



**Halbjahresschrift für die Didaktik
der deutschen Sprache und
Literatur**

<http://www.didaktik-deutsch.de>
12. Jahrgang 2007 – ISSN 1431-4355
Schneider Verlag Hohengehren
GmbH

*Andreas Voss
Inge Blatt
Kerstin Kowalski*

**ZUR ERFASSUNG
ORTHOGRAPHISCHER
KOMPETENZ IN IGLU 2006:
DARGESTELLT AN EINEM
SPRACHSYSTEMATISCHEN
TEST AUF GRUNDLAGE VON
DATEN AUS DER IGLU-
VORUNTERSUCHUNG**

In: Didaktik Deutsch. Jg. 12. H. 23. S. 15-33.

Die in der Zeitschrift veröffentlichten Beiträge sind urheberrechtlich geschützt. Alle Rechte, insbesondere das der Übersetzung in fremde Sprachen, vorbehalten. Kein Teil dieser Zeitschrift darf ohne schriftliche Genehmigung des Verlages in irgendeiner Form – durch Fotokopie, Mikrofilm oder andere Verfahren – reproduziert oder in eine von Maschinen, insbesondere von Datenverarbeitungsanlagen, verwendbare Sprache übertragen werden. – Fotokopien für den persönlichen und sonstigen eigenen Gebrauch dürfen nur von einzelnen Beiträgen oder Teilen daraus als Einzelkopien hergestellt werden.

Andreas Voss
Inge Blatt
Kerstin Kowalski

ZUR ERFASSUNG ORTHOGRAPHISCHER KOMPETENZ IN IGLU 2006: DARGESTELLT AN EINEM SPRACHSYSTEMATISCHEN TEST AUF GRUND- LAGE VON DATEN AUS DER IGLU-VORUNTERSUCHUNG

1 Einleitung

Der Erfolg des Rechtschreibunterrichts an deutschen Schulen steht nach Ergebnissen empirischer Untersuchungen in einem Missverhältnis zum zeitlichen Aufwand. Die Hälfte der Grundschul Kinder in Jahrgangsstufe vier investiert mindestens zwei, ein Drittel sogar drei Unterrichtsstunden pro Woche für das Erlernen der deutschen Orthographie.

Die Ergebnisse der Ergänzungsstudie zu der im Jahre 2001 durchgeführten *Internationalen Grundschul-Leseuntersuchung* (IGLU-E), in der die Rechtschreibkompetenz von Viertklässlern erhoben wurde, zeigen, dass trotz dieser zeitintensiven Bemühungen ein erheblicher Anteil der Kinder nicht von diesem Unterricht profitieren kann (vgl. Valtin, Badel, Löffler, Meyer-Schepers & Voss, 2003; Valtin, Löffler, Meyer-Schepers & Badel, 2004; Löffler & Meyer-Schepers, 2005). Auch die Untersuchung von Schülertexten im Rahmen der IGLU-E-Aufsatzstudie hat eine zum Teil gravierende Beeinträchtigung der Lesbarkeit aufgrund von Rechtschreibfehlern ergeben (Blatt, Voss & Matthießen, 2005, S. 143-151).

Ein besonders bedenkliches Zeichen ist die im Laufe der Grundschulzeit beträchtliche Zunahme der Variantenschreibungen, die in allen Leistungserhebungen nachgewiesen wurde (vgl. Zusammenstellung von Variantenschreibung bei Schönweiss). Dies lässt auf eine erhebliche Verunsicherung der Kinder durch den Rechtschreibunterricht schließen.

Diese Befunde haben dazu geführt, dass der Erfassung orthographischer Kompetenz in der zweiten IGLU-Studie (2006) eine erhöhte Aufmerksamkeit zukam. Während im ersten IGLU-Zyklus aufgrund eines rotierten Testdesigns nur eine Teilstichprobe in ausgewählten Bundesländern getestet wurde, wurde in der zweiten IGLU-Studie die orthographische Kompetenz von fast 8000 Kindern bundesweit erhoben. Dabei kamen drei Tests mit unterschiedlicher theoretischer Verortung zum Einsatz. Neben dem im Jahr 2001 eingesetzten Test *Dortmunder Schriftkompetenz-Ermittlung* (D-SE), der mittlerweile zum Instrument *gutschrift* weiter entwickelt wurde, wurden die *Deutsche Schreibprobe* (DSP)¹ und ein *sprachsystematischer Rechtschreibtest* (SRT) eingesetzt. Jeder dieser drei Tests basiert auf einem Kompetenzmodell, nach dem jeweils differenzielle orthographische Teilfähigkeiten ausgewiesen werden.

¹ Dieser Test geht auf die Hamburger Schreibprobe (HSP) von Peter May zurück.

Der sprachsystematische Test beruht auf Ergebnissen graphematischer Forschung, und zwar in erster Linie auf Arbeiten von Peter Eisenberg. Die Graphematik betrachtet das Schriftsystem als einen eigenständigen wissenschaftlichen Gegenstand (Eisenberg 1989; Butt & Eisenberg 1990; Eisenberg & Fuhrhop 2007). Sie untersucht die Systeme der gesprochenen und geschriebenen Sprache gesondert und erforscht die Beziehung zwischen beiden Systemen. Dabei geht die Graphematik der Frage nach, „aus welchen Gründen das Schriftsystem in einem gegebenen Fall anders strukturiert ist als das Lautsystem.“ (Eisenberg 1989, S. 59). Sie vertritt die Interdependenzthese. Danach ist „Orthographie als Schriftsystem [...] integraler Bestandteil der Grammatik des Deutschen, sodass das Schriftsystem der Lautsprache gegenüber nicht einfach sekundär ist. Beide, Lautsprache und Schrift, entwickeln sich gemeinsam und nach vergleichbaren Gesetzmäßigkeiten. Orthographie, die sprachlich richtige Schreibung, ist somit nicht das Ergebnis bewusster Festlegungen von „oben her“, nicht das Werk von Grammatikern, sondern hat sich durch den Sprachgebrauch herausgebildet.“ (Hinney & Menzel, 1998, S. 265).

Die graphematische Forschung unterscheidet zwischen einem Kern- und Peripheriebereich der Wortschreibung (vgl. auch im Folgenden Butt & Eisenberg, 1990). Nach den vorliegenden Ergebnissen ist das Schriftsystem und sein Bezug zur gesprochenen Sprache im Kernbereich durch wenige Prinzipien geregelt (phonographisch, silbisch, morphologisch, wortübergreifend). Die graphematische Forschung stellt weiterhin die Bedeutung der Rechtschreibung als Hilfe für den Leser heraus, um den Sinn von Texten schnell erfassen zu können (z. B. Eisenberg, Spitta & Voigt, 1994, S. 15). Damit eröffnet sich eine neue Sicht auf Rechtschreibung, die keinen Selbstzweck mehr hat, sondern eine für die Lese- und Sprachkompetenz funktionale Bedeutung gewinnt (Eisenberg & Fuhrhop, 2007, S. 18).

Die Bedeutung der graphematischen Forschung für die Orthographiedidaktik wird erst in jüngerer Zeit erkannt (Eisenberg & Fuhrhop 2007, S. 16-19). Hinney (1997) legt eine grundlegende Untersuchung zur Passung von sprachwissenschaftlichen Theorieannahmen und lernpsychologischen Bedingungen vor, in der sie für eine „Neuorientierung der Rechtschreibinhalte“ auf graphematischer Basis plädiert.

Bei dem graphematisch basierten sprachsystematischen Rechtschreibtest handelt es sich um eine Neuentwicklung. Damit wird das Ziel verfolgt, den graphematischen Forschungsansatz systematisch auf seine didaktische Eignung hin zu untersuchen. Bislang liegen keine größeren empirischen Untersuchungen dazu vor. In der IGLU-Studie werden die theoretischen Annahmen des Tests mit den erhobenen orthographischen Leistungsdaten der Kinder auf der Grundlage von Modellgeltungstests überprüft. Die Ergebnisse geben Aufschluss darüber, wie reliabel die nach dem theoretischen Modell ausdifferenzierten Teilfähigkeiten erfasst werden können und ob es sich um voneinander unterscheidbare Teilfähigkeiten handelt.

Der sprachsystematische Test wurde bereits in der Voruntersuchung zu IGLU 2006 in 25 Schulen eingesetzt und ausgewertet. Auf Grundlage der Analyseergebnisse wurde er für den Einsatz in IGLU 2006 weiterentwickelt. Im Folgenden wird die Konzeption des sprachsystematischen Rechtschreibtests vorgestellt, die Analyseme-

thoden werden beschrieben und ausgewählte Ergebnisse aus der IGLU-Voruntersuchung berichtet.

2. Sprachsystematisch basierter Rechtschreibtest

2.1 Theoretisches Rahmenkonzept

Die international vergleichenden Leistungsstudien PISA und IGLU haben neue Standards für die Erfassung der Lesekompetenz gesetzt. Die theoretische Rahmenkonzeption dieser Studien und die Untersuchung der Leistungsdaten erlauben einen differenzierten Einblick in unterscheidbare Anforderungs- und Fähigkeitsbereiche der Lesekompetenz (vgl. Campell, Kelly, Mullis, Martin & Sainsbury, 2001; Deutsches PISA-Konsortium, 2000; oder auch Kirsch & Jungeblut, 1986).

Analog dazu wurde für den sprachsystematischen Rechtschreibtest ein Rahmenkonzept auf graphematischer Basis entwickelt. Dieses Modell einer voll ausgebildeten Rechtschreibkompetenz weist fünf Anforderungsbereiche aus (Tabelle 1):

Tabelle 1: Modellierung der Rechtschreibkompetenz nach einem sprachsystematischen Kompetenzmodell

Orientierung an Prinzipien	Phonographisches und silbisches Prinzip im Kernbereich	Morphologisches Prinzip im Kernbereich	Peripheriebereich	Prinzipien der Wortbildung	Wortübergreifendes Prinzip
Teilfähigkeiten	Bezug herstellen zwischen Schrift- und Lautstruktur unter Berücksichtigung der silbenstrukturellen Informationen (Silbenanfangs u. -endrand u. Silbenschnitt)	Vererbte silbenschriftliche Informationen in flektierten + abgeleiteten Formen herleiten; Flexionsmorpheme kennen u. anwenden	Markierungen in offenen Silben setzen und vererbte Schreibweisen herleiten; Transfer bei Sonderfällen und Lernwörtern; Fremdwortschreibung	Wortarten u. Wortbildungsmorpheme kennen und in Ableitungen und Komposita produktiv anwenden	Wortarten u. Syntaxstrukturen kennen u. für Groß-/Klein-, dass-Schreibung u. Kommasetzung anwenden

In dem Modell wird eine Unterscheidung getroffen zwischen Anforderungsbereichen auf der Ebene der Schreibung einfacher und flektierter Wörter und Wortformen, der Schreibung abgeleiteter Wörter und der Schreibung auf Grundlage des wortübergreifenden Prinzips.

Der Kernbereich der Rechtschreibung stellt den umfangsmäßig größten Teil der Wortschreibungen dar, der verlässlich nach dem phonographischen, silbischen und morphologischen Prinzip geregelt ist (vgl. Eisenberg & Fuhrhop, 2007). Der Peripheriebereich umfasst dagegen einen zahlenmäßig geringen Bereich von Ausnahmen. Dazu zählen das sog. Dehnungs-h, die Verdoppelung des Vokalgraphems, nicht ableitbare einsilbige Wörter, unregelmäßige Wortschreibungen und Fremdwörter. Während die Schreibungen im Kernbereich auf Verstehens- und Wissenstrans-

ferbasis gelernt werden können, müssen die Wörter im Peripheriebereich überwiegend durch Üben eingepägt werden.

Die Rechtschreibung von abgeleiteten Wörtern verlangt die Fähigkeit zur Analyse von Stamm- und Wortbildungsmorphemen. Das wortübergreifende Prinzip bezieht sich auf die Groß- und Kleinschreibung und die das/dass-Schreibung. Die Großschreibung kann wortartenbezogen oder syntaxbezogen erklärt werden (vgl. Röber-Siekmeyer, 1999, S. 60 ff.). Die das/dass-Schreibung und die Kommasetzung setzen eine syntaktische Analyse voraus.²

Wie sich die Teilfähigkeiten einer voll ausgebildeten Rechtschreibkompetenz aufbauen, dazu liegen derzeit keine wissenschaftlich überprüften Erkenntnisse vor. Die vorherrschende Auffassung von einem stufenweisen Aufbau, wonach sich Wissen, Fähigkeiten und Fertigkeiten in aufeinanderfolgenden und sich ablösenden Stufen mit steigendem Schwierigkeitsgrad ausbilden, wird zunehmend in Frage gestellt (vgl. Bremerich-Vos, 2004, S. 85 f.).

Ein hierarchisch-integratives Modell, wie es Bereiter (1980) für die Schreibentwicklung vorlegt, steht für den Aufbau von Rechtschreibkompetenz noch aus. Es beruht auf der Annahme, dass Lernende bei einer erfolgreichen Kompetenzentwicklung den Lerngegenstand bereits zu Beginn des Lernens in einer dem Gegenstand entsprechenden komplexen Weise mental repräsentieren. Im Verlauf des Lernens differenzieren sie die einzelnen Teilkomponenten aus und erweitern ihr mentales Modell um weitere Teilkomponenten (vgl. dazu Kliewer & Pohl, 2006, S. 656-659).

Die unterschiedlichen Theorieannahmen führen auch zu unterschiedlichen didaktischen Konsequenzen. Wenn Erwerbs- und Entwicklungsverläufe in Stufen erfolgen, so muss sich der Unterricht an der Stufenfolge ausrichten. Die Lernenden lernen zunächst die Laut-Buchstabenzuordnung und anschließend die orthographischen Regeln.

Geht man jedoch davon aus, dass sich Kompetenzen hierarchisch-integrativ entwickeln, so wäre die alphabetische Stufe keine notwendige Entwicklungsstufe, sondern eventuell sogar lernhinderlich, weil Kinder dadurch eine falsche Vorstellung vom Lerngegenstand entwickeln, die sie später revidieren müssen. Unter Annahme eines hierarchisch-integrativen Modells müssten die Kinder schon im Anfangsunterricht Gelegenheit erhalten, ein adäquates mentales Modell des Lerngegenstandes Rechtschreibung aufzubauen. Aus graphematischer Sicht hieße dies, sie mit dem phonographischen, silbischen und morphologischen Prinzip vertraut zu machen.

2.2 Das Testinstrument

Auf Grundlage des vorgestellten theoretischen Rahmenkonzepts wurde eine zusammenhängende Geschichte als Testinstrument konzipiert, die vollständig diktiert wurde. Es handelt sich dabei um eine fiktive Geschichte, die in der Natur spielt. Tiere sind die Protagonisten und kommunizieren miteinander. Diese Inhalte fügen

² Auf die Abprüfung dieser Teilfähigkeiten wurde bewusst verzichtet, da dafür grammatische C-Tests eine geeignetere Grundlage bieten.

sich in den Kontext des Deutschunterrichts in der Grundschule ein, in dem fiktive und phantastische Texte vorherrschen, sowie in den Kontext des Sachunterrichts, in dem Natur und Tiere thematisiert werden.

Dass sich der Rechtschreibtest damit von üblichen Diktattexten unterscheidet, war aus zwei Gründen beabsichtigt. Zum einen sollte durch die ungewöhnlichen Inhalte negativen Diktatassoziationen vorgebeugt werden,³ zum anderen sollte Rechtschreiben nicht isoliert, sondern im Zusammenhang einer Geschichte abgeprüft werden. Dies entspricht der Theorieannahme, dass Rechtschreibung für den Leser da ist.

Mit der Wahl eines Diktats als Testinstrument soll jedoch keinesfalls das herkömmliche Diktatschreiben im Unterricht befürwortet werden. Es muss vielmehr zwischen Leistungstests in empirischen Studien und der Lernkontrolle im Unterricht unterschieden werden.

Um mit einem Rechtschreibtest in einer vergleichenden Leistungsstudie verlässliche Aussagen treffen zu können, werden in Abhängigkeit von der Schwierigkeit der Testwörter für die in Abschnitt 2.1 ausgewiesenen fünf Teilfähigkeiten rund 100 Wörter benötigt. Dies ist im Lernzusammenhang des Unterrichts völlig anders. Hier können kleinere Tests zur Überprüfung nach jeder Lerneinheit geschrieben werden. Hierbei müssen auch nicht notwendigerweise alle Fehler gezählt werden, sondern es können Punkte für diejenigen richtig geschriebenen Wörter vergeben werden, deren Schreibung Gegenstand der Lerneinheit war. Dies eröffnet Kindern und Lehrkräften die Gelegenheit, einen Zusammenhang von Unterricht und Lernerfolg herzustellen. Aber auch im Unterricht sollte den Schülerinnen und Schülern stets die Funktion der Rechtschreibung für den Leser deutlich werden. Dies schließt ein, keine überdehnte Diktatsprache zu verwenden.

Die Hinweise der Testdurchführung für die Testleiter sahen daher auch vor, das gesamte Diktat zunächst sinnbetonend vorzulesen, bevor die Sätze einzeln, nach Einheiten gegliedert, diktiert werden sollten. Dabei sollten die Satzzeichen von den Testleitern mit diktiert werden. Falls auf Seiten der Kinder Bedarf bestand, konnten Wörter und Interpunktionszeichen noch einmal wiederholt werden. Am Ende sollte das ganze Diktat noch einmal vorgelesen werden.

Die zu schreibenden Wörter sind, durch die Form des Tests bedingt, in einen formalen syntaktischen und textuellen Kontext sowie in einen Bedeutungszusammenhang eingebettet. Sie werden in der Grundform, in flektierten Formen und als Ableitungen und Komposita dargeboten. Auch dies sprach für die Wahl eines zusammenhängenden Textes.

Die Testwörter entstammen zum Teil dem Grundwortschatz und sind somit im Unterricht geübte Wörter, bei denen ein Übungseffekt angenommen werden kann, zum Teil sind es ungebräuchlichere Wörter, die jedoch mit Hilfe von Regeln zu erschließen sind, die nach den Rahmenplänen in der Grundschule vermittelt werden.

Folgende Auflistung zeigt die Übereinstimmung der Testwörter mit dem Grundwortschatz in Bayern (gesamt 28):

³ Dass dies von Kindern auch so verstanden wurde, darauf deutet der Zusatz „Ende der Geschichte“ unter einem Diktat hin.

alle, älter(es), Bilder, beginnen (flektiert: beginnt), bunt(e), fröhlich, Fuchs, Haare (als Teil eines Kompositums: Schnurrbarthaaren), kommen, (hoch)leben, leuchten (Partizip Präsens: leuchtende), malen (flektiert: malt), Mittag (als Teil eines Kompositums: Mittagssonne), nimmt, Pferd, Pferde, Reh(e), schläft, schließlich, schütteln, Stift, süß(e), schmücken, vielleicht, weiß(en), Wiese, will(st), zwei.

Bezüglich der Groß- und Kleinschreibung wurden Wörter ausgewählt, die aufgrund der Wortart groß- bzw. kleingeschrieben werden. Auf Substantivierungen von Verben und Adjektiven wurde bis auf eine Konversion ganz verzichtet. Dies lag darin begründet, dass in erster Linie erhoben werden sollte, inwieweit die Kinder von der im Unterricht überwiegend praktizierten wortartenbezogenen Vermittlung der Großschreibung profitieren.

Die diktierte Geschichte besteht aus 112 Analyseeinheiten (Wörtern). Jedes Wort dient als Indikator für mehrere Prinzipien bzw. Teilfähigkeiten, womit die Informationsdichte gegenüber den ansonsten üblichen Lücken- und Satzdictaten maximiert wird, die in der Regel nach Wortfehlern ausgewertet werden.

3. Datengrundlage und Auswertungsverfahren

Datengrundlage für die im Folgenden dargestellten Analysen sind die IGLU-Voruntersuchungsdaten aus dem Erhebungsjahr 2005. Neben der Pilotierung der internationalen Leseverständnisinstrumente wurde im Sommer 2005 auch ein Test der im Rahmen von IGLU durchgeführten nationalen Ergänzungsstudie Orthographie eingesetzt und erprobt. Bei dieser Voruntersuchung handelt es sich nicht um eine repräsentative Stichprobe für deutsche Grundschul Kinder am Ende der vierten Jahrgangsstufe. Es beteiligten sich die Bundesländer Baden-Württemberg, Berlin, Niedersachsen, Saarland und Sachsen. In jedem dieser Bundesländer wurden Daten aus fünf Schulen mit jeweils zwei Schulklassen mit einem rotierten Testdesign erhoben. Insgesamt wurden 486 Kinder getestet.

Die für die Analyse der Testdaten eingesetzten Auswertungsverfahren verbinden qualitative und quantitative Elemente. Die Erfassung der getreuen Wortschreibungen mit allen Schreibvarianten erfolgte in einem ersten Schritt mit qualitativen inhaltsanalytischen Verfahren. Die transkribierten Daten wurden im nächsten Schritt in ein Programm für qualitative Datenanalysen eingelesen. In diesem Programm erfolgte eine automatisierte Verkodung der Analysekatoren an den entsprechenden Schülerdaten. Aus dem Programm wurden für jedes Kind mehrere hundert Zeichen lange Indikatorlisten mit Informationen zu den entsprechenden Teilfähigkeiten herausgeschrieben, die im nächsten Arbeitsschritt als Datengrundlage in eine Skalierungssoftware eingelesen werden konnten. Die Datenskalierung erfolgte auf Grundlage probabilistischer Testverfahren, wie sie in groß angelegten Leistungsvergleichsstudien wie PISA oder IGLU Anwendung finden. Diese Verfahren bieten den Vorteil, dass sich mit ihnen theoretische Annahmen zum Aufbau einer fachbezogenen Leistung in Form eines Kompetenzmodells ausdifferenzieren lassen (vgl. BMBF, 2003, S. 74 f.). Der probabilistische Analyseansatz ermöglicht es, Personenfähigkeiten und Aufgabenschwierigkeiten voneinander zu trennen und als eigen-

ständige Größen aus den Testdaten zu schätzen. Diese Quantifizierung der Aufgabenschwierigkeiten ist aus inhaltlichen Gründen für eine fachwissenschaftliche Interpretation wichtig, da durch eine anschließende Interpretation der Ergebnisse die strukturellen Komponenten charakterisiert werden können, die zu diesen Schwierigkeiten geführt haben. Bei den groß angelegten Studien wird diese Eigenschaft probabilistischer Testverfahren für die Kompetenzniveaudifferenzierung genutzt.

Die Modelle erlauben es also, mehrere Teilfähigkeiten zu formulieren und bei der Datenmodellierung zu berücksichtigen. Bei der Anpassung der Daten an die theoretischen Modelle wird durch eine Skalierungssoftware ermittelt, wie zuverlässig die theoretisch unterschiedenen Teilfähigkeiten sind und wie diese Teilfähigkeiten miteinander im Zusammenhang stehen. Aus psychometrischer Sicht ist es erstrebenswert, die in einem Kompetenzmodell ausdifferenzierten Teilfähigkeiten reliabel zu erfassen. Weiterhin sollten diese gegeneinander abgrenzbar sein, d. h., es sollten nur moderate Zusammenhänge zwischen ihnen bestehen.

Zudem lässt sich die Datenpassung von einfachen und komplexen Kompetenzmodellen durch so genannte informationsstatistische Kennzahlen (LR-, CAIC-Statistik) miteinander vergleichen. Mit diesen Modellgeltungstests kann aus den empirischen Daten entschieden werden, ob die theoretischen Annahmen, die zur Ausdifferenzierung des Kompetenzmodells geführt haben, sinnvoll sind, oder ob diese Ausdifferenzierung zu überdenken ist.

Alle dargestellten Ergebnisse wurden auf der Grundlage von einparametrischen logistischen Modellen für dichotome und ordinale Daten⁴ mit der Skalierungssoftware *ConQuest* ermittelt (vgl. Wu, Adams & Wilson, 1998). Die Eigenschaften dieser Modellklasse sind bei Fischer (1974) und Rost (2004) beschrieben.

4. Ergebnisse

Die Darstellung der Ergebnisse geschieht in mehreren Teilschritten. Zunächst erfolgt eine deskriptive Darstellung der Testergebnisse. Hierfür werden die Fehlerhäufigkeiten und die Schreibvarianten von ausgewählten Wörtern dargestellt. In einem ersten Analyseschritt werden die Skalierungsergebnisse zu den fünf orthographischen Teilfähigkeiten erläutert. Neben der Fragestellung, wie zuverlässig sich die im Kompetenzmodell ausgewiesenen Teilfähigkeiten messen lassen, werden auch die latenten – d. h. nicht unmittelbar beobachtbaren – Zusammenhänge zwischen ihnen untersucht. In einem weiteren Analyseschritt werden die orthographischen Teilfähigkeiten gemeinsam mit den Daten des IGLU-Lesetests skaliert und die Skalierungsergebnisse dargestellt. Welcher Zusammenhang zwischen den Testleistungen der Kinder und Kovariaten wie Geschlecht, Bildungshintergrund, Migrationsstatus etc. besteht, wird abschließend behandelt.

4.1 Auswertung und Analyse der Leistungsdaten

⁴ Dichotom: zweifach gestufte Antwortausprägungen (richtig – falsch);
ordinal: mehrfach gestufte Antwortausprägungen mit eindeutiger Ordnungsrelation (falsch – teilrichtig – richtig).

Deskriptive Befunde

Die 112 Testwörter der IGLU-Voruntersuchung 2006 wurden von den 486 getesteten Kindern in insgesamt 2442 Varianten geschrieben. Zu den schwierigen Wörtern zählen Wörter wie *Schnurrbarthaaren*, *Vergnügen* und *herrlichen*. Ersteres wurde nur von drei Kindern (0,6%) richtig geschrieben, zweiteres von 38 (7,8%) und letzteres von 116 (23,9%). Leichte Wörter sind hingegen z. B. *wie* (97,9% Richtig-schreibungen), *nach* und *es* (jeweils 96,9% Richtig-schreibungen). Von den insgesamt 112 Testwörtern konnten 103 Wörter als Analysewörter genutzt werden. Wörter, die keine Trennschärfe aufwiesen und sich deshalb nicht als Analysewörter eigneten, waren z. B. die Artikel *der*, *die*, *das*, *ein* und die Konjunktion *oder*. Sie wurden von fast allen Kindern richtig geschrieben.

Das variantenreichste Wort ist *Schnurrbarthaaren* und wurde von den Kindern in 146 unterschiedlichen Varianten verschriftet. Das Wort *quaken* kommt in 63 verschiedenen Schreibweisen vor, wobei anhand der in Tabelle 2 dargestellten Schreibvarianten der Kinder deutlich wird, dass ihnen die Buchstabenverbindung *qu* nicht geläufig ist.

Tabelle 2: Variantenschreibungen des Wortes *quaken*

ckuken	gagen	gaken	gargen
guaken	guarken	Gwargen	kaaken
kargen	kfaken	knarken	kquaken
kragen	kraken	kwacken	kwagen
Kwaken	kwaken	Kwaken	kwarken
puacken	puaken	puarcken	Qacken
qacken	qaken	qarcken	Qargen
Qarken	qarken	Qarken	Qpacken
Qqaken	Qquaken	Qua	Quacken
quacken	Quagen	quagen	Quagken
Quaken	quaken	Quaqen	Quaqken
quarcken	Quarcken	quargen	Quargen
Quarke	Quarken	quarken	quarqen
qucken	qugen	Quraken	Qurken
Quwarken	Qvarken	Qwaken	Qwarken
wogen	gwagen	guaken	

Die Auswertung auf Wortebene ergab, dass nur zwei der 486 Kinder alle Wörter der Geschichte richtig geschrieben haben. Im Durchschnitt wurden 16,6 Wörter von den Kindern falsch geschrieben. Rund ein Drittel der Kinder haben insgesamt nur weniger als zehn Fehler gemacht, zwei Drittel machten nicht mehr als 18 Fehler. 20 Prozent der Kinder machten auf Wortebene mehr als 25 Fehler. Zehn Prozent machten mehr als 33 Fehler, schrieben also fast jedes vierte diktierete Wort falsch. Die leistungsschwächsten fünf Prozent der Kinder machten mehr als 40 Fehler auf Wortebene und das leistungsschwächste Kind schrieb 70 der insgesamt 112 Testwörter falsch.

Skalierungsergebnisse zum orthographischen Kompetenzmodell

Für das mehrdimensionale Kompetenzmodell werden zunächst einige wichtige Statistiken berichtet, mit denen die Operationalisierung der fünf Teilfähigkeiten dargestellt wird.

Tabelle 3: Auszählung der Leistungsdaten auf Basis von Wortstruktureinheiten und ermittelte Reliabilitäten für die fünf Teilfähigkeiten

	Anzahl der Strukturein- heiten im Test	Min.	Max.	Mittelwert	Variations- koeffizient	Prozent korrekt	Relia- bilität
Phonographisches und silbisches Prinzip im Kernbereich	58	21	58	52.9	0.11	90%	0.94
Morphologisches Prinzip im Kernbereich	46	14	46	37.6	0.16	82%	0.95
Peripheriebereich	22	1	22	17.3	0.23	78%	0.87
Wortbildung	18	4	18	15.0	0.17	84%	0.97
Wortübergreifendes Prinzip	26	5	26	22.0	0.18	84%	0.89

Zentral für die Beschreibung der Testgüte sind die in der rechten Spalte ausgewiesenen Reliabilitäten der fünf Teilfähigkeiten, die über die Korrelationen von unabhängigen Plausible-Value-Ziehungen für das orthographische Kompetenzmodell ermittelt wurden (vgl. Adams & Carstensen, 2002, S. 152). Die Werte geben Auskunft darüber, wie zuverlässig der Test unter Anwendung des Testmodells die Personenfähigkeit misst (vgl. Rost, 2004, S. 316; Wu, Adams & Wilson, 1998, S. 138 f.). Für jede der fünf Teilfähigkeiten liegen die ermittelten Reliabilitäten um 0,9 und zeigen damit an, dass sich die der Theorie nach unterschiedenen Teilfähigkeiten reliabel empirisch erfassen lassen. Die in Spalte vier dargestellten Mittelwerte geben Auskunft über die durchschnittliche Anzahl korrekter Antworten in der jeweiligen Teilfähigkeit. Der Wert „Prozent korrekt“ zeigt den Quotienten aus „Anzahl der Struktureinheiten“ und „Mittelwert“ an, wodurch eine direkte Vergleichbarkeit der Teilfähigkeiten in Hinblick auf die durchschnittliche Richtigschreibung ermöglicht wird. Die Tabelle macht deutlich, dass die erhobenen Teilfähigkeiten sich im Schwierigkeitsgrad unterscheiden. Es ergibt sich eine Abstufung vom *Peripheriebereich* (78%), über das *morphologische Prinzip im Kernbereich* (82%), über die *Wortbildung* und das *wortübergreifende Prinzip* (jeweils 84%) bis hin zum *phonographisch-silbischen Prinzip im Kernbereich* (90%). Diese Werte verdeutlichen, dass für die differenzielle Messung von Teilfähigkeiten auf der Grundlage von kurzen Wortstruktureinheiten eine umfangreiche Datenbasis für eine reliable Messung der jeweiligen Teilfähigkeiten erforderlich ist.

Zur Operationalisierung der fünf Teilfähigkeiten wurden zwischen 18 und 58 Wortstruktureinheiten eingesetzt. Diese große Anzahl an Indikatoren für die jeweiligen Teilfähigkeiten ergibt sich aus dem Verhältnis zwischen den Anforderungsniveaus

der jeweiligen Wortstruktureinheiten und den Personenfähigkeiten der Kinder. „Prozent korrekt“-Werte im Bereich zwischen 78 Prozent und 90 Prozent verdeutlichen, dass die entsprechenden Wortstruktureinheiten von der überwiegenden Mehrzahl der Kinder richtig geschrieben werden. Bei probabilistischen Testverfahren sollten die relativen Lösungshäufigkeiten jedoch bei 50 Prozent liegen, damit eine Passung von Aufgabenschwierigkeiten und Personenfähigkeiten erreicht wird (vgl. Crocker & Algina, 1986). Bei den Aufgaben zum Leseverständnis aus der IGLU-Voruntersuchung 2006 ist dieses Ziel fast gegeben. Die relative Lösungshäufigkeit liegt hier bei 54 Prozent. Dass bei der Erfassung der Rechtschreibkompetenz mit einem differenziellen Testinstrument eine solch hohe Anzahl von Indikatoren zur Operationalisierung von Teilfähigkeiten erforderlich ist, liegt am Testgegenstand (vgl. Valtin et al., 2004, S. 147-150).

Die Leistungsvariation (siehe Spalte Variationskoeffizient) der Kinder ist im *Peripheriebereich* am größten. Mit Abstand folgen das *wortübergreifende Prinzip*, die *Wortbildung* und das *morphologische Prinzip im Kernbereich*. Am homogensten ist die Testleistung der Kinder bei den gemessenen Struktureinheiten zum *phonographischen und silbischen Prinzip im Kernbereich*. Damit wird deutlich, dass das letztgenannte Prinzip von der Mehrzahl der getesteten Kinder beherrscht wird, wohingegen beim *Peripheriebereich* eine Diskrepanz zwischen leistungsschwachen und -starken Kindern besteht.

Tabelle 4: Latente Interkorrelationen (Korrelationskoeffizienten) der fünf Teilfähigkeiten im untersuchten Rechtschreibkompetenzmodell

	Phonographisch/ silbisch Kernbereich	Morpho- logischer Kernbereich	Peripherie- bereich	Wort- bildung	Wortüber- greifendes Prinzip
Phonographisch/ silbisch Kernbereich	1				
Morphologischer Kernbereich	0,99	1			
Peripheriebereich	0,76	0,79	1		
Wortbildung	0,97	0,96	0,78	1	
Wortübergreifen- des Prinzip	0,84	0,84	0,70	0,91	1

In Tabelle 4 sind die Zusammenhänge zwischen den latenten Teilfähigkeiten dargestellt. Die höchsten Korrelationen ergeben sich für die beiden Fähigkeiten im *Kernbereich* (0,99). Die gemeinsame Varianz liegt hier bei 98 Prozent, womit aus analytischer Sicht deutlich wird, dass die Unterscheidung dieser Teilfähigkeiten nicht sinnvoll ist und unter diesem Aspekt für die Daten der Hauptuntersuchung eine Zusammenlegung dieser beiden Bereiche zu empfehlen ist. Demgegenüber sprechen die Untersuchungsbefunde, dass sich die latenten Korrelationen der beiden *Kernbereiche* zum *Peripheriebereich* unterscheiden – 0,76 für das *phonographisch-silbische Prinzip* und 0,79 für das *morphologische Prinzip*, was einer gemeinsamen

Varianz von 58 Prozent bzw. 62 Prozent entspricht – für eine gesonderte Betrachtung. Die geringste latente Korrelation (0,70) ergibt sich für das *wortübergreifende Prinzip* und den *Peripheriebereich*. Der gemeinsame Varianzanteil liegt für diese beiden Inhaltsbereiche bei unter 50 Prozent, womit die theoretische Annahme, die zu einer Ausdifferenzierung in diese beiden Teilfähigkeiten führte, aus analytischer Sicht bestätigt wird.

Im Vergleich dazu liegen die im Rahmen der Studien IGLU (vgl. Bos et al., 2003, S. 79-84) und PISA (vgl. Artelt, Stanat, Schneider & Schiefele, 2001, S. 80-84) ermittelten latenten Korrelationen zwischen verschiedenen Textsorten bzw. Leseverständnisaspekten (fachspezifische Subskalen) in der Regel über 0,90. Der in der PISA-Untersuchung ermittelte Zusammenhang zwischen den fachübergreifenden Skalen ‚Lesen‘, ‚Mathematik‘ und ‚Naturwissenschaft‘ fällt ebenfalls hoch aus – die Korrelationen liegen im Bereich zwischen 0,77 und 0,89 (vgl. Adams & Carstensen, 2002, S. 152-154).

Um die Dimensionalität der Rechtschreibkompetenzdaten empirisch zu überprüfen, wurden die Deviance-Werte der Kompetenzmodelle miteinander verglichen. Der Deviance-Wert⁵ des mehrdimensionalen Modells liegt mit 48.934 um 1.065 Punkte unter dem des alternativen eindimensionalen Modells (49.999 Punkte). Die Differenz der Deviance-Werte folgt einer Chi-Quadrat-Verteilung mit 14 Freiheitsgraden. Aus analytischer Sicht kann das komplexere fünfdimensionale orthographische Kompetenzmodell die in den erfassten Schülerdaten enthaltenen Informationen besser darstellen als ein Generalfaktormodell, in dem nur nach Richtig- und Falschschreibung unterschieden wird. Dieser Unterschied ist signifikant.

Den dargestellten Analysen zufolge ergibt sich also eine empirische Evidenz für das differenzielle Rechtschreibkompetenzmodell, wobei noch nicht alle theoretisch möglichen Modellvarianten berechnet wurden.

Für die Interpretation der Ergebnisse sind die ermittelten und in Tabelle 5 angegebenen Aufgabenschwierigkeiten von zentraler Bedeutung. Sie werden für ausgewählte Wörter dargestellt und erläutert.

⁵ Der statistischen Auswertung von Daten liegen theoretische Modelle mit spezifischen Annahmen über die Kompetenzstruktur zugrunde. Diese Annahmen lassen sich zu einem Kompetenzmodell zusammenfassen. Der Deviance-Wert ist ein Wert zur Überprüfung der Modellgüte. Er macht eine Aussage über die empirische Datenpassung in Bezug auf das zugrundeliegende Kompetenzmodell.

Tabelle 5: Aufgabenschwierigkeiten von Wörtern derselben Kategorie

Kategorien	Leichteres Wort	Relative Lösungshäufigkeit in %	Schwereres Wort	Relative Lösungshäufigkeit in %
Ganzes Wort	<u>F</u> uchs	98	<u>G</u> ans	89
Silbengelenk	kom <u>mm</u> en	96	Schnur <u>r</u> -	34
Silbeninitiales h	Re <u>h</u>	95	frö <u>h</u> lich	70
Dehnungs-h	<u>i</u> hre	96	Fo <u>h</u> len	58
<ß>	sü <u>ß</u> e	81	schlie <u>ß</u> lich	57
<f>	<u>F</u> uchs	98	Fo <u>h</u> len	50
<v>	<u>v</u> or	92	<u>V</u> order	68
<ch>	n <u>ach</u>	98	hoch <u>ch</u> leben	93
Umlaut	n <u>ä</u> chsten	84	gl <u>ä</u> nzenden	56
Flexions-				
morphem # t	frag <u>t</u>	98	begin <u>nt</u>	86
Großschreibung	<u>K</u> atze	97	<u>V</u> ergnügen	9

In der linken Spalte sind ausgewählte sprachwissenschaftliche Kategorien dargestellt, die hier betrachtet werden. In den Spalten rechts davon sind ausgewählte Testwörter mit ihren dazugehörigen Schwierigkeiten aufgelistet („relative Lösungshäufigkeit der Kinder“). Es zeigen sich große Unterschiede in den relativen Lösungshäufigkeiten als Indikatoren für die Aufgabenschwierigkeit, und zwar nicht nur zwischen den Kategorien „Leichteres und Schwereres Wort“, sondern auch innerhalb dieser Kategorien. Bei den „Leichteren Wörtern“ ist die β - und Umlautschreibung am schwierigsten. Bei den „Schwereren Wörtern“ ist das Bild uneinheitlicher, und zwar sowohl untereinander als auch im Vergleich zu den „Leichteren Wörtern“. Während die Schreibung des Wortes *Gans*, die *ch*-Schreibung in *hochleben* und das Flexionsmorphem *t* in *beginnt* mit 86-93 Prozent Richtigschreibungen nur relativ schwierigere Wörter sind, zählen die Umlautschreibung in *glänzenden*, die β -Schreibung in *schließlich*, die *f*- und Dehnungs-*h*-Schreibung in *Fohlen*, die Silbengelenkschreibung in *Schnurr-* und die Großschreibung von *Vergnügen* zu den absolut schwierigen Wörtern. Die *v*-Schreibung in *vorder-* und die Schreibung des vererbten silbeninitialen *h* in *fröhlich* sind Beispiele von mittlerer Schwierigkeit.

Eine vergleichende Analyse ergibt Hinweise auf mögliche Ursachen für die Unterschiede im Grad der Aufgabenschwierigkeit. Zum einen ist die Aufgabenschwierigkeit ein und desselben Phänomens höher, wenn dieses in einer flektierten oder abgeleiteten Wortform vorkommt (*kommen* – *Schnurr-*, *Reh* – *fröhlich*, *süße* – *schließlich*, *nach* – *hochleben*). Zum anderen wird ein und dasselbe Phänomen in bekannteren Wörtern häufiger richtig geschrieben als in unbekannteren (*ihre* – *Fohlen*, *Fuchs* – *Fohlen*, *vor* – *vorder* als Teil von *Vorderhuf*, *nächsten* – *glänzenden*, *fragt* – *beginnt*).

Die unterschiedliche Schwierigkeit in den bekannten Wörtern *Fuchs* und *Gans* kann als eine Folge der Ranschburg'schen Hemmung angesehen werden, da das gleichzeitige Üben von *ganz* und *Gans* in Sprachbüchern verbreitet ist.

Der gravierendste Unterschied betrifft die Großschreibung der beiden Wörter *Katze* und *Vergnügen*. Hierfür sind mehrere Ursachen zu vermuten. Dass 98 Prozent der Kinder *Katze* großschreiben, kann damit erklärt werden, dass es sich um ein bekanntes Wort und ein Konkretum handelt. Bei *Vergnügen* dagegen handelt es sich um ein unbekanntes abstraktes Nomen, das von der Form her eine Konversion darstellt.

Skalierungsergebnisse zum Zusammenhang von Orthographie- und Lesekompetenz

In Ergänzung zur Skalierung der Orthographiedaten wurde ein Modell berechnet, in dem die fünf orthographischen Teilfähigkeiten zusammen mit den Aufgaben aus dem IGLU-Lesetest und den Wortfehlern, d. h. der orthographischen Richtigschreibung auf Wortebene, analysiert wurden (Tabelle 6). Die latenten Zusammenhänge zwischen der mit den IGLU-Tests ermittelten *Lesekompetenz* und den fünf Teilfähigkeiten liegen zwischen 0,50 für den *Peripheriebereich* und 0,77 für die *Wortbildung*. Die latenten Zusammenhänge zwischen Lesekompetenz und den beiden *Kernbereichen* bzw. dem *wortübergreifenden Prinzip* liegen einheitlich zwischen 0,65 und 0,68.

Tabelle 6: Latente Zusammenhänge (Korrelationskoeffizienten) der fünf Teilfähigkeiten mit der Leseleistung und den Wortfehlern (Richtigschreibung auf Wortebene)

	Phonographisch/ silbisch Kernbereich	Morpho- logischer Kernbereich	Peripherie- bereich	Wort- bildung	Wortüber- greifendes Prinzip	Lese- kompetenz	Wortfehler
Lesekompetenz	0,65	0,65	0,50	0,77	0,68	1	
Wortfehler	0,97	0,98	0,80	0,98	0,93	0,66	1

In Tab. 5 sind auch für den prinzipienübergreifenden Indikator einer orthographischen Kompetenz auf Wortebene (*Wortfehler*) die latenten Zusammenhänge mit den ausdifferenzierten Teilfähigkeiten und der Lesekompetenz dargestellt. Der Zusammenhang zwischen Wortfehlern und den orthographischen Teilleistungen im *Kernbereich* bzw. bei der *Wortbildung* liegt sehr nahe beim maximal zu erreichenden Korrelationswert von 1 und verdeutlicht aus analytischer Sicht, dass bei diesen Teilfähigkeiten alternativ ohne größeren Informationsverlust mit dem Indikator *Wortfehler* gearbeitet werden kann. Geringere Zusammenhänge bestehen zwischen den Kategorien *Wortfehler*, *wortübergreifendes Prinzip* und *Peripheriebereich*, wobei sich letztere – wie bereits bei der Leseleistung – als eigenständige Analysekatgorie darstellt.

4.2 Zusammenhänge von Kovariaten und Leistungsdaten

Mit dem Skalierungsprogramm *ConQuest* lassen sich Schätzwerte für die orthographische Fähigkeit jedes Kindes erzeugen, die als Grundlage für weiterführende Bedingungsanalysen herangezogen werden können. An dieser Stelle soll untersucht werden, inwieweit Merkmale wie Geschlecht, Bildungs- und Migrations-

hintergrund sowie das Leseverständnis der Kinder mit der Rechtschreibkompetenz auf Wortebene im Zusammenhang stehen. Um die Interpretationsergebnisse vergleichbar zu machen, wurden die Werte der Logit-Metrik⁶ analog zur IGLU-Auswertung transformiert, so dass die Werte einer Verteilung mit einem Mittelwert von 500 und einer Standardabweichung von 100 folgen.

Bildungshintergrund

Zwischen dem Bildungshintergrund der Kinder – gemessen an der Anzahl der im Elternhaus zur Verfügung stehenden Kinderbücher – und den Testleistungen ergibt sich ein streng monotoner Zusammenhang. Kinder, denen in ihren Elternhäusern bis maximal zehn Bücher zur Verfügung stehen, erzielen Testleistungen, die mit rund 400 Punkten in den beiden *Kernbereichen* und bei der *Großschreibung* eine Standardabweichung unter dem Durchschnitt liegen. Diese Leistungsdifferenz entspricht zirka zwei Schuljahren. Für den *Peripheriebereich* ist dieser Zusammenhang aus Bildungshintergrund und Orthographieleistung nicht ganz so eng. Mit 445 Punkten liegt die Differenz zur Durchschnittsleistung bei gut einem Schuljahr. Kinder aus bildungsnahen Haushalten, die mehr als 100 Kinderbücher haben, erzielen eine durchschnittliche Testleistung von 560 Punkten. Die soziale Kopplung aus Bildungshintergrund und orthographischer Leistung ist für den Bereich der Orthographie im gleichen Maß gegeben wie für den Bereich der Leseleistung, wobei in Abhängigkeit von der Anzahl zur Verfügung stehender Kinderbücher ein Leistungsunterschied von bis zu drei Schuljahren in den Daten zu finden ist. Dieser Leistungsunterschied verringert sich jedoch für die Testleistung im Peripheriebereich auf zwei Schuljahre.

Geschlecht

Die Testleistungen der Jungen liegen mit Ausnahme des Peripheriebereichs rund 30 Punkte unter den Leistungen der Mädchen. Für den Peripheriebereich vergrößert sich dieser Unterschied auf 70 Punkte und liegt bei etwa 1 ½ Schuljahren.

Migrationstatus

Die orthographischen Leistungen von Kindern, die nicht in Deutschland geboren sind, liegen rund 30 Punkte unter den Leistungen ihrer deutschen Mitschülerinnen und Mitschüler, wobei der Unterschied für den Bereich der Wortbildung mit 40 Punkten besonders groß ist (entspricht knapp einem Schuljahr) und für den Peripheriebereich mit 15 Punkten eher gering ausfällt (entspricht rund 1/3 Schuljahr). Kinder, die zur Einschulung nach Deutschland kamen, erzielen vergleichbare Leistungen wie ihre deutschen Mitschülerinnen und Mitschüler. Umso älter die Grundschulkinder bei ihrer Ankunft in Deutschland sind, desto schlechter sind ihre Testleistungen.

⁶ Auf der Logit-Metrik werden die von *ConQuest* geschätzten Parameter der Schülerfähigkeit und der Wortschwierigkeit in einem Verhältnis zueinander gesetzt (vgl. Ludlow & Haley, 1995).

5. Fazit

Das abschließende Fazit fasst die Ergebnisse im Hinblick auf die forschungsleitenden Fragen zusammen:

- 1) Modellierung der Rechtschreibkompetenz und Schwierigkeitsniveaus der Teilfähigkeiten
- 2) Schlussfolgerungen in Bezug auf den Aufbau von Rechtschreibkompetenz
- 3) Zusammenhang von Rechtschreib- und Lesekompetenz
- 4) Einflussfaktoren auf die Rechtschreibleistung

Zu 1) Die Ergebnisse der empirischen Dimensionalitätsprüfung bestätigen die Annahme, dass das mehrdimensionale Modell mit den fünf Teilfähigkeiten die Rechtschreibkompetenz zuverlässiger erfasst als ein eindimensionales Modell auf Wortebene. Es konnte gezeigt werden, dass die fünf Teilkompetenzen reliabel zu messen sind. Die Interkorrelationen der fünf Teilfähigkeiten machen deutlich, dass die Fähigkeit zur Wortschreibung im Peripheriebereich sich am deutlichsten von den anderen Teilfähigkeiten unterscheidet. Die Wortschreibungen im Kernbereich stellen ein und dieselbe Kompetenzleistung dar, wobei der Schwierigkeitsgrad variiert. Er ist – bezogen auf das jeweilige Phänomen – im morphologischen Kernbereich höher als im phonographisch-silbischen. Dies ist erwartungskonform, weil die Schreibung von flektierten Formen und Ableitungen eine zusätzliche kognitive Operation erfordert. Innerhalb des phonographischen Prinzips stellt die Schreibung von f, v, pf, qu, g, k, s und ß den höchsten Schwierigkeitsgrad dar, was sich mit den damit zusammenhängenden besonderen Anforderungen erklären lässt (Sonderstellung von v, Mehrgraphen, Unterscheidung von stimmhaften und stimmlosen Lauten).

Weiterhin besteht ein hoher Zusammenhang zwischen Wortbildung und Wortschreibung im Kernbereich. Dies kann damit zusammenhängen, dass auch der Wortbildung verlässliche Regularitäten zugrunde liegen.

Dass das wortübergreifende Prinzip mit der Wortbildung höher korreliert als mit der Wortschreibung im Kernbereich, erklärt sich dadurch, dass die Wortbildungsmorpheme Hinweise auf die Groß- bzw. Kleinschreibung liefern.

Innerhalb der Großschreibung konnte die größte Bandbreite im Schwierigkeitsgrad nachgewiesen werden. Zwischen den Beispielen *Katze* und *Vergnügen* liegt der Unterschied in der relativen Lösungshäufigkeit bei 89 Prozent. Ein Vergleich mit dem Schwierigkeitsgraden der Wörter in dieser Kategorie deutet darauf hin, dass die Geläufigkeit eines Wortes, seine grammatische Form und die Zuordnung zu Konkreta oder Abstrakta schwierigkeitsbestimmende Faktoren sind.

Zu 2) Die Frage, ob sich Rechtschreibkompetenz stufenweise oder hierarchisch-integrativ aufbaut, lässt sich mit einem Test nicht beantworten. Die Ergebnisse der Untersuchung lassen jedoch einige Rückschlüsse zu. Die Zusammenhänge zwischen den Teilfähigkeiten und die Unterschiede im Schwierigkeitsgrad deuten darauf hin, dass sich die Rechtschreibkompetenz über die Einsicht in die Regularitäten der Schrift aufbaut. Dafür spricht vor allem, dass sich die Teilfähigkeiten im Kern- und Peripheriebereich substantiell unterscheiden. Im Kernbereich der Wortschreibung

und im Bereich der Wortbildung können Kinder Regelmäßigkeiten erkennen und gewonnene Einsichten transferieren. Gelingt ihnen dies, so verfügen sie über eine ausbaufähige Basis. Das trifft nicht für den auf Übung basierenden Peripheriebereich zu. Dass bei Kindern mit Migrationshintergrund die Teilleistungen in Kern- und Peripheriebereich geringere Unterschiede aufweisen, könnte anzeigen, dass diese Kinder noch keine Regelmäßigkeiten entdeckt haben, sondern Richtigschreibungen allein auf Übung zurückzuführen sind.

Die Teilfähigkeit Groß- und Kleinschreibung basiert zumindest zu einem Teil auf grammatikalischen Einsichten in die Wortbildung. Zur Beantwortung der Frage, inwieweit Kinder die Bestimmung der Wortart oder syntaktische Informationen heranziehen, liefern die Testergebnisse keine Aufschlüsse.

Zu 3) Die unterschiedlichen Zusammenhänge der einzelnen Teilfähigkeiten mit der Lesekompetenz lassen sich sehr schlüssig interpretieren. Dass der Bereich Wortbildung den höchsten Zusammenhang mit der Leseleistung aufweist, ergibt sich aus der Ökonomie dieses Prinzips für das Lesen. Viertklässler erlesen sich nicht mehr die Lautform eines Wortes, um daraus die Bedeutung zu erschließen, wie zu Beginn des Schriftspracherwerbs, sondern sie konstruieren die Bedeutung aus dem Lexem und den grammatischen Morphemen. Der geringste Zusammenhang besteht erwartungsgemäß mit dem Peripheriebereich, schon allein deswegen, weil Wörter aus diesem Bereich weniger häufig vorkommen. Im Hinblick auf den Zusammenhang mit dem wortübergreifenden Prinzip erscheinen Vergleichsuntersuchungen mit Rechtschreibexperten von Interesse. Bei dieser Gruppe wäre ein höherer Zusammenhang zu erwarten, da Experten vermutlich die syntaxbezogenen Informationen in höherem Maße nutzen.

Zu 4) Die Kontrollvariablen Bildungshintergrund, Geschlecht und Migrationsstatus erwiesen sich erwartungsgemäß als förderliche bzw. hinderliche Einflussfaktoren. Da in der Voruntersuchung keine detaillierten Daten zum Rechtschreibunterricht aus Lehrerfragebögen vorliegen, konnten keine Zusammenhänge mit dem jeweiligen Unterricht ermittelt werden. Erst eine Untersuchung der unterrichtlichen Einflussfaktoren auf die Entwicklung der Rechtschreibkompetenz kann jedoch Antworten auf die Frage liefern, inwieweit auch didaktogene Ursachen für die Schwierigkeiten der Kinder beim Rechtschreiblernen verantwortlich sind. Aus den Ergebnissen der vorliegenden Studie, wonach Kinder ein und dasselbe Rechtschreibphänomen abhängig vom Bekanntheitsgrad der Wörter richtig bzw. falsch schreiben, lässt sich schließen, dass der Unterricht mehr auf übendes als auf verstehendes Lernen ausgerichtet ist. Dass dies den Rechtschreiberwerb nicht fördert, ergibt sich aus dem Befund, dass es sich bei den Teilfähigkeiten im Kern- und Peripheriebereich um unterschiedliche Kompetenzen handelt, wobei die Teilfähigkeiten im Kernbereich die Rechtschreibleistung vorwiegend bestimmen und in einem hohem Zusammenhang zur Leseleistung stehen.

Literatur

Adams, R. J. (2002). Scaling PISA cognitive data. In R. J. Adams & M. L. Wu (Hg.), *PISA 2000 technical report* (S. 99-108). Paris: OECD.

- Adams, R. J. & Carstensen, C. (2002). Scaling outcomes. In R. J. Adams & M. L. Wu (Hg.), *PISA 2000 technical report* (S. 149-162). Paris: OECD.
- Artelt, C., Stanat, P., Schneider, W. & Schiefele, U. (2001). Lesekompetenz: Testkonzeption und Ergebnisse. In J. Baumert, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele, W. Schneider, P. Stanat, K.-J. Tillmann & M. Weiß (Hg.), *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (S. 69-137). Opladen: Leske + Budrich.
- Bereiter, C. (1980). Development in writing. In: L. W. Gregg & E. R. Steinberg (Hg.), *Cognitive Processes in Writing* (S. 73-93). Hillsdale: Erlbaum.
- Blatt, I., Voss, A. & Matthießen, I. (2005). Kinder schreiben Briefe aus der Zukunft. Qualitative Analysen von Kindertexten und fachdidaktische Diskussion. In W. Bos, E.-M. Lankes; M. Prenzel, K. Schwippert, R. Valtin & G. Walther (Hg.), *IGLU. Vertiefende Analyse zu Leseverständnis; Rahmenbedingungen und Zusatzstudien* (S. 109-158). Münster: Waxmann.
- Bos, W., Lankes, E.-M., Schwippert, K., Valtin, R., Voss, A., Badel, I. & Plaßmeier, N. (2003). Lesekompetenzen deutscher Grundschülerinnen und Grundschüler am Ende der vierten Jahrgangsstufe im internationalen Vergleich. In W. Bos, E.-M. Lankes, M. Prenzel, K. Schwippert, G. Walther & R. Valtin (Hg.), *Erste Ergebnisse aus IGLU. Schülerleistungen am Ende der vierten Jahrgangsstufe im internationalen Vergleich* (S. 69-142). Münster: Waxmann.
- Bremerich-Vos, A. (2004). Bildungsstandards, Rechtschreibkompetenz und IGLU – einige Anmerkungen. In A. Bremerich-Vos, C. Löffler, K.-L. Herné (Hg.), *Neue Beiträge zur Rechtschreibtheorie und -didaktik* (S. 85-104). Freiburg: Fillibach.
- Bundesministerium für Bildung und Forschung (2003). *Expertise zur Entwicklung nationaler Bildungsstandards*, Berlin: BMBF. (Bildungsreform; Bd. 1).
- Butt, M. & Eisenberg, P. (1990). Schreibsilbe und Sprechsilbe. In: C. Stetter (Hg.), *Zu einer Theorie der Orthographie. Interdisziplinäre Aspekte gegenwärtiger Schrift- und Orthographieforschung* (S. 33-64). Tübingen: Max Niemeyer (Reihe Germanistische Linguistik, Bd. 99).
- Campbell, J. R., Kelly, D. L., Mullis, I. V. S., Martin, M. O. & Sainsbury, M. (2001). *Framework and specifications for PIRLS assessment 2001* (2. Aufl.). Chestnut Hill, MA: Boston College.
- Crocker, L., & Algina, J. (1986). *Introduction to classical & modern test theory*. New York: Holt, Reinhart and Winston.
- Deutsches PISA-Konsortium (Hg.) (2000). *Schülerleistungen im internationalen Vergleich. Eine neue Rahmenkonzeption für die Erfassung von Wissen und Fähigkeiten*. Berlin: Max-Planck-Institut für Bildungsforschung.
- Eisenberg, P. (1989). Die Schreibsilbe im Deutschen. In: P. Eisenberg & H. Günther (Hg.), *Schriftsystem und Orthographie* (S. 57-84). Tübingen: Max Niemeyer. (Reihe Germanistische Linguistik, Bd. 97).
- Eisenberg, P. & Fuhrhop, N. (2007). Schulorthographie und Graphematik. *Zeitschrift für Sprachwissenschaft*, 26, 15-41.
- Eisenberg, P., Spitta, G. & Voigt, G. (1994). Schreiben: Rechtschreiben. In: *Praxis Deutsch* 124, 14-25.
- Fischer, G. H. (1974). *Einführung in die Theorie psychologischer Tests: Grundlagen und Anwendungen*. Bern: Huber.

- Hinney, G. (1997). *Neubestimmung von Lerninhalten für den Rechtschreibunterricht. Ein fachdidaktischer Beitrag zur Schriftaneignung als Problemlöseprozess*. Frankfurt am Main: Peter Lang.
- Hinney, G. & Menzel, W. (1998). Didaktik des Rechtschreibens. In: G. Lange, K. Neumann & W. Ziesenis (Hg.), *Taschenbuch des Deutschunterrichts. Grundfragen und Praxis der Sprach- und Literaturdidaktik* (6. vollständig überarbeitete Aufl.) S. 258-304. Baltmannsweiler: Schneider Hohengehren. (Grundlagen – Sprachdidaktik – Mediendidaktik, Bd. 1).
- Kirsch, I. S. & Jungeblut, A. (1986). *Literacy: Profiles of America's Young Adults*. Report (Nr. 16-PL-02). National Assessment of Educational Progress.
- Kirsch, I. S., Jungeblut, A. & Mosenthal, P. B. (1998). The measurement of adult literacy. In S. T. Murray, I. S. Kirsch & L. B. Jenkins (Hg.), *Adult literacy in OECD countries: Technical report on the First International Adult Literacy Survey* (S. 105-134). Washington, DC: U.S. Department of Education.
- Kliwer, H. J. & Pohl, I. (Hg.) (2006). *Lexikon Deutschdidaktik*. 2 Bde. Baltmannsweiler: Schneider Verlag. (Bd. 1).
- Kliwer, H. J. & Pohl, I. (Hg.) (2006). *Lexikon Deutschdidaktik*. 2 Bde. Baltmannsweiler: Schneider Verlag. (Bd. 2).
- Löffler, I. & Meyer-Schepers, U. (2005). Orthographische Kompetenzen: Ergebnisse qualitativer Fehleranalysen, insbesondere bei schwachen Rechtschreibern. In W. Bos, E.-M. Lankes, M. Prenzel, K. Schwippert, G. Walther & R. Valtin (Hg.), *IGLU. Vertiefende Analysen zum Leseverständnis, Rahmenbedingungen und Zusatzstudien* (S. 81-108). Münster: Waxmann.
- Ludlow, L. H. & Haley, S. M. (1995). Rasch model logits: Interpretation, use, and transformation. *Educational and psychological measurement*, 55 (6), 967-975.
- Röber-Siekmeyer, C. (1999). *Ein anderer Weg zur Groß- und Kleinschreibung*. Leipzig: Klett.
- Rost, J. (2004). *Lehrbuch Testtheorie – Testkonstruktion* (2., vollständig überarbeitete und erweiterte Aufl.). Bern: Hans Huber.
- Schönweiss, F. *Förderdiagnostisches Konzept*. Verfügbar unter: http://www.lernserver.de/coRED/static/lehrer_hintergrund_konzept.php?cat=3:3:2 [Stand = 30.05.2007].
- Valtin, R., Badel, I., Löffler, I., Meyer-Schepers, U. & Voss, A. (2003). Orthographische Kompetenzen von Schülerinnen und Schülern der vierten Klasse. In W. Bos, E.-M. Lankes, M. Prenzel, K. Schwippert, G. Walther & R. Valtin (Hg.), *Erste Ergebnisse aus IGLU. Schülerleistungen am Ende der vierten Jahrgangsstufe im internationalen Vergleich* (S. 227-264). Münster: Waxmann.
- Valtin, R., Löffler, I., Meyer-Schepers, U., Badel, I. (2004). Orthographische Kompetenzen von Schülerinnen und Schülern der vierten Klasse im Vergleich der Länder. In W. Bos, E.-M. Lankes, M. Prenzel, K. Schwippert, R. Valtin & G. Walther (Hg.), *Einige Länder der Bundesrepublik Deutschland im nationalen und internationalen Vergleich* (S. 141-164). Münster: Waxmann.
- Wu, M. L., Adams, R. J. & Wilson, M. R. (1998). *ACER ConQuest. Generalised item response modelling software*. Melbourne: Acer Press.

Anschriften der Verfasser:

Dr. Andreas Voss, Universität Dortmund, Institut für Schulentwicklungsforschung (IFS), Vogelpothsweg 78, 44227 Dortmund, voss@ifs.uni-dortmund.de

Prof. Dr. Inge Blatt, Universität Hamburg, Fachbereich Erziehungswissenschaft, Sektion 4, Von-Melle-Park 8, 20146 Hamburg, blatt@erzwiss.uni-hamburg.de

Dipl. Päd. Kerstin Kowalski, Universität Dortmund, Institut für Schulentwicklungsforschung (IFS), Vogelpothsweg 78, 44227 Dortmund, kowalski@ifs.uni-dortmund.de

Rüdiger Vogt

MÜNDLICHE ARGUMENTATIONSKOMPETENZ BEURTEILEN Dimensionen, Probleme, Perspektiven

Die in der aktuellen Diskussion um fachliche „Kompetenzen“ thematisierte Frage, in welcher Weise diese modelliert werden können, kann in unterschiedlicher Weise beantwortet werden. Aktuell ist der Rückgriff auf die Definition Weinerts gängig, die im Paradigma der pädagogischen Psychologie auf die kognitive Herausbildung von Problemlösungsfähigkeiten und -fertigkeiten abzielt. In der aktuellen Diskussion um Tests wird ein Konzept favorisiert, das den Kompetenzbegriff eng an das anbindet, was in der quantitativen Forschung am wichtigsten ist, nämlich ihre Messung. So wird im Rahmen der sprachliche Kompetenzen fokussierenden DESI-Studie die folgende Definition zugrunde gelegt: „Sprachkompetenz wird dabei definiert als Komplex von Teilfähigkeiten, die durch den schulischen Unterricht vermittelt werden sollen, die Kompetenzmessung nimmt dabei die Rolle der Lernerfolgsüberprüfung ein.“ (Jude/Klieme 2007, 11)

Die in dieser Formulierung zum Ausdruck kommende Reduktion von Sprachkompetenz auf das, was in der Schule unterrichtet wird, kann nicht überzeugen. Vielmehr wäre für die Inhalte des Faches Deutsch ein Kompetenzkonzept angemessen, das zunächst die sprachlichen Fähigkeiten systematisch in den Blick nimmt, um sie dann auf die besonderen Bedingungen der Kompetenzvermittlung in der Institution Schule zu beziehen. Und schließlich muss gefragt werden, welche Kompetenzen unter diesen Bedingungen vermittelt werden können, bzw. welche Faktoren die sprachlichen Handlungen der Probanden beeinflussen.

Der Bereich der mündlichen Kommunikation erweist sich unter den Bedingungen des Testens als besonders schwierig, da dialogische Aktivitäten nur unter Schwierigkeiten bzw. großem Aufwand getestet werden können: Man kann miteinander sprechende Schüler entweder durch geschulte Rater in der Situation beobachten lassen oder aber ihre Aktivitäten audiovisuell dokumentieren und anschließend auswerten. Dazu bedarf es aber geeigneter Beobachtungskategorien: Die vorliegenden Beurteilungsraster müssten empirisch validiert werden. Eine weitere Möglichkeit wäre die Transkription der Dokumentation, die dann gesprächsanalytisch aufbereitet wird. Kein Wunder, dass sich die Verantwortlichen der sog. *large-scale*-Untersuchungen mit der Nutzung dieser Instrumente schwer tun: So wird beispielsweise bei DESI argumentative Kompetenz als schriftliche Formulierung eines Gesprächsbeitrags modelliert. Das mag für einen solchen Zusammenhang als angemessen