

Wilfried Bos, Eva-Maria Lankes, Manfred Prenzel,
Knut Schwippert, Renate Valtin, Gerd Walther (Hrsg.)

IGLU

Einige Länder der Bundesrepublik Deutschland
im nationalen und internationalen Vergleich



Waxmann Münster / New York
München / Berlin

Inhalt

I.	IGLU –Einige Länder der Bundesrepublik Deutschland im Vergleich.....	1
	<i>Wilfried Bos, Eva-Maria Lankes, Manfred Prenzel, Knut Schwippert, Renate Valtin, Gerd Walther</i>	
1.	Vorbemerkung	1
2.	Finanzierung und politische Administration.....	3
3.	Die Organisation von IGLU	4
II.	Anlage und Durchführung der Internationalen Grundschul-Lese-Untersuchung (IGLU) und ihrer Erweiterung um Mathematik und Naturwissenschaften (IGLU-E)	7
	<i>Eva-Maria Lankes, Wilfried Bos, Nike Plafßmeier, Knut Schwippert, Andreas Voss</i>	
1.	Die Untersuchungspopulation.....	7
1.1	Stichprobe und Vergleichsgruppen.....	7
1.2	Stichprobenziehung, Gewichtung und die Schätzung von Standardfehlern	10
1.3	Nationale Beteiligungsquoten.....	12
2.	Testdesign und Testdurchführung	16
2.1	Test- und Fragebogenentwicklung.....	16
2.2	Aufbau der Untersuchung.....	16
2.3	Skalierung.....	18
2.4	Durchführung der Erhebung.....	19
III.	Lehr- und Lernbedingungen in einigen Ländern der Bundesrepublik Deutschland und im internationalen Vergleich.....	21
	<i>Eva-Maria Lankes, Nike Plafßmeier, Wilfried Bos, Knut Schwippert</i>	
1.	Personelle und materielle Ausstattung.....	22
1.1	Schul- und Klassengrößen	22
1.2	Ausstattung mit Bibliotheken und Computern	24
2.	Lernbiographie der Viertklässler	27
2.1	Einschulungsalter und Testalter.....	27

2.2	Unterrichtszeit	30
3.	Individualisierung im Leseunterricht der Grundschule	32
3.1	Organisation des Leseunterrichts	32
3.2	Differenzierung durch Material im Leseunterricht	34
3.3	Leseaktivitäten im Unterricht	36
4.	Förderung innerhalb der Schule und des Unterrichts	37
4.1	Förderung zusätzlich zum regulären Unterricht	38
4.2	Förderung innerhalb des regulären Unterrichts	39
5.	Schulleitung und Kollegium	40
5.1	Die Rolle der Schulleitung	40
5.2	Kooperation im Kollegium	42
5.3	Stellung der Schule zu Schulerfolg und Schulversagen	43
6.	Zusammenschau	44
IV.	Lesekompetenzen am Ende der vierten Jahrgangsstufe in einigen Ländern der Bundesrepublik Deutschland im nationalen und internationalen Vergleich	49
	<i>Wilfried Bos, Renate Valtin, Eva-Maria Lankes, Knut Schwippert, Andreas Voss, Isolde Badel, Nike Pflaßmeier</i>	
1.	Einleitung	49
2.	Zum Konstrukt und zur Erfassung von Lesekompetenz	50
2.1	Das IGLU-Lesekompetenzmodell	51
2.2	Zur Modellierung der Lesekompetenz in IGLU	53
2.3	Kompetenzstufen des Leseverständnisses	53
3.	Ergebnisse der internationalen und nationalen Vergleiche	55
3.1	Zur Interpretation der Testwerte	55
3.2	Zum Niveau der Lesekompetenz	57
3.3	Zur Breite der Leistungsverteilungen	62
3.4	Leseleistungen nach Kontrolle von Jahrgangsstufe und Alter	65
3.5	Leseleistungen bei Informations- und literarischen Texten	67
3.6	Leistungen von Mädchen und Jungen in der Lesekompetenz	71
3.7	Anteile der Schülerinnen und Schüler auf den Kompetenzstufen	74
3.8	Vergleich der Ergebnisse von IGLU und PISA	78
3.9	Einstellungen zum Lesen	80
3.10	Vorschulische Erfahrungen	83
4.	Zusammenschau	85

V.	Naturwissenschaftliche Kompetenz am Ende der Grundschulzeit: Vergleiche zwischen einigen Ländern der Bundesrepublik Deutschland.....	93
	<i>Manfred Prenzel, Helmut Geiser, Rolf Langeheine, Kirstin Lobemeier</i>	
1.	Naturwissenschaftliche Grundbildung in der Grundschule	94
2.	Der Untersuchungsansatz zur naturwissenschaftlichen Grundbildung	96
2.1	Die Testkonzeption bei TIMSS	96
2.2	Der Testansatz bei IGLU-E	96
3.	Ergebnisse.....	101
3.1	Naturwissenschaftsleistungen an deutschen Grundschulen im nationalen und internationalen Vergleich	101
3.2	Leistungsverteilungen im Naturwissenschaftstest nach Ländern.....	106
3.3	Motivation und Selbstkonzept in den Ländern	110
4.	Zusammenschau	112
VI.	Mathematische Kompetenzen am Ende der vierten Jahrgangsstufe in einigen Ländern der Bundesrepublik Deutschland	117
	<i>Gerd Walther, Helmut Geiser, Rolf Langeheine, Kirstin Lobemeier</i>	
1.	Mathematische Grundbildung.....	118
2.	Erhebung mathematischer Grundbildung	119
2.1	Testansatz bei IGLU-E	119
2.2	Testaufbau und Items.....	120
2.3	Stufen der mathematischen Kompetenz im Grundschulalter.....	121
3.	Ergebnisse.....	124
3.1	Mathematikleistungen einiger Länder der Bundesrepublik Deutschland im nationalen und internationalen Vergleich	125
3.2	Mathematikleistungen in einigen Ländern der Bundesrepublik Deutschland am Ende der Grundschulzeit und am Ende der Sekundarstufe I.....	128
3.3	Leistungsunterschiede im Mathematiktest.....	130
3.4	Motivation und Selbstkonzept	134
4.	Zusammenschau	136

VII. Orthographische Kompetenzen von Schülerinnen und Schülern der vierten Klasse im Vergleich der Länder	141
<i>Renate Valtin, Ilona Löffler, Ursula Meyer-Schepers, Isolde Badel,</i>	
1. Zur Messung der orthographischen Kompetenz in IGLU-E.....	141
2. Ergebnisse zu orthographischen Kompetenzen	145
2.1 Quantitative und qualitative Fehleranalysen.....	145
2.2 Kompetenzstufen der Orthographie.....	150
3. Zum Rechtschreibunterricht in vierten Klassen.....	152
3.1 Zeit, Medien, Übungsformen und Differenzierung	152
3.2 Grundwortschatz.....	158
4. Zusammenschau	160
VIII. Heterogenität und Chancengleichheit am Ende der vierten Jahrgangsstufe in den Ländern der Bundesrepublik Deutschland und im internationalen Vergleich.....	165
<i>Knut Schwippert, Wilfried Bos, Eva-Maria Lankes</i>	
1. Strukturmerkmale von Familien	165
1.1 Sozialschicht.....	166
1.2 Familien mit Migrationsgeschichte	168
2. Leistungsbezogene Vergleiche – national	173
3. Leistungsbezogene Vergleiche – international	183
4. Zusammenschau	187
IX. Schullaufbahneempfehlungen von Lehrkräften für Kinder am Ende der vierten Jahrgangsstufe.....	191
<i>Wilfried Bos, Andreas Voss, Eva-Maria Lankes, Knut Schwippert, Oliver Thiel, Renate Valtin</i>	
1. Kompetenz als Kriterium der Schullaufbahneempfehlung.....	193
2. Schulnoten als Kriterium der Schullaufbahneempfehlung	204
3. Soziale Merkmale als Kriterium der Schullaufbahneempfehlung.....	211
4. Zusammenfassendes Prognosemodell der Schullaufbahneempfehlung	215
5. Schullaufbahneempfehlung und späterer Bildungsabschluss	218
6. Zusammenschau	223
Abbildungsverzeichnis	229
Tabellenverzeichnis.....	233

I. IGLU – Einige Länder der Bundesrepublik Deutschland im Vergleich

Wilfried Bos, Eva-Maria Lankes, Manfred Prenzel, Knut Schwippert, Renate Valtin, Gerd Walther

1. Vorbemerkung

Mit dem hier vorgelegten Band stellen wir aus der Perspektive eines innerdeutschen Vergleichs Ergebnisse der Internationalen Grundschul-Lese-Untersuchung (IGLU) und der Erweiterungsstudie (IGLU-E) vor, in der Kompetenzen in Mathematik, Naturwissenschaften und Orthographie erhoben wurden. Dies ist der zweite von drei Berichtsbänden, der weitere Erkenntnisse über die Leistungen in den getesteten Fachkompetenzen, über deren Einordnung in die internationalen und nationalen Befunde sowie über ausgewählte Bedingungen des lernrelevanten Hintergrunds liefert. Der dritte Band wird eine noch ausführlichere Darstellung der Leistungen im Kontext der international und national zusätzlich erhobenen Einfluss- und Bedingungsvariablen enthalten.

Weltweit beteiligten sich 35 Staaten mit repräsentativen Stichproben an diesem Projekt. National war nicht geplant, repräsentative Stichproben für einen Bundesländervergleich zu erheben. Einzelne Bundesländer meldeten bei der nationalen Projektleitung eigene erweiterte Forschungsfragestellungen an, die im Rahmen von IGLU zusätzlich untersucht wurden. So war ein Bundesland besonders an dem Verhältnis von Testleistung, Notengebung und erteilten Übergangsempfehlungen an weiterführende Schulen interessiert; ein anderes war neben den in IGLU getesteten Kompetenzen an dem Stand der motorischen Entwicklung seiner Viertklässler interessiert; ein weiteres hatte ein besonderes Interesse daran, den Beitrag von Leistungsrückmeldungen für die Schulentwicklungsprozesse der betroffenen Schulen zu untersuchen. Einige Länder waren explizit auch an Leistungsvergleichen interessiert. Um diese einzelnen Fragestellungen verfolgen zu können, wurde in diesen Ländern die Stichprobe so ausgeweitet, dass entsprechend der Fragestellung zufallskritisch abgesicherte Antworten gegeben werden können. Insgesamt wurde in sechs Ländern der Bundesrepublik Deutschland die Stichprobe so vergrößert, dass den länderspezifischen Fragestellungen nachgegangen werden konnte. In einem Flächenland war bereits die einfache Stichprobe groß genug, um differenziert einzelnen Fragestellungen nachgehen zu können. Da mit der Vergrößerung der Stichprobe auch Ländervergleiche bezüglich der in IGLU erhobenen Kompetenzen und Rahmenbedingun-

gen möglich wurden, beauftragten die nun insgesamt sieben Länder die Projektleitung, einen derartigen Vergleich vorzunehmen und darüber einen entsprechenden Bericht vorzulegen.

Die IGLU-Konzeption für die Grundschule beruht auf einer Vorstellung von Grundbildung (*Literacy*) und betont die kulturelle Bedeutung von Bildungsinhalten. Die Testaufgaben sprechen Kompetenzen im Leseverständnis, in Mathematik und in den Naturwissenschaften an, die in der Lebenswelt von Grundschulkindern bedeutsam sind und zugleich eine wichtige Basis für das anschließende Lernen in der Sekundarstufe I darstellen. Mit Hilfe von Fragebögen an Eltern, Lehrer, Schulleiter und Kinder wurde außerdem ein breites Spektrum von Zusatzinformationen erhoben. Diese Informationen geben Aufschluss über Schulen, Unterricht, Unterstützungssysteme, Einstellungen und Interessen der Befragten aus insgesamt 35 nationalen Bildungssystemen in den teilnehmenden Staaten.

Die Befunde aus IGLU und IGLU-E zeigen, dass Schülerinnen und Schüler am Ende der vierten Jahrgangsstufe in Deutschland internationale Vergleiche nicht scheuen müssen. Die Ergebnisse liefern keine Hinweise auf generelle Leistungsdefizite. Gleichwohl zeichnen sich für einige der besonders schwerwiegenden Probleme, die mit PISA für das Ende der Sekundarstufe I sichtbar wurden, die Anfänge bereits in der Grundschule ab: Kinder aus Familien mit Migrationshintergrund liegen bereits am Ende der Grundschulzeit in ihren Leistungen weit hinter Kindern ohne Migrationshintergrund zurück; der Zusammenhang zwischen Sozialschichtzugehörigkeit und Leistung ist deutlich; eine zwar kleine, aber angesichts des allgemein bildenden Auftrags der Grundschule nicht unerhebliche Anzahl von Risikokindern wird dem Unterricht in den weiterführenden Schulen nur schwer folgen können. Die Grundschule ist die Einrichtung, die im Rahmen des deutschen Bildungssystems für die Förderung aller Schülerinnen und Schüler unabhängig von ihrer sozialen und kulturellen Herkunft und den bereits erbrachten Vorleistungen zuständig ist. Einerseits ist sie für die Grundlegung der Bildung verantwortlich, andererseits hat sie am Ende der Grundschulzeit Schullaufbahneempfehlungen für den Besuch einer weiterführenden Schulart auszusprechen. Angesichts der international vergleichsweise geringen Dauer der Grundschule in Deutschland hat sie somit gerade im deutschen Bildungswesen eine zentrale Förder- und Ausgleichsfunktion, die weiter entwickelt werden muss. Deutlich wird aber auch, dass sich die Probleme auf Ebene der Sekundarstufe I massiv verschärfen und zuspitzen. Deshalb reicht eine Optimierung auf Grundschulebene allein nicht aus.

In dem vorliegenden Band werden die entsprechenden Befunde für einige Länder der Bundesrepublik Deutschland vorgelegt und in die internationalen Ergebnisse

eingeordnet. Das Design der Studie, die Methoden der Auswertung und die der Studie zugrunde liegenden Rahmenkonzeptionen wurden im ersten Band ausführlich dargestellt. Erklärungen zum Hintergrund der Studie, die für das Verständnis der hier berichteten Ergebnisse unerlässlich sind, werden aus dem ersten Band übernommen bzw. zusammengefasst dargestellt. Für darüber hinausgehende Informationen wird an den entsprechenden Stellen auf die Kapitel des ersten Bandes verwiesen.

2. Finanzierung und politische Administration

Mit dem Konstanzer Beschluss vom 24. Oktober 1997 hat die Kultusministerkonferenz (KMK) die Frage nach dem Leistungsvermögen von Schülerinnen und Schülern in Deutschland thematisiert und die Beteiligung an entsprechenden internationalen Vergleichsuntersuchungen als wichtige Ergänzung der länderbezogenen Qualitätssicherungsmaßnahmen im Bildungsbereich eindeutig befürwortet. In Konsequenz dieses Beschlusses hat die KMK am 25. Mai 2000 der Durchführung von IGLU zugestimmt. Der Durchführung des Ländervergleichs stimmte die Amtschefkonferenz auf ihrer 172. Sitzung am 7. November 2002 zu.

IGLU und IGLU-E sind durch Drittmittel geförderte Zuwendungsforschungen. Die Durchführung von IGLU wurde je zur Hälfte durch Zuwendungen des Bundesministeriums für Bildung, Wissenschaft, Forschung und Technologie (BMBF, Förderkennzeichen: IGL2000) und der Ständigen Konferenz der Kultusminister der Länder gefördert. IGLU-E wurde anteilmäßig von den zwölf Ländern gefördert, die sich daran beteiligten. Die Federführung für die Administration von IGLU-E und dem Ländervergleich wurde vom Land Schleswig-Holstein übernommen. Der Ländervergleich wurde im Auftrag der sieben zu vergleichenden Länder durchgeführt und von diesen finanziert. Es bestehen Berichtspflichten gegenüber den Zuwendungsgebern und Auftraggebern. Die wissenschaftliche Publikationsfreiheit ist jedoch nicht eingeschränkt. Ebenso wenig wurde von den Zuwendungsgebern und Auftraggebern in irgendeiner Weise Einfluss auf Form und Inhalt der Berichterstattung ausgeübt. Der vorliegende Bericht wird ausschließlich von den beteiligten Wissenschaftlerinnen und Wissenschaftlern verantwortet.

3. Die Organisation von IGLU

IGLU wird weltweit unter der Bezeichnung *Progress in International Reading Literacy Study* (PIRLS) durchgeführt. International wird diese Studie – wie bereits TIMSS – von der *International Association for the Evaluation of Educational Achievement* (IEA) verantwortet. Die IEA hat mit der Durchführung und Organisation Michael O. Martin und Ina V.S. Mullis am *International Study Center* (ISC) des Boston College in Chestnut Hill / MA, USA betraut. Die Arbeitsgruppe war verantwortlich für Entwicklung, Design und Implementation der Studie, für die internationale Koordination der Entwicklung der Instrumente und der Erhebungsprozeduren, für die Qualität der Datenerhebung, für die Datenanalyse und den internationalen Bericht der Ergebnisse. Marc Joncas von *Statistics Canada* verantwortete die Stichprobenziehung, die Anpassung des internationalen Designs an nationale Besonderheiten, die Dokumentation der nationalen Stichproben und die Berechnung der internationalen Stichprobengewichte.

In jedem Teilnehmerland verantwortet international und national ein *National Research Coordinator* (NRC) die nationale Durchführung und Leitung der Studie. In Deutschland übernahm diese Aufgabe Wilfried Bos. Zur Durchführung der Studie wurde in Deutschland ein nationales Konsortium eingerichtet, dem folgende Mitglieder angehören:

- Wilfried Bos, Lehrstuhl für Quantitative Methoden und Internationale Bildungsforschung am Institut für International und Interkulturell Vergleichende Erziehungswissenschaft des Fachbereichs Erziehungswissenschaft an der Universität Hamburg;
- Manfred Prenzel, Geschäftsführender Direktor des Leibniz-Instituts für die Pädagogik der Naturwissenschaften (IPN) an der Christian-Albrechts-Universität zu Kiel;
- Renate Valtin, Lehrstuhl für Grundschulpädagogik am Institut Erziehungswissenschaft an der Humboldt-Universität zu Berlin;
- Gerd Walther, Lehrstuhl für Didaktik der Mathematik am Mathematischen Seminar der Christian-Albrechts-Universität zu Kiel.

Die Projektleitung am Lehrstuhl für Quantitative Methoden und Internationale Bildungsforschung an der Universität Hamburg obliegt Eva-Maria Lankes, das Datenmanagement und die Koordination der Datenanalysen verantwortet hier Knut Schwippert. Mit der Organisation der Datenerhebung und Datenverarbeitung wurde international und in Deutschland auch für die Übernahme der Feldarbeit das *Data Processing Center* (DPC) der IEA in Hamburg beauftragt.

Zum IGLU-Team an der Universität Hamburg gehören weiterhin: Ingola Mohr, Nike Plabmeier und Andreas Voss. Weitere Mitarbeiterinnen und Mitarbeiter im Projekt sind Isolde Badel (Humboldt-Universität zu Berlin), Helmut Geiser, Rolf Langeheine und Kirstin Lobemeier (IPN Kiel). Für die Herstellung und Gestaltung der Druckvorlage war Heike Poppendieker (Universität Hamburg) zuständig.

Für IGLU-Arbeiten sind im DPC Hamburg zuständig: Svenja Bundt, Steffen Knoll, Heiko Sibberns und Anja Waschk.

Bei der Auswahl und Konstruktion der nationalen Testinstrumente waren zwei Arbeitsgruppen beteiligt:

Arbeitsgruppe Mathematikdidaktik:

- Gerd Walther (Sprecher, Mitglied des Konsortiums), Universität Kiel
- Peter Bender, Universität Paderborn
- Detlef Lind, Universität Wuppertal
- Petra Scherer, Universität Bielefeld
- Christoph Selter, Pädagogische Hochschule Heidelberg
- Elsbeth Stern, Max-Planck-Institut für Bildungsforschung, Berlin
- Bernd Wollring, Universität Kassel

Beratend:

- Gabriele Kaiser, Universität Hamburg
- Jörg Voigt, Universität Münster
- Erich Christian Wittmann, Universität Dortmund

Arbeitsgruppe Didaktik der Naturwissenschaften:

- Manfred Prenzel (Sprecher, Mitglied des Konsortiums), Universität Kiel
- Horst Bayrhuber, Universität Kiel
- Reinhard Demuth, Universität Kiel
- Reinders Duit, Universität Kiel
- Wolfgang Einsiedler, Universität Erlangen-Nürnberg
- Manfred Euler, Universität Kiel
- Gabriele Faust-Siehl, Universität Frankfurt
- Hans Fischer, Universität Dortmund
- Helmut Geiser, Universität Kiel
- Uwe Hameyer, Universität Kiel
- Joachim Kahlert, Universität München
- Walter Köhnlein, Universität Hildesheim
- Rolf Langeheine, Universität Kiel

- Eva-Maria Lankes, Universität Hamburg
- Roland Lauterbach, Universität Leipzig
- Gisela Lück, Universität Bielefeld
- Kornelia Möller, Universität Münster
- Ilka Parchmann, Universität Kiel
- Helmut Schreier, Universität Hamburg

Die Arbeit wurde beratend begleitet von einem wissenschaftlichen Beirat, dem folgende Personen angehören:

- Jürgen Baumert, Max-Planck-Institut für Bildungsforschung, Berlin
- Hans Brügelmann (zum Sept. 2002 ausgeschieden), Universität Siegen
- Reinhard Demuth, Universität Kiel
- Hartmut Ditton, Universität Osnabrück
- Wolfgang Einsiedler, Universität Erlangen-Nürnberg
- Ingrid Gogolin, Universität Hamburg
- Sabine Gruehn, Universität Berlin
- Petra Hanke, Universität Münster
- Heinz-Günter Holtappels, Universität Dortmund
- Dagmar Killus, Universität Potsdam
- Klaus Klemm, Universität Gesamthochschule Essen
- Olaf Köller, Universität Erlangen-Nürnberg
- Rainer Lehmann, Universität Berlin
- Rainer Peek, Landesinstitut für Schule, Soest
- Hans-Günter Rolff, Universität Dortmund
- Hans-Günther Roßbach, Universität Bamberg
- Gerheid Scheerer-Neumann, Universität Potsdam
- Thomas Trautmann, Universität Hamburg

An der Vorbereitung, Durchführung und Auswertung der Studie waren viele studentische Hilfskräfte beteiligt, die hier nicht alle namentlich erwähnt werden können. Für ihre engagierte Unterstützung und Mitarbeit sei ihnen an dieser Stelle freundlich gedankt.

Ganz besonderer Dank soll auch all denen ausgesprochen werden, um die es in dieser Studie geht, den Schülerinnen und Schülern, ihren Eltern, ihren Lehrerinnen und Lehrern und den Schulleitungen. Ohne ihre hohe Bereitschaft, die Tests zu bearbeiten, die Fragebögen auszufüllen und wichtige Informationen zur Verfügung zu stellen, wäre diese Studie nicht möglich gewesen.

II. Anlage und Durchführung der Internationalen Grundschul-Lese-Untersuchung (IGLU) und ihrer Erweiterung um Mathematik und Naturwissenschaften (IGLU-E)

Eva-Maria Lankes, Wilfried Bos, Nike Plafßmeier, Knut Schwippert, Andreas Voss

In der Internationalen Grundschul-Lese-Untersuchung (IGLU) wurden die Fähigkeiten von Kindern im Leseverständnis am Ende der vierten Jahrgangsstufe untersucht. International nahmen 35 Staaten¹ an dieser von der *International Association for the Evaluation of Educational Achievement* (IEA) unter dem Namen *Progress in International Reading Literacy Study* (PIRLS) initiierten und organisierten Studie teil. In Deutschland wurde die Studie um einen zweiten Testtag erweitert, an dem Tests in Mathematik, Naturwissenschaften und Rechtschreiben durchgeführt wurden und ein Aufsatz geschrieben wurde (IGLU-E). Der vorliegende Band ergänzt die Berichterstattung um eine neue Perspektive: Wichtige Befunde der Studie werden für einige Länder der Bundesrepublik Deutschland im Vergleich dargestellt. Die Anlage der Studie wird im ersten Ergebnisbericht (Bos et al., 2003) ausführlich beschrieben. Im folgenden Kapitel werden die internationalen Verfahren und Prozeduren (vgl. Lankes et al., 2003b) soweit zusammengefasst, wie sie für das Verständnis der dargestellten Befunde notwendig sind. Der Schwerpunkt der Ausführungen liegt auf den Besonderheiten der nationalen Anlage und der Durchführung der Studie in den hier berichteten sieben Ländern der Bundesrepublik Deutschland.

1. Die Untersuchungspopulation

1.1 Stichprobe und Vergleichsgruppen

Weltweit wurden im Jahr 2001 im Lesen 146.490 Grundschülerinnen und -schüler aus 35 Staaten getestet. In Deutschland nahmen 10.571 Schülerinnen und Schüler aus allen 16 Ländern an der internationalen Lesestudie (1. Testtag) teil. Alle Grundschulen, deren Kinder der vierten Jahrgangsstufe in der Lage waren, den Test zu bearbeiten (also auch Kinder an Sonderschulen), gingen in die Stichprobenziehung

1 Zur eindeutigen Unterscheidung wird in diesem Band international von ‚Staaten‘, national von ‚Ländern der Bundesrepublik Deutschland‘ oder auch nur von ‚Ländern‘ gesprochen.

ein. Die internationalen Vorgaben erlaubten es, Schulen mit geistig, körperlich oder mehrfach behinderten Schülerinnen und Schülern sowie sehr kleine Schulen nach festgelegten Kriterien vor der Stichprobenziehung von der national angestrebten Stichprobe auszuschließen. Der Ausschluss durfte allerdings nicht mehr als 5 Prozent der Population betragen, für die eine Aussage gemacht werden sollte, damit das Land noch in der internationalen Berichtlegung verbleiben konnte.

Tabelle II.1: Schulstichprobe in den Ländern der Bundesrepublik Deutschland

Land	Intern. Stichprobe	Oversampling (Zufallsstichprobe)	Gesamtanzahl der gezogenen Schulen	nachträglich ausgeschlossen	nicht teilgenommen	dafür Ersatzschule	in die internationale Auswertung einbezogen	zusätzliches Oversampling (keine Zufallsstichprobe)
BW	20	5	25		1		24	
BY	23	2	25				25	
BE	5		5				5	
BB	4	21	25		1	1	25	
HB	2	23	25		2	1	24	12
HH	2		2				2	
HE	11	14	25				25	
MV	3		3		1		2	
NI	15		15	1			14	
NW	35		35		1		34	
RP	8		8				8	
SL	2		2				2	
SN	7		7				7	
ST	4		4				4	
SH	6		6				6	
TH	4		4				4	19
Gesamt	151	65	216	1	6	2	211	31

Baden-Württemberg (BW), Bayern (BY), Berlin (BE), Brandenburg (BB), Bremen (HB), Hamburg (HH), Hessen (HE), Mecklenburg-Vorpommern (MV), Niedersachsen (NI), Nordrhein-Westfalen (NW), Rheinland-Pfalz (RP), Saarland (SL), Sachsen (SN), Sachsen-Anhalt (ST), Schleswig-Holstein (SH), Thüringen (TH)

IEA: Progress in International Reading Literacy Study

© IGLU-Germany

In Deutschland waren lediglich sehr kleine Schulen und diejenigen Sonderschulen ausgeschlossen, deren Kinder aus körperlichen oder geistigen Gründen den Test nicht durchführen konnten. Die Ausschlussrate aus diesen beiden Gründen betrug in Deutschland 0,8 Prozent.² Für die internationale Stichprobe wurden 151 Schulen aus allen Ländern der Bundesrepublik Deutschland proportional zur Einwohnerzahl des

2 Detaillierte Informationen zur Stichprobenziehung in der PIRLS-Untersuchung in Martin, Mullis & Kennedy, 2003.

jeweiligen Landes gezogen (vgl. Tab. II.1). Weil einige Länder ihre Stichprobe erweiterten (*Oversampling*), umfasste die den Zufallskriterien genügende Stichprobe in Deutschland 216 Schulen. Unter den 216 Schulen befand sich eine Förderschule, die eigentlich unter die Ausschlusskriterien gefallen wäre, aber nicht als solche erkannt worden war. Sie wurde nachträglich ausgeschlossen, da sich im Verlauf des Tests erwies, dass die Schülerinnen und Schüler nicht in der Lage waren, den Test zu bearbeiten. Unter der Kategorie ‚Nicht teilgenommen‘ befinden sich drei Schulen, die die international vorgeschriebene Mindestteilnahmequote von 50 Prozent nicht erreichten und drei, die nicht bereit waren, am Test teilzunehmen. Für zwei Schulen konnten Ersatzschulen gezogen werden. In der internationalen Auswertung des ersten Testtages wurden somit 211 Schulen (98 %) berücksichtigt.

Tabelle II.2: Vergleichsgruppen

Deutschland	VG 1	VG 2	
Baden-Württemberg (BW)	Deutschland (DEU)	Bulgarien	Niederlande
Bayern (BY)	England (ENG)	Deutschland	Norwegen
Brandenburg (BB)	Frankreich (FRA)	England	Rumänien
Bremen (HB)	Griechenland (GRC)	Frankreich	Schottland
Hessen (HE)	Italien (ITA)	Griechenland	Schweden
Nordrhein-Westfalen (NW)	Niederlande (NDL)	Island	Slowakei
Thüringen (TH)	Schottland (SCO)	Italien	Slowenien
	Schweden (SWE)	Kanada	Tschechien
		Lettland	Ungarn
		Litauen	USA
		Neuseeland	Zypern

IEA: Progress in International Reading Literacy Study

© IGLU-Germany

Im vorliegenden Bericht werden Ergebnisse aus den Ländern der Bundesrepublik berichtet, die mit einer ausreichend großen Zufallsstichprobe an den Tests teilgenommen haben. Das sind die Länder Baden-Württemberg, Bayern, Brandenburg, Bremen, Hessen und Nordrhein-Westfalen. In zwei Ländern wurden darüber hinaus bestimmte, ausgewählte Schulen (insgesamt 31 Schulen) unter besonderen Kriterien mit in die Stichprobe aufgenommen, die jedoch nicht dem Anspruch einer Zufallsstichprobe genügen.

In *Bremen* sollten in einer zusätzlichen wissenschaftlichen Untersuchung zwei Gruppen von Schulen miteinander verglichen werden. Zu diesem Zweck wurden zusätzlich zu 25 zufällig gezogenen Schulen noch zwölf weitere Schulen gezielt ausgewählt, die weder bei den internationalen Vergleichen noch in diesem Band berücksichtigt werden. Die in Bremen gezogene Zufallsstichprobe von 25 Schulen entspricht den Vorgaben für den Ländervergleich und liefert repräsentative Daten.

In *Thüringen* sollten mit der Erweiterung der Stichprobe Schulentwicklungsprozesse unterstützt werden. Deshalb konnten sich alle Grundschulen um eine Teilnahme bewerben. Zusätzlich zu vier zufällig gezogenen Schulen wurden vom Ministerium des Landes aus den 43 Bewerbungen nach Zufallszahlen 19 Schulen für eine Teilnahme ausgewählt. An diesen Schulen nahmen dann alle Klassen an der Erhebung teil. Diese Schulen erhielten in der internationalen Auswertung ein Gewicht von 0 und blieben damit unberücksichtigt. In diesem Band werden die Thüringer Ergebnisse der 23 Schulen berichtet, die teilgenommen haben. Weil es sich bei diesen Schulen nicht um eine Zufallsstichprobe handelt, werden die Ergebnisse ungewichtet und entsprechend internationalen Gepflogenheiten am Ende der jeweiligen Tabellen und Abbildungen mit einer Fußnote versehen berichtet.

Die Befunde zum Lesen für die sieben hier berichteten Länder (vgl. Kap. IV) werden in diesem Band zu verschiedenen Kennwerten in Beziehung gesetzt (Tab. II.2):

- Innerhalb Deutschlands können die sieben Länder miteinander sowie mit dem bundesdeutschen Mittelwert (Deutschland) verglichen werden.
- International werden die Befunde der sechs Länder (ohne Thüringen) eingereiht in die Darstellung aller Staaten außer den Ländern der Vergleichsgruppe 3 (vgl. Lankes et al., 2003b, S. 8 f.).
- Der Kennwert VG 2 enthält den Mittelwert der Vergleichsgruppe 2, zu der alle an IGLU teilnehmenden Länder der OECD – ohne die Türkei – und die IGLU-Länder gehören, die bis zum Jahre 2007 Mitgliedsländer der Europäischen Union werden (vgl. Lankes et al., 2003b, S. 9).
- Der Kennwert VG 1 enthält den Mittelwert der Vergleichsgruppe 1, das sind die acht Länder der europäischen Union, die an der Studie teilgenommen haben.

Die Befunde in Mathematik und Naturwissenschaften für die sechs Länder, die am zweiten Testtag teilgenommen haben (außer Brandenburg), werden auf der Basis der Staaten verglichen, die an der TIMS-Grundschulstudie teilgenommen haben (vgl. Kap. V und VI).

1.2 Stichprobenziehung, Gewichtung und die Schätzung von Standardfehlern

Bei einer internationalen Vergleichsuntersuchung muss die Stichprobe, anhand derer ein Schulsystem beschrieben werden soll, groß genug sein, um statistisch verlässliche Aussagen machen zu können. Da die IGLU-Studie in die international durchgeführte PIRLS-Untersuchung eingebettet ist, wurden für die internationale Stichprobe von IGLU die in der IEA vorgegebenen Standards für die Stichprobenziehung an-

gewendet. Die Stichproben in den einzelnen teilnehmenden Nationen waren so wählen, dass Mittelwerte $\pm 1/10$ Standardabweichung mit einer Irrtumswahrscheinlichkeit von ≤ 5 Prozent zufallskritisch abgesichert werden können. Als Planungsgrundlage für den Stichprobenumfang in der PIRLS-Untersuchung wurde von der internationalen Studienleitung vorgegeben, dass mindestens 150 Schulen in der Stichprobe enthalten sein sollen. Für den Vergleich von Ländern der Bundesrepublik Deutschland wurde ein etwas liberaleres Qualitätskriterium angelegt. Hier sollten Mittelwertsunterschiede mit $\pm 1/5$ Standardabweichung mit 5-prozentiger Irrtumswahrscheinlichkeit bestimmt werden können.

Die Stichprobenziehung in IGLU basierte auf einem zweistufigen Verfahren: Grundsätzlich wurden zuerst je Land Schulen gezogen, dann Klassen innerhalb der Schulen (Lankes et al., 2003b). Um länderspezifische Untersuchungen zu ermöglichen, lässt das Untersuchungsdesign zu, dass die Stichproben der Länder jeweils so weit erhöht werden, dass interessierende Fragestellungen zufallskritisch abgesichert werden können. Fünf Länder der Bundesrepublik Deutschland haben sich für eine derartige überproportionale Stichprobe (*Oversampling*) entschieden (Baden-Württemberg, Bayern, Brandenburg, Bremen und Hessen). In Nordrhein-Westfalen hatte die Zufallsstichprobe bereits den für den Ländervergleich erforderlichen Umfang. Um in den Analysen die Proportionen in den realen Schülerzahlen der Länder beizubehalten, wurden die Länder mit ausgeweiteter Zufallsstichprobe gewichtet. Die erweiterte Stichprobe Thüringens wird aufgrund des Auswahlverfahrens nicht in den direkten Vergleich einbezogen, jedoch werden dort, wo es möglich ist, die Befunde denen der anderen Länder gegenüber gestellt. Die für Thüringen berichteten Befunde basieren allerdings auf ungewichteten Analysen. Ein Vergleich der IGLU-Leistungswerte zwischen zufällig und systematisch ausgewählten Schulen Thüringens und Befunden aus den 2003 in Thüringen durchgeführten Kompetenztests deuten jedoch darauf hin, dass die hier vorgelegten Befunde die landestypischen Kompetenzen der Schülerinnen und Schüler tendenziell eher etwas überschätzen.³

Bei disproportionalen, einfach stratifizierten Klumpenstichproben, wie der in IGLU vorliegenden, können die üblichen Analyseverfahren den Standardfehler systematisch unterschätzen. Wie sehr diese Unterschätzung von den tatsächlichen Werten abweicht, hängt von der Homogenität der Cluster ab. Dieses Problem kann mit so genannten *Jackknife*-Verfahren angemessen berücksichtigt werden. Sie bestimmen die Variabilität der Schätzung von Populationskennwerten durch das wiederholte Schätzen dieser Werte aus Substichproben und bieten damit die Möglichkeit, Stich-

3 Wir danken Herrn Prof. Nachtigall von der Friedrich-Schiller-Universität in Jena für die vorzeitige Bereitstellung der Informationen aus seinem Projekt.

probenfehler zu schätzen ohne die Annahme einfacher Zufallsstichproben voraussetzen zu müssen. Durch die Reduzierung der Freiheitsgrade führt diese Methode zu konservativen Schätzungen. In IGLU werden alle Standardfehler durch *Jackknifing* (*Jackknife Repeating Replication Technique*, JRR) geschätzt (Baumert, Bos & Lehmann, 2000; Gonzalez & Foy, 2000; Johnson & Rust, 1992; Wolter, 1985).

1.3 Nationale Beteiligungsquoten

Auch innerhalb der Schulen gab es eng definierte Möglichkeiten, einzelne Kinder vom Test auszuschließen: Schüler fremder Muttersprache, die weniger als ein Jahr in der Landessprache unterrichtet worden sind; körperlich und geistig behinderte Schüler, die nicht in der Lage waren, den Test auszuführen oder (z.B. in integrierten Klassen) nicht an den regulären Klassenarbeiten teilnehmen. In Deutschland waren in den 211 Schulen insgesamt 8.997 Schülerinnen und Schüler in die Stichprobe einbezogen. 58 Kinder wurden aufgrund der bereits genannten Kriterien vom Test ausgeschlossen, 27 Kinder waren verzogen und wurden – um Doppelzählungen zu vermeiden – aus der Schülerstichprobe ausgeschlossen. Die Beteiligungsquote der Schülerinnen und Schüler beträgt ungewichtet 86,7 Prozent (gewichtet 88,1 %). Die gesamte Ausschöpfung über die Schulen sowie Schülerinnen und Schüler hinweg beträgt ungewichtet 85,1 Prozent (gewichtet 86,4 %) der zu erreichenden Zielpopulation und entspricht damit den internationalen Vorgaben. Nach der Gewichtung sind im Datensatz 7.633 Schülerinnen und Schüler berücksichtigt.

Eine nach internationalen Vorgaben zufriedenstellende Stichprobe (Tab. II.3) muss *entweder* (a) eine Schulteilnahme von mindestens 85 Prozent ohne Ersatzschulen *und* gleichzeitig eine Schülerteilnahmequote von mindestens 85 Prozent aufweisen (Bayern, Brandenburg, Hessen und Nordrhein-Westfalen) *oder* (b) in Kombination der beiden Quoten (Produkt aus gewichteter Schulteilnahme ohne Ersatzschulen und gewichteter Schülerteilnahme) mindestens einen Wert von 75 Prozent erreichen (Baden-Württemberg). Erreicht ein Land diese Vorgaben erst nach Einbeziehung von Ersatzschulen, dann wird dieses Land in den Berichten mit einem Stern gekennzeichnet (Bremen). Die Ausschlüsse von Schülerinnen und Schülern durch Lehrkräfte im Vorfeld der Tests liegen in allen Ländern der Bundesrepublik im international vertretbaren Rahmen und unterscheiden sich kaum vom deutschen Mittelwert. Die Teilnahmequoten für Thüringen sind ohne Bedeutung, da es sich hier nicht um eine Zufallsstichprobe handelt und damit die Repräsentativität ohnehin nicht gegeben ist.

Thüringen wird in den Berichten immer abgesetzt und mit einer Fußnote erläutert (Einzelheiten zu den Bestimmungen siehe Martin et al., 2003, S. 125).⁴

Tabelle II.3: Teilnahmequoten in den hier berichteten Ländern der Bundesrepublik Deutschland am 1. Testtag

	Schülerteilnahmequote		Schulteilnahmequote				Kombination aus gewichteter Schüler- und gewichteter Schulteilnahmequote	
	(gewichtet)	(ungewichtet)	(gewichtet)		(ungewichtet)		Original-schulen	Mit Ersatz-schule
			Original-schulen	Mit Ersatz-schule	Original-schulen	Mit Ersatz-schule		
BW	83%	81%	96%		96%		80%	
BY	88%	87%	100%		100%		88%	
BB	93%	93%	96%	100%	96%	100%	88%	93%
HB	81%	77%	89%	96%	92%	96%	71%	77%
HE	93%	92%	100%		100%		93%	
NW	88%	88%	97%		97%		86%	

Abkürzungen vgl. Tabelle II.2.

IEA: Progress in International Reading Literacy Study

© IGLU-Germany

Analog zu PISA (Baumert et al., 2002; Baumert, Bos & Watermann, 2000) wurde geprüft, ob es Hinweise darauf gibt, dass geringere Beteiligungsquoten zu einer Überschätzung der Leistungswerte führen könnten. Die höchste Schülerbeteiligung (gewichtet) wurde mit 93 Prozent in Brandenburg und Hessen erreicht, die niedrigste in Bremen mit 81 Prozent. Die Tabelle II.4 geht von dem *Worst-Case-Scenario* eines perfekten Zusammenhangs zwischen Beteiligungsbereitschaft und Testleistung aus. Die Tabelle vermittelt einen Eindruck von der Veränderung der Leistungsmittelwerte und der Leistungsstreuungen, wenn sukzessive die schwächsten Schülerinnen und Schüler aus den Stichproben ausgeschlossen werden. Unter dieser mit Sicherheit unzutreffenden Annahme würde der Abstand von Bremen zu den übrigen hier berichteten Ländern in allen Fächern noch etwas größer – ein Hinweis darauf, dass die Mittelwerte für Bremen möglicherweise leicht überschätzt werden. Realistischer ist jedoch die Annahme, dass die Leistungsstreuung der Grundschulen in Bremen ein ähnliches Ausmaß haben sollte wie etwa in Brandenburg, einem Land mit ebenfalls niedrigen Leistungsergebnissen. Der Kennwert von Brandenburg bei einer annähernd vergleichbaren Leistungsstreuung liegt im Lesen 19 Punkte, aber nicht signifikant über dem von Bremen (vgl. Tab. II.4).

4 Zu internationalen Stichprobenbesonderheiten vgl. Lankes et al., 2003b, S. 11.

Tabelle II.4: Vergleich der Leistungen im Lesen, in Mathematik und Naturwissenschaften in Abhängigkeit von der Beteiligungsrate unter der Annahme eines perfekten Zusammenhangs zwischen Teilnahmebereitschaft und Testleistung (Mittelwert der fünf *Plausible Values*)

Lesen												
	81 Prozent			83 Prozent			88 Prozent			93 Prozent		
	Mittelwert	(SE)	SD	Mittelwert	(SE)	SD	Mittelwert	(SE)	SD	Mittelwert	(SE)	SD
HB	507	(8,5)	72									
BW	554	(3,9)	57	549	(4,3)	63						
NW	544	(4,5)	57	540	(4,6)	60	531	(5,2)	69			
BY	559	(2,5)	53	555	(2,5)	57	546	(3,4)	67			
HE	561	(2,8)	49	558	(2,9)	51	552	(2,7)	56	544	(3,1)	65
BB	545	(3,3)	53	541	(3,5)	56	534	(3,6)	62	526	(4,4)	71

Mathematik												
	81 Prozent			83 Prozent			88 Prozent			93 Prozent		
	Mittelwert	(SE)	SD	Mittelwert	(SE)	SD	Mittelwert	(SE)	SD	Mittelwert	(SE)	SD
HB	512	(9,8)	83									
BW	571	(4,6)	72	565	(5,2)	78						
NW	551	(5,6)	70	546	(5,9)	73	537	(6,2)	81			
BY	561	(3,5)	69	556	(3,5)	72	547	(4,5)	82			
HE	566	(3,8)	63	562	(3,9)	65	555	(4,3)	70	547	(5,2)	78

Naturwissenschaften												
	81 Prozent			83 Prozent			88 Prozent			93 Prozent		
	Mittelwert	(SE)	SD	Mittelwert	(SE)	SD	Mittelwert	(SE)	SD	Mittelwert	(SE)	SD
HB	526	(11,0)	85									
BW	580	(5,3)	74	574	(5,7)	80						
NW	563	(5,7)	71	559	(6,1)	74	549	(6,7)	83			
BY	580	(3,7)	69	575	(3,6)	72	566	(4,8)	82			
HE	585	(4,4)	67	581	(4,8)	69	574	(4,9)	75	564	(5,4)	84

Abkürzungen vgl. Tabelle II.2.

IEA: Progress in International Reading Literacy Study

© IGLU-Germany

In Mathematik und den Naturwissenschaften fehlt eine solche Vergleichsebene, weil Brandenburg am zweiten Testtag nicht teilgenommen hat. Auch für Baden-Württemberg können derartige Überlegungen angestellt werden. Schneidet man die Stichprobe bei den von Baden-Württemberg erreichten 83 Prozent, so erreicht Bayern im Lesen ein um sechs und Hessen ein um neun Punkte besseres Ergebnis gegenüber dem baden-württembergischen Wert. Die Unterschiede zwischen Baden-

Württemberg, Bayern und Hessen sind jedoch auch zwischen den bei 83 Prozent geschnittenen Stichproben nicht signifikant. Bei einer annähernd vergleichbaren Leistungsstreuung zeigen sich keine signifikanten Leistungsunterschiede zwischen Baden-Württemberg, Bayern und Hessen. Dasselbe Bild bietet sich in den beiden anderen getesteten Fächern. Aus den Daten ergeben sich keine Hinweise auf eine Verzerrung der Stichprobe durch die etwas niedrigere Schülerteilnahmequote in Baden-Württemberg.

Tabelle II.5: Anteil an Schülern, für die Daten aus den verschiedenen Fragebögen vorliegen, in den hier berichteten Ländern der Bundesrepublik Deutschland (ungewichtet)

	Alle Bundesländer	BW	BY	BB	HB	HE	NW	TH
Schülerfragebogen	84%	79%	86%	83%*	75%	92%	86%	85%
Elternfragebogen	76%	74%	82%	83%	64%	83%	72%	78%
Schulleiterfragebogen	95%	92%	96%	92%	92%	100%	94%	100%
Deutschlehrerfragebogen	92%	95%	94%	92%	72%	98%	94%	100%
Mathematiklehrerfragebogen	93%	93%	96%	-	74%	98%	94%	100%
Sachunterrichtslehrerfragebogen	88%	91%	90%	-	66%	100%	93%	100%

* nur internationaler Teil des Schülerfragebogens (1. Testtag).
Abkürzungen vgl. Tabelle II.2.

IEA: Progress in International Reading Literacy Study

© IGLU-Germany

Von den 16 Ländern der Bundesrepublik Deutschland entschieden sich zwölf Länder für einen zweiten Testtag: Die Länder Brandenburg, Mecklenburg-Vorpommern, Niedersachsen und Sachsen-Anhalt haben an der Erweiterung um Mathematik und Naturwissenschaften (IGLU-E) nicht teilgenommen. Die Teilnahmequoten am zweiten Testtag unterscheiden sich nur geringfügig von denen am ersten Testtag. Der Rücklauf der Fragebögen war in den Ländern leicht unterschiedlich (vgl. Tab. II.5). Besonders auffallend sind die niedrigen Rücklaufquoten der Fragebögen für Eltern und Lehrkräfte in Bremen.

2. Testdesign und Testdurchführung

2.1 Test- und Fragebogenentwicklung

Die Entwicklung der Tests in IGLU und IGLU-E orientierte sich an einem Rahmenkonzept von Grundbildung, das Kompetenzen definiert, die über den unterrichtlichen Rahmen hinaus für die Lebensbewältigung in konkreten Anwendungssituationen nützlich sind. Im Bereich Deutsch wurden neben dem Leseverständnis am zweiten Testtag die Rechtschreibkompetenz und die Fähigkeit im Textschreiben erhoben. In der Erweiterung am zweiten Testtag wurden außerdem grundlegendes mathematisches Wissen und Können in inner- und außermathematischen Anwendungsfeldern getestet. Ein weiterer Bereich des zweiten Testtags nahm Kenntnisse und Kompetenzen über Naturphänomene und erste naturwissenschaftliche Konzepte in den Blick. Kompetenzen, die bei IGLU erhoben wurden, sind auch zu einem bedeutsamen Teil Gegenstand der Grundschul-Lehrpläne. Damit liefern die IGLU-Befunde Aussagen, die für die Weiterentwicklung von Unterrichtskonzepten und Lehrplänen der Grundschule eine wichtige Rolle spielen können.

Die Leistungen der Schülerinnen und Schüler werden dabei als Ergebnis eines Lernprozesses betrachtet, der nicht nur durch den Unterricht, sondern auch durch außerschulische und familiäre Merkmale und Bedingungen beeinflusst wird. Um das Geschehen im Unterricht und die vielfältigen Bedingungen, die die Schülerleistungen beeinflussen, möglichst mehrdimensional und mehrperspektivisch in den Blick nehmen zu können, wurden alle am Lerngeschehen beteiligten Gruppen befragt: Eltern und Kinder gaben Auskunft über die familiären Hintergründe der Schülerinnen und Schüler und über lern- und leistungsrelevante Merkmale des Kindes. Schüler, Lehrkräfte und Schulleitungen wurden nach ihrer jeweiligen Einschätzung und Wahrnehmung des Unterrichtsgeschehens befragt. Lehrkräfte und Schulleitungen lieferten Informationen über leistungsrelevante Rahmenbedingungen an der Schule (vgl. Lankes, Bos, Mohr, Plabmeier & Schwippert, 2003a).

2.2 Aufbau der Untersuchung

Am ersten Testtag (vgl. Tab. II.6) wurde in allen Ländern der Bundesrepublik Deutschland der internationale Lesetest durchgeführt, am zweiten Testtag (vgl. Tab. II.7) nahmen Kinder in zwölf Ländern an Tests in Mathematik und Naturwissenschaft sowie an einem Diktat oder Aufsatz teil. Ziel der Testauswertung ist es, aus der Vielzahl der beantworteten Testfragen für jeden Schüler einen Kennwert für das jeweils getestete Lese-, Mathematik- und Naturwissenschaftsverständnis sowie für

seine Fähigkeiten in Orthographie und Aufsatz zu ermitteln, um Aussagen über Leistungsunterschiede zwischen Gruppen von Schülerinnen und Schülern machen zu können. Damit die Fachbereiche in genügender Breite abgebildet, die Schüler aber gleichzeitig nicht über Gebühr belastet werden, kam ein Testdesign mit rotierten Aufgabenblöcken (*Multi Matrix Design*) zur Anwendung.

Tabelle II.6: Ablauf der Untersuchung und eingesetzte Instrumente in Deutschland am ersten Testtag (alle Länder der Bundesrepublik)

Testhefte												
Minuten	1	2	3	4	5	6	7	8	9	10	11	12
40	Li1	Li2	Li3	Li(R)	In1	In2	In3	Li(R)	Li1	In2	Li3	Li(R)
10	Pause											
40	Li2	Li3	In1	In(R)	In2	In3	Li1	In(R)	In1	Li2	In3	In(R)
15	Pause											
10	Test zu Analogien und Worterkennung											
20	Internationaler Schülerfragebogen											

Li = Lesen mit literarischer Absicht; In = Lesen mit informierender Absicht; (R) = Reader⁵.

IEA: Progress in International Reading Literacy Study

© IGLU-Germany

Tabelle II.7: Ablauf der Untersuchung und eingesetzte Instrumente in Deutschland am zweiten Testtag (12 Länder der Bundesrepublik)

Testhefte												
Minuten	1	2	3	4	5	6	1	2	3	4	5	6
20	N1	M2	N3	M4	N5	M6	N1	M2	N3	M4	N5	M6
20	M1	N2	M3	N4	M5	N6	M1	N2	M3	N4	M5	N6
10	Pause											
20	WO	WO	WO	WO	WO	WO	S	IRL	S	IRL	S	IRL
15	Pause											
10	Test zu Analogien											
20	Nationaler Schülerfragebogen											

M = Aufgaben zur Mathematik; N = Aufgaben zur Naturwissenschaft; IRL = International Reading Literacy; S = Schreibaufgabe; WO = Wortdiktat.

IEA: Progress in International Reading Literacy Study

© IGLU-Germany

Die Testhefte des zweiten Testtages enthielten in den Bereichen Mathematik und Naturwissenschaften überlappende Testaufgaben. Auch der Schülerfragebogen ent-

5 Die Testhefte 4, 8 und 12 bestanden im Gegensatz zu allen anderen Heften, in denen die Aufgaben direkt hinter dem Text angeordnet waren, aus zwei Teilen, einem bunt bebilderten Textheft (*Reader*) und einem Aufgabenheft.

hielt am zweiten Testtag rotierte Fragen. Alle Schülerinnen und Schüler bearbeiteten auf diese Weise Aufgabenblöcke zu den Fächern Lesen, Mathematik und Naturwissenschaft. Jede zweite Klasse erhielt am zweiten Testtag ein Wortdiktat, die anderen Klassen bearbeiteten je zur Hälfte eine Schreibaufgabe oder eine Aufgabe aus der IEA-Studie *International Reading Literacy Study* (Elley, 1992, 1994; Lehmann, Peek, Pieper & Stritzky, 1995). Die Tests zu figuralen und verbalen Analogien und Worterkennung sowie der internationale Schülerfragebogen wurden von allen Schülerinnen und Schülern in derselben Form bearbeitet. Der gesamte Test dauerte am ersten Testtag inklusive Pausen etwas mehr als zwei Stunden, am zweiten Testtag etwa zwei Stunden. Auch wenn ein Kind an einer Stelle nicht mehr weiter kam, gab es immer die Möglichkeit an einer anderen Stelle erfolgreich wieder einzusteigen.

2.3 Skalierung

Im Rahmen des angewendeten probabilistischen Testmodells werden Schülerfähigkeiten und Aufgabenschwierigkeiten simultan bestimmt. Ein Vorteil des probabilistischen Modells besteht darin, dass sich Schülerfähigkeiten und Aufgabenschwierigkeiten auf der gleichen Skala abbilden lassen. Personenfähigkeiten und Itemschwierigkeiten sind somit direkt aufeinander bezogen. Ist die Personenfähigkeit größer als die Fragenschwierigkeit, kann davon ausgegangen werden, dass der Schüler die Frage mit höherer Wahrscheinlichkeit richtig als falsch beantwortet.

Dieses Auswertungsverfahren eröffnet die Möglichkeit des so genannten *Proficiency Scaling* (Klieme, Baumert, Köller & Bos, 2000). Ausgangspunkt sind theoretische Vorstellungen über Merkmale, auf denen die zu messende Kompetenz beruht. Die Aufgaben werden nach Schwierigkeit und Anforderungen analysiert. Damit können auf der Fähigkeitsskala inhaltlich definierte Stufen der Kompetenz beschrieben und durch Itembeispiele illustriert werden. Im Rahmen der IGLU-Studie wurden sowohl für das Leseverständnis als auch für die mathematischen und naturwissenschaftlichen Fähigkeiten entsprechende Kompetenzstufen formuliert (Lankes et al., 2003b).

Um anhand der Testinstrumente valide Schätzer für die zu erfassenden Kompetenzen (Lesen, Mathematik, Naturwissenschaften) zu erhalten,⁶ müssen Tests über eine ausreichende Länge und genügend Leistungsfragen verfügen. Da aber Grundschulkindern solch theoretisch geforderte Testdauer forschungspraktisch nicht zugemutet werden kann, bekommt jeder Schüler nur einen Teil der im Gesamttest enthaltenen Aufgaben. Um unter Berücksichtigung der Fragenauswahl und der in der Stichprobe

6 Zur Beschreibung der unterschiedlichen Skalierungsverfahren im Leseverständnis und den mathematischen bzw. naturwissenschaftlichen Kompetenzen vgl. Lankes et al., 2003b.

erfassten Probanden die Schülerfähigkeiten angemessen zu bestimmen, werden diese unter Berücksichtigung des Messfehlers abgeschätzt. Diese *Plausible Values* dienen nicht als bester Schätzer für eine spezifische Personenfähigkeit, sondern werden benötigt, um angemessene Aussagen über die Population oder um Populationsvergleiche vornehmen zu können (Mislevy, 1984; Mislevy, Beaton, Kaplan & Sheehan, 1992; Rost, 1996; Wu, Adams & Wilson, 1998). Auf Länderebene wird wegen der kleineren Stichproben stets auf Basis der fünf *Plausible Values* berichtet.

2.4 Durchführung der Erhebung

Die Durchführung der Erhebung wurde dem *IEA Data Processing Center* (DPC) in Hamburg übertragen, das in Zusammenarbeit mit der Projektleitung an der Universität Hamburg die internationalen Vorgaben und nationalen Besonderheiten umsetzte (Lankes et al., 2003b). Die Verfahren der Datenerhebung und Datenverarbeitung wurden in enger Zusammenarbeit mit den Datenschutzbeauftragten der Länder den gesetzlichen Vorgaben entsprechend gestaltet.

Literatur

- 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.
- Baumert, J., Bos, W. & Lehmann, R. H. (Hrsg.). (2000). *TIMSS/III. Dritte Internationale Mathematik- und Naturwissenschaftsstudie – Mathematische und naturwissenschaftliche Bildung am Ende der Schullaufbahn, Bd. 2: Mathematische und physikalische Kompetenzen am Ende der gymnasialen Oberstufe*. Opladen: Leske + Budrich.
- Baumert, J., Bos, W. & Watermann, R. (2000). Mathematische und naturwissenschaftliche Grundbildung im internationalen Vergleich. In J. Baumert, W. Bos & R. H. Lehmann (Hrsg.), *TIMSS/III: Dritte Internationale Mathematik- und Naturwissenschaftsstudie – Mathematische und naturwissenschaftliche Bildung am Ende der Schullaufbahn, Bd. 1: Mathematische und naturwissenschaftliche Grundbildung am Ende der Pflichtschulzeit* (S. 135–197). Opladen: Leske + Budrich.
- Bos, W., Lankes, E.-M., Prenzel, M., Schwippert, K., Walther, G. & Valtin, R. (Hrsg.). (2003). *Erste Ergebnisse aus IGLU. Schülerleistungen am Ende der vierten Jahrgangsstufe im internationalen Vergleich*. Münster: Waxmann.
- Elley, W. B. (1992). *How in the world do students read? IEA study of reading literacy* (2. Aufl.). The Hague: International Association for the Evaluation of Educational Achievement.

- Elley, W. B. (1994). *The IEA study of reading literacy: achievement and instruction in thirty-two school systems*. Oxford: Pergamon. (International studies in educational achievement, 11).
- Gonzalez, E. J. & Foy, P. (2000). Estimation of sampling variance. In M. O. Martin, K. D. Gregory & S. E. Stemler (Hrsg.), *TIMSS 1999 technical report. IEA's repeat of the third international mathematics and science study at the eighth grade* (S. 203–222). Chestnut Hill, MA: Boston College.
- Johnson, E. G. & Rust, K. F. (1992). Population references and variance estimation for NAEP data. *Journal of Educational Statistics*, 17, 175–190.
- Klieme, E., Baumert, J., Köller, O. & Bos, W. (2000). Mathematische und naturwissenschaftliche Grundbildung: Konzeptuelle Grundlagen und die Erfassung und Skalierung von Kompetenzen. In J. Baumert, W. Bos & R. H. Lehmann (Hrsg.), *TIMSS/III. Dritte Internationale Mathematik- und Naturwissenschaftsstudie – Mathematische und naturwissenschaftliche Bildung am Ende der Schullaufbahn, Bd. 1: Mathematische und naturwissenschaftliche Grundbildung am Ende der Pflichtschulzeit* (S. 85–133). Opladen: Leske + Budrich.
- Lankes, E.-M., Bos, W., Mohr, I., Plaßmeier, N. & Schwippert, K. (2003a). Lehr- und Lernbedingungen in den Teilnehmerländern. In W. Bos, E.-M. Lankes, M. Prenzel, K. Schwippert, G. Walther & R. Valtin (Hrsg.), *Erste Ergebnisse aus IGLU. Schülerleistungen am Ende der vierten Jahrgangsstufe im internationalen Vergleich* (S. 29–67). Münster: Waxmann.
- Lankes, E.-M., Bos, W., Mohr, I., Plaßmeier, N., Schwippert, K., Sibberns, H. & Voss, A. (2003b). Anlage und Durchführung der Internationalen Grundschul-Lese-Untersuchung (IGLU) und ihrer Erweiterung um Mathematik und Naturwissenschaften (IGLU-E). In W. Bos, E.-M. Lankes, M. Prenzel, K. Schwippert, G. Walther & R. Valtin (Hrsg.), *Erste Ergebnisse aus IGLU. Schülerleistungen am Ende der vierten Jahrgangsstufe im internationalen Vergleich* (S. 7–28). Münster: Waxmann.
- Lehmann, R. H., Peek, R., Pieper, I. & Stritzky, R. v. (1995). *Leseverständnis und Lesegewohnheiten deutscher Schüler und Schülerinnen*. Weinheim: Beltz.
- Martin, M. O., Mullis, I. V. S. & Kennedy, A. M. (Hrsg.). (2003). *PIRLS 2001 Technical Report*. Chestnut Hill, MA: Boston College.
- Mislevy, R. J. (1984). Estimating latent distributions. *Psychometrika*, 49 (3), 359–381.
- Mislevy, R. J., Beaton, A. E., Kaplan, B. & Sheehan, K. M. (1992). Estimation population characteristics from sparse matrix of item responses. *Journal of Educational Measurement*, 29 (2), 133–161.
- Rost, J. (1996). *Lehrbuch Testtheorie, Testkonstruktion*. Bern: Huber.
- Wolter, K. M. (1985). *Introduction to variance estimation*. New York: Springer.
- Wu, M. L., Adams, R. J. & Wilson, M. R. (1998). *ACER ConQuest. Generalised item response modelling software*. Melbourne: Acer Press.