Lese-Untersuchung mit dem Stolperwörter-Test

Abschlussbericht [1] des Projekts LUST-1 [2]

von

Hans Brügelmann (2003k) [3]

(Universität Siegen)

1. Ziele und Kontext des Projekts LUST-1

IGLU-Day brachte die Wende. Am 8. April wurden die Ergebnisse der Internationalen Grundschul-Lese-Untersuchung veröffentlicht. Von einem Tag auf den anderen entzogen die Befunde [4] dem bildungspolitischen Konsens der Diskussion nach PISA den Boden. Monatelang hatten Ministerien, Parteien und Verbände einen Vorschlag nach dem anderen aus dem Boden gestampft, was in Vor- und Grundschule geändert werden müsste, um das „blamable Abschneiden“ bei PISA auszumerzen. Mit IGLU konnte endlich die abenteuerliche Logik außer Kraft gesetzt werden: Wenn 15-Jährige nicht gut genug lesen können, dann muss sich der Anfangsunterricht ändern, denn Lesen lernt man im ersten und evtl. noch im zweiten Schuljahr. Dass Lesenlernen ein lebenslanger Prozess ist und dass zum Lesenlernen mehr gehört als Wörter zu entziffern und geübte Texte sinngerecht intoniert vorzutragen, dämmerte Vielen erst, als deutlich wurde, dass bereits ViertklässlerInnen Texte nutzen können, um komplexe Fragen und Aufgaben inhaltlich zu bearbeiten. Wozu neben IGLU also noch LUST?

In der aktuellen Diskussion über Evaluation von Schule und Unterricht dominieren Studien zum „System Monitoring“. Das Ziel von TIMSS, PISA und IGLU ist eine Bestandsaufnahme „des Systems“. Adressaten sind die Bildungspolitik und andere Funktionsträger, die *allgemeine* Entscheidungen zu treffen haben.

---

[1] Dieser Bericht sowie weitere Ergebnisse und Interpretationen werden zum Download veröffentlicht unter
➔ **www.uni-siegen.de/~agprim/lust/index.htm** .

[2] **L**ese-**U**ntersuchung mit dem **S**tolperwörter-**T**est (s. Anm. 6). Diese Studie wurde durchgeführt und ausgewertet von Hans Brügelmann, Siegen, in Kooperation mit Wilfried Metze, Berlin (Autor der Tests) und mit Erika Brinkmann, Schwäbisch Gmünd (Parallelstudie Baden-Württemberg, in Vorb.). Die Auswertung von LUST wurde finanziell gefördert von der Zukunftsstiftung Bildung, Bochum, und vom Grundschulverband – Arbeitskreis Grundschule e.V., Frankfurt. Beiden Förderern danken wir für sehr rasche und unbürokratische Hilfe, als uns der unerwartet hohe Rücklauf aus den Schulen zu überwältigen drohte. Die Nachfolge-Studien LUST-2 (2003/2004) und LUST-3 (2004-2006), die zum Teil echte Längsschnitte enthalten werden, sind zur Zeit in Vorbereitung (Koordination: Axel Backhaus).

[3] Ich danke ganz herzlich Katrin Belz, Sina Gerlach, Oliver von Keutz, Steffi Maxa, Verena Reich, Sara Roth, Markus Spannan und Petya Todorova, die gemeinsam mit mir das fast endlose Geschäft des Kodierens übernommen haben.

[4] s. Kasten 1

Um die Qualität von Unterricht in der einzelnen Klasse zu verbessern, braucht aber die Lehrperson vor Ort Informationen über die Leistung ihrer Klasse. Der Grundschulverband hat deshalb bereits vor einigen Jahren ein mehrstufiges Modell der Rechenschaftspflichten und der Evaluationsformen vorgeschlagen (vgl. Bartnitzky u. a. 1999). In diesem Modell hat eine interne Evaluation, die externe Kennwerte zur Einordnung der eigenen Daten nutzt, Vorrang (vgl. zur Begründung ausführlicher Brügelmann 2000; Rolff 1999; 2003). Dieser Anspruch hat Folgen für Design und Methoden der Untersuchung:

- Statt landesweit repräsentativer Stichproben sind Vollerhebungen von ganzen Klassen in einer Schule bzw. Region erforderlich.

- Statt zeitlich und methodisch hohen Aufwands für Durchführung und Auswertung werden robuste und unaufwändige Verfahren benötigt, die von den KollegInnen im Alltag mit Gewinn für ihre Alltagsarbeit eingesetzt werden können.

Da die Schulministerien zur Zeit Großstudien favorisieren [5] hat der Grundschulverband ein Projekt der Universität Siegen unterstützt, in dem in drei Schulamtsbezirken Nordrhein-Westfalens [6] allen LehrerInnen der 2., 3. und 4. Klassen ein einfacher Lesetest [7] und – für die Erstdurchführung [8] – auch deren Auswertung angeboten wurde [9].

Im Folgenden werden die Daten von 18.083 Kindern der Klassen 2 (6.039), 3 (6.198) und 4 (5.846) berichtet. Ihre LehrerInnen haben den Test im Januar bzw. Februar 2003 durchgeführt, also zur Schuljahresmitte [10].

Damit die im Folgenden berichteten Daten nicht missverstanden werden, schicken wir einige Hinweise zur Interpretation der Kennwerte vorweg:

---

[5] Vgl. neben TIMSS und PISA speziell für den Grundschulbereich IGLU.

[6] Ich danke den beteiligten Schulämtern für ihre Unterstützung, insbesondere meinen Kontaktpersonen SAD'in Karin Brügelmann (Rhein-Sieg-Kreis), SAD'in Heidemarie Goßmann (Märkischer Kreis) und SAD Wolf Kuhnke (Kreis Siegen-Wittgenstein) sowie den Schulleiterinnen und LehrerInnen, die in unerwartet großem Umfang an der Untersuchung und zum Teil mit zusätzlichem persönlichen Engagement an der Auswertung mitgewirkt haben.

[7] Dieser Test ist von Wilfried Metze, Berlin, für 1. bis 4. Klassen entwickelt worden (vgl. zu den Aufgaben, ihrem Hintergrund und ersten Ergebnissen ➔ www.wilfriedmetze.de/lesetest oder www.lesetest1-4.de). In der Aufgabe geht es darum, in 45 bzw. 60 einzelnen Sätzen mit 4-15 Wörtern ein jeweils nicht passendes Wort zu streichen.

[8] Dies war erforderlich, um den Schulen rasch Vergleichswerte für die Interpretation der eigenen Daten anbieten zu können.

[9] LUST ist insofern als Komplementäruntersuchung zur Internationalen Grundschul-Lese-Untersuchung (IGLU/PIRLS) zu lesen, deren Ergebnisse am 8.4.03 publiziert werden. Während IGLU sich auf das Bildungssystem bezieht und eine Bestandsaufnahme im Sinne des „System Monitoring" versucht, ist LUST als praxisbezogene Evaluation angelegt: einzelne LehrerInnen erhalten ein unaufwändiges Instrument sowie Vergleichsdaten, um den Leistungsstand einzelner SchülerInnen, aber auch ihrer Klasse insgesamt besser einschätzen zu können. IGLU ist eine bundesweit repräsentative Stichprobe ausgewählter Schulen, LUST vom Ansatz her eine Vollerhebung in drei Schulbezirken. Und anders als bei IGLU („Textverständnis") geht es bei LUST um die grundlegenden „Taktiken" des Lesens : *rasches* und *genaues Satz*lesen durch Verbindung verschiedener Zugriffe auf Schrift (s. a. Anm. 6 und ergänzend zu diesem Ansatz : Arbeitsgruppe Leseförderung 1978).

[10] 31 Kinder aus einer ersten Klasse bleiben ebenso außen vor wie weitere über 3.000 SchülerInnen, die den Test vor oder nach der spezifizierten Phase durchgeführt haben.

- Der Test stellt eine Art „Warnlampe“ für den Stand der Leseentwicklung dar, er bietet keine spezifische Diagnose des Leistungs*profils*. Die Ergebnisse punktueller Erhebungen sind immer nur eine Information unter anderen. Das gilt auch für standardisierte Tests. Sie können begleitende Beobachtungen in wichtiger Weise ergänzen, aber sie können keine höhere Autorität beanspruchen als andere Daten (z. B Beobachtungen der Lehrperson über ein ganzes Schuljahr hinweg).

- Da es sich um eine punktuelle Aufgabe handelt, kann auch das Ergebnis einzelner SchülerInnen von ihrer sonst üblichen Leistung abweichen. Solche Abweichungen sollten für die Lehrperson Anlass zum Nachdenken und ggf. zu einer genaueren Beobachtung des Kindes sein. Aus ihrer längeren Kenntnis des einzelnen Kindes muss sie die punktuellen Testergebnisse ggf. aber auch relativieren – positiv wie negativ.

- Auch die Klassenwerte sind interpretationsbedürftig. Ein Vergleich mit anderen Klassen ist hilfreich, um die Binnenwahrnehmung und die eigenen Maßstäbe zu überprüfen. Aber die Voraussetzungen, die die Kinder in den Unterricht mitgebracht haben, und die Bedingungen, unter denen er stattfindet, kennen nur die vor Ort Beteiligten. Ggfs. ist also auch das Klassenergebnis entsprechend zu relativieren.

- Was Unterschiede im Test für Leistungen im Alltag bedeuten ist ebenfalls nicht leicht einzuschätzen. Bewertungen wie „ein Schuljahr zurück“ oder „80% aller Klassen sind besser“ können dramatische Unterschiede, aber auch geringfügige Differenzen ausdrücken. Man muss die konkreten Maßstäbe und die absoluten Zahlen betrachten, um solche Aussagen zutreffend einschätzen zu können.
  Wenn z. B. die Leistungen der meisten SchülerInnen eng beieinander liegen, kann schon ein einziger Fehler einen großen Sprung auf der Rangskala bedeuten [11] .
  Wenn andererseits der Test sehr schwierig ist, können sich die absoluten Werte deutlich voneinander unterscheiden – ohne dass diese Unterschiede im Alltag relevant sein müssen.

Im Folgenden habe ich versucht,

a) solche kritischen Stellen der Urteilsbildung ausdrücklich zu markieren
und
b) unsere eigenen Maßstäbe und Annahmen immer durchsichtig zu machen,

damit die LeserInnen unsere Einschätzungen selbst überprüfen können.

Einen wichtigen Bezugspunkt unserer Studie bilden die Befunde der Internationalen Grundschul-Lese-Untersuchung (IGLU). Wir haben deshalb vorweg noch einmal deren zentrale Ergebnisse in einigen Thesen zusammengefasst (vgl. Kasten 1).

[11] Dies ist z. B. in der „Hamburger Schreibprobe“ von May (2002) bei der Auswertung nach einzelnen Rechtschreibstrategien häufig der Fall.

[Kasten 1]

Hauptbefunde aus IGLU – eine kurze Übersicht zur Lesekompetenz in Klasse 4 [12]

- In der **Durchschnittsleistung** erreichen die deutschen SchülerInnen unter 35 Ländern Platz 11. Sie liegen damit im Mittel der EU-Länder, etwas höher als der OECD-Durchschnitt. Statistisch signifikant sind nur Schweden, die Niederlande und Bulgarien besser. Korrigiert um Alter und Dauer der Schulzeit liegt Deutschland gemeinsam mit sieben anderen Ländern auf Platz 5-12. Sowohl die EU- als auch die OECD-Gruppe ist mit den entsprechenden PISA-Gruppen vergleichbar. Da Deutschland dort unter dem Mittelwert lag, scheint die Grundschule im internationalen Vergleich besser abzuschneiden als die Sekundarstufe. (S. 10, 17)

- Schon die mittleren zwei Drittel der SchülerInnen streuen über drei Schuljahre durchschnittlicher Leseleistung. Diese **Streuung** ist aber in den meisten anderen Ländern noch wesentlich höher. In Deutschland ist sie auch unter den 15-Jährigen wesentlich höher. (S. 12)

- Der Anteil von **Risikokindern** auf Lesestufe I wird mit 10% geringer angesetzt als in den meisten anderen Ländern (zum Vergleich: Niederlande 4%, Schweden 5-6%, aber USA 13-14%, Norwegen 25%). (S. 15). Er ist vor allem deutlich geringer als die bei PISA geschätzten 25% am Ende der Schulzeit (vgl. Baumert u. a. 2001).

- Aber auch die **Spitzengruppe** der deutschen SchülerInnen ist nicht so ausgeprägt: 18% auf Kompetenzstufe IV (zum Vergleich: England 30%, Schweden 28% -- aber in Frankreich 14% und Norwegen 9%). (S. 16)

- **Jungen** liegen im Durchschnitt etwa 4 Monate hinter den Mädchen. Das ist weniger als in den meisten anderen Ländern. Zudem ist der Unterschied erheblich geringer als am Ende der Sekundarstufe (= PISA). (S. 35)

- Die Leseleistung korreliert hoch mit dem **sozio-ökonomischen** Hintergrund der Familien (S. 33). Konkret liegen Kinder aus Elternhäusern mit weniger als 100 Büchern gegenüber denen mit mehr als 100 Büchern um knapp ein Schuljahr zurück. Das ist allerdings weniger als in vielen anderen Ländern. Und der Unterschied ist erheblich geringer als am Ende der Sekundarstufe (= PISA). (S. 36)

- Die IGLU-Ergebnisse weisen aber auch nach, dass soziale Faktoren bei der **Schulempfehlung** nach der vierten Klasse mit entscheiden, nicht nur die Leistungsfähigkeit der Schüler selbst. Bei gleichem Testergebnis erhalten 40 Prozent der Kinder eine Empfehlung für die Realschule, 33 Prozent für das Gymnasium und 22 Prozent für die Hauptschule (Bos u. a. 2003, 130-134).

- Kinder, deren Eltern beide im **Ausland** geboren sind, liegen in der 4. Klasse über ein Schuljahr hinter denen mit zwei in Deutschland geborenen Eltern zurück. Das ist *mehr* als in den meisten anderen Ländern. Aber der Unterschied ist dennoch erheblich *geringer* als am Ende der Sekundarstufe (= PISA). (S. 37)

---

[12] Zusammengefasst anhand des IGLU-Kurzberichts, der unter ➔ www.erzwiss.uni-hamburg.de/IGLU/home.htm) zugänglich ist; s. ergänzend zum Kontext des IGLU-Projekts Brügelmann (2003a). Der internationale Bericht findet sich unter: ➔ http://timss.bc.edu/pirls2001.html; die deutsche Langfassung ist publiziert als: Bos, W., u. a. (Hrsg.) (2003): Erste Ergebnisse aus IGLU. Schülerleistungen am Ende der vierten Jahrgangsstufe im internationalen Vergleich. Waxmann: Münster u. a.

2. Maße und Kategorien der Auswertung des Stolperwörter-Lesetests

Die Leistung im Test kann zusammenfassend durch die Zahl der richtig bearbeiteten Sätze beschrieben werden („Rohpunkte“ in der Anweisung des Testautors Wilfried Metze). In dieses Maß gehen allerdings zwei Teilleistungen ein, die bei gleichem Summenwert sehr unterschiedlich aussehen können: Die Geschwindigkeit (Zahl der *bearbeiteten Sätze*) und die Genauigkeit (Zahl der *Fehler*).

Ein Beispiel: Bearbeitet ein Viertklässler in 4 min. 50 Sätze und macht dabei 15 Fehler, kommt er auf 35 Punkte. Das gilt aber auch für jemanden, der 37 Sätze bearbeitet und zwei Fehler macht. Trotz gleichen Punktwerts handelt es sich um sehr unterschiedliche Leistungen. Entsprechend unterschiedlich müsste die Rückmeldung an diese beiden Kinder aussehen, und auch eventuelle Fördermaßnahmen sollten spezifisch ausfallen [13].

In unserer eigenen Auswertung haben wir deshalb beide Größen erfasst. Soweit wir in unseren Berechnungen auf die Selbstauswertungen der LehrerInnen zurückgreifen, sind nur die „Rohpunkte“ verfügbar. Da in unseren Auswertungen dieser Wert als Differenz von „bearbeiteten Aufgaben minus Fehler“ ebenfalls vorliegt, ist die Zahl der Fälle für die berichteten Rohpunkte größer als für die differenzierte Teilauswertung nach bearbeiteten Sätzen und Fehlern.

Zur Beurteilung der Testqualität werden vor allem zwei Maße herangezogen: die Validität, d. h. die inhaltliche Aussagekraft eines Tests [14] und die Reliabilität, d. h. die Verlässlichkeit, mit der er die Daten erhebt.

Die Validität ist vor allem deshalb zu überprüfen, weil der Stolperwörter-Lesetest ungewöhnlich konstruiert ist, um die *grundlegende Lesefähigkeit* zu erfassen. Es geht um mehr als nur um rasches Worterkennen und um weniger als das Verständnis von Textzusammenhängen. Andererseits fordert der Test mehr als die reine Leseleistung: Das *inhaltliche* Satzverständnis muss in eine bewusste *sprachform*bezogene Entscheidung übersetzt werden. Wir haben die inhaltliche Aussagekraft deshalb in zwei Außenvergleichen überprüft [15] :

- Vergleich mit den Ergebnissen in parallel eingesetzten Tests zum raschen Worterkennen (.41** bzw. .76** mit dem IEA-O40-Worttest [16]) und zum Textverständnis (r = .61**, .87* bzw. .86** mit dem HAMLET [17]);

---

[13] Nur selten sind Fördermaßnahmen so klug gestaltet wie im Computerprogramm LALIPUR, das kostenlos herunterzuladen ist von der Homepage unser Didaktischen Entwicklungs- und Prüfstelle für Lernsoftware Primarstufe : www.uni-siegen.de/~agprim/DEPnew/index.htm . Dieses Programm (allerdings nur für MS-DOS, nicht für WINDOWS ausgelegt) erlaubt den Kindern (fordert aber auch von ihnen), die gegenläufigen Anforderungen von Genauigkeit und Geschwindigkeit selbst zu optimieren, um in einem Adventure Game Erfolg zu haben.

[14] Zentrale Frage: „Misst der Test wirklich, was er zu messen behauptet?“

[15] Entsprechend den üblichen statistischen Konventionen markieren wir die statistische „Signifikanz“ von Korrelationen und Unterschieden bei einer Fehlerwahrscheinlichkeit von höchstens 5% mit einem „*“ und bei einem Fehlerrisiko von höchstens 1% mit „**“.

[16] N = 43 (Projekt LUST-1) bzw. N = 12 (Projekt LUST-2) [Stand 4.12.03]

[17] N = 44 und N = 8 (Projekt LUST-1) bzw. N = 23 (METZE-Studie) [Stand: 4.12.03]

- Vergleich mit den Lesenoten, die die LehrerInnen vergeben (r = .55** bis .64** [18]), und mit ihren Urteilen über einzelne Aspekte der Leseleistung ihrer SchülerInnen (r = . 56** bis .74**).

Auch wenn es sich nur um kleine Stichproben handelt und die Zusammenhänge unterschiedlich stark ausfallen, liegen die Korrelationen im Rahmen der üblichen Werte, so dass die im Folgenden berichteten Befunde als guter Indikator für die grundlegende Lesefähigkeit betrachtet werden können. Dennoch planen wir für weitere Studien eine Differenzierung des Tests durch die Kombination verschiedener Teilaufgaben, vor allem um seine diagnostische Aussagekraft zu steigern [19] .

Die Reliabilität des Tests haben wir auf zwei Wegen geprüft. Die Stabilität der Testwerte wird durch die gute Vorhersage der Rangfolge in der Leseleistung eine Woche (r = .90**) und ein halbes Jahr später bestätigt (r = .81**) [20] . Als Maß für die interne Konsistenz des Tests haben wir zusätzlich Cronbachs Alpha berechnet. Die Werte liegen für zweite und vierte Klassen und für die beiden Teilformen A und B durchgängig über .88** [21] – ebenfalls ein sehr befriedigender Wert für die Verlässlichkeit des Tests.

Blicken wir man nun auf die Klasse als Auswertungseinheit. Hier ergibt sich ein zusätzliches Problem. Mittelwerte sind eine Größe, die die Leistung der Gruppe auf einen Blick vermitteln kann. Aber derselbe Durchschnitt kann aus ganz unterschiedlichen Verteilungen resultieren.

Wieder ein Beispiel: In einer Klasse mit einem arithmetischen Mittel von 35 Punkten könnte der schwächste Schüler 10 und der stärkste 60 Punkte haben. Derselbe Mittelwert kann sich z. B. aber auch ergeben, wenn der schwächste Schüler 30 und der leistungsstärkste 40 Punkte hat. Das Leistungsniveau der beiden Klassen kann also durch den Mittelwert allein nicht ausreichend beschrieben werden.

Wir geben deshalb neben dem arithmetischen Mittel (aM) immer noch in Klammern die Standardabweichung (SD [22]) an. Sie ist ein Maß für die Streuung. Bei normal verteilten Werten („Glockenkurve“ mit Häufung der Fälle im mittleren Bereich wie bei unseren Werten) kann als Faustformel gelten: je 1/3 aller Fälle liegt im Bereich von +/- einer Standardabweichung *über* bzw. *unter* dem Mittelwert. Oder anders ausgedrückt: 2/3 der Stichprobe liegt zwischen den Prozenträngen 16 und 84.

In einer gesonderten Auswertung weisen wir zusätzlich die Ergebnisse der *im Ergebnis* schwächsten und der stärksten Klasse aus – immer mit dem bereits erwähnten Vorbehalt, dass wir über die Bedingungen, unter denen diese Leistungen erzielt wurden, nichts wissen. Die gro-

---

18 Innerhalb der einzelnen Klassen höher, s. unten Kap. 9

19 Vgl. Backhaus/ Brügelmann (2003).

20 Die ersten Auswertungen der kurzfristigen Wiederholung zeigen einen Leistungsanstieg von rund 30%. Ob der Übungs- bzw. Vertrautheitseffekt sich auf das Format oder den Inhalt bezieht, lässt sich nicht klären. Allerdings bleibt die Rangfolge der SchülerInnen sehr stabil, wie die Korrelation von .93** zeigt (Stand: 20.12.; N = 162). Die Korrelation bei der Erhebung nach einem halben Jahr lag in acht von elf Klassen bei > .80**, sie war in zwei Klassen > .70** und lag in einer weiteren bei .64** (Stand: 24.10.2003; N = 270).

21 Für die Form A beträgt Cronbachs Alpha bezogen auf die ersten 32 Items .94 in den 2. Klassen (N = 59) und .88 in den 4. Klassen (N = 51), für die Form B betragen die Werte .95 in Klasse 2 (N = 56) und .93 in Klasse 4 (N = 77). (s. a. stolper.satz.03.trennschärfe/ 09-30 / 05-06)

22 SD ist das in der Statistik übliche Kürzel für „Standard Deviation“.

ße Bandbreite der Klassenwerte kann also sowohl die Folge unterschiedlicher Lernvoraussetzungen bzw. Lernbedingungen als auch unterschiedlichen Unterrichts bzw. seiner Bedingungen sein. Hinzu kommt, dass die Tests von den LehrerInnen selbst durchgeführt worden sind. Trotz standardisierter Instruktion und Auswertungsanleitung sind Abweichungen im Einzelfall nicht auszuschließen.

Da die SchülerInnen auf den einzelnen Jahrgangsstufen unterschiedlich viel Zeit hatten, um die 60 Sätze der Aufgabe zu bearbeiten, sind die Ergebnisse nicht direkt vergleichbar. Wir haben deshalb zusätzlich errechnet, wie viele Sätze die Schüler im Mittel *pro Minute bearbeitet* haben und wie viele davon *pro Minute richtig.* Diese Kennwerte sind zugleich – neben der *Fehlerquote* [23] – sehr anschauliche Maße für die Leistung einzelner Kinder und für den Vergleich verschiedener Gruppen.

Beispiel: In mehreren Probeversuchen mit kleinen Gruppen von LehrerInnen und LehramtsstudentInnen zeigte sich z. B., dass diese sehr geübten LeserInnen knapp 3 sek. pro Satz brauchten. Selbst in dieser homogenen Gruppe kompetenter LeserInnen streute das Lesetempo allerdings zwischen 2 und 5 Sekunden, also mit einer Differenz von 150%: Sie schafften zwischen 12 und 30 Sätzen pro Minute.

Um die Ergebnisse einzelner SchülerInnen bzw. Klassen genauer einschätzen zu können, schlüsseln wir die Ergebnisse im Folgenden zusätzlich weiter nach folgenden Gruppen auf:

- Jahrgang
- Leistung
- Geschlecht
- Muttersprache der Eltern
- Familienkonstellation
- Klasse
- Schulbezirk.

Bedeutsam für solche Vergleiche ist der Standardfehler (SE [24]), den wir in einigen Tabellen zusätzlich angeben. Da die verrechneten Daten aus Stichproben stammen, stellen sie immer nur eine Schätzung der Werte in der Grundgesamtheit dar. Wie in den meisten Studien üblich beschreibt der hier berichtete Standardfehler (x 2) den Bereich, innerhalb dessen der „wahre Wert“ mit 95%-iger Wahrscheinlichkeit liegt. Vergleicht man beispielweise zwei Mittelwerte, so ist nicht jeder zahlenmäßige Unterschied zwischen ihnen schon statistisch „signifikant“. Dies gilt erst, wenn sich auch die beiden Sicherheitsbereiche (Mittelwerte +/- 2 SE) nicht überlappen.

Wegen der Größe unserer Stichproben werden allerdings auch sehr kleine Unterschiede (statistisch) „signifikant“. Wir geben deshalb zusätzlich die Effektstärke ES als ein Maß für die inhaltliche Bedeutung des Unterschieds an [25] .

---

[23] Anteil der ausgelassenen bzw. falsch bearbeiteten Sätze – bezogen auf die insgesamt bearbeiteten Sätze.

[24] SE ist das in der Statistik übliche Kürzel für „Standard Error“. Streng genommen ist dieses Maß nur bei Zufallsstichproben anwendbar. In unserem Fall haben die LehrerInnen bzw. Schulen freiwillig (aber in einem sehr großen Umfang) teilgenommen, so dass der SE (nur) als Näherungswert für die Fehlermarge genommen werden darf.

[25] Die „Effektstärke“ (ES als Differenz der Mittelwerte geteilt durch den Mittelwert der beiden Standardabweichungen oder durch die SD der Hauptgruppe) ist ein gebräuchliches Maß für die quantitative Einschätzung der Be-

Ergänzende Auswertungen beziehen sich auf das Verhältnis von Noten und Testleistung und auf die Bedeutung der einzelnen Kriterien, nach denen eine Teilgruppe von LehrerInnen ihre SchülerInnen bewertet hat.

Außerdem werden wir Auswertungen auf Klassen-Ebene durchführen, um die Bedeutung von Faktoren wie Klassengröße, Anteil von Kindern mit anderer Muttersprache im Elterhaus usw. zu untersuchen.

Fazit zu Kap. 2:

Die Leistung der Kinder kann durch drei Maße ausgedrückt werden: das Tempo der Bearbeitung, die Quote der Fehler und – als zusammenfassender Wert – durch die Zahl der richtigen Sätze pro Minute.

Nach ersten Analysen in ausgewählten Teilstichproben halten wir den Stolperwörter-Lesetest sowohl für zureichend valide, also für inhaltlich aussagekräftig, als auch für reliabel, also verlässlich.

Neben dem Mittelwert ist die Standardabweichung (SD) als Maß Streuung der Werte ein wichtiger Indikator für das Leistungsniveau einer Gruppe (rund 2/3 der Gruppe liegt innerhalb von +/- einer SD um den Mittelwert).

Vergleicht man Mittelwerte, sind die Standardfehler (SE) hinzuzurechnen, um festzustellen, ob ein Unterschied statistisch signifikant, d. h. nicht zufallsbedingt ist.

Die Effektstärke (ES) gibt an, wie groß die quantitative Bedeutung eines Unterschieds einzuschätzen ist.

deutung von Mittelwertunterschieden. Eine ES von .50 gilt als stark, bedeutet sie doch eine Positionsverschiebung z. B. vom Mittelwert (= PR 50) in Stichprobe I auf PR 69 in Stichprobe II.

3. Ergebnisse der Jahrgänge im Vergleich [26]

Zwei Bezugswerte sind zur Einschätzung der Daten über die Grundschulzeit hinweg interessant: zum einen das Niveau von SchülerInnen am Ende der ersten Klasse, sozusagen als „Ausgangswert“ von Leseanfängern, zum anderen die Leistungen von lesekundigen Erwachsenen, sozusagen der „Zielwert“. Für Daten zu den ersten Klassen konnten wir auf eine Erhebung von Metze am Ende des ersten Schuljahres zurückgreifen [27]. Um einen Maßstab für die Zielperspektive zu gewinnen, haben wir den Test selbst mit einigen Studierenden und LehrerInnen durchgeführt [28] -- ein sehr anspruchsvoller Maßstab, wenn man bedenkt, dass immer noch weniger als 30% eines Jahrgangs die allgemeine Hochschulreife erwirbt. Die folgende Tabelle 3.1a zeigt die Bezugswerte im Überblick:

Tab. 3.1a: Mittelwerte von ErstklässlerInnen und Erwachsenen

Altersstufe	Bearbeitete Sätze pro Minute	Anteil Fehler / bearbeitete Aufgaben	Richtige Sätze pro Minute
Ende 1. Klasse	2.2	19.7 %	1.8
Erwachsene	18.4	0.8 %	18.2

Vor diesem Hintergrund von „Ausgangs- und Zielwerten“ lassen sich die Leistungen der Zweit- bis ViertklässlerInnen nicht nur quantitativ, sondern auch qualitativ einschätzen.

Eine Übersicht über die erreichten Leistungen zu verschiedenen Zeitpunkten gibt die folgende Tab. 3.1b [29] :

Tab. 3.1b: Mittelwerte der Jahrgänge 2 bis 4 im Vergleich

Jahrgangsstufe	Bearbeitete Sätze pro Minute	Anteil Fehler / bearbeitete Aufgaben	Richtige Sätze pro Minute
Mitte 2. Klasse	4.8	12.7 %	4.1
Mitte 3. Klasse	6.6	6.3 %	6.1
Mitte 4. Klasse	8.7	4.5 %	8.1

[26] Bei unserer Studie handelt es sich um drei gleichzeitig erhobene Querschnitte. Insofern kann man im strengen Sinne nicht von einer „Entwicklung“ (einzelner Kinder) sprechen. Die Daten zeigen aber, welches Niveau Schülergruppen verschiedenen Alters unter gleichen Bedingungen zu den genannten Zeitpunkten erreichen, so dass mit großer Plausibilität eine Durchschnittsentwicklung erschlossen werden kann.

[27] Die Daten wurden errechnet aus Angaben in dem Bericht auf seiner Homepage www.lesetest1-4.de [Zugriff af das Manuskript 10.8.2003]. Für das erste Schuljahr handelt es sich um 1.536 Kinder aus 72 Klassen.

[28] Stand der Auswertungen von N = 41 : 11.8.2003 [plus eine Gruppe von 19 Schweizer LehrerInnen, die 14.9 richtige Sätze pro Minute schafften, also rund 4 Sekunden pro Satz brauchten)

[29] Auszug aus Tab. 3.1c im Anhang. Der Übersichtlichkeit halber sind die statistischen Kennwerte nur in den Volltabellen im technischen Anhang enthalten. Auch unter Berücksichtigung der Standardfehler sind alle Mittelwertunterschiede hoch signifikant.

Mitte 2. Klasse ist in den vielen Klassen der Leselehrgang „abgeschlossen“, wie es in den meisten Lehrplänen heißt. Dies ist der früheste Termin unserer Erhebungen (Januar/ Februar 2003).

Die **Zweitklässler** schaffen im Durchschnitt in sechs Minuten 25 Sätze *richtig*. Damit verdoppeln sie nur sechs Monate später das Ergebnis der Vergleichskinder Ende erster Klasse, brauchen aber vier- bis fünfmal so lange wie die von uns getesteten Erwachsenen. Wie auch die Fehlerquote zeigt, liegt für den durchschnittlichen Zweitklässler die Hauptschwierigkeit nur noch teilweise im Lesen-*Können*: Knapp 13% der Sätze werden falsch bearbeitet, gegenüber immerhin noch fast 20% Ende erste Klasse. Die besondere Anforderung liegt also in der Lese*geschwindigkeit*.

Drittklässler bearbeiten in fünf Minuten [30] durchschnittlich 30 Sätze *richtig*. Sie brauchen etwa ein Drittel weniger Zeit als die Zweitklässler pro Satz. Zusätzlich halbiert sich die Fehlerquote. Damit bearbeiten sie pro Minute fast 50% mehr Sätze richtig als die ZweitklässlerInnen.

Im Durchschnitt bearbeiten die **Viertklässler** unserer Stichprobe fast 9 Sätze pro Minute. Sie erreichen damit schon die Hälfte dessen, was erwachsene VielleserInnen bewältigen. Zugleich sinkt die Fehlerquote noch einmal um ein Drittel, so dass am Ende der vierten Klasse durchschnittlich 8 Sätze pro Minute *richtig* bearbeitet werden – gegenüber 18 richtigen Sätzen bei den mindestens 15 Jahre älteren Erwachsenen. Für die im „Stolperwörter-Test“ geprüfte *grundlegende* Leseleistung ist das ein gutes Ergebnis, zumal wenn man die Differenz von rund drei Jahren Lesepraxis der GrundschülerInnen vs. 20 bis 50 Jahren bei den Erwachsenen bedenkt.

Fazit zu Kap. 3.1 :

Im Durchschnitt brauchen Kinder der 2. Klassen 15 sek pro *richtig* bearbeitetem Satz. Drittklässler benötigen nur noch 10 sek und Viertklässler im Durchschnitt 7-8 sek. Bezogen auf die durchschnittlich 3 sek, die von akademisch gebildeten LeserInnen benötigt werden, bedeutet das einen rapiden Lernfortschritt in nur zwei Schuljahren und ein hohes Niveau grundlegender Teilleistungen des Lesens zum Ende der Grundschule.

Dabei verdoppelt sich die Lesegeschwindigkeit von Klasse 2 bis Klasse 4 (nur noch 7 sek statt 14 sek pro bearbeitetem Satz). Zusätzlich schrumpft die Fehlerquote von rund 13% auf 4-5 %.

[30] Nach der Testanweisung des Autors Wilfried Metze wird den Kindern von Klassenstufe zu Klassenstufe weniger Zeit zur Bearbeitung des Tests eingeräumt.

3.2 Vergleich der oberen und der unteren Leistungsgruppen

Die Mittelwerte geben nur einen sehr verdichteten Eindruck vom Leistungsniveau auf den verschiedenen Klassenstufen. Nicht weniger wichtig ist die Bandbreite, in der die Leistungen um diese Mittelwerte streuen [31]. Die hohe Standardabweichungen in Tab 3.1c zeigen, dass die Leistungen auf den einzelnen Jahrgangsstufen erheblich streuen. Einen ersten Überblick vermittelt die folgende Tab. 3.2a:

Tab. 3.2a: Extremgruppenvergleiche der Jahrgänge 2 bis 4 nach richtigen Sätzen pro Min. in Prozentrang-Gruppen (PZR)

Jahrgangsstufe	PZR 1-5	PZR 6-10	PZR 11-15	16-84	PZR 86-90	PZR 91-95	PZR 96-100
Mitte 2. Klasse	0.0 – 0.9	1.0 – 1.5	1.6 – 1.8	...	6.5 – 7.2	7.3 – 8.3	8.5 – 10.5
Mitte 3. Klasse	0.0 – 2.5	2.6 – 3.2	3.4 – 3.7	...	8.5 – 9.2	9.4 – 10.2	10.4 – 12.0
Mitte 4. Klasse	0.5 – 3.8	4.0 – 4.5	4.7 – 5.2	...	11.0 – 11.7	12.0 – 13.0	13.2 – 15.5

Alle Leistungsgruppen werden in der untersuchten Aufgabe von Jahr zu Jahr besser, wobei die Fortschritte (in absoluten Zuwächsen richtiger Sätze pro Minute) in den oberen Prozentranggruppen etwas höher sind.

Diese Übersicht soll zunächst in einzelnen Schritten für die drei Jahrgänge entfaltet werden. Dabei können zusätzlich die Fehlerquoten berücksichtigt werden.

Tab 3.2b: Extremgruppenvergleich Klasse 2

Prozentränge	Richtige Sätze pro Minute	Anteil der Fehler
1 - 5	0.0 – 0.9	43.7 - 100.0
6 - 10	1.0 – 1.5	31.2 - 41.7
11 - 15	1.6 – 1.8	25.0 - 31.0
16 - 85	...	...
86 - 90	6.7 – 7.2	0.0 - 0.0
91 - 95	7.5 – 8.3	0.0 - 0.0
96 - 100	8.5 – 10.5	0.0 - 0.0

In den **zweiten** Klassen bewältigen die besten fünf Prozent also mindestens zehnmal so viele Aufgaben wie die schwächsten fünf Prozent. Auch wenn sich die Werte in den beiden Spalten nicht auf dieselben Personen beziehen müssen, dürfte neben dem Lesetempo die unterschiedliche Genauigkeit beim Lesen zu dieser großen Bandbreite der Leistungen beitragen [32] .

Rein quantitativ sind die Unterschiede im oberen Bereich mit einer Spannbreite von 4 Sätzen/ min. größer als im unteren Bereich (1.5 Sätze/min.). Genauer betrachtet liegen die oberen 15% aber enger beieinander als die unteren 15%. Im unteren Bereich schaffen einige Kinder gar keinen Satz oder landen nur Zufallstreffer, während andere zwar langsam, aber doch erfolgreich

[31] Vgl. zum Folgenden die Ur-Tabellen 3.1e-3.1j im Anhang mit den Häufigkeitsverteilungen der Rohwerte

[32] Vgl. auch die deutliche Korrelation von r = -.34** zwischen Zahl der bearbeiteten Sätze und Fehlerquote.

ein bis zwei Sätze pro Minute bearbeiten – ein erheblicher qualitativer Unterschied. Im oberen Bereich dagegen unterscheiden sich die Leistungen nur graduell – denn ob jemand 7 oder 9 sek pro Satz braucht, beeinflusst seine Lesekompetenz nicht wesentlich. Eine Differenz von fünf Prozenträngen bedeutet im unteren Bereich also ganz andere Leistungsunterschiede als im oberen Bereich.

Schließlich erstaunt, dass die besten 15% Ende der zweiten Klasse nicht weit vom Lesetempo der Erwachsenen entfernt sind (10 vs. 18 Sätze/ min.). Für die Grundleistungen „rasches und genaues Wortlesen“ sowie „Integration der Wortbedeutungen auf Satzebene“ scheint in dieser Teilgruppe der Mastery-Level bereits fast erreicht. Die leistungsstärksten ZweitklässlerInnen könnten unter diesem Aspekt selbst im vierten Jahrgang noch im oberen Leistungsdrittel mithalten! Abgesehen von einer weiteren Steigerung der Geläufigkeit durch häufiges Lesen werden diese Kinder eher in anderen Dimensionen bzw. auf anderen Ebenen der Lesefähigkeit (Textverständnis) machen müssen.

Betrachten wir nun die Werte der einzelnen Leistungsgruppen in **Klasse 3**:

Tab 3.2c: Extremgruppenvergleich Klasse 3

Prozentränge	Richtige Sätze pro Minute	Anteil der Fehler
1 - 5	0.0 – 2.5	22.2 - 84.0
6 - 10	2.6 – 3.2	16.0 - 22.0
11 - 15	3.4 – 3.7	12.5 - 15.8
16 - 85	...	...
86 - 90	8.5 – 9.2	0.0 - 0.0
91 - 95	9.4 – 10.2	0.0 - 0.0
96 - 100	10.4 – 12.0	0.0 - 0.0

Der absolute Zuwachs ist in beiden Extremgruppen mit zwei Sätzen pro Minute etwa gleich groß. Relativ gesehen haben sich die besten fünf Prozent allerdings nur wenig gegenüber der vergleichbaren Gruppe ein Jahr vorher verbessert (Zuwachs um 10-20%). SchülerInnen auf den Prozenträngen 5-15 lesen jetzt (mit 16-23 sek pro richtigem Satz) fast genauso schnell und genau wie der Durchschnitt der ZweitklässlerInnen (15 sek). Dazu dürften neben einem höheren Lesetempo vor allem die Halbierung der Fehlerquote in dieser Gruppe beigetragen haben [33]. Der Abstand zu den unteren 15% ist damit qualitativ geschrumpft, wenn man einmal von den allerschwächsten LeserInnen absieht.

Und die Streuung in **Klasse 4**?

[33] Vgl. auch die deutliche Korrelation von r = -.31** zwischen Zahl der bearbeiteten Sätze und Fehlerquote.

Tab 3.2d: Extremgruppenvergleich Klasse 4		
Prozentränge	Richtige Sätze pro Minute	Anteil der Fehler
1 - 5	0.5 – 3.8	16.7 - 80.0
6 - 10	4.0 – 4.5	11.1 - 16.1
11 - 15	4.7 – 5.2	8.8 - 10.9
16 - 85	...	...
86 - 90	11.0 – 11.7	0.0 - 0.0
91 - 95	12.0 – 13.0	0.0 - 0.0
96 - 100	13.2 – 15.5	0.0 - 0.0

Relativ gesehen schaffen die oberen 15% jetzt nur noch viermal so viele Aufgaben wie die unteren 15% (gegenüber einer Relation von 10:1 in Klasse 2). Absolut gesehen hat sich der Abstand nicht wesentlich verändert.

Wieder sind die oberen 15% um etwa 20% besser als ein Jahr vorher. Sie erreichen sogar das Niveau der Erwachsenen in unserer kleinen Zusatzerhebung.

Zwei Drittel der Stichprobe liegen zwischen 5 und 11 Sekunden pro richtigem Satz. Das ist eine deutliche Differenz – aber die Leistungsunterschiede sind gegenüber der dritten und erst recht der zweiten Klasse qualitativ nicht mehr so bedeutsam. Das zeigt vor allem die erneute Abnahme der Fehlerquote im unteren Prozentrangbereich.

Vergleicht man die äußersten Extreme, lässt sich das Bild noch einmal präzisieren. Mit den oberen bzw. unteren 5% erfassen wir näherungsweise „den besten“ bzw. „den schlechtesten“ Schüler einer durchschnittlichen Klasse.

„Der schwächste Schüler“ einer durchschnittlichen 4. Klasse schafft mit 3,3 richtigen Sätzen pro Minute nur ein Fünftel der Leistung seines besten Mitschülers. Er bearbeitet zwar knapp vier Sätze pro Minute. Damit benötigt er nur drei- bis viermal so lange wie der beste Mitschüler. Gleichzeitig macht er aber neunmal so viele Fehler (15.2 vs. 1.7).

Andererseits erscheint eine Quote von 15% falschen Urteilen in der Gruppe besonders leistungsschwacher SchülerInnen vergleichsweise gering, wenn man an die dramatischen Meldungen über angeblich wachsende „Analphabeten“- und „Legastheniker“-Zahlen denkt. Dennoch zeigen sich auf beiden Indikatoren – Lesetempo und Lesegenauigkeit – deutliche Unterschiede zwischen den Leistungsextremen. Deren alltags- und unterrichtspraktische Bedeutung muss noch gesondert diskutiert werden (s. das folgende Kap. 4).

Fazit zu Kap. 3.2 :

Die leistungsstärksten 15% [34] brauchen bereits Mitte der 2. Klasse nur 8-9 sek, erreichen also zu diesem Zeitpunkt schon das Durchschnittsniveau der 4. Klasse. Ihre Leistung verbessert sich bis Mitte 3. Klasse um 10-20% auf 6-7 sek., bis Mitte 4. Klasse um weitere 20% auf 4.5-5.5 sek. Damit erreicht diese Teilgruppe bereits vor dem Ende der Grundschulzeit das Niveau der akademisch gebildeten LeserInnen. Diese Kinder dürften in einem herkömmlichen Lese"unterricht" massiv unterfordert sein.

Die leistungsschwächsten 15% [35] brauchen Mitte der 2. Klasse mehr als 30 sek., um einen Satz zu erlesen und das nicht passende Wort zu finden, teilweise schaffen sie nur Zufallstreffer (0-2 Sätze in 6 min.). Auffallend ist die hohe Streuung selbst innerhalb dieser Teilgruppe. Damit brauchen diese SchülerInnen das 5- bis mehr als 10-fache der Zeit, die die leistungsstärksten SchülerInnen aufwenden. Mitte 3. Klasse benötigen die unteren 15% mindestens 18 sek, Mitte 4. Klasse mindestens 11 sek. Allerdings gibt es immer noch Kinder, die nur Zufallstreffer landen. Damit nimmt die Streuung *innerhalb* der Teilgruppe der unteren 15% von Jahr zu Jahr zu.

Die schwächsten 5% bearbeiten auch am Ende der Grundschulzeit noch jeden siebten Satz falsch und sie brauchen zusätzlich drei- bis viermal so viel Zeit wie die leistungsstärksten 5% .

[34] ➔ Tab. 3.2a, 3.2b, 3.2c, 3.2d

[35] ➔ Tab. 3.2a, 3.2b, 3.2c, 3.2d

4 Die Entwicklung der Leistungen im unteren Bereich

Die durchschnittlichen Leistungen und erst recht das Niveau der Spitzengruppe werfen nach den bisher berichteten Befunden keine besonderen Probleme auf. Genauer zu prüfen sind aber die Ergebnisse im unteren Leistungsbereich. Damit stellt sich die Frage nach den sog. „Mindeststandards“. Konkret: wie viele Sätze muss ein Kind am Ende der Grundschulzeit pro Minute richtig bearbeiten, damit man von einer „tragfähigen Grundlage“ für seine weitere Entwicklung im Lesen sprechen kann?

4.1 Die Leistungen der zwei, drei schwächsten SchülerInnen pro Klasse

Solange es keine Längsschnitte über die Entwicklung von SchülerInnen auf verschiedenen Niveaus in den anschließenden Schuljahren gibt, kann jede Entscheidung zwischen verschiedenen Leistungsanforderungen am Ende der Grundschule nur Plausibilitätscharakter beanspruchen. Dies umso mehr, als aus der Forschung sehr unterschiedliche Entwicklungsverläufe des Schriftspracherwerbs bekannt sind (vgl. u. a. Brinkmann 1997).

Damit sich die LeserInnen ein eigenes Urteil bilden können, berichten wir im Folgenden Daten zu vier verschiedenen Kriterien, die unterschiedlich hohe Anforderungen an eine „grundlegende Lesefähigkeit“ definieren [36]:

0-2 Sätze/ min. kann nicht lesen (30-60% falsch, richtige Lösungen allenfalls durch Zufallstreffer bei einer Ratewahrscheinlichkeit von durchschnittlich 10-15% pro Satz);

2-3 Sätze/ min. kann Sätze zwar eigenständig, aber nur sehr langsam (20-30 sek pro Satz) und nur mit großen Schwierigkeiten (etwa 20-30% Fehler) erlesen;

3-4 Sätze/ min. kann Sätze erlesen, aber immer noch langsam (15 sek) und mühevoll (im Durchschnitt noch mehr als 10% Fehler);

4-6 Sätze/ min. kann Sätze zunehmend fehlerfrei (6-8%) und zügiger, aber auch immer noch relativ langsam (etwa 12 sek pro Satz) erlesen.

In der folgenden Tabelle haben wir deshalb die Anteile der Jahrgänge in allen vier Kategorien ausgewiesen, so dass man sich je nach Maßstab ein eigenes Bild vom Niveau der SchülerInnen im unteren Leistungsbereich machen kann [37] :

[36] Die genau definierten Schwellen sind: 0-1,9; 2,0-2,9; 3;0-3,9; 4,0-5,9 richtige Sätze pro Minute (vgl. Tab 4.1d-f zur Variable „Niveau“ im Anhang).

[37] Vgl. ergänzend die Auswertung nach vier Niveaus für die Gesamtgruppe in [s:stolper.03.desc niveaus by stufkorr].

Tab. 4.1a: Jahrgangsvergleich der Anteile (Rohdaten) im unteren Leistungsbereich				
richtige Sätze pro Minute Jahrgang	0-2	0-3 [38]	0-4	0-6
2. Klasse	14.8 %	32.1 %	51.4 %	80.5 %
3. Klasse	1.5 %	6.8 %	17.4 %	53.0 %
4. Klasse	0.1 %	1.2 %	4.0 %	20.3 %

Diese Übersicht vermittelt auf den ersten Blick ein sehr positives Bild: Die Anteile in den unteren Leistungsgruppen nehmen von Jahrgang zu Jahrgang rapide ab. Die Gruppe der Leseunfähigen liegt schon Mitte 3. Klasse, bei einem härteren Maßstab Mitte 4. Klasse unter 2% .

Allerdings ist daran zu erinnern, dass wir es nicht mit einem echten Längsschnitt zu tun haben. So können wir nicht davon ausgehen, dass die drei Querschnitte in ihrer Zusammensetzung identisch sind. Da es sich um Erhebungen in denselben Schulbezirken handelt, können wir zwar davon ausgehen, dass die Zusammensetzung der Schülerschaft auf den verschiedenen Jahrgangsstufen vergleichbar ist. Wir haben aber beim Vergleich der unteren Leistungsbereiche selektive Abgänge durch Sitzenbleiben und durch den Wechsel auf Sonderschulen zu berücksichtigen. Zur Überprüfung dieser Unsicherheit ist unbedingt eine echte Längsschnittstudie erforderlich, die in Teilstichproben auch mit vergleichsweise geringem Aufwand möglich wäre.

Vorläufig kann man mit folgenden Schätzwerten arbeiten: pro Jahr bleiben in NRW in der Grundschule etwa 1.5% der SchülerInnen sitzen (ddp 2003). Da es sich dabei um besonders leistungsschwache Kinder handelt, muss statistisch gesehen der Anteil der leistungsschwachen Gruppe in Klasse 2 gegenüber Klasse 4 um bis zu 3% höher angesetzt werden. Zusätzlich wechseln in der Grundschulzeit etwa 2% in die Schulen für Lernbehinderte, so dass insgesamt rund 5% „Dropouts“ in den unteren Leistungsgruppen zu vermuten sind und die dortigen Werte entsprechend zu erhöhen sind, um einen fairen Längsschnitt zu konstruieren.

Die Verteilung der Korrekturwerte auf die Kategorien im unteren Leistungsbereich muss nach persönlicher Einschätzung erfolgen. Sie ist so vorgenommen worden, dass sie das Ausmaß der Leseschwierigkeiten eher über- als unterschätzt [39] :

[38] Vgl. die Tabellen 4.1g-i im Anhang. Die Werte sind hier kumulativ berechnet, die höheren schließen also die jeweils vorhergehende Gruppe mit ein.

[39] In die beiden niedrigsten Gruppen wurden pro Jahrgang jeweils 1% dazu berechnet, in der dritten noch einmal 0.5% . Daraus ergeben sich kumulativ plus 2.5% pro Jahrgang = plus 5% in Klasse 4 in der Gruppe 0-6.

Tab. 4.1b: Jahrgangsvergleich der Anteile (korrigiert) im unteren Leistungsbereich				
richtige Sätze pro Minute Jahrgang	0-2	0-3 [40]	0-4	0-6
2. Klasse	14.8 %	32.1 %	51.4 %	80.5 %
3. Klasse	2.5 %	8.8 %	19.9 %	55.5 %
4. Klasse	2.1 %	5.2 %	9.0 %	29.3 %

In dieser Übersicht wird das Bild differenzierter. Zum einen ist immer noch erkennbar, dass der Anteil der leistungsschwachen SchülerInnen deutlich abnimmt, in den untersten Kategorien schon von Klasse 2 auf 3, bei schärfer formulierten Anforderungen (nicht mehr als 12 sek pro Satz) auch bis Klasse 4. Gleichzeitig ist aber festzustellen, dass im untersten Bereich (0-2) von Klasse 3 auf 4 mit einer Stagnation zu rechnen ist, die für die betroffenen Kinder erhebliche Schwierigkeiten in der Sekundarstufe befürchten lässt.

Einen etwas anderen Zugang zu Entwicklung im unteren Leistungsbereich verschafft die folgende Übersicht, in der die Daten nicht nach *inhaltlichen* Leistungs*stufen*, sondern nach *prozentual* aus der Stichprobe definierten Leistungs*gruppen* (Prozentränge = PZR) aufgegliedert werden [41] . Dabei beschränkt sich die Darstellung auf die unteren 30%:

Tab. 4.1c: Jahrgangsvergleich der Leistungswerte im unteren Leistungsbereich (30%)									
Richtige Sätze pro Minute Jahrgang	Durch-schnitt	PZR 1 - 5	PZR 6 -10	PZR 11-15	PZR 16-20	PZR 21-25	PZR 26-30	....	PZR 95-100
2. Klasse	4.1	.0-0.9	1.0-1.5	1.6-1.8	2.0-2.2	2.2-2.5	2.6-2.8		8.3-10.5
3. Klasse	6.1	.0-2.6	2.8-3.2	3.4-3.6	3.8-4.0	4.2-4.3	4.4-4.6		10.4-12.0
4. Klasse	8.1	.5-4.0	4.2-4.8	5.0-5.2	5.5-5.7	6.0-6.2	6.3-6.5		13.0-15.5
absolute Zuwächse	+ 4.0	+ 2.0	+ 2.8	+ 3.0	+ 3.2	+ 3.4			+ 4.8
proport. Zuwächse [42]	1:2.0	1:2.7	1:2.6	1:2.4	1:2.6	1: 2.3			1 :1.5

[40] Die Werte sind kumulativ berechnet, schließen also die jeweils vorhergehende Gruppe mit ein.

[41] Vgl. die Rohtabellen 3.2e, g, i im Anhang.

[42] Berechnet als Differenz zwischen den Werten zweier Jahrgänge, dividiert durch den Ausgangswert

Betrachtet man die Durchschnitte, sieht man einen deutlichen Leistungszuwachs von Klasse 2 über 3 nach 4. Analog gibt es Fortschritte in den einzelnen Prozentrang-Gruppen – proportional am deutlichsten in den unteren Prozenträngen, absolut gesehen stärker in den oberen.

Zwei Einschränkungen sind allerdings zu beachten:

- In der untersten Leistungsgruppe gibt es einerseits auf allen Klassenstufen SchülerInnen, die kaum einen Satz richtig bearbeiten, andererseits öffnet sich hier das Leistungsspektrum nach oben (von 0.0-0.9 auf 0.5-3.8 – was in der zweiten Klassen immerhin einem Prozentrang von 50, also einer durchschnittlichen Leistung entsprochen hätte). In allen anderen Teilgruppen verdoppelt oder verdreifacht sich die Leistung.

- Andererseits kann dieser vertikale Vergleich in die Irre führen, da wir ja einen Abgang von 5% durch Sitzenbleiben und Wechsel auf die Sonderschule berücksichtigen müssen. Unterstellt man, dass es sich bei diesen SchülerInnen in der Regel auch im Lesen um die leistungsschwächsten Kinder handelt, muss man die Werte diagonal vergleichen [43], wie die parallele Färbung verdeutlicht.

Nach dieser Korrektur ergeben sich (von links nach rechts aufgelistet) *innerhalb des unteren Leistungsdrittels* absolute Zuwächse von 2.0 in der schwächsten Gruppe bis 3.4 in der stärksten Gruppe. Berechnet man die Zuwächse dagegen proportional, erhält man die in der untersten Zeile angegebenen Relationen von durchschnittlich 1:2.5 .

Zum Vergleich sind in der rechten Spalte die Zuwächse der *leistungsstärksten 5%* dargestellt: In absoluten Zahlen ist der Zuwachs höher, proportional gesehen geringer. Der Abstand von Prozentrang 95 zu Prozentrang 5 [44] beträgt in Klasse 2 rund 7 richtige Sätze pro Minute, in Klasse 4 (korrigiert) dagegen rund 9-10 richtige Sätzen pro Minute (s. Tab. 3.2b).

Aber: Alle Gruppen machen Fortschritte [45] , und proportional am stärksten die unteren Leistungsgruppen. Das bedeutet: Das Verhältnis der Leistungen der oberen zu den unteren 15% beträgt in Klasse 2 noch 1:9.4, in Klasse 4 dagegen nur noch 1:3.8 .

Aus der Perspektive der einzelnen Lehrperson „bleiben die Schwachen schwach“. Durch den Vergleich allein mit der Bezugsgruppe, die sich ja auch weiter entwickelt, werden aber die individuellen Fortschritte unterbewertet. Das nennen wir den „Karawanen-Effekt“ des Lernens in Klassen [46]. Die unteren 5% oder 15% einer Gruppe sind definitionsgemäß *immer* schlechter als

---

[43] Diese Konzentration aller Abgänge im Prozentrangbereich 1-5 ist die für meine im Folgenden entwickelten Argumentation ungünstigste Variante, die ich gewählt habe, um meine Folgerungen möglichst gut abzusichern.

[44] Das entspricht in etwa dem Vergleich des leistungsstärksten mit dem leistungsschwächsten Schüler einer Durchschnittsklasse, s. oben ➔ 3.2.

[45] Auch hier ist allerdings wieder zu betonen: Diese Folgerung hat solange nur hypothetischen Charakter, wie sie nicht durch entsprechende Belege aus einem echten Längsschnitt abgesichert ist. Erste Längsschnittdaten aus einer Teilstichprobe von 198 Kindern sprechen allerdings mit Korrelationen von durchschnittlich .83** pro Klasse schon jetzt dafür, dass sich Ränge der einzelnen Kinder kaum verändern (s. im einzelnen Tab. 4.1j).

[46] Analog hat May (1995) für die Rechtschreibentwicklung festgestellt, dass sich die Fortschritte der leistungsstarken und der leistungsschwachen SchülerInnen auch qualitativ gleichen – wenn man sie auf die jeweilige Aus-

der Durchschnitt. Mit der Fixierung des Blicks auf ihren Platz in der Bezugsgruppe wird aber übersehen, dass *alle* SchülerInnen von Jahr zu Jahr Fortschritte machen. Pädagogisch gesehen sind die *Fortschritte* jeder *Teil*gruppe bedeutsamer als die *Abstände* innerhalb der *Gesamt*-gruppe. Konkret: Wer in der vierten Klasse auf Prozentrang 15 pro Minute 5.2 Sätze schafft, hat sich gegenüber Klasse 2 um 3.4 Sätze verbessert – sein Abstand zum Durchschnitt beträgt aber nur 2.9 Sätze. Die pädagogische Frage ist, welcher Bezugspunkt bei der Leistungsbewertung dominieren sollte.

Bei einer Karawane verwundert es niemanden, wenn der, der zuletzt auf die Reise gegangen ist, auch als letzter ankommt. Bedeutsamer ist der Weg, den die Karawane als *ganze* geschafft hat. Man muss bedenken, dass sich schon die Schulanfänger in ihren schriftsprachlichen Voraussetzungen bis zu drei, vier Jahren durchschnittlicher Entwicklung unterscheiden. Dann überrascht der Unterschied von rund drei Entwicklungsjahren in der Leistung eines Viertklässlers auf Prozentrang 10 (4.9 Sätze pro Minute; korrigiert um die Abgänge: 4.0) und einem auf Prozentrang 90 (11.7 Sätze pro Minute) nicht. Damit stellt sich aber die Frage, welchen Sinn die aktuell diskutierten „Bildungsstandards“ machen, wenn sie als verbindliche Niveaus für alle formuliert werden.

Mit diesem Hinweise sollen keinesfalls die in der untersten Leistungsgruppe beobachteten Schwierigkeiten verdrängt werden. Auch wenn die Größenordnung von rund 5% besonders gefährdeten SchülerInnen [47] nicht mit den bei PISA berichteten 25% „leseschwachen“ SchülerInnen zu vergleichen ist, stimmt die Art der Probleme besorgt, da es sich beim Stolperwörter-Lesetest anders als bei IGLU (und erst recht bei PISA) um grundlegende Teilleistungen handelt.

Ebenso deutlich muss man andererseits sagen, dass Maßnahmen zur Behebung dieses Problems nicht Inhalt und Stil des Unterricht für die übrigen 95% bestimmen dürfen [48]. Diese Warnung ist auch deshalb wichtig, weil sich die in PISA in einer größeren Breite, d. h. bei 25% der SchülerInnen, beobachteten Schwächen auf das Textverständnis beziehen. Dass die Kinder der Grundschule schon in dieser Hinsicht relativ besser abschneiden, haben die IGLU-Ergebnisse gezeigt (vgl. Kasten in Kap. 1).

4.2 Die Leistungsverteilung in Schulen in sozialen Brennpunkten

Da unsere drei Stichproben aus eher ländlichen und kleinstädtischen Regionen stammen, haben wir ergänzend eine großstädtische Sonderstichprobe von 25 vierten Klassen aus dem Köl-

---

gangsposition bezieht. Am Vergleich von deutschsprachigen SchülerInnen und Kindern anderer Muttersprache können wir zeigen, dass sich dieses Muster auch in der Sekundarstufe zeigt (Sekundärauswertung der NRW-Kids-2001-Studie, Brügelmann 2003, i.V.).

[47] Auf die Altersgruppe bezogen: plus vermutlich weiteren 5%, die den Jahrgang oder die Grundschule verlassen haben.

[48] Mehr noch: Es kann aus unseren Ergebnissen auch für die gefährdete Teilgruppe der unteren 5% nicht geschlossen werden, eine effektive Förderung müsse sich auf einfache Teilleistungen konzentrieren. Intensivstudien (vgl. Peschel 2002) zeigen, dass die Förderung anspruchsvollen Lesens, also der selbstständige Umgang mit Texten vom ersten Schuljahr an, gerade bei schwachen SchülerInnen ein hohes Niveau auch in den Basisleistungen sichern kann.

ner Raum einbezogen [49]. Bei dieser Sonderstichprobe handelt es sich um Schulen in sog. sozialen „Brennpunkten“, deren SchülerInnen in der Regel ungünstigere Lernvoraussetzungen und –bedingungen mitbringen und die häufig auch während der Schulzeit zusätzliche Belastungen zu verarbeiten haben.

Im Februar und März 2003, also einen Monat nach unserer Hauptstichprobe, wurden die Daten in 8 Schulen mit insgesamt 539 Kindern erhoben. Diese Teilstichprobe weicht in zwei zentralen Merkmalen tatsächlich von unserer Hauptstichprobe ab:

- der Anteil der Kinder von Alleinerziehenden ist doppelt so hoch (14.8% statt 7.3% sonst);
- der Anteil der Kinder mit anderer Muttersprache ist dreimal so hoch (64.2% statt 21.3% sonst) .

In den Durchschnittsleistungen (richtig bearbeitete Sätze) schneiden diese Klassen insgesamt schwächer ab:

- arithmetisches Mittel 7.6 [50] (SD 2.5) vs. 8.7 (SD 2.7) in der Hauptstichprobe;

Anders als nach den Umfeldbedingungen erwartet sind allerdings die Leistungen im unteren Bereich nur wenig schlechter:

- 7.0 % der Kinder unter Prozentrang 5 der Gesamtstichprobe;
- 1.6 % haben weniger als 3 Sätze pro Minute richtig (vs. 1.2% in der Gesamtstichprobe);

Insbesondere ist fest zu halten:

- Die deutschsprachigen Kinder schneiden um weniger als 10% schlechter ab als in der Hauptstichprobe (8.1 vs. 8.7 richtige Sätze pro Minute), obwohl sie aus einem sozialen Brennpunkt kommen und in den Klassen mit einem Anteil von einem Drittel (gegenüber mehr als drei Vierteln in der Hauptstichprobe) die Minderheit darstellen.
- die Kinder anderer Muttersprache schneiden unter diesen Bedingungen mit durchschnittlich 7.2 richtigen Sätzen (SD 2.5; SE 0.15) sogar geringfügig besser ab als sonst (6.9 richtige Sätze [SD 2.4; SE 0.08] in der Hauptstichprobe[51]).

---

[49] Ich danke den KollegInnen SAD’in Monika Baum und SAD Hans Wielpütz vom Schulamt der Stadt Köln für ihre rasche und unbürokratische Hilfe.

[50] Die Kölner Stichprobe war dabei noch insofern benachteiligt, als in einem Teil der Klassen versehentlich der Bogen für erste Schuljahre eingesetzt wurde, der nur 45 statt 60 Sätze enthält, so dass die Spitzengruppe ihr Potenzial gar nicht ausschöpfen konnte. Belegt wird diese Annahme durch den sprunghaften Anstieg der Schülerzahl bei 45 bearbeiteten Sätzen: 13.7% erreichten dieses Niveau (gegenüber 1.7% bei 44 bearbeiteten Sätzen). Und 6.2% bearbeiteten diese 45 Sätze auch richtig, weitere 4.7% noch 44 Sätze. Die Teilgruppe (N = 74), die den 60er Bogen bearbeitet hat, erreicht in der Tat 8.3 Punkte (gegenüber 8.0 in der Gruppe mit den 45er Bögen [N = 124]).

[51] Auch eine Teilauswertung auf Klassenebene zeigt (Stand der Auswertung: 4.4.2003) : Linear betrachtet verändert sich in der Gesamtstichprobe (einschließlich der Kölner Sonderstichprobe, zur Zeit zusammen N = 136 Klassen) die Zahl richtiger Sätze weder in der Gruppe mit anderer Muttersprache in Abhängigkeit von ihrem Anteil an der Kinderzahl in der Klasse (r = um .10) noch in der Gruppe der deutschsprachigen SchülerInnen (r = um .00). Ein

Der schlechtere Gesamtwert dieser Sonderstichprobe ergibt sich also nicht aus schlechteren Leistungen ihrer Teilgruppen. Er ist vielmehr rein rechnerisch bedingt und resultiert daraus, dass die in *beiden* Stichproben schwächer abschneidenden Migrantenkinder in den Brennpunkt-Klassen einen deutlich höheren Anteil darstellen (ausführlicher Kap. 5.3). Bedenkt man die zusätzlichen sozio-ökonomische Belastungen in den sog. „Brennpunkten“, fällt selbst der Unterschied in der deutschsprachigen Gruppe erstaunlich gering aus.

Diese Befunde sind ein wichtiger Hinweis darauf, dass es einer Reihe von Grundschulen gelingt, Kindern mit anderer Muttersprache auch unter schwierigen Bedingungen gerecht zu werden. Dabei könnten objektive Rahmenbedingungen (Klassengröße [52], Anzahl der Förderstunden), aber auch pädagogische und didaktische Konzepte der LehrerInnen eine bedeutsame Rolle spielen.

höherer Anteil von Migrantenkindern hat also wider Erwarten auf beide Gruppen keinen negativen Einfluss (teilweise sogar einen positiven, s. die Klassifikationsanalyse in Kap. 5.3).

[52] Im Mittel hatten die vierten Klassen in der Kölner Stichprobe 21.5 SchülerInnen, in den anderen Bezirken mit 22.4 kaum mehr. Die Zahl richtiger Sätze ist im Übrigen sowohl bei den deutschsprachigen Schülern als auch bei den Kindern mit anderer Muttersprache unabhängig von der Klassengröße (r = um .00 bis -.10).

<u>Fazit 4 zu Kap.</u>:

Aus den bisher berichteten Befunden folgt: der Abstand zwischen leistungsschwachen und leistungsstarken SchülerInnen verringert sich von Klassenstufe zu Klassenstufe deutlich – von etwa 1:8 in Klasse 2 auf 1:3 in Klasse 3 und Klasse 4.

Dieser Befund darf aber nicht als „Aufhol"-Effekt missverstanden werden, da rund 5% der Zweitklässler durch Nichtversetzung oder Wechsel in die Sonderschule den Anteil leistungsschwacher LeserInnen verringern. Berücksichtigt man diese rein „statistische Verbesserung" speziell der untersten Teilgruppen, so öffnet sich die Schere zwischen Leistungsstarken und Leistungsschwachen über die Grundschulzeit hinweg (sog. „Matthäus-Effekt": *Wer hat, dem wird gegeben*) [53].

Andererseits machen *alle* Gruppen deutliche Fortschritte. Bedenkt man, dass sich bereits Schulanfänger in ihren schriftsprachlichen Voraussetzungen um rund 3-4 Jahre durchschnittlicher Entwicklungszeit unterscheiden, überrascht nicht, dass auch Mitte 4. Klasse die leistungsstarken GrundschülerInnen auf Prozentrang 90 den leistungsschwachen auf Prozentrang 10 um rund drei Jahre durchschnittlicher Leseentwicklung voraus sind [54] . Wir interpretieren beide Befunde zusammen als „Karawanen-Effekt": Im Vergleich zum Durchschnitt bleiben die Leistungsschwachen am Ende der Gruppe – bezogen auf ihre je unterschiedlichen Voraussetzungen machen sie bedeutsame Fortschritte. Das gilt allerdings nicht für die leistungsschwächsten 5-10% [55]. Diese bleiben immer mehr zurück – vermutlich auch eine Folge der ständig entmutigenden Rückmeldung durch eine vergleichende Leistungsbewertung und fehlende Förderung von ihrem individuellen Entwicklungsstand aus.

[53] ➔ Tab. 4.1b

[54] ... sofern man Lernen nicht als einen mechanischen Prozess missversteht, in dem sich Entwicklungszeit beliebig verkürzen lässt, wenn man die Intensität des Unterrichts entsprechend steigert. Insbesondere ist zu bedenken, dass entwicklungsförderliche bzw. hinderliche Bedingungen auch parallel zur Schule wirken. Eine US-amerikanische Studie zeigt, dass leistungsstarke LeserInnen außerhalb der Schule rund 100-mal so viel Text lesen wie leistungsschwache (Anderson u. a. 1988). Diesen beiläufigen Übungsvorteil kann kein Förderunterricht ausgleichen. Vielmehr kommt es darauf an, auch die leseschwachen SchülerInnen „zum Lesen zu verlocken" – in ihrer Freizeit

[55] ➔ Tab. 4.1c

5 Muttersprache der Eltern

Die Muttersprache der Eltern ist ein besserer Indikator für abweichende Sprachvoraussetzungen als die früher erhobene Staatsangehörigkeit. Wie in anderen Studien (TIMSS, PISA, IGLU) auch fasst das Merkmal „Sprache der Eltern“ die Sprachlernsituation der Kinder spezifischer als rechtliche Kategorien. Auf diese Weise werden auch Kinder aus „deutschen“ Aussiedlerfamilien und ebenso eingebürgerte Kinder von ArbeitsmigrantInnen erfasst.

Andererseits differiert die konkrete Sprachpraxis (z. B. mono- oder bilingual; Trennung nach Familien- vs. Umweltsprache oder nach personenbezogenen Sprachen; Niveau der Mutter- und der deutschen Sprache) innerhalb der so erfassten Gruppen stark. Wie die Streuungen (SD) zeigen, ist die Heterogenität der Leseleistungen innerhalb der so definierten „Migrantenkinder“ dennoch nicht größer als unter den deutschsprachigen Kindern (vgl. Tab. 5.1b-d).

5.1 Grunddaten für die Klassenstufen 2 bis 4

Tab. 5.1a: Richtige Wörter pro Minute auf den verschiedenen Klassenstufen aufgegliedert nach Sprache der Eltern [56]

	Klasse 2	Klasse 3	Klasse 4
beide Eltern Muttersprache Deutsch [57]	4.3**	6.3**	8.5**
ein Elternteil andere Muttersprache [58]	2.9	5.1	7.5
beide Eltern andere Muttersprache	3.1	4.8	6.9**

Mitte 2. Klasse macht die Muttersprache der Eltern einen beträchtlichen Unterschied für die Lesefähigkeit aus: Kinder aus Elterhäusern, in denen eine andere Sprache gesprochen wird, schaffen mit 3.1 bzw. 2.9 vs. 4.3 deutlich weniger richtige Sätze als die deutschsprachigen Kinder (auch statistisch nicht nur ein statistisch signifikanter, sondern auch ein quantitativ bedeutsamer Unterschied: ES = .55 [59]).

Der Unterschied resultiert sowohl aus einem langsameren Lesetempo (4.1 vs. 5.0 Sätze) als auch aus einer höheren Fehlerquote (21% vs. 11% der bearbeiteten Sätze falsch).

---

[56] Auszug aus Tab. 5.1b-d im technischen Anhang, wo sich auch die statistischen Kennwerte zur Beurteilung der Unterschiede finden.

[57] 78-80% der Gesamtgruppe in den Jahrgängen 2 bis 4

[58] Die Gruppe der MigrantInnen ist in den Tabellen zwar detaillierter aufgegliedert, aber die Teilgruppe der Kinder mit nur einem Elternteil anderer Muttersprache macht je nach Jahrgang nicht mehr als 4-8% der Kinder mit Migrationshintergrund aus, so dass beide Teilgruppen in der Interpretation zusammengefasst werden, zumal sich ihre Leistungen nur in Klassenstufe 4 statistisch signifikant unterscheiden.

[59] Die Effektstärke ES wird berechnet aus der Differenz der Mittelwerte, geteilt durch die Streuung der Hauptstichprobe oder der Kontrollgruppe – hier der deutschsprachigen Kinder.

Rund 30% Differenz bedeuten einen deutlichen Abstand. Umgerechnet entsprechen 1.2 Sätze pro Minute einer durchschnittlichen Entwicklung von etwa einem halben Schuljahr.

Die Unterschiede sind ein Jahr später, also Mitte der **3. Klasse**, noch deutlicher (auch quantitativ bedeutsam: ES = .68). In absoluten Zahlen entsprechen sie weiterhin etwa einem halben Schuljahr, aber relativ sind sie von rund 30% auf etwa 20% geschrumpft:

In beiden Gruppen hat sich die Fehlerquote gegenüber den Zweitklässlern etwa halbiert und das Arbeitstempo deutlich beschleunigt.

Kinder mit Eltern, die eine andere Muttersprache sprechen, lesen auch am Ende der Grundschulzeit schlechter als deutschsprachig aufgewachsene Kinder: Sie brauchen Mitte der **4. Klasse weiterhin** gut 20% länger für einen richtigen Satz (ES = .59).

Auf den ersten Blick ist das Bild analog wie bei den leseschwächeren Kindern insgesamt: Auch die Migrantenkinder machen deutliche Fortschritte, aber da die deutschsprachigen Kinder ebenfalls besser werden, bleibt der Abstand erhalten.

5.2 Leistungsvergleich im unteren Bereich in Abhängigkeit von der Muttersprache

Es könnte aber sein, dass dieser Eindruck – analog zu Kap. 4.1 – trügt. Der Anteil der Kinder mit Eltern, die eine andere Sprache sprechen, sinkt von 22.1% in Klasse 2 bzw. 22.7% in Klasse 3 auf 20.0% in Klasse 4. Das bedeutet: Die Gruppe der Kinder mit anderer Muttersprache ist in Klasse 4 um rund 10% kleiner als in Klasse 2. Geht man davon aus, dass diese Differenz komplett durch einen Wechsel in niedrigere Klassen oder in Sonderschulen, also durch unzureichende Schulleistungen – dann vermutlich gerade auch im sprachlichen Bereich – zu erklären ist, müsste man am Ende der Grundschulzeit den Anteil auf den unteren Lesestufen um rund 9.1% [60] erhöhen. Da im Grundschulbereich etwa 7% der ausländischen Kinder in Sonderschulen sind (vgl. BMB+F 2001, 80), erscheint eine Quote von rund 10% für Nichtversetzungen und Überweisungen in Sonderschulen durchaus realistisch [61]. Für deutschsprachige SchülerInnen müsste nach analoger Berechnung ein Wert von rund 4 % angesetzt werden.

Das bedeutet, dass wir in den unteren Leistungsbereichen für die deutschsprachigen Kinder in Klasse 4 zusammen um 4% höhere Anteile ansetzen müssen, als unsere Rohstatistik ausweist, für die Kinder anderer Muttersprache rund 10%. Die folgende Tabelle macht diesen Korrekturprozess durchsichtig [62] :

---

[60] 9.1% werden errechnet aus 10% : [100+10]

[61] Die Schätzung ist noch schwieriger als in dem (in Kap. 5.1) allgemein diskutierten Fall des Ausscheidens durch schlechte Schulleistungen, da einerseits Kinder anderer Muttersprache Deutschland wieder verlassen (ohne dass sie deshalb Leseschwierigkeiten haben müssten), andererseits Seiteneinsteiger (mit schlechteren Leistungen) nicht zu Lasten des (fehlenden) Fortschritts der Teilgruppe verrechnet werden dürfen. Hier kann nur ein echter Längsschnitt verlässliche Zahlen liefern.

[62] Auszüge aus den Tab. 5.2e-h im statistischen Anhang.

Tab. 5.2a: Jahrgangsvergleich von Kindern deutscher bzw. anderer Muttersprachen

richtige Sätze pro Minute [63] Jahrgang/ Muttersprache	PR 50	0-2	0-3 [64]	0-4	0-5
deutsch	4.0	14.3 %	30.7 %	50.4 %	68.1 %
2. Klasse					
andere	2.8	31.8 %	54.6 %	73.2 %	88.0 %
Rohwert	8.0	0.3 %	1.0 %	3.6 %	10.3 %
deutsch		+1.5	+2.5	+3.5	+4.0
korrigiert		1.8 %	3.5 %	7.1 %	14.3 %
4. Klasse [65]					
Rohwert	6.6	0.3 %	4.1 %	10.3 %	23.8 %
andere		+5.0	+7.0	+ 9.0	+10.0
korrigiert		5.3 %	11.1 %	19.3 %	33.8 %

Die Tabelle enthält eine schlechte und eine gute Nachricht.

Sie zeigt zum einen, dass die Kinder anderer Muttersprache in den unteren Bereichen sowohl in Klasse 2 als auch in Klasse 4 deutlich stärker vertreten sind als die deutschsprachigen SchülerInnen (je nach Teilgruppe mit Relationen von bis zu 2:1 in der zweiten Klasse und bis zu 3:1 in der vierten Klasse).

Sie zeigt zum anderen, dass in beiden Gruppen die Anteile in den unteren Niveaugruppen erheblich schrumpfen: in absoluten Zahlen *durchgängig* stärker bei den Kinder anderer Muttersprache [66] , in der relativen Abnahme *zum Teil* stärker bei den deutschsprachigen Kindern. Insofern lässt sich auch hier von einem „Karawanen-Effekt“ sprechen: Alle Gruppen machen Fortschritte – aber von ihrem jeweils unterschiedlichen Ausgangsniveau aus [67].

Die folgende Tabelle weist die Fortschritte der Teilgruppen konkret in Leistungswerten aus. In dieser Übersicht sind die Daten der Kinder anderer Muttersprache nach Prozentrang-Gruppen aufgegliedert, in den Zellen finden sich die Werte für die richtigen Sätze pro Minute. Da wir unterstellen, dass rund 10% der ZweitklässlerInnen anderer Muttersprache bis zur vierten Klasse aus den Gruppen ausgeschieden sind, vergleichen wir unterschiedliche Prozentrang-Gruppen, um den Zuwachs abzuschätzen: die farblich jeweils gleich markierten Prozentränge 1-5 in Klasse 4 mit Prozentrang 10-15 in Klasse 2, Prozentrang 6-10 in Klasse 4 mit Prozentrang 16-20 in

---

[63] s. Anm. 7

[64] Die Werte sind kumulativ berechnet, schließen also die jeweils vorhergehende Gruppe mit ein.

[65] Die zusätzlichen 4.0 % für die deutschsprachigen Kinder wurden auf die vier unteren Leistungsgruppen mit + 1.5, +1.0, + 1.0 und + 0.5 %-Punkten umgelegt, wobei sich die Werte von links nach rechts wegen der kumulativen Darstellung addieren. Für die Kinder anderer Muttersprache lauten die Korrekturwerte +5.0, +2.0, +2.0, + 1.0, ebenfalls von links nach rechts kumulativ ausgewiesen.

[66] z.B. in der Teilgruppe 0-2 von 31.8% auf 5.3% bei den Migrantenkindern vs. 14.3% auf 1.8% bei den deutschsprachigen SchülerInnen.

[67] S. dazu einschränkend Anm. 45. Für das erste Schuljahr kommt Schründer-Lenzen (2003) im Berliner Längsschnitt allerdings zu demselben Ergebnis.

Klasse 2 usw. Dieses Berechnungsverfahren unterschätzt die erreichten Lernfortschritte eher, als dass es sie überschätzt.

Tab. 5.2b: Zuwächse von Kindern anderer Muttersprache in verschiedenen Leistungsgruppen

Richtige Sätze pro Minute **Jahrgang**	Durch-schnitt PZR 50	PZR 1 - 5	PZR 6 -10	PZR 11-15	PZR 16-20	PZR 21-25	PZR 26-30	PZR 31-90	PZR 91-100
2. Klasse	2.8	0.0-0.5	0.6-0.8	1.0-1.2	1.3-1.5	1.5-1.7	1.7-2.0	...	5.8- 9.7
4. Klasse	6.6	0.5-3.2	3.5-4.0	4.1-4.5	4.5-4.7	4.8-5.0	5.1-5.5	...	10.2-15.5
absolute Zuwächse	+3.8	+ 0.7	+ 2.3	+ 2.7	+ 2.8				+ 5.1
proport. Zuwächse	1:2.4	1:1.6	1:2.6	1:2.7	1:2.5				1 :1.6

Im unteren Bereich sind die Zuwächse absolut gesehen geringer, aber proportional gesehen teilweise höher als im oberen Bereich. Im Vergleich mit den deutschsprachigen Kindern sind die Zuwächse im unteren Bereich absolut gesehen geringer, proportional gesehen etwas höher (abgesehen von der untersten Leistungsgruppe).

Tab. 5.2c: Zuwächse von Kindern deutscher Muttersprache in verschiedenen Leistungsgruppen

Richtige Sätze pro Minute **Jahrgang**	Durch-schnitt PZR 50	PZR [68] 1 - 5	PZR 6 -10	PZR 11-15	PZR 16-20	PZR 21-25	PZR 26-30	PZR 31-90	PZR 91-100
2. Klasse	4.0	0.0-1.1	1.2-1.7	1.8-2.1	2.2-2.3	2.5-2.7	2.8-3.0	...	7.5-10.5
4. Klasse	8.0	1.0-4.2	4.5-5.0	5.2-5.7	5.7-6.1	6.2-6.5	6.7-6.9	...	12.2-15.5
absolute Zuwächse	+4.0	+ 1.1	+ 2.8	+ 3.2	+ 3.3	+ 3.4			+ 4.8
proport. Zuwächse	1:2.0	1:1.8	1:2.4	1:2.4	1:2.3	1: 2.2			1 :1.5

Insgesamt ist die Verteilung der Kinder anderer Muttersprache um fünf bis fünfzehn Prozentränge nach unten verschoben, d. h. ein Kind anderer Muttersprache auf Prozentrang 20 bringt etwa die Leistung, die ein deutschsprachiges Kind auf Prozentrang 10 zeigt.

Es muss noch einmal betont werden, dass es sich bei dieser Berechnung von Lernzuwächsen aus Querschnittsdaten um Schätzungen handelt. Teilt man unsere Annahmen, so sieht man, dass die Entwicklung der unteren Leistungsgruppen in beiden Gruppen besorgt stimmen muss.

[68] Achtung: Da für die Gruppe der deutschsprachigen Kinder anders für die Kinder anderer Muttersprache eine Verlustquote von nur rund 4% unterstellt wird, sind die Vergleichswerte für die 2. vs. 4. Klasse farbig in 5er und nicht in 10er Schritten gestaffelt.

Bei den Kindern anderer Muttersprache ist dies rund ein Fünftel: neben den (vermutlich besonders schwachen) 10%, die eine Klasse wiederholen oder auf die Sonderschule überwiesen worden sind, weitere 5%, die nur sehr geringe, und noch einmal 5%, die jedenfalls absolut gesehen unterdurchschnittliche Fortschritte machen.

Bei den deutschsprachigen Kindern ist dies rund ein Zehntel: neben den (vermutlich besonders schwachen) knapp 5%, die eine Klasse wiederholen oder auf die Sonderschule überwiesen worden sind, weitere 5%, die zumindest absolut gesehen unterdurchschnittliche Fortschritte machen.

Im Übrigen gilt: Alle Kinder machen deutliche Fortschritte [69] . Dabei ist der proportionale Lernzuwachs höher in den unteren, der absolute höher in den oberen Leistungsgruppen. Damit öffnet sich die Leistungsschere in beiden Gruppen.

5.3 Durchschnittsleistungen von Klassen in Abhängigkeit vom Anteil der Kinder mit anderer Muttersprache

Aus PISA-3 wird in der Presse berichtet, die Leseleistungen würden bei einem Anteil von Kindern mit anderer Muttersprache über 20% „sprunghaft“ schwächer, ab 40% ergäben sich aber erstaunlicherweise keine weiteren Veränderungen. Schon der Urtext (Baumert u. a. 2003, 56) ist in seinen Aussagen missverständlich. Gesprochen wird dort von einem „Anteil von SchülerInnen und Schülern mit Migrationshintergrund in Schulen“, also nicht in der einzelnen *Lerngruppe*. Dennoch wird der unzulängliche „Umgang mit Heterogenität“ im *Unterricht* beklagt.

Nach der unserer Auswertung von 696 Klassen ergibt sich in LUST für die Grundschule ein etwas anderer Befund [70] . Zwar nimmt die durchschnittlich Leseleistung der Klassen mit wachsendem Anteil an Kindern anderer Muttersprache ab. Gliedert man die Leistungen aber nach Teilgruppen auf, differenziert sich das Bild (s. Tab. 5.3a).

Es sind – bezogen auf den wachsenden Anteil von Kindern anderer Muttersprache – keine linearen Veränderungen zu erkennen. Zwei Punkte lassen sich festhalten, wobei diese Aussagen vorläufig wegen der zum Teil zu geringen Zellenbesetzung im Extrembereich auf die Schwellenwerte 0-80% bzw. 0-60% Anteil von Migrantenkindern begrenzt werden müssen:

- *Deutschsprachige* Kinder schneiden in Klassen *ohne* Kinder anderer Muttersprache *nicht* besser ab als in Klassen mit Kindern anderer Muttersprache [71].

---

[69] Auch May (2000, 179-180) berichtet aus dem Hamburger PLUS-Projekt, dass Migrantenkinder vergleichbare Fortschritte machen (dort in der Rechtschreibung) wie die deutschsprachigen Kinder.

[70] Detaillierte statistische Kennwerte finden sich in Tab. 5.3b+c im statistischen Anhang. Faustformel: Die farbig markierten Werte liegen innerhalb desselben Bereichs, d. h. die Mittelwerte unterscheiden sich auf dem 5%-Niveau (= +/- 2 SE) statistisch nicht signifikant voneinander. Die abweichen Werte außerhalb des farbigen Bereiche sind wegen der geringen Zellenbesetzungen nicht verlässlich einzuschätzen.

[71] Lediglich in den Gruppen 60+ bzw. 80+ deutet sich eine leichte Verschlechterung an, die wegen der geringen Besetzung dieser Zellen zur Zeit aber nur als Tendenz vermerkt werden kann. Fasst man die drei Jahrgänge zusammen und nimmt man die Klassen *ohne* Migrantenkinder als Baseline, so beträgt die maximale Abweichung über alle Anteilsgruppen minus 0.5 Punkte, und zwar in der Gruppe 60+ (die Teilgruppe 80-100 ist wegen der gerin-

- Für die Leseleistung Kinder mit *anderer Muttersprache* macht es ebenfalls keinen Unterschied, wie groß ihr Anteil in der Klasse ist. Lediglich in der 3. Klassenstufe lässt sich tendenziell eine leichte lineare Verschlechterung ausmachen [72] .

Ein größerer Anteil von Kindern mit anderer Muttersprache bedeutet danach weder für diese Gruppe noch für die deutschsprachigen Kinder eine Gefährdung ihrer Leseentwicklung [73] .

Tab. 5.3a: Leseleistung [74] (= „richtige Wörter pro Minute“) von deutsch- und fremdsprachig aufgewachsenen Kindern in Abhängigkeit vom Anteil der Kinder mit anderer Muttersprache in den Klassen

%-Anteil von Kindern and Sprache	RI_MIN_D[75] 2. Klasse N = 212	RI_MIN_D 3. Klasse N = 236	RI_MIN_D 4. Klasse N = 247	RI_MIN_A[76] 2. Klasse N = 176	RI_MIN_A 3. Klasse N = 191	RI_MIN_A 4. Klasse N = 199
0	3.9	6.2	8.5	-	-	-
01 – 09	4.5	# 6.4	8.4	3.3	# 5.3	6.8
10 – 19	4.6	6.5	8.5	3.2	5.0	7.1
20 – 39	4.4	6.1	8.2	3.3	4.8	6.8
40 – 59	4.0	6.0	8.1	3.1	4.6	6.8
60 – 79	(3.8)	# 5.8	# 7.5	(3.2)	# 4.6	# 6.8
80 –100	(3.2)	(5.4)	(7.6)	(1.2)	(3.7)	(5.9)
40 –100	3.9	5.9	8.0	3.1	4.5	6.7

---

gen Zahl auch in der Zusammenfassung nicht zuverlässig bewertbar). Sollte sich diese Tendenz bei größeren Zahlen bestätigen, bedeutete aber auch sie nur einen Rückstand von etwa zwei bis drei Schulmonaten gegenüber den Klassen ohne jeden Migrantenanteil.

[72] Mittelt man auch bei den Migrantenkindern die Werte über die drei Jahrgänge hinweg, um Zufallsschwankungen auszugleichen, kommt man zu einer maximalen Abweichung von nur 0.3 Punkten in der Anteilsgruppe 40+. Und auch diese Differenz – sollte sie sich in größeren Stichproben bestätigen – entspräche nur ein bis zwei Schulmonaten Rückstand.

[73] So auch für die Schweiz Rüesch (1999, 43) und für eine bundesweite Auswertung von Erhebungen mit dem Stolperwörter-Test Metze (2003) – anders dagegen die ersten Daten aus dem Berliner Längsschnitt von Schründer-Lenzen (2003).
Gegen das gängige Vorurteil spricht auch der folgende Befund aus einer Studie in 14 hessischen Grundschulklassen, auf die mich mein Siegener Kollege *Jürgen Zinnecker* aufmerksam gemacht hat: Je mehr eingewanderte Kinder in einer Schulklasse waren, um so aufmerksamer waren die Schüler im Unterricht, denn: "...befinden sich viele oder auch sehr viele Kinder aus Migrantenfamilien in einer Klasse, so scheinen alle Kinder der Klasse, also deutsche wie nicht-deutsche, besonders zur Mitarbeit und damit auch tendenziell zu besseren Lernleistungen angeregt zu werden." (Walter 2001, 118) Im Hamburger BLK-Modellversuch „Elementare Schriftkultur“ fielen die Leistungen erst bei einem Anteil von mehr als 80% Kindern anderer Muttersprache ab – dann aber sehr deutlich (pers. Mitteilung von Mechthild Dehn per e-mail v. 4.4.03).

[74] RI_MIN_D : richtige Wörter pro Minute (deutschsprachige Kinder), RI_MIN_A : dto. (andere Muttersprache)

Alle Zellen sind mit 20+ Schulen besetzt, außer in den mit * markierten Zellen mit 10-19 Schulen. Eingeklammerte Werte deuten an, dass in dieser Zelle N < 5 ist. Alle anderen N sind > 9.

[75] S. Anm. 73.

[76] S. Anm. 73 .

Der Eindruck einer Benachteiligung in solchen Klassen entsteht fälschlich dadurch, dass die Kinder mit anderer Muttersprache im Durchschnitt schwächere Leseleistungen erbringen. Wenn ihr Anteil in einer Klasse ansteigt, müssen also auch die Leistungen der Gruppe insgesamt sinken, selbst wenn die beiden Teilgruppen (deutsch- und fremdsprachige Kinder) gleiche Leistungen wie sonst auch erbringen.

Die für die Anteilsgruppen 60+ vorsichtig angedeuteten Negativ-Tendenzen sind zusätzlich dadurch zu relativieren, dass Klassen mit hohem Migrantenanteil mehrheitlich in sozioökonomisch schwächeren Stadtteilen liegen, so dass selbst in den Extremgruppen eine schlechtere Durchschnittsleistung nicht ohne Weiteres und jedenfalls nicht allein dem Migrantenanteil zugerechnet werden kann.

Wie stark die Kategorie „andere Muttersprache“ die sehr unterschiedlichen Voraussetzungen der gemeinten SchülerInnen vereinfacht, wird deutlich, wenn man bedenkt, dass in 15-20% der Klassen (mit durchaus unterschiedlich hohen Anteilen von Kindern anderer Muttersprache) die deutschsprachigen Kinder sogar schlechtere Leseleistungen erbringen als ihre KlassenkameradInnen mit Migrationshintergrund...

<u>Fazit zu Kap. 5</u> :

Kinder, deren Eltern eine andere Muttersprache sprechen (rund 20%), schaffen zu allen Zeitpunkten weniger Sätze als die deutschsprachigen SchülerInnen. Aber auch in dieser Gruppe sind die Leistungen der ViertklässlerInnen doppelt so gut wie die der ZweitklässlerInnen [77].

Da Migranten-Kinder deutlich häufiger nicht versetzt oder auf eine Sonderschule überwiesen werden als deutschsprachige SchülerInnen, schönt der Querschnittsvergleich allerdings auch hier das Bild. Rechnet man eine Verringerung der unteren Leistungsgruppen um 10% ein, so bleiben rund 20% Kinder mit nur geringen (oder gar keinen?) Fortschritten – gegenüber rund 10% unter den deutschsprachigen Kindern [78]. In Klassen mit einem hohen Anteil an Migrantenkindern sind entgegen dem gängigen Vorurteil weder deren Leistungen noch die der deutschsprachigen Kinder schlechter.

Beachtung verdient schließlich der Befund, dass Migrantenkinder bei gleichem Testergebnis durchgängig schlechter beurteilt werden als deutschsprachig aufgewachsene Kinder – außer im untersten Leistungsbereich [79] .

[77] ➔ Tab. 5.1a, 5.1b, 5.1c

[78] ➔ Tab. 5.2a, 5.2b, 5.2c

[79] ➔ Tab. 9.4b

6 Mädchen und Jungen im Vergleich

6.1 Die Leistungsunterschiede im Überblick

Wie aus vielen anderen Studien bekannt (vgl. die Zusammenfassung bei Richter 1996), schneiden auch in unserer Untersuchung die Jungen im Vergleich mit den Mädchen beim Lesen schlechter ab: Sie brauchen mehr Zeit für einen richtig beurteilten Satz.

Tab. 6.1a [80] : Richtige Wörter pro Minute nach Geschlecht und Jahrgang

	Klasse 2	Klasse 3	Klasse 4
Mädchen	4.2 **	6.2 **	8.4 **
Jungen	4.0	5.9	7.8

Die auf dem 1%-Niveau (**) signifikanten Leistungsdifferenzen lassen sich wie folgt auffächern [81] :

- Jungen machen insbesondere in der zweiten Klasse mehr Fehler.
- Umgekehrt nehmen die Unterschiede im Arbeitstempo zum Ende der Grundschulzeit zu.

Da sich immer wieder die Frage stellt, ob die Differenzen zwischen Mädchen und Jungen bereits vor der Schule angelegt sind oder durch den Unterricht gefördert werden, ist es interessant, die beiden Gruppen über die verschiedenen Jahrgänge hinweg zu vergleichen.

6.2 Die Leistungsunterscheide nach Jahrgangsstufen

Im einzelnen ergibt sich auf den drei Jahrgangsstufen folgendes Bild:

Tab. 6.2a : Mittelwerte von Mädchen vs. Jungen in Klasse 2

	Bearbeitete Sätze pro Minute	Anteil Fehler / bearbeitete Aufgaben	Richtige Sätze pro Minute
Mädchen	4.9 **	11.5 % **	4.2 **
Jungen	4.8	14.0 %	4.0

Mitte der **2. Klasse** schneiden die Jungen im Gesamtergebnis nur wenig schwächer [82] ab. Sie bearbeiten fast genau so viele Sätze, machen aber rund 20% mehr Fehler als die Mädchen.

[80] Auszug aus den Tab. 6.1b-d im statistischen Anhang.

[81] S. im einzelnen Tab. 6.1b-d im Anhang und Kap. 6.2 .

Mitte der **3. Klasse** bearbeiten die Jungen weiterhin etwas weniger Sätze als die Mädchen, sie machen jetzt allerdings weniger als 20 % mehr Fehler. Die Fehlerquote hat in beiden Gruppen erheblich abgenommen und der Unterschied zwischen ihnen fällt auch absolut noch geringer [83] aus als in Klasse 2:

Tab. 6.2b : Mittelwerte von Mädchen vs. Jungen in Klasse 3			
	Bearbeitete Sätze pro Minute	Anteil Fehler / bearbeitete Aufgaben	Richtige Sätze pro Minute
Mädchen	6.7 **	5.8 % **	6.2 **
Jungen	6.5	6.9 %	5.9

Mitte der **4. Klasse** machen die Jungen nur noch 15 % mehr Fehler als die Mädchen, aber sie arbeiten langsamer. Auch im Gesamtergebnis sind ihnen die Mädchen damit etwas voraus [84] :

Tab. 6.2c : Mittelwerte von Mädchen vs. Jungen in Klasse 4			
	Bearbeitete Sätze pro Minute	Anteil Fehler / bearbeitete Aufgaben	Richtige Sätze pro Minute
Mädchen	9.0 **	4.2 % **	8.4 **
Jungen	8.5	4.8 %	7.8

Insgesamt überraschen die geringen Unterschiede [85] . Vermutlich lassen sie sich damit erklären, dass Mädchen bei Aufgaben unter Zeitdruck generell schwächere Leistungen zeigen als in reinen Power-Tests, während Jungen von der Speed-Bedingung weniger beeinflusst sind (vgl. zuletzt Ratzka 2003).

Eine klare Antwort auf unsere Ausgangsfrage, ob die Unterschiede schon in die Schule mitgebracht oder erst durch den Unterricht erzeugt werden, geben die Ergebnisse nicht.

Auf alle Fälle kann aber festgehalten werden, dass der (soziale) Faktor „Muttersprache“ wesentlich größere Unterschiede zur Folge hat, als der gemeinhin stärker biologisch interpretierte Faktor „Geschlecht“.

---

[82] Die Effektstärke von 0.10 weist auf einen auch statistisch wenig bedeutsamen Unterschied hin.

[83] Auch die Effektstärke beträgt nur = 0.14 .

[84] Aber die Effektstärke von 0.26 weist auch diesen Unterschied als noch geringfügig aus.

[85] Für die Zahl der richtigen Sätze pro Minute beträgt die Effektstärke selbst in der vierten Klasse nur 0.22 .

Fazit zu Kap. 6 :

Wie schon in vielen anderen Studien auch festgestellt, lesen die Jungen *etwas* langsamer (Mitte vierter Klasse 8.5 : 9.0 bearbeitete Sätze pro Min.) und sie machen etwas *mehr* Fehler (4.8% : 4.2%) [86]. Insgesamt überraschen aber die geringen Unterschiede. Vermutlich lassen sie sich damit erklären, dass Mädchen bei Aufgaben unter Zeitdruck (wie in unserem Test) generell schwächere Leistungen zeigen als sonst, während Jungen von dieser Testbedingung weniger beeinflusst sind. Die Frage, ob die Leistungsunterschiede schon in die Schule mitgebracht oder erst durch den Unterricht erzeugt werden, kann durch die Ergebnisse des Vergleichs der Klassenstufen 2 bis 4 nicht beantwortet werden.

Auf alle Fälle lässt sich aber festhalten, dass der (soziale) Faktor „Muttersprache" wesentlich größere Unterschiede in der Entwicklung der Leseleistung zur Folge hat als der zumindest teilweise biologisch interpretierte Faktor „Geschlecht".
Beachtung verdient schließlich der Befund, dass Mädchen bei gleichem Testergebnis durchgängig besser benotet werden als Jungen – außer im untersten Leistungsbereich [87] .

[86] ➔ Tab. 6.2a, 6.2b, 6.2c

[87] ➔ Tab. 9.4a

7 Familienkonstellation

Je nach Jahrgang leben 7-9% der Kinder unserer Stichprobe mit nur einem Elternteil zusammen. Innerhalb der Gruppe der allein erzogenen Kinder ist eine durchgängig höhere Streuung als in der Gesamtstichprobe zu erwarten, da sie nach anderen bildungsrelevanten Kriterien, z. B. Einkommen, in sich sehr heterogen ist

7.1 Die Ausgangsfrage

Immer wieder wird die Besorgnis geäußert, dass Kinder alleinerziehender Eltern in ihrer Entwicklung gefährdet seien. Wegen der Korrelation mit ungünstigen sozio-ökonomischen Bedingungen in einer bekanntermaßen starken Teilgruppe (vgl. Geißler 2002, 417 ff.) lässt sich in der Tat eine geringere Leistung auch der Gesamtgruppe erwarten.

Wir haben deshalb auch den Zusammenhang dieses Faktors mit der Leseleistung untersucht.

7.2 Die Befunde im einzelnen nach Jahrgängen gestaffelt

In unserer Stichprobe hat diese besondere Familienkonstellation nur eine geringe Bedeutung für die Leseleistung, auch wenn die Unterschiede im Hauptkriterium (richtige Sätze/min.) auf dem 1%-Nivau (**) statistisch signifikant sind.

Wie die Ergebnisse der **2. Klassen** zeigen [88], arbeiten die Kinder alleinerziehender Eltern etwas langsamer. Obwohl sie tendenziell weniger Fehler machen, schaffen sie im Ergebnis auch weniger richtige Sätze, allerdings ist der Unterschied gering [89].

Tab. 7.2a : Mittelwerte von Kindern alleinerziehender Eltern in Klasse 2

	Bearbeitete Sätze pro Minute	Anteil Fehler / bearbeitete Aufgaben	Richtige Sätze pro Minute
zwei Eltern	4.9 **	12.8 %	4.2 **
allein-erziehend	4.4	12.1 %	3.6

Mitte der **3. Klasse** sind die Unterschiede ebenfalls marginal [90] , obwohl die Kinder alleinerziehender Eltern jetzt in der Tendenz auch mehr Fehler machen [91] :

[88] S. Tab. 7.1a im statistischen Anhang

[89] Effektstärke = .22

[90] Effektstärke = .17

[91] S. Tab. 7.1b im statistischen Anhang

Tab. 7.2a : Mittelwerte von Kindern alleinerziehender Eltern in Klasse 3			
	Bearbeitete Sätze pro Minute	Anteil Fehler / bearbeitete Aufgaben	Richtige Sätze pro Minute
zwei Eltern	6.7 **	6.3 %	6.1 **
allein-erziehend	6.1	6.8 %	5.7

Die Ergebnisse **Mitte 4.** Klasse bestätigen die Befunde der vorergehenden Klassenstufen: Kinder allein erziehender Eltern brauchen am Ende der Grundschulzeit 5-10% länger für einen richtigen Satz [92].

Tab. 7.2c : Mittelwerte von Kindern alleinerziehender Eltern in Klasse 4			
	Bearbeitete Sätze pro Minute	Anteil Fehler / bearbeitete Aufgaben	Richtige Sätze pro Minute
zwei Eltern	8.8 **	4.5 %	8.2 **
allein-erziehend	8.1	4.6 %	7.6

Insgesamt zeigen sich also nur geringfügige Unterschiede – etwa in der Größenordnung wie zwischen Mädchen und Jungen und deutlich geringer als zwischen Kindern deutscher oder anderer Muttersprache. Überraschend ist aber, dass die Streuungen zu allen drei Zeitpunkten in der Gruppe der Alleinerziehenden gleich oder sogar *geringer* ist als in den Familien mit zwei Eltern. Hier hätte man wegen der breit streuenden sozio-ökonomischen Bedingungen [93] eher größere Unterschiede erwarten können.

[92] Mit einer Effektstärke von .22 bleibt der Unterschied wenig bedeutsam. Etwas höher sind die Effektstärken in der Untersuchung von Metze (2003) mit .29 bis .31 ausgefallen. Da die Metze-Stichprobe generell schwächer abschneidet, bietet sich als eine mögliche Erklärung an, dass dort anders als bei LUST „alleinerziehend“ stärker mit „sozio-ökonomisch benachteiligt“ korreliert (s. Anm. 88).

[93] Vgl. zur Korrelation mit wirtschaftlichen Notlagen in einer bedeutsamen Teilgruppe: Geißler 2002, 276 ff.

Fazit zu Kap. 7 :

Kinder, die mit nur einem Elternteil zusammen leben (7-9% in unserer Stichprobe), unterscheiden sich in ihren Leseleistungen nur wenig von anderen Kindern. Von der zweiten Klasse bis zur vierten Klasse schneiden sie nur geringfügig schwächer ab (7.9 zu 7.3 Sekunden pro richtigem Satz) [94].

[94] ➔ Tab. 7.2a, 7.2b, 7.2c

8 Unterschiede zwischen Bezirken und Klassen

8.1 Vergleich mit der METZE-Erhebung

Als erstes bietet sich ein Vergleich mit den von Metze (2003) berichteten Daten aus seiner über das Internet gewonnenen Stichprobe an.

Tab. 8.1a : Vergleich der Leseleistungen in zwei Stichproben		
METZE / LUST	Richtige Sätze pro Minute	absolute Zuwächse
Ende 1. Klasse	1.8	
Mitte 2. Klasse	4.2	
Ende 2. Klasse	4.2	+ 2.4
Mitte 3. Klasse	6.1	+ 1.9
Ende 3. Klasse	6.3	+ 2.1
Mitte 4. Klasse	8.1	+ 2.0
Ende 4. Klasse	7.8	+ 1.5

Dieses Ergebnis überrascht. Die LUST-Stichprobe erreicht jeweils ein halbes Jahr früher die Werte der METZE-Stichprobe. Von der Gewinnung der ausgewerteten Klassen her würde man eher das umgekehrte Ergebnis erwarten: LUST hat in drei Schulbezirken große Anteile der LehrerInnen gewinnen können, so dass die Erhebung fast flächendeckend war. In der METZE-Erhebung haben sich einzelne KollegInnen aus dem ganzen Bundesgebiet selbst gemeldet, so dass bei der geringeren Stichprobengröße der Effekt der Selbstselektion höher sein müsste [95] . Dieser führt wegen des besonderen Interesses und Engagements von LehrerInnen, die an solchen Erhebungen mitmachen, eher zu einer positiven Verzerrung von Stichproben.

Bei ansonsten gleichartigen Testbedingungen und Auswertungsformen ist die einzige plausible Erklärung für einen negativen Selektionseffekt, dass sich vor allem LehrerInnen, die unter schwierigen Bedingungen arbeiten oder die besonders unsicher sind, das Angebot des Selbsttests angenommen haben.

8.2 Unterschiede zwischen Schulamtsbezirken

Die folgende Tabelle schlüsselt die Leistungen der drei Jahrgänge nach den drei an der Untersuchung beteiligten Schulbezirken auf:

Tab. 8.2a : Leseleistungen in drei Schulbezirken im Vergleich [96]			
Richtige Wörter pro Minute	2. Klasse	3. Klasse	4. Klasse
Rhein-Sieg-Kreis	4.2 **	6.3 *	8.4 **
Kreis Siegen-Wittgenstein	4.5	6.2	8.8
Märkischer Kreis	3.7 **	5.6 **	7.4 **

[95] Mit N = 1.536 (1. Klasse), N = 1.708 (2. Klasse), N = 1.705 (3. Klasse) und N = 1.570 (4. Klasse) ist die Stichprobe auch in absoluten Zahlen deutlich kleiner als die entsprechenden LUST-Untergruppen mit 3.206 bis 3.688 Kindern pro Jahrgang.

[96] Die genaueren Daten finden sich in den Tab. 8.2b-d im statistischen Anhang

In der Gesamttendenz schneidet die Klassen aus dem Bezirk Siegen-Wittgenstein sowohl beim Lesetempo als auch bei der Fehlerquote und damit auch bei der Zahl richtiger Sätze pro Minute am besten, die Klassen aus dem Märkischen Kreis am schwächsten ab.

Dieser Vergleich ist allerdings nur mit großen Vorbehalten zu interpretieren, da die Beteiligung in den drei Bezirken für die Monate Januar/ Februar sehr unterschiedlich ausgefallen ist [97] . Insbesondere die vom Muster der 2. und 4. Klassen abweichenden Ergebnisse im 3. Jahrgang (Rhein-Sieg-Kreis vor Siegen-Wittgenstein) lassen Verzerrungen in den Teilstichproben vermuten.

Wesentlich deutlicher, zum Teil schon dramatisch sind auf allen Stufen die Unterschiede zwischen den *Klassen* eines Jahrgangs.

8.3 Unterschiede zwischen einzelnen Klassen

Im einzelnen streuen die Extreme der Durchschnittswerte einzelner Klassen für die Zahl richtig korrigierter Sätze pro Minute erheblich [98]

Tab. 8.3a : Streuung der Durchschnittsleistungen zwischen ganzen Klassen

Jahrgang	richtige Sätze pro Minute
2	2.0 bis 7.1
3	4.0 bis 9.3
4	5.8 bis 12.9

Die Übersicht zeigt: Die schwächste vierte Klasse schneidet schlechter ab als die beste zweite Klasse. Dieser Befund deckt sich mit den Ergebnissen von Metze (2003). Auch er stellt fest, dass die fünf seiner 81 zweiten Klassen besser abschneiden als die schwächste vierte Klasse.

Solche schlagzeilenträchtigen Relationen sind aber nur vernünftig interpretierbar, wenn man die Voraussetzungen der SchülerInnen der betroffenen Klassen am Schulanfang und wenn man die Unterrichtsbedingungen während der Schulzeit kennt. Selbst dann dürfte nach aller Erfahrung noch ein bedeutsamer Teil der Leistungsunterschiede auf Unterschiede der Unterrichtsgestaltung zurückzuführen sein. Hier müssten weiterführende Studien ansetzen, die anders als

---

[97] Im Kreis Siegen-Wittgenstein ist der Rücklauf insgesamt etwas geringer ausgefallen als im Rehin-Sieg-Kreis (4.039 von 9.759 SchülerInnen = 41.4% gegenüber 9.821 = 48.1% von 20.413 SchülerInnen im Rhein-Sieg-Kreis), aber deutlich höher als im Märkischen Kreis (6.870 = 33.3 % von 20.614 SchülerInnen). Rückläufe aus dem Dezember bzw. aus März/ April konnten in diesen – zeitlich auf die Schuljahresmitte begrenzten – Vergleich nicht mit einbezogen werden. Im einzelnen haben im Januar/ Februar aus dem Rhein-Sieg-Kreis 8.310 SchülerInnen teilgenommen, aus Siegen-Wittgenstein 3.131 und aus dem Märkischen Kreis 5.856 .

[98] Stand der Zusatzauswertungen auf Klassenebene: 4.4.2002. Aktuelle Rohtabelle wird nach Zusatzauswertung im Anhang noch eingefügt.

LUST, IGLU und PISA konkrete Arrangements von Unterricht und ihre Wahrnehmung durch die Kinder untersuchen. Wir brauchen insbesondere ethnografische Studien, die auch die Sicht der Kinder repräsentieren, um besser zu verstehen, was verschiedene Aufgaben und Lernbedingungen für einzelne Kinder(gruppen) bedeuten (vgl. für den Schriftspracherwerb z. B. Panagiotopoulou 2002).

<u>Fazit zu Kap. 8</u> :

Zum Teil dramatisch unterscheiden sich die Mittelwerte der einzelnen Klassen: Die beste Klasse 2 ist leistungsstärker als die schwächste Klasse 4 [99]. Ehe man ein kritisches Urteil über den Unterricht in den betreffenden Klassen fällt, müsste man allerdings die Lernvoraussetzungen der Kinder am Schulanfang und die jeweiligen Rahmenbedingungen des Unterrichts kennen. Dennoch dürfte immer noch ein bedeutsamer Anteil der Unterschiede auf Differenzen im Unterricht zurückzuführen sein.

[99] ➔ Abb. 8.3a

9 Beziehung zwischen Testergebnissen und Noten

Im Gefolge von PISA und IGLU ist eine breite Lehrerschelte betrieben worden: Erstens würden LehrerInnen [100] schwache LeserInnen nicht erkennen und zweitens streuten ihre Noten bei gleicher Leistung [101] der Kinder unvertretbar weit (Baumert u. a 2003, 70-72; Bos u. a. 2003, 133-134).

9.1 Korrelationen über die Kinder aller Klassen hinweg

In dieser Studie konnte die Übereinstimmung von Lesenote und Testergebnis für eine Teilstichprobe von 9.376 Kindern (nur 3. und 4. Klassen) überprüft werden.

Über die verschiedenen Klassen hinweg ergab sich für die SchülerInnen der dritten Klassen eine Korrelation [102] von .63**, in den vierten Klassen eine Korrelation von .56** . Die Korrelation des Tests mit der Rechtschreibnote ist jeweils etwas niedriger (.52** bzw. .49**) [103].

Für diese – wie in anderen Studien auch – relativ geringe Übereinstimmung kann es zwei Gründe geben:

- Erstens könnte es sein, dass die LehrerInnen die Leseleistung der Kinder über- oder unterschätzen, dass sie also schon nicht richtig wahrnehmen, was einzelne SchülerInnen (nicht) können.

- Es ist andererseits auch denkbar, dass sie die Leistungen richtig erfassen, dass aber in verschiedenen Klassen dieselbe Leistung unterschiedlich benotet wird, weil die LehrerInnen unterschiedliche Maßstäbe anlegen.

9.2 Korrelationen innerhalb einzelner Klassen

Deshalb ist es wichtig, die Korrelationen zusätzlich für jede Klasse einzeln zu berechnen (vgl. Köller 2002). Mittelt man die so gewonnenen klassenbezogenen Korrelationen, erhält man ein genaueres Bild, wie weit es den LehrerInnen gelingt, innerhalb ihrer Klasse die Leistungsverteilung zutreffend abzubilden.

Bei diesem Vorgehen erhält man folgende Verteilung :

---

100 ... zumindest in der Sekundarstufe (Baumert u. a. 2001, 119-120).

101 Unterstellt wurde dabei, dass die „wahre Leistung“ durch den zweistündigen Test erfasst worden sei...

102 Vgl. Tab. 9.1a-b im Anhang

103 Aber Lese- und Rechtschreibnote korrelieren untereinander mit jeweils .68 am höchsten.

Tab 9.2a Korrelationen zwischen Lesenote und Testleistung		
	Häufigkeit in der Stichprobe (N = 140 Klassen)	
	abs.	Anteil
.00 - .09	0	.0 %
.10 - .19	1	0.7 %
.20 - .29	0	.0 %
.30 - .39	2	1.4 %
.40 - .49	14	10.0 %
.50 - .59	13	9.3 %
.60 - .69	29	20.1 %
.70 - .79	44	31.4 %
.80 - .89	34	24.3 %
.90 - .99	3	2.1 %

Betrachtet man diese Verteilung, so sieht man, dass mehr als die Hälfte der Korrelationen über .70 liegt. *Innerhalb* ihrer Klassen kommen die meisten LehrerInnen also zu ähnlichen Beurteilungen wie der Test, was die Leistungspositionen der einzelnen Kinder im Verhältnis zueinander betrifft [104]. Allerdings weichen Noten und Testergebnis in einem Fünftel der Fälle *auch innerhalb* der einzelnen Klassen erheblich voneinander ab.

9.3 Vergleich von Testleistungen und Noten in verschiedenen Klassen

Im Anschluss an PISA (vgl. Baumert u. a. 2002, 19) ist besonders interessant zu sehen, welche Noten LehrerInnen bei Testleistungen im unteren Bereich geben, anders gesagt: ob sie wahrnehmen, wenn die Leistungsentwicklung von einzelnen Kindern gefährdet ist.

Die folgende Tabelle zeigt die Notenanteile für dieselbe Testleistung (a) in dritten Klassen (schwarze Zahlen oben links in der betreffenden Zelle) und (b) in vierten Klassen (rote Zahlen unten rechts) [105] :

[104] Für die selbst bei Korrelationen von .70 (und höher) noch verbleibenden Abweichungen ist zu berücksichtigen dass eine unvollständige niedrige Korrelation teilweise daher rühren kann, dass LehrerInnen (anders als ein punktueller Test) bei ihrer Bewertung auch die Lerngeschichte der SchülerInnen berücksichtigen, also Einzelergebnisse relativieren, bzw. – durchaus im Sinne der Richtlinien – bei demselben Ergebnis auch die unterschiedliche Anstrengung bewerten

[105] Auszüge aus den Tab. 9.3b-c im statistischen Anhang.

Tab. 9.3a : Klassifikation von Kindern nach Niveaus der Testleistung und Noten							
Niveau [106]	Note ➔	1	2	3	4	5	6
1	Klasse 3 (N = 40)	60.0%	37.5%	2.5%	-	-	-
	Klasse 4 (N = 489)	36.4%	49.9%	11.5%	2.2%	-	-
2	Klasse 3 (N = 2.130)	18.9%	56.5%	22.3%	2.2%	0.0%	-
	Klasse 4 (N = 3.293)	9.7%	47.2%	36.2%	6.8%	0.1%	-
3	Klasse 3 (N = 1.710)	2.3%	32.4%	51.8%	13.2%	0.4%	-
	Klasse 4 (N = 746)	1.1%	14.6%	50.9%	32.6%	0.8%	-
4	Klasse 3 (N = 489)	0.2%	9.6%	54.0%	33.7%	2.5%	-
	Klasse 4 (N = 128)	0.8%	5.5%	41.4%	47.7%	4.7%	-
5	Klasse 3 (N = 245)	0.8%	3.3%	33.5%	54.7%	7.3%	0.4%
	Klasse 4 (N = 41)	-	(7.3%)	(26.8%)	(61.0%)	(4.9%)	-
6	Klasse 3 (N = 61)	-	(3.3%)	(19.7%)	(57.4%)	(19.7%)	-
	Klasse 4 (N = 4)	-	-	-	(100.0%)	-	-

Interessant ist schon der Vergleich der schwarzen und der roten Zahlen innerhalb einzelner Zellen. Sie zeigen, dass es relativ viele Kinder mit gleicher Testleistung gibt, die auch dieselbe Lesenote haben – obwohl ein ganzes Schuljahr zwischen ihnen liegt. In der Regel würde man erwarten, dass ein Jahr später ein strengerer Maßstab angelegt wird, d. h. dass bei gleicher Testleistung eine schlechtere Note vergeben wird [107] .

Von links nach rechts gelesen zeigt die Tabelle getrennt für jede Klassenstufe, welche Note Kinder mit gleicher Testleistung von ihren LehrerInnen bekommen haben. Diese Lesart verdeutlicht, dass durchgängig rund die Hälfte der Urteile auf dieselbe Note entfallen. Zudem liegt in jeweils 79-98% der Fälle eine Differenz von nur einer Note vor und in 93% bis 98% der Fälle streuen die Beurteilungen nur um +/- eine Note. Die Tabelle macht aber auch sichtbar, dass die Noten in der Regel über vier bis fünf Stufen streuen.

Nicht minder interessant ist analog der vertikale Vergleich in der Spalten: Kinder mit der derselben Note schneiden im Test ganz unterschiedlich ab. Dabei ist es wichtig darauf hinzuweisen, dass sich die Note nicht auf die Beurteilung der Testleistung bezieht, dass also NICHT ein bekanntes (identisches) Testergebnis unterschiedlich bewertet wurde. Insofern können Abweichungen in zwei Richtungen gedeutet werden: Der punktuelle Test erfasst nur unzureichend, wie gut ein Kind lesen kann [108], LehrerInnen dagegen urteilen valider, denn sie können auf ein breiteres Spektrum an Beobachtungen zurückgreifen; oder aber: die Testleistung und die Aus-

[106] Definition der Niveaus nach Zahl der richtig bearbeiteten Wörter pro Minute (s. zur Klassifikation auch oben Kap. 4.1).

[107] Als mögliche Erklärung für manche unerwartete Bewertung kommt noch in Betracht, dass LehrerInnen einzelne Aspekte des Lesens (leise vs. laut; Worterkennen vs. Textverständnis) auf verschiedenen Jahrgangsstufen unterschiedlich gewichten.

[108] Vgl. auch die eindrucksvolle Darstellung von Ratzka (2003) zu ihrer Untersuchung mit dem TIMSS-Test in Grundschulen, nach weniger als die Hälfte der Kinder in verschiedenen parallel eingesetzten Mathematik-Tests (trotz in zwei Fällen gleichen Schwerpunkts) gleiche Rangpositionen erreichte.

wertung sind verlässlicher, weil – von situativen Besonderheiten abgesehen – die gleiche Aufgabe gestellt und bei der Auswertung die gleichen Maßstäbe angelegt wurden. Beide Annahmen dürften jeweils einen Teil der Abweichungen erklären.

Insgesamt deuten die Ergebnisse darauf hin, dass die meisten Grundschul-LehrerInnen die Leseleistung der Kinder in ihrer Klasse und deren relative Position durchaus einzuschätzen wissen, dass aber bei der Bewertung dieser richtig erkannten Leistung, also bei ihrer Benotung in verschiedenen Klassen unterschiedliche Maßstäbe angelegt werden.

Untersuchungen wie die vorliegende können helfen, die individuellen Kriterien zu überprüfen: Durch die Rückmeldung der Ergebnisse erhalten die LehrerInnen Anhaltspunkte, wie bestimmte Leistungen von anderen KollegInnen eingeschätzt werden. Eine solche Kalibrierung der Maßstäbe erscheint uns sinnvoller, als das Urteil der LehrerInnen durch punktuelle Tests oder zentrale Prüfungen mit allen ihren Schwächen zu ersetzen. Auf diese Weise könnte man die Stärken der längerfristigen Beobachtung durch die Lehrpersonen und ihres Hintergrundwissens über die Kinder erhalten und zugleich den Blick der LehrerInnen über ihre Lerngruppe hinaus weiten.

9.4 Vergleich von Testleistungen und Noten in verschiedenen Teilgruppen

Klassenübergreifende Vergleiche können auch Hinweise auf gruppenspezifische Verzerrungen geben. So deuten unsere Analysen darauf hin, dass Mädchen bei gleicher Testleistung generell etwas bessere Noten bekommen [109] :

Tab. 9.4a : Noten bei gleicher Testleistung in Klasse 3 und 4 – differenziert nach Geschlecht

Jahrgang	3. Klasse		4. Klasse	
Geschlecht	J	M	J	M
12+ Sätze richtig/ min.	1.5	1.4	1.9	1.7
6 -12	2.2	2.0	2.5	2.3
4 - 6	2.8	2.7	3.2	3.1
3 - 4	3.4	3.2	3.6	3.3
2 - 3	3.6	3.7	3.6	3.6
0 - 2	4.1	3.8	4.0	4.0

Auch wenn die Unterschiede gering sind, weisen fast alle Differenzen in dieselbe Richtung [110] . Lediglich im untersten Leistungsbereich sind keine bzw. keine einheitlichen Unterschiede nachweisbar.

Deutlicher sind die Unterschiede zwischen Kindern mit und ohne Migrationshintergrund (definiert durch die andere Muttersprache mindestens eines Elternteils [111]):

[109] Auszüge aus den Tab. 9.4c-f im statistischen Anhang.

[110] Die Effektstärken liegen bei .15 bis .30 .

Tab. 9.4b : Noten bei gleicher Testleistung in Klasse 3 und 4 – differenziert nach Migrationsstatus [112]				
Jahrgang	3. Klasse		4. Klasse	
Status	Migrant	deutsch	Migrant	deutsch
12+ Sätze richtig/ min.	2.0	1.4	2.3	1.7 **
6 -12	2.4	2.0 **	2.7	2.4 **
4 - 6	3.0	2.7**	3.4	3.1 **
3 - 4	3.4	3.2**	3.7	3.3 **
2 - 3	3.7	3.6	3.7	3.6
0 - 2	3.9	3.9	4.0	4.0

Abgesehen vom untersten Leistungsbereich bekommen Kinder anderer Muttersprache bei gleicher Testleistung schlechtere Noten. Diese Benachteiligung ist besonders deutlich im obersten Leistungsbereich [113].

<u>Fazit zu Kap. 9</u> :

Die Korrelationen zwischen Noten und Testergebnis über alle Kinder eines Jahrgangs hinweg sind mäßig (um .55). Dieses Ergebnis bestätigt die Befunde anderer Studien, dass in verschiedenen Klassen unterschiedliche Anforderungen und Bewertungsmaßstäbe gelten [114]. Interessant ist aber, dass die Korrelationen *innerhalb einzelner* Klassen – von einigen Besorgnis erregenden Ausnahmen abgesehen [115] – deutlich höher liegen (.60 - .90). Dieses Ergebnis deutet darauf hin, dass die meisten Grundschul-LehrerInnen die Leistungspositionen der Kinder in ihrer Klasse durchaus einzuschätzen wissen, dass sie bei der Benotung aber in verschiedenen Klassen unterschiedliche Maßstäbe anlegen. Untersuchungen wie die vorliegende können helfen, die individuellen Kriterien zu überprüfen. Im Übrigen erhalten bei gleichem Testergebnis Mädchen besser Noten als Jungen, deutschsprachige Kinder bessere Noten als Migrantenkinder – außer jeweils im untersten Leistungsbereich.

---

[111] Vgl. die Hinweise in Kap. 5 .

[112] Auszüge aus den Tab. 5.4g-j .

[113] Die Effektstärke schwankt hier immerhin um 1.0 – eine ganz erhebliche Differenz. Aber auch in den mittleren Leistungsgruppen betragen die Effektstärken noch .30 bis .60

[114] ➔ Tab. 9.2a, 9.3a . Allerdings kann eine niedrige Korrelation – zumindest teilweise – auch daher rühren, dass LehrerInnen (anders als ein punktueller Test) bei ihrer Bewertung auch die Lerngeschichte der SchülerInnen berücksichtigen, also zufällige Einzelergebnisse durch längerfristige Beobachtungen relativieren, bzw. – durchaus im Sinne der Richtlinien – bei demselben Ergebnis auch die unterschiedliche Anstrengung bewerten.

[115] Es gab sogar Klassen, in denen die Korrelation zwischen Noten und Testergebnissen unter .40 lag! Von 140 Einzelkorrelationen lagen allerdings nur 3, also knapp 2.1%, unter .40, weitere 10.0% unter .50.

10. Fazit

Am Ende der Grundschulzeit können 90% der Kinder unbekannte Sätze einigermaßen zügig (mindestens fünf Sätze/ Min.) und ohne größere Schwierigkeiten (weniger als 10% Fehler) erlesen und auf ihre Stimmigkeit prüfen. Allerdings streuen die Leistungen am Ende der Grundschulzeit über mehrere Schuljahre hinweg. Diese Ergebnisse von LUST bestätigen damit die über einen anderen methodischen Zugang gewonnenen zentralen Befunde von IGLU [116] .

Weitere 5% der Viertklässler (Prozentrang 5-10) lesen ebenfalls selbstständig, wenn auch recht langsam (12-15 sek pro Satz). Die schwächsten 5% können die Aufgabe entweder nur mit erheblichen Schwierigkeiten bewältigen (mehr als 15 sek pro Satz und im Durchschnitt mehr als 15% Fehler) oder gar nicht richtig lesen (ca. 0.5%).

Bei dieser Einschätzung der Leseleistungen der 10-Jährigen *insgesamt* ist allerdings zu berücksichtigen, dass knapp 5% der Schulanfänger schon vor Schulbeginn oder im Laufe der Grundschulzeit auf Sonderschulen wechseln (in der Regel auch wegen ihrer sehr schwachen schriftsprachlichen Leistungen). Der Anteil der in ihrer Leseentwicklung gefährdeten Kinder ist insofern um rund 3-4% höher, d. h. je nach Härte des Erfolgskriteriums mit 10 bis 15% anzusetzen.

In Bezug auf die große Mehrheit der Kinder, auch die mit unterdurchschnittlichen Leistungen, ist andererseits zu berücksichtigen, was wir als „Karawanen-Effekt“ bezeichnen. Die unteren 10%, 20% oder 30% einer Gruppe sind *definitionsgemäß* immer schlechter als der Durchschnitt. Mit der Fixierung des Blicks auf ihren Platz in der Bezugsgruppe wird leicht übersehen, dass *alle* SchülerInnen von Jahr zu Jahr Fortschritte machen – bezogen auf ihre jeweiligen Voraussetzungen und Lernbedingungen. Auch wenn unser Befund aus dem Querschnittsvergleich noch in einem echten Längsschnitt abzusichern ist: Pädagogisch gesehen sind die *Fortschritte* jeder *Teil*gruppe bedeutsamer als die *Abstände* innerhalb der *Gesamt*gruppe. Das gilt für leistungsschwache SchülerInnen generell und es gilt im Besonderen auch für Migrantenkinder.

Diese Einsicht muss auch Konsequenzen für die Leistungsbewertung haben. Erstaunlich ist, dass die unteren 10-20% trotz ständig negativer Rückmeldung überhaupt noch Fortschritte machen. Aber wie unsere Ergebnisse für die *untersten* 5% zeigen, bleiben diese immer mehr zurück, während sonst alle Teilgruppen vergleichbare Fortschritte machen.

Bei einer Karawane verwundert es niemanden, wenn der, der zuletzt auf die Reise gegangen ist, auch als letzter ankommt. Bedeutsamer ist der Weg, den jedeR Einzelne und die Karawane als *ganze* geschafft haben. In dieser Hinsicht sind unsere Befunde zu den Fortschritten der meisten Kinder im Lesen während der Grundschulzeit ermutigend. Allerdings: Diese Förderung muss in der Sekundarstufe fortgesetzt werden, um die Lesentwicklung derjenigen zu stabilisieren, die als letzte gestartet sind. Denn der entscheidende Faktor scheint für die meisten Kinder „mehr Zeit zum Lernen“ zu sein: Dieselben Ziele werden von fast allen Kindern erreicht – nur von den später gestarteten zu einem späteren Zeitpunkt. „Bildungsstandards“ in Form verbindlicher Niveaustufen für alle zu demselben Termin machen vor diesem Hintergrund keinen Sinn.

[116] S. oben Kasten 1 .

Mehr getan werden muss auf jeden Fall für die leistungsschwächsten 5-10% – aber was? Wenn in PISA von rund 25% leseschwachen SchülerInnen die Rede ist, bedeuten deren Schwierigkeiten beim *Textverständnis* nach den Ergebnissen unserer Studie nicht, dass der Grundschulunterricht *generell Grundfertigkeiten* stärker üben muss. Die zentrale Frage ist vielmehr, was wir unter „Förderung" verstehen, insbesondere ob das im Anfangsunterricht und in der Sonderpädagogik immer noch verbreitete Teilleistungskonzept eine Überwindung oder gar Vermeidung ihrer Schwierigkeiten verspricht. Unsere Erfahrungen sprechen eher dafür, dass eine frühe und selbstständige Beschäftigung mit anspruchsvollen Texten auch für diese Kinder förderlicher ist (vgl. Konzepte wie „Lesewelt Schule" und „freies Schreiben eigener Texte von Anfang an"). Die PISA-Ergebnisse zum geringen Niveau des Textverständnisses, zur geringen Lesehäufigkeit und zur wenig verbreiteten Lesefreude außerhalb der Schule macht außerdem deutlich, dass dieser Ansatz auch für alle anderen Kinder, also für den Unterricht *insgesamt* eine große Bedeutung haben könnte. Insbesondere die massiven Leistungsunterschiede – schon in Klasse 2 – erfordern von Anfang an in der Gestaltung des Unterrichts Raum für Aktivitäten und Aufgaben auf ganz unterschiedlichen Niveaus.

Auch für die Sekundarstufe sind die Ergebnisse bedeutsam. Sie muss sich wie die Grundschule auf die unterschiedlichen Voraussetzungen ihrer SchülerInnen einlassen und Raum für unterschiedliche Lernschritte geben. Erste Analysen zur *Rechtschreib*entwicklung in der Sekundarstufe I zeigen, dass sich der Karawanen-Effekt bis Klasse 10 fortsetzt. Hierauf muss sich der Unterricht einstellen.

Als positives Ergebnis des Projekts LUST kann schließlich festgehalten werden, dass mit sehr geringem Aufwand (Testzeit pro Klasse 4-10 min., Auswertungsbudget pro Bezirk < 5.000 € [117]) produktive Prozesse in den Kollegien der beteiligten Schulen angeregt werden konnten. Dieses individuelle Nachdenken und der Austausch innerhalb von Kollegien und in Schulbezirken lässt sich mit Hilfe der hier vorgelegten Referenzdaten auch an anderer Stelle in Gang setzen. Der Stolperwörter-Leselest lässt sich nach den Erfahrungen in unserer LUST-Studie von LehrerInnen leicht selbst durchführen und auswerten. Eine differenziertere Version des Stolperwörter-Leselests zur genaueren Analyse des Leistungsprofils einzelner Kinder wird von uns gegenwärtig entwickelt. Er soll 2003 bis 2006 in zwei echten Längsschnitten (LUST-2 und LUST-3) erprobt werden, wenn wir die erforderliche finanzielle Förderung absichern können.

[117] Dieser finanzielle Aufwand für die Auswertung fällt nur an, wenn Vergleichswerte für den gesamten Bezirk gewonnen und differenziertere Analysen durchgeführt werden sollen. Wenn einmal eine Vollerhebung durchgeführt worden ist, kann jede Lehrperson den Test auch selbst durchführen, wie eine Reihe von LehrerInnen dies auch schon (vorweg) in unserer Erhebung getan haben.

Literaturnachweise

Anderson, R.C., et al. (1988a): Growth in reading and how children spend their time outside school. In: Reading Research Quarterly, Vol. 23, No. 3, 285-303.

Arbeitsgruppe Leseförderung (1978): Taktiken des Lesens. In: Grundschule, 10. Jg., H. 7, 299-303.

Auernheimer, G., u. a. (Hrsg.) (2001) : Interkulturalität im Arbeitsfeld Schule. Interkulturelle Studien Bd.8. Leske + Budrich: Opladen.

Balhorn, H./ Brügelmann, H. (Hrsg.) (1995): Rätsel des Schriftspracherwerbs. Neue Sichtweisen der Forschung. "Auswahlband Theorie" der DGLS-Jahrbücher 1-5. Libelle: CH-Lengwil.

Bartnitzky, H., u. a. (1999): Zur Qualität der Leistung. 5 Thesen zur Evaluation und Rechenschaft der Grundschularbeit. Grundschulverband – Arbeitskreis Grundschule e. V.: Frankfurt. (auch in: Schmitt 1999, 164 ff.)

Baumert, J., u. a. (Hrsg.) (2001): PISA 2000 – Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich. Leske + Budrich: Opladen.

Baumert, J., u. a. (Hrsg.) (2002): PISA 2000 – Die Länder der Bundesrepublik Deutschland im Vergleich. Leske + Budrich: Opladen.

Baumert, J., u. a. (2003): PISA 2000. Ein differenzierter Blick auf die Länder der Bundesrepublik Deutschland. Zusammenfassung zentraler Befunde. Max-Planck-Institut für Bildungsforschung: Berlin.

BMB+F (Hrsg.) (2001): Grund- und Strukturdaten 1997/98. Bundesministerium für Bildung, Wissenschaft, Forschung und Technologie: 53170 Bonn.

Brinkmann, E. (1997): Rechtschreibgeschichten -- Zur Entwicklung einzelner Wörter und orthographischer Muster über die Grundschulzeit hinweg. Bericht No. 35 des Projekts OASE, FB 2 der Universität, 57078 Siegen.

Brügelmann, H. (2000): IGLU: Glasklare Information über den aktuellen Leistungsstand – oder droht den Grundschulen wieder eine pädagogische Eiszeit? In: Grundschulverband aktuell, Nr. 72, 3-8.

Brügelmann, H. (2003a): IGLU-2001: Zu Risiken und Nebenwirkungen... In: Grundschulverband aktuell. Nr. 82 (April 2003, 12-16).

Brügelmann, H. (2003b): Der Karawanen-Effekt beim Rechtschreiblernen. Kinder anderer Muttersprache lernen genauso schnell, aber von einem anderen Startpunkt aus. Bericht aus der Reanalyse der freien Texte von 4.- bis 10-Klässlern in der NRW-Kids Studie 2001. Vervielf. Ms. FB 2 der Universität: Siegen (i. V.).

Brügelmann, H./ Backhaus A. (2003): Entwicklung der Leseleistung in verschiedenen Leistungsgruppen. Lese-Untersuchung mit dem Stolperwörter-Test (LUST-3). Antrag an die Deutsche Forschungsgemeinschaft (Entwurf: Stand 30.9.). FB 2 der Universität: Siegen.

ddp (2003): Drei Prozent „Sitzenbleiber“. In: Süddeutsche Zeitung v. 12.3.03, Nr. 59, 50.

Geißler, R. (2002): Die Sozialstruktur Deutschlands. Die gesellschaftliche Entwicklung vor und nach der Vereinigung, Hrsgg. von der Bundeszentrale für politische Bildung. Westdeutscher Verlag: Wiesbaden.

Heinzel, F./ Prengel, A. (Hrsg.): Heterogenität, Integration und Differenzierung in der Primarstufe. Jahrbuch Grundschulforschung Bd. 6. Leske+Budrich: Opladen.

Köller, O. (2002): Des Schülers Leid, des Lehrers Freud. Schulnoten sind nötig und besser als ihr Ruf. In: Schule – Wissen – Bildung. Klett ThemenDienst Nr. 16: Dezember 2002, 7-10.

May, P. (1995): Kinder lernen rechtschreiben: Gemeinsamkeiten und Unterschiede guter und schwacher Lerner. In: Balhorn, H./ Brügelmann, H. (1995, 220-229). [Nachdruck aus 1990]

May, P. (2000): Lernförderlichkeit im schriftsprachlichen Unterricht. Effekte des Klassen- und Förderunterrichts in der Grundschule auf den Lernerfolg. Ergebnisse der Evaluation des Projekts „Lesen und Schreiben für alle“ (PLUS). Juni 2000. Behörde für Schule, Jugend und Berufsbildung: Hamburg.

May, P. (2002b): Hamburger Schreibprobe 1+. Hinweise zur Durchführung und Auswertung. Verlag für Pädagogische Medien: Hamburg.

Metze, W. (2003): Stolperwörter-Lesetest. Ergebnisse der Stichprobenerhebung (Stand 3.8.2003). Vervielf. Ms. Auch zugänglich über: www.lesetest1-4.de .

Panagiotopoulou, A. (2002): Lernbiografien von Schulanfängern im schriftkulturellen Kontext. In: Heinzel, F./ Prengel, A. (2002, 235-241).

Peschel, F. (2002): Offener Unterricht -- Idee - Realität - Perspektive und ein praxiserprobtes Konzept zur Diskussion. Teil I: Allgemeindidaktische Überlegungen. Schneider Verlag Hohengehren: Baltmannsweiler.

Peschel, F. (2002): Offener Unterricht - Idee, Realität, Perspektive und ein praxiserprobtes Konzept in der Evaluation. Dissertation. FB 2 der Universität: Siegen. Schneider Verlag Hohengehren: Baltmannsweiler.

Ratzka, N. (2003): Mathematische Fähigkeiten und Fertigkeiten am Ende der Grundschulzeit – Empirische Studien im Anschluss an TIMSS. Dissertation im FB 2 der Universität: Siegen.

Richter, S. (1996): Unterschiede in den Schulleistungen von Jungen und Mädchen. Geschlechtsspezifische Aspekte des Schriftsprachenerwerbs und ihre Berücksichtigung im Unterricht. S. Roderer: Regensburg.

Rolff, H.-G. (1999): Schulentwicklung in der Auseinandersetzung. In: Pädagogik, 51. Jg., H. 4, 37-40.

Rolff, H.-G. (2003): Bildungsstandards sind attraktiv - und problematisch. In: Frankfurter Rundschau, 12.03., WB 5.

Rüesch, P. (1999): Gute Schulen im multikulturellen Umfeld. Hrsgg. von der Bildungsdirektion des Kantons Zürich. Orell Füssli: Zürich.

Schmitt, R. (Hrsg.) (1999): An der Schwelle zum dritten Jahrtausend. BundesGrundschulKongress 1999. Grundschulverband – Arbeitskreis Grundschule: Frankfurt.

Schründer-Lenzen, A. (2003): Lesekompetenzentwicklung von Migrantenkindern. Erste Ergebnisse einer Langzeitstudie zur Schulleistungsentwicklung von Kindern mit Migrationshintergrund im Verlauf der Grundschule. Vervielf. Ms. zum Vortrag auf der Jahrestagung „Grundschulforschung“ der DGfE am 1.1.2003. Universität: Bremen.

Walter, P. (2001): pädagogische Kompetenz und Erfahrung in kulturell heterogenen Grundschulen: In: Auernheimer u.a. (2001).

➔ Ergänzende Informationen

Informationen des Projekts IGLU:

www.erzwiss.uni-hamburg.de/IGLU/Info/info.htm

www.erzwiss.uni-hamburg.de/IGLU/Info/publika.htm

Informationen des Projekts LUST

z. B. der technische Anhang mit den vollständigen statistischen Kennwerten:

www.uni-siegen.de/~agprim/lust/index.htm

Aktuelle Kommentare:

www.grundschulverband.de ➔ „PISA/IGLU“