insbesondere in Lesen und Schreiben kann infolge des hohen Anteils dieser Kinder mit nicht auswertbarer Leistung (21 Prozent in Lesen und 16 Prozent in Schreiben) von einem substantziellen Unterschied gesprochen werden.

Die Ergebnisse bestätigen die Vermutung, dass Merkmale der Sprachherkunft für Schülerinnen und Schüler bereits in der Klassenstufe 4 mit erheblichen Leistungsunterschieden gekoppelt sind (vgl. Schwippert, Bos & Lankes, 2003). Es zeigt sich die Notwendigkeit, für Schülerinnen und Schüler mit nicht-deutschem Sprachhintergrund möglichst früh mit entsprechenden systematischen Fördermaßnahmen zu beginnen.

#### 4.1.6 Diagnosegenauigkeit im Fach Mathematik

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

Man muss diese beiden Aspekte der Diagnosegenauigkeit unbedingt unterscheiden. Eine Lehrerin kann z.B. alle Aufgaben konstant etwas über- oder unterschätzen, aber gleichwohl kann die Rangordnung der von ihr geschätzten Aufgabenschwierigkeit identisch mit der Rangordnung der gelösten Aufgabe sein – oder umgekehrt. Eine hohe Korrelation der geschätzten mit der empirischen Aufgabenrangreihe sagt etwas über die *fachdidaktische* Kompetenz der Lehrkraft aus, ist am ehesten Ausdruck einer zutreffenden Orientiertheit über Schwierigkeitsunterschiede zwischen Aufgaben. Eine geringe Abweichung der durchschnittlichen geschätzten Aufgabenschwierigkeit von der empirischen Schwierigkeit sagt dagegen eher etwas über die *pädagogisch-psychologische* Diagnostik aus, d.h. wie gut die Lehrkraft über das durchschnittliche Leistungsniveau der Klasse im Bilde ist. Wir sprechen deshalb im Folgenden vereinfacht von der fachdidaktischen und der pädagogischen Komponente der Diagnosegenauigkeit. Die beiden Kennwerte sind nicht nur konzeptuell, sondern auch statistisch vollkommen unabhängig voneinander: Die Korrelation beträgt über alle Länder hinweg  $r = ,02$ .

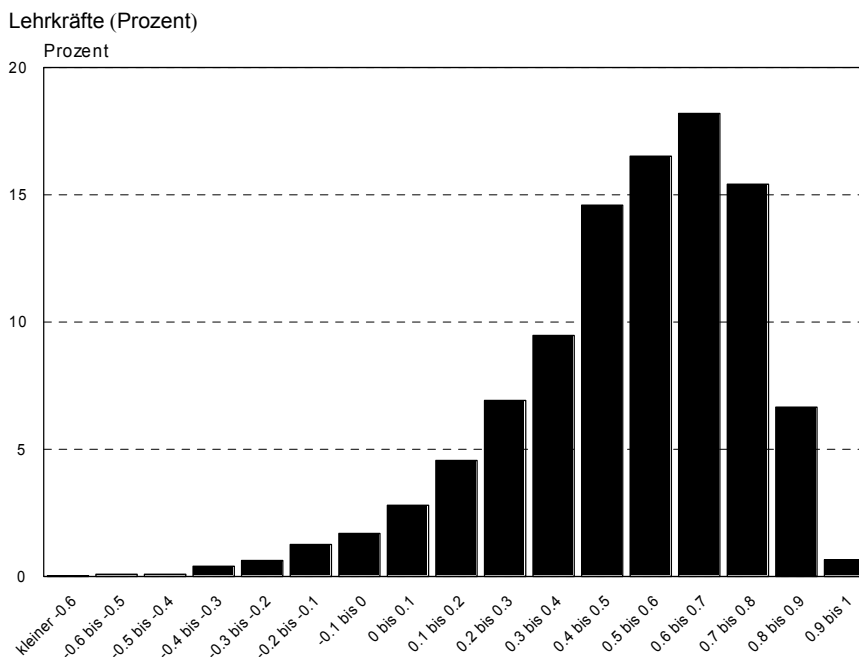
Die Frage der Diagnosegenauigkeit zu vertiefen, würde allerdings den Rahmen dieses Berichtes sprengen. Hierzu liegen nicht nur Handreichungen für die an VERA beteiligten Lehrkräfte, sondern auch Publikationen der VERA-Autoren vor<sup>6</sup>.

*a) Fachdidaktische Komponente: Vergleich der Aufgaben-Rangordnungen*

Im Land Nordrhein-Westfalen liegen verwertbare Angaben von 5718 Lehrkräften zur Diagnosegenauigkeit vor.

Die durchschnittliche Korrelation liegt bei  $r = ,51$ . Wenn man das Auflösungs-niveau erhöht, zeigt sich eine Verteilung wie in Abbildung 8.

Es zeigt sich ein sehr großes Spektrum, von Korrelationen um Null herum (dies ergibt sich, wenn man nach Zufall antwortet) und sogar einigen wenigen negativen Korrelationen, bis hin zu wenigen Klassen, deren Lehrkräfte eine ausgezeichnete Diagnosegenauigkeit aufweisen (Korrelationen zwischen  $r = ,90$  und  $1,00$ ). Korrelationen in Höhe von  $r = ,50$  und höher würden wir als akzeptabel, Korrelationen in Höhe von  $r = ,70$  und höher als gut bezeichnen, d.h. ca. 25 Prozent der Lehrkräfte verfügen über eine gute oder sehr gute Diagnosegenauigkeit, ca. 60 Prozent verfügen über eine akzeptable Diagnosegenauigkeit, aber ca. 40 Prozent der Lehrkräfte liegen mit ihrer Diagnosegenauigkeit z.T. deutlich unterhalb akzeptabler Werte.



**Abbildung 8:** Verteilung der Korrelationskoeffizienten: Ähnlichkeit geschätzter und empirischer Aufgabenschwierigkeiten (fachdidaktischer Aspekt der Diagnosegenauigkeit)

<sup>6</sup> Helmke, A., Hosenfeld, I. & Schrader, F.-W. (2003). Diagnosekompetenz in Ausbildung und Beruf entwickeln. *Karlsruher Pädagogische Beiträge* (55), 15-34.

Helmke, A., Hosenfeld, I. & Schrader, F.-W. (2004). Vergleichsarbeiten als Instrument zur Verbesserung der Diagnosekompetenz von Lehrkräften. In R. Arnold & C. Griese (Hrsg.), *Schulleitung und Schulentwicklung* (S. 119-144). Hohengehren: Schneider-Verlag.

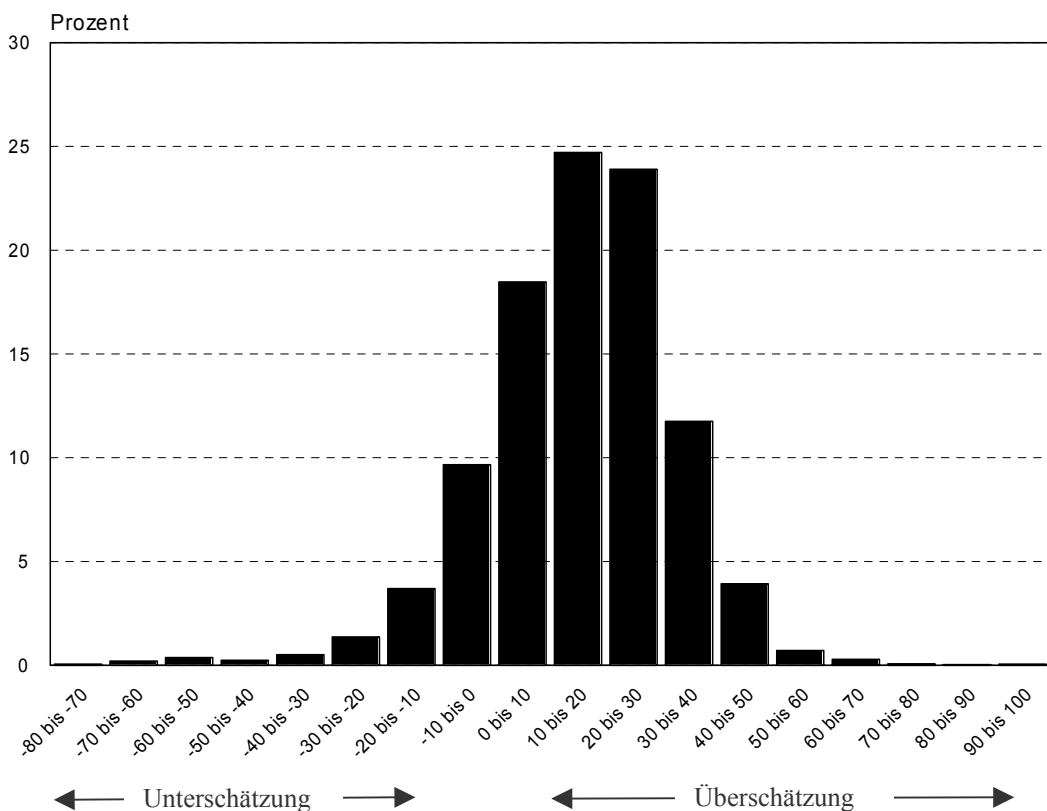
*b) Pädagogische Komponente: Unter- vs. Überschätzungstendenz*

Wir berichten in der folgenden Abbildung das Ausmaß, in dem die Lehrkräfte die Leistungen ihrer Schülerinnen und Schüler unter- oder überschätzen. Hierzu liegen verwertbare Angaben von 5926 Lehrkräften vor.

Abbildung 9 zeigt, dass die Lehrkräfte die Leistungen ihrer Klassen wesentlich häufiger *überschätzen* (im Maximalfall um 60-70 Prozent) als *unterschätzen* (im Maximalfall um 70-60 Prozent). Im Durchschnitt wird die tatsächliche Schülerleistung um 14,6 Prozent überschätzt. Es bleibt abzuwarten, ob die Nutzung der Zusatzauswertungen zur Diagnostischen Kompetenz und die verstärkte Beschäftigung mit Fragen der Pädagogischen Diagnostik in den nächsten Jahren zu einer Verbesserung dieses Musters führen.

Legt man das Ergebnis (die durchschnittliche Einschätzung) unter die Lupe und fragt, in welchem der drei Teilgebiete der Mathematik die geringsten und die größten Überschätzungen stattfinden, dann zeigt sich folgendes: Die Schülerleistungen im Bereich Sachrechnen/Größen werden um 19,6 Prozent überschätzt, die im Geometrie um 13,3 Prozent und die in der Bereich der Arithmetik um 10,9 Prozent. Es ist also ganz eindeutig so, dass vor allem die Schülerfähigkeiten im Bereich von texthaltigen Aufgaben, in dem es um Sachrechnen und Größen geht, seitens der Lehrkräfte überschätzt werden, während dies für die Schülerfähigkeiten im Bereich der Arithmetik am

WLehrkräfte (Prozent)



**Abbildung 9:** Prozentuale Unter- vs. Überschätzung des mathematischen Leistungsniveaus der Klasse (pädagogisch-psychologischer Aspekt der Diagnosegenauigkeit)

## 4.2 Rahmenbedingungen

Die großen Surveys der letzten Jahre, insbesondere PISA und IGLU/PIRLS, haben gezeigt, dass Merkmale des sozialen, ökonomischen und kulturellen Kapitals von Familien einen überwiegenden Einfluss auf die Leistungsfähigkeit der Kinder ausüben. Auf der Ebene von Klassen und Schulen entspricht dies einer wichtigen Rolle des Schuleinzugsgebietes und der Klassenzusammensetzung. Von Schulen „im sozialen Brennpunkt“ spricht man – obwohl es keine verbindliche 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 Verteilung ausgewählter Kontextmerkmale.

Es werden zunächst die auf Klassenebene (Ausnahme: Sprachdominanz) erfragten Daten zum sozioökonomischen und Sprachhintergrund berichtet (Tabelle 14). Anschließend folgt mit Tabelle 15 eine Darstellung der Zusammenhänge zwischen Kontextmerkmalen und Schülerleistungen. Tabelle 16 zeigt die Schülervariablen.

Der Anteil von Schülerinnen und Schülern, für die unabhängig von Nationalität und Geburtsort Deutsch nicht die dominante Sprache ist, beträgt in Nordrhein-Westfalen 11,7 Prozent. Das entspricht durchschnittlich zweieinhalb Schülern pro Klasse. Dabei ergibt sich eine Korrelation mit der Schülerleistung in Lesen von  $r = -,38$ .

Fast 23 Prozent aller beteiligten Klassen liegen im so genannten „sozialen Brennpunkt“. Den Lehrerangaben zufolge sind 8,5 Prozent der Familien ihrer Schülerinnen und Schüler von Arbeitslosigkeit betroffen. 55 Prozent aller Lehrkräfte gaben an, bei der Angabe dieser Einschätzung ziemlich bis absolut sicher gewesen zu sein.

**Tabelle 14: Kontextvariablen (Vergleich NRW – Gesamtstichprobe)\*; durchschnittliche Klassenanteile in Prozent**

	Deutsch nicht dominante Sprache <sup>1,a)</sup>	Schule liegt in sozialem Brennpunkt <sup>2)</sup>	Familie gehört zur Grundschrift <sup>2,b)</sup>	Familie von Arbeitslosigkeit betroffen <sup>2)</sup>	Familie bezieht Sozialhilfe <sup>2)</sup>
<b>NRW</b>	11,7	22,9	25,9	8,5	7,6
<b>Gesamt</b>	11,3	32,2	28,7	15,8	13,9

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

1) erfragt auf Individualebene in der Population, in Brandenburg in der Zentralstichprobe

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 angelehrte Arbeiter, Landarbeiter, alle un- und angelehrten Berufe aus dem manuellen Bereich sowie Dienstleistungstätigkeiten mit weitgehend manuellem Charakter und geringem Anforderungsniveau.“

**Tabelle 15: Zusammenhänge zwischen Kontextmerkmalen und Schülerleistungen (Schulklassenebene)**

	Arithmetik	Geometrie	Sach-rechnen	Lesen	Schreiben	Sprach-betrachtung	Ortho-graphie	min. N
<b>Prozentsatz Grundschrift</b>	-,18	-,24	-,33	-,40	-,31	-,38	-,46	207
<b>Prozentsatz Erhalt von Sozialhilfe</b>	-,15	-,23	-,26	-,32	-,33	-,31	-,34	205
<b>Prozentsatz Arbeitslosigkeit in der Familie</b>	-,16	-,17	-,20	-,30	-,30	-,31	-,33	200
<b>Prozentsatz deutsch nicht dominant</b>	-,19	-,20	-,27	-,38	-,28	-,37	-,31	7879

Bei den Zusammenhängen zwischen Kontextmerkmalen und Schülerleistungen wurden insgesamt niedrige bis mittlere Korrelationen gefunden, wobei die höchsten (zwischen  $r = -,28$  und  $r = -,46$ ) im Bereich Deutsch auftreten.

**Tabelle 16: Schülervariablen (Vergleich NRW – Gesamtstichprobe)\*; durchschnittliche Klassenanteile in Prozent**

	ungenügende Sprachbeherrschung <sup>1,a)</sup>	nicht anwesend beim Deutshtest <sup>1)</sup>	nicht anwesend beim Mathetest <sup>1)</sup>	Klassenwiederholer <sup>1)</sup>
<b>NRW</b>	1,1	3,1	3,3	1,7
<b>Gesamt</b>	2,0	4,5	3,8	1,8

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

1) erfragt auf Individualebene in der Population, in Brandenburg in der Zentralstichprobe

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

#### 4.2.2 Verteilung der Zentralstichprobe auf die Kontextgruppen

In Tabelle 17 ist die prozentuale Verteilung der Zentralstichprobe auf die Kontextgruppen dargestellt (zur Bildung der Kontextgruppen siehe Kapitel 3.3, S.20).

*Tabelle 17: Verteilung auf die Kontextgruppen (Vergleich NRW – Gesamtstichprobe), Angaben in Prozent*

	Gruppe 1	Gruppe 2	Gruppe 3	Gruppe 4	N (Klassen)
NRW	23,8	34,7	27,2	14,4	202
Gesamt	19,2	29,1	29,4	22,3	992

Die Verteilung der Zentralstichprobe auf die Kontextgruppen fällt außer für die Gruppe vier nahezu kongruent aus. Die Kontextgruppe vier ist als Indikator für eher ungünstige Bedingungen hier mit einem Anteil von 14,4 Prozent deutlich unterdurchschnittlich vertreten.

### 4.3 Lehrerfragebogen (ausgewählte Ergebnisse)

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 2004 zumindest mittelfristig zu einer erhöhten Kooperation zwischen den beteiligten Lehrkräften geführt hat.

#### 4.3.1 Kontinuität des Unterrichts in Mathematik und Deutsch und Fähigkeitsniveaus

Klassen in Schulen mit *regulärem* Lehrerwechsel nach der zweiten Jahrgangsstufe zeigen in einigen Bereichen (Geometrie, Sachrechnen, Lesen) eine deutliche Überrepräsentation des Fähigkeitsniveaus zwei, welche mit einer geringeren Anzahl Schüler im unteren Fähigkeitspektrum und im Lesen auch mit weniger Schülern auf höchstem Fähigkeitsniveau einhergeht. Das insgesamt recht günstige Profil dieser Gruppe relativiert sich jedoch bei Betrachtung wichtiger Kontextvariablen (Anteil Grundschrift, Arbeitslose, Sozialhilfeempfänger, Deutsch nicht dominant): Sie weist hier verhältnismäßig günstigere Bedingungen auf als die beiden anderen Gruppen. In Klassen aus Schulen mit regulärem Lehrerwechsel nach der dritten Klasse finden sich fast durchgängig weniger Schüler auf dem höchsten Fähigkeitsniveau, bei gleichzeitig höherem Anteil auf Fähigkeitsniveau eins. Diese Konstellation geht nicht auf benachteiligende Kontextbedingungen zurück, möglicherweise also wirken sich zeitlich kurz zurückliegende Lehrerwechsel tatsächlich eher ungünstig auf die Leistungen der Schüler aus. Einschränkend muss hier allerdings auf die Geometrie verwiesen werden: Hier schneiden die Klassen mit Lehrerwechsel nach der dritten Jahrgangsstufe deutlich besser ab als die Klassen der anderen beiden Gruppen. Weiterhin sind die Befunde zur Gruppe mit Wechsel nach der 3. Klasse aufgrund der geringen Fallzahl (N=8) wenig repräsentativ.



**Tabelle 18:** Lehrerwechsel während der ersten vier Grundschuljahre im Fach Mathematik, Angaben in Prozent

Bereich		n.a.L.*	FN 1	FN 2	FN 3	N (Klassen)
Arithmetik	kein Lehrerwechsel	2,9	15,9	31,9	49,3	134
	Lehrerwechsel nach der 2. Klasse	1,5	13,2	36,6	48,7	23
	Lehrerwechsel nach der 3. Klasse	4,2	15,7	36,8	43,3	8
Geometrie	kein Lehrerwechsel	1,9	16,7	50,1	31,2	134
	Lehrerwechsel nach der 2. Klasse	0,6	8,9	62,5	28,0	23
	Lehrerwechsel nach der 3. Klasse	0,6	12,6	43,9	42,9	8
Sachrechnen	kein Lehrerwechsel	3,9	35,4	34,1	26,6	134
	Lehrerwechsel nach der 2. Klasse	1,7	28,0	43,8	26,5	23
	Lehrerwechsel nach der 3. Klasse	4,5	37,0	35,5	23,0	8

\* nicht auswertbare Leistung

**Tabelle 19:** Lehrerwechsel während der ersten vier Grundschuljahre im Fach Deutsch, Angaben in Prozent

Bereich		n.a.L.*	FN 1	FN 2	FN 3	N
Lesen	kein Lehrerwechsel	7,3	29,7	35,1	28,0	134
	Lehrerwechsel nach der 2. Klasse	4,4	26,3	45,2	24,1	23
	Lehrerwechsel nach der 3. Klasse	9,1	44,4	32,9	13,6	8
Schreiben	kein Lehrerwechsel	8,6	13,0	40,3	38,2	134
	Lehrerwechsel nach der 2. Klasse	7,0	8,8	43,7	40,5	23
	Lehrerwechsel nach der 3. Klasse	9,1	14,3	43,7	32,8	8
Sprachbetrachtung	kein Lehrerwechsel	1,7	16,0	49,6	32,7	134
	Lehrerwechsel nach der 2. Klasse	1,4	13,2	54,2	31,3	23
	Lehrerwechsel nach der 3. Klasse	2,3	27,9	46,6	23,2	8
Orthographie	kein Lehrerwechsel	1,5	12,5	57,7	28,4	134
	Lehrerwechsel nach der 2. Klasse	1,1	8,9	56,5	33,5	23
	Lehrerwechsel nach der 3. Klasse	3,0	21,4	54,0	21,6	8

\* nicht auswertbare Leistung

Betrachtet man die Frage nach der Zahl der durch die einzelnen Lehrkräfte unterrichteten Halbjahre, so kann keine einheitliche Tendenz festgestellt werden. Die Klassen, welche seit der zweiten Jahrgangsstufe durch dieselbe Lehrkraft unterrichtet werden, schneiden im Mittel im Bereich Lesen besser, im Bereich Geometrie hingegen deutlich schlechter ab als die anderen Gruppen. Die in den Deutsch-Bereichen durchgängig schlechteren Leistungen der Gruppe mit Wechsel in der 4. Klasse finden keine Entsprechung in der Mathematik und sollten hinsichtlich ihrer Implikationen vorsichtig beurteilt werden: Von einer unmittelbaren kausalen Interpretation ist abzusehen, da unklar bleibt, ob schwächere Klassen wegen eines Lehrerwechsels schlechtere Leistungen zeigen, oder Lehrerwechsel wegen der Leistungsdefizite gerade in schwächeren Klassen vollzogen werden.

**Tabelle 20:** Durch Lehrer unterrichtete Halbjahre im Fach Mathematik, Angaben in Prozent

<b>Bereich</b>		<b>n.a.L.*</b>	<b>FN 1</b>	<b>FN 2</b>	<b>FN 3</b>	<b>N (Klassen)</b>
Arithmetik	Beginn / Mitte 1. Klasse	3.16	15.37	32.67	48.81	118
	Beginn / Mitte 2. Klasse	1.94	17.93	36.10	44.03	17
	Beginn / Mitte 3. Klasse	1.82	16.29	33.25	48.63	48
	Beginn 4. Klasse	1.58	16.79	30.83	50.80	29
Geometrie	Beginn / Mitte 1. Klasse	1.72	15.60	49.70	32.99	118
	Beginn / Mitte 2. Klasse	4.80	22.18	54.02	19.00	17
	Beginn / Mitte 3. Klasse	1.27	13.46	57.26	28.01	48
	Beginn 4. Klasse	0.30	16.60	57.74	25.36	29
<b>Bereich</b>		<b>n.a.L.*</b>	<b>FN 1</b>	<b>FN 2</b>	<b>FN 3</b>	<b>N (Klassen)</b>
Sachrechnen	Beginn / Mitte 1. Klasse	4.02	36.06	33.56	26.35	118
	Beginn / Mitte 2. Klasse	4.26	38.98	30.32	26.43	17
	Beginn / Mitte 3. Klasse	3.89	32.29	37.69	26.12	48
	Beginn 4. Klasse	5.65	29.33	35.82	29.20	29

\* nicht auswertbare Leistung

**Tabelle 21:** Durch Lehrer unterrichtete Halbjahre im Fach Deutsch, Angaben in Prozent

<b>Bereich</b>		<b>n.a.L.*</b>	<b>FN 1</b>	<b>FN 2</b>	<b>FN 3</b>	<b>N (Klassen)</b>
Lesen	Beginn / Mitte 1. Klasse	6.88	29.65	36.72	26.75	136
	Beginn / Mitte 2. Klasse	5.47	31.18	29.90	33.45	14
	Beginn / Mitte 3. Klasse	5.10	27.27	41.36	26.27	40
	Beginn 4. Klasse	9.38	37.80	32.72	20.10	21
Schreiben	Beginn / Mitte 1. Klasse	8.55	12.76	39.56	39.13	136
	Beginn / Mitte 2. Klasse	6.72	14.30	42.83	36.14	14
	Beginn / Mitte 3. Klasse	6.17	11.02	42.52	40.29	40
	Beginn 4. Klasse	6.54	20.54	45.09	27.83	21

Bereich		n.a.L.*	FN 1	FN 2	FN 3	N (Klassen)
Sprachbetrachtung	Beginn / Mitte 1. Klasse	1.70	15.74	48.71	33.86	136
	Beginn / Mitte 2. Klasse	0.34	20.16	47.03	32.47	14
	Beginn / Mitte 3. Klasse	1.61	12.48	53.90	32.02	40
	Beginn 4. Klasse	3.48	21.42	52.64	22.46	21
Orthographie	Beginn / Mitte 1. Klasse	1.56	13.22	57.35	27.86	136
	Beginn / Mitte 2. Klasse	1.10	9.18	59.39	30.33	14
	Beginn / Mitte 3. Klasse	1.42	9.77	58.29	30.52	40
	Beginn 4. Klasse	2.48	12.05	60.22	25.24	21

\* 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. Für diese Vorgehensweise spricht auch die Tatsache, dass die Verteilungen der Unterrichtserfahrung in Abhängigkeit von den Fähigkeitsniveaus das gleiche Bild widerspiegeln, wie es bereits im Zusammenhang mit der länderspezifischen Verteilung der Fähigkeitsniveaus (siehe Abschnitt 4.1.1) dargestellt wurde.

**Tabelle 22:** Durch die Lehrkraft unterrichtete Jahre im Fach Mathematik, Angaben in Prozent

Bereich		n.a.L.*	FN 1	FN 2	FN 3	N (Klassen)
Arithmetik	weniger als 2 Jahre	1,9	19,3	27,2	51,7	8
	2 - 5 Jahre	3,2	16,3	33,4	47,0	27
	6 - 10 Jahre	1,5	15,9	36,4	46,2	26
	11 - 15 Jahre	4,1	16,9	31,0	48,1	29
	16 - 20 Jahre	2,5	16,5	34,8	46,2	15
	mehr als 20 Jahre	2,3	15,3	32,5	49,9	107
Geometrie	weniger als 2 Jahre	1,5	16,9	59,0	22,6	8
	2 - 5 Jahre	3,3	14,1	50,4	32,2	27
	6 - 10 Jahre	2,0	17,0	55,5	25,5	26
	11 - 15 Jahre	2,1	19,0	52,0	26,9	29
	16 - 20 Jahre	2,4	15,6	51,6	30,4	15
	mehr als 20 Jahre	1,0	15,0	52,8	31,3	107

<b>Bereich</b>		<b>n.a.L.*</b>	<b>FN 1</b>	<b>FN 2</b>	<b>FN 3</b>	<b>N (Klassen)</b>
Sachrechnen	weniger als 2 Jahre	5,1	30,4	37,9	26,6	8
	2 - 5 Jahre	5,4	32,6	34,6	27,3	27
	6 - 10 Jahre	4,5	35,9	31,5	28,1	26
	11 - 15 Jahre	5,2	32,0	34,9	27,9	29
	16 - 20 Jahre	3,2	35,1	35,2	26,6	15
	mehr als 20 Jahre	3,7	35,6	34,8	25,9	107

\* nicht auswertbare Leistung

**Tabelle 23:** Durch die Lehrkraft unterrichtete Jahre im Fach Deutsch, Angaben in Prozent

<b>Bereich</b>		<b>n.a.L.*</b>	<b>FN 1</b>	<b>FN 2</b>	<b>FN 3</b>	<b>N (Klassen)</b>
Lesen	weniger als 2 Jahre	3,0	24,0	36,3	36,8	7
	2 - 5 Jahre	7,7	34,1	40,6	17,6	28
	6 - 10 Jahre	7,1	27,5	33,7	31,7	21
	11 - 15 Jahre	6,9	28,6	35,1	29,4	33
	16 - 20 Jahre	5,0	28,6	35,6	30,7	16
	mehr als 20 Jahre	6,8	30,7	37,1	25,5	106
Schreiben	weniger als 2 Jahre	6,0	8,7	47,2	38,1	7
	2 - 5 Jahre	6,6	14,8	40,6	38,1	28
	6 - 10 Jahre	8,8	13,5	42,9	34,7	21
	11 - 15 Jahre	6,8	14,5	39,0	39,7	33
	16 - 20 Jahre	7,1	7,0	35,2	50,7	16
	mehr als 20 Jahre	8,4	13,8	41,6	36,2	106
Sprachbetrachtung	weniger als 2 Jahre	0,5	10,1	50,9	38,6	7
	2 - 5 Jahre	2,0	17,4	50,9	29,7	28
	6 - 10 Jahre	3,3	16,8	51,6	28,4	21
	11 - 15 Jahre	2,1	12,6	49,5	35,8	33
	16 - 20 Jahre	1,6	13,6	46,3	38,5	16
	mehr als 20 Jahre	1,4	17,2	50,1	31,3	106
Orthographie	weniger als 2 Jahre	2,3	8,7	63,3	25,7	7
	2 - 5 Jahre	2,3	13,8	62,7	21,2	28
	6 - 10 Jahre	2,3	13,1	58,1	26,5	21
	11 - 15 Jahre	2,0	10,6	56,0	31,5	33
	16 - 20 Jahre	1,7	8,3	58,1	31,9	16
	mehr als 20 Jahre	1,1	12,9	56,9	29,1	106

\* nicht auswertbare Leistung

**Tabelle 24:** Mathematik als grundständiges Fach, Angaben in Prozent

Bereich		n.a.L.*	FN 1	FN 2	FN 3	N (Klassen)
Arithmetik	nein	2,85	15,26	33,10	48,78	70
	ja	2,39	16,50	32,61	48,50	138
Geometrie	nein	1,40	14,79	54,05	29,77	70
	ja	1,76	16,05	51,98	30,22	138
Sachrechnen	nein	2,62	34,00	36,02	27,36	70
	ja	1,76	16,05	51,98	30,22	138

\* nicht auswertbare Leistung

**Tabelle 25:** Deutsch als grundständiges Fach, Angaben in Prozent

Bereich		n.a.L.	FN 1	FN 2	FN 3	N (Klassen)
Lesen	nein	5,95	28,74	36,91	28,40	51
	ja	6,98	30,52	36,66	25,85	159
Schreiben	nein	6,56	11,35	38,73	43,35	51
	ja	8,22	13,70	41,75	36,33	159
Sprachbetrachtung	nein	1,05	15,46	50,69	32,80	51
	ja	2,01	16,24	49,79	31,96	159
Orthographie	nein	1,43	11,62	57,33	29,63	51
	ja	1,66	12,40	58,11	27,83	159

\* nicht auswertbare Leistung

### 4.3.3 Kooperation bei der Aufgabenauswahl und -auswertung

Neben Informationen zum sozialen Kontext wurde im Lehrerfragebogen auch erfragt, inwiefern die Auswahl der Wahlaufgaben sowie die Auswertung der Vergleichsarbeiten in Kooperation mit Kolleg/innen erfolgte. Da die Kooperation (z.B. zur Besprechung der VERA-Ergebnisse) für die Tiefe der Verarbeitung wichtig ist und gerade die Auseinandersetzung mit den Inhalten und Konzepten der Vergleichsarbeiten den Blick auf Ansatzpunkte für Unterrichtsentwicklung eröffnet, wurde eine Kooperation durch das Prozedere der Aufgabenauswahl (pro Schule nur eine Aufgabekombination erlaubt) und der Vergleichsangebote in den Ergebnisrückmeldungen forciert. Die Antworten der Lehrkräfte zu diesen Fragen kann als Hinweis zu den schulischen Bedingungen, insbesondere dem Evaluations- und Kooperationsklima interpretiert werden. Tabelle 26 belegt, dass bei der *Aufgabenauswahl* in beiden Fächern in den meisten Fällen kooperiert wurde. Demgegenüber erfolgte die *schulinterne Auswertung* deutlich seltener in Zusammenarbeit mit Kolleg/innen. Die von VERA intendierte Kooperation im Zusammenhang mit den Vergleichsarbeiten ist hier noch nicht überzeugend gelungen. Ein Grund könnte sein, dass die Lehrkräfte sich für die zeitintensivere Auswertung außerhalb des Unterrichts verabreden mussten. Dies lässt sich mit dem Arbeitsalltag von Lehrkräften in vielen Fällen vermutlich 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.

**Tabelle 26:** Kooperation im Rahmen der Vergleichsarbeiten; Angaben in Prozent

		ja	nein	N (Klassen)
Aufgabenauswahl im Team	Deutsch	88,1	11,9	210
	Mathematik	87,1	12,9	209
Auswertung im Team	Deutsch	27,4	72,6	208
	Mathematik	22,0	78,0	209

## 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 in den Folgejahren im Interesse einer Verringerung der Arbeit für die beteiligten Lehrkräfte noch weiter ausgebaut und zugleich vereinfacht werden wird. Die Konzeption der Vergleichsarbeiten in den kommenden Jahren sieht eine noch stärkere Orientierung an den inzwischen gültigen bundesländerübergreifenden Bildungsstandards vor.

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 Unternehmen, das als jährlicher Zyklus angelegt ist, lässt sich erst dann beurteilen, wenn belegbar ist, ob und in welchem Umfang 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) beigetragen hat.

Wir denken, dass das Spektrum der genannten Ziele sehr breit ist und betonen an dieser Stelle nochmals mit Nachdruck, dass die Bereitstellung vielfältiger Ergebnisdarstellungen und Vergleichsmöglichkeiten letztendlich der Verbesserung des Lehrens und Lernens dienen soll. Dagegen sind Ländervergleiche bei VERA aus methodischen Gründen mit größter Vorsicht zu behandeln.

Es reicht nach unserer Ansicht allerdings nicht aus, eine solche Zielerreichung bloß zu erhoffen, geschweige denn, sie bereits als gegeben anzusehen. Seit der „empirischen Wende“ gilt der Grundsatz der Wirkungsorientierung: Das Ausmaß der Erreichung (oder Verfehlung) bildungspolitischer Ziele muss nachgewiesen, also empirisch überprüft werden. Erste Schritte in diese Richtung haben wir in Gestalt von ausführlichen internetbasierten Lehrerbefragungen zur Rezeption und zum Nutzen von VERA, deren zweite soeben begonnen hat, bereits getan. Über die Ergebnisse dieser Surveys und über mögliche Konsequenzen daraus - sowohl für die Vergleichsarbeiten selbst als auch für Rahmenbedingungen und Unterstützungssysteme wird an anderer Stelle ausführlich zu berichten sein.

## 6 Literatur

- Artelt, C., Schneider, W., & Schiefele, U. (2002). Ländervergleich zur Lesekompetenz. In J. Baumert, C. Artelt, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele, W. Schneider, K.-J. Tillmann & M. Weiß (Hrsg.), *PISA 2002 - Die Länder der Bundesrepublik Deutschland im Vergleich*. Opladen: Leske + Budrich.
- Baumert, J., Artelt C., Klieme, E., Neubrand, M., Prenzel, M., Schiefele, U., Schneider, W., Tillmann, K.-J. & Weiß M. (Hrsg.) (2002): *PISA 2000 - Die Länder der Bundesrepublik Deutschland im Vergleich*. Opladen: Leske + Budrich.
- Blum, W., Neubrand, M., Ehmke, T., Senkbeil, M., Jordan, A., Ulfig, F & Carstensen, C. H. (2004). Mathematische Kompetenz. In M. Prenzel, J. Baumert, W. Blum, R. Lehmann, D. Leutner, M. Neubrand, R. Pekrun, H.-G. Rolff, J. Rost & U. Schiefele (Hrsg.), *PISA 2003: Der Bildungsstandard der Jugendlichen in Deutschland – Ergebnisse des zweiten internationalen Vergleichs* (S. 47-92). Münster: Waxmann Verlag.
- Bos, W., Lankes, E.-M, Prenzel, M., Schwippert, K., Valtin R. & Walther, G. (Hrsg.) (2004). *IGLU: Einige Länder der Bundesrepublik Deutschland im nationalen und internationalen Vergleich*. Münster: Waxmann Verlag.
- Ditton, H. (1990). Anmerkungen zum praktischen Umgang mit Anteilen erklärter Varianz. *Empirische Pädagogik*, 4, 289-309.
- Helmke, A. & Hosenfeld, I. (2004). Vergleichsarbeiten - Kompetenzmodelle - Standards. In M. Wosnitza, A. Frey & R. S. Jäger (Hrsg.), *Lernprozesse, Lernumgebungen und Lerndiagnostik. Wissenschaftliche Beiträge zum Lernen im 21. Jahrhundert* (S. 56-75). Landau: Verlag Empirische Pädagogik.
- Helmke, A., Hosenfeld, I. & Schrader, F.-W. (2004). Vergleichsarbeiten als Instrument zur Verbesserung der Diagnosekompetenz von Lehrkräften. In R. Arnold & C. Griese (Hrsg.), *Schulleitung und Schulentwicklung* (S. 119-144). Hohengehren: Schneider-Verlag.
- Helmke, A., Hosenfeld, I. & Schrader, F.-W. (2003). *Diagnosekompetenz in Ausbildung und Beruf entwickeln*. Karlsruher Pädagogische Beiträge (55).
- Helmke, A., Hosenfeld, I., Schrader, F.-W. & Wagner, W. (2002). Sozialer und sprachlicher Hintergrund. In A. Helmke.& R. S. Jäger (Hrsg.), *Das Projekt Markus: Mathematikgesamterhebung Rheinland-Pfalz: Kompetenzen, Unterrichtsmerkmale, Schulkontext* (S. 71-154). Landau: Verlag Empirische Pädagogik.
- Helmke, A., Hosenfeld, I. & Schrader, F.-W. (2002). Unterricht, Mathematikleistung und Lernmotivation. In A. Helmke.& R. S. Jäger (Hrsg.), *Das Projekt Markus: Mathematikgesamterhebung Rheinland-Pfalz: Kompetenzen, Unterrichtsmerkmale, Schulkontext* (S. 413-480). Landau: Verlag Empirische Pädagogik.
- Helmke, A. & Reich, H. H. (2001). *Die Bedeutung der sprachlichen Herkunft für die Schulleistung*. *Empirische Pädagogik*, 15(4), S. 567-600.
- Helmke, A. & Weinert F. E. (1997). Bedingungsfaktoren schulischer Leistungen. In F. E. Weinert (Hrsg.), *Psychologie des Unterrichts und der Schule* (Enzyklopädie der Psychologie, Serie Pädagogische Psychologie, Bd. 3). S. 71-176. Göttingen: Hogrefe.
- Hosenfeld, I., Helmke, A., Ridder, A., & Schrader, F.-W. (2001). Eine mehrbenenanalytische Betrachtung von Schul- und Klasseneffekten. *Empirische Pädagogik*, 15(4), 513-534.
- Hosenfeld, I., Helmke, A., Ridder, A. & Schrader, F.-W. (2002). Die Rolle des Kontextes. In A. Helmke & R. S. Jäger (Hrsg.), *Das Projekt MARKUS. Mathematik-Gesamterhebung*

- Rheinland-Pfalz: Kompetenzen, Unterrichtsmerkmale, Schulkontext* (S. 155-256). Landau: Verlag Empirische Pädagogik.
- Klieme, E., Avenarius, H., Blum, W., Döbrich, P., Gruber, H., Prenzel, M., Reiss, K., Riquarts, K., Rost, J., Tenorth, H.-E. & Vollmer, H. J. (2003). *Zur Entwicklung nationaler Bildungsstandards. Eine Expertise*. Frankfurt a. M.: DIPF.
- Rasch, G. (1960). *Probabilistic Models for Some Intelligence and Attainment Tests* (Studies in Mathematical Psychology). Copenhagen: Nielsen & Lydiche.
- Schwippert, K., Bos, W. & Lankes, E.-M. (2003). Heterogenität und Chancengleichheit am Ende der vierten Jahrgangsstufe im internationalen Vergleich. In W. Bos, E.-M Lankes M., Prenzel, K. Schwippert, G. Walther, & R. Valtin (Hrsg.). *Erste Ergebnisse aus IGLU* (S. 265-302). Münster: Waxmann Verlag.
- Wu, M. L., Adams, R. J. & Wilson, M. R. (1998). *ConQuest - Generalised Item Response Modelling Software*. Melbourne: Australian Council for Educational Research (ACER).
- Zimmer, K., Burba, D. & Rost, J. (2004). Kompetenzen von Jungen und Mädchen. In M. Prenzel, J. Baumert, W. Blum, R. Lehmann, D. Leutner, M. Neubrand, R. Pekrun, H.-G. Rolff, J. Rost & U. Schiefele (Hrsg.), *PISA 2003: Der Bildungsstandard der Jugendlichen in Deutschland – Ergebnisse des zweiten internationalen Vergleichs* (S. 211-222). Münster: Waxmann Verlag.



## 7 Anhang

Allgemein:

- Grundauszählung des Lehrerfragebogens (Zentralstichprobe)
- Druckversion des Lehrerfragebogens
- Elternfragebogen
- Online-Eingabemaske für Schülerdaten (WWW)
- Offline-Eingabemaske für Schülerdaten (Excel)
  
- Handreichung zur Durchführung
- Handreichung zur Analyse der Falschlösungen
- Handreichung zur Pädagogischen Nutzung
- Vorlage zur Rückmeldung der Ergebnisse an die Eltern

Mathematik:

- Beschreibung der Fähigkeitsniveaus
- Handreichung zur Diagnostischen Kompetenz
- Didaktische Erläuterungen der Zentralaufgaben
- Lehrplanbezüge der Zentralaufgaben

Deutsch:

- Beschreibung der Fähigkeitsniveaus
- Didaktische Erläuterungen der Zentralaufgaben
- Lehrplanbezüge der Zentralaufgabe