

Projekt LUST¹

Zwischenbericht (3.4.2003)²

von

Hans Brügelmann³

(Universität Siegen)

0. Vorwort

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.

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 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 eingesetzt werden können.

¹ Lese-Untersuchung mit dem Stolperwörter-Test (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 wird 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 in einer Situation, die uns selbst überwältigt hat.

² **Zwischenauswertung von mehr als 6.500 der insgesamt rund 15.000 beteiligten SchülerInnen.** Weitere Ergebnisse und Interpretationen werden veröffentlicht unter → www.uni-siegen.de/~agprim/lust/index.htm .

³ Ich danke ganz herzlich Sina Gerlach, Oliver von Keutz, Steffi Maxa, Sara Roth, Markus Spannan, Petya Todorova, die gemeinsam mit mir das fast endlose Geschäft des Kodierens übernommen haben (s. im Übrigen Anm. 1)

Da die Ministerien zur Zeit Großstudien favorisieren⁴ hat der Grundschulverband ein Projekt der Universität Siegen unterstützt, in dem in drei Schulamtsbezirken Nordrhein-Westfalens⁵ allen LehrerInnen der 2., 3. und 4. Klassen ein einfacher Lesetest⁶ und – für die Erstdurchführung⁷ – auch deren Auswertung angeboten wurde⁸.

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

- Der Test stellt eine Art „Warnlampe“ dar, er bietet keine spezifische Diagnose. 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 stellen keine höhere Autorität dar.
- 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. Sie kann die punktuellen Testwerte aus ihrer längeren Kenntnis des einzelnen Kindes ggf. auch relativieren.
- 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. Ggf. ist auch das Klassenergebnis entsprechend zu relativieren.

⁴ Vgl. neben TIMSS und PISA für den Grundschulbereich die bevorstehenden Publikation der Befunde der *Internationalen Grundschul-Lese-Untersuchung (IGLU)*.

⁵ 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.

⁶ 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 60 einzelnen Sätzen mit 4-15 Wörtern ein jeweils nicht passendes Wort zu streichen. Zur Aussagekraft des Tests ist ein spezifischer Bericht in Vorbereitung.

⁷ Dies war erforderlich, um den Schulen rasch Vergleichswerte für die Interpretation der eigenen Daten anbieten zu können.

⁸ LUST ist insofern als Komplementäruntersuchung zur Internationalen Grundschul-Lese-Untersuchung (IGLU/PIRLS) zu lesen, deren Ergebnisse am 8.4. 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 eine Vollerhebung in drei Schulbezirken. Und anders als bei IGLU („Textverständnis“) geht es bei LUST um die grundlegenden „Taktiken“ des Lesens : *rasches* und *genaues Satzlesen* durch Verbindung verschiedener Zugriffe auf Schrift (s. a. Anm. 6 und ergänzend zu diesem Ansatz : Arbeitsgruppe Leseförderung 1978).

- 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. Wenn andererseits der Test sehr schwierig ist, können sich die absoluten Werte deutlich unterscheiden – aber diese Unterschiede sind vielleicht im Alltag irrelevant.

Wir haben im Folgenden versucht

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

damit Sie unsere Einschätzungen selbst überprüfen können.

2. Maße und Kategorien der Auswertung

Die Leistung im Test kann zusammenfassend durch die Zahl der richtig bearbeiteten Sätze beschrieben werden („Rohpunkte“ in der Anweisung des Testautors Wilfried Metzke). In dieses Maß gehen allerdings zwei Teilleistungen ein, die bei gleichem Summenwert sehr unterschiedlich aussehen können: Die Geschwindigkeit (Zahl der *bearbeiteten Sätze*) und 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 ausfallen, und auch Fördermaßnahmen sollten spezifisch ausfallen⁹.

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

Blickt man auf die Klasse als Auswertungsgröße, ergibt sich ein zusätzliches Problem. Mittelwerte sind eine Größe, die die Leistung der Gruppe auf einen Blick vermitteln kann. Aber auch derselbe Durchschnitt kann aus ganz unterschiedlichen Verteilungen resultieren.

⁹ Nur selten sind Fördermaßnahmen so klug gestaltet wie im Computerprogramm LALIPUR, das kostenlos herunterzuladen ist von der Homepage unserer 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.

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. Die „Leistung“ 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) 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 der Schlussauswertung werden wir zusätzlich die Ergebnisse der *im Ergebnis* schwächsten und der stärksten Klasse ausweisen – immer mit dem bereits erwähnten Vorbehalt, dass wir über die Bedingungen, unter denen diese Leistungen erzielt wurden, nichts wissen.

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, d. h. wie viele *bearbeitete* Sätze und wie viele *richtige* Sätze sie im Durchschnitt in einer Minute geschafft haben. Diese Kennwerte sind zugleich – neben der *Fehlerquote* – 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
- Schulbezirk

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.

3. Ergebnisse der Jahrgänge im Vergleich ¹⁰

Im Durchschnitt brauchen Kinder der 2. Klassen 14 sek pro *richtig* bearbeitetem Satz. Drittklässler benötigen nur noch 10 sek und Viertklässler im Durchschnitt 7 sek. Bezogen auf die durchschnittlich 2-4 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 ¹¹.

Von Klasse 2 bis Klasse 4 nimmt sowohl die Lesegeschwindigkeit erheblich zu (nur noch 7 sek statt 14 sek pro bearbeitetem Satz) als auch die Fehlerquote erheblich ab (von fast 13% auf gut 4%) ¹².

3.1 Mittelwerte und Streuungen der grundlegenden Kennwerte

Mitte 2. Klasse, also in den meisten Klassen „nach Abschluss des Leselehrgangs“, wie es in vielen Lehrplänen heißt, ist der früheste Termin unserer Erhebungen.

Beginnen wir mit verschiedenen zusammenfassenden Kennwerten:

Tab 3.1a: Durchschnittsleistungen Mitte 2. Klasse

2. Klassen Grund- Daten	aM (SD)	Roh- Punkte	richtig = bearbeitet - falsch	Fehler- quote	Bearbeitete Sätze pro Minute	Richtige Sätze pro Minute	Sek pro richtigem Satz
N = Mitte des Schuljahres	2.095 ¹³	25.3 (13.4)				4.2 (2.2)	14.2
N =	1.623 ¹⁴		25.8 (13.7)	12.9 (15.8)	4.8 (2.2)		

¹⁰ 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 (Vollerhebungen aus denselben Bezirken) zeigen aber, welches Niveau Schülergruppen verschiedenen Alters zu den genannten Zeitpunkten erreichen, so dass mit großer Plausibilität eine Durchschnittsentwicklung erschlossen werden kann.

¹¹ → Tab. 3.1a, 3.1b, 3.1c

¹² → Tab. 3.1a, 3.1b, 3.1c

¹³ Da ein Teil der LehrerInnen schon selbst die Punkte (= bearbeitet – falsch) errechnet hat, haben wir für den Gesamtwert und für die Teilwerte unterschiedlich große Stichproben. Der Vergleich der Kennwerte „Rohpunkte“ mit „richtig“ zeigt zum jetzigen Zeitpunkt der Auswertungen nur für die 2. Klassen einen (allerdings geringfügigen) Unterschied. Noch geringer sind die Unterschiede in Klasse 3 und 4 – ein Beleg, dass LehrerInnen den Test ohne große Vorbereitung alleine auswerten können.

¹⁴ s. Anm. 5: in dieser Zeile finden sich die Kennwerte der nur von uns ausgewerteten Teilgruppe.

Die Zweitklässler schaffen im Durchschnitt in sechs Minuten 25 Sätze richtig. Sie brauchen also knapp 15 sek pro Satz, d. h. etwa fünfmal so lange wie die von uns getesteten Erwachsenen. Wie die Rohpunkte und die Fehlerquote zeigen, liegt für den durchschnittlichen Zweitklässler die Hauptschwierigkeit nicht mehr im Lesen-Können (nur 13% der Sätze werden falsch bearbeitet), sondern in der Lesegeschwindigkeit.

Auch für **Mitte 3. Klasse** haben wir die grundlegenden Kennwerten zusammen gestellt:

Tab 3.1b: Durchschnittsleistungen Mitte 3. Klasse

3. Klassen Grund-Daten	aM (SD)	Roh-Punkte	richtig = bearbeitet - falsch	Fehler-Quote	Bearbeitete Sätze pro Minute	Richtige Sätze pro Minute	Sek pro richtigem Satz
N = Mitte des Schuljahres	2.203	30.5 (11.6)				6.1 (2.3)	9.8
N =	1.718		30.5 (11.7)	6.2 (8.4)	6.4 (2.3)		

Die Drittklässler schaffen in fünf Minuten durchschnittlich 30 Sätze richtig. Sie brauchen also etwa 10 sek pro Satz, d. h. etwa 35% weniger Zeit als die Zweitklässler. Die Fehlerquote ist dabei auf etwa die Hälfte geschrumpft und auch die Zeit pro bearbeitetem Satz hat sich deutlich reduziert. Die Standardabweichung ist immer noch hoch, aber die Leistungen streuen nicht mehr so breit wie in der zweiten Klasse.

Die folgende Tabelle gibt einen Überblick über die Kennwerte **für Mitte 4. Klasse** – „am Ende der Grundschulzeit“:

Tab 3.3 a: Durchschnittsleistungen Mitte 4. Klasse

4. Klassen Grund-Daten	aM (SD)	Roh-Punkte	richtig = bearbeitet - falsch	Fehler-Quote	Bearbeitete Sätze pro Minute	Richtige Sätze pro Minute	Sek pro richtigem Satz
N = Mitte des Schuljahres	2.361	34.0 (11.1)				8.5 (2.8)	7.1
N =	1.664		34.2 (11.2)	4.2 (5.7)	8.9 (2.7)		

Im Durchschnitt brauchen die Viertklässler unserer Stichprobe also 7 sek für einen richtig verstandenen Satz und die entsprechende Korrektur. Das ist nur zweimal so lang wie die oben erwähnten akademischen VielleserInnen – für die im „Stolperwörter-Test“ geprüfte *grundlegende* Leseleistung ein gutes Ergebnis, zumal wenn man die Differenz von rund drei Jahren Lesepaxis vs 20 bis 50 Jahren bei den Erwachsenen bedenkt.

3.2 Vergleich der oberen und der unteren Leistungsgruppen

Die leistungsstärksten 15% ¹⁵ brauchen bereits Mitte der 2. Klasse nur 6-9 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% ¹⁶ 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 hohe Standardabweichung in Tab 3.1a zeigt, dass die Leistungen in der **2. Klasse** erheblich streuen. Dies veranschaulichen auch die Werte zu den einzelnen Leistungsgruppen in der folgenden Tabelle:

Tab 3.2a: Extremgruppenvergleich Klasse 2

2. Klasse Prozentränge	Richtige Sätze pro Minute	Sekunden pro richtigem Satz
1 - 5	0.0 – 0.9	67 -
6 - 10	1.0 – 1.5	40 - 60
11 - 15	1.6 – 1.8	33 - 37
....		
85 - 90	6.7 – 7.2	8.3 - 9.0
91 - 95	7.5 – 8.3	7.5 - 8.0
96 - 100	8.5 – 10.5	6.0 - 7.0

Die besten fünf Prozent lesen also mindestens zehnmal so schnell wie die schwächsten fünf Prozent.

Man sieht aber auch, dass die oberen 15% enger beieinander liegen als die unteren 15%. Im unteren Bereich schaffen einige Kinder gar keinen Satz oder landen nur Zufallstreffer, während andere zwar langsam, aber doch erfolgreich ein bis zwei Sätze pro Minute bearbeiten – ein er-

¹⁵ → Tab. 3.2a, 3.2b, 3.2c, 3.2d

¹⁶ → Tab. 3.2a, 3.2b, 3.2c, 3.2d

heblicher qualitativer Unterschied. Im oberen Bereich dagegen unterscheiden sich die Leistungen nur graduell – und ob jemand 7 oder 9 sek pro Satz braucht beeinflusst seine Lesekompetenz nicht. Eine Differenz von fünf Prozenträngen bedeutet im unteren Bereich also ganz andere Leistungsunterschiede als im oberen Bereich.

Und schließlich erstaunt, dass die besten 15% Ende der zweiten Klasse nicht weit vom Lesetempo der Erwachsenen entfernt sind (6-8 sek vs. 2-5 sek). Für die Grundleistung „rasches und genaues Wortlesen sowie Integration der Wortbedeutungen auf Satzebene“ scheint in dieser Teilgruppe bereits der Mastery-Level erreicht. Fortschritte werden diese Kinder eher in anderen Dimensionen bzw. auf anderen Ebenen (Textverständnis) machen.

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

Tab 3.2b: Extremgruppenvergleich Klasse 3

3. Klasse Prozentränge	richtige Sätze pro Minute	Sekunden pro richtigem Satz
1 - 5	2.6 - ...	23 - ..
6 - 10	2.8 – 3.2	19 - 21
11 - 15	3.4 – 3.7	16 - 18
....		
85 - 90	8.6 – 9.0	6.5 - 7.0
91 - 95	9.2 – 10.2	6.0 - 6.5
96 - 100	10.4 - 12.0	5.0 - 5.8

Der absolute Zuwachs ist in beiden Extremgruppen etwa gleich groß. Die besten fünf Prozent sind allerdings nur wenig besser als die vergleichbare Gruppe ein Jahr vorher (Verbesserung um 10-20%). Der Abstand zu den unteren 15% ist erheblich geschrumpft (wenn man einmal von den allerschwächsten LeserInnen absieht). 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).

Und die Streuung in **Klasse 4**?

Tab 3.2c: Extremgruppenvergleich Klasse 4

4. Klasse Prozentränge	richtige Sätze pro Minute	Sekunden pro richtigem Satz
1 - 5	0.0 - 3.8	16 - ...
6 - 10	4.0 - 4.8	12 - 15
11 - 15	5.0 - 5.5	11 - 12
....		
85 - 90	11.3 – 12.0	5.0 - 5.3
91 - 95	12.2 – 13.2	4.5 - 4.9
96 - 100	13.5 – 15.0	4.0 - 4.4

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

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 erheblich geschrumpft. Selbst schwache LeserInnen (Prozentrang 11-15) brauchen nur zwei- bis dreimal so lange wie die sehr guten LeserInnen.

Geht man noch weiter in die Extreme, ergibt sich folgendes Bild:

Tab 3.2d: Extremgruppenvergleich Klasse 4 (untere und obere 5%)

4. Klassen nach Leistung ¹⁷	aM (SD)	Roh-Punkte	richtig = bearbeitet - falsch	Fehler-Quote	Bearbeitete Sätze pro Minute	Richtige Sätze pro Minute	Sekunden pro richtigem Satz
Untere 5 % N = 136		11.4 (4.8)	12.6 (3.9)	13.6 (11.3)	3.7 (1.3)	2.9 (1.1)	20.1
Obere 5 % N = 100		57.4 (2.1)	57.3 (2.1)	1.5 (1.9)	14.6 (5.1)	14.3 (1.0)	4.2

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 braucht mit 20sek für einen richtigen Satz fünfmal so lange wie der beste Mitschüler. Er bearbeitet zwar knapp vier Sätze pro Minute. Damit benötigt er nur viermal so lange wie der beste Mitschüler. Gleichzeitig macht er aber mehr als neunmal so viele Fehler.

Die Quote von 11% falschen Urteilen erscheint für besonders leistungsschwache SchülerInnen gering, wenn man an die dramatischen Meldungen über wachsende Analphabeten- und Legasthenikerquoten denkt. Dennoch zeigen sich auf beiden Indikatoren – Lesetempo und Lesegenauigkeit – deutliche Unterschiede zwischen den Leistungsextremen. Deren alltagspraktische Bedeutung muss noch gesondert diskutiert werden (s. das folgende Kap. 4).

¹⁷ gesplittet nach Variable „pr_5_95“

4 Die Leistungen im unteren Bereich

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*)¹⁸.

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 sind¹⁹. 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%²⁰. 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.

Die durchschnittlichen Leistungen und erst recht das Niveau der Spitzengruppe werfen nach den oben berichteten Daten 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 schwächsten zwei, drei SchülerInnen pro Klasse

¹⁸ → Tab. 4.1b

¹⁹ ... 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.). Diesen beiläufigen Übungsvorteil kann kein Förderunterricht ausgleichen. Vielmehr kommt es darauf an, auch die leseschwachen SchülerInnen „zum Lesen zu verlocken“ – in ihrer Freizeit

²⁰ → Tab. 4.1c

Solange es keine Längsschnitte über die weitere Entwicklung von SchülerInnen auf verschiedenen Niveaus gibt, kann jede Entscheidung zwischen verschiedenen Leistungsanforderungen nur Plausibilitätscharakter beanspruchen. Dies umso mehr, als aus der Forschung sehr unterschiedliche Entwicklungsverläufe zum Schriftspracherwerb bekannt sind.

Damit sich die LeserInnen ein eigenes Urteil bilden können, berichten wir im Folgenden Daten zu vier verschiedenen Kriterien, die unterschiedlich hohe Anforderungen definieren²¹:

- 0-2 Sätze/ min. kann nicht lesen (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 10-20% Fehler) erlesen;
- 3-4 Sätze/ min. kann Sätze erlesen, aber immer noch langsam (15-20 sek) und mühevoll (im Durchschnitt 5-10% Fehler);
- 4-5 Sätze/ min. kann Sätze weitgehend fehlerfrei (3-5%) und zügiger, aber auch immer noch relativ langsam (12-15 sek) erlesen.

In der folgenden Tabelle haben wir deshalb die Anteile in allen drei Kategorien für verschiedene Teilgruppen ausgewiesen, so dass man sich je nach Maßstab ein eigenes Bild machen kann²²:

Tab. 4.1a: Jahrgangvergleich der Anteile (Rohdaten) im unteren Leistungsbereich

richtige Sätze pro Minute ²³	0-2	0-3 ²⁴	0-4	0-5
Jahrgang				
2. Klasse N = 2.095	14.6 %	31.0 %	50.1 %	65.8 %
3. Klasse N = 2.203	1.9 %	6.9 %	17.2 %	31.7 %
4. Klasse N = 2.361	1.1 %	2.0 %	4.5 %	9.7 %

²¹ Die genau definierten Schwellen sind: 0-1,9; 2,0-2,9; 3;0-3,9; 4,0-4,9 .

²² Vgl. ergänzend die Auswertung nach vier Niveaus für die Gesamtgruppe in [s.stolper.03.desc niveaus by stufkorr].

²³ s. Anm. 7

²⁴ Die Werte sind kumulativ berechnet, schließen also die jeweils vorhergehende Gruppe mit ein.

Diese Übersicht vermittelt 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 Vollerhebungen ganzer Schulbezirke 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 Wechsel auf eine Sonderschule 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% 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 muss nach persönlicher Einschätzung erfolgen. Sie ist so vorgenommen worden, dass sie das Ausmaß der Leseschwierigkeiten eher über- als unterschätzt:

Tab. 4.1b: Jahrgangvergleich der Anteile (korrigiert) im unteren Leistungsbereich

richtige Sätze pro Minute ²⁵ Jahrgang (korrigiert)	0-2	0-3 ²⁶	0-4	0-5
2. Klasse N = 2.095	14.6 %	31.0 %	50.1 %	65.8 %
3. Klasse ²⁷ N = 2.203	2.9 %	8.9 %	19.7 %	34.2 %
4. Klasse ²⁸ N = 2.361	3.1 %	6.0 %	9.0 %	14.7 %

²⁵ s. Anm. 7

²⁶ Die Werte sind kumulativ berechnet, schließen also die jeweils vorhergehende Gruppe mit ein.

²⁷ Die zusätzlichen 2.5 % wurden auf die drei unteren Leistungsgruppen mit + 1.0, +1.0 und + 0.5 %-Punkten umgelegt, wobei sich die Werte von links nach rechts wegen der kumulativen Darstellung addieren.

²⁸ Die zusätzlichen 5.0 % wurden auf die vier unteren Leistungsgruppen mit + 2.0, +2.0, + 0.5 und + 0.5 %-Punkten umgelegt, wobei sich die Werte von links nach rechts wegen der kumulativen Darstellung addieren.

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

Einen etwas anderen Zugang zu diesem Problem verschafft die folgende Übersicht, in der die Daten nicht nach *inhaltlichen* Leistungsstufen, sondern nach *prozentual* auf die Stichprobe definierten Leistungsgruppen (Prozentränge = PZR) aufgegliedert werden. Dabei beschränkt sich die Darstellung auf die unteren 30%:

Tab. 4.1c: Jahrgangvergleich der Leistungswerte im unteren Leistungsbereich (30%)

Richtige Sätze pro Minute Jahrgang	Durchschnitt	PZR 1 - 5	PZR 6 -10	PZR 11-15	PZR 16-20	PZR 21-25	PZR 26-30	PZR 95-100
2. Klasse N = 2.095	3.8	.0-0.9	1.0-1.5	1.6-1.8	2.0-2.2	2.2-2.5	2.6-2.8		8.5-10.5
3. Klasse N = 2.203	5.8	.0-2.6	2.8-3.2	3.4-3.6	2.8-4.0	4.1-4.3	4.4-4.6		10.4-12.0
4. Klasse N = 2.361	8.2	.0-3.8	4.0-4.7	5.0-5.5	5.7-5.9	6.0-6.4	6.5-6.7		13.5-15.0
absolute Zuwächse	+ 4.4	+ 1.2	+ 2.6	+ 3.3	+ 3.5	+ 3.5			+ 4.8
proport. Zuwächse	1:2.1	1:2.7	1:2.5	1:2.5	1:2.5	1: 2.3			1 :1.5

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 keinen Satz richtig bearbeiten, andererseits öffnet sich das Leistungsspektrum nach oben (von 0.9 auf 3.8 – was in der zweiten Klassen immerhin einem Prozentrang von 50, also einer durchschnittlichen Leistung entsprochen hätte). In allen anderen Teilgruppen verdoppelt, verdreifacht oder vervierfacht 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 vergleichen²⁹, wie die parallele Färbung verdeutlicht.

²⁹ Dies ist die für meine im Folgenden entwickelten Argumentation ungünstigste Variante, die ich gewählt habe, um meine Folgerungen möglichst gut abzusichern.

Nach dieser Korrektur ergeben sich (von links nach rechts aufgelistet) absolute Zuwächse von 1.2 in der schwächsten Gruppe bis 3.5 in der stärksten Gruppe (der unteren 30% !). 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 Werten ist der Zuwachs höher, proportional gesehen geringer. Betrachtet man die Leistungszuwächse in absoluten Werten, gewinnen die leistungsstärkeren SchülerInnen mehr. Der Abstand von Prozentrang 95 zu Prozentrang 5³⁰ beträgt in Klasse 2 knapp 7 richtigen Sätze pro Minute, in Klasse 4 (korrigiert) dagegen knapp 10 richtige Sätzen pro Minute.

Aber: Alle Gruppen machen Fortschritte, und proportional am stärksten die unteren Leistungsgruppen. Das bedeutet: Das Verhältnis der Leistungen der oberen zu den unteren 15% beträgt in Klasse 2 noch 1:10.5, in Klasse 4 dagegen nur noch 1:5.1 .

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 Klassen³¹. Die unteren 5% oder 15% einer Gruppe sind definitionsgemäß *immer* schlechter als 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 *Teilgruppe* bedeutsamer als die *Abstände* innerhalb der *Gesamtgruppe*. Konkret: Wer in der vierten Klasse auf Prozentrang 15 pro Minute 5.5 Sätze schafft, hat sich gegenüber Klasse 2 um 3.3 Sätze verbessert – sein Abstand zum Durchschnitt beträgt aber nur 2.7 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 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 (12.2 Sätze pro Minute) nicht.

Mit diesem Hinweis sollen aber die in der untersten Leistungsgruppe beobachteten Schwierigkeiten nicht verdrängt werden. Auch wenn die Größenordnung von rund 5% nicht mit den bei PISA berichteten 25% „leseschwachen“ SchülerInnen zu vergleichen ist, stimmt die Art der Probleme besorgt, da es sich hier um grundlegende Teilleistungen handelt. Ebenso deutlich muss man sagen, dass Maßnahmen zur Behebung dieses Problems nicht Inhalt und Stil des

³⁰ Das entspricht in etwa dem Vergleich des leistungsstärksten mit dem leistungsschwächsten Schüler einer Durchschnittsklasse.

³¹ 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 Ausgangsposition 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.).

Unterricht für die übrigen 95% bestimmen darf³². Diese Perspektive 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. Wie die Kinder der Grundschule in dieser Hinsicht abschneiden, werden die IGLU-Ergebnisse zeigen.

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 Kölner Raum einbezogen³³. Bei dieser Sonderstichprobe handelt es sich um Schulen in sog. sozialen „Brennpunkten“, deren SchülerInnen besonders ungünstige Lernvoraussetzungen mitbringen und die häufig auch während der Schulzeit zusätzliche Belastungen zu verarbeiten haben.

[Das Folgende wird nach vollständiger Auswertung ergänzt]

Nachdem 20 Klassen mit insgesamt 434 Kindern ausgewertet wurden, lässt sich als **Zwischenbilanz** über diese Teilstichprobe sagen, dass sie in zwei zentralen Merkmalen tatsächlich von unserer Hauptstichprobe abweicht:

- der Anteil der Kinder von Alleinerziehenden ist mehr als dreimal so hoch (gut 15% statt 5% sonst);
- auch der Anteil der Kinder mit anderer Muttersprache ist drei- bis viermal so hoch (67% statt sonst knapp 18%) .

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

- arithmetisches Mittel 7.6³⁴ (SD 2.5) vs. 8.5 (SD 2.8) in der Gesamtstichprobe;
- 5.5 % der Kinder unter Prozentrang 5 der Gesamtstichprobe;
- 1.7 % haben weniger als 3 Sätze pro Minute richtig (vs. 1.1% in der Gesamtstichprobe);

³² 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.

³³ Ich danke den KollegInnen SAD'in Monika Baum und SAD Hans Wielpütz vom Schulamt der Stadt Köln für ihre rasche und unbürokratische Hilfe.

³⁴ 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]).

- 5.5% haben 0-4 Sätze pro Minute richtig (vs. 3.6% in der Gesamtstichprobe) .

Anders als nach den Umfeldbedingungen erwartet sind also die Leistungen im unteren Bereich kaum schlechter. Insbesondere ist fest zu halten:

- Die deutschsprachigen Kinder schneiden nur wenig schlechter ab als in der Hauptstichprobe (8.3 vs. 8.8 richtige Sätze pro Minute), obwohl sie aus einem sozialen Brennpunkt kommen und in den Klassen mit einem Anteil von 33% (gegenüber 82% in der Hauptstichprobe) die Minderheit darstellen.
- die Kinder anderer Muttersprache schneiden unter diesen Bedingungen mit durchschnittlich 7.2 richtigen Sätzen sogar eher etwas besser ab als sonst (7.0 richtige Sätze in der Hauptstichprobe³⁵).

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.

Sollten sich diese Befunde in der vollständigen Auswertung bestätigen, wären sie ein wichtiger Hinweis darauf, dass es einer Reihe von Grundschulen gelingt, Kindern mit anderer Muttersprache auch unter schwierigen Bedingungen gerecht zu werden. Dabei dürften objektive Rahmenbedingungen (Klassengröße³⁶, Anzahl der Förderstunden), aber auch pädagogische und didaktische Konzepte der LehrerInnen eine bedeutsame Rolle spielen.

³⁵ Auch eine Teilauswertung auf Klassenebene zeigt: 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 ($r = \text{um } .10$) noch in der Gruppe der deutschsprachigen SchülerInnen ($r = \text{um } .00$). Ein 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).

³⁶ Die Zahl richtiger Sätze ist allerdings sowohl bei den deutschsprachigen Schülern als auch bei den Kindern mit anderer Muttersprache unabhängig von der Klassengröße ($r = \text{um } .00$ bis $-.10$).

5 Muttersprache der Eltern

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 ZweitklässlerInnen³⁷. Da Migranten-Kinder deutlich häufiger nicht versetzt oder auf eine Sonderschule überwiesen werden als deutschsprachige SchülerInnen, schön 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 Kindern³⁸. In Klassen mit einem hohen Anteil an Migrantenkindern sind entgegen dem gängigen Vorurteil weder deren Leistungen noch die der deutschsprachigen Kinder schlechter.

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 besser als rechtliche Kategorien. Auf diese Weise werden auch Kinder aus „deutschen“ Aussiedlerfamilien und auch eingebürgerte Kinder von ArbeitsmigrantInnen erfasst. Aber 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 differiert doch sehr stark. Wie die Streuungen (SD) zeigen, ist die Heterogenität in der Regel nicht größer als unter den deutschsprachigen Kindern.

5.1 Grunddaten für die Klassenstufen 2 bis 4

Mitte 2. Klasse macht die Muttersprache der Eltern einen beträchtlichen Unterschied für die Lesefähigkeit aus: Kinder aus Elternhäusern, in denen eine andere Sprache gesprochen wird, schaffen mit 3.3 vs. 4.4 deutlich weniger richtige Sätze als die deutschsprachigen Kinder (auch statistisch nicht nur ein signifikanter, sondern auch ein bedeutsamer Unterschied: $ES = .50$ ³⁹).

Tab. 5.1a: Grunddaten für Kinder deutscher bzw. anderer Muttersprachen in Klasse 2

2. Klassen nach Muttersprache	aM (SD)	Roh-Punkte	richtig = bearbeitet - falsch	Fehler-Quote	Bearbeitete Sätze pro Minute	Richtige Sätze pro Minute
beide Eltern N =	1.241	26.3 (13.3)				4.4 (2.2)

³⁷ → Tab. 5.1a, 5.1b, 5.1c

³⁸ → Tab. 5.2a, 5.2b, 5.2c

³⁹ Die Effektstärke ES wird berechnet aus der Differenz der Mittelwerte, geteilt durch die Streuung der Kontrollgruppe – hier der deutschsprachigen Kinder.

deutschsprachig N =	890		26.9 (13.7)	10.2 (14.0)	4.9 (2.2)	
ein oder zwei Eltern andere Sprache ⁴⁰ N =	335	20.1 (12.6)				3.3 (2.1)
	282		20.6 (12.7)	22.1 (19.2)	4.2 (2.2)	

Rund 30% Differenz bedeuten einen deutlichen Abstand. Umgerechnet entsprechen 1.1 Sätze pro Minute einer durchschnittlichen Entwicklung von etwa einem halben Schuljahr

Die Unterschiede sind auch ein Jahr später, also Mitte der **3. Klasse**, immer noch deutlich (auch statistisch bedeutsam: ES = .54). In absoluten Zahlen entsprechen sie weiterhin etwa einem halben Schuljahr, aber relativ sind sie von rund 30% auf etwa 20% geschrumpft:

Tab. 5.1b: Grunddaten für Kinder deutscher bzw. anderer Muttersprachen in Klasse 3

3. Klassen nach Muttersprache	aM (SD)	Roh-Punkte	richtig = bearbeitet - falsch	Fehler-Quote	Bearbeitete Sätze pro Minute	Richtige Sätze pro Minute
beide Eltern deutschsprachig N =	1.415	31.7 (11.8)				6.3 (2.4)
	1.021		31.7 (12.0)	5.0 (6.9)	6.6 (2.4)	
ein oder zwei Eltern andere Sprache N =	352	25.2 (10.5)				5.0 (2.1)
	280		25.3 (10.6)	10.8 (12.0)	5.6 (2.0)	

In beiden Gruppen hat sich die Fehlerquote gegenüber den Zweitklässlern 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** gut 20% länger für einen richtigen Satz (auch statistisch bedeutsam: ES = .64).

Tab. 5.1c: Grunddaten für Kinder deutscher bzw. anderer Muttersprachen in Klasse 4

4. Klassen nach Muttersprache	aM (SD)	Roh-Punkte	richtig = bearbeitet - falsch	Fehler-Quote	Bearbeitete Sätze pro Minute	Richtige Sätze pro Minute
beide Eltern deutschsprachig N =	1.531	35.2 (11.4)				8.8 (2.8)
	995		36.4 (11.0)	3.5 (4.7)	9.4 (2.7)	
ein oder zwei Eltern	326	28.2 (10.6)				7.0 (2.7)

⁴⁰ Da die Zahl von Kindern mit nur einem Elternteil anderer Muttersprache bei nur 5-20 liegt, haben wir sie nicht gesondert ausgewiesen. In der Tendenz liegen ihre Werte zwischen den beiden anderen Gruppen (beide Eltern deutsch vs. beide Eltern fremdsprachig), mal näher bei der einen, mal näher bei der anderen Gruppe.

andere Sprache N =	238	28.4 (10.1)	6.8 (8.4)	7.6 (2.5)	
-----------------------	-----	----------------	---------------	---------------	--

Auf den ersten Blick ist das Bild analog wie bei den leseschwächeren Kindern insgesamt: Sie machen deutliche Fortschritte, aber da die leistungsstärkeren 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. 5.1 – trügt. Der Anteil der Kinder mit Eltern, die eine andere Sprache sprechen, sinkt von 21.2% in Klasse 2 über 19.9% in Klasse 3 auf 17.6% in Klasse 4. Das bedeutet: Die Gruppe der Kinder mit anderer Muttersprache ist in Klasse 4 um rund 17% 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 14.5% erhöhen (= 17 : [100+17]). Da im Grundschulbereich rund 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 realistisch⁴¹. Für deutschsprachige SchülerInnen müsste analog ein Wert von rund 4 % angesetzt werden. Das bedeutet, dass wir in den unteren Leistungsbereichen für die deutschsprachigen Kinder in Klasse 4 zusammen um 3% höhere Anteile ansetzen müssen, als unsere Rohstatistik ausweist, für die Kinder anderer Muttersprache rund 10%. Die folgende Tabelle macht diesen Korrekturprozess durchsichtig:

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

richtige Sätze pro Minute ⁴² Jahrgang/ Muttersprache	PR 50	0-2	0-3 ⁴³	0-4	0-5
deutsch N = 1.303 2. Klasse	4.2	14.0 %	30.2 %	45.4 %	65.7 %
andere N = 342	3.0	31.6 %	51.8 %	70.5 %	80.7 %
Roh deutsch N = 1.531 4. Klasse ⁴⁴	8.6	1.1 %	1.8 %	4.1 %	9.5 %
korr		+1.5 2.6 %	+2.0 3.8 %	+2.5 6.6 %	+3.0 12.5 %
Roh andere N = 326	6.7	1.8 %	4.9 %	11.0 %	22.4 %
korr		+5.0 6.8 %	+7.0 11.9 %	+ 9.0 20.0 %	+10.0 32.4 %

⁴¹ Die Schätzung ist noch schwieriger als im 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.

⁴² s. Anm. 7

⁴³ Die Werte sind kumulativ berechnet, schließen also die jeweils vorhergehende Gruppe mit ein.

Die Tabelle vermittelt 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 (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 Besetzung erheblich schrumpft: in absoluten Zahlen *durchgängig* stärker bei den Kinder anderer Muttersprache, 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 ⁴⁵.

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 unterschiedlichen Prozentrang-Gruppen, um den Zuwachs abzuschätzen: Prozentrang 1-5 in Klasse 4 mit Prozentrang 10-15 in Klasse 2, Prozentrang 6-10 in Klasse 4 mit Prozentrang 16-20 in Klasse 2 usw. Dies ist ein sehr konservatives Verfahren, dass die erreichten Lernfortschritte eher unterschätzt.

Die farbige Auszeichnung macht deutlich, welche Gruppen jeweils verglichen wurden.

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

Richtige Sätze pro Minute Jahrgang	Durchschnitt PZR 50	PZR 1 - 5	PZR 6 -10	PZR 11-15	PZR 16-20	PZR 21-25	PZR 26-30	PZR 90-100
2. Klasse N = 342	3.0	0.2-0.3	0.5-0.8	1.0-1.2	1.3-1.5	1.6-1.7	1.8-1.9		6.5- 9.7
4. Klasse N = 326	6.7	.0-3.0	3.2-3.8	4.0-4.5	4.6-4.8	5.0-5.2	5.5-5.6		11.0-14.5
absolute Zuwächse	+3.7	+ 0.4	+ 2.0	+ 2.8	+ 2.8				+ 4.6
proport. Zuwächse	1:2.2	1:1.4	1:2.6	1:3.0	1:2.5				1 :1.6

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

⁴⁴ Die zusätzlichen 3.0 % für die deutschsprachigen Kinder wurden auf die vier unteren Leistungsgruppen mit + 1.0, +0.5, + 0.5 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.

⁴⁵ s. dazu auch Anm. 22

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

Richtige Sätze pro Minute Jahrgang	Durchschnitt PZR 50	PZR ⁴⁶ 1 - 5	PZR 6 -10	PZR 11-15	PZR 16-20	PZR 21-25	PZR 26-30	PZR 90-100
2. Klasse N = 1.303	4.2	.0-1.0	1.2-1.7	1.8-2.1	2.2-2.3	2.5-2.7	2.8-3.1		7.7-10.5
4. Klasse N = 1.531	8.6	0.2-4.4	4.5-5.1	5.2-5.7	6.0-6.3	6.5-6.8	7.0-7.2		12.5-15.0
absolute Zuwächse	+4.4	+ 1.4	+ 2.8	+ 2.4	+ 3.5	+ 3.7			+ 4.6
proport. Zuwächse	1:2.0	1:2.7	1:2.5	1:2.8	1:2.4	1: 2.3			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 besorgt stimmen muss.

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. Dabei ist der proportionale Lernzuwachs höher in den unteren, der absolute höher in den oberen Leistungsgruppen.

5.3 Leseleistung 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 von einem „Anteil von SchülerInnen und Schülern mit Migrationshintergrund in Schulen“, also nicht in der einzelnen *Lerngruppe*. Dennoch wird auf den unzulänglichen „Umgang mit Heterogenität“ in Bezug auf den Unterricht gesprochen

⁴⁶ Achtung: Da für die Gruppe der deutschsprachigen Kinder anders für die Kinder anderer Muttersprache eine Verlustquote von nur 5% unterstellt wird, sind die Vergleichswerte farbig in 5er und nicht in 10er Schritten gestaffelt.

Nach der Auswertung von 268 Klassen ergibt sich in LUST für die Grundschule ein etwas anderer Befund. Zwar nimmt die Leseleistung der Klassen mit höherem Anteil an Kindern anderer Muttersprache im Durchschnitt ab. Gliedert man die Leistungen aber nach Teilgruppen auf, differenziert sich das Bild:

Tab. 5.3a: Leseleistung⁴⁷ (= „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 2. Klasse N = 80	RI_MIN_D 3. Klasse N = 84	RI_MIN_D 4. Klasse N = 104	RI_MIN_A 2. Klasse N = 71	RI_MIN_A 3. Klasse N = 70	RI_MIN_A 4. Klasse N = 87
0	3.9	6.1	8.8	-	-	-
01 – 09	3.9	6.5	8.5	2.7	4.2	6.7
10 – 19	4.8	6.5	8.7	3.2	5.2	7.1
20 – 39	5.0	6.2	9.2	3.6	4.9	7.5
40 – 59	4.1	(6.4)	(8.3)	3.2	(5.2)	(6.0)
60 – 79	(3.8)	(6.5)	(7.6)	(3.9)	(4.9)	(7.2)
80 – 100		(5.5)	8.1		(4.0)	6.9
40 – 100	4.1	6.4	8.1	3.3	5.0	6.8

Es sind keine linearen Veränderungen zu erkennen. Aber zwei Punkte lassen sich festhalten:

- Deutschsprachige Kinder schneiden in Klassen ohne Kinder anderer Muttersprache *nicht* besser ab als in Klassen mit Kindern anderer Muttersprache;
- Kinder mit anderer Muttersprache schneiden am schlechtesten ab, wenn ihr Anteil in der Klasse *unter* 10% liegt.

Auch wenn weitere Ergebnisse das Bild noch im Detail vereinheitlichen dürften⁴⁸, lässt sich als Gesamttendenz schon jetzt mit großer Sicherheit bestätigen, was wir in Kap. 4.2 kurz angedeutet haben: ein größerer Anteil von Kindern mit anderer Muttersprache bedeutet weder für diese Gruppe noch für die deutschsprachigen Kinder eine Gefährdung ihrer Leseentwicklung⁴⁹ (so auch für die Schweiz: Rüesch 1999, 43).

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 sinken, selbst wenn die beiden Teilgruppen (deutsch- und fremdsprachige Kinder) gleiche Leistungen wie sonst auch erbringen.

⁴⁷ RI_MIN_D : richtige Wörter pro Minute (deutschsprachige Kinder), RI_MIN_A : dto. (andere Muttersprache)

⁴⁸ evtl. kleine Änderungen in den eingeklammerten Zellen, bedingt durch zur Zeit noch zu geringe Zahlen

⁴⁹ 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)

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 schlechtere Leseleistungen erbringen als ihre KlassenkameradInnen mit Migrationshintergrund...

6 Mädchen und Jungen im Vergleich

Wie schon in vielen anderen Studien auch festgestellt, lesen die Jungen *etwas* langsamer (Mitte vierter Klasse 8.6 : 9.2 bearbeitete Sätze pro Min.) und sie machen *etwas mehr* Fehler (4.4% : 4.1%)⁵⁰. 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 biologische Faktor „Geschlecht“.

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.
- Sie machen insbesondere mehr Fehler.
- Im Arbeitstempo gibt es aber kaum Unterschiede.

Anders als in der Lesegeschwindigkeit zeigt sich also in der Genauigkeit der zentrale Unterschied.

Im Detail und vor allem über die Klassenstufen hinweg ist das Bild aber nicht so eindeutig:

6.2 Die Leistungsunterscheide nach Jahrgangsstufen

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.

Mitte der **2. Klasse** schneiden die Jungen im Gesamtergebnis nur wenig schwächer⁵¹ ab. Sie bearbeiten genau so viele Sätze, machen aber rund 25% mehr Fehler:

⁵⁰ → Tab. 6.2a, 6.2b, 6.2c

⁵¹ Die Effektstärke von 0.14 weist auf einen auch statistisch wenig bedeutsamen Unterschied hin.

Tab 6.2a: Leistungsunterschiede zwischen Mädchen und Jungen in Klasse 2

2. Klassen nach: Geschlecht	aM (SD)	Roh-Punkte	richtig = bearbeitet - falsch	Fehler-Quote	Bearbeitete Sätze pro Minute	Richtige Sätze pro Minute	Sek pro richtigem Satz
Jungen N = 1.049	24.7 (13.2)					4.1 (2.2)	14.6
			25.2 (13.4)	14.5 (16.3)	4.8 (2.2)		
Mädchen N = 995	26.1 (13.6)					4.4 (2.3)	13.6
			26.5 (13.9)	11.4 (15.1)	4.8 (2.2)		

Mitte der **3. Klasse** bearbeiten die Jungen weiterhin genau so viele Sätze wie die Mädchen, sie machen aber auch jetzt noch 25 % mehr Fehler. Da die Fehlerquote aber in beiden Gruppen erheblich abgenommen hat, fällt der Unterscheid zwischen beiden Gruppen noch geringer⁵² aus als in Klasse 2:

Tab 6.2b: Leistungsunterschiede zwischen Mädchen und Jungen in Klasse 3

3. Klassen nach: Geschlecht	aM (SD)	Roh-Punkte	Bearbeitet - Fehler	Fehler-Quote	Bearbeitete Sätze pro Minute	Richtige Sätze pro Minute	Sek pro richtigem Satz
Jungen N = 1.080	30.2 (11.9)					6.0 (2.4)	10.0
			30.2 (12.1)	6.8 (8.6)	6.4 (2.3)		
Mädchen N = 1.064	30.9 (11.5)					6.2 (2.3)	9.7
			30.8 (11.5)	5.5 (8.3)	6.5 (2.2)		

Mitte der **4. Klasse** machen die Jungen nur noch knapp 10% mehr Fehler als die Mädchen, aber sie arbeiten langsamer. Auch im Gesamtergebnis sind ihnen die Mädchen damit um rund 10% voraus⁵³:

Tab 6.2c: Leistungsunterschiede zwischen Mädchen und Jungen in Klasse 4

4. Klassen nach:	aM (SD)	Roh-Punkte	Bearbeitet - Fehler	Fehler-Quote	Bearbeitete Sätze	Richtige Sätze	Sek pro richtigem
------------------	---------	------------	---------------------	--------------	-------------------	----------------	-------------------

⁵² Effektstärke = 0.08

⁵³ Aber die Effektstärke von 0.26 weist auch diesen Unterschied als noch geringfügig aus.

Geschlecht					pro Minute	pro Minute	Satz
Jungen	N = 1.147	32.3 (10.8)				8.1 (2.7)	7.4
	N = 801		33.2 (10.7)	4.4 (5.6)	8.6 (2.6)		
Mädchen	N = 1.202	35.1 (11.8)				8.8 (2.9)	6.8
	N = 851		35.4 (11.4)	4.1 (5.8)	9.2 (2.8)		

Insgesamt überraschen die geringen Unterschiede. Vermutlich lassen sie sich damit erklären, dass Mädchen bei Aufgaben unter Zeitdruck generell schwächere Leistungen zeigen, während Jungen von dieser Testbedingung weniger beeinflusst sind (vgl. Ratzka 2003, i.V.).

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

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

7 Familienkonstellation

Kinder, die mit nur einem Elternteil zusammen leben (knapp 5%) unterscheiden sich in ihren Leistungen nur wenig von anderen Kindern. In der zweiten Klasse sind sie kaum, in der vierten Klasse geringfügig schwächer (7.8 : 7.1 sek pro richtigem Satz)

⁵⁴

7.1 Die Ausgangsfrage

Immer wieder wird die Besorgnis geäußert, dass Kinder alleinerziehender Eltern in ihrer Entwicklung gefährdet werden. Insbesondere ist eine durchgängig höhere Streuung innerhalb der Gruppe der allein erzogenen Kinder zu erwarten, da sie nach anderen Kriterien sehr heterogen ist. Insbesondere Insofern die Korrelation mit sozio-ökonomischen Bedingungen in einer bekanntermaßen starken Teilgruppe (vgl. Geißler 2002, 417 ff.) lässt 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

Weniger als 5% der Kinder unserer Stichprobe leben mit nur einem Elternteil zusammen. In unserer Stichprobe hat diese besondere Familienkonstellation nur eine geringe Bedeutung für die Leseleistung, wie die Ergebnisse der **2. Klassen** zeigen ⁵⁵ :

Tab. 7.2a Leistungen in der 2. Klasse nach Familienkonstellation

2. Klassen nach Familiensituation	aM (SD)	Roh-Punkte	Bearbeitet - Fehler	Fehler-Quote	Bearbeitete Sätze pro Minute	Richtige Sätze pro Minute	Sek pro richtigem Satz
zwei Eltern N = 1.932	25.5 (13.5)	26.0 (13.8)	13.0 (16.0)	4.8 (2.2)	4.2 (2.2)	14.3	
alleinerziehend N = 163	22.8 (12.0)	23.1 (12.7)	4.3 (2.1)	3.8 (2.0)	15.8		
	N = 132						

Auch Mitte der **3. Klasse** sind die Unterschiede marginal ⁵⁶ :

⁵⁴ → Tab. 7.2a, 7.2b, 7.2c

⁵⁵ Effektstärke = .18

Tab. 7.2b Leistungen in der 3. Klasse nach Familienkonstellation

3. Klassen nach Familiensituation	aM (SD)	Roh-Punkte	Bearbeitet - Fehler	Fehler-Quote	Bearbeitete Sätze pro Minute	Richtige Sätze pro Minute	Sek pro richtigem Satz
zwei Eltern N =	2.089	30.7 (11.6)				6.1 (2.3)	9.8
	N = 1.625		30.7 (11.7)	6.1 (8.2)	6.5 (2.3)		
alleinerziehend N =	114	27.9 (11.4)				5.6 (2.3)	10.7
	N = 93		27.8 (11.4)	7.8 (11.4)	5.9 (2.1)		

Die Ergebnisse **Mitte 4.** Klasse bestätigen die Befunde der vorerghenden Klassenstufen: Kinder allein erziehender Eltern brauchen am Ende der Grundschulzeit 10% länger für einen richtigen Satz ⁵⁷.

Tab. 7.2c Leistungen in der 4. Klasse nach Familienkonstellation

4. Klassen nach Familiensituation	aM (SD)	Roh-Punkte	Bearbeitet - Fehler	Fehler-Quote	Bearbeitete Sätze pro Minute	Richtige Sätze pro Minute	Sek pro richtigem Satz
zwei Eltern N =	2.127	34.0 (11.5)				8.5 (2.9)	7.1
	N = 1.480		34.6 (11.2)	4.1 (5.7)	9.0 (2.8)		
alleinerziehend N =	234	30.7 (10.2)				7.7 (2.6)	7.8
	N = 184		31.0 (10.4)	5.0 (5.6)	8.1 (2.5)		

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 Streuung 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 ⁵⁸ eher größere Unterschiede erwartet.

⁵⁶ Effektstärke = .22

⁵⁷ Mit einer Effektstärke von .28 nähert sich der Unterschied allmählich einer mäßig bedeutsamen Differenz. Sie entspricht etwa drei, vier Schulmonaten durchschnittlicher Entwicklung.

⁵⁸ Vgl. zur Korrelation mit wirtschaftlichen Notlagen in einer bedeutsamen Teilgruppe: Geißler 2002, 276 ff.

8 Unterschiede zwischen Bezirken und Klassen

8.1 Unterschiede zwischen Schulamtsbezirken

[noch in Arbeit: bisher zu kleine Auswertungsgrundlage]

Mitte der **2.Klasse** schneiden die SchülerInnen aus dem Kreis Siegen-Wittgenstein am besten ab. Noch auffälliger ist die geringere Streuung (fast nur halb so groß wie in den beiden anderen Bezirken):

2. Klassen nach Bezirken	N = aM (SD)	Roh-Punkte	Bearbeitet - Fehler	Fehler-Quote	Bearbeitete Sätze pro Minute	Richtige Sätze pro Minute	Sek pro richtigem Satz
Rhein-Sieg-Kreis							
Siegen-Wittgenstein							
Märkischer Kreis							

Mitte der **3. Klasse** liegen dagegen alle drei Bezirke nahe beieinander. Doch auch dieses Mal ist die Streuung in Siegen-Wittgenstein wesentlich geringer als in den beiden anderen Bezirken:

3. Klassen nach Bezirken	N = aM (SD)	Roh-Punkte	Bearbeitet - Fehler	Fehler-Quote	Bearbeitete Sätze pro Minute	Richtige Sätze pro Minute	Sek pro richtigem Satz
Rhein-Sieg-Kreis							
Siegen-Wittgenstein							
Märkischer Kreis							

Die Unterschiede zwischen den Schulbezirken betragen am **Ende der Schulzeit** 10% bzw. 20% -- wie schon im 2. Jahrgangsstufe mit leichten Vorteilen für Siegen-Wittgenstein, wo auch dieses Mal die Streuung deutlich geringer als in den beiden anderen Bezirken.

4. Klassen nach Bezirken	N = aM (SD)	Roh-Punkte	Bearbeitet - Fehler	Fehler-Quote	Bearbeitete Sätze pro Minute	Richtige Sätze pro Minute	Sek pro richtigem Satz
Rhein-Sieg-Kreis							
Siegen-Wittgenstein							
Märkischer Kreis							

Wie man diese Differenzen inhaltlich einzuschätzen hat, werden wir erst sagen können, wenn die Teil-Stichproben größer sind. Auffallend ist die geringe Streuung im Kreis Siegen, was für eine homogenere Schülerschaft als in den beiden anderen (großstadtnäheren?) Bezirken spricht.

Deutlicher sind auf allen Stufen die Unterschiede zwischen den *Klassen* eines Jahrgangs.

8.2 Unterschiede zwischen einzelnen Klassen

Zum Teil dramatisch unterscheiden sich die Mittelwerte der einzelnen Klassen: Die beste Klasse 2 ist leistungstärker als die schwächste Klasse 4⁵⁹. 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 sehr bedeutsamer Anteil der Unterschiede auf Differenzen im Unterricht zurückzuführen sein.

Im einzelnen liegen die Extreme der Durchschnittswerte einzelner Klassen für die Zahl richtig korrigierter Sätze pro Minute zwischen

Tab 8.2a Unterschiede der Mittelwerte von Klassen pro Jahrgang

Jahrgang	richtige Sätze pro Minute	Sekunden pro Satz
2	2.0 bis 7.1	8.4 bis 30.0
3	4.0 bis 9.3	6.4 bis 15.0
4	5.8 bis 12.9	4.6 bis 10.3

Die Übersicht zeigt: Die schwächste vierte Klasse schneidet schlechter ab als die beste zweite Klasse. Aber solche schlagzeilenträchtigen Relationen sind nur vernünftig interpretierbar, wenn man die Voraussetzungen der SchülerInnen beider Klasse am Schulanfang und wenn man die Unterrichtsbedingungen während der Schulzeit kennt. Aber 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 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).

⁵⁹ → Abb. 8.2a

9 Beziehung zwischen Testergebnissen und Noten

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 gelten⁶⁰. Interessant ist aber, dass die Korrelationen *innerhalb einzelner* Klassen – von einigen Besorgnis erregenden Ausnahmen abgesehen⁶¹ – 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.

9.1 Korrelationen über die Kinder aller Klassen hinweg

In dieser Zwischenauswertung konnte die Übereinstimmung von Lesenote und Testergebnis für eine Teilstichprobe von 3.127 Kindern (nur 3. und 4. Klasse) überprüft werden. Über die 140 Klassen hinweg ergab sich für die SchülerInnen der dritten Klassen (N = 1.379) eine Korrelation von .63, in den vierten Klassen (N = 1.788) eine Korrelation von .53. 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 wichtig, die Korrelationen zusätzlich für jede Klasse einzeln zu berechnen (vgl. Köller 2002). Dabei erhält man folgende Verteilung:

⁶⁰ → 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.

⁶¹ 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.

Tab 9.2a Korrelationen zwischen Lesenote und Testleistung

Korrelation Lesenote – 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 betrifft⁶². Allerdings weichen Noten und Testergebnis in einem Fünftel der 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.

Tab. 9.3a Klassifikation von Kindern nach Niveaus der Testleistung und Noten

Note	1	2	3	4	5	6
Niveau ⁶³						

⁶² 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

⁶³ Definition der Niveaus nach Zahl der richtig bearbeiteten Wörter pro Minute:

If (rich_min > 11.99)	niveau = 1.
If (rich_min > 5.99 and rich_min < 12)	niveau = 2.
If (rich_min > 4.99 and rich_min < 6)	niveau = 3.
If (rich_min > 2.99 and rich_min < 5)	niveau = 4.
if (rich_min > 1.99 and rich_min < 3)	niveau = 5.
if (rich_min < 2)	niveau = 6.

1	Klasse 3 (N = 12)	25.0%	66.7%	8.3%	-	-	-
	Klasse 4 (N = 244)	30.3%	57.8%	9.4%	2.5%	-	-
2	Klasse 3 (N = 668)	17.7%	53.4%	25.7%	3.0%	0.1%	-
	Klasse 4 (N = 1.204)	9.1%	47.3%	35.4%	8.3%	-	-
3	Klasse 3 (N = 263)	2.7%	36.1%	53.6%	7.2%	0.4%	-
	Klasse 4 (N = 132)	1.4%	15.5%	54.2%	28.2%	0.7%	-
4	Klasse 3 (N = 342)	0.9%	17.0%	52.9%	27.2%	2.0%	-
	Klasse 4 (N = 130)	-	13.1%	39.2%	44.6%	3.1%	-
5	Klasse 3 (N = 67)	-	1.5%	26.9%	61.2%	9.0%	1.5%
	Klasse 4 (N = 7)	-	-	14.2%	85.8%	-	-
6	Klasse 3 (N = 27)	-	3.7%	18.5%	63.0%	11.1%	3.7%
	Klasse 4 (N = -)	-	-	-	-	-	-

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.

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 rund die Hälfte der Urteile auf dieselbe Note entfallen und in jeweils rund 70% der Fälle eine Differenz von nur einer Note vorliegt. Sie verdeutlicht aber, dass die Noten in der Regel über vier bis fünf Stufen streuen.

Nicht minder interessant ist analog der Vergleich in der Spalten: Kinder mit der derselben Note schneiden im Test ganz unterschiedlich ab.

Es ist wichtig darauf hinzuweisen, dass sich die Note nicht auf die Testleistung bezieht, dass also NICHT dasselbe Testergebnis unterschiedlich bewertet wurde. Insofern können Abweichungen in zwei Richtungen gedeutet werden: Der punktuelle Test erfasst nur unzureichend, wie gut ein Kind lesen kann⁶⁴, LehrerInnen dagegen urteilen valider, denn sie können auf ein breiteres Spektrum an Beobachtungen zurückgreifen; oder aber: die Testleistung und die Auswertung sind verlässlicher, weil – von situativen Besonderheiten abgesehen -- die gleiche Aufgabe gestellt und die gleichen Maßstäbe bei der Beurteilung 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 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

⁶⁴ Vgl. auch die eindrucksvolle Darstellung von Ratzka (2003) zu ihrer Untersuchung mit dem TIMS-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.

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. So 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.

Dafür können auch Hinweise auf gruppenspezifische Verzerrungen hilfreich sein. So deuten erste Analysen darauf hin, dass Kinder anderer Muttersprache bei gleicher Testleistung eher bessere Noten bekommen⁶⁵. Mädchen und Jungen dagegen erhalten im Durchschnitt gleiche Noten⁶⁶.

⁶⁵ In Klasse 3 beträgt die Effektstärke .24, in Klasse 4 liegt sie sogar bei .38

⁶⁶ Die Effektstärken liegen unter .10

10. Fazit

Am Ende der Grundschulzeit können 90% der Kinder unbekannte Sätze einigermaßen zügig (mindestens fünf Sätze/ Min.) und ohne größere Schwierigkeiten (weniger als 5% Fehler) lesen. Weitere 5% der Viertklässler lesen ebenfalls selbstständig, wenn auch recht langsam (12-15 sek pro Satz). Die letzten 5% können die Aufgabe entweder nur mit erheblichen Schwierigkeiten bewältigen (mehr als 15 sek pro Satz und im Durchschnitt mehr als 10% Fehler) oder sogar gar nicht richtig lesen (ca. 1%). Für die Einschätzung der Leseleistungen der 10-Jährigen *insgesamt* ist außerdem zu berücksichtigen, dass knapp 5% der Schulanfänger schon vor Schulbeginn oder im Laufe der Grundschulzeit auf Sonderschulen wechseln (in der Regel – zumindest 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 Kriterium mit 5 bis 15% anzusetzen.

In Bezug auf die große Mehrheit der Kinder, auch die mit unterdurchschnittlichen Leistungen, ist aber 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. Pädagogisch gesehen sind die Fortschritte jeder Teilgruppe bedeutsamer als die Abstände innerhalb der Gesamtgruppe. Das gilt für leistungsschwache SchülerInnen generell und für Migrantenkinder im Besonderen – und das muss auch Konsequenzen für die Leistungsbewertung haben. Erstaunlich ist, dass die schwachen SchülerInnen trotz ständig negativer Rückmeldung überhaupt noch Fortschritte machen. Aber wie unsere Ergebnisse für die unterste Leistungsgruppe zeigen, bleibt 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 am spätesten gestartet sind. Denn der entscheidende Faktor scheint für die meisten Kinder „mehr Zeit zum Lernen“ zu sein. Wenn in PISA von rund 25% leseschwachen SchülerInnen die Rede ist, bedeutet deren Schwierigkeiten beim *Textverständnis* also nicht, dass der Grundschulunterricht *generell Grundfertigkeiten* stärker üben muss.

Mehr getan werden muss auf jeden Fall für die leistungsschwächsten 5-10%. Die zentrale Frage ist allerdings, ob das im Anfangsunterricht und in der Förderung 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 wenig verbreiteten Lesefreude außerhalb der Schule macht außerdem deutlich, dass diese Ansatz auch für alle anderen Kinder, also für den Unterricht *insgesamt* eine große Bedeutung hat. Insbesondere die massiven Leistungs-

unterschiede -- schon in Klasse 2 -- erfordern darüber hinaus 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 auf die unterschiedlichen Voraussetzungen ihrer SchülerInnen einlassen und Raum für unterschiedliche Lernschritte geben. Unsere Analysen zur Rechtschreibentwicklung 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 €⁶⁷) produktive Prozesse in den Kollegien der beteiligten Schulen in Gang gesetzt werden konnten.

Dieses individuelle Nachdenken und der Austausch innerhalb von Kollegien und in Schulbezirken wird noch einmal zusätzlich angeregt werden können, wenn im Sommer unsere Auswertung auf Klassenebene fortgesetzt werden kann.

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:

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

Aktuelle Kommentare:

www.grundschulverband.de → „PISA/IGLU“

⁶⁷ 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

- 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.) (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.
- 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.).
- 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). [reprinted from 1990]
- 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.
- Ratzka, N. (2003): Mathematische Fähigkeiten und Fertigkeiten am Ende der Grundschulzeit – Empirische Studien im Anschluss an TIMSS. Dissertation im FB 2 der Universität: Siegen (i.V.).
- Richter, S. (1996): Unterschiede in den Schulleistungen von Jungen und Mädchen. Geschlechtsspezifische Aspekte des Schriftspracherwerbs 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.

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