

Um das bildungspolitische Wissen darüber zu verbessern, welche Reformen tatsächlich zu besseren Bildungsergebnissen führen und welche nicht, wäre es wichtig, bildungspolitische Reformen in Zukunft immer wissenschaftlich fundierten Ergebnissevaluationen zu unterziehen. Aus jeder der zahlreichen Reformen, die in den verschiedenen Bundesländern regelmäßig vorgenommen werden – sei es in Bezug auf das Prüfungssystem oder die Notengebung, die Selbstständigkeit von Schule oder die Erhöhung von Wahlfreiheit oder irgendeine sonstige Veränderung der Rahmenbedingungen –, könnte bei einer begleitenden empirischen Untersuchung herausgefunden werden, ob die Reform wirklich zu besseren Schülerleistungen geführt hat oder nicht. Auf diese Weise könnte auch die Bildungspolitik selbst noch viel lernen.

*Verf.: Prof. Dr. Ludger Wößmann, Lehrstuhl für Volkswirtschaftslehre, insb. Bildungsökonomik, Ludwig-Maximilians-Universität München, Bereichsleiter Humankapital und Innovation, ifo Institut für Wirtschaftsforschung, Poschingerstr. 5, 81679 München*

*Hendrik Jürges/Kerstin Schneider*

## **Leistungsgerechtere Bezahlung von Lehrern: Sinnvolle Option oder sinnlose Ökonomisierung?**

### **1** Die Krise und ihre Chancen

Das deutsche Bildungssystem steht auf dem Prüfstand. Nach den international vergleichenden Schulstudien der vergangenen Jahre wird in Deutschland über neue, ungewohnte Gestaltungsmöglichkeiten nachgedacht. Reformen des Bildungssystems sind denkbar geworden, die es ohne die Debatte um PISA vielleicht nicht gegeben hätte. So liegt in der Krise die Chance, Reformen voranzubringen, die das deutsche Schulsystem dauerhaft leistungsfähiger machen können. In diesem Beitrag greifen wir nur einen Aspekt aus der Diskussion heraus und betrachten ihn aus bildungsökonomischer Sicht: die Struktur der Lehrerbesehung.

Um dies vorwegzunehmen: Deutschland ist in seinen Bemühungen um ein modernes Schulsystem nicht allein. Auch in anderen Ländern reagiert man auf Probleme mit zum Teil heftig umstrittenen Reformvorschlägen und mehr oder weniger mutigen Versuchen, alte und vertraute Strukturen aufzubrechen. Ein Unterschied zur Situation in Deutschland besteht darin, dass die Probleme in anderen Ländern häufig früher bekannt waren und daher schon seit einigen Jahren diskutiert werden.

Nun sind Erfahrungen aus anderen Ländern, selbst die unserer europäischen Nachbarn, nicht ohne weiteres auf Deutschland zu übertragen. Aber es ist sicherlich kein Nachteil, über Reformen, deren Erfolg naturgemäß unsicher ist, erst entscheiden zu müssen, nachdem andere Länder Reformen bereits gewagt haben und erste Ergebnisse über Erfolg und Misserfolg vorliegen. Es ist daher eher ein Vorteil für die deutsche Bildungspolitik, keine Pionierarbeit leisten zu müssen und von Reformanstrengungen anderer Länder lernen zu können.

Soll das Bildungssystem leistungsfähiger werden, so müssen zunächst die für den Bildungserfolg relevanten und durch Bildungspolitik beeinflussbaren Faktoren bzw. Inputs identifiziert werden. So lässt sich die sozio-ökonomische Herkunft der Schülerinnen und Schüler nicht

durch Bildungspolitik verändern. Wie aber Unterschiede im sozialen Status durch das Bildungssystem aufgefangen werden, ist sehr wohl eine bildungspolitische Frage. Neben dem Elternhaus ist die „Qualität“ der Lehrerinnen und Lehrer von größter Bedeutung für den Bildungserfolg.<sup>1</sup> Wie empirische Studien gezeigt haben, weist die Qualität von Lehrkräften aber eine große Varianz auf. So wurden in einer US-Studie Leistungsunterschiede bei Schülern und Schülerinnen in Abhängigkeit von der Qualität der Lehrkräfte in der Größenordnung von bis zu einem Schuljahr geschätzt.<sup>2</sup> Gute Lehrerinnen und Lehrer können sogar Unterschiede, die durch die soziale Herkunft von Kindern entstehen, ausgleichen. Hat ein Kind aus einkommensschwachen Verhältnissen für fünf Jahre einen sehr guten Lehrer, so gleichen sich seine Bildungschancen im Vergleich zu einem Kind aus einer wohlhabenden Familie mit durchschnittlich gutem Lehrer aus. Ein bildungspolitisches Ziel höchster Priorität ist daher, geeignete Lehramtskandidaten zu finden, sie gut auszubilden und als Lehrkräfte dauerhaft zu engagierter und guter Lehre zu motivieren.

Nun ist aber die Qualität von Lehrerinnen und Lehrern nicht verlässlich durch beobachtbare und traditionelle (Input-)Indikatoren wie deren akademische Abschlüsse oder Examensnoten zu messen.<sup>3</sup> Nicht einmal die Berufserfahrung der Lehrkräfte (gemessen in Jahren) scheint einen messbaren Effekt auf Schülerleistung zu haben.<sup>4</sup> Das bedeutet nun nicht, dass Lehrerqualität gar nicht messbar ist. Es bedeutet lediglich, dass andere Maße als die traditionellen herangezogen werden müssen.

Wenn die Qualität der Lehrerinnen und Lehrer aber nur unzureichend durch Inputindikatoren gemessen werden kann, so bleibt alternativ zu prüfen, ob es Bildungsergebnisse oder Outputs gibt, die als Indikatoren für die Leistung der Lehrkräfte herangezogen werden können. Auch bei der outputorientierten Qualitätsmessung gibt es erheblichen Diskussionsbedarf. Insbesondere die Mehrdimensionalität des Outputs bereitet dabei Probleme. Denn was sind die Ziele, die im Primar- und Sekundarbereich erreicht werden sollen?<sup>5</sup> Zu nennen sind hier: die Vermittlung von Basiskompetenzen in den Bereichen, Sprache, Mathematik, Natur- und Gesellschaftswissenschaften, Unterstützung in der Entwicklung der Kinder und Jugendlichen, Vorbereitung auf das Arbeitsleben, Vorbereitung auf das Leben in Verantwortung für sich und die Gemeinschaft, Ausgleich der herkunftsbedingten sozio-ökonomischen Unterschiede und die Schaffung eines gewaltfreien und gesunden Lernumfeldes. Zwar schließen sich die genannten Ziele nicht gegenseitig aus, aber das Erreichen jedes Zieles verlangt eine andere Verteilung der vorhandenen Ressourcen. Im Mittelpunkt der outputorientierten Leistungsbewertung von Lehrerinnen und Lehrern durch Bildungsökonomern steht häufig der Erfolg in der Vermittlung von Basiskompetenzen, wie sie auch in den internationalen Schulleistungsuntersuchungen gemessen wurden.

Wenn in einem ersten Schritt ein akzeptiertes Maß für Lehrerleistung gefunden ist, so ist im zweiten Schritt zu prüfen, ob und wie ein solches Maß als Basis für Zielvereinbarungen an den Schulen und schließlich für eine leistungsgerechtere Entlohnung von Lehrkräften genutzt werden kann. Variable Gehaltskomponenten, die auf Zielvereinbarungen basieren, sind heute be-

<sup>1</sup> Hanushek, E./Rivkin, S., Teacher Quality, in: Hanushek, E./Welsh, F. (Hrsg.), Handbook of the Economics of Education, Bd. 2, 2006.

<sup>2</sup> Rivkin, S./Hanushek, E./Kain, J. F., Teachers, schools, and Academic Achievement, *Econometrica* 73 (2), 2005, S. 417–458. Hanushek, E., The Failure of Input-Based Schooling Policies, *The Economic Journal* 113, 2003, S. F64–F98.

<sup>3</sup> Hanushek E., The Economics of Schooling: Production and Efficiency in Public Schools, *The Journal of Economic Literature*, 1989; Rockoff, J., The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data, *The American Economic Review* 94 (2), 2004, S. 247–252.

<sup>4</sup> Rivkin, S./Hanushek, E./Kain, J. F., a.a.O., 2005.

<sup>5</sup> Vgl. Dixit, A., Incentives and Organizations in the Public Sector: An interpretative review, *Journal of Human Resources* 37 (4), 2002, S. 696–727 und Diskussion in *Dolton, P./McIntosh, S./Chevalier, A.*, Teacher Pay and Performance, Bedford Way Papers, Institute of Education, University of London 2003, S. 26 ff.

reits als leistungsfördernd für viele Berufsgruppen akzeptiert.<sup>6</sup> Im öffentlichen Dienst und insbesondere im Schuldienst gibt es jedoch größere Widerstände gegen eine variable Gehaltsstruktur. Dennoch hat die OECD in einem Länderbericht eine stärkere Differenzierung bei der Lehrerbesoldung in Deutschland empfohlen und beschreibt das bestehende Karriereprofil wie folgt:

„Erstens ist die Lehrerbesoldung nur von einer sehr begrenzten Anzahl an Faktoren abhängig – Ausbildung, Alter und Seniorität. Die Übernahme zusätzlicher Aufgaben wird bis auf wenige Ausnahmen finanziell nicht belohnt. Die Besoldung eines Lehrers ist kaum an die Überprüfung seiner Leistung geknüpft, und die Teilnahme an beruflicher Fortbildung hat keinen Einfluss auf die Beförderungschancen. Auch wird bei der Besoldung nicht nach unterschiedlichen Aufgaben, die sich aus Besonderheiten des Lehrumfelds, der Lage der Schule, oder des unterrichteten Fachs ergeben, differenziert.

Zweitens sind Beförderungsmöglichkeiten äußerst begrenzt, eine Tatsache, die oft als Ursache für die mangelnde Attraktivität des Lehrerberufs genannt wird. Zwei Aspekte erscheinen als besonders problematisch. Während Gymnasial- und Berufsschullehrer in zwei höhere Besoldungsgruppen befördert werden können, haben Lehrer an Grund- und Sonderschulen, Hauptschulen und Realschulen keine Beförderungschancen, es sei denn sie übernehmen Funktionen des Schulmanagements. Eine Grundlage oder Rechtfertigung dieses Unterschieds konnten wir nicht erkennen. Ferner erfolgt die Beförderung zum Oberstudienrat im Allgemeinen nach mehreren Dienstjahren automatisch und ist nicht mit der Übernahme neuer Pflichten verbunden. Wenn heutzutage von Schulen verlangt wird, neue Aufgaben und Verantwortlichkeiten zu übernehmen, um den sich wandelnden Bedürfnissen der Schüler gerecht zu werden, dann erscheint der Mangel an beruflichen Entwicklungsmöglichkeiten für Lehrer als schwer zu rechtfertigen.“<sup>7</sup>

## 2 Theoretische Argumente für leistungsabhängige Bezahlung

Bevor wir im Folgenden die Struktur der Lehrerbesoldung im internationalen Vergleich und die Erfahrungen anderer Länder mit leistungsabhängiger Bezahlung darstellen, diskutieren wir zunächst theoretische Argumente für neue Formen der Lehrerbesoldung.

Löhne können input- oder outputorientiert sein. Ein Beispiel für Beschäftigte, die outputorientiert entlohnt werden, sind Mitarbeiter im Außendienst. Neben einem Fixum sind die Mitarbeiterinnen und Mitarbeiter am Umsatz oder Gewinn, den sie für das Unternehmen erwirtschaften, beteiligt. Im öffentlichen Dienst und damit auch bei der Lehrerbesoldung ist das nicht der Fall. Hier ist die inputorientierte Entlohnung vorherrschend. Etwas vereinfachend dargestellt ergibt sich die Höhe der Besoldung in Deutschland aus der erreichten Qualifikation (Lehrkraft für Primarstufe, Sekundarstufe I, Sekundarstufe II) und dem Alter der Lehrkraft als Maß für Berufserfahrung, ergänzt um Gehaltskomponenten in Abhängigkeit vom Familienstand. Eine systematisch ergebnisorientierte Entlohnung oder auch eine differenzierte Entlohnung nach Fächern gibt es nicht. Die derzeit praktizierte inputorientierte Entlohnung für Lehrerinnen und Lehrer ist administrativ in der Umsetzung unproblematisch, da die Inputs gut beobachtbar sind. Das ist bei der outputorientierten Entlohnung von Lehrkräften nicht der Fall.

<sup>6</sup> Lazear, E., Performance Pay and Productivity, *The American Economic Review* 90 (5), 2000, S. 1346–1361; Lazear, E., Teacher Incentives, *Swedish Economic Policy Review* 10, 2003, S. 179–214.

<sup>7</sup> OECD, Attracting, Developing and Retaining Effective Teachers. Country Note: Germany, 2004, S. 36. (Eigene Übersetzung durch die Autoren).

Was wären die zu erwartenden Vorteile einer stärker ergebnisorientierten Entlohnung von Lehrkräften?<sup>8</sup> Zunächst einmal ist festzuhalten, dass Lehrer in Deutschland als Beamte oder Angestellte im öffentlichen Dienst eine ausgeprägte Arbeitsplatzsicherheit genießen. Gepaart mit wenig Evaluation von Lehrern und einem fehlenden Accountability-Konzept der Schulen lassen sich kaum extrinsische Anreize für Schulen und Lehrkräfte erkennen, ihre Arbeit zu intensivieren.<sup>9</sup> Die bestehende intrinsische Motivation der Lehrerinnen und Lehrer<sup>10</sup> könnte jedoch durch extrinsische Motivation verstärkt werden. Problematisch sind Anreize von außen nur dann, wenn sie die intrinsische Motivation der Lehrer verdrängen. Dies kann zunächst nicht ausgeschlossen werden. Daher ist ein sorgfältiges Prüfen alternativer Besoldungsformen unabdingbar.

Welche Effekte sind durch variable Gehaltskomponenten zu erwarten? Hier sind zum einen Anreizeffekte und zum anderen aber auch nicht zu unterschätzende Selektionseffekte zu nennen. Die Anreizeffekte liegen auf der Hand. Können sich die Beteiligten darauf einigen, wie der Output gemessen werden, dann profitieren eigentlich alle Beteiligten. Die Lehrkräfte wissen besser einzuschätzen, was von ihnen erwartet wird und haben einen Anreiz, den Schülern gezielter zu helfen, die definierten Ziele zu erreichen.

Der zweite Aspekt betrifft die Selektion qualifizierter Kandidaten in den Lehrberuf. Die Herausforderung an die Bildungspolitik besteht nämlich nicht nur darin, die bereits aktiven Lehrerinnen und Lehrer zu mehr Engagement zu bewegen, sondern auch gute Kandidaten für den Lehrberuf zu gewinnen und an den Schulen zu halten. Die bestehende rechtliche oder faktische Unkündbarkeit, hohe Flexibilität durch weitreichende Möglichkeiten der Teilzeitbeschäftigung, ein relativ hohes Einkommensniveau in Kombination mit geringen Aufstiegschancen, aber leider auch zunehmender Imageverlust charakterisieren derzeit den Lehrberuf in Deutschland. Von den genannten Punkten werden häufig die Arbeitsplatzsicherheit und die Flexibilität als wichtiges Argument für die Wahl des Lehrerberufes genannt. Es ist allerdings fraglich, ob durch dieses Selektionskriterium die besten Lehrkräfte gewonnen werden können.<sup>11</sup> Eine Befragung von Schweizer Maturanden hat ergeben, dass sich vor allem Schülerinnen und Schüler, die bei der Wahl der Fächer am Gymnasium eine eher musische oder sozialwissenschaftliche Ausrichtung gewählt haben, für den Lehrberuf interessieren. Es liegt auf der Hand, dass Lehrer mit den genannten Neigungen nicht für alle Fächer die geeignetste Wahl darstellen.<sup>12</sup>

### 3 Besoldung der Lehrkräfte im internationalen Vergleich

Ob und wie potentielle Lehramtskandidaten und schon qualifizierte Lehrerinnen und Lehrer auf monetäre Anreize durch leistungsabhängige Bezahlung reagieren, ist nicht klar. Unstrittig ist jedoch, dass monetäre Aspekte bei der Entscheidung, Lehrer zu werden und zu bleiben, von nicht unerheblicher Bedeutung sind.<sup>13</sup>

Abbildung 1 zeigt die kaufkraftbereinigten Gehälter im Sekundarbereich I nach 15 Jahren Berufserfahrung und Mindestausbildung in den OECD-Ländern. Deutschland nimmt dabei mit

<sup>8</sup> Lazeur, a.a.O. (2003).

<sup>9</sup> OECD, Attracting, Developing and Retaining Effective Teachers; Country Note: Germany, 2004, S. 34.

<sup>10</sup> OECD, Teachers Matter: Attracting, Developing and Retaining Effective Teachers, 2005, S. 67 ff.

<sup>11</sup> Denzler, S./Fiechter, U./Wolter, S., Die Lehrkräfte von morgen. Eine empirische Untersuchung der Bestimmungsfaktoren des Berufswunsches bei bernischen Maturanden, Diskussionspapier Nr. 6, Universität Bern, Volkswirtschaftliches Institut, 2005.

<sup>12</sup> Denzler, S./Fiechter, U./Wolter, S., a.a.O. (2005).

<sup>13</sup> Santiago, P., Teacher Demand and Supply: Improving Teaching Quality and Addressing Teacher Shortages, OECD Education Working Paper Nr. 1, Abschnitt 4.3.2, 2002.

\$48.167 eine Position in der Spitzengruppe ein. Im Vergleich dazu erhalten Lehrer im OECD-Durchschnitt \$37.488, in den USA \$40.088 und in Finnland, dem PISA-Spitzenreiter, nur \$38.318.<sup>14</sup>

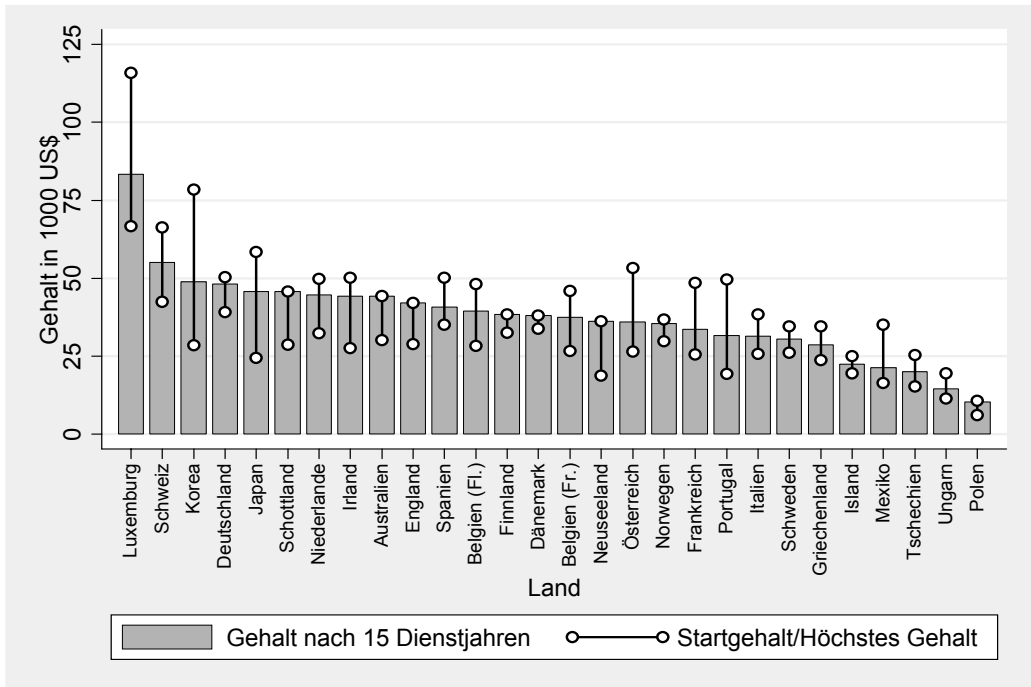


Abb. 1: Lehrergehälter (US-Dollar, kaufkraftbereinigt) im Sekundarbereich I, 2004, Quelle: *OECD, Education at a glance, OECD Indicators 2006, S. 378.*

Aber die Gehälter in Deutschland sind nicht nur im Vergleich mit anderen OECD Ländern hoch, sie sind auch gemessen am BIP pro Kopf hoch, also im Vergleich mit den „durchschnittlichen“ Einkommen eines Landes.<sup>15</sup> Eine Lehrkraft der Sekundarstufe II beispielsweise verdient das 1,8-fache des BIP pro Kopf. Der OECD-Durchschnitt liegt beim 1,42-fachen des BIP pro Kopf. Höhere relative Einkommen erzielen nur noch die Lehrkräfte in Korea (2,36), der Türkei (2,30) und der Schweiz (1,99). Lehrer in Finnland haben dagegen mit dem 1,46-fachen des BIP pro Kopf eine deutlich schlechtere relative Einkommensposition. Auch scheint es keinen Zusammenhang zwischen der relativen Bezahlung von Lehrkräften und Schülerleistungen in der PISA-Studie zu geben, wie Abbildung 2 zeigt.

<sup>14</sup> Die Gehaltsangaben sind kaufkraftbereinigt.

<sup>15</sup> *OECD, Education at a Glance, 2006, S. 374.* Eine detaillierte Studie zu relativen Löhnen von US-amerikanischen Lehrkräften, die Mikrodatsätze verwendet, ist *Podgursky, M., Is Teacher Pay „Adequate“?* Harvard University, Research Working Paper WP05–10, 2005.

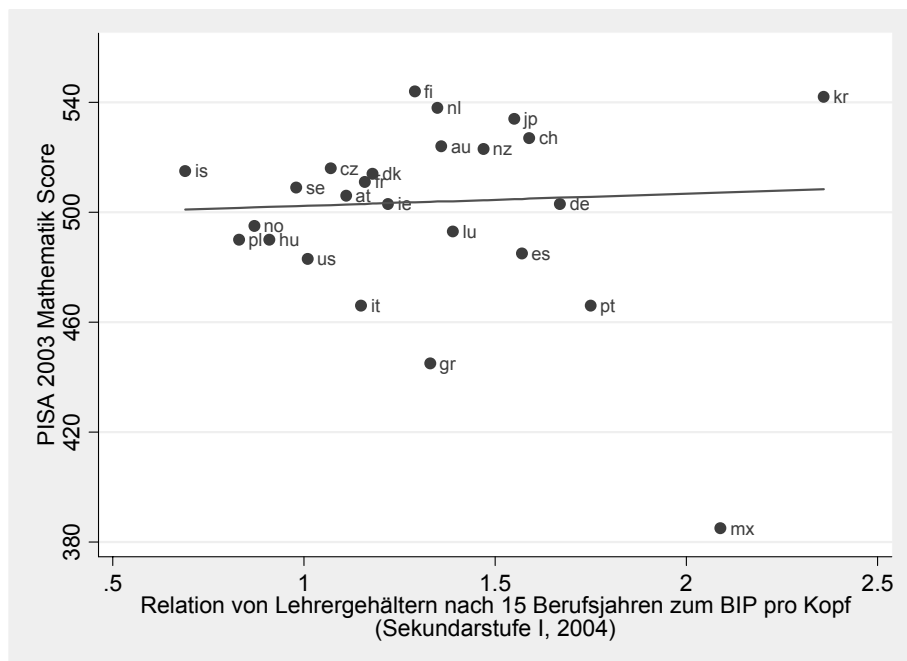


Abb.2: Lehrergehälter und Schülerleistungen. Datenbasis: *OECD* (2006): *Education at a Glance*.

Es ist zudem fraglich, ob eine weitere Erhöhung der Einkommen das Angebot an Lehrern erhöhen würde. Die geschätzten Arbeitsangebotselastizitäten für Lehrer in der Schweiz liegen nur zwischen 0.12 und 0.18.<sup>16</sup> Aber selbst wenn ein höheres Lohnniveau zu einem höheren Angebot an Lehrern führt, bedeutet mehr Quantität nur dann mehr Qualität, wenn auch bei der Auswahl der Lehrerinnen und Lehrer funktionierende Mechanismen am Werk sind.

Ein weiteres Argument gegen eine generelle Erhöhung der Bezüge ist der nicht gleichmäßig über die Fächer verteilte Bedarf an Lehrkräften. So geht aus einem Bericht des Landes Nordrhein-Westfalens hervor, dass der Bedarf an neuen Lehrkräften bis 2025 zwar groß ist, aber längst nicht jedes Fach gleich stark nachgefragt ist.<sup>17</sup> Da ein stark nachgefragter Mathematiklehrer aber nicht durch einen weniger stark nachgefragten Biologielehrer der Sekundarstufe I ersetzt werden kann, ist eine einheitliche Einkommenserhöhung für Lehrerinnen und Lehrer aller Fächer keine Lösung. Sie würde eine große Belastung für die öffentlichen Haushalte darstellen und das Problem der Lehrerknappheit in einigen Fächern und das Überangebot in anderen Fächern nicht lösen.

Analysiert man die Gehaltsstruktur weiter, so fällt auf, dass Deutschland zwar vergleichsweise hohe Gehälter zahlt, die Varianz in den Gehältern aber sehr gering ist.<sup>18</sup> Abbildung 1 zeigt auch das Karriereprofil in den Ländern der OECD. Im OECD-Durchschnitt steigt das Einkommen

<sup>16</sup> Wolter, S./Denzler, S., Wage Elasticity of the Teacher Supply in Switzerland, IZA Discussion paper Nr. 733, 2003. Für andere Länder werden höhere Werte berichtet. Eine Erklärung dafür dürfte das positive Lohndifferential in der Schweiz sein, wohingegen die relative Einkommensposition in anderen Ländern häufig schlechter ist.

<sup>17</sup> Ministerium für Schule und Weiterbildung des Landes Nordrhein-Westfalen, Lehrereinstellungsangebot und Lehrereinstellungsbedarf, NRW 2006–2025, Statistische Materialien, Nr. 7.

<sup>18</sup> Die Angaben beziehen sich jeweils auf Gehälter von Lehrkräften mit Mindestausbildung.

für Lehrerinnen und Lehrer in der Sekundarstufe I von \$27.560 auf ein Maximum von \$45.277, was einem Anstieg von etwa 70 % entspricht. Besonders flach ist das Profil in Deutschland und Dänemark. In Deutschland liegt das Einkommen für Berufsanfänger in der Sekundarstufe I bei einem Einkommen von \$39.132 und kann bis auf \$50.284 steigen. Die Differenz zwischen dem geringsten und dem höchsten Gehalt, das eine Lehrkraft der Sekundarstufe I beziehen kann, beträgt in Deutschland nur 28 %. Deutschland zahlt seinen Lehrerinnen und Lehrern somit vergleichsweise hohe aber wenig differenzierte Gehälter.

Interessant sind auch Informationen darüber, welche Gehaltskomponenten zu Variation in den Lehrergehältern führen. Dabei sind in der OECD erhebliche Unterschiede erkennbar.<sup>19</sup> Für Lehrkräfte in Deutschland resultieren Einkommenszuwächse, wenn zusätzliche Verwaltungsaufgaben übernommen werden, mehr Stunden unterrichtet werden als vertraglich vorgesehen, sich der Familienstand ändert und wenn die nächste Altersstufe erreicht wird. Es sind aber weder Zuschläge für das Arbeiten mit besonderen pädagogischen Anforderungen vorgesehen, noch finden Kriterien der Qualifikation, Ausbildung oder Leistung der Lehrkräfte Anwendung. Monetäre Anreize für besonders leistungsstarke Lehrer gibt es in Deutschland im Allgemeinen nicht. Zwar ist Deutschland damit keine Ausnahme, es ist jedoch festzuhalten, dass immerhin in 13 von 32 Ländern der OECD Leistungskomponenten bei der Entlohnung berücksichtigt werden. Die bereits bestehenden Anreizsysteme unterscheiden sich stark zwischen den Ländern. Wir werden im Folgenden die Anreizsysteme und die Erfahrungen mit leistungsgerechterer Bezahlung in ausgewählten Ländern beschreiben.<sup>20</sup>

#### 4 Erfahrungen mit merit pay

Es gibt drei Grundmodelle für Systeme leistungsabhängiger Vergütung von Lehrern.<sup>21</sup> Erstens sind das Systeme mit individueller leistungsabhängiger Entlohnung (merit pay; oft aber auch als performance pay beschrieben), die z.B. auf Indikatoren der Schülerleistungen beruht oder der Evaluationen durch Dritte zugrunde liegen. Zweitens können individuelle Zusatzqualifikationen und Weiterbildung belohnt werden (knowledge-and skill-based) und drittens gibt es in einigen Ländern auch Leistungsprämien auf Schulebene. Die umstrittenste aber auch die direkteste Form der finanziellen Leistungsanreize ist die individuelle leistungsabhängige Entlohnung, auf die sich die Darstellung im Folgenden beschränkt.<sup>22</sup> Dabei liegt der Schwerpunkt auf den Programmen, für die uns wissenschaftliche Auswertungen vorliegen.

Kritik an der Einführung von merit pay wird vorwiegend von den Gewerkschaften geübt. Aber auch Wissenschaftler äußern sich zum Teil recht skeptisch. Die Argumente, die gegen die Einführung von merit pay ins Feld geführt wurden, sind: die Schwierigkeit, die Leistung von Lehrerinnen und Lehrern angemessen zu erfassen und zuzuordnen (Stichwort: Mehrdimensionalität des Lernerfolgs); das Resultieren neuer, unerwünschter Hierarchien an den Schulen, die demotivierend für die Lehrkräfte sind; Wettbewerb, der zu Neid und sozialem Konkurrenzverhalten zwischen den Lehrkräften führen kann; erhebliche Mehrkosten für die öffentlichen Haushalte. Folgt man diesen Argumenten, so ist nicht auszuschließen, dass finanzielle Anreizsysteme zu unerwünschten Ergebnissen führen, weil das Lehr- und Lernumfeld nicht verbessert

<sup>19</sup> OECD, a.a.O. (2006), Tabellen D3.2a–D3.2d.

<sup>20</sup> Einen Überblick bietet *Eurydice*, Der Lehrerberuf in Europa: Profil, Tendenzen und Anliegen. Bericht III: Beschäftigungsbedingungen und Gehälter, 2003.

<sup>21</sup> OECD, *Teachers Matter*, 2005, S. 184

<sup>22</sup> OECD, Absatz 6.2.3 und auch 6.2., 2005.

wird, sondern sich eher verschlechtert.<sup>23</sup> Die folgenden Beispiele von merit-pay-Systemen sollten auch vor dem Hintergrund dieser Kritikpunkte beurteilt werden.

Ein europäisches Beispiel für die Einführung leistungsabhängiger Bezahlung für Lehrer ist das performance-related pay (PRP) in England und Wales. Dort gibt es seit dem akademischen Jahr 1999/2000 leistungsabhängige Bezahlung für Lehrkräfte. PRP besteht aus zwei Komponenten.<sup>24</sup> Erstens gibt es nach Erreichen der obersten Gehaltsstufe der allgemeinen Gehaltsskala eine Leistungsschwelle, den Performance Threshold, jenseits dessen zweitens eine zusätzliche Gehaltsskala, die Upper Pay Scale, eingeführt wurde. Die Überschreitung des Performance Threshold ist für die Lehrkräfte mit einer dauerhaften Gehaltserhöhung von £2.000 verbunden, die auch in die Berechnung der Pensionsansprüche eingeht. Ist eine Lehrkraft in der Upper Pay Scale angelangt, so ist die weitere Gehaltsentwicklung leistungsabhängig. „Points may be awarded to recognise substantial and sustained performance and contribution to the school as a teacher.“<sup>25</sup>

Ziel der Reform war es, ein funktionierendes Anreizsystem zu schaffen, den Lehrerberuf für Berufseinsteiger attraktiv zu machen und qualifizierte Lehrerinnen und Lehrer im Schuldienst zu halten. Vor der Reform lag das Einkommen der Lehrkräfte einheitlich zwischen £14.658 und £23,193 und hing von der Qualifikation und der Berufserfahrung der Lehrkräfte ab. Typischerweise erhielten die Lehrkräfte eine jährliche Lohnerhöhung und 1999 waren etwa 75 % der Lehrerinnen und Lehrer in der höchsten Einkommenskategorie, die nach etwa 5 Jahren Berufserfahrung erreicht wird. Nach der Reform konnten die Einkommen auf bis zu £30.000 pro Jahr steigen. Die durchschnittlichen Einkommen von Angestellten lagen zum Vergleich bei etwas über £25.000.

Um den Performance Threshold zu überschreiten, müssen sich die Lehrkräfte der höchsten ursprünglichen Einkommenskategorie zunächst bewerben und sich in fünf Bereichen ausweisen: „knowledge and understanding of teaching; teaching management and assessment; wider professional effectiveness; professional characteristics; and pupil progress“.<sup>26</sup> Dabei sollte der Lernfortschritt der Schülerinnen und Schüler nicht unter dem Durchschnitt vergleichbarer Schüler in England liegen. Über die Anträge der Lehrkräfte entscheiden im Wesentlichen die Schulleitungen, zum Teil unterstützt durch externe Gutachter. Den Schulleitungen die Entscheidungen über die Beförderung von Lehrerinnen und Lehrern zu überlassen, ist aber vor dem Hintergrund der Kritik an PRP nicht unproblematisch, denn sie birgt ein nicht unerhebliches Konfliktpotential für das Kollegium.

Anders als von den meisten Lehrkräften erwartet, lösten die Schulleiter nach der Einführung des Programms das Problem zunächst durch weitestgehende Befürwortung der Anträge. In der ersten Welle im Jahr 2000 stellten 80 % der antragsberechtigten Lehrerinnen und Lehrer einen Antrag, der auch in 95 % der Fälle positiv beschieden wurde. Es ist also fraglich, ob dieser Teil der Reform tatsächlich eine Leistungskomponente beinhaltet oder nicht. Offenbar handelte es sich ex post um eine allgemeine Lohnerhöhung, die nur wenig positive Anreizeffekte hat. Erste

<sup>23</sup> Eine Zusammenfassung der Argumente für und gegen performance pay findet sich in *Harvey-Beavis, O.*, Performance-Based Rewards for Teachers: A Literature Review, Working Paper OECD, 2003.

<sup>24</sup> *Marsden, D./Belfield, R.*, Performance Pay for Teachers: Linking Individual and Organisational-Level Targets, CEP Discussion Paper No 703, London School of Economics, 2005; *Atkinson, A./Burgess/S./Croxon, B./Gregg, P./Propper, C./Slater, H./Wilson, D.*, Evaluating the Impact of Performance Related Pay for Teachers in England, CMPO Working Paper Nr. 04/113, University of Bristol 2004.

<sup>25</sup> *School Teachers' Pay and Conditions of Employment 2001: Revised Guidance*, DfES/0696/2001. [http://www.teachernet.gov.uk/\\_doc/4013/pay\\_guidance.doc](http://www.teachernet.gov.uk/_doc/4013/pay_guidance.doc).

<sup>26</sup> *Atkinson et al.*, a.a.O. (2004), S. 10.

wissenschaftliche Untersuchungen zeigen, ob und wie das neue Anreizsystem wirkt.<sup>27</sup> Marsden und Belfield haben eine Wiederholungs-Befragung von Lehrkräften und Schulleitern in England und Wales ausgewertet. Die Daten wurden zu drei Zeitpunkten erhoben: vor Einführung der Reformen, nach der Einführung der Reform, aber vor den ersten Entscheidungen über die Beförderungen, und nachdem das System schon eine Zeitlang etabliert war. Waren die befragten Lehrer und auch die Schulleiter anfangs sehr skeptisch, so äußerten sie sich im Verlauf der Zeit deutlich positiver über PRP (vgl. Tabelle 1).<sup>28</sup> Sowohl die generelle Abwehrhaltung als auch die Sorge, das System könne nicht fair sein, schwinden nach den ersten Erfahrungen zunehmend. Problematisch bleibt jedoch die Einschätzung von immerhin 76 % der Befragten, dass der Beitrag Einzelner nur schwer zu ermitteln sei.

**Tabelle 1: Einstellung der englischen Lehrkräfte zu Performance Management**

Lehrkräfte	Stimme nicht zu	Weiß nicht	Stimme zu
<b>Prinzipiell ist die Idee, Lehrer leistungsabhängig zu bezahlen, gut</b>			
Welle 1	65	13	22
Welle 2	55	16	29
Welle 3	44	18	39
<b>Schulleiter werden ihre Günstlinge bevorzugen</b>			
Welle 1	16	28	55
Welle 2	22	36	42
Welle 3	48	32	20
<b>Es ist schwierig, den Beitrag Einzelner zu identifizieren</b>			
Welle 1	4	6	90
Welle 2	8	7	84
Welle 3	14	10	76

Quelle: Marsden/Belfield, a.a.O., S. 9, Table 1

Die befragten Lehrkräfte gaben zwar nicht an, sich durch das monetäre Anreizsystem wirklich motiviert zu fühlen, aber es gibt Hinweise dafür, dass PRP dazu beiträgt, die gemeinsamen Ziele deutlicher herauszuarbeiten und umzusetzen. Auch Anzeichen für einen positiven Einfluss auf die Leistungen der Schülerinnen und Schüler durch die Reformen lassen sich finden.<sup>29</sup> Sowohl das Niveau der Testergebnisse als auch die Leistungszuwächse stiegen durch die Einführung des PRP signifikant an.

<sup>27</sup> Marsden/Belfield, a.a.O. (2005).

<sup>28</sup> Marsden/Belfield, a.a.O. (2005), Tabelle 1, S. 9.

<sup>29</sup> Atkinson et al., a.a.O. (2004).

Trotz dieser recht optimistischen Einschätzung des Programms dürfen Schwächen der Reform nicht unerwähnt bleiben: die Anreize für die Schulleiter und die fiskalischen Konsequenzen.<sup>30</sup> Da der Aufstieg in die Upper Pay Scale ursprünglich ausschließlich von Kriterien der Leistungsmessung und der Einschätzung durch die Schulleiter abhängen sollte, war die Anzahl der Lehrkräfte, die sich für zusätzliche Gehaltskomponenten qualifizieren würden, und damit auch der finanzielle Umfang des Programms nicht festgelegt. Schulleiter, die innerschulische Konflikte fürchteten, reagierten, indem sie alle formal qualifizierten Lehrer für die nächste Gehaltsstufe empfahlen. Die Mehrkosten der Reform drohten aus dem Ruder zu laufen. Deshalb ergänzte das Department for Education and Skills seinen Kriterienkatalog bereits kurz nach der Einführung um den Zusatz der Finanzierbarkeit. Die zur Verfügung stehenden Mittel wurden festgelegt und der Aufstieg in die nächste Gehaltsgruppe damit um ein weiteres Element ergänzt. Es bestand nun Konkurrenz zwischen den Lehrerinnen und Lehrern um knappe Mittel und damit die Erfordernis, innerhalb einer Schule ein Lehrerranking zu erstellen. Nach dem Versuch, relativ konfliktfrei in die leistungsabhängige Lehrbesoldung einzusteigen, wuchs also recht schnell die Einsicht, dass das Ziel der Reform nur erreicht werden kann, wenn auch stärker konfliktträchtige Elemente in der Besoldung der Lehrkräfte zugelassen würden. Natürlich hat die Reform in England ihr Ziel noch nicht erreicht und sie hat bislang wohl eher zu einer generellen Gehaltserhöhung für Lehrkräfte geführt und weniger zu leistungsabhängiger Bezahlung. Es ist jedoch durch die neuen Strukturen, die laufend weiterentwickelt werden, eine solide Grundlage für leistungsgerechtere Bezahlung gelegt worden.

Deutlichere Hinweise auf die positive Wirkung von leistungsabhängiger Bezahlung finden sich in den Studien von Lavy.<sup>31</sup> In den Klassen 10 bis 12 israelischer Schulen wurde eine experimentelle Studie mit relativen Leistungsturnieren in verschiedenen Fächern durchgeführt. In jedem Leistungsturnier konnten die Lehrkräfte bis zu \$7.500 hinzuverdienen, was bei einem Jahresgehalt von \$30.000 durchaus erheblich ist. Lavy zeigt, dass durch die Anreize sowohl die Prüfungsleistungen stiegen als auch die Schulabbrecherquote sank. Insbesondere Schülerinnen und Schüler aus sozial schwachen Verhältnissen profitierten von den Leistungsanreizen für ihre Lehrerinnen und Lehrer. Die finanziellen Anreize haben also in der Tat zu besseren Leistungen der Lehrkräfte geführt und die Lehrerinnen und Lehrer erhöhten ihre Anstrengungen dort, wo die erwarteten Erträge am höchsten waren: bei den schwächeren Schülern und Schülerinnen. In einem anderen Experiment wurden ganze Schulen und nicht einzelne Lehrkräfte für besondere Leistungen (durch eine bessere Ausstattung) belohnt.<sup>32</sup> Auch hier ließen sich positive Effekte auf verschiedene Erfolgsmaße finden. Lavy zeigt jedoch, dass Programme, die mit den Anreizen beim individuellen Lehrer und nicht bei der Schule ansetzen, kosteneffektiver waren.<sup>33</sup> Dieses Ergebnis bestätigt, dass individuelle Anreize stärker sind als schulbezogene. Dies ist auch theoretisch zu erwarten. Insbesondere dann, wenn die durch gute Leistungen zusätzlich erworbenen Ressourcen nur der Schule allgemein und nicht der Lehrkraft persönlich zugute kommen, entsteht ein typisches Gefangenendilemma.

Die Kosten des Programms in Israel beliefen sich auf \$170 pro Schüler und die Teilnahme am Programm erhöhte die Wahrscheinlichkeit, einen Schulabschluss zu erreichen, um 4 %-Punkte. Nach Angrist und Lavy<sup>34</sup> liegt der geschätzte Ertrag eines Schulabschlusses bei \$4.025 pro Jahr.

<sup>30</sup> *Cutler, T./Waine, B.*, Not so Seamless? Performance Related Pay and Financial Control in English Schools, Public Policy & Management, January, 2005, S. 67–71.

<sup>31</sup> *Lavy, V.*, Evaluating the Effect of Teachers' Group Performance Incentives on Pupil Achievement, Journal of Political Economy, 110, 2002, S. 1286–1317; *Lavy, V.*, Performance Pay and Teachers' Effort, Productivity and Grading Ethics, NBER Working Paper 10622, 2004.

<sup>32</sup> *Lavy, a.a.O.* (2002).

<sup>33</sup> *Lavy, a.a.O.* (2004), S. 30.

<sup>34</sup> *Angrist, J./Lavy, V.*, New Evidence on Classroom Computers and Pupil Learning, The Economic Journal, 112, 2002, S. 735–76.

4 % dieses Betrages sind \$161, d.h. das Programm amortisiert sich gesamtwirtschaftlich nach ungefähr einem Jahr. Die Kosten des Programms lassen sich also gut rechtfertigen.

Auch aus den USA gibt es Erfahrungen mit merit pay-Systemen. Allerdings gibt es in den USA, anders als in England, kein einheitliches System, sondern eine große Vielfalt verschiedener, häufig kleinräumiger Projekte. Entsprechend vielfältig ist auch die Beurteilung der Wirkung von merit pay auf die Schülerleistungen. Bereits 1989 gab es in den USA in 12 % aller Schulbezirke, verteilt auf 29 Bundesstaaten, merit pay-Systeme, die durch die Lehrergewerkschaften harsch kritisiert und oft nach kurzer Zeit wieder aufgegeben oder aber grundlegend verändert wurden, häufig weil sie konzeptionell schlecht waren und auch die praktische Umsetzung nicht zufriedenstellend war.<sup>35</sup> So betrug die Leistungskomponente durchschnittlich nur 2 % des Grundgehältes eines Lehrers.<sup>36</sup> Der Versuch, alle „neueren“ merit pay-Systeme in den USA ausführlich darzustellen, sprengt den Rahmen dieses Beitrags. Deshalb greifen wir hier nur einige Ergebnisse, von denen auch deutsche Reformbemühungen profitieren können, heraus.

Eine Fallstudie aus Michigan zeigt, dass Lehrerinnen und Lehrer in ihren Anstrengungen durchaus auf finanzielle Anreize reagieren.<sup>37</sup> Andererseits zeigt sich aber auch, dass Anreizsysteme nicht erfolgreich sind, wenn sie nicht zielgenau formuliert sind. In der genannten Fallstudie wurde eine geringere Abbrecherquote belohnt, nicht aber bessere Noten oder geringere Durchfallquoten. Im Ergebnis verbesserte sich auch nur die Abbrecherquote und die anderen Indikatoren verbesserten sich nicht parallel, wie ursprünglich erhofft. Die Wahl des Leistungsindikators ist also von großer Bedeutung und es werden sich auch nicht alle Ziele über die Fokussierung auf einen einzigen Output steuern lassen.

Ein wichtiges Kriterium für den Erfolg von finanziellen Anreizsystemen ist die Frage, ob durch das System wirklich die leistungsstarken Lehrkräfte identifiziert werden können.<sup>38</sup> Im Career Ladder System in Tennessee konnten Lehrkräfte sich zunächst entscheiden, ob sie an dem Programm teilnehmen wollten. Bei Teilnahme bestand dann die Möglichkeit, einen Bonus von bis zu \$7.000 hinzuverdienen. Über das Erreichen der nächsten Karrierestufe wurde auf den unteren Stufen in der Regel von den Schulleitern entschieden; bei den oberen Stufen waren es dann dreiköpfige Teams von Lehrkräften aus anderen Schulbezirken. Ähnlich wie in England und Wales zeigte sich, dass gerade die Evaluationen durch die Schulleitung sehr hohe Erfolgsraten (95 %) produzierten. Warum sollten Schulleiter ihren Lehrkräften den Aufstieg verwehren? Sobald externe Gutachter hinzukamen, ging die Erfolgsquote auf 79 % zurück, blieb aber dennoch sehr hoch. Die Wirkung des Programms in Bezug auf die Lernleistungen war nicht einheitlich. So konnte ein signifikanter Anstieg der Mathematikleistungen festgestellt werden, wenn Schüler von Lehrkräften unterrichtet wurden, die am Career Ladder System teilnahmen. Bei den Leseleistungen waren die Effekte weniger stark ausgeprägt.

Ein immer wieder kritisch diskutierter Aspekt ist die Einschätzung von Lehrerleistungen durch die Schulleiter. Diese hat den Vorteil, der Mehrdimensionalität des Outputs besser Rechnung tragen zu können. Schulleiter können auch wichtige Eigenschaften wie die Teamfähigkeit oder den Erfolg der Lehrkraft bei der Vermittlung sozialer Kompetenzen beurteilen. Leistungstests von Schülern hingegen können „nur“ den Erfolg in der Vermittlung von Basiskompetenzen in den Schulfächern bewerten. Eine Beurteilung möglicher Verfahren zur Erfassung von Lehrer-

<sup>35</sup> OECD, *Teachers matter*, 2005, S. 184.

<sup>36</sup> Dee, T. S./Keys, B. J., Does Merit Pay Reward Good Teachers? Evidence from a Randomized Experiment, *Journal of Policy Analysis and Management* 23 (3), 471–488.

<sup>37</sup> Eberts, R./Hollenbeck, K./Stone, J., Teacher Performance Incentives and Student Outcomes, *The Journal of Human Resources*, 2002, S. 913–927.

<sup>38</sup> Dee/Keys, a.a.O.

qualität findet sich in einer Studie von Jacob und Lefgren<sup>39</sup>. In dieser Studie wurden die Einschätzungen der Lehrerleistung durch den Schulleiter mit den traditionellen Determinanten der Lehrerbildung – Bildungsabschlüsse und Berufserfahrung – sowie mit value-added Maßen der Schülerleistungen verglichen. Indikatoren, die auf der formalen Qualifikation beruhen, waren deutlich schlechter als die beiden anderen Indikatoren, wenn sie zur Prognose der Leistungen von Schülern und Schülerinnen herangezogen wurden. Die Einschätzung der Lehrqualität durch die Schulleiter war dagegen am genauesten, wenn die Gruppen der schwächsten und der stärksten Lehrkräfte identifiziert werden sollen. Innerhalb des Mittelfeldes der Lehrkräfte sind Prognosen auf Basis von value-added Maßen am besten geeignet. Die Autoren schließen daraus, dass Einschätzungen durch Schulleiter sehr wohl als Basis für merit pay fungieren könnten. Nun muss aber bedacht werden, dass die der Studie zugrunde liegende Befragung ohne Konsequenzen für alle Beteiligten war. Die Ergebnisse der Beurteilung durch die Schulleiter wurden den Lehrerinnen und Lehrern nicht einmal mitgeteilt. So lässt sich aus der Studie zwar schließen, dass die Schulleiter die Qualität ihrer Lehrkräfte durchaus kennen, es ist aber nicht klar, wie die Anreize für Schulleiter gesetzt sein müssen, diese „objektive“ Einschätzung auch mit allen Konsequenzen zu vertreten. Keines der bisherigen Programme, die sich auf die Einschätzung durch Schulleiter verlassen haben, lässt erkennen, dass differenzierte Leistungsbewertungen zu erwarten sind.

Ein erster Versuch, den Zusammenhang zwischen merit pay-Systemen und Schülerleistungen großräumig für die USA zu analysieren, zeigt, dass Schülerleistungen steigen, wenn Lehrerinnen und Lehrer finanzielle Leistungsanreize bekommen.<sup>40</sup> Insbesondere Schülerinnen und Schüler aus einkommensschwachen Familien profitieren von den Anreizen für ihre Lehrkräfte. Für nicht-monetäre Anreize, wie Unterrichtsbesuche oder Entlassungen von Lehrkräften konnte dagegen kein statistisch gesicherter Zusammenhang mit Schülerleistungen gefunden werden.

## 5 Indikatoren: Schülerleistung

Neben der Bewertung durch Schulleiter ist der aus unserer Sicht wichtigste aber auch umstrittenste Indikator für *Lehrerfolg* der *Lernerfolg* der Schüler und Schülerinnen. Voraussetzung für die Vergleichbarkeit von Lernerfolg über den Klassenverband hinaus sind regelmäßige standardisierte Leistungsüberprüfungen, sei es auf Länder- oder auf Bundesebene, deren Ergebnisse systematisch ausgewertet werden. Diese Voraussetzung ist in Deutschland – anders als in vielen anderen Ländern – bislang noch nicht ausreichend gegeben.

Die oben diskutierte Literatur zeigt, dass Lehrkräfte auf finanzielle Anreize reagieren, sich aber nicht unbedingt alle gewünschten Ziele automatisch und gleichzeitig erreichen lassen. Entscheidet man sich für den Lernerfolg der Schülerinnen und Schüler als primäre Zielgröße, so bleibt zu entscheiden, ob Lernerfolg nominal (zum Beispiel durch Schulabschlüsse) oder kontinuierlich (durch die erreichte Note oder Punktzahl) gemessen werden soll. Auch diese Entscheidung ist für den Erfolg des Systems wichtig. Wenn das formulierte Ziel nämlich die Anzahl der bestandenen Prüfungen ist, bestehen Anreize, sich vorrangig um die marginalen Schülerinnen und Schüler (d.h. Schülerinnen und Schüler an der Bestehensgrenze) zu kümmern, was zu Lasten der leistungsschwächeren aber auch der leistungsstärkeren Schülerinnen und Schüler geht. Diese unerwünschten Effekte lassen sich durch stärker differenzierende Leistungsmaße vermeiden.<sup>41</sup>

<sup>39</sup> Jacob, B./Lefgren, L., Principals as Agents: Subjective Performance Measurement in Education, RWP05–040, Harvard University 2005.

<sup>40</sup> Figlio, D./Kenny L., Individual Teacher Incentives and Student Performance, NBER Working Paper 12627, 2006.

<sup>41</sup> Burgess, S./Propper, C./Slater, H./Wilson, D., Who wins and who loses from school accountability? The distribution of educational gains in English secondary schools, University of Bristol 2005.

Erheblicher Diskussionsbedarf besteht des Weiteren bei der Zurechnung von Schülerleistungen zu den einzelnen Lehrerinnen und Lehrern. Schülerleistungen, die zu einem Zeitpunkt gemessen werden, sind das Ergebnis des gesamten Bildungsprozesses, an dem nicht nur eine einzelne Lehrkraft beteiligt, sondern der das Ergebnis von impliziter „Teamarbeit“ ist. Auch darf nicht vergessen werden, dass das Leistungsniveau in erheblichem Maße den sozio-ökonomischen Hintergrund der Schülerinnen und Schüler widerspiegelt. Reine Leistungsniveaus, zu einem Zeitpunkt gemessen, sind als Grundlage für die Beurteilung von Lehrkräften unangemessen.

Dennoch ist die Messung von Schülerleistungen die Voraussetzung für Maße, die eine Bewertung der Leistung einzelner Lehrkräfte ermöglichen können. Um dies vorwegzunehmen: eine hundertprozentige Zurechnung zu einzelnen Lehrkräften ist wohl nicht zu erreichen, aber durch value-added Maße, die beispielsweise in den USA zentraler Bestandteil für die Leistungsbewertung von Schulen sind, können gehaltvolle Aussagen getroffen werden. Die Grundidee von value-added Maßen ist, dass zwar nicht das Leistungsniveau von Schülern und Schülerinnen einer einzelnen Lehrkraft oder einer Schule zugerechnet werden kann, wohl aber die Leistungszuwächse. Die korrekte Implementierung von value-added Maßen ist nicht trivial, wie eine Vielzahl von wissenschaftlichen Studien zeigt.<sup>42</sup> Alternativen zu value-added Maßen sind jedoch bislang nicht bekannt, und es wird, trotz der anerkannten Schwächen, selbst von kritischen Stimmen nicht gefordert, die Leistungsmessung durch value-added wieder aufzugeben.

Mit der Möglichkeit der Lehrqualitätsmessung durch Schülerleistungen in Deutschland befasst sich eine Studie von Jürges und Schneider (2006).<sup>43</sup> Die Studie verwendet die deutschen Daten der PIRLS Studie und sucht Antworten auf die Frage nach einem fairen Ranking von Lehrerinnen und Lehrern. Kritik an performance pay wird nämlich häufig an der Frage festgemacht, ob Schülerleistungen wirklich die Qualität der Lehrkräfte widerspiegeln oder nicht einfach nur Resultat der sozialen Herkunft der Schüler sind. Dieser Punkt ist von großer Bedeutung. Wird der soziale Hintergrund der Kinder nicht kontrolliert, sind keine validen Aussagen über die Qualität der Lehrerinnen und Lehrer möglich. Die Informationen, die erforderlich sind, um den sozio-ökonomischen Hintergrund der Schüler zu erfassen, sind leicht zu erheben. Wichtig ist dabei, dass die Informationen über die Schüler auf individueller Ebene vorliegen. Im Wesentlichen handelt es sich um Informationen, wie sie im derzeit diskutierten Schülerregister vorgesehen sind. Es ist nicht ausreichend, die Zusammensetzung der Schülerschaft einer Schule im Aggregat aus einer Befragung der Lehrer oder Schulleiter zu erheben. Liegen die Informationen nur auf Schulebene vor, so kann zum einen die Variation der Lehrqualität innerhalb einer Schule nicht erfasst werden. Zum anderen bestehen auch Anreize für die Schulleitung, die Zusammensetzung der Schülerschaft zu „schönen“.

Nutzt man die relevante Information, um ein faires Lehrerranking für die Lehrkräfte in der PIRLS-Studie zu erstellen, so können drei Gruppen statistisch gesichert unterschieden werden. Die durchschnittlichen, die unterdurchschnittlichen und die überdurchschnittlichen Lehrkräfte. In die Gruppe der überdurchschnittlichen Lehrer fallen 21 % der Lehrkräfte und 15 % werden als unterdurchschnittlich klassifiziert. Durch die Berücksichtigung der Kontextvariablen ändert sich das Ranking der Lehrerinnen und Lehrer deutlich im Vergleich zu einem Ranking, das die reinen Testergebnisse zur Grundlage hat. Weniger als zwei Drittel der Lehrer finden sich nach

<sup>42</sup> Ladd, H./Walsh, R., Implementing Value-Added Measure of School Effectiveness: Getting the Incentives Right, *Economics of Education Review*, 21, 2002, S. 1–17; Kane, T./Staiger, D., Volatility in School Test Scores: Implications for Test-Based Accountability Systems, *Brookings Papers on Education Policy*, 2002a, S. 235–283; Kane, T./Staiger, D., The Promise and Pitfalls of Using Imprecise School Accountability Measures, *Journal of Economic Perspectives* 16, 2002b, S. 91–114.

<sup>43</sup> Jürges, H./Schneider, K., Fair Ranking of Teachers, erscheint in *Empirical Economics*, 2007.

dem fairen Ranking in derselben Gruppe wieder. Durch die faire Korrektur des Rankings profitieren vor allem die Lehrerinnen und Lehrer, die in schwierigeren Lernumfeldern arbeiten. Damit greift ein faires Ranking die berechtigte Sorge vieler Kritiker auf.

## 6 Voraussetzungen in Deutschland und erste Schritte

Internationale Bildungsexperten empfehlen Deutschland, die Besoldungsstruktur für Lehrerinnen und Lehrer zu reformieren, um stärkere Leistungsanreize zu setzen. Deutschland setzt bei seiner Lehrerbesoldung weniger auf finanzielle Anreize als andere Länder und lässt auch eine Kultur der Lehrerevaluation und der Leistungsmessung vermissen. Regelmäßige Evaluationen und die Messung von Lernerfolg durch Vergleichsarbeiten wären aber in einem ersten Schritt einzuführen.

Die Öffentlichkeit hat längst begonnen, mehr Transparenz im Schulsystem zu fordern. Freie Schuwahl ist für Eltern von zweifelhaftem Wert, wenn die Qualität einer Schule kaum objektiv beurteilt werden kann. Einen ersten Schritt, diese Forderung nach mehr Transparenz umzusetzen, erkennt man in der zunehmenden Akzeptanz zentraler Abschlussprüfungen und auch in der regelmäßigen Durchführung von Vergleichsarbeiten. Auch die von der KMK geplante Einführung eines nationalen Bildungsregisters mit der Erhebung von Informationen über den sozioökonomischen Hintergrund der Schülerinnen und Schüler ist vor diesem Hintergrund eindeutig zu begrüßen. Mit der Einführung des Bildungsregisters wäre eine systematische wissenschaftliche Analyse von Bildungsbiographien und Stärken und Schwächen des Bildungssystems möglich. Erst die vollständige Erfassung von Schülerleistungen und wichtiger Hintergrundinformationen macht eine wirksame Reform der Besoldung überhaupt möglich.

Solange die Ergebnisse nur erhoben werden, ist die Anreizwirkung auf Schulen und Lehrkräfte nicht optimal genutzt. Gefordert ist vielmehr eine Veröffentlichung der Ergebnisse. Die Literatur zum Thema School Accountability, also dem Versuch, die Leistung von Schulen zu messen und die Ergebnisse zu veröffentlichen, ist umfangreich und es gibt viele kritische Stimmen. Aber so groß die Skepsis gegen das Messen von Leistung – einem integralen Bestandteil von Accountability – auch ist, die positiven Aspekte scheinen aus Sicht der Wissenschaft und auch der Öffentlichkeit zu überwiegen.<sup>44</sup>

Ist dieser erste Schritt getan, fällt es leichter darüber nachzudenken, wie der Lehrerberuf durch moderne Entlohnungsstrukturen wieder für qualifizierte Kräfte attraktiv gemacht und gehalten werden kann. Das erfordert auch, nach dem Karrierebeginn noch hinreichend Raum für differenzierte Entlohnung zu bieten. Dem Einwand, dass extrinsische Motivation die intrinsische verdrängt, begegnet man mit sorgfältig entwickelten, fairen Leistungsindikatoren, die das Arbeitsumfeld der Lehrkraft bestmöglich berücksichtigen. Bevor in Deutschland finanzielle Anreize implementiert werden, kann zudem versucht werden, mit leistungsabhängigen nicht-finanziellen Anreizen zu arbeiten. Allein die Messung von Leistung führt (kausal) zu besseren Leistungen, wie ein Vergleich der Schülerleistungen zwischen Ländern mit und ohne Zentralprüfungen zeigt.<sup>45</sup> Die Veröffentlichung (fairer) Rankings beispielsweise in Form von Bestenlisten ist eine weitere Form der nicht-monetären Leistungsanreize. Hier kann in einem ersten Schritt auch

<sup>44</sup> Hoxby, C., The Cost of Accountability, NBER Working Paper 8855, 2002.

<sup>45</sup> Jürges, H./Richter, W./Schneider, K., Teacher Quality and Incentives: Theoretical and Empirical Effects of Standards on Teacher Quality, FinanzArchiv 61, 2005, S. 298–326; Jürges, H./Schneider, K./Büchel, F., The Effect of Central Exit Examinations on Student Achievement: Quasi-experimental Evidence from TIMSS Germany, Journal of the European Economic Association, 3, 2005, S. 1134–1155.

behutsam vorgegangen werden, indem etwa nur die überdurchschnittlichen Lehrkräfte genannt werden. Im letzten Schritt kann dann leistungsgerechtere Besoldung eingeführt werden.

Die Erfahrungen mit den Reformen in anderen Ländern haben gezeigt, dass die Skepsis der unmittelbar Betroffenen zunächst groß ist und größte Sorgfalt auf die Entwicklung geeigneter Indikatoren verwendet werden muss. Vielleicht ist es auch angesichts der Messprobleme vergeblich, auf ein optimales System zu hoffen. Aber jeder Versuch, die Transparenz im Bildungssystem zu erhöhen, ist ein Schritt in die richtige Richtung. Wenn zusätzliche Leistungsanreize auch in Deutschland eingeführt werden sollen, so ist eine gute Aufklärungsarbeit im Vorfeld wichtig. Es gilt herauszustellen, dass leistungsabhängige Besoldung in erster Linie eine Chance für produktive Lehrerinnen und Lehrer ist, für ihre gute und wichtige Arbeit auch angemessen honoriert zu werden.

*Verf.: Dr. Hendrik Jürges, Mannheim Research Institute for the Economics of Aging (MEA), Fakultät für Rechtswissenschaft und Volkswirtschaftslehre, Universität Mannheim, 68131 Mannheim,*

*Prof. Dr. Kerstin Schneider, Fachbereich B, Bergische Universität Wuppertal, 42097 Wuppertal*