

Knut Schwippert, Wilfried Bos & Eva-Maria Lankes

# Lesen Mädchen anders?

Vertiefende Analysen zu Geschlechtsdifferenzen auf Basis der Internationalen Grundschul-Lese-Untersuchung IGLU

### Zusammenfassung

Die Gender bezogene Forschung der letzten Jahre setzte sich insbesondere mit Differenzen in mathematischen und naturwissenschaftlichen Kompetenzen auseinander, während die Unterschiede im Leseverständnis weniger Beachtung fanden. Dabei sind national wie auch international die Befunde von groß angelegten Schulvergleichsstudien konsistent: Mädchen lernen schneller und besser lesen, und auch wenn die Jungen in der Sekundarstufe aufholen, so erreichen sie den Stand der Mädchen auch im Jugendalter noch nicht. Anhand der 2001 durchgeführten Internationalen Grundschul-Lese-Untersuchung (IGLU) wird untersucht, ob der Vorsprung der Mädchen im Leseverständnis schon in der Grundschule angelegt ist. Anhand differenzieller Item Analysen wird der Frage nachgegangen, ob sich auch bei einzelnen Fragen systematische Unterschiede zwischen Mädchen und Jungen nachweisen lassen. Anhand der Analysen kann gezeigt werden, dass geringe Geschlechtsdifferenzen bezüglich des Frageformats (offenes Antwortformat vs. *Multiple Choice*) und den Leseleistungen bei literarischen und Informationstexten bestehen. Bei Betrachtung der in IGLU getesteten Verstehensaspekte ergeben sich keine Differenzen zwischen Jungen und Mädchen. Hingegen ist bezüglich der Aufgabenschwierigkeiten ein Zusammenhang mit geschlechtsspezifischen Lösungshäufigkeiten festzustellen, was ältere Befunde stützt, die gezeigt haben, dass Mäd-

### Summary

Over the last few years research has particularly concerned itself with gender differences between competencies in mathematics and natural sciences, whilst differences in reading comprehension have had little attention. At the same time, national and international evidence from large-scale school comparisons has shown consistently that girls learn to read faster and better. Even when boys catch up at secondary level, they do not reach the girl' standard in their teens. On the basis of the international reading study IGLU carried out in 2001, this contribution will ask whether the girl' head-start can already be observed at primary school level. Using differential item-analysis, the question of whether answers to individual questions show systematic differences will be investigated. The analysis shows only small gender differences regarding question format (open questions vs. multiple choice) and reading performance for literary and informational texts. Also, no differences between boys and girls can be found in the aspects of comprehension tested for in the IGLU-study. However, there is a connection between the level of task difficulty and the frequency of solving tasks by gender, which supports previous evidence that girls read more proficiently. A possible consequence of this study could be the using of reading incentives in class, which are particularly aimed at boys, to encourage their reading. Further studies would have to in-

chen routinierter lesen. Als mögliche Konsequenz dieser Untersuchung könnte die Anregung gegeben werden, im Unterricht vermehrt Leseanreize zu geben, welche Jungen eher ansprechen, um sie so zum vermehrten Lesen zu führen. In späteren Untersuchungen wäre dann zu klären, ob die Jungen so mehr Sicherheit bei routinierten Leseaufgaben erreichen könnten und ggf. zu dem von Mädchen gezeigten Niveau des Leseverständnisses aufschließen könnten.

investigate whether boys achieve an improved confidence in tasks calling for reading proficiency and are therefore able to reach the standards of reading comprehension set by the girls.

Groß angelegte Systemmonitoring-Studien haben in den letzten zehn Jahren in Deutschland mehr Aufmerksamkeit erregt als in den vorherigen Dekaden. Deutschland hat sich seit der ersten internationalen Schulvergleichsstudie der *International Association of Educational Achievement* (IEA) an solchen Studien meist nur mit nicht repräsentativen Stichproben beteiligt. Zum steten Repertoire der Tests gehörten die Leseleistung sowie mathematische und naturwissenschaftliche Kompetenzen, die zum Teil wiederholt getestet wurden (vgl. BOS/SCHWIPPERT 2002). Aufgrund der traditionell geisteswissenschaftlichen Orientierung in den 1970er- und 1980er-Jahren wurden die Ergebnisse jedoch kaum zur Kenntnis genommen. Erst die Ergebnisse der 1991 durchgeführten *Reading Literacy Study* (vgl. LEHMANN u.a. 1995) sind in deutschen Fachkreisen auf breiteres Interesse gestoßen und die Mitte der 90er-Jahre getesteten mathematischen und naturwissenschaftlichen Kenntnisse führten auch in einer breiten Öffentlichkeit zum so genannten TIMSS-Schock. Die Ergebnisse waren aus deutscher Sicht ernüchternd, Deutschland lag nicht im vorderen Leistungsbereich (vgl. BAUMERT u.a. 1997). Und PISA schließlich zeigte auf, dass es auch im neuen Jahrtausend nicht zum Besten mit Schülerleistungen am Ende der Sekundarstufe I stand (vgl. BAUMERT u.a. 2001). Im Leseverständnis lagen die deutschen Fünfzehnjährigen unter dem Durchschnitt und in Mathematik und Naturwissenschaften sah es nicht besser aus. Die knapp zwei Jahre später veröffentlichte Internationale Grundschul-Lese-Untersuchung IGLU zeigte dann jedoch, dass nicht das gesamte deutsche Schulsystem durch Defizite gekennzeichnet war, sondern dass es qualitative Unterschiede zwischen Primar- und Sekundarstufe gab. Nur drei Länder lagen hier signifikant vor Deutschland (vgl. BOS u.a. 2003).

Alle Studien legten die Basis für anschließende vertiefende Analysen. Ziel war es, Unterschiede in den Leistungen zu erkennen und ggf. pädagogische Konzepte zur Verringerung von Differenzen bzw. zur Steigerung von Kenntnissen zu finden. Ein Analyse-schwerpunkt war in allen diesen Studien der Unterschied zwischen Schülerinnen und Schülern. In diesen Untersuchungen zeigte sich, dass die Schülerinnen in mathematischen und naturwissenschaftlichen Kompetenzen Defizite hatten, während die Schüler im Leseverständnis hinter den Schülerinnen zurück blieben. Die Gender bezogene Forschung der letzten Jahre setzte sich insbesondere mit Differenzen in mathematischen und naturwissenschaftlichen Kompetenzen auseinander, während die Unterschiede im Leseverständnis weniger Beachtung fanden (vgl. STANAT/KUNTER 2001). Dabei sind national wie auch international die Befunde von groß angelegten Schulvergleichsstudien konsistent: Mädchen lernen schneller und besser lesen. Wenngleich die Jungen in der Sekundarstufe aufholen (vgl. ELLEY 1994; LEHMANN 1994; WAGEMAKER 1996), so erreichen sie den Stand der Mädchen auch im Jugendalter noch nicht (vgl. OECD 1995). Auch die

beiden jüngsten groß angelegten Leseuntersuchungen, an denen Deutschland sich beteiligt hat, bestätigen diese Befunde. In der Internationalen Grundschul-Lese-Untersuchung (IGLU), in der unter anderem das Leseverständnis von Viertklässlern untersucht wurde, haben in allen 35 teilnehmenden Ländern die Mädchen ein besseres Leseverständnis gezeigt als die Jungen (vgl. MULLIS u.a. 2003). Das gleiche Bild zeigt sich auch in den Befunden des *Programme for International Student Assessment* (PISA). Auch hier schnitten die fünfzehnjährigen Mädchen in allen 32 Ländern besser ab als ihre gleichaltrigen Klassenkameraden (vgl. OECD 2001).

Da die empirische Evidenz für Unterschiede im Leseverständnistest zwischen Mädchen und Jungen nicht so verbreitet ist wie in anderen Kompetenzbereichen, soll am Beispiel der IGLU-Studie untersucht werden, wie der in Deutschland eingesetzte Test zur Klärung dieser Differenzen herangezogen werden kann. Da sich auch das Leseverständnis und die Lesemotivation zwischen Jungen und Mädchen unterscheiden, könnte damit gerechnet werden, dass Leseanreize durch unterschiedliche Textgattungen zu unterschiedlichen Ergebnissen in den anschließenden Tests führen könnten. So haben STANAT und KUNTER (vgl. 2001) darauf hingewiesen, dass der abnehmende Lesevorsprung von neunjährigen Kindern zu vierzehnjährigen Jugendlichen, der im Rahmen der internationalen *Reading Literacy Study* 1991 festgestellt wurde (vgl. LEHMANN 1994), unter anderem auf unterschiedlich umfangreich verwendete Textgattungen, die Jungen und Mädchen unterschiedlich ansprechen, zurückzuführen sein kann: Ein Befund, der auch durch die Untersuchung von RICHTER und BRÜGELMANN (vgl. 1995) bestätigt wird, die einen generellen Nachteil der Jungen im Lesen nicht sehen und eine Betrachtung von Lesekompetenzen differenziert nach Textgattungen anregen.

Aber nicht nur der Leseanreiz, auch die zum Gelesenen gestellten Fragen werden näher betrachtet werden. Untersuchungen der letzten Jahre haben gezeigt, dass Mädchen den Jungen nicht nur im Lesen, sondern auch im Schreiben überlegen sind (vgl. VALTIN u.a. 2003). Wenn nun bei Fragen die Antworten frei zu formulieren sind, könnte dies den Mädchen einen weiteren Vorsprung verschaffen. Als Konsequenz könnten Frageformen offener (*Constructed Response*) oder geschlossener Formen (*Multiple Choice*) zu abweichenden Befunden führen. Jedoch weist RICHTER (vgl. 1996) in ihrer Arbeit auf die Bedeutung von „Jungen- und Mädchenwörtern“ beim Schreiben hin. Sie belegt, dass „Jungenwörter“ von Jungen relativ häufiger richtiger geschrieben wurden als von Mädchen, sodass von einer generellen Überlegenheit der Mädchen nicht gesprochen werden dürfe (vgl. auch: RICHTER/BRÜGELMANN 1994). Da in der IGLU-Studie unterschiedliche Frageformen eingesetzt wurden, können hier spezifische Schwächen und Stärken der beiden Geschlechter analysiert werden.

LEHMANN (vgl. 1994) hat in der vor rund zehn Jahren durchgeführten Leseuntersuchung festgestellt, dass das bessere Abschneiden der Mädchen nicht nur auf ein besseres Beantworten aller Fragen im Gesamttest zurückzuführen ist, sondern dass sie insbesondere bei Routineaufgaben sicherer antworten. Ob sich dieser Befund replizieren lässt wird ebenfalls untersucht.

Im folgenden Beitrag wird somit der IGLU-Lesetest daraufhin untersucht, inwieweit mit ihm auf Testebene Unterschiede im Leseverständnis von Jungen und Mädchen im Grundschulalter erklärt werden können. Anhand differenzieller Item Analysen (*Differential Item Functioning* DIF) wird anhand der deutschen Stichprobe vertiefend analysiert, ob Fragenformate, Textgattungen und unterschiedliche Anforderungsniveaus der Tests geschlechtsspezifische Unterschiede aufweisen. Ergänzend werden Beispielaufgaben vor-

gestellt, bei denen signifikante Unterschiede in den Lösungshäufigkeiten von Jungen und Mädchen bestehen.

## 1 Die IGLU-Studie

In der Internationalen Grundschul-Lese-Untersuchung (IGLU) wurden die Fähigkeiten von Kindern im Leseverständnis am Ende der vierten Jahrgangsstufe im Rahmen der von der *International Association for the Evaluation of Educational Achievement* (IEA) initiierten Untersuchung überprüft. International ist diese Studie unter dem Namen *Progress in International Reading Literacy Study* (PIRLS 2001) bekannt (vgl. CAMPBELL u.a. 2001). Die Zielgruppe dieser Untersuchung war der höhere der beiden Jahrgänge, welchen die meisten Neunjährigen besuchen. In den meisten Ländern und auch in Deutschland entspricht diese Definition der vierten Jahrgangsstufe. Im Frühsommer 2001 (nördliche Hemisphäre) bzw. im Herbst desselben Jahres (südliche Hemisphäre) wurden rund 146.490 Grundschülerinnen und -schüler aus 35 Staaten getestet (vgl. BOS u.a. 2003).

Die Entwicklung der Tests in IGLU orientiert sich an einem Rahmenkonzept, das sich auf die internationale Diskussion über Grundbildung oder grundlegende Lese- und Schreibkompetenz bezieht. In pragmatischer Absicht werden grundlegende Kompetenzen definiert, die über den unterrichtlichen Rahmen hinaus für die Lebensbewältigung in konkreten Anwendungssituationen nützlich sind. Dieses Konzept kommt dem sehr nahe, was in den letzten Jahren in der grundschulpädagogischen Diskussion um den Bildungsauftrag von Grundschule erörtert wird.

Das Modell der Lesekompetenz von IGLU basiert auf dem kognitionspsychologischen Forschungsansatz von van DIJK und KINTSCH (vgl. 1983) und dem psychometrischen Ansatz von KIRSCH und MOSENTHAL (vgl. 1989–1991). Textverständnis kann hiernach als ein informationsverarbeitender Prozess verstanden werden, in dem der Leser die im Text enthaltene Information unter Einsatz verschiedener Lesestrategien aktiv mit seinem Vor- und Weltwissen verbindet (vgl. auch GROEBEN/HURRELMANN 2002) und eine mentale Repräsentation des Gelesenen konstruiert.

Bei dem in der IGLU-Studie eingesetzten Leseverständnistest wurden den Kindern kurze kontinuierliche Texte vorgelegt, die je nach Gestaltung der Texte und der Anreicherung mit graphischen Elementen zwischen 2 bis 4 Seiten umfassen. Die geschätzte Lesedauer war je nach Lesegeschwindigkeit der Kinder auf 15 bis 20 Minuten angelegt. Inklusive der anschließenden Bearbeitungszeit der 11 bis 14 Testfragen wurde den Kindern jeweils 40 Minuten Zeit gegeben. Mit einer Unterbrechung von 15 Minuten wurde den Kindern zwei dieser Tests in einem Testheft vorgelegt. Die insgesamt 10 Testhefte wurden aus 8 verschiedenen Texten zusammengestellt. Vier der Texte gehörten zur Textgattung der literarischen Texte, die vier anderen wurden als Informationstexte konzipiert. Dieses im internationalen Lesetest verwendete Testdesign wird als Test mit rotierten Aufgabenblöcken bezeichnet.

Um anhand der Testinstrumente zuverlässige (valide) Schätzer für die zu erfassende Lesekompetenz zu erhalten, müssen Tests über eine ausreichende Länge und somit über genügend Leistungsfragen verfügen. Die an der Entwicklung der Studie beteiligten Leseexperten gehen von einer Testdauer von ca. 2 Zeitstunden je Leseschwerpunkt aus (vgl. CAMPBELL u.a. 2001). Da aber gerade Grundschulkindern solch theoretisch geforderte

Testdauer forschungspraktisch nicht zugemutet werden kann, bekommt jeder Schüler nur einen Teil der im Gesamttest enthaltenen Aufgaben – nämlich die zuvor beschriebenen Testhefte mit jeweils nur zwei Testblöcken. Aufgrund des angewendeten Skalierungsverfahrens ist sichergestellt, dass alle Testblöcke die gleiche Kompetenz erfassen, zusammengekommen die geforderte Validität besitzen und die Kinder nicht über Gebühr belasten. Dieses als *Multi Matrix Design* bezeichnete Forschungsdesign ist seit einigen Jahren in der empirischen Bildungsforschung etabliert.

## 2 Die Skalierung des IGLU-Tests

Im Rahmen des angewendeten probabilistischen Testmodells werden Schülerfähigkeiten und Fragenschwierigkeiten beim so genannten Prozess der „Skalierung“ simultan bestimmt. Ein Vorteil des probabilistischen Modells besteht darin, dass sich Schülerfähigkeiten mit Aufgabenschwierigkeiten auf der gleichen Skala abbilden lassen. Personenfähigkeiten und Aufgabenschwierigkeiten sind somit direkt aufeinander bezogen (vgl. BOS u.a. 2003).

Das im Folgenden zur Identifizierung von geschlechtsspezifischen Besonderheiten herangezogene probabilistische Testmodell beruht auf den Arbeiten von Georg RASCH (vgl. 1980). Wenn von ‚dem Rasch-Modell‘ gesprochen wird, ist implizit immer das einparametrische Rasch-Modell gemeint. Die Bezeichnung einparametrisch rührt daher, dass die Frageneigenschaften anhand eines einzelnen Kennwertes – nämlich der Fragenschwierigkeit – charakterisiert werden. Wenn Analysen nach dem Rasch-Modell geplant sind, sind verschiedene Voraussetzungen zu prüfen. Das Rasch-Modell hat bei Gültigkeit unter anderem folgende Eigenschaften:

Schwere Fragen werden durchgängig von Schülern aller Fähigkeitsgruppen systematisch seltener richtig beantwortet als leichtere (parallele [überschneidungsfreie] Itemfunktionen).

Ob Fragen richtig oder falsch beantwortet werden hängt ausschließlich von der gemessenen Personenfähigkeit ab (hier z.B. des Leseverständnisses) (lokale stochastische Unabhängigkeit).

Auch anhand von Teiltests können die Personenfähigkeiten bestimmt und mit dem Gesamttest verglichen werden. Hierbei ist es ohne Bedeutung, welche Auswahl der Testfragen gestellt wird (spezifische Objektivität).

Der IGLU-Test besteht zum einen aus Fragen im *Multiple-Choice-Format*, zum anderen aus Fragen mit offenem Antwortformat. Innerhalb der Aufgaben mit offenem Antwortformat gibt es Fragen, bei denen die Probanden auch Teilantworten geben können und die somit ggf. als nur teilweise richtig beantwortet eingeschätzt werden können (*Partial Credit*). Bei den Analysen können diese Teilantworten jeweils als unterschiedlich schwierige Einzelaufgaben interpretiert werden, die allerdings voneinander abhängig sind.

Modelle sind einfachere Abbilder der Wirklichkeit. Bei der Nutzung von Modellen zur Erklärung beobachteter Phänomene ist zu prüfen, ob die vereinfachte Abbildung der Wirklichkeit gerecht wird oder ob einige Besonderheiten, die im einfachen Modell nicht berücksichtigt wurden, im Rahmen von komplexeren Modellen aufgenommen werden müssen. Ob die Nutzung komplexerer Modelle sinnvoll oder notwendig ist, kann im

Rahmen von probabilistischen Test-Modellen geprüft werden. Bei *Differential Item Functioning* (DIF) Analysen wird die Eigenschaft der lokalen stochastischen Unabhängigkeit einer besonderen Prüfung unterzogen (vgl. ANGOFF 1993; DORANS/HOLLAND 1993). Hierbei wird geprüft, ob die Lösung der Aufgaben tatsächlich nur – wie in einer der Modellvoraussetzungen formuliert – von der Personenfähigkeit abhängig ist oder ob andere beobachtbare Personeneigenschaften mit zu unterschiedlichen Lösungshäufigkeiten führen. Die DIF-Analyse soll im Folgenden genutzt werden, um systematisch zu untersuchen, ob einzelne Fragen eher von Jungen oder eher von Mädchen richtig gelöst werden. Die Überprüfung der Testfragen auf differenzielle Funktionen wird nur dann vorgenommen, wenn die Überprüfung der allgemein gesetzten Voraussetzungen positiv ausgefallen ist, also das Rasch-Modell gilt. Von der fragenspezifischen Betrachtung ist das Phänomen ggf. vorhandener genereller Fähigkeitsvorsprünge für eine der Geschlechtergruppen abzulösen.

### 3 DIF-Analysen

Um die DIF-Analysen durchführen zu können, wurden die deutschen Daten der IGLU-Studie neu skaliert. Die Analysen wurden mit dem Programm ConQuest vorgenommen (vgl. WU/ADAMS/WILSON 1998). Zur Vorbereitung der Analysen werden alle Fragen dichotomisiert als richtig oder falsch beantwortet. Bei offenen Fragen wird die Frage schon beim Erreichen von Teilpunkten als richtig beantwortet gewertet. Für dieses Vorgehen spricht, dass die im Folgenden betrachteten Frageeigenschaften sich jeweils auf den Fragestamm und nicht auf die Teillösungen der Aufgaben beziehen. Da die *Partial Credit Items* somit aus analysepraktischen Erwägungen ebenso wie die *Multiple Choice Items* dichotomisiert wurden, war mit einer leichten Veränderung der Itemkennwerte zu rechnen. Die Korrelation der Itemschwierigkeiten des Modells, in dem die offenen Fragen als *Partial Credit Items* mehrstufig aufgenommen wurden, und dem Modell, in dem diese dichotomisiert Berücksichtigung fanden, ist mit  $r = .88$  ( $n = 98$  /  $p < 0,001$ ) so hoch, dass in beiden Modellen die Aufgabenschwierigkeiten substantiell dieselben sind. – Nur bei einigen Fragen hat es leichte (erwartete) Veränderungen gegeben.

Eine zusätzliche Prüfung des komplexeren Testmodells, in dem die für *Partial Credit* Aufgaben ermittelten Schwellenparameter zum einen geschlechtsneutral und zum anderen geschlechtsspezifisch modelliert wurden, ergab keine signifikanten Unterschiede in der Schwellenstruktur zwischen den Geschlechtern.

Mit dem ConQuest Programm werden folgende Modell-Parameter bestimmt: (1) die generelle Item-Schwierigkeit, (2) gruppenspezifische Unterschiede zwischen Jungen und Mädchen und schließlich (3) fragenspezifische Unterschiede (DIF) zwischen Jungen und Mädchen.

Die Modellprüfung wird auf Basis des deutschen Teildatensatzes des IGLU-Lesetests durchgeführt. Hierzu stehen 7.633 Beobachtungen von Jungen (50,2%) und Mädchen (49,8%) zur Verfügung. Die Unterscheidung des Geschlechts erfolgte nach Selbstausskunft der Kinder, wo diese nicht eindeutig im Fragebogen angegeben war, wurde die Information aus Klassenlisten ergänzt. Altersgemäß werden sich somit die Angaben zum Geschlecht auf primäre Geschlechtsmerkmale beziehen. Eine differenzierte Erfragung des „Gender“ anhand verschiedener Indikatorvariablen ist aus forschungspraktischen Grün-

den nicht vorgenommen worden, da bei 10-Jährigen die Gender-Identität weitestgehend noch an eigenen primären Geschlechtsmerkmalen orientiert ist.

Nach dem Analyselauf wurden die Aufgabenschwierigkeiten, die DIF-Parameter und Informationen zum Fragenformat, zur Textgattung und die durch die Fragen angesprochenen Kompetenzebenen (die *ex ante* definiert wurden) auf Frageebene zusammengespielt. Für die Analysen stehen für 98 Einzelfragen die beschriebenen Informationen zur Verfügung.

## 4 Befunde

Die ersten Ergebnisse der IGLU-Studie wurden im Frühjahr 2003 vorgestellt (vgl. BOS u.a. 2003). Im internationalen Vergleich nehmen die deutschen Grundschüler einen Rangplatz im ersten Leistungsdrittel der Länder ein. Nur in den drei Ländern Schweden, Niederlande und England, weisen die Kinder ein signifikant besseres Leseverständnis auf. Im Vergleich mit sieben ebenfalls an der IGLU-Studie teilnehmenden EU-Ländern, die auch an der PISA-Untersuchung teilgenommen haben, nehmen die Grundschüler einen mittleren Platz ein. Wird die Vergleichsgruppe um weitere 14 OECD-Länder ergänzt, liegen die Leistungen in Deutschland über dem Durchschnittswert dieser Vergleichsgruppe. Der Unterschied im Leseverständnis beim Lesen von Informationstexten und literarischen Texten ist in Deutschland nur sehr schwach ausgeprägt. Nur in fünf Ländern ist die numerische Differenz geringer (vgl. MULLIS u.a. 2003). Der Leistungsvorsprung der Mädchen im Leseverständnis ist in allen teilnehmenden Ländern zu beobachten. In Deutschland ist die Differenz in Relation zu den anderen Ländern jedoch eher schwach ausgeprägt. Differenziert man den Vorsprung der Mädchen nach Textgattungen, ist zu beobachten, dass dieser bei literarischen Texten etwas ausgeprägter ist als bei den Informationstexten.

Im ersten Schritt der vertiefenden Analysen kann bestätigt werden, dass das zugrunde gelegte Rasch-Modell, welches die Personenfähigkeiten und Aufgabenschwierigkeiten auf einer gemeinsamen Metrik verankert, die vorliegenden Daten angemessen beschreibt. Alle 98 in das Modell aufgenommenen dichotomisierten Fragen weisen eine akzeptable FIT-Statistik auf. Nur eine Frage erscheint für die Gesamtgruppe tendenziell etwas zu leicht. Ansonsten ist in keiner der Gruppen auf Fragenebene ein substanzieller Boden- bzw. Deckeneffekt zu beobachten. Das bedeutet, dass der Lesetest auch noch besonders schwache und besonders gute Leser zuverlässig unterscheidet. Erwartungsgemäß zeigt sich auch bei dieser Analyse, dass den Mädchen die Aufgaben im Gesamttest signifikant leichter gefallen sind als den Jungen. Dies bedeutet, dass für Mädchen im Leseverständnis ein systematischer Vorsprung festgestellt werden konnte. Dieser Befund ist jedoch noch kein Hinweis auf ein mögliches DIF-Phänomen. Es besagt lediglich, dass die Mädchen die Aufgaben im Lesetest systematisch leichter lösen konnten als die Jungen. Ob dieser Vorsprung aufgrund eines generellen Vorteils, höherem Interesse, stärkerer Motivation oder besserem Leseverständnis zustande kommt, kann anhand dieser Analysen nicht geklärt werden. Verschiedene dieser Phänomene wirken sicherlich zusammen und müssen anhand von multivariaten Verfahren analysiert werden. Dies soll hier jedoch nicht erfolgen und bleibt somit vertiefenden Analysen, wie sie z.B. RICHTER (vgl. 1995, 1996) vorgestellt hat, vorbehalten.

Im nächsten Schritt wird überprüft, ob einzelne Fragen im Kontext des gesamten Lesetests signifikant häufiger von einer der DIF-Gruppen (Jungen vs. Mädchen) gelöst wer-

den. Hierzu werden für jede einzelne Frage die tatsächliche und die erwartete Lösungshäufigkeit unter Berücksichtigung der Fragenschwierigkeit für die betrachteten Gruppen ermittelt. Die ermittelten Differenzen werden schließlich noch Signifikanztests unterzogen, um festzustellen, ob sich die beobachteten numerischen Differenzen zwischen den Gruppen auch zufallskritisch abgesichert unterscheiden lassen. Erst nach dieser Überprüfung ist die Interpretation der beobachteten Unterschiede sinnvoll.

Was bedeutet dies praktisch? – Da, wie zuvor festgestellt wurde, die Mädchen tendenziell etwas besser lesen als die Jungen, könnte damit gerechnet werden, dass sie bei jeder gestellten Aufgabe besser abschneiden als die Jungen. Um dieses Phänomen nicht bei der Betrachtung einzelner Fragen immer wieder neu zu interpretieren, wird der generelle Vorteil der Mädchen bei der Betrachtung der einzelnen Fragen per se ausgeklammert. Das heißt, nur wenn bei einer Frage der Vorteil, den Mädchen haben, größer ist als ihr erwarteter Vorteil, wird diese Frage als für Mädchen spezifisch leichter interpretiert. Ist bei einer anderen Frage der Vorteil der Mädchen geringer als ihr erwarteter, so wird diese Frage als relativ leichter für die Jungen gewertet. Neben der Feststellung dieser relativen Vor- bzw. Nachteile – also der differentiellen Item Funktionen (DIF) – wird noch geprüft, ob die beobachteten Abweichungen zufällig oder als statistisch signifikant zu interpretieren sind. Dieses Vorgehen greift die von RICHTER (VGL. 1996) beschriebene Problematik bei der Testkonstruktion auf, da analytisch zwischen Unterschieden zwischen den Geschlechtern bedingt durch die Textauswahl und die Formulierung der einzelnen Fragen unterschieden wird.

In dem untersuchten IGLU-Test wurden für rund die Hälfte der Fragen signifikante Unterschiede zwischen den Gruppen festgestellt. Dies scheint auf den ersten Blick viel zu sein. Es ist aber zu berücksichtigen, dass mit 7.633 Observationen eine sehr große Gesamtstichprobe zur Verfügung steht, und da die Testhefte im rotierten Design vorgelegt wurden, lagen jeweils rund 1.850 Antworten von Schülerinnen und Schülern pro Frage vor. Bei diesen Stichprobengrößen werden auch kleinste Unterschiede als statistisch signifikant identifiziert, da bei den Signifikanztests i.d.R. eine Zufallsauswahl von Probanden vorausgesetzt wird; eine Voraussetzung, die in dem IGLU-Design aber *nicht* vorliegt. Hier wurden die Kinder im so genannten Klumpenstichprobenverfahren ausgewählt (es wurden intakte Klasse getestet), was zu einer Verringerung der effektiven Stichprobengröße führt und somit die gängigen Signifikanztests zu liberal ausfallen lässt (vgl. FOY/JONCAS 1999; KISH 1995).

In dem hier gewählten Analysemodell weisen DIF-Werte, die kleiner als 0 sind, auf einen Vorteil der Mädchen hin. Das bedeutet, unabhängig von der generellen höheren Lösungshäufigkeit der Mädchen, sind einige der Fragen für sie nochmals besonders leicht. Entsprechende Fragen (mit positivem DIF-Parameter) werden auch für die Jungen identifiziert.

Die wirkliche Ausprägung des DIF lässt sich besser abschätzen, wenn die für die Fragen ermittelten DIF-Parameter in Relation zu den Fragenschwierigkeiten betrachtet werden. Die Fragenschwierigkeiten liegen in der original *logit*-Metrik zwischen -2,70 (leichte Frage) und + 2,66 (schwierige Frage) mit einer Streuung von 1,2 (Standardabweichung) um die mittlere Schwierigkeit 0. Der größte beobachtete DIF-Parameter liegt mit 0,41 bei ca. einem Drittel dieser Standardabweichung zugunsten der Jungen. Diese Frage lösen Jungen zu 38,3% und Mädchen zu 28,2% richtig. Bei einem mittleren (signifikanten) DIF-Parameter von ca. 0,18 *logits* beträgt der Unterschied in den Lösungshäufigkeiten zwischen den Geschlechtern etwas weniger als 4 Prozentpunkte. Dieser geringe Unter-

schied ließe sich in einer normalen Klasse mit 24 Kindern (bei jeweils 12 Jungen und Mädchen) praktisch nicht beobachten, da hier nur ein Junge mehr oder weniger, der die Fragen richtig löst, schon 8 Prozentpunkte Differenz ausmachen würde. Dies deutet also darauf hin, dass bei den als signifikant identifizierten DIF-Werten eine mittlere Ausprägung in einer realen Schulklasse praktisch keinen Unterschied ausmachen würde, wohl aber bei Betrachtungen der untersuchten Population berücksichtigt werden muss. Je nach Ausprägung der beobachteten Effekte und deren Zusammenhang mit zusätzlich erfassten Informationen sind diese schließlich auch aus didaktischer Sicht als relevant zu betrachten und zu diskutieren.

Wie einzelne Aufgaben aussehen, die im IGLU-Leseverständnistest einen geschlechtsbezogenen Vorteil für Jungen bzw. Mädchen aufweisen wird im folgenden Abschnitt vorgestellt.

#### 4.1 Beispiele

Die ausgewählten Beispiele für Items, für die nach dem zuvor vorgestellten Verfahren eine differenzielle Itemfunktion (DIF) für Jungen und Mädchen festgestellt werden konnte, sind aus den IGLU-Tests entnommen, die zur Veröffentlichung von der internationalen Studienleitung der PIRLS-Untersuchung freigegeben wurden (die Hälfte der Texte und Fragen wird geheim gehalten, um die Möglichkeit einer Messwiederholung und somit einer Verankerung der Studien im zweiten PIRLS-Zyklus 2006 zu gewährleisten). Bei den ausgewählten Texten handelt es sich um einen literarischen Text (*Der Hase kündigt das Erdbeben an*) und um einen Informationstext (*Die Nacht der Papageientaucher*). Den Beispielaufgaben sind die jeweils relevanten Textpassagen gegenübergestellt. Die vollständigen Textpassagen werden hier aus Platzgründen nicht wiedergegeben und können dem IGLU-Bericht (BOS u.a. 2003, S. 86 und S. 87) entnommen werden.

Im ersten Beispiel liegt ein DIF (-0,24 *logits*) zugunsten der Mädchen vor. Die Lösungshäufigkeit ist bei den Mädchen 91,9 und bei den Jungen 85,9%. Bei dieser Frage sollte eine im Text explizit angegebene Information identifiziert werden.

##### Beispiel 1:

Textausschnitt: Hase

„... Seine größte Sorge war, dass es eines Tages ein Erdbeben geben könnte. ...“

Frage: H02M (L1) M

Was war die größte Sorge des Hasen?

- (A) ein Löwe
- (B) ein lautes Krachen
- (C) ein Erdbeben
- (D) eine Frucht, die herunterfiel

Auch im zweiten Beispiel handelt es sich um eine Aufgabe, bei der die Mädchen einen relativen Vorsprung haben (DIF = -0,13). Diese Aufgabe lösten die Jungen zu 88,2 und die Mädchen zu 92,6 Prozent richtig. Bei dieser Aufgabe bestand der Verstehensaspekt darin, eine einfache Schlussfolgerung zu ziehen.

*Beispiel 2:*

Textausschnitt: Hase

„ ... Er nahm sie in den Mund, kletterte auf einen Felsen und warf sie wieder auf den Boden. RUMS! Der Hase sprang in die Luft. ‚Ein Erdbeben! Schnell – renn weg – gerade ist es wieder passiert!‘ Aber da merkte er plötzlich, dass der Löwe laut lachte. Und dann sah er die Frucht, die bis zu seinen Füßen gerollt war.“

Frage: H05M (L2) M

Warum ließ der Löwe die Frucht auf den Boden fallen?

- (A) Er wollte den Hasen in die Flucht schlagen.
- (B) Er wollte dem Hasen helfen, an die Frucht zu kommen.
- (C) Er wollte dem Hasen zeigen, was eigentlich passiert war.
- (D) Er wollte den Hasen zum Lachen bringen.

Der in den Beispielen 1 und 2 belegte Vorteil der Mädchen bei den Aufgaben kann auf verschiedene mögliche Ursachen zurückgeführt werden. Zum einen handelt es sich bei dem Hasentext um einen literarischen Text, der dem Genre nach eher Mädchen als Jungen interessieren dürfte. Darüber hinaus sind in den beiden Beispielen relativ einfache Aufgaben aufgegriffen worden, bei deren Beantwortung routiniertes und sorgfältiges Lesen vorausgesetzt wird. Das generelle Anforderungsniveau ist bei beiden eher gering.

Beispiel 3 zeigt eine Frage, bei der die Jungen einen relativen Vorteil aufweisen (DIF = 0,22). Der Anteil der richtigen Antworten ist bei den Jungen 91,9 und bei den Mädchen 92,8%. An dieser Stelle wird nochmals die analytische Trennung des generellen Geschlechtsunterschieds im gesamten Lesetest und der fragenspezifischen Identifizierung von Unterschieden bei Jungen und Mädchen deutlich. Obwohl die absolute Lösungshäufigkeit bei den Mädchen höher ist als bei den Jungen, ist diese Frage für Jungen als relativ „leichter“ identifiziert worden, als für sie generell erwartet wurde. Wie im Beispiel 1 war auch hier eine explizit im Text gegebene Information zu finden.

*Beispiel 3:*

Textausschnitt: Papageientaucher

„ ... Nun kehren sie zurück zu Hellas Insel und zu den benachbarten unbewohnten Inseln, um Eier zu legen und ihre Küken aufzuziehen. ...“

Frage: N03M (L1) J

Weshalb kommen die Papageientaucher zur Insel?

- (A) Um gerettet zu werden.
- (B) Um nach Nahrung zu suchen.
- (C) Um Eier zu legen.
- (D) Um fliegen zu lernen.

Beispiel 4 schließlich beschreibt eine Frage, bei der nicht nur ein relativer, sondern auch ein absoluter Vorteil bei den Jungen beobachtet wurde. Hier liegt im Vergleich zu Beispiel 3 auch die absolute Lösungshäufigkeit der Jungen mit 38,3% um ca. 10 Prozentpunkte über der der Mädchen (28,2%). Bei diesem Beispiel handelt es sich um eine Frage, bei der Schlussfolgerungen gezogen werden sollten und das Gelesene interpretiert werden musste.

*Beispiel 4:*

Textausschnitt: Die Nacht der Papageientaucher

„ ... Von den hohen Klippen aus flattern sie los, aber sie kommen noch nicht weit. Die meisten plumpsen unversehrt ins Meer. Aber einige lassen sich vom Lichtschein des Dorfes verwirren – vielleicht halten sie ihn für Mondlicht, das sich auf dem Wasser spiegelt. Jede Nacht machen Hunderte von jungen Papageientauchern eine Bruchlandung im Dorf. ...“

Frage: N06M (L3) J

Was könnten die Leute aus dem Dorf tun, um zu verhindern, dass die Jungvögel versehentlich dort landen?

- (A) Das Licht ausmachen.
- (B) Pappkartons bereithalten.
- (C) Katzen und Hunde einsperren.
- (D) Mit ihren Taschenlampen in den Himmel leuchten.

Eine mögliche Ursache, warum im Text zum Papageientaucher die Jungen unter anderem diese beiden Fragen (relativ) einfacher zu lösen scheinen, könnte darin liegen, dass dieser Text vornehmlich informelle Elemente enthält und eine reale Begebenheit beschreibt, in

die auch die Jungen sich gut hineinversetzen können. Im letzten Beispiel ist darüber hinaus aus den gegebenen Informationen eine praktische Problemlösung zu finden. Hier kann die technische Orientierung der Frage als möglicher Hinweis auf die erhöhte Lösungswahrscheinlichkeit der Jungen vermutet werden.

#### 4.2 Die Frage des Frageformats

Wenn Kompetenztests wie in IGLU entwickelt werden, stellt sich immer auch die Frage, wie diese zu erfassen sind. Schließlich soll mit den vorgelegten Fragen hier das Leseverständnis möglichst valide erfasst werden. In internationalen Untersuchungen haben sich als Standard zwei Frageformen etabliert. Es sind zum einen Fragen mit Mehrfach-Wahlantworten (*Multiple Choice*) oder Fragen, die ein offenes Antwortformat (*Constructed Response*) haben. Da die Beantwortung der offenen Fragen potentiell durch die Schreibkompetenzen konfundiert sein kann, werden auch *Multiple-Choice*-Aufgaben vorgelegt. Kritiker dieser Frageform weisen in Deutschland darauf hin, dass die Kinder hier diese Form der Aufgabengestaltung nicht gewohnt seien. Da jede der Frageformen spezifische Vor- und Nachteile aufweist, hat man sich entschieden, die Tests aus einer Kombination beider Frageformate zu bilden (vgl. ELLEY/MANGUBHAI 1992). Die Aufgabenschwierigkeiten sind bei der gewählten Kodierung der *Constructed-Response*-Fragen (dichotomisiert auf richtig bzw. falsch) unabhängig vom Fragenformat. Da bei den Jungen wie beim Leseverständnis auch bei den schriftlichen Leistungen eine geringere Kompetenz nachgewiesen werden konnte (vgl. VALTIN u.a. 2003), könnte vermutet werden, dass Jungen gerade bei den Fragen, bei denen Antworten aufzuschreiben sind, systematisch benachteiligt sein könnten. Zur Klärung dieser Vermutung wird anhand von Varianzanalysen geprüft, ob die beobachteten geschlechtsspezifischen DIF-Parameter in einem Zusammenhang mit dem Fragenformat stehen. Anhand einer einfaktoriellen Varianzanalyse konnte ein tendenzieller Zusammenhang festgestellt werden ( $F = 3,91 / df = 1 / p = 0,051$ ). Bei Mittelwertvergleichen zeigt sich, dass Mädchen einen kleinen Vorteil bei der Bearbeitung von *Multiple-Choice*-Fragen haben. Der Vorteil ist mit 0,03 *logits* jedoch ausgesprochen klein.

#### 4.3 Die Frage der Textgattung

Neben dem Frageformat wurde vermutet, dass auch die Textgattungen, welche den Kindern vorgelegt wurden, Jungen und Mädchen unterschiedlich ansprechen würden. Bei getrennten Analysen von Texten zur Informationsermittlung und Texten literarischen Inhalts, die beide im gleichen Umfang eingesetzt wurden, hat sich im internationalen Vergleich gezeigt, dass die Mädchen einen größeren Vorsprung bei literarischen als bei Informationstexten vor den Jungen aufweisen. In 23 der 35 teilnehmenden Ländern ist dieser Vorsprung bei literarischen höher als bei Informationstexten, in 5 Ländern besteht kein Unterschied und in 7 Ländern ist der Vorsprung zugunsten der Informationstexte ausgeprägt (vgl. PIRLS 2001). Der internationale Durchschnitt dieser Differenzen ist ebenfalls zugunsten der Mädchen bei den literarischen Texten größer als bei den Informationstexten. Dieser Zusammenhang konnte für die deutsche Teilstichprobe ebenfalls wieder anhand einer einfaktoriellen Varianzanalyse belegt werden. Hier zeigt sich ein si-

gnifikanter Haupteffekt zwischen den 98 DIF-Parametern und den beiden Textgattungen ( $F = 11,28 / df = 1 / p = 0,001$ ). Bei Mittelwertvergleichen zeigt sich, dass die Mädchen einen kleinen Vorteil bei der Bearbeitung von Aufgaben zu literarischen Texten haben. Der Vorteil ist mit 0,04 *logits* jedoch ebenfalls sehr klein.

#### 4.4 Die Frage der Verstehensaspekte

Als Zwischenergebnis lässt sich somit festhalten, dass Mädchen eher bei Fragen im *Multiple-Choice*-Format profitieren und wenn ihnen Texte zur Bearbeitung vorgelegt werden, die einen literarischen Inhalt haben. Da Tests, wie sie bei IGLU eingesetzt werden, ein weites Spektrum an Lesern testen sollen, müssen Aufgaben unterschiedlicher Verstehensaspekte und Schwierigkeitsstufen in den Test aufgenommen werden. In IGLU werden vier Aspekte unterschieden (BOS u.a. 2003, S. 79):

- 1) Erkennen und wiedergeben explizit angegebener Informationen.
- 2) Einfache Schlussfolgerungen ziehen.
- 3) Komplexe Schlussfolgerungen ziehen und begründen; Interpretieren des Gelesenen.
- 4) Prüfen und bewerten von Inhalt und Sprache.

Für jeden Verstehensaspekt wurden jeweils unterschiedlich schwierige Aufgaben entwickelt. Trotz ausgewogenem Design unterscheiden sich die mittleren Aufgabenschwierigkeiten in diesen vier Verstehensaspekten hoch signifikant. Die Fragen unter 1) sind die am häufigsten und unter 4) die am seltensten richtig beantworteten Fragen ( $F = 9,72 / df = 3 / p < 0,001$ ). Um zu prüfen, ob es einen geschlechtsspezifischen Unterschied bei der Bearbeitung der Fragen zu den verschiedenen Verstehensaspekten gibt, wird anhand einer einfaktoriellen Varianzanalyse geprüft, ob die entsprechenden DIF-Parameter in einem Zusammenhang mit den vier Gruppen von Verstehensaspekten stehen. Dieser Zusammenhang konnte nicht festgestellt werden ( $F = 0,01 / df = 3 / p = 0,809$ ). Somit haben weder Jungen noch Mädchen spezifische Stärken bei der Bewältigung von Aufgaben zu verschiedenen Verstehensaspekten im Leseverständnistest.

#### 4.5 Die Frage der Aufgabenschwierigkeiten

Die Aufteilung der Fragen in vier verschiedene Gruppen erscheint bei der Betrachtung der Vielschichtigkeit der Aufgaben inhaltlich gerechtfertigt, aber auch als relativ grob. Da sich die Aufgabenschwierigkeiten wie zuvor beschrieben in den Gruppen systematisch unterscheiden und eine Zusammenfassung von Fragen zu Fragengruppen mit einem gewissen Informationsverlust einher geht, soll zum Abschluss der Zusammenhang zwischen den DIF-Parametern und den Aufgabenschwierigkeiten auf Frageebene vorgestellt werden. Als Zusammenhangsmaß wird auf die Korrelation dieser beiden Merkmale zurückgegriffen. Bei den Analysen zeigt sich, dass beide Maße hoch signifikant zusammenhängen ( $r = .29 / n = 98 / p = 0,004$ ). Aufgrund der gewählten Kodierung ist ein positiver DIF-Parameter als relativ schwieriger für Mädchen zu interpretieren (der DIF-Parameter ist in *logits* angegeben und additiv zu den Itemschwierigkeiten für die Mädchen hinzuzurechnen, positive DIFs weisen somit auf eine zusätzliche Schwierigkeit hin). Die positive Korrelation ist schließlich so zu interpretieren, dass Mädchen gegenüber Jungen mit zu-

nehmender Aufgabenschwierigkeit zunehmend seltener richtig lösen. Mit anderen Worten, die Mädchen lösen in dem Test (wenn ihr genereller Vorsprung herauspartialisiert wird) die einfacheren Aufgaben systematisch sicherer als die Jungen, während die Jungen bei den schwierigeren Aufgaben bei der Lösung relativ besser abschneiden. Dieser Befund deckt sich mit dem Befund, den LEHMANN (vgl. 1994) bei Analysen am deutschen Datensatz der *IEA-Reading Literacy Study* nachgewiesen hat. Er konnte feststellen, dass die Mädchen vor allem weitgehend routinisierte Prozesse schneller und genauer bearbeiten.

Wie einzelne Aufgaben aussehen, die im IGLU-Leseverständnistest einen geschlechtsbezogenen Vorteil für Jungen bzw. Mädchen aufweisen, wird im folgenden Abschnitt vorgestellt.

## 5 Zusammenfassung und Diskussion

In groß angelegten Schulvergleichsstudien hat sich die Skalierung von Schülerantworten nach dem probabilistischen Testmodell etabliert. Das einparametrische Rasch-Modell macht restriktive Voraussetzungen, die anhand von Teststatistiken überprüft werden müssen. Die Überprüfung, ob das Rasch-Modell zur Skalierung des deutschen Datensatzes der Internationalen Grundschul-Lese-Untersuchung herangezogen werden darf, konnte prinzipiell bestätigt werden. Das bedeutet, dass sowohl Personenfähigkeiten wie auch Aufgabenschwierigkeiten auf einer gemeinsamen Metrik abgebildet werden können und vergleichende Analysen zwischen den Probanden zulässig sind. Da – wie in älteren Leseverständnissuntersuchungen – auch in IGLU ein Vorsprung der Mädchen beobachtet wurde, sollte zusätzlich überprüft werden, ob dieser Vorsprung im Leseverständnis auch durch die Anlage der Untersuchung, die Auswahl der Texte oder des Frageformats begünstigt wurde.

Anhand der vorgenommenen Analysen konnte für eine Reihe von Fragen eine differenzielle Item-Funktion für Jungen und Mädchen festgestellt werden. Doch ist das beobachtete DIF-Phänomen ein Problem? – Modelle sollen die Realität beschreiben. Hierbei haben differenziertere Modelle im Vergleich zu einfacheren einen höheren Informationsgehalt und beschreiben aufgrund gesteigerter Komplexität die Wirklichkeit facettenreicher. Welches Modell zur Erklärung der Wirklichkeit herangezogen wird, hängt von der Fragestellung ab, der nachgegangen werden soll. In einigen Fällen sind einfache Modelle angemessen, in anderen Fällen ist auf komplexere zurückzugreifen. Generell ließ sich zeigen, dass das einparametrische Rasch-Modell die Daten angemessen modelliert, dass aber komplexere Modelle auch differenzielle Einblicke erlauben würden. Die Modellierung von geschlechtsspezifischen DIF-Parametern ist eine dieser Möglichkeiten, die hier genutzt wurde.

*Tabelle 1: Zusammenfassung der relativen Unterschiede zwischen Jungen und Mädchen*

Geschlechtsspezifische Unterschiede betreffend:	Signifikanz	Relevanz
Leseverständnis	ja	ja
Frageformat (multiple choice vs. offenes Format)	ja	gering
Textgattung (literarische vs. Informationstexte)	ja	gering
Verstehensaspekte	nein	-
Aufgabenschwierigkeiten	ja	ja

Die relativen Unterschiede zwischen Jungen und Mädchen bei einzelnen Fragen erwiesen sich in der großen Stichprobe als signifikant, absolut betrachtet jedoch als relativ klein (siehe zusammenfassend Tabelle 1). So wurde festgestellt, dass für Fragen mit durchschnittlichem (signifikanten) DIF z.B. zugunsten der Mädchen in einer realen Schulklasse praktisch kein Unterschied beobachtbar sein würde. Erst Analysen in großen Felduntersuchungen lassen eine solche Identifizierung zu. Alle vorgestellten Beispiele belegen statistisch signifikante Unterschiede im relativen Lösungsverhalten von Jungen und Mädchen. Betrachtet man für diese die absoluten Lösungshäufigkeiten, wird man jedoch feststellen können, dass die Differenzen mit knapp bzw. deutlich weniger als 10 Prozentpunkten Unterschied relativ gering ausfallen und somit substantielle Vor- bzw. Nachteile weder für Jungen noch für Mädchen bei den einzelnen Testaufgaben bestehen. Als eine wesentliche Voraussetzung, die Fragen richtig zu beantworten, ist somit auch bei diesen nicht das Geschlecht anzusehen, sondern letztendlich – wie schließlich auch intendiert – das Leseverständnis selbst.

Für den Gesamttest lässt sich feststellen, dass zwar signifikante, aber im Einzelnen keine relevanten Geschlechtsdifferenzen bezüglich des Frageformats und den textformspezifischen Leseleistungen zu beobachten sind. Bei Betrachtung der in IGLU getesteten Verstehensaspekte ergaben sich gar keine Differenzen zwischen Jungen und Mädchen. Lediglich bezüglich der Aufgabenschwierigkeiten konnte ein Zusammenhang mit den ermittelten DIF-Parametern festgestellt werden. Hier bestätigte sich der von LEHMANN (vgl. 1994) für die 1991 durchgeführte *Reading Literacy Study* ermittelte Befund, dass Mädchen einfache Aufgaben routinierter und somit sicherer lösen.

Fastet man die vorgestellten Befunde zusammen, so lässt sich feststellen, dass der Vorsprung, den die Mädchen im IGLU-Leseverständnistest aufweisen, nur ansatzweise durch Frageformat und Textgattung, nicht jedoch durch erfragte Verstehensaspekte begünstigt wird. Betrachtet man die Größe des Vorsprungs, so kann gesagt werden, dass die ermittelten Differenzen praktisch nicht auf die Gestaltung des Tests, sondern auf ein besseres Leseverständnis der Mädchen zurückzuführen sind. Die in IGLU auch getestete Textgattung „literarische Texte“ scheint die Mädchen etwas mehr anzusprechen und somit die dazu gestellten Fragen von diesen auch mit mehr Freude beantwortet zu werden. In weiteren Untersuchungen wäre zu prüfen, ob eine verstärkte Nutzung von Informationstexten im Leseunterricht förderlich ist, um das Defizit von Jungen im Leseverständnis abzubauen – eine Anregung, die auch RICHTER (vgl. 1996) bei Vorstellung der „ökologischen Didaktik“ formuliert. Zudem wäre zu prüfen, ob ein häufigeres Lesen von Informationstexten verstärkt den Jungen Leseanreize bieten könnte, um Interesse und Motivation zu heben und sie so mittelbar besser und auch routinierter lesen lernen zu lassen. Wodurch die in der Grundschule festgestellten Differenzen im Leseverständnis zwischen Jungen und Mädchen im Detail bedingt sind, wird somit in vertiefenden Untersuchungen zu klären sein. Erklärungen zu den Differenzen bei 15-Jährigen haben STANAT und KUNTER (vgl. 2002) für die in PISA getestete Stichprobe beschrieben. Die von ihnen vorgelegten Befunde können mit dem in IGLU erhobenen Datenmaterial jetzt auch für die Grundschulen überprüft werden.

## Literatur

- ANGOFF, W. H. (1993): Perspectives on differential item functioning methodology. In: HOLLAND, P. W./WAINER, H. (Eds.): *Differential Item Functioning*. – Hillsdale, S. 3-23.
- BAUMERT u.a. 1997 = BAUMERT, J./LEHMANN, R. H./LEHRKE, M./SCHMITZ, B./CLAUSEN, M./HOSENFELD, I./KÖLLER, O./NEUBRAND, J. (1997): TIMSS - Mathematisch-naturwissenschaftlicher Unterricht im internationalen Vergleich. Deskriptive Befunde. – Opladen.
- BAUMERT u.a. 2001 = BAUMERT, J./KLIEME, E./NEUBRAND, M./PRENZEL, M./SCHIEFELE, U./SCHNEIDER, W./STANAT, P./TILLMANN, K.-J./WEIB, M. (2001): PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich. – Opladen.
- BOS, W./SCHWIPPERT, K. (2002): Vom Sinn und Unsinn internationaler Schulleistungsuntersuchungen. In: *Bildung und Erziehung*, 55. Jg., S. 5-23.
- BOS u.a. 2003 = 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.
- CAMPBELL u.a. 2001 = CAMPBELL, J. R./KELLY, D. L./MULLIS, I. V. S./MARTIN, M. O./SAINSBURY, M. (2001): *PIRLS Progress in International Reading Literacy Study*. – 2<sup>nd</sup> ed. – Boston.
- DIJK, T. A. VAN/KINTSCH, W. (1983): *Strategies of discourse comprehension*. – New York.
- DORANS, N. J./HOLLAND, P. W. (1993): DIF Detection and Description: Mantel-Haensel and Standardization. In: HOLLAND, P. W./WAINER, H. (Eds.): *Differential Item Functioning*. – Hillsdale, S. 35-66.
- ELLEY, W. B. (1994): *The IEA Study of Reading Literacy: Achievement and instruction in thirty-two school systems*. – Exeter.
- ELLEY, W. B./MANGUBHAI, F. (1992): Multiple-choice and open-ended items in reading tests: Same or different? In: *Studies in Educational Evaluation*, Vol. 18., S. 191-199.
- FOY, P./JONCAS, M. (1999): *School Sampling Manual*. – Ottawa.
- GROEBEN, N./HURRELMANN, B. (Hrsg.) (2002): *Lesekompetenz. Bedingungen, Dimensionen, Funktionen*. – Weinheim.
- KIRSCH, I. S./MOSENTHAL, P. B. (1989–1991): *Understanding documents. A monthly column appearing in the Journal of Reading*. – Newark.
- KISH, L. (1995): *Survey Sampling*. – New York.
- LEHMANN, R. H. (1994): Lesen Mädchen wirklich besser? Ergebnisse aus der internationalen IEA-Lesestudie. In: RICHTER, S./BRÜGELMANN, H. (Hrsg.): *Mädchen lernen ANDERS lernen Jungen: Geschlechtsspezifische Unterschiede beim Schriftspracherwerb*. – Lengwil, S. 99-109.
- LEHMANN u.a. 1995 = LEHMANN, R. H./PEEK, R./PIEPER, I./STRITZKY, R. V. (1995): *Leseverständnis und Lesegewohnheiten deutscher Schüler und Schülerinnen*. – Weinheim.
- MULLIS u.a. 2003 = MULLIS, I. V. S./MARTIN, M. O./GONZALEZ, E. J./KENNEDY, A. M. (2003): *PIRLS 2001 International Report*. – Boston.
- OECD (1995): *Literacy, Economy and Society. Results of the first international adult literacy survey*. – Paris: OECD.
- OECD (2001): *Knowledge and Skills for Life: First results from PISA 2000, education and skills*. – Paris: OECD.
- RASCH, G. (1980): *Probabilistic Models for Some Intelligence and Attainment Test*. – Chicago.
- RICHTER, S. (1995): Mädchen lesen (und schreiben) Anders als Jungen. In: Balhorn, H. (Hrsg.): *Bücher – eine Welt zum Lesen, Schreiben, Träumen...* – Vechta, S. 66–75.
- RICHTER, S. (1996): *Unterschiede in den Schulleistungen von Mädchen und Jungen*. – Regensburg.
- RICHTER, S./BRÜGELMANN, H. (Hrsg.) (1994): *Mädchen lernen anders lernen Jungen*. – Konstanz.
- RICHTER, S./BRÜGELMANN, H. (1995): Jungen und Mädchen lernen verschieden. In: *Beiträge zur Lehrerbildung*, 13. Jg., Nr. 1, S. 75–76.
- STANAT, P./KUNTER, M. (2001): Geschlechtsunterschiede in Basiskompetenzen. In: BAUMERT, J. u.a. (Hrsg.): *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich*. – Opladen, S. 249–269.
- STANAT, P./KUNTER, M. (2002): Geschlechterspezifische Leistungsunterschiede bei Fünfzehnjährigen im internationalen Vergleich. In: *Zeitschrift für Erziehungswissenschaft*, 4. Jg., S. 28–48.
- VALTIN u.a. 2003 = VALTIN, R./BADEL, I./LÖFFLER, I./MEYER-SCHEPERS, U./VOSS, A. (2003): *Orthographische Kompetenzen von Schülerinnen und Schülern der vierten Klasse*. In: Bos, W. u.a. (Hrsg.):

Erste Ergebnisse aus IGLU. Schülerleistungen am Ende der vierten Jahrgangsstufe im internationalen Vergleich. – Münster, S. 227–264.

WAGEMAKER, H. (1996): Are Girls Better Readers? Gender Differences in Reading Literacy in 32 Countries. – Wellington.

WU, M. L./ADAMS, R. J./WILSON, M. R. (1998): Acer ConQuest. Generalised item response modelling software. – Victoria.

*Anschriften der Verfasser:* PD Dr. Knut Schwippert, Prof. Dr. Wilfried Bos, Dr. Eva-Maria Lankes, Universität Hamburg, FB Erziehungswissenschaft, Institut für International und Interkulturell Vergleichende Erziehungswissenschaft, Sedanstr. 19, 20146 Hamburg, Tel. Schwippert: 040-42838-7094, e-mail: Schwippert@erzwiss.uni-hamburg.de