

Grundlegende Leseleistungen und der „Karawanen-Effekt“ in der Grundschule

Zentrale Befunde aus dem Projekt LUST¹ an der Universität Siegen

von

Hans Brügelmann²

1. Ziele und Rahmenbedingungen 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 Befunde³ 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, den 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 IGLU ist (wie bei TIMSS und PISA) 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. Dieser Anspruch hat Folgen für Design und Methoden von entsprechenden Untersuchungen:

¹ Lese-Untersuchung mit dem Stolperwörter-Test (s. Anm. 5). 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.

² 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.

³ s. Kasten 1

- 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 hat der Grundschulverband ein Projekt der Universität Siegen unterstützt, in dem in drei Schulamtsbezirken Nordrhein-Westfalens allen LehrerInnen der 2., 3. und 4. Klassen ein einfacher Lesetest⁵ und – für die Erstdurchführung⁶ – auch deren Auswertung angeboten wurde.

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.

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

- 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 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.

⁴ 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.

⁵ Dieser Test ist von Wilfried Metze, Berlin, für 1. bis 4. Klassen entwickelt worden (vgl. seinen Beitrag in diesem Heft).

⁶ Dies war erforderlich, um den Schulen rasch Vergleichswerte für die Interpretation der eigenen Daten anbieten zu können.

- 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 voneinander unterscheiden – ohne dass diese Unterschiede im Alltag relevant sein müssen.

Diese Einschränkungen sind zu bedenken, wenn im Folgenden einzelne Befunde vorgestellt und diskutiert werden – erst recht aber bei der Interpretation von Testergebnissen einzelner Klassen oder Kinder.

Besonders wichtig aber ist folgender Punkt: 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. In dem so relativierten Sinne sind Begriffe wie „verbessert“, „abgenommen“ usw. zu verstehen.

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.

In unserer eigenen Auswertung haben wir deshalb beide Größen erfasst. Soweit wir in unseren Berechnungen auf die Selbstausswertungen 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.

Eine Langfassung dieses Beitrags, einschließlich des Literaturverzeichnisses und eines Anhangs mit allen statistischen Kennwerten findet sich auf der Homepage www.grundschulverband.de unter → „PISA/IGLU“.

3. Ergebnisse der Jahrgänge im Vergleich

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 Metzke am Ende des ersten Schuljahres zurückgreifen. Um einen Maßstab für die Zielperspektive zu gewinnen, haben wir den Test selbst mit einigen Studierenden und LehrerInnen durchgeführt – ein sehr anspruchsvoller Maßstab, wenn man bedenkt, dass immer noch weniger als 30% eines Jahrgangs die allgemeine Hochschulreife erwirbt. Die folgende Tabelle zeigt die Bezugswerte im Überblick:

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.

Die Veränderung der Durchschnittsleistungen von Klasse 2 bis 4

Eine Übersicht über die erreichten Leistungen zu verschiedenen Zeitpunkten und den deutlichen Fortschritt von Klassenstufe zu Klassenstufe gibt die folgende Tabelle:

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

Mitte 2. Klasse ist in den vielen Klassen der Leselehrgang „abgeschlossen“, wie es in den meisten Lehrplänen heißt.

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 *Lesegeschwindigkeit*.

Drittklässler bearbeiten in fünf Minuten 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.

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.

Die Streuung der Leistungen auf den Jahrgangsstufen 2 bis 4

Die folgende Tabelle schlüsselt die Werte für den zweiten Jahrgang nach Leistungsgruppen (Prozentrang 1-5, 6-10 usw.) auf:

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 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 .

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 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**:

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. Der Abstand zu den oberen 15% ist damit qualitativ geschrumpft, wenn man einmal von den allerschwächsten LeserInnen absieht.

Und die Streuung in **Klasse 4**?

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).

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?

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).

Im Folgenden haben wir die Anteile der Jahrgänge in vier verschiedenen Kategorien ausgewiesen, so dass man sich je nach Maßstab ein eigenes Bild vom Niveau der SchülerInnen im unteren Leistungsbereich machen kann.

Jahrgangvergleich der Anteile (korrigiert) im unteren Leistungsbereich				
richtige Sätze pro Minute Jahrgang	0-2	0-3 ⁷	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 %

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 erhält man wenn man einzelne Leistungsgruppen betrachtet. Auch dann zeigt sich: Alle Gruppen machen Fortschritte ⁸, und proportional am stärksten die unteren Leistungsgruppen.

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. 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 5.2 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.

⁷ Die Werte sind kumulativ berechnet, schließen also die jeweils vorhergehende Gruppe mit ein.

⁸ 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 152 Kindern stützen allerdings unsere Annahme: eine Korrelation der Leistungen im Januar und im Juli von .91** spricht dafür, dass sich die Ränge der einzelnen Kinder trotz eines durchschnittlichen Leistungszuwachses von 1.9 richtigen Sätzen kaum verändern.

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 Hinweis sollen aber keinesfalls die in der untersten Leistungsgruppe beobachteten Schwierigkeiten verdrängt werden. Auch wenn die Größenordnung von rund 5% besonders gefährdeten SchülerInnen⁹ (in LUST) 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 dürfen¹⁰. 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 selbst auf diesem Niveau relativ besser als die 15-Jährigen abschneiden, haben die IGLU-Ergebnisse gezeigt (vgl. Grundschulverband aktuell Nr. 83).

5. Befunde zu einigen besonderen Aspekten

Kinder, deren Eltern eine andere Muttersprache sprechen (20-22%), 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. 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 manchmal gar keinen?) Fortschritten – gegenüber rund 10% unter den deutschsprachigen Kindern. 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.

⁹ Auf die Altersgruppe insgesamt bezogen: plus vermutlich weiteren 5%, die den Jahrgang (durch Sitzenbleiben) oder die Grundschule (durch Überweisung in die Sonderschule) verlassen haben.

¹⁰ 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.

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%). 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.

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).

Zum Teil dramatisch unterscheiden sich die Mittelwerte der einzelnen **Klassen**: Die beste Klasse 2 ist leistungsstärker als die schwächste Klasse 4. 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.

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 ¹¹. Interessant ist aber, dass die Korrelationen *innerhalb einzelner* Klassen – von einigen Besorgnis erregenden Ausnahmen abgesehen – 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.

6. Zusammenfassung und Folgerungen

Am Ende der Grundschulzeit können 90% der Kinder unbekannte Sätze einigermaßen zügig (mindestens fünf Sätze/ Min.) und ohne größere Schwierigkeiten (weniger als 10% Fehler) erle-

¹¹ Allerdings kann eine niedrige Korrelation 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.

sen 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 ¹².

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 *Teilgruppe* bedeutsamer als die *Abstände* innerhalb der *Gesamtgruppe*. 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.

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 viel-

¹² S. oben Kasten 1 .

mehr, 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 machen 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 *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 €¹³) 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-Lesetest kann nach den Erfahrungen in unserer LUST-Studie von LehrerInnen leicht selbst durchgeführt und ausgewertet werden. Eine differenziertere Version des Tests zur genaueren Analyse des Leistungsprofils einzelner Kinder wird von uns gegenwärtig entwickelt.

Literaturnachweise

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.)

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.

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.

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

[oder:] Das Literaturverzeichnis wird auf Anfrage an den Autor zugeschickt. Es kann auch direkt eingesehen werden in der Langfassung auf der Homepage des Grundschulverbands.

¹³ 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.