

Martin Wellenreuther

Konsequenzen aus PISA in unterrichtspraktischer Sicht

– empirisch geprüfte Modelle des Umgangs mit Heterogenität¹

Zusammenfassung: Zentrales Ergebnis der internationalen Vergleichsstudien ist, dass es Deutschland derzeit nicht gelingt, eine breite Elite zu fördern. Stattdessen ist die Gruppe der Schüler mit sehr gering entwickelten Kompetenzen besonders groß. Um diese Situation zu verbessern, werden verschiedene Methoden eines effektiven Umgangs mit der Heterogenität der Schüler diskutiert und anhand empirischer Untersuchungen erläutert. Der wichtigste Ansatzpunkt für eine Verbesserung der schulischen Lernergebnisse liegt in einer Optimierung des alltäglichen Unterrichts. Dabei spielen verschiedene Punkte zusammen: Eine andere Kultur des Erklärens und Einführens neuer Inhalte, eine breitere und tiefere Verankerung der zentralen Kenntnisse und Kompetenzen, wobei ein gänzlich anderer Umgang mit Tests erforderlich erscheint. Tests sollten vor allem Wissenslücken identifizieren, auf die im nachfolgenden Unterricht eingegangen werden kann („assessment for learning“).

Um die grundlegenden Prinzipien einer anderen Kultur des Lernens und Förderns zu verdeutlichen, wird nach der Diskussion einer Optimierung des „normalen“ Unterrichts auf verschiedene Methoden des Arbeitens mit Gruppen eingegangen. Zuerst wird die Gruppenrallye von Slavin behandelt, bei der im Anschluss an die Einführung neuer Inhalte im Rahmen direkter Instruktion in heterogenen Gruppen Aufgaben bearbeitet werden, wobei eine Gruppe nur gut abschneiden kann, wenn alle Schüler sich in den anschließenden individuellen Tests in ihren Leistungen verbessern. Eine andere Form der Gruppenarbeit findet bei dem Modell des Lesenslernens in Neuseeland statt. Hier werden in Klassen viele leistungshomogene Gruppen gebildet, die jeweils an Unterrichtsmaterialien ihrer Kompetenzstufe unter Anleitung des Lehrers arbeiten. Schüler können in eine höhere Kompetenzstufe aufsteigen, wenn sie durch einen individuellen Test gezeigt haben, dass sie die Aufgaben ihrer Kompetenzstufe sicher beherrschen.

Zum Abschluss wird diskutiert, welche Punkte eines solches Methodenarsenals sich am ehesten auf die deutsche Schule übertragen lassen. Dies scheint am leichtesten für eine Optimierung der direkten Instruktion möglich zu sein. Allerdings ist eine solche Optimierung kaum durch den einzelnen Lehrer leistbar, sondern setzt kontinuierliche und im Schulalltag fest institutionalisierte Kooperationen unter Lehrern voraus. In diesen Kooperationen sollten gemeinsam Unterrichtseinheiten zusammen mit Erklärungen, Aufgaben, Tests und Zusammenfassungen entwickelt, im Unterricht eingesetzt und aufgrund der gemachten Erfahrungen überarbeitet werden.

1. Problemstellung

1.1 PISA und die deutsche „Bildungsmisere“

Internationale Vergleichsuntersuchungen wie TIMSS² und PISA zeigen: In Deutschland wird eine vergleichsweise schmale Eli-

te gefördert, während die Gruppe der Schüler mit insgesamt niedrigem Leistungsniveau vergleichsweise groß ist. Zwar gibt es innerhalb der verschiedenen Bundesländer noch erhebliche Unterschiede, wobei sich Baden-Württemberg zu den deutschen Spitzenländern zählen darf. Ver-

¹ Der folgende Beitrag ist die schriftliche Fassung eines Vortrags, den ich am 2. Juni 2005 in Bad Urach auf der Baden-Württembergischen Regionaltagung des Bundesarbeitskreises der Seminar- und Fachleiter/innen e.V. gehalten habe.

² TIMSS steht für „Third International Mathematics and Science Study“, PISA für „Programmes International for Student Achievement“.

glichen mit Finnland sehen die Ergebnisse Baden-Württembergs aber immer noch eher bescheiden aus: In Baden-Württemberg gelangen 10% in die höchste Kompetenzstufe im Lesen (Finnland 18%), in der niedrigsten Kompetenzstufe befinden sich allerdings etwa doppelt so viele Schüler (19%), verglichen mit 10% in Finnland. Dieses Ergebnis ist deshalb schwer hinnehmbar, weil Jugendliche mit niedrigem Kompetenzniveau größere Probleme haben werden, ein selbstbestimmtes Leben zu führen als Jugendliche mit hohem Kompetenzniveau.

1.2 Konkrete Bedingungen der deutschen „Bildungsmisere“

Bevor man Maßnahmen zur Reform des Bildungswesens ergreift, sollte man konkrete Bedingungen und zu vermutende tiefere Ursachen genauer analysieren. Leider liefern die durchgeführten internationalen Vergleichsstudien dafür nur begrenzt Anhaltspunkte.³ Man kann nur vergleichend prüfen, in welchen Bereichen Deutschland im Vergleich zu den bekannten Spitzenländern Defizite aufweist.

Folgende konkrete Bedingungen spielen m. E. eine Rolle:

(1) **Halbtagschulsystem in Deutschland, in PISA-Spitzenländern Ganztagschulsystem:** Dadurch erhalten deutsche Schüler in manchen Fächern deutlich weniger Unterricht als in anderen Ländern (in Mathematik z. B. bis zur 10. Klasse ein ganzes Schuljahr weniger als in Japan, vgl. Helmke & Hesse 2002, S. 453f.) Zweitens bestehen in Ganztagschulen günstigere Voraussetzungen für ein verteiltes Üben schulischer Inhalte am Nachmittag. Vor allem Jungen sind zu Hause nachmittags schwer für die Erledigung schulischer Arbeiten zu motivieren. Sie erliegen leicht der Versuchung, stattdessen am Computer zu spielen. Die in der Schule gelernten Inhalte werden dann im Gedächtnis überschrie-

ben, eine feste Verankerung unterbleibt (vgl. Scheich 2005, S. 33). Drittens ergeben sich bei einem Ganztagsbetrieb für Lehrer leichter Möglichkeiten, miteinander zu kooperieren (vgl. Wellenreuther 2005, S. 476ff.).

(2) **Keine systematische Qualitätssicherung der Schulbücher und Lehrmaterialien.** In Deutschland werden Schulbücher und Lehrmaterialien in der Regel formal überprüft, nicht aber bezüglich der für das Lernen relevanten Fragen. Es interessiert die Kultusbürokratien nicht, wie verständlich Erklärungen in Schulbüchern sind, wie gut Inhalte aufeinander aufbauen, ob das Wesentliche zusammen gefasst wird, ob Tests zur Selbstüberprüfung enthalten sind, ob sich die Lernmaterialien zum Nachbereiten oder eigenständigen Erarbeiten von Inhalten eignen. Daraus ergeben sich dann auch eingeschränkte Möglichkeiten einer effektiven individuellen Förderung aller Schüler.

(3) **Falsche pädagogische Konzeption des Lernens zentraler Inhalte:** In Deutschland werden bestimmte kognitive Inhalte jedes Jahr erneut, aber oberflächlich behandelt, wobei diese Vorgehensweise mit dem Argument begründet wird, dadurch die Inhalte immer tiefer verstehen zu können (Spiralcurriculum). In Wahrheit werden die Inhalte dann jedes Schuljahr fast völlig neu und unvollkommen gelernt und dann wieder vergessen. Es wäre viel ökonomischer, die Inhalte beim ersten Mal viel intensiver und länger zu behandeln, bis sie fest im Langzeitgedächtnis integriert sind. Dies würde insgesamt Unterrichtszeit einsparen.

(4) **Veraltetes System schulischer Leistungsmessung:** Unser System der Leistungsmessung bezieht sich vorrangig auf die Feststellung von Leistungsunterschieden zwischen Schülern einer Klasse (summativ Leistungsmessung mit Hilfe normbezogener Tests) und nicht auf die

³ Zur Problematik von Ursachenanalysen im Rahmen von internationalen Vergleichsuntersuchungen vgl. Wellenreuther 2005, Kap. 1 und 2.

Diagnose des Lernstands der Schüler, um den nachfolgenden Unterricht daran zu orientieren (formative Leistungsmessung, „assessment for learning“; standardbezogene, kriterienorientierte Tests). Die Leistungsmessungen werden hier nicht systematisch in den Dienst einer Verbesserung des Unterrichts gestellt (vgl. Black & Wiliam 1998 a, b). Ferner taugt dieses System nicht zum präzisen und objektiven Erfassen von Kompetenzen. Außerdem werden keine standardbezogenen externen Leistungsprüfungen durchgeführt, die unter bestimmten Bedingungen auch Aussagen über die pädagogische Wirksamkeit von Lehrern erlauben würden. Bei niedriger Lehrerleistung könnten die Ursachen dafür genauer diagnostiziert werden, um daraufhin maßgeschneiderte Unterstützungsleistungen (z. B. bestimmte Weiterbildungsangebote) anbieten zu können.⁴

1.2 Vermutete tiefere Ursachen

Hinter diesen konkreten Bedingungen für das schlechte Abschneiden unseres Bildungssystems gibt es vermutlich eine Reihe von tiefer liegenden Ursachen. Dazu gehört m. E. in der Pädagogik eine gewisse Empiriefeindlichkeit, die sich in vielfältiger Weise äußert: Schulpädagogen, die an Hochschulen berufen werden, brauchen genau so wenig wie Lehrerstudenten vertiefte Kenntnisse in empirischen quantitativen Forschungsmethoden nachzuweisen, in den meisten Bundesländern brauchen Lehrerstudenten überhaupt keine Kenntnisse in Forschungsmethoden zu erwerben. Eine Berücksichtigung des international erreichten Forschungsstands durch die

Schulpädagogik ist deshalb in Deutschland nicht üblich. Theorien und Methoden werden danach beurteilt, ob sie zum Zeitgeist passen, und nicht nach dem Kriterium der empirischen Fundiertheit. So passt Stationenarbeit zum Zeitgeist, direkte Instruktion dagegen nicht, obwohl direkte Instruktion vermutlich für das Lernen oft erheblich effizienter ist (vgl. Aebli 1968; Hinrichs 2003).⁵

Eine weitere Ursache liegt in der in Deutschland vorherrschenden Lernkultur: Bei uns überwiegt eine *Selektions-* statt einer *Förder- oder Kooperationskultur*. Für schlechte Leistungen wird in der Regel die mangelnde Begabung des Schülers und nicht die mangelnde Anstrengungsbereitschaft von Schülern (und Lehrern) verantwortlich gemacht. Entsprechend werden in unserem selektiven, gegliederten Schulsystem Schüler vor allem von oben (Gymnasium) nach unten befördert, eine Förderung von unten nach oben findet kaum statt. Leistungsmessungen werden kaum zum Justieren des nachfolgenden Unterrichts verwendet, sondern indizieren den Abschluss einer Unterrichtseinheit, gleichgültig, wie viel Prozent der Schüler die grundlegenden Inhalte so gelernt haben, dass sie den nachfolgenden Unterricht begreifen können.⁶ Ein Aufbrechen dieser Selektionskultur durch Intensivierung einer *institutionalisierten Kooperation zwischen Lehrern* (vgl. Paine & Ma 1993), die gemeinsam Unterricht planen, diese Entwürfe erproben und gemeinsam überarbeiten, ist in einem Halbtagsschulsystem schwer realisierbar.⁷

⁴ Über die Voraussetzungen und Wirkungen eines konsequent an Standards orientierten Bildungssystems vgl. Resnick & Zurawsky (2005) sowie Briars & Resnick (2000).

⁵ Dem wird immer gleich das Argument entgegengebracht, durch direkte Instruktion würde das soziale Lernen weniger gefördert. Die empirische Befundlage stützt dieses Argument nicht.

⁶ Ich werde auf diesen Punkt noch genauer auf der Grundlage experimenteller Forschung eingehen.

⁷ Zu einer Förderkultur gehört auch ein anderes Verständnis der Bildungsamkeit der Schüler. Wer die Anstrengung des Schülers in den Mittelpunkt stellt, für den ist auch das Erreichen konkreter Standards und ein darauf gerichteter Unterricht selbstverständlich (vgl. Resnick 1999). In einer Selektionskultur geht es dagegen hauptsächlich um die Auswahl der Fähigkeiten; wer scheitert, verfügt nicht über die erforderlichen Begabungen. Dabei wird die Klasse bzw. der Jahrgang als Norm genommen, wobei bei normorientierten Tests immer ein größerer Teil der Schüler unter dem Durchschnitt liegen werden. Bei standardorientierten Messungen ist es dagegen prinzipiell möglich, dass alle Schüler anspruchsvolle Standards erreichen.

Eine andere, damit zusammenhängende Ursache liegt darin, dass eine Verbesserung der direkten Instruktion, die immer noch zu etwa 80 % den Unterrichtsalltag in Deutschland bestimmt, in den pädagogischen Diskursen vernachlässigt wurde. Statt diese normale Unterrichtsform zu optimieren, wurde viel Zeit und Energie in die Etablierung neuer Lehr- und Lernformen investiert. Es wurde dabei unterstellt, dass eine Einführung solcher neuer Lehr- und Lernformen quasi automatisch zu einer Verbesserung des Unterrichts führen müsse, ohne dass dafür jemals ernstzunehmende empirische Belege vorgelegt wurden. Zum gängigen pädagogischen Leitspruch wurde, Schüler könnten nur dann etwas effektiv lernen, wenn sie es selbst entdeckt bzw. konstruiert hätten. Reusser (1996, S. 16) bemerkte dazu: *„Wir laufen Gefahr, die essentielle Rolle der direkten und indirekten Führung durch gute Lehrer und gut strukturierte Unterrichtsmaterialien zu unterschätzen.“*

Man muss dabei allerdings differenzieren zwischen Entdecken im Sinne einer Aneignung neuen Wissens und Entdecken im Sinne von Bezüge herstellen zwischen theoretischem Wissen und konkreten Anwendungen. Das Zitat von Reusser bezieht

sich auf die erste Art des Entdeckens, dem Entwickeln und Konstruieren neuen Wissens. Hier macht es wenig Sinn, Schüler auf die lange Reise der Entwicklung neuen Wissens zu schicken, für die die Menschheit Jahrtausende benötigt hat.

Tuovinen und Sweller (1999) haben zu dieser Lernsituation ein instruktives Experiment durchgeführt. In diesem Experiment wurden Studenten in den Gebrauch eines Datenbankprogramms (FileMaker Pro) eingeführt. Man verglich zwei Lernmethoden:

- eine strukturierte Methode, in der die Reihenfolge der Bearbeitung genau festgelegt war und in der mit Lösungsbeispielen gearbeitet wurde, und
- eine offene Methode, in der die Schüler über die Reihenfolge selbst entscheiden konnten.

Auch bei der offenen Methode bestand die Möglichkeit, sich über noch fehlende Wissensbausteine zu informieren, insofern handelte es sich keineswegs um ein wirklich reines „entdeckendes“ Lernen.

Es wurde zwischen Studierenden, die noch keinerlei Erfahrungen mit Datenbankprogrammen hatten, und denen mit Erfahrung differenziert.

Tab. 1 Strukturierte Instruktion mit Hilfe von ausgearbeiteten Lösungsbeispielen versus freie, durch offene Fragen angeregte Exploration

	Strukturierte Instruktion mit Lösungsbeispielen	Durch offene Fragen angeregte Exploration
Keine vorherigen Erfahrungen mit einem Datenbankprogramm	29,6 ⁸ (16) (n = 8)	15,1 (9,6) (n = 9)
Vorherige Erfahrung mit einem Datenbankprogramm	30,9 (12,7) (n = 8)	35,9 (15,3) (n = 7)

⁸ Arithmetischer Mittelwert; in Klammern dahinter steht jeweils die zugehörige Standardabweichung.

Interpretation: Der einzige statistisch signifikante Effekt ist der Unterschied in den beiden Gruppen ohne einschlägiges Vorwissen: Bei fehlendem Vorwissen ist eine strukturierte Vorgehensweise mit Lösungsbeispielen deutlich wirksamer als entdeckendes Lernen! Es macht offensichtlich wenig Sinn, von Schülern eine Auswahl und Strukturierung von Wissens-elementen zu verlangen, wenn sie noch keine Vorstellung von den zu vermittelnden Inhalten und ihrer Struktur besitzen. Wenn Schüler schon einiges an Vorwissen mitbringen, scheint die offene Methode etwas günstiger zu sein – allerdings ist dieser Effekt nicht statistisch signifikant (was bei den kleinen Stichproben kein Wunder ist)⁹. Schüler überblicken, was ihnen an Wissen noch fehlt und wählen dazu die entsprechenden Lektionen aus.

Übertragen auf den Unterricht könnte das bedeuten: In der Schule ist geringes Vorwissen die gängige Ausgangssituation. In dieser Situation ist ein strukturiertes Vorgehen deutlich wirksamer als ein offenes Vorgehen. Entdeckendes Lernen benachteiligt Schüler mit niedrigem Vorwissen, also vor allem Schüler aus bildungsfernen Schichten. Möglicherweise begünstigt der konstruktivistische Zeitgeist Kinder, die von Hause aus mit viel Wissen versorgt werden ...

2. Reaktionen auf Heterogenität im Hinblick auf kognitives Lernen

Bei schulischen Lernangeboten wird in der Regel von der stillschweigenden Voraussetzung ausgegangen, die Schüler einer Klasse seien bezüglich ihrer Vorkenntnisse so homogen, dass ein gleiches methodisches Vorgehen für alle Schüler sinnvoll ist. Wir sprechen hier von einem gleichschrittigen Unterricht. Ein solcher Unterricht hat den Vorteil, dass der Lehrer den Unterricht für alle Schüler in gleicher Weise vorbereiten kann. Dabei wird allerdings übersehen,

dass eine große Heterogenität in allen Schulformen des gegliederten Schulsystems die Regel und nicht die Ausnahme ist.

Die Heterogenität der Schüler manifestiert sich in folgenden Merkmalen:

a. *Wissensbasis:* Schüler einer Klasse verfügen in den verschiedenen Wissensbereichen über unterschiedliche Kenntnisse, so dass für den einzelnen Schüler die jeweils zu lernende Informationsmenge unterschiedlich ist.

b. *Intelligenz:* Schüler unterscheiden sich außerdem darin, wie schnell sie Informationen aufnehmen, wie viele Informationen sie im Arbeitsgedächtnis speichern und wie effizient sie Informationen in ihr Langzeitgedächtnis integrieren können.

c. *Motivation:* Ferner differieren Schüler in ihrer Lernlust, ihren Ängsten und in ihren Motivationen. Dies hat Auswirkungen auf den Umfang der Lerntätigkeiten in den verschiedenen Bereichen sowie auch auf die Fähigkeit, effektiv Informationen zu verarbeiten.

d. *Meta-Kognition:* Für das Lernen sind ferner metakognitive Unterschiede in den Strategien und Verfahrensweisen der Problembearbeitung, der Problemlösung sowie der Fähigkeit, die Güte der eigenen Problemlösung kritisch zu beurteilen, bedeutsam.

Durch Unterricht ist vor allem (a), (c), und (d) beeinflussbar.

Die Heterogenität der Schüler hat also für schulisches Lernen Folgen. Schüler mit einem breiteren Vorwissen, großer Lernfreude, mit hochentwickelten metakognitiven Fähigkeiten oder mit großem intellektuellem Potenzial lernen bei gleichem Lernangebot mehr. Dieser Effekt ist unter der Bezeichnung „Matthäuseffekt“ bekannt, getreu dem Motto aus dem Matthäusevangelium: „Wer hat, dem wird gegeben.“ Eine wirksame Förderung muss gewährleisten, dass alle Schüler im Bereich ihrer „Zone

⁹ „Statistisch nicht signifikant“ bedeutet, dass per Zufall ein solcher Unterschied auch häufig auftreten würde.

der nächsten Entwicklung“ (Wygotski) gefördert werden. Dies kann nur geschehen, wenn der Lehrer seinen Unterricht an den Lernständen der verschiedenen Schülergruppen orientiert.

Im Folgenden werde ich auf verschiedene Möglichkeiten eingehen, durch einen adaptiven Umgang mit Heterogenität die Wirksamkeit des Unterrichts zu erhöhen (vgl. Roßbach & Wellenreuther 1992). Es handelt sich dabei um

1. Adaptive direkte Instruktion,
2. Bildung heterogener Lerngruppen (Gruppenrallye nach Slavin),
3. Bildung homogener Lerngruppen (jahrgangsübergreifende Differenzierung und die Methode des Lesenlernens in Neuseeland).

2.1 Adaptive direkte Instruktion und die Rolle von Tests

Eine Verbesserung der Unterrichtsqualität sollte bei den Lernarrangements anknüpfen, die am häufigsten im Unterricht angewendet werden. Dieses Lernarrangement ist in etwa 80% des Unterrichts die direkte Instruktion. Allerdings wird diese hochkomplexe Unterrichtsform häufig nicht in wirksamer Weise eingesetzt. In Helmkes Längsschnittuntersuchung erzielten 6 von 39 Lehrern ein vergleichsweise gutes Lernergebnis, ohne dass sich zwischen dem 5. und 6. Schuljahr die Leistungsunterschiede zwischen leistungsstarken und -schwachen Schülern noch weiter vergrößerte. Diese Optimalklassen unterschieden sich von den anderen Klassen in folgenden Punkten:

- Sehr effiziente Form der Klassenführung (vgl. Kounin 1976, Evertson, Emmer, Clements, & Worsham 1994³),
- hohe Instruktionsintensität (kaum Zeitvergeudung für außerfachliche Belange),
- hohe diagnostische Sensibilität bezüglich des Eingehens der Lehrer auf die affektiven Bedürfnisse der Schüler,

- hohes Maß an Förderorientierung. Die Hauptsorge der Lehrer gilt den leistungsschwachen Schülern.

Eine adaptive direkte Instruktion sollte die Weichen dafür stellen, dass möglichst alle Schüler die geforderten Mindeststandards erreichen. Für die Phase der Einführung komplexer neuer Inhalte sind nach den Ergebnissen der Grundlagenforschung unter dem Gesichtspunkt der geringen Kapazität des Arbeitsgedächtnisses folgende Methoden wichtig:

- die mehrfache Verwendung von Lösungsbeispielen, durch die sich jeder Schüler adaptiv das noch Unverstandene aneignen kann,
- Grafiken mit integriertem Text, die für sich – ohne untenstehenden Text – verständlich sind,
- verständliche mündliche und schriftliche Erklärungen und Darstellungen.

Für die Phase der festen Verankerung eignen sich danach folgende Methoden:

- Verteiltes statt massiertes Üben, wobei die Übungsphase insgesamt drei- bis viermal so viel Zeit in Anspruch nehmen sollte wie die Einführung der Inhalte,
- die gemeinsame mehrfache Erarbeitung der zu Grunde liegenden Struktur (vgl. Dumke 1984),
- durch kleine Tests, die Missverständnisse aufdecken, angeleitetes *adaptives Üben*, wobei sinnvolles Üben immer einen Vergleich mit Erklärungen oder Lösungsbeispielen bzw. entsprechendes Feedback durch Lehrer oder Schüler (Chefs, Gruppenarbeit) einschließt,
- die vielfache Verankerung in der Lebenswelt des Schülers durch Verwendung lebensnaher Beispiele (vgl. Anand & Ross 1987) und

- herausfordernde Transferaufgaben, Durchführung von Experimenten, Projekten und Exkursionen (vgl. MacKenzie & White 1982; Riegel 2004).¹⁰

Das Potenzial der direkten Instruktion ergibt sich gerade dadurch, dass hier der Lehrer die Möglichkeit hat, sukzessive mit den Schülern ein kognitives Gerüst aufzubauen und dieses in fast jeder Stunde zu wiederholen und zu vertiefen (vgl. Aebli 1968; Helmke 2003; Wellenreuther 2005, Kap. 4, 5, 7 und 9). Auch kann er hier Schüler mit spezifischen Verständnisschwierigkeiten zu Gruppen zusammenfassen, um diese Unklarheiten zu beseitigen. Solche Möglichkeiten gibt es im Rahmen der „neuen Lehr- und Lernformen“ kaum.¹¹

Zur Überprüfung des erreichten Kenntnisstands sollten von Fachleuten entwickelte Tests zur Verfügung gestellt werden, um sicher zu stellen, dass alle Schüler die aufeinander aufbauenden zentralen Inhalte sicher beherrschen. Es macht keinen Sinn, am Ende einer Unterrichtseinheit einen abschließenden Test durchzuführen, der drei Wochen danach zurückgegeben wird, um dann kurz nachbesprochen zu werden. Der Test wird hierbei nicht als Ressource für nachfolgendes Lernen, sondern ausschließlich als Messinstrument eingesetzt. In der Regel bekommen in solchen Tests etwa 20 % der Schüler bestätigt, dass sie keine ausreichenden Kenntnisse erworben haben. Wenn die Schüler dies erfahren, wurde meist schon längst mit der nächsten Unterrichtseinheit angefangen.

Eine alternative, adaptive Vorgehensweise nutzt Tests als Ressource für das nachfolgende Lernen. Tests informieren, welche Inhalte mehr oder weniger gut verstanden bzw. beherrscht werden. Der Lehrer kann diese Informationen dann für seinen nachfolgenden Unterricht nutzen. Eine systematische Nutzung solcher Testergebnisse führt zu einer erheblichen Steigerung des Unterrichtserfolgs. Statt Tests nur zum Abschluss einer Unterrichtseinheit einzusetzen, gibt es vielfältige alternative Nutzungsmöglichkeiten.

- Vor dem abschließenden Test werden *mehrere kleine Tests* geschrieben, um den nachfolgenden Unterricht zu orientieren. Z.B. kann der Lehrer Schüler nach den jeweils festgestellten Schwierigkeiten in Gruppen einteilen und in ihnen die fraglichen Inhalte nochmals erklären. Er kann auch Schüler, die diese Inhalte schon gut beherrschen, als Erklärhelfer (Tutoren) in diesen Gruppen einsetzen. Forschungen belegen, dass allein die Durchführung kleiner vorbereitender Tests mit jeweils kurzer Nachbesprechung die Effektivität des Unterrichts erheblich steigert (vgl. Dempster & Perkins 1993, Dempster 1996, Black & Wiliam 1998 a, b).
- Sowohl für diese kleineren Tests als auch für den abschließenden Test werden Wiederholungsmöglichkeiten auf freiwilliger Basis vorgesehen, was den Aufbau einer umfangreichen Testbibliothek mit jeweils mehreren Paralleltests voraussetzt.¹²

¹⁰ Insofern ist in der Phase der festen Verankerung von Wissen die selbstständige Arbeit in Projekten sehr sinnvoll. Es ist somit kein Zufall, dass Schulen wie die Helene Lange Schule in Wiesbaden, die in PISA erwartungswidrig gute Ergebnisse erzielen, sowohl der gängigen Zensierungspraxis sehr ablehnend gegenüber stehen als auch eine sehr intensive Projektarbeit betreiben (vgl. Riegel 2004).

¹¹ Hierdurch erklärt sich im Wesentlichen auch die geringe Lernwirksamkeit der Stationenarbeit (vgl. Hinrichs 2003): Hier ist es kaum möglich, in jeder Stunde an das Vorwissen anzuschließen und das Wichtigste am Anfang und am Ende der Stunde zusammenzufassen – wie auch, wenn bei 10–20 Stationen die Schüler an völlig verschiedenen Problemen arbeiten. Auch eine spezifische Förderung von Schülern in einer Gruppe, die bestimmte Inhalte noch nicht verstanden haben, muss aus solchen Gründen unterbleiben.

¹² Dies ist ein wichtiger Bereich, in dem zentrale Testinstitute Schulen in ihrer Arbeit unterstützen können. So ist in Schweden das Verfügbarmachen von Vergleichsarbeiten durch das zentrale Testinstitut üblich. Der Vorteil solcher Vergleichsarbeiten ist, dass durch sie die vorgegebenen Standards konkretisiert werden. Erst dadurch ist es für Lehrer möglich, den Unterricht an solchen Standards zu orientieren.

Tab. 2 Ergebnisse in der abschließenden Klausur (Mittlere Lösungsprozentsätze, Standardabweichungen und Stichprobengröße in Klammern)

Qualifikation des Lehrers	Häufigkeit des Testeinsatzes	
	Einmal	häufiger bei Bedarf
erfahrener, hochqualifizierter Lehrer	75,88 (15,27) (n = 23)	82,44 (12,16) (n = 27)
normaler Lehrer	62,61 (19,88) (n = 18)	82,55 (10,54) (n = 20)

- Die Tests sollten kriterienorientiert und unterrichtsvalide sein. Wichtig ist, wie gut ein Schüler die wesentlichen Inhalte beherrscht und nicht, wie gut er im Vergleich zu seinen Mitschülern abschneidet.

Wie bedeutsam ein adaptiver Einsatz von Tests für das Lernen von Schülern ist, belegen viele empirische Untersuchungen¹³. Im Folgenden möchte ich auf zwei Experimente eingehen, die sich dieser Thematik widmen (Martinez & Martinez 1992 und Butler 1988).

Im ersten Experiment ging es um die Wirkung der *Wiederholbarkeit von Teiltests* auf das Ergebnis im Endtest. Eine Gruppe von Lernern erhielt die Möglichkeit, bei schlechtem Abschneiden eines Teiltests diesen in Form von Paralleltests zu wiederholen. Dazu wurden Mastery-Tests entwickelt. Bei solchen Tests sollten Schüler mindestens 80% der Aufgaben richtig lösen, um den Inhalt zu beherrschen. In der anderen Gruppe bestand diese Möglichkeit zur Wiederholung der Teiltests nicht. Zusätzlich wurde noch das Qualifikationsniveau des Lehrers variiert (normaler Lehrer / Lehrer, der im früheren Fachunterricht immer sehr gute Ergebnisse erzielt hatte).

Es zeigte sich ein signifikanter Effekt der Möglichkeit zur Testwiederholung. Zwar erreichte der erfahrene Lehrer schon beim einmaligen Testen gute Ergebnisse (Lö-

sungsprozentsatz 76%). Bei der Möglichkeit zur Wiederholung der Teiltests verschwand allerdings dieser Unterschied zwischen den Lehrern. Das unterschiedliche Qualifikationsniveau von Lehrern spielt unter dieser Bedingung keine Rolle mehr! Es ist zu vermuten, dass durch die Möglichkeit zur Testwiederholung jeder Schüler vor dem Schreiben des Endtests motiviert wird, seine individuellen Wissenslücken zu schließen. Ein solches Schließen von Wissenslücken erscheint vor allem dann wichtig, wenn die gelernten Inhalte, die durch die Teiltests geprüft werden, hierarchisch aufeinander aufbauen.

Ein zweites Experiment bezog sich auf die Wirkung von Noten auf die nachfolgende Motivation und Leistung (vgl. Butler 1988). Es wurden hierbei drei Gruppen miteinander verglichen: Eine Gruppe, in der nach dem Schreiben des Tests nur ein Kommentar zum erzielten Ergebnis gegeben wurde, eine zweite Gruppe, in der sowohl Kommentar und Noten auf die Leistung gegeben wurde, und eine dritte Gruppe, in der die Leistung lediglich benotet wurde. Nach diesen „Rückmeldungen“ sollte nachfolgend eine ähnliche Leistung erbracht werden. Es zeigte sich, dass in der ersten Gruppe (nur Kommentar) eine signifikante Leistungssteigerung eintrat, während in den anderen beiden Gruppen die Leistung signifikant absackte.

¹³ Eine hervorragende Literaturübersicht zu diesem Thema geben Black & Wiliam (1998a).

Auch hier kann man vermuten, dass Kommentare den Schüler auf zu behobende Wissenslücken hinweisen, die er durch zusätzliche Anstrengungen beheben kann. Noten enthalten dagegen immer Vergleichsinformationen und legen dem Schüler nahe, er habe aufgrund mangelnder Fähigkeiten versagt. Dadurch wird seine Sachmotivation untergraben und die erforderlichen zusätzlichen Anstrengungen bleiben aus. Es scheint somit sinnvoll zu sein, sich mehr auf kriterienorientierte Mastery-Tests zu konzentrieren, weil diese das fachliche Können zum Vergleichsmaßstab heranziehen.¹⁴

Adaptiver Unterricht benötigt somit eher mehr als weniger Tests. Diese Tests werden hauptsächlich zur Steuerung des Lernverhaltens des Schülers und zur Justierung des nachfolgenden Unterrichts eingesetzt. Die genaue Erfassung des Leistungsstands der Schüler ist in einem solchen Rahmen auch wichtig, allerdings stärker in Bezug auf die zu erreichenden Kompetenzmaßstäbe bzw. festgelegten Standards und weniger in Bezug auf den relativen Leistungsstand im Vergleich zu den Mitschülern in der Klasse.

2.2 Bildung heterogener Lerngruppen (Gruppenrallye)

Nicht jede Form der Gruppenarbeit ist wirksam. Eine Form der Gruppenarbeit, deren Wirksamkeit empirisch gut belegt ist, ist die Gruppenrallye nach Slavin¹⁵. In dieser wird *Heterogenität als Ressource für das Lernen* aller Schüler, insbesondere der

schwächeren Schüler, genutzt. Gruppenrallyes eignen sich vor allem für die *Einübung* und *Festigung* von Inhalten, die zuvor vom Lehrer eingeführt und erklärt wurden. Sie eignen sich damit für ganz normalen Unterrichtsstoff. Insbesondere in Fächern wie Mathematik, Deutsch oder in den Fremdsprachen ergibt sich häufig die Notwendigkeit einer gezielten Einübung und Festigung von Inhalten, die zuvor vom Lehrer erklärt wurden.

Zunächst werden *heterogene* Gruppen von vier bis fünf Schülern gebildet; alle Gruppen sollen im Durchschnitt gleich leistungsstark sein. Eine Fünfergruppe kann z. B. aus einem sehr leistungsstarken Schüler, drei mittleren Schülern und einem leistungsschwachen Schüler bestehen. Diese Gruppen erhalten Namen (z. B. Quadrate, Mathe-Monster). Eine heterogene Zusammensetzung ist aus verschiedenen Gründen wichtig:

- Sie stellt sicher, dass die stärkeren Schüler den schwächeren gezielt helfen können.
- Sie schafft bessere Möglichkeiten zum Abbau ethnischer oder sozialer Vorurteile, weil Heterogenität auch soziale Heterogenität umfasst.

Der normale Ablauf einer Gruppenrallye sieht folgendermaßen aus:

- *Klassenunterricht*, 1–2 Stunden pro Woche:¹⁶ Hierbei werden neue Inhalte eingeführt, Verfahrensweisen diskutiert und die Regeln der Gruppenarbeit besprochen.

¹⁴ Warum in der Schule ständig das vergleichende Messen im Vordergrund steht, und nicht das Fördern der Schüler, um sie an die geforderten Leistungen heranzuführen, bleibt schwer nachvollziehbar. Erst am Ende der Schullaufbahn, also am Ende der zehnten Klasse, ist eine präzise Messung des Wissens und der Fähigkeiten der Schüler wichtig. Vorher sollten Tests vorrangig zuverlässige und gültige Informationen für eine Verbesserung des nachfolgenden Unterricht geben und Aussagen darüber zulassen, wie weit der einzelne Schüler von den geforderten Standards noch entfernt ist.

¹⁵ Die Gruppenrallye ist die Methode der Gruppenarbeit, die am strengsten überprüft wurde. Ihre Wirksamkeit liegt um 0,32 Standardeinheiten (Effektstärke) über dem normalen Klassenunterricht; nur eine Methode der Gruppenarbeit scheint noch wirksamer zu sein: Das mit der Gruppenrallye verwandte Gruppenturnier. Das beliebte Gruppenpuzzle ist mit 0,12 Standardeinheiten über dem normalen Klassenunterricht deutlich weniger wirksam. Allerdings bezieht sich die Gruppenrallye auch auf andere Inhalte wie das Gruppenpuzzle: In der Gruppenrallye werden normale Unterrichtsinhalte nachbereitet, im Gruppenpuzzle offene komplexe Probleme von verschiedenen Seiten entwickelt (vgl. Slavin 1995; Wellenreuther 2005).

¹⁶ Die Zeitangaben beziehen sich auf Fächer mit vier bis fünf Stunden Fachunterricht in der Woche (z. B. in einem Hauptfach wie Mathematik oder Deutsch).

- **Gruppenlernen**, 1–2 Stunden: Das Gruppenlernen findet in den heterogenen Vierer- oder Fünfergruppen statt. Die Schüler erhalten Aufgabenblätter und zugehörige Lösungsblätter¹⁷ (pro Gruppe jeweils zwei). Innerhalb der Gruppen werden in Zweier- oder Dreiergruppen die Aufgaben bearbeitet, um danach sofort die Lösung mit den anderen Schülern zu vergleichen. Wenn ein Schüler eine Aufgabe nicht lösen kann, dann sind die anderen Mitglieder der Gruppe dafür verantwortlich, ihm die Lösung in einer Weise zu erklären, dass der Schüler diese Erklärung selbst mit eigenen Worten rekonstruieren kann. Der Lehrer hat darauf zu achten, dass die Schüler die Arbeitsblätter genau beantworten und sie die Lösungen nach den ausgehändigten Antwortbogen auch kontrollieren.
- **Leistungstest**, Dauer $\frac{1}{2}$ bis höchstens 1 Schulstunde. Ziel ist eine individuelle Leistungsprüfung. Für die Durchführung wird die Sitzordnung verändert, zur Vermeidung von Mogeln werden möglicherweise Paralleltests ausgehändigt. Eine Zusammenarbeit bzw. ein gegenseitiges Helfen ist hierbei nicht erlaubt. Durch ein spezielles Berechnungsverfahren wird sichergestellt, dass schwächere Schüler den gleichen Betrag zum Gruppenergebnis beisteuern können wie gute Schüler, indem individuelle Leistungsverbesserungen für das Gruppenergebnis am meisten zählen. Dies wird aus der folgenden Berechnungstabelle deutlich, bei der Testergebnisse in Form von Lösungsprozentsätzen ausgedrückt werden. Das Ergebnis im individuellen Test wird jeweils mit dem Ergebnis in einem früheren Test (Basistest) verglichen.

Berechnung der Punkte für das einzelne Team aufgrund der Testergebnisse (vgl. Slavin 1995, S. 80)

Individueller Test – Basistest (Vergleich von Lösungsprozentsätzen)	Team erhält
10 Punkte oder mehr schlechter als im Basistest	5 Punkte
9 Punkte bis 0 Punkte schlechter als im Basistest	10 Punkte
1 bis 9 Punkte besser als im Basistest	20 Punkte
10 und mehr Punkte besser oder optimales Ergebnis als im Basistest	30 Punkte

¹⁷ Die Vergabe von Antwortblättern hat den Sinn, den Schülern eine Möglichkeit zur selbstständigen Kontrolle ihrer Lösungen an die Hand zu geben. Die Schüler müssen natürlich den verantwortungsbewussten Umgang mit diesen Antwortblättern lernen. Wenn sie die Lösungen abschreiben, können sie im anschließenden Test nicht erfolgreich sein. Vielleicht müssen sie diese Erfahrung selbst machen. Im Unterricht kann der Lehrer natürlich auf den sinnvollen Umgang mit den Antwortbögen eingehen.

- Für jede Gruppe wird aufgrund der individuellen Testergebnisse nach dieser Tabelle ein Gesamtpunktwert gebildet. Wenn eine Gruppe dabei einen hohen Durchschnittswert erzielt, bekommt sie eine Urkunde, die in der Klasse ausgehängt wird. Der Lehrer sollte die Tests schnell, d. h. bis zur nächsten Stunde, auswerten und die Ergebnisse mitteilen.
- **Rückmeldung an die Gruppen:** Sowohl die individuellen Ergebnisse als auch die Gruppenergebnisse werden mitgeteilt. Alle Schüler mit maximalem Verbesserungswert sollten namentlich aufgeführt werden. Wichtig ist nicht nur der individuelle Erfolg des einzelnen Schülers, sondern der Erfolg der gesamten Gruppe, da die Gruppe ja auch etwas mit dem individuellen Erfolg zu tun hat. Entsprechend wird das Gruppenergebnis hervorgehoben (z. B. „Mathe-Monster wieder vorn“ oder „Die Quadrate haben aufgeholt“). Die Rückmeldung sollte sich allerdings nicht nur auf die Leistung, sondern auch auf den Prozess, wie die Leistung in den einzelnen Gruppen zustande gekommen ist, beziehen. Dazu ist erforderlich, dass der Lehrer sich während der Gruppenarbeit dazu Notizen macht. Zusätzlich kann am Ende der Gruppensitzungen ein Beurteilungsbogen an die Schüler ausgehändigt werden, auf dem sie ihre persönlichen Erfahrungen mit der Gruppe durch Ankreuzen deutlich machen können.

Die Interaktionsformen lassen insbesondere am Anfang meist zu wünschen übrig. Der Lehrer kann in der Klasse über einige Verhaltensregeln sprechen, die für die Gruppenarbeit bedeutsam sind, z. B. „hören wir einander zu“; „lassen wir einander ausreden“; „ist es zu laut, stören wir die anderen durch unser Verhalten?“; „fragen wir gegenseitig um Hilfe, wenn wir etwas nicht verstanden haben?“; „geben wir uns Erklärungen, die der andere auch verstehen kann?“; „behandle ich meine Gruppenmit-

glieder immer so, wie ich selbst von ihnen behandelt werden möchte?“

Durch die Festlegung der Gruppenzusammensetzung wird in bestimmtem Umfang auch festgelegt, welche Gruppen ein hohes und welche ein eher niedrigeres Entwicklungspotenzial haben. Manche Gruppen harmonisieren sehr gut, manche weniger gut. Deshalb sollte nach einer bestimmten Zeit – Huber (1985) schlägt sechs Wochen vor – die Zusammensetzung geändert werden, damit die Mitglieder der weniger erfolgreichen Gruppen eine neue Chance für einen Gruppenerfolg bekommen.

Nach den Ergebnissen der Metaanalyse von Slavin handelt es sich bei der Gruppenrallye um eine effektive Unterrichtsmethode, die im sozialen und kognitiven Bereich zu deutlich besseren Ergebnissen als alternative Unterrichtsmethoden führt. Dieser Erfolg der Gruppenrallye kann auf verschiedene Faktoren zurückgeführt werden:

1) **Die Belohnung der Gruppen mit hohen durchschnittlichen individuellen Lernzuwächsen:** Im Mittelpunkt steht der individuelle Lernfortschritt eines jeden Gruppenmitglieds. Die Gruppe kann nur ein gutes Ergebnis erzielen, wenn alle, auch die schwächeren, viel dazulernen. Dieses Belohnungssystem lässt zu, dass alle Gruppen gut abschneiden. Dies scheint aus Motivationsgründen durchaus sinnvoll.

Wenn Gruppenprodukte zu erstellen sind, dann fühlen sich meist nur wenige Mitglieder der Gruppe für das Ergebnis verantwortlich. Die schwächeren Schüler werden häufig abgehängt, weil sie zum Gruppenprodukt wenig beitragen können (vgl. Salomon & Globerson 1989).

2) **Optimale Lernsequenzierung:** Die Strukturierung des Lernens in a) Präsentation und Erarbeitung neuer Inhalte im Rahmen direkter Instruktion, b) Einübung und Anwendung der neuen Inhalte in der Gruppenarbeit und c) Prüfung des individuellen Kenntnisstands entspricht einer optimalen Lernsequenzierung.

Für die Gruppenrallye benötigt man geeignete, dieser Arbeit angepasste Unterrichtsmaterialien. Dazu gehören insbesondere Arbeitsblätter mit den zu bearbeitenden Aufgaben, Lösungsblätter, die den Teams zur Kontrolle ausgehändigt werden, und Tests, die das behandelte Wissen überprüfen. Um diese Entwicklungsarbeit zu leisten, scheint es erforderlich, dass Lehrer untereinander über einen längeren Zeitraum intensiv miteinander kooperieren, um arbeitsteilig die benötigten Unterrichtsmaterialien zu entwickeln.¹⁸

2.3 Bildung homogener Lerngruppen (Lesenlernen in Neuseeland)

Als Musterbeispiel für eine effektive Förderung nach Bildung leistungshomogener Gruppen möchte ich auf die Strukturierung des Leseunterrichts in Neuseeland eingehen. Dabei beziehe ich mich auf einen Artikel von Wilkinson & Townsend (2000) über „best practice classrooms“. Neuseeland gehört nach PISA zu den Spitzenländern bezüglich der Lesekompetenz. Wilkinson und Townsend (2000) haben Experten-Lehrer untersucht, von denen aufgrund der Lernergebnisse ihrer Klassen aus früheren Jahren bekannt war, dass sie einen vorzüglichen Leseunterricht durchführen.

Folgende Punkte sind für diesen Unterricht kennzeichnend:

1) *Einteilung der Unterrichtsmaterialien nach Kompetenzstufen*: Die Lesebücher in Neuseeland werden nach dem Farbenrad grob in drei Schwierigkeitsstufen eingeteilt: Anfänger, erstes Lesen und flüssiges

Lesen. Innerhalb dieser Gruppen werden dann nach den Farben noch weitere Unterteilungen vorgenommen. Alle Lesematerialien außerhalb des Schulbuchs sind nach dem Farbenrad eingestuft.

Die Entwicklung angepasster Unterrichtsmaterialien hat im Rahmen einer solchen Förderung die Funktion, den Lehrer zu entlasten, damit er sich auf die gezielte Förderung seiner Schüler konzentrieren kann. Die Lehrer werden nämlich durch die differenzierenden Arbeitsaufträge an die Gruppen schon genügend gefordert.¹⁹

2) *Einteilung der Schüler in bis zu sieben homogene Lerngruppen*: Um Schüler in die passenden Lerngruppen einzuordnen, werden Berichte von Lehrern, die die Schüler zuvor unterrichteten, eigene Beobachtungen in der ersten Unterrichtswoche sowie Ergebnisse von Tests berücksichtigt. Durch die stärkere Differenzierung in viele homogene Lerngruppen pro Klasse werden die Lehrer möglicherweise für die Größe der bestehenden Unterschiede stärker sensibilisiert.

3) *Lernziele des Leseunterrichts und zusätzliche Maßnahmen*: Jeder Schüler soll in einer bestimmten Zeit zum Stadium des flüssigen Lesens geführt werden. Bei größeren Defiziten kümmert sich an jeder Schule zusätzlich eine Fachkraft individuell so lange um das Kind, bis die Defizite aufgeholt sind (vgl. Faust, 2002). Ferner wird der Lehrer in seinem Unterricht in der Regel noch durch eine Hilfskraft zusätzlich unterstützt. (Wilkinson & Townsend, 2000, S. 465).

¹⁸ Zur Verdeutlichung der für die Gruppenrallye benötigten Unterrichtsmaterialien vgl. die im Anhang dargestellten Beispiele sowie Slavin (1995, S. 168f.).

¹⁹ Dies ist einer der Gründe, warum eine systematische Qualitätssicherung der Schulbücher und zugelassenen Lehrmaterialien in Deutschland sehr wichtig wäre.

Beim Lernziel „flüssiges Lesen“ handelt es sich um eine absolute Norm. Das Erreichen der Zwischenstufen zum Erreichen dieses Endziels ist klar geregelt: Ein Schüler muss beim Lesen einer Geschichte seiner Kompetenzstufe 95 % der Wörter richtig lesen, um in die nächst höhere Kompetenzstufe aufzusteigen.²⁰ Schüler steigen so von Kompetenzstufe zu Kompetenzstufe auf.

Klassenmanagementprobleme: Eine effektive Unterrichtung von so vielen unterschiedlichen homogenen Lerngruppen, die nach einheimischen neuseeländischen Baumarten Rata, Karaka, Ponga, Manuka, Totara, Kowhai und Rim benannt werden, setzt eine präzise Aufgabenverteilung im Morgenkreis voraus. Mrs. Brigham beginnt: „Gut, lasst uns zuerst auf die **Aufgabentafel** für heute sehen. Karaka soll sich die Geschichte „König Midas und das Goldene Vlies“ anhören. Ihr bekommt vier Blätter, und darauf sollt ihr die vier wichtigsten Ereignisse zeichnen. Über die Zeichnungen schreibt ihr eine passende Überschrift. Dann sollt ihr das Gedicht „der haarige Hund“ für eine öffentliche Darbietung lernen und danach habt ihr im Textbuch zu arbeiten. Ich möchte, dass Rimu zu mir zum Lesen kommt. Danach habt ihr Partnerlesen aus der Wahlbox, danach habt ihr Wortsuchen ... Totari und Kowhai nehmen sich Gedichte vor und lesen gemeinsam still, danach kommt ihr zu mir zum Lesen. Ponga und Manuka holen sich aus ihrem Regal ein Buch, lesen gemeinsam still ... Ben und Jordan, ihr sollt, falls ihr dazu Zeit habt, die Geschichte am Computer fertig schreiben ...“ (Wilkinson & Townsend, 2000, S. 460)

Vergleich: Neuseeländischer und deutscher Leseunterricht:

1) In Neuseeland werden Lernmaterialien nach Schwierigkeit geordnet. „Im Gleichschritt marsch“ gilt in Neuseeland nur für die Lernaktivitäten der homogenen Lerngruppe. Wenn ein Schüler die Schwierigkeiten seiner Kompetenzstufe gemeistert hat, kann er in die nächst höhere Kompetenzstufe aufsteigen. Dadurch wird sichergestellt, dass sich keine größeren Lücken im systematischen Wissensaufbau ergeben.

2) In Neuseeland geht es um das Erreichen flüssigen Lesens für möglichst *alle* Schüler. Das ist eine absolute, keine relative Norm. Schüler vergleichen sich hauptsächlich mit dieser absoluten Norm. Das Erreichen der Zwischenstufen zum Erreichen dieses Endziels ist klar geregelt: Ein Schüler muss, um in die nächst höhere Kompetenzstufe aufzusteigen, beim Lesen einer Geschichte seiner Kompetenzstufe 95 % der Wörter richtig lesen. Falls das Erreichen des End-Ziels „flüssiges Lesen“ gefährdet ist, werden massive Maßnahmen wie z. B. zusätzlicher individueller Unterricht durch eine dafür abgestellte Lehrkraft ergriffen.

Deutsche Lehrer können für ihre Klassen am Ende der 2. Klasse gar nicht sicher angeben, wie viele Schüler *nicht* flüssig lesen können. In deutschen Schulen wird als normales Ergebnis hingenommen, wenn etwa 20 % der Schüler in den ersten zwei bis drei Schuljahren nicht lernen, Texte wirklich flüssig zu lesen.

Vermutlich hängen die Erfolge des Lesenslernens in Neuseeland damit zusammen, dass in den leistungshomogenen Gruppen die Schüler in ihrer Zone der nächsten Entwicklung arbeiten können. Allerdings schafft der Lehrer eine sinnvolle Aufgabenstellung in den Gruppen nur dadurch, dass die Lehrmaterialien nach Schwierigkeitsstufen vorsortiert sind. Die Durchfüh-

²⁰ In Deutschland wird eine soziale, relative Norm verwendet: Es reicht den Lehrern, wenn die Mehrzahl der Schüler erfolgreich die Fibelaufgaben bewältigt. Da die kognitiven Voraussetzungen sehr unterschiedlich sind, gibt es im Hinblick auf diese relative Norm immer viele Schüler, die beim Erreichen dieser Norm Schwierigkeiten haben und versagen. Durch die Überprüfung der erreichten Kompetenz der einzelnen Schüler werden die Lehrer stärker für das Erreichen der Kompetenzstufen und der absoluten Norm „flüssiges Lesen“ sensibilisiert. In deutschen Schulen wird als normales Ergebnis hingenommen, wenn zwischen 20 und 30 % der Schüler in den ersten drei Schuljahren nicht lernen, Texte wirklich flüssig zu lesen.

rung von kompetenzorientierten individuellen Lesetests vermitteln den Schülern das Gefühl, ein konkretes Ziel bis zur Meisterschaft erreicht zu haben, ohne dabei das zu erreichende Endziel aus dem Blick zu verlieren. Sowohl die Gruppenrallye als auch das Modell des Lesenlernens in Neuseeland orientieren sich an der individuellen oder an der sachlichen Bezugsnorm. Es zählt nicht eine durch Vergleich mit der Klasse erzielte benotete Leistung, sondern eine im Vergleich zu einem absoluten Standard (Kompetenz im Vergleich zu definierten Kompetenzstufen) oder zu einem früheren Testergebnis erbrachten Leistung (Gruppenrallye; Vergleich mit einem individuellen Bezugsrahmen). Dadurch kann auf identitätsverletzende Informationen zum Rangplatz in der Klasse verzichtet werden. Diese Interpretation steht in Übereinstimmung mit den Ergebnissen der experimentellen Studie von Butler (1988).²¹

Wenn die Wirksamkeit des Leseunterrichts in Deutschland erhöht werden soll, müssen einige wesentliche Ideen des neuseeländischen Modells des Lesenlernens auf Deutschland übertragen werden. Dazu zählt die Bereitstellung geeigneter, nach Schwierigkeit geordneter Lese- und Unterrichtsmaterialien und die Bereitschaft der Lehrer zur Bildung und differenzierten Unterrichtung leistungshomogener Gruppen. Entsprechend müssten leistungshomogene Gruppen gebildet werden, wobei die Anzahl der Gruppen nach dem Umfang der Heterogenität in der Klasse festzulegen ist. Ferner müsste durch individuelle Mastery-Tests in regelmäßigen Abständen geprüft werden, ob Schüler um eine Kompetenzstufe aufsteigen können.

Zusätzlich wäre wichtig, von Anbeginn an intensiv die Gruppe von Schülern zu fördern, die bei Beginn der ersten Klasse noch keine ausreichenden Voraussetzungen für den Leseunterricht mitbringen.²² Deshalb müsste schon bei Schulbeginn jeder Schüler danach getestet werden, ob er hinsichtlich des Lesen- und Schreibenlernens die erforderliche phonologische Bewusstheit schon entwickelt hat (vgl. Jansen, Mannheim, Marx & Skowronek 2002). Alle Schüler mit unzureichender phonologischer Bewusstheit müssten im ersten Jahr in diesem Bereich gezielt, z. B. mit Hilfe des empirisch geprüften Programms von Forster und Martschinke, gefördert werden (vgl. Forster & Martschinke 2001, Küspert 2001). Nach Forschungen, die wir in Lüneburg durchgeführt haben, können dann fast alle Schüler in den ersten beiden Schuljahren erfolgreich Lesen und Schreiben lernen (vgl. Böer 2003).

3. Empirische Forschung und innere Schulreform

Spätestens nach PISA stellt sich die Frage nach einer grundlegenden inneren Reform unseres Schulsystems, um die Lernmöglichkeiten aller Schüler zu verbessern. Diesem Ziel dienen die drei Möglichkeiten des Umgangs mit Heterogenität, die in diesem Beitrag vorgestellt wurden:

- Optimierung der direkten Instruktion, also des alltäglichen Unterrichts, vor allem durch Erhöhung seiner Adaptivität,
- Durchführung von Gruppenarbeit mit heterogenen Gruppen im Anschluss an eine Phase direkter Instruktion (Gruppenrallye), und

²¹ In dieser Studie wurden zunächst nach dem Bielefelder Screening die Schüler von Versuchs- und Kontrollgruppen ausgewählt, deren phonologische Bewusstheit für den Lese- und Schreibunterricht zu gering ausgeprägt war. Nachfolgend wurden die ausgewählten Schüler der Versuchsgruppe nach dem Rundgang durch Hörhäuser gezielt geschult, die Kinder der Kontrollgruppe (vergleichbare Schüler aus Parallelklassen) erhielten traditionellen Förderunterricht. Durch die gezielte Förderung konnten die Schüler der Versuchsgruppe in ihren Leistungen sich so stark verbessern, dass sie normal beschulbar waren. Die Ergebnisse waren trotz relativ kleiner Gruppen signifikant.

²² Dieses Problem wäre durch ein entsprechendes spielerisches Training im Kindergarten natürlich sinnvoller lösbar (vgl. Lundberg, Frost und Peterson 1988, Küspert 2001).

→ Durchführung von Unterricht in kleinen leistungshomogenen Gruppen (Lesenlernen in Neuseeland).

Hinter allen vorgestellten Maßnahmen steht eine zentrale Idee: Unterricht kann dann wirksam sein, wenn er sich auf die Vorkenntnisse und Lebenswelten der Schüler einstellt und sie in ihrem Bereich der Zone der nächsten Entwicklung (Wygotski) fördert. Dieser Idee widerspricht (1) ein gleichschrittiges Unterrichten der ganzen Klasse, (2) ein entdeckendes Lernen, wenn zuerst einmal neue Schemata eingeführt werden müssen²³ und (3) eine Nutzung von Tests nur zur summativen Leistungsmessung.

Abschließend stellt sich die Frage, unter welchen Bedingungen die hier vorgeschlagenen Methoden des Unterrichtens, die alle streng empirisch überprüft sind, von Lehrern aufgenommen und im Unterricht angemessen eingesetzt werden. Aufgrund der Lernforschung ist bekannt, dass Erkenntnisse erst dann in Handeln umgesetzt werden können, wenn sie auf alltägliche Problemlösungen konkret angewendet werden müssen. Diese Kompetenzen müssten Lehrer erwerben können „ wenn sie

1. regelmäßig im Rahmen professioneller Lehrerfortbildungen mit den neuesten Erkenntnissen der Grundlagenforschung vertraut gemacht und

2. wenn sie sich außerdem in Teams zusammenfinden würden, in denen z. B. in jeder Woche zwei Stunden am Nachmittag kooperativ Unterrichtseinheiten und andere Reformmaßnahmen konkret entwickelt und dann anschließend erprobt werden (Lehrer hospitieren dabei). Danach werden die Unterrichtsentwürfe aufgrund der gemachten Erfahrungen überarbeitet.²⁴

Von den hier dargestellten Methoden lassen sich m. E. am leichtesten die Methoden einer Optimierung der direkten Instruktion auf die alltägliche Unterrichtspraxis anwenden. Allerdings setzen alle genannten Methoden einen grundlegenden Wandel von einer Selektionskultur zu einer Förder- und Lernkultur voraus. Tests haben hier vor allem die Funktion, nachfolgendes Lernen zu steuern und das Erreichen von gesetzten Standards zu prüfen. Entscheidend ist, wie viel die Schüler am Ende des Schullebens für das Leben in Beruf und Gesellschaft gelernt haben. Schüler kommen mit sehr unterschiedlichen Voraussetzungen in die Schule, sie besitzen unterschiedliche Kenntnisse und Fähigkeiten, und es macht keinen Sinn, alle Schüler immer gleich zu behandeln. Die Erfahrungen in anderen Ländern zeigen, dass eine effektive Förderkultur häufig in Gesamtschulen realisiert wird, in denen Schüler bis zum zehnten Schuljahr in heterogenen Klassen gemeinsam ganztags unterrichtet werden. Solche heterogene Klassenverbände bieten den Vorteil, dass für alle Schüler ein *anspruchsvolles Lernklima* erhalten wird, in dem durch vielfältige Differenzierungsmaßnahmen und den Einsatz stärkerer Schüler als Erklärhelfer in Gruppen alle Schüler individuell gefördert werden können.

Dieses Lernklima wird zerstört, wenn Unterrichtseinheiten wie bisher durch vergleichende abschließende Klassenarbeiten „abgeschlossen“ werden. Hier werden Lernchancen verspielt, die bei den Schülern mit grundlegenden Wissenslücken noch bestehen. Wenn geklärt ist, welche Lerngebiete zentral sind, u. a. weil sie aufeinander aufbauen, dann muss der Unterrichtsaufbau dieses auch berücksichtigen. Er muss vermehrt Möglichkeiten bereit-

²³ Schüler in ihrer Zone der nächsten Entwicklung zu fördern bedeutet ja, sie in einem Bereich zu fordern, der im Rahmen einer didaktischen Anleitung und Unterstützung durch Lehrer vermittelt werden kann. Die Zone der nächsten Entwicklung meint nicht den Bereich, den ein Schüler selbst entdecken oder sich erarbeiten kann.

²⁴ Solche Formen der institutionalisierten Kooperation unter Lehrern sind in den asiatischen Ländern üblich (Paine & Ma 1993). Stigler und Hiebert (1999) gehen davon aus, dass tiefgreifende innerschulische Reformen ohne konkrete minutiöse gemeinsame Planung und Erprobung von Reformkonzepten im Unterricht nicht erfolgreich sind.

stellen, dass Schüler mit höchstens ausreichenden Leistungen Möglichkeiten zum nochmaligen intensiven Nacharbeiten erhalten, um dann erneut den fraglichen Test zu schreiben. Um eine solche Förderkultur an unseren Schulen zu etablieren, bedarf es grundlegender Korrekturen auch an der Stundenverteilung und der Organisation des Unterrichts. Solche andere Formen der Zusammenarbeit sind allerdings leichter in Ganztagschulen als in den in Deutschland immer noch vorherrschenden Halbtagschulen möglich.

Literatur

- Aebli, Hans: Psychologische Didaktik. Stuttgart 1968³.
- Anand, P.G. & Ross, S.M.: Using Computer-Assisted Instruction to Personalize Arithmetic Materials for Elementary School Children. *Journal of Educational Psychology*, Vol. 79, No. 1, 1987, 72–78.
- Black, Paul & Wiliam, Dylan: Assessment and Classroom Teaching. *Assessment and Education*, 1998a, 5, 7–73.
- Black, Paul & Wiliam, Dylan: Inside the Black Box. Raising Standards Through Classroom Assessment. Phi delta kappan: A Journal for the promotion of leadership in education. Vol. 80, 1998b. S. 139–148.
- Böer, K.: Diagnose und Förderung phonologischer Bewusstheit in Grundschulen – diskutiert an einer praktischen Erprobung in einer ersten Klasse der Grundschule Mühlenweg, Hannover 2003. Unveröffentlichte Examensarbeit, Universität Lüneburg.
- Briars, D.J. & Resnick, L.B.: Standards, Assessments – and What Else? The Essential Elements of Standards-Based School Improvement. CSE Technical Report 528. 2000. (Im Internet bei Google unter dem Titel abrufbar)
- Butler, R.: Enhancing and undermining intrinsic motivation: the effects of task-involving and ego-involving evaluation on interest and performance. *British Journal of Educational Psychology*, 58, 1988, 1–14
- Dempster, F.N., Perkins, P.G.: Revitalizing Classroom Assessment: Using Tests to Promote Learning. *Journal of Instructional Psychology*, 1993, 20, 197–203
- Dempster, F.N.: Distributing and Managing the Conditions of Encoding and Practice. In: Bjork, E.L. & Bjork, R.A.: *Memory*. Academic Press, San Diego, 1996², 317–344.
- Dumke, D.: Die hierarchische Strukturierung von Unterrichtsinhalten als Lernhilfe in der Grundschule. *Psychologie in Erziehung und Unterricht*, Jg. 31, 1984, 43–49.
- Evertson, C.M., Emmer, E.T., Clements, B.S. & Worsham, M.E.: *Classroom Management for Elementary Teachers*. Boston, Allyn & Bacon 1994³.
- Faust, G.: Pisa und die Grundschule. Interpretation der Befunde und mögliche Konsequenzen. *Die Deutsche Schule*, Bd. 94, No. 3, 2002, 300–317.
- Forster, M. & Martschinke, S.: Leichter lesen und schreiben lernen mit der Hexe Susi. Übungen und Spiele zur Förderung der phonologischen Bewusstheit. Diagnose und Förderung im Schriftspracherwerb, Band 2. Donauwörth 2001.
- Helmke, Andreas: Leistungssteigerung und Ausgleich von Leistungsunterschieden in Schulklassen: unvereinbare Ziele? *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, Bd. 20, Heft 1, 1988, 45–76.
- Helmke, Andreas: Unterrichtsqualität – erfassen, bewerten, verbessern. Seelze 2003.
- Helmke, Andreas, Hesse, Hermann-Günter: Kindheit und Jugend in Asien. In: Krüger, H.-H. & Grunert, C. (Hrsg.): *Handbuch der Kindheits- und Jugendforschung*. Opladen 2002, 440–471.
- Hinrichs, Tatjana: Stationenarbeit oder direkte Instruktion – ein empirischer Vergleich zweier Lernarrangements in vierten Grundschulklassen am Beispiel der Unterrichtseinheit „Die Spinne“. Unveröffentlichte Examensarbeit, Universität Lüneburg. 2003.
- Huber, G.L. (Hrsg.): 1: Lernen in Schülergruppen : A, Grundlagen. B, Organisationsmodelle und Materialien. In: Rotering-Steinberg, S. ... [Bearb.]. *Pädagogisch-psychologische Grundlagen für das Lernen in Gruppen: Studienbrief* [Bearb.]. Deutsches Institut für Fernstudien an der Universität Tübingen, Tübingen: Dt. Inst. für Fernstudien 1985.
- Jansen, H., Mannhaupt, G., Marx, H. & Skowronek, H.: BISC. Bielefelder Screening zur Früherkennung von Lese- und Rechtschreibschwierigkeiten. Göttingen: Hogrefe 2002².
- Kounin, Jacob S.: *Techniken der Klassenführung*. Stuttgart 1976.

- Küspert, Petra: Wie Kinder leicht lesen und schreiben lernen. Neue Strategien gegen Legasthenie. Ratingen 2001.
- Lundberg, I., Frost, J. & Petersen, O.-P.: Effects of an extensive program for stimulating phonological awareness in preschool children. *Reading Research Quarterly*, Vol. 23, No. 3, 1988, 263–284.
- MacKenzie, A. A., White, R. T.: Fieldwork in Geography and Long-term Memory Structures. *American Educational Research Journal*, Vol. 19, No. 4, 1982, 623–632.
- Martinez, J. G. R. & Martinez, N. C.: Re-examining repeated testing and teacher effects in a remedial mathematics course. *British Journal of Educational Psychology*, Vol. 62, 1992, 356–363.
- Paine, L. & Ma, L.: Teachers Working together: A Dialogue on organizational and cultural perspectives of chinese teachers. *International Journal of educational Research*, 19/8, 1993, 675–718.
- Reusser, K.: From cognitive Modeling to the Design of Pedagogical Tools. Vosniadou, S., Corte, E., Glaser R., & Mandl, H. (Hrsg.): *International Perspectives on the Design of Technology-Supported Learning Environments*. Mahwah, NJ, 1996, 81–103
- Resnick, Lauren B.: Making Amerika smarter. *Education Week Century Series*. 18 (40), 1999, 38–40. (Im Internet bei Google unter „instituteforlearning“, „publications“ abrufbar)
- Resnick, Lauren B. & Zurawsky, Chris: *Standards-Based Reform and Accountability. Getting Back on Course*. *American Educator*, Spring 2005. (Im Internet bei Google unter „instituteforlearning“, „publications“ abrufbar)
- Riegel, Enja: *Schule kann gelingen. Wie unsere Kinder wirklich fürs Leben lernen*. Frankfurt am Main: S. Fischer 2004.
- Roßbach, H.-G. & Wellenreuther, M.: Empirische Forschungen zur Wirksamkeit von Methoden der Leistungsdifferenzierung in der Grundschule. In: Heinzel, F. & Prengel, A. (Hrsg.): *Heterogenität, Integration und Differenzierung in der Primarstufe*. *Jahrbuch Grundschulforschung* 6. Opladen 2002, 44–57.
- Salomon, G. & Globerson, T.: When Teams do not Function the way they ought to. *International Journal of Educational Research*, Vol. 13, 1989, 89–99.
- Scheich, Henning: *Neurobiologische Grundlagen des Lernens*. Seminar 1, *Das Lehren Lernen*. 2005, S. 33–34.
- Slavin, R.E.: *Cooperative Learning. Theory, Research, and practice*. Ayllon and Bacon. Boston et al. 1995².
- Stigler, James W. & Hiebert, James: *The Teaching Gap. Best Ideas from the World's Teachers for Improving Education in the Classroom*. Free Press, New York 1999.
- Tuovinen, J. E. & Sweller, J.: A Comparison of Cognitive Load associated with discovery learning and worked examples. *Journal of Educational Psychology*, Vol. 91, No. 2, 1999, 334–341.
- Wellenreuther, M.: *Lehren und Lernen – aber wie? Empirisch-experimentelle Forschungen zum Lehren und Lernen im Unterricht*. *Grundlagen der Schulpädagogik*, Band 50. Schneider Verlag Hohengehren 2005. Zweite, korrigierte und überarbeitete Auflage.
- Wilkinson, I. A. G. & Townsend, M. A. R.: From Rata to Rimu: Grouping for instruction in best practice New Zealand classrooms. *The Reading Teacher*, Vol. 53, No. 6, 2000, 460–471.

Dr. Martin Wellenreuther
 Institut für Pädagogik, Universität Lüneburg