Peter M. Roeder und Fritz Sang

„Über die institutionelle Verarbeitung von Leistungsunterschieden“

Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie 2/1991, S.159-170.

Professor Dr. Peter M. Roeder war 1973 als Direktor an das MPIB berufen worden und seit 1983 verantwortlich für den Forschungsbereich „Schule und Unterricht. Sein Mitarbeiter Dr. Fritz Sang war als Mathematiker zuständig für die Berechnungen und Roeder war zuständig für den Text des Aufsatzes. Das konnte durch Anfragen in Erfahrung gebracht werden.

Diese Untersuchung aus dem Jahre 1991 basiert ebenfalls auf den Daten des -Projektes „Schulleistung“. Das war in den Jahren 1968 bis 1970 am „Max-Planck-Institut für Bildundsforschung (MPIB)“ durchgeführt worden. Die Ernte: *„verarbeitbare Datensätze für 12594 Schüler aus 427 Klassen und für insgesamt 1130 Deutsch-, Englisch- und Mathematiklehrer“*, mit Auskünften zu ihren Unterrichtsstrategien, anhand deren *„Kennziffern zur Quantität und Qualität der Instruktion“* gebildet werden konnten (Baumert, Roeder Sang und Schmitz 1986, S.644). Unter den Schülern waren 1049 Gymnasiasten aus Berlin und Bremen, die erst nach sechs Jahren Grundschule aufs Gymnasium überwechseln konnten.

Der Aufsatz enthält wichtige Erkenntnisse zum niedrigeren Lerntempo von undifferenzierten oder unzureichend differenzierten Klassen und zur unterschiedlichen Effizienz der verschiedenen Unterrichtsstile.

Im Jahre 1972 hatte der schwedische Wissenschaftler Ulf P. Lundgren nachgewiesen, dass Lehrer *„sich in ihrem Unterrichtsverhalten in vieler Hinsicht an einer Gruppe von Schülern orientieren, die in der Intelligenzverteilung innerhalb der Klasse zwischen dem 10. und dem 25. Perzentil angesiedelt sind.“* Den Hinweis auf diese von Lundgren entdeckte **„Steuerungsgruppe“** nimmt Roeder zum Anlass, mitzuteilen, dass Lundgrens Entdeckung auch für Deutschland gelte (s.163).

Der erste Satz der Zusammenfassung des Aufsatzes (S.168), der in keiner Weise den Befunden entsprach, ist ein erstaunlich freimütiges und eindeutiges Bekenntnis zu der am MPIB praktizierten zurückhaltenden Informationspolitik: *„Wir wollen diesen Aufsatz bewusst nicht mit einem Plädoyer für die vier jährige Grundschule beschließen...“* Dieser Satz brachte den Text um seine Wirkung.

Ins Präteritum gesetzt, ist er ein Geständnis: *„Wir wollten diesen Aufsatz bewusst nicht mit einem Plädoyer für die vier jährige Grundschule beschließen...“*

In den anschließenden Sätzen lässt Roeder dann durchblicken, dass er den Aufsatz eigentlich mit einem Plädoyer für die vier jährige Grundschule hätte enden lassen müssen. Denn ein „advanced placement“ erhalten an amerikanischen High Schools nur jene Schüler, denen anspruchsvollere Aufgaben und Aufträge zugemutet werden können. Mit diesem Fachausdruck wird von Roeder der Leistungsvorsprung jener Gymnasiasten beschrieben, welche nach vierjähriger Grundschule im 5. und 6. Jahrgang schon das Gymnasium besuchen konnten. Er lag in Mathematik und Englisch mit *„etwa einer Standardabweichung“* bei etwa zwei Schuljahren und in Deutsch bei etwa einem Schuljahr.

In seinem Aufsatz *über „Entwicklung vor, während und nach der Grundschulzeit“* war dieser Aufsatz von 1991 für Roeder die einzige ihm bekannte *„repräsentative Studie“* zu der Frage, *„inwieweit sich die Erwartungen erfüllen, die sich mit der Verlängerung der Grundschuldauer verbinden“* (1997, S.406f)

Unsere Frage: Wer oder was hinderte die Wissenschaftler daran, nach Kant *„von ihrer Vernunft öffentlichen Gebrauch zu machen“*?

Im Jahre 2002, mehr als 30 Jahre nach Abschluss der Datenerhebung, kommentieren dann sein Nachfolger Jürgen Baumert und Olaf Köller jene von ihm auf Seite 167 veröffentlichte Tabelle wie folgt:

„Unübersehbar zeigen sich deutliche Leistungsvorteile der Gymnasiasten aus den Ländern mit vierjähriger Grundschule. Die frühere Differenzierung scheint also leistungsstärkere Schüler in allen drei Fächern besser zu fördern.“ (Oerter/Montada 2002, S.470, gleichlautend im Oerter/Montada 2008, S.450)

Weitere Informationen unter „MPIB-Projekte“ / „Die Geschichte der Tabelle 1“

Zeitschrift f. Entwicklungspsychologie u. Pädagogische Psychologie
1991, Band XXIII, Heft 2, 159—170

1991

Über die institutionelle Verarbeitung von Leistungsunterschieden

Peter M. Roeder und Fritz Sang
Max-Planck-Institut für Bildungsforschung, Berlin

Fast alle entwickelten nationalen Schulsysteme reagieren mit Formen äußerer und curricularer Differenzierung auf die Heterogenität der Lernvoraussetzungen. Wo dies nicht geschieht, entwickelt sich in privaten Schulsystemen ein funktionales Äquivalent für solche organisatorische Differenzierung. Die häufig als Alternative empfohlene Binnendifferenzierung hat, wie die Unterrichtsforschung zeigt, im deutschen Sekundarschulwesen keine Tradition. Es fehlt deshalb auch eine empirisch begründete Diskussion der Hypothese, daß längerfristig durchgeführte Binnendifferenzierung zu einer vergleichbar starken Auseinanderentwicklung der Schulleistungen führt wie äußere Differenzierung. Daß aber auch äußere Leistungsdifferenzierung den Lehrer nicht der Aufgabe enthebt, eine Balance zwischen Chancenausgleich — operationalisiert als Verringerung der Varianz der Schulleistungen in der Klasse — und Leistungsförderung zu finden, wird am Beispiel der Leistungsentwicklung von Schülern des 7. Schuljahrs von Gymnasien demonstriert. Da die Stichprobe der Schulen auch Gymnasien enthält, die Schüler nach sechs Jahren Grundschule aufnehmen, läßt sich in einem Quasi-Experiment der enge Zusammenhang zwischen Leistungserwartungen der Lehrer, die nach sechsjähriger Grundschule deutlich niedriger liegen, ihren curricularen Planungen für das 7. Schuljahr und den entsprechend unterschiedlichen Schülerleistungen sowohl zu Beginn als auch noch am Ende des 7. Schuljahrs nachweisen.

Auch in selektiven Sekundarschulsystemen ist also das Problem der Heterogenität der Lernvoraussetzungen nicht schon gelöst und sehen sich Lehrer einer großen Spannbreite von Leistungen und Interessen in ihren Klassen konfrontiert. Das gilt selbst für die Berliner Hauptschule, die nur noch von 10 % der deutschen Schüler eines Schuljahrgangs besucht wird, wie schon die Abgangsstatistiken belegen. Immerhin erreichen 10 % ihrer Schüler noch die Mittlere Reife, während auf der anderen Seite des Leistungsspektrums weitere 10 % nicht einmal den Hauptschulabschluß schaffen. Wie groß die Spannbreite der Lernvoraussetzungen ist, verdeutlichen auch unsere Beobachtungen in den Leistungskursen des in drei Niveaus erteilten Englisch- und Mathematikunterrichts dieser fünf Hauptschulen. Die Lehrer reagierten darauf mit enormen Unterschieden in der Schwierigkeit der Aufgabenstellung und im Unterrichtstempo (Roeder, 1985). Das Ausmaß der Unterschiede in Lernvoraussetzungen verdeutlicht auch die Untersuchung von Barr und Dreeben über Binnendifferenzierung im Anfangsleseunterricht. Die Zahl der in den oberen Leistungsgruppen aus der Fibel gelernten Wörter betrug im allgemeinen ein Vielfaches der in den unteren Leistungsgruppen gelernten (Barr & Dreeben, 1983, Tabelle 6.2; Sørensen & Hallinan, 1986).

Dies dokumentiert eindrucksvoll, wie durch die Schule Leistungsunterschiede produziert werden, aber auch, welchen Anteil vorausliegende Unterschiede in den Lernvoraussetzungen daran haben und schließlich, wie schwierig die Aufgabe der Binnendifferenzierung ist, wenn mit ihr sowohl Egalisierung als auch Individualisierung des Leistungsfortschritts angestrebt werden. Es macht zugleich verständlich, warum die Mehrzahl der von uns befragten Hauptschul- und Gesamtschullehrer meinten, unter gegebenen Bedingungen Binnendifferenzierung nicht praktizieren zu können. Bei aller berechtigten Kritik an äußerer Leistungsdifferenzierung wäre es wohl gegenwärtig eine Illusion zu glauben, die Mehrzahl der Lehrer sei imstande, in der heterogenen Lerngruppe optimale Lernbedingungen für alle Schüler zu schaffen.

Lernvoraussetzungen der Schüler — Leistungserwartungen und curriculare Entscheidungen der Lehrer

Dieses skeptische Urteil wird mindestens teilweise durch Untersuchungen gestützt, die der Frage nach der Vereinbarkeit möglicherweise konfligierender, aber gleich wichtiger Unterrichtsziele nachgehen, nämlich der optimalen Förderung bei gleichzeitigem Leistungsausgleich auf der Ebene der Schulklasse. Daß wir hier möglicherweise einem Dilemma konfrontiert sind, wird schon durch Lundgrens Analysen der Rolle der „Steuerungsgruppe" im Unterricht nahegelegt. Er beobachtete, daß schwedische Oberschullehrer sich in ihrem Unterrichtsverhalten in vieler Hinsicht an einer Gruppe von Schülern orientieren, die in der Intelligenzverteilung innerhalb der Klasse zwischen dem 10. und dem 25. Perzentil angesiedelt sind. Diese Steuerungsgruppe bestimmt sozusagen das Tempo der Instruktion, aber auch viele andere Facetten der Interaktion zwischen Lehrer und Schülern (Lundgren, 1972). Diese Beobachtung impliziert, daß die Lehrer ihre Aufmerksamkeit auf die Schüler konzentrieren, deren Leistungen zumeist deutlich unter dem Durchschnitt liegen. Ihnen ist es wichtiger, daß im Unterrichtsverlauf niemand auf der Strecke bleibt, als die ohnehin leistungsstarken Schüler ihren Möglichkeiten entsprechend zu fördern. Zu erwarten wäre also eine Angleichung der Schülerleistungen in diesen Klassen, wobei die Reduktion der Leistungsvarianz auf Kosten der leistungsstarken Schüler erreicht wird.

Diese Erwartung wird durch eine ältere Untersuchung des MPI für Bildungsforschung bestätigt (Baumert, Roeder, Sang & Schmitz, 1986).

Es handelt sich hier um eine Längsschnittuntersuchung mit einer repräsentativen Stichprobe von über 400 7. Gymnasialklassen. Zu Anfang und am Ende des Schuljahres wurden die Schülerleistungen in den Fächern Deutsch, Englisch und Mathematik erhoben, und zwar mit Parallelformen von eigens zu diesem Zweck mit dem Ziel hoher Curriculumvalidität konstruierten Tests. Mit Hilfe eines umfangreichen Lehrerfragebogens wurden insbesondere Unterrichtsziele und unterrichtsmethodische Orientierungen und Praktiken erhoben. Der egalisierende Einfluß des Unterrichts läßt sich in dieser Studie als Entwicklung der Leistungsvarianz innerhalb der Klassen operationalisieren.

Die Leistungssteigerung im Laufe des Schuljahrs — unter Kontrolle der Eingangsleistung und der Intelligenz — kann als Indikator für die Förderungsleistung des Unterrichts bei Schülern unterschiedlicher Lernvoraussetzungen dienen.

Vergleicht man Klassen, in denen sich die Leistungsvarianz im Laufe des Schuljahrs vergrößerte, mit solchen, in denen sie abnahm, so ergibt sich für alle drei Schulfächer prinzipiell das gleiche Bild, obgleich es für Mathematik am schärfsten konturiert ist: Der Leistungszuwachs ist durchweg höher in den Klassen, in denen auch die Streubreite der Leistungen zunimmt. Beim Vergleich der Leistungsgruppen (drei Niveaus) zeigt sich, daß die Verringerung der Leistungsvarianz innerhalb der Klassen vor allem auf Kosten der leistungsstarken Schüler erreicht wird. In Klassen mit Leistungsausgleich im Mathematikunterricht bleiben sie um etwa eine Standardabweichung hinter dem leistungsbesten Drittel der Schüler in Klassen mit steigender Varianz zurück. Im Englisch- und Deutschunterricht beträgt diese Differenz noch etwa eine dreiviertel Standardabweichung. Dies sind sicher nicht nur statistisch, sondern auch praktisch höchst bedeutsame Unterschiede. Nur im Englisch- und Deutschunterricht werden sie durch höhere Leistungszuwächse der Schüler im unteren Leistungsdrittel in gewissem Umfang aufgewogen.

Eine etwas detailliertere Analyse zeigt ein nicht ganz so trübes Bild. Immerhin gelingt es einem Teil der Lehrer, trotz Leistungsausgleichs einen überdurchschnittlichen Leistungszuwachs — gemessen jeweils am Mittelwert unserer Stichprobe — in ihren Klassen zu erzielen. Dies trifft für etwa 16 % der Klassen im Englisch- und Mathematikunterricht zu und für einen fast doppelt so hohen Anteil der Klassen im Deutschunterricht. Auch bei diesem Vergleich schneiden Schüler in Klassen mit zunehmender Varianz besser ab, gelingt also der Leistungsausgleich nicht ganz ohne Verluste. Aber immerhin haben die Lehrer in diesen Klassen eine Balance zwischen den beiden konkurrierenden Zielsetzungen von Leistungsausgleich und -förderung gefunden, die für viele von ihnen akzeptabel sein dürfte.

Selbst unter den günstigen Bedingungen des Gymnasiums ist es also eine schwierige Aufgabe, eine befriedigende Balance zwischen optimaler Förderung und Leistungsausgleich zu finden. Sie wird nur von einer Minderheit der Lehrer wirklich bewältigt (vgl. auch Treiber & Weinert, 1985; Helmke, 1988; Baumert, Schmitz, Roeder & Sang, 1989).

Die Studie des MPI zur Leistungsentwicklung in 7. Gymnasialklassen erlaubt wegen der Größe der Stichproben zwar keine sehr detaillierte Analyse der Faktoren, die im Unterricht die Balance zwischen Leistungsförderung und Leistungsausgleich bedingen.

Tabelle 1 zeigt, daß die unterschiedlichen Leistungserwartungen und entsprechenden curricularen Entscheidungen der Lehrer sich deutlich in unterschiedlichen Schülerleistungen niederschlagen. ?

Die Testleistungen sind in Prozentpunktwerten angegeben; Unterschiede können also als Differenzen in Prozent des maximal erreichbaren Testwerts interpretiert werden. In der letzten Zeile sind die Residuen des Leistungszuwachses unter Kontrolle der Ausgangsleistung, der Intelligenz und des Familienbackgrounds wiedergegeben. Sie sind also ein möglichst direktes Maß für die Förderungsleistung des Unterrichts im 7. Schuljahr des Gymnasiums.

Die Unterschiede zwischen den beiden Gruppen, die wir hier vergleichen — Gymnasiasten nach sechsjähriger Grundschule und solche nach vierjähriger Grundschule plus zwei Jahre Gymnasium —, sind nicht nur höchst signifikant,

Tab. 1. Leistungsunterschiede und Unterschiede der Leistungsentwicklung zwischen Gymnasialschülern des 7. Schuljahres aus sechsjährigen Grundschulen und solchen aus vierjährigen Grundschulen plus zwei Jahre Gymnasium (N = 12.252 — insgesamt)

			Englisch	Mathematik	Deutsch	N
Anfang des 7. Schuljahres	sechsjährige Grundschule	$\bar{x}$	38.8	38.0	49.5	1.049
		s	9.4	11.9	11.6	
	vierjährige Grundschule	$\bar{x}$	52.6	47.1	54.8	11.203
		s	12.4	13.9	12.6	
Ende des 7. Schuljahres	sechsjährige Grundschule	$\bar{x}$	42.0	40.9	39.7	1.049
		s	10.1	11.9	10.4	
		Residuen	—.80	.70	—.84	
	vierjährige Grundschule	$\bar{x}$	53.9	46.0	44.5	11.203
		s	12.8	14.2	11.9	
		Residuen	.08	—.07	.08	

was bei einer Stichprobe von mehr als 12.000 Schülern kaum überraschen kann, sondern in einer Größenordnung, die auch praktisch bedeutsam ist.[1] Im Englischen beträgt der durchschnittliche Leistungsunterschied zu Beginn des Schuljahrs etwas mehr als eine Standardabweichung, im Mathematikunterricht etwa eine dreiviertel Standardabweichung, im Deutschen etwas weniger als eine halbe Standardabweichung. Die Unterschiede zwischen den Fächern sind plausibel: Der Deutschunterricht dürfte weniger linear und kumulativ aufgebaut sein als der Englisch- und Mathematikunterricht, und der Deutschtest mißt auch Leistungen, die nicht nur Ergebnis des Unterrichts sind, sondern auch von kulturellen Gewohnheiten des Elternhauses abhängen. Daß die Unterschiede im Englischen am größten ausfallen, dürfte u. a. auf den besonderen Stellenwert des Lehrbuchs in diesem Fach zurückzuführen sein: Das in der Grundschule normalerweise verwendete Buch hat einen wesentlich niedrigeren Schwierigkeitsgrad als die für das Gymnasium geschriebenen und in der Regel dort benutzten Lehrbücher.

Die zu Beginn der 7. Klasse bestehenden Leistungsunterschiede werden, wie die Tabelle zeigt, im Laufe des Schuljahres nur teilweise ausgeglichen. Dies ist nach den Aussagen der Lehrer über ihre Leistungserwartungen und ihre entsprechende Unterrichtsplanung bzw. -gestaltung auch nicht zu erwarten. Im Englischen liegt der Unterschied immer noch bei etwa einer Standardabweichung, bleibt also fast unverändert bestehen. Im Mathematik- und Deutschunterricht geht die Differenz auf etwa eine drittel Standardabweichung zurück, ist also immer noch auch praktisch bedeutsam.

Wie die Residualwerte für den Leistungszuwachs belegen, bleiben die Klassen nach sechsjähriger Grundschule im Englischen und im Deutschen etwas hinter dem durchschnittlichen Leistungsfortschritt zurück; in Mathematik liegen sie etwas darüber. Diese Unterschiede sind zwar statistisch, aber kaum noch praktisch signifikant. Beide Schülergruppen machen insgesamt durchaus vergleichbare Lernfortschritte, wodurch freilich auch die zu Beginn des Jahres bestehenden Leistungsunterschiede weitgehend erhalten bleiben. Die Aussagen der Lehrer über ihre Unterrichtsplanung für das 7. Schuljahr und über die Themen

---

[1] Da die Zuweisung der Schüler zur vier- bzw. sechsjährigen Grundschule nicht nach dem Zufallsprinzip erfolgt ist, stellt sich natürlich die Frage, ob die Unterschiede in der Fachleistung nicht auch auf Fähigkeitsunterschiede der Schüler zurückgeführt werden können, zumal in den Regionen mit sechsjähriger Grundschule ein höherer Anteil eines Altersjahrgangs aufs Gymnasium übertritt. Dieser Einwand kann mit zwei Befunden weitgehend entkräftet werden: Erstens besteht in unserer Stichprobe kein Zusammenhang zwischen relativem Gymnasialbesuch und Fachleistung (Roeder, Baumert, Sang & Schmitz, 1986). Zweitens sind die Intelligenztestwerte zwischen beiden Gruppen zwar hochsignifikant verschieden (durchschnittlicher IQ nach sechsjähriger Grundschule 108,5, SD 6,99, nach vierjähriger Grundschule 109,7, SD 7,44); aber die statistische Signifikanz ist eine Funktion der Größe der Stichprobe von über 12 000 Schülern. Praktisch dürfte dieser Unterschied bedeutungslos sein und sicher nicht ausreichen, die beträchtlichen Fachleistungsunterschiede zu erklären.