

Gundlach, Erich; Wößmann, Ludger

Book Part — Manuscript Version (Preprint)

Bildungsressourcen, Bildungsinstitutionen und Bildungsqualität: Makroökonomische Relevanz und mikroökonomische Evidenz

Provided in Cooperation with:

Kiel Institute for the World Economy – Leibniz Center for Research on Global Economic Challenges

Suggested Citation: Gundlach, Erich; Wößmann, Ludger (2004) : Bildungsressourcen, Bildungsinstitutionen und Bildungsqualität: Makroökonomische Relevanz und mikroökonomische Evidenz, In: Blaga, Steffen Backes-Gellner, Uschi (Ed.): Ökonomie der Evaluation von Schulen und Hochschulen, ISBN 3-428-11569-4, Duncker & Humblot, Berlin, pp. 15-52

This Version is available at:

<https://hdl.handle.net/10419/3311>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

**Bildungsressourcen, Bildungsinstitutionen und Bildungsqualität:
Makroökonomische Relevanz und mikroökonomische Evidenz**

Erich Gundlach^a und Ludger Wößmann^b

revidiert, Oktober 2003

^aInstitut für Weltwirtschaft, Postfach 4309, 24100 Kiel, Tel.: 0431 8814-284,
E-mail: egundlach@ifw.uni-kiel.de

^bIfo Institut für Wirtschaftsforschung, Poschingerstr. 5, 81679 München, Tel.: (089)
9224-1699, E-mail: woessmann@ifo.de

A. Einleitung und Zusammenfassung

Nach dem sogenannten PISA-Schock wird insbesondere in Deutschland intensiv diskutiert, wie die Qualität der schulischen Ausbildung verbessert werden kann. Dabei herrscht die Überzeugung vor, dass die Qualität der schulischen Ausbildung eine wichtige Determinante des langfristigen makroökonomischen Produktivitätswachstums ist, die zumindest teilweise mit einer Variation der staatlichen Bildungsausgaben gesteuert werden kann. Investitionen in Humankapital gelten folgerichtig ganz allgemein als wirtschaftspolitischer Schlüssel für den Erhalt der internationalen Wettbewerbsfähigkeit einer zunehmend wissensbasierten Volkswirtschaft. Darüber hinaus wird auch diskutiert, ob nicht eher institutionelle Reformen statt weiterer Ausgabensteigerungen im Schulsektor nötig sind, um die Qualität der schulischen Ausbildung zu verbessern.

Im Gegensatz zur öffentlichen Diskussion ist man in der makroökonomischen empirischen Forschung in verschiedenen neueren Studien häufig zu dem Ergebnis gekommen, dass andere Faktoren als das Humankapital für die großen internationalen Unterschiede in der Produktivität des Faktors Arbeit verantwortlich sein müssen. Danach wäre von wirtschaftspolitischen Maßnahmen im Bildungssektor, wie etwa einer verbesserten finanziellen Ausstattung oder gar einer grundlegenden institutionellen Reform, kein wesentlicher Beitrag zum langfristigen Produktivitätswachstum der Volkswirtschaft zu erwarten. Die aktuelle Debatte zu PISA und den Folgen wäre dann lediglich für eine effiziente Gestaltung des Bildungssektors von Bedeutung, aber nicht darüber hinaus.

Als Ausgangspunkt für unseren Beitrag versuchen wir deshalb zunächst, den aktuellen Stand der internationalen empirischen Forschung zur makroökonomischen Rolle des Humankapitals nachzuzeichnen. In der makroökonomischen empirischen Forschung wird Humankapital im Wesentlichen anhand des durchschnittlichen schulischen Ausbildungsstandes der Arbeitskräfte erfasst. Nicht berücksichtigt wird dabei in der Regel, dass sich die Qualität der schulischen Ausbildung von Land zu Land zum Teil erheblich unterscheidet, was nicht erst seit PISA bekannt ist. Von daher gesehen ist es

kein unbedingt überraschendes Ergebnis, dass ein rein quantitatives Konzept des Humankapitals nur wenig zu einer statistischen Aufklärung der großen internationalen Einkommensunterschiede beitragen kann.

Wenn wir demgegenüber die durchschnittliche Anzahl der Schuljahre mit einem Qualitätsindikator gewichten, der auf den Ergebnissen verschiedener standardisierter internationaler Vergleichstests der Schülerleistung in Mathematik und Naturwissenschaft beruht, dann lassen sich rein statistisch gesehen rund 50 Prozent der internationalen Produktivitätsunterschiede mit solch einem qualitätsbereinigten Humankapitalkonzept erklären. Für die Gruppe der OECD-Länder, für die die verlässlichsten Daten vorliegen, lässt sich zeigen, dass bereits allein mit der Variation im Humankapital rein statistisch gesehen die gesamte Variation der Arbeitsproduktivität erklärt werden kann.

Für sich genommen bestätigen diese Befunde, dass gerade die Berücksichtigung der Bildungsqualität, hier festgemacht am durchschnittlichen Leistungsstand der Schüler eines Landes, hilft, die Relevanz des Faktors Humankapital für Wachstum und Entwicklung im internationalen Vergleich nachzuweisen. Wenn Unterschiede in der Bildungsqualität berücksichtigt werden lässt sich für die Makroebene darüber hinaus empirisch belegen, dass ein höherer Bestand an Humankapital unter sonst gleichen Bedingungen einhergeht mit einem höheren durchschnittlichen Einkommen des ärmsten Teils der Bevölkerung. Schließlich kann man unter Berücksichtigung der Bildungsqualität für einen internationalen Querschnitt auch zeigen, dass die gesamtwirtschaftliche Ertragsrate des Humankapitals tendenziell über der aus vielen Mikrostudien bekannten privaten Mincerschen Ertragsrate in der Nähe von 10 Prozent liegen könnte.

Aus unserer Sicht führen diese makroökonomischen Befunde unmittelbar zur Frage nach den wirtschafts- beziehungsweise bildungspolitischen Maßnahmen, die gegebenenfalls geeignet wären, die Bildungsqualität in einem Land nachhaltig zu verbessern. Die für viele Politiker naheliegende Option besteht offenkundig darin, die Ressourcenausstattung innerhalb eines gegebenen Schulsystems zu verbessern. Eine

weitergehende Option, die sicher nicht ohne erheblichen Widerstand der betroffenen Interessengruppen durchzusetzen wäre, besteht darin, bei einer gegebenen Ressourcenausstattung das Schulsystem in Bezug auf seine institutionelle Struktur zu verändern.

Wir berichten zunächst über neuere mikroökonomische Ergebnisse zum Zusammenhang zwischen der Bildungsressource „Klassengröße“ und dem Leistungsstand der Schüler. Selbst wenn die potentielle Endogenität der Klassengröße bei den empirischen Schätzungen durch eine Kombination von Fixed-Effects- und Instrument-Variablen-Schätzung berücksichtigt wird, zeigen sich für die überwiegende Anzahl der betrachteten Industrie- und Schwellenländer keine statistisch signifikanten Ressourceneffekte. Nur in vier von insgesamt 36 ausgewählten Fällen, und zwar nur in Ländern mit relativ niedrigen Lehrergehältern, hat die Klassengröße einen negativen Einfluss auf die Schülerleistungen. Dieser Befund legt nahe, dass Klassengrößeneffekte offenbar nur dann auftreten, wenn die betreffende Lehrerschaft im Durchschnitt über ein eher niedriges Qualifikationsniveau verfügt.

Im Gegensatz dazu belegen weitere mikroökonomische Schätzungen, dass insbesondere die Bildungsinstitution „Zentralprüfung“ einen statistisch signifikanten und quantitativ relevanten Einfluss auf den Leistungsstand der Schüler ausübt. So weisen Mittelstufenschüler in Ländern mit zentralen Abschlussprüfungen gegenüber Mittelstufenschülern aus Ländern ohne solche Prüfungen einen Leistungsvorsprung auf, der etwa einem ganzen Schuljahr entspricht. Verschiedene statistische Tests legen nahe, dass der geschätzte Effekt tatsächlich von der Existenz beziehungsweise Nicht-Existenz von Zentralprüfungen hervorgerufen wird und nicht andere länderspezifische Charakteristika widerspiegelt.

Unsere empirischen Ergebnisse belegen auch, dass die Existenz von zentralen Prüfungen den Einfluss des elterlichen Bildungsstandes auf gemessene Schülerleistungen verringern und somit zu besseren Bildungschancen für Schüler mit nachteiligem familiärem Hintergrund beitragen. Darüber hinaus scheinen zentrale Prüfungen aufgrund ihrer Anreizwirkungen für Schüler, Lehrer, Schulen und

Schulbehörden die unabdingbare Voraussetzung für ein leistungsfähiges Funktionieren von ansonsten dezentral organisierten Schulsystemen zu sein. Insgesamt betrachtet verdeutlichen unsere empirischen Ergebnisse, dass nicht die Bildungsressource „Klassengröße“, sondern die Bildungsinstitution „Zentralprüfung“ die Bildungsqualität erhöht, die wiederum eine zentrale Rolle für das Wachstum und die Entwicklung einer Volkswirtschaft spielt.

B. Die Bildungsqualität als Determinante der gesamtwirtschaftlichen Entwicklung

I. Humankapital als Komponente der Arbeitsproduktivität

In der makroökonomischen Literatur firmiert derzeit unter dem Stichwort „Development Accounting“ eine Debatte, bei der über das Gewicht des Humankapitals für eine statistische Erklärung des gesamtwirtschaftlichen Entwicklungsniveaus eines Landes gestritten wird. Dabei wird der Bestand an Humankapital in der Regel anhand des durchschnittlichen Ausbildungsstands der Arbeitskräfte gemessen. Verschiedene empirische Untersuchungen lassen Zweifel aufkommen, ob der Ausbildungsstand der Arbeitskräfte wirklich eine entscheidende Rolle für die Erklärung internationaler Einkommens- und Wachstumsunterschiede spielt. Klenow und Rodriguez-Clare (1997), Prescott (1998) sowie Hall und Jones (1999) kommen aufgrund ihrer Berechnungen beispielsweise zu dem Ergebnis, dass internationale Unterschiede im Ausbildungsstand der Arbeitskräfte nur einen sehr kleinen Teil der internationalen Einkommensunterschiede erklären können. Nach ihren Befunden scheinen vielmehr nicht weiter spezifizierte internationale Technologieunterschiede – die als empirisch als Residualgröße bestimmt werden – die wesentliche Ursache für das internationale Wohlstandsgefälle zu sein. Von einer Reform oder gar einer Ausweitung des Bildungssystems wären demnach für sich genommen keine nennenswerten Impulse für das langfristige Wirtschaftswachstum zu erwarten.

Eine solche Schlussfolgerung widerspricht nahezu allen wirtschaftspolitischen Programmen, die derzeit im nationalen Rahmen sowie von multilateralen

Organisationen wie Weltbank und Währungsfonds vertreten werden. Deshalb wird im Folgenden untersucht, woran es liegen kann, dass dem Faktor Humankapital in den genannten Studien keine bedeutende produktive Rolle zugeschrieben wird. Neben der benutzten Methodik kommt natürlich der jeweiligen Messung des Bildungsstands der Arbeitskräfte eine unmittelbare Bedeutung für die Ergebnisse zu, was sich insbesondere dann zeigt, wenn versucht wird, neben internationalen Unterschieden in der Bildungsquantität auch internationale Unterschiede in der Bildungsqualität zu erfassen wird.

Das methodische Verfahren beim „Development Accounting“ ist denkbar einfach: Das Bruttoinlandsprodukt je Person im erwerbsfähigen Alter, im Folgenden auch als Arbeitsproduktivität eines Landes bezeichnet, wird mit Hilfe einer Produktionsfunktion beschreiben als das Ergebnis von Technologie und Faktorinput, wobei sich der Faktorinput wiederum aus Sachkapital sowie aus Humankapital in Form von Bildung zusammensetzt. Für den einfachsten Fall einer Cobb-Douglas-Produktionsfunktion mit Harrod-neutraler Technologie lässt sich ein solcher Zusammenhang darstellen als logarithmisch-lineare Gleichung:

$$(1) \quad \ln(Y/L) = \alpha \ln(K/L) + \beta \ln(AH/L) \quad ,$$

wobei Y für das reale Bruttoinlandsprodukt steht, L für die Anzahl der Personen im erwerbsfähigen Alter, K für den Bestand an Sachkapital und H für den Bestand an Humankapital. Die beiden Koeffizienten α und β entsprechen unter den Annahmen konstanter Skalenerträge und vollständigen Wettbewerbs den jeweiligen Anteilen der beiden Produktionsfaktoren am Sozialprodukt, also der Lohnquote und Gewinnquote. Die Lohnquote beträgt in den meisten Ländern rund 70 Prozent (Gollin 2002), so dass für die Gewinnquote ein Wert von rund 30 Prozent angenommen werden kann. Sind die Bestände von Sach- und Humankapital bekannt, so lässt sich zusammen mit den Verteilungsquoten ihr jeweiliges Gewicht für eine statistische Erklärung der beobachteten internationalen Variation der Arbeitsproduktivität bestimmen. Bei diesem Verfahren wird der verbleibende Anteil der Variation der Arbeitsproduktivität als Restgröße dem Faktor Technologie A zugeschlagen.

Der jeweils gemessene Produktivitätsbeitrag der Bildung hängt somit im Wesentlichen davon ab, ob und gegebenenfalls wie die theoretische Variable Humankapital in eine empirisch zugängliche Messziffer übersetzt werden kann. Hall und Jones (1999) schlagen beispielsweise vor, den volkswirtschaftlichen Bestand an Humankapital mit Hilfe der sogenannten Mincer-Gleichung (Mincer 1974) zu bestimmen. Danach wäre Humankapital das Produkt aus dem Faktor Arbeit und einem exponentiellen Faktor, der dem Produkt aus der Anzahl der durchschnittlichen Schuljahre (S) und der Ertragsrate auf die schulische Ausbildung (r) entspricht:

$$(2) \quad H = L \cdot e^{r \cdot S} .$$

Diese Spezifikation liefert eine theoretisch überzeugende Möglichkeit, die Variable Humankapital empirisch zu erfassen. Schulische Ertragsraten sind in den letzten 20 Jahren für nahezu alle Länder geschätzt und zusammenfassend ausgewertet worden, zuletzt von Psacharopoulos und Patrinos (2002). International vergleichbare Angaben über die durchschnittliche Anzahl der Schuljahre sind ebenfalls vorgelegt worden, zuletzt von Barro und Lee (2001). Die genannten empirischen Studien von Klenow und Rodriguez-Clare (1997) sowie Hall und Jones (1999) kamen mit verschiedenen Varianten dieses Ansatzes jedoch ganz überwiegend zu dem Ergebnis, dass Humankapital in Form von Bildung nur eine unwesentliche Rolle für eine Erklärung der internationalen Unterschiede im Entwicklungsniveau spielt.

Allerdings vernachlässigen diese Studien bei der Berechnung des Humankapitals die internationalen Unterschiede in der Qualität der schulischen Ausbildung, wie sie etwa von Hanushek und Kimko (2000) ermittelt worden sind. Auch die „Third International Mathematics and Science Study (TIMSS)“ (IEA 1998) belegt, dass es selbst innerhalb der relativ homogenen Gruppe der führenden OECD-Länder signifikante Unterschiede der durchschnittlichen Schülerleistungen gibt, was im Rahmen der ersten PISA-Studie (OECD 2001) eindrucksvoll bestätigt wurde. Von daher erscheint die implizite Annahme in Gleichung (2), dass ein Schuljahr in Japan genauso zu gewichten sei wie ein Schuljahr in einem beliebigen Entwicklungsland, als wenig plausibel. Um die

offenkundigen internationalen Unterschiede in der Qualität der Ausbildung zu berücksichtigen, könnte man beispielsweise die durchschnittliche Anzahl der Schuljahre mit einem normierten Qualitätsindex Q $[0,1]$ multiplizieren, so dass

$$(3) \quad H = L \cdot e^{r \cdot (S \cdot Q)}$$

gilt, wobei dann für das Land mit $Q = 1$ die tatsächliche Anzahl der Schuljahre in die Berechnung einfließen würde, während für Länder mit im Durchschnitt schlechteren oder besseren Schülern die Anzahl der Schuljahre entsprechend dem jeweiligen Wert des Qualitätsindex gewichtet werden würde.¹

Ausgangspunkt für die Berechnung des qualitätsgewichteten Humankapitalindex nach Gleichung (3) sind die von Hanushek und Kimko (2001) berechneten durchschnittlichen Testergebnisse der Schüler eines Landes in internationalen Vergleichstests in Mathematik und Naturwissenschaft, wie sie von der „International Association for the Evaluation of Educational Attainment (IEA)“ und im Rahmen des „International Assessment of Educational Progress (IEAP)“ für viele Länder und für verschiedene Jahre durchgeführt wurden. Hanushek und Kimko (2001) berechnen anhand aller bis zum Jahr 1991 vorliegenden Test-Statistiken für jedes Land ihrer Stichprobe eine durchschnittliche Kennziffer für die Qualität der schulischen Ausbildung Q . Bei unseren Berechnungen normieren wir Q auf den Wert für die Vereinigten Staaten. Als Werte für die Ertragsrate r benutzen wir die nach der sogenannten "elaborate method" geschätzten "sozialen" Ertragsraten, die die gesamtwirtschaftlichen Kosten der schulischen Ausbildung (aber keine möglichen externen Erträge) berücksichtigen. Psacharopoulos (1994) bietet eine Zusammenstellung der in der Literatur nach der "elaborate method" geschätzten länderspezifischen Ertragsraten, die nach dem jeweiligen Ausmaß der schulischen Ausbildung (primär, sekundär, tertiär) differenziert werden können. Unsere Stichprobe von Ländern enthält darüber hinaus auch geschätzte

¹ Zu den Einzelheiten der Berechnung des Qualitätsindex vgl. Gundlach et al. (2002) sowie Wößmann (2003a).

Werte für Q , die auf Basis der jeweiligen Durchschnittswerte für Ländergruppen nach Regionen sowie nach Einkommensklassen ermittelt wurden.

Die von uns gewählte Spezifikation des Zusammenhangs zwischen Bildungsquantität (S) und Bildungsqualität (Q) besitzt den Vorteil, dass der Bestand des Humankapitals nunmehr ohne Obergrenze in Abhängigkeit von Qualitätsverbesserungen wachsen kann, was bei einer Beschränkung auf die zumindest asymptotisch begrenzte Dimension Quantität nicht der Fall ist. In diesem Sinne wird unser Humankapitalmaß eher vergleichbar mit Sachkapitalmaßen, die ebenfalls keiner Obergrenze unterliegen. Ein weiterer Vorteil unseres Humankapitalmaßes ist die direkte Anbindung an den Faktor Arbeit, die plausibler erscheint als eine Spezifikation mit Arbeit, Bildungsqualität sowie Bildungsquantität als jeweils unabhängigen Faktoren.

Problematisch erscheint demgegenüber die gewählte lineare Verknüpfung der Quantität mit der Qualität, die keiner speziellen theoretischen Begründung folgt. Die lineare Verknüpfung impliziert, dass die Quantität denselben Einkommenseffekt bewirkt wie die Qualität. Über die Plausibilität dieser Annahme lässt sich in Ermangelung klarer theoretischer Vorgaben sicher streiten. Entscheidend ist im vorliegenden Zusammenhang zunächst, ob mit unserer Spezifikationsannahme die Relevanz der Faktors Humankapital tendenziell überzeichnet wird. Nach den empirischen Ergebnissen von Hanushek und Kimko (2000) sowie Barro (1999) scheint das jedoch nicht der Fall zu sein. Beide Studien finden, dass die Bildungsqualität in Wachstumsregressionen ein stärkeres Gewicht besitzt als die Bildungsquantität.

Tabelle 1 zeigt, wie der statistische Erklärungsgehalt der Variable Humankapital für die internationalen Unterschiede in der Arbeitsproduktivität in Abhängigkeit von dem jeweils benutzten Humankapitalkonzept variiert. Die erste Zeile reproduziert das zentrale Ergebnis der Studie von Klenow und Rodriguez-Clare (1997), wonach internationale Humankapitalunterschiede gemäß Gleichung (2) lediglich 11 Prozent der internationalen Unterschiede in der Arbeitsproduktivität erklären können. Internationale Sachkapitalunterschiede erklären nach ihrem Befund 23 Prozent, so dass rund zwei Drittel der beobachteten Varianz unerklärt bleiben beziehungsweise als internationale

Technologieunterschiede betrachtet werden. Zeile 2 zeigt, dass sich der Erklärungsanteil des Faktors Humankapital verdoppelt, wenn ein aktualisierter Datensatz benutzt wird und der Faktor Humankapital so wie bei Hall und Jones (1999) als Intensität (H/L) statt als Koeffizient (H/Y) spezifiziert wird. Insgesamt gesehen dominieren internationale Technologieunterschiede aber auch hier die internationale Varianz der Arbeitsproduktivität.

In den Zeilen 3 und 4 wird dieselbe Methodik und dieselbe Stichprobe von Ländern wie in Zeile 2 benutzt, beim empirischen Messen des Humankapitals werden jedoch verschiedene konzeptionelle Verbesserungen berücksichtigt (vgl. dazu im Einzelnen Gundlach et al. 2002). Zeile 3 verdeutlicht, dass man mit länderspezifischen sozialen Ertragsraten der Bildung keine wesentlich anderen Ergebnisse erhält als mit internationalen Durchschnittswerten wie in Zeile 2, d.h., länderspezifische Schätzfehler gleichen sich vermutlich mehr oder weniger aus. Falls allerdings länderspezifische Unterschiede in der Qualität der schulischen Ausbildung berücksichtigt werden wie in Zeile 4, steigt der statistische Erklärungsgehalt der Variable Humankapital auf 46 Prozent an, so dass der Erklärungsanteil des Faktors Technologie auf 35 Prozent fällt. Dabei gilt es allerdings zu berücksichtigen, dass dem Faktor Technologie als restgröße zugerechnete Beitrag zum Teil auch durch Messfehler bei den Faktorinputs bedingt sein könnte.

Deshalb bietet es sich an, die „Development Accounting“-Methode auf eine Gruppe von Ländern zu beschränken, bei denen a priori keine substantiellen Technologieunterschiede zu vermuten sind. Für die Gruppe der OECD-Länder könnte eine solche Annahme unter Umständen zulässig sein, da alle OECD-Länder relativ offen gegenüber der internationalen Mobilität von Gütern, Kapital und Technologie sind. Gleichzeitig besitzen die OECD-Länder gegenüber der Gruppe der Entwicklungsländer weitaus verlässlichere Statistiken, so dass auch von daher die denkbaren Messfehler beim Faktorinput für diesen Fall minimiert werden.

Tabelle 2 wiederholt die in Tabelle 1 angestellten Berechnungen für eine Stichprobe von 23 OECD-Ländern. Die Ergebnisse bestätigen die Vermutung, dass die

Residualgröße Technologie innerhalb der Gruppe der OECD-Länder einen wesentlich kleineren Teil der Unterschiede in der Arbeitsproduktivität erklärt als in der gesamten Stichprobe von 113 Ländern in Tabelle 1. Wenn Unterschiede in der Qualität der schulischen Ausbildung mit ins Bild genommen werden, lässt die Varianz der Arbeitsproduktivität zwischen den OECD-Ländern vollständig mit Hilfe der Varianz der Humankapitalbestände aufklären (Zeile 3). Aus rein statistischer Sicht erklären die Varianz von Sach- und Humankapital für diese Länder zusammengenommen über 100 Prozent der Varianz in der Arbeitsproduktivität, so dass sich ein negativer Restbeitrag für die Residualgröße ergibt.

Natürlich lässt sich mit Hilfe der Methode des „Development Accounting“ keine letzte Gewissheit über die makroökonomische Rolle des Humankapitals gewinnen. Zumindest zweifelhaft erscheint jedoch nach unseren Ergebnissen, ob Humankapital wirklich eine so geringe Rolle spielt, wie das einige neuere Studien nahe legen. Unsere Ergebnisse deuten vielmehr darauf hin, dass man neben der Bildungsquantität auch und gerade die Bildungsqualität berücksichtigen muss, wenn der Beitrag des Humankapitals zur statistischen Erklärung der internationalen Unterschiede der Arbeitsproduktivität bestimmt werden soll. Dabei sind neben den direkten auch die indirekten Produktivitätseffekte des Humankapitals von Interesse. Solche möglichen indirekten Produktivitätseffekte des Humankapitals werden in der Literatur in jüngster Zeit beispielsweise im Zusammenhang mit dem Einfluss des durchschnittlichen Humankapitalbestands auf die Einkommen der Armen sowie mit der Frage nach der gesamtwirtschaftlichen Ertragsrate des Humankapitals diskutiert.

II. Humankapital und das Einkommen der Armen

In einer vielbeachteten empirischen Studie haben Dollar und Kraay (2002) einen engen statistischen Zusammenhang zwischen dem durchschnittlichen Wirtschaftswachstum und dem Einkommenswachstum des ärmsten Teils der Bevölkerung ermittelt. Nach ihren Ergebnissen geht im internationalen Vergleich ein um 10 Prozent höheres Durchschnittseinkommen der Gesamtbevölkerung mit einem um 10 Prozent höheren Durchschnittseinkommen der Armen einher. Gleichzeitig stellen sie fest, dass

vermeintlich armutsbekämpfende Maßnahmen, wie etwa höhere öffentliche Ausgaben für Gesundheit und Bildung, offenbar keinen eigenen Einfluss auf die Einkommen der Armen haben, der über ihre allgemeinen Wachstumswirkungen hinausgehen würde. Danach sollten Strategien zur Armutsbekämpfung vornehmlich auf ein höheres Wirtschaftswachstum setzen, nicht aber beispielsweise auf Bildungsinvestitionen.

Wie im vorhergehenden Abschnitt lässt sich auch dieser Befund überprüfen, indem bei der Humankapital-Variable wie oben beschrieben die internationalen Unterschiede in der Qualität der schulischen Ausbildung berücksichtigt werden. International vergleichbare Daten zum Einkommen der Armen erhält man anhand eines internationalen Datensatzes zur Einkommensverteilung, der erstmals von Deininger und Squire (1996) aufbereitet wurde. Dieser Datensatz enthält Gini-Koeffizienten und kumulative Einkommens-Quintile für einen Zeitraum von 40 Jahren für 111 Länder. Analog zu Dollar und Kraay (2002) wird im Folgenden das durchschnittliche Bruttoinlandsprodukt je Einwohner (Pro-Kopf-Einkommen) der Armen entsprechend dem ersten Einkommens-Quintil definiert, also als Anteil des Einkommens der ärmsten 20 Prozent der Bevölkerung am gesamtwirtschaftlichen Einkommen.

Mit diesen Ausgangsdaten für das durchschnittliche Einkommen der Armen in 111 Ländern in 1990 lässt sich ein fundamentales Ergebnis von Dollar und Kraay (2002) reproduzieren, die eine Stichprobe von 269 gepoolten Querschnitts- und Zeitreihenbeobachtungen benutzen. Die OLS-Schätzgleichung mit Daten für das Jahr 1990

$$(4) \quad \ln y_p = a_0 + 1,06 \ln y_c \quad \bar{R}^2 = 0,86, \quad n = 111$$

bestätigt, dass das Durchschnittseinkommen (y_c) und das Einkommen der Armen (y_p) statistisch hoch korreliert sind. Danach geht eine Zunahme des Durchschnittseinkommens mit einer Zunahme des Einkommens der Armen in gleicher Höhe einher (Standardfehler des Regressionskoeffizienten: 0,04). Dieser Befund deutet für sich genommen darauf hin, dass der Anteil des Einkommens der Armen am Durchschnittseinkommen im internationalen Vergleich nicht systematisch mit der Höhe

des Durchschnittseinkommens variiert. Die hier zu diskutierende Frage ist, ob die nach Berücksichtigung des Durchschnittseinkommens verbleibende internationale Varianz im Einkommen Armen mit Hilfe zusätzlicher Variablen, wie etwa dem Humankapital, aufgeklärt werden kann. Dabei wird hier im Folgenden wiederum ein umfassendes Konzept des Humankapitals unter Einschluss der Bildungsqualität benutzt, während Dollar und Kraay den Bestand an Humankapital lediglich die quantitative Dimension des Humankapitals berücksichtigen und sich zudem auf die primäre Ausbildung beschränken.

Um den potentiellen (zusätzlichen) Einfluss des durchschnittlichen Humankapitalbestands und anderer Variablen auf das Einkommen der Armen zu bestimmen, wird Schätzgleichung (4) erweitert auf

$$(5) \quad \ln y_p = a_0 + a_1 \ln y_c + a_2 \ln(H/L) + a_k X_k \quad \text{mit } k = 3, \dots, m,$$

wobei X für sonstige Einflussfaktoren auf das Einkommen der Armen steht.² Wenn man zunächst nur Humankapital als zusätzliche Variable berücksichtigt, erhält man für den Koeffizienten a_1 in etwa das gleiche Ergebnis wie in Gleichung (4), also eine Bestätigung des Befunds eines direkt proportionalen Zusammenhangs zwischen dem Durchschnittseinkommen und dem Einkommen der Armen (Tabelle 3, Spalte 1). Im Gegensatz zum Befund in Dollar und Kraay (2002) erweist sich die hier benutzte Humankapital-Variable als eine statistisch signifikante Einflussgröße für das Einkommen der Armen. Die Punktschätzung impliziert, dass ein Anstieg dieser Variable um 10 Prozent, unabhängig vom Effekt eines höheren Durchschnittseinkommens, das durchschnittliche Einkommen der Armen zusätzlich um 3,2 Prozent erhöhen würde. Wenn man dieses Regressionsergebnis zum Nennwert nimmt, sollten Bildungsinvestitionen ein wesentlicher Bestandteil einer jeden Strategie zur Armutsbekämpfung sein.

² Zu den Einzelheiten der im Folgenden diskutierten Schätzungen vgl. Gundlach et al. (2004).

Um die Robustheit dieses Ergebnisses zu überprüfen, können zusätzliche Variablen in die Schätzgleichung aufgenommen werden. In den meisten empirischen Wachstumsanalysen hat sich beispielsweise ein Maß für die Sachkapitalakkumulation, in der Regel die Investitionsquote, als eine robuste Variable erwiesen (Levine und Renelt 1992). Die Investitionsquote (*INV*) wird hier gemessen als durchschnittlicher Anteil der realen Investitionen am realen Bruttoinlandsprodukt im Zeitraum 1960-1990 (siehe PWT 1994). Diese Variable weist im vorliegenden Zusammenhang allerdings einen quantitativ unbedeutenden und statistisch insignifikanten negativen Regressionskoeffizienten auf (Spalte 2). Eine Erklärung dafür ist, dass das Durchschnittseinkommen möglicherweise bereits den Beitrag der Sachkapitalakkumulation für eine Erklärung der internationalen Unterschiede im Einkommen der Armen enthält. Für den Beitrag des Humankapitals gilt dies aber offensichtlich nicht: Der geschätzte Regressionskoeffizient bleibt statistisch signifikant und quantitativ nahezu unverändert.

Die nächsten beiden Spezifikationen in Tabelle 3 enthalten Variablen, die in einem mehr oder weniger direkten Zusammenhang mit dem Einkommen der Armen stehen. Ein hoher Wertschöpfungsanteil von Bergbau und ähnlichen extraktiven Aktivitäten am Sozialprodukt (*MINING*) könnte zu einer ungleichen Einkommensverteilung und wegen des damit häufig einhergehenden „rent seeking“ zu einem langsameren Wirtschaftswachstum führen (Rodriguez und Sachs 1999), was letztlich auch zu einem niedrigeren absoluten Einkommen der Armen führen müsste als es ohne den wachstumsbremsenden Einfluss des rent-seeking der Fall sein könnte. Neben der ineffizienten Ausbeutung von Rohstoffen könnte auch das Vorkommen von Malaria (*MALARIA*) wegen ihres Einflusses auf den Gesundheitsstand und die Ausfallzeiten der Arbeitskräfte die wirtschaftliche Entwicklung eines Landes nachhaltig hemmen. Bloom und Sachs (1998) halten beispielsweise das Vorkommen von Malaria für den wesentlichen Grund für die wirtschaftliche Malaise Afrikas. Die Ergebnisse in den Spalten (3) und (4) der Tabelle 2 zeigen jedoch, dass diesen beiden Variablen kein statistisch signifikanter eigenständiger Erklärungsgehalt für die internationalen Unterschiede im Einkommen der Armen zukommt.

Dieses Ergebnis bleibt auch bestehen, wenn alle zusätzlichen Variablen zusammen in die Schätzgleichung aufgenommen werden (Spalte (5)). Die Humankapital-Variable hat nach wie vor einen statistisch signifikanten positiven Effekt auf das Einkommen der Armen, zusätzlich zu dem 1:1-Effekt des Durchschnittseinkommens. Mit Hilfe von Beta-Koeffizienten kann man den Einfluss der beiden statistisch signifikanten Variablen direkt miteinander vergleichen. Bei einer Standardabweichung von 1,21 für die abhängige Variable ($\sigma(y_p)$) ergeben sich Beta-Koeffizienten von 0,18 für die Humankapital-Variable und von 0,51 für das Durchschnittseinkommen. Somit könnte eine Zunahme der Humankapital-Variable um eine Standardabweichung in etwa ein Drittel des Effekts einer Zunahme des Durchschnittseinkommens um eine Standardabweichung ausmachen.

Ein Problem für die Interpretation der bisherigen Ergebnisse könnte dadurch entstehen, dass die Humankapital-Variable über den politischen Prozess umgekehrt wie in Gleichung (5) angenommen auch von der Höhe des Einkommens der Armen abhängt. So könnten beispielsweise in denjenigen Ländern, in denen das Einkommen der Armen relativ hoch ist, mehr Ressourcen für Bildungsinvestitionen zur Verfügung stehen als in Ländern, in denen das Einkommen der Armen relativ niedrig ist. Wenn es sich bei der Humankapital-Variable demnach um eine endogene Größe handelt, was eher wahrscheinlich ist, dann könnte die Signifikanz und die Höhe des Regressionskoeffizienten bei einer OLS-Schätzung wie in Tabelle 3 überzeichnet werden. Eine umgekehrte Kausalität könnte natürlich auch den Regressionskoeffizienten für das Durchschnittseinkommen beeinflussen. Dollar und Kraay (2002) finden jedoch auf der Basis von Paneldaten, dass die mögliche Endogenität des Durchschnittseinkommens ihren geschätzten Regressionskoeffizienten in der Nähe von 1 nicht nach oben verzerrt hat.

Da der hier geschätzte Regressionskoeffizient ebenfalls in der Nähe von 1 liegt, lässt sich der Befund von Dollar und Kraay hinsichtlich der Endogenität des Durchschnittseinkommens als Restriktion in die Schätzgleichung (5) einbauen, wobei die bislang zusätzlich berücksichtigten Variablen aufgrund ihrer fehlenden statistischen Signifikanz nicht mehr enthalten sind:

$$(6) \quad \ln y_p - \ln y_c = a_0 + a_2 \ln(H / L) .$$

Gleichung (6) wird geschätzt mit zwei Instrument-Variablen (IV), dem absoluten Abstand eines Landes vom Äquator (*DISTANCE*) (siehe Hall und Jones 1999) und der Durchschnittstemperatur eines Landes (*MEANTEMP*) (siehe McArthur und Sachs 2001). Beide Variablen werden sicherlich nicht vom Einkommen der Armen (oder vom Durchschnittseinkommen) beeinflusst, und beide Variablen können als plausible Instrumente für den Humankapitalbestand betrachtet werden, weil sie eng mit den institutionellen Rahmenbedingungen eines Landes zusammenhängen, so wie es die Studien von Hall und Jones (1999) und Acemoglu et al. (2001) nahe legen. Demnach sollten beide Variablen mit der Humankapital-Variable, nicht aber mit dem Störterm in Gleichung (6) korreliert sein.³

Tabelle 4 enthält die Ergebnisse der IV-Schätzung. In allen gewählten Spezifikationen ist der geschätzte Einfluss der Humankapital-Variable auf das Einkommen der Armen statistisch signifikant und quantitativ relevant. Wenn beide Instrumente zusammen benutzt werden, liefert ein Chi-Quadrat-Test auf Überidentifikation keine Anhaltspunkte dafür, dass eines der Instrumente mit dem Störterm korreliert wäre und somit direkt in die Schätzgleichung aufgenommen werden müsste.⁴ Im Durchschnitt

³ Die beiden Variablen können nicht als geeignete Instrumente betrachtet werden, wenn sie selbst das Einkommen der Armen direkt beeinflussen und somit in die Schätzgleichung gehören. Eine solche Möglichkeit lässt sich prinzipiell nicht ausschließen, kann aber nach dem gegenwärtigen Stand der Literatur als nicht plausibel betrachtet werden. Hall und Jones (1999) sowie Acemoglu et al. (2001), letztere insbesondere im Hinblick auf McArthur und Sachs (2001), finden bei ihren Schätzungen, dass sich geographischer Variablen bei Aufnahme in die Schätzgleichung als statisch insignifikant erweisen. Hinzu kommt, dass aus theoretischer Sicht nicht klar ist, wie ein direkter Zusammenhang zwischen dem Einkommen der Armen und einer Geographie-Variable zustande kommen könnte. Ein indirekter Zusammenhang über die Faktorakkumulation erscheint demgegenüber leichter vorstellbar.

⁴ Die Aussagefähigkeit des empirischen Tests auf Überidentifikation sollte nicht überbewertet werden, da nicht überprüft wird, ob alle jeweils benutzten Variablen geeignete Instrumente sind. Die Testergebnisse basieren jeweils auf der Annahme, dass mindestens eine der benutzten Variablen tatsächlich ein geeignetes Instrument darstellt.

besagen die drei Punktschätzungen, dass eine Zunahme der Humankapital-Variable um 10 Prozent zu einer Zunahme des Einkommens der Armen um 3,7 Prozent führt, und zwar zusätzlich zu einer Zunahme durch ein höheres Durchschnittseinkommen. Der mit der IV-Methode geschätzte Humankapital-Effekt ist somit größer als der mit der OLS-Methode geschätzte Humankapital-Effekt. Eine potentielle Überschätzung des Koeffizienten aufgrund der Endogenität des Humankapitals scheint also überkompensiert zu werden von einer potentiellen Unterschätzung aufgrund von Messfehlern bei der hier benutzten Humankapital-Variable.

Insgesamt gesehen unterstützen diese Ergebnisse für einen internationalen Datensatz die Hypothese für sich genommen, dass eine Zunahme des Humankapitalbestands einer Volkswirtschaft einen substantiellen positiven Effekt auf das durchschnittliche Einkommen der Armen hat. Wie ein Vergleich mit den Ergebnissen von Dollar und Kraay zeigt, gilt diese Einschätzung allerdings nur dann, wenn neben der Quantität der Bildung auch die Qualität der Bildung als Komponente des Humankapitals berücksichtigt wird. Offen bleiben muss an dieser Stelle die Frage, wie robust der vorliegende Befund im Hinblick auf denkbare alternative Spezifikationen der Humankapitalvariable und im Hinblick auf alternative Instrumentvariablen ist.

III. Die makroökonomische Ertragsrate des Humankapitals

Wenn sich neben den Wirkungen des Humankapitals auf das Einkommen der Armen auch positive externe Effekte der Humankapitalakkumulation nachweisen lassen, wie sie in vielen Modellen der endogenen Wachstumstheorie postuliert werden, dann hätte man ein weiteres Argument für die makroökonomische Relevanz bildungspolitischer Maßnahmen gewonnen. Ein Vergleich der privaten mit der makroökonomischen Ertragsrate einer Bildungsinvestition sollte Aufschluss darüber geben können, ob die in allen Ländern der Welt zu beobachtende Subventionierung der schulischen Bildung zu hoch, zu niedrig oder gerade angemessen ist.

Der Test auf Überidentifikation bietet insofern keinen umfassenden Schutz vor

Bislang gibt es allerdings erstaunlich wenig empirische Evidenz zum Niveau der gesamtwirtschaftlichen Ertragsrate des Humankapitals. Psacharopoulos und Patrinos (2002) fassen zusammen, was an empirischen Ergebnissen für einzelne Länder auf Basis einer mehr oder weniger einheitlichen methodischen Grundlage (Mincersche Ertragsraten-Funktion und sogenannte „elaborate method“) ermittelt worden ist. Dabei ist zu beachten, dass die in der mikroökonomischen Literatur ermittelten „sozialen“ Ertragsraten definitionsgemäß immer niedriger sein müssen als die privaten Ertragsraten, weil sie die Kosten der öffentlichen Subvention der schulischen Bildung mit ins Bild nehmen, nicht aber deren denkbare positive externe Effekte. Um die externen Effekte des Humankapitals bestimmen zu können, die sich in der makroökonomischen Ertragsrate spiegeln müssten, übertragen Heckman und Klenow (1997) die Mincersche Ertragsraten-Gleichung auf die Makroebene. Durch einen Vergleich der aus der Literatur bekannten privaten Ertragsraten mit der zu ermittelnden makroökonomischen Ertragsrate lässt sich überprüfen, ob von staatlicher Seite mehr oder weniger in die Bildung investiert werden sollte.

Die Mincersche Ertragsraten-Gleichung postuliert eine Beziehung zwischen dem logarithmierten Einkommen einer Person und der Anzahl ihrer Schuljahre (und anderer Faktoren) (Mincer 1974). Unter der vereinfachenden Annahme, dass die schulische Ausbildung keine direkten Kosten hat, also beispielsweise vom Staat gebührenfrei angeboten wird, lässt sich der Unterschied im individuellen Einkommen, der mit und ohne Schulbildung erzielt werden kann, als Ertragsrate (r) der schulischen Ausbildung darstellen. Für eine über alle Schuljahre hinweg konstante Ertragsrate folgt der gesuchte halblogarithmische Zusammenhang zwischen dem Einkommen (y) und der Anzahl der Schuljahre (S) als

$$(7) \quad \ln y_S = y_0 + rS \quad .$$

Wenn diese Beziehung als Schätzgleichung auf die gesamtwirtschaftliche Ebene übertragen wird, wo y für das Bruttoinlandsprodukt pro Person und S für die Anzahl der

Fehlspezifikationen, er kann aber durchaus Fehlspezifikationen offenlegen.

Schuljahre pro Person stehen, dann reflektiert der Regressionskoeffizient r nicht nur die individuellen Erträge der schulischen Ausbildung wie bei mikroökonomischen Analysen, sondern eben auch möglicherweise anfallende Externalitäten, weil alle Personen und alle Einkommen einer Volkswirtschaft zusammen mit dem durchschnittlichen Ausbildungsstand erfasst werden.

Gleichung (7) muss natürlich auf der rechten Seite um weitere Faktoren ergänzt werden, die ebenfalls einen Einfluss auf die abhängige Variable haben. Dazu zählen in der klassischen Spezifikation von Mincer (1974) beispielsweise die im Berufsleben erworbene Erfahrung der Arbeitskräfte („experience“) und andere sozio-ökonomische Faktoren wie das Geschlecht, der Wohnort oder der Beruf. Die meisten sozio-ökonomischen Variablen können bei einer internationalen Querschnittsbetrachtung wegen der Durchschnittsbildung als Erklärungsfaktoren vernachlässigt werden. Nicht vernachlässigt werden darf bei einer solchen Betrachtung aber, dass es im internationalen Vergleich große Unterschiede in der Kapitalintensität und darüber hinaus möglicherweise auch in der Technologie gibt.

Betrachtet man Gleichung (7) vor dem Hintergrund einer Produktionsfunktion mit Technologie sowie Sach- und Humankapital als Inputs, dann wäre beispielsweise

$$(8) \quad Y = K^\alpha \left(A L e^{rS} e^{qE} \right)^{1-\alpha}$$

eine denkbare Ausgangsgleichung, wobei Y für das Bruttoinlandsprodukt, K für den Sachkapitalbestand, A für das technologische Niveau, L für die Anzahl der Einwohner und q für die Ertragsrate auf die Erfahrung E stehen. Als Schätzgleichung in der Pro-Kopf-Version (y_c) folgt dann

$$(9) \quad \ln y_c = \alpha \ln k + (1-\alpha)A + (1-\alpha)rS + (1-\alpha)qE \quad ,$$

wobei k für die Kapitalintensität (K/L) steht. Diese Darstellung verdeutlicht, dass der zu schätzende Regressionskoeffizient für die Anzahl der Schuljahre je Einwohner (S) nicht

mehr als direktes Maß für die Ertragsrate der Ausbildung interpretiert werden darf. Die soziale Ertragsrate der schulischen Ausbildung ergibt sich nach Gleichung (9) vielmehr als Quotient aus Regressionskoeffizient und Produktionselastizität (beziehungsweise Verteilungsquote). Diese Restriktion wird von Heckman und Klenow (1997) bei ihrer Schätzung der sozialen Ertragsrate der schulischen Ausbildung übersehen.

Die makroökonomische Ertragsrate der schulischen Ausbildung wird auf Basis von Gleichung (9) für einen Querschnitt von Industrie- und Entwicklungsländern im Jahr 1990 geschätzt.⁵ Die benutzten Daten stammen aus verschiedenen Quellen. Das reale Sozialprodukt pro Person (y_c) und die Kapitalintensität (k) stammen aus PWT (1994). Das technologische Niveau eines Landes wird von Heckman und Klenow (1997) anhand der Lebenserwartung bei der Geburt (*LIFE*) approximiert, die aus World Bank (1999) stammt. Als alternative Variable für das technologische Niveau wird der absolute Abstand eines Landes vom Äquator (*DISTANCE*) benutzt. Mit dieser exogenen Variable können internationale Unterschiede in der institutionellen Infrastruktur und bei den Krankheitsrisiken für die Bevölkerung approximiert werden (Hall und Jones 1999; McArthur und Sachs 2001). Die durchschnittliche Anzahl der Schuljahre (S) der Bevölkerung im Alter von 15-65 Jahren stammt aus Barro und Lee (1996). Alternativ wird auch die mit dem Qualitätsindex von Hanushek und Kimko (2000) gewichtete Anzahl der Schuljahre (SQ) verwendet (siehe Abschnitt B. II). Die Variable Erfahrung (E) wird analog zu Heckman und Klenow (1997) als durchschnittliche Anzahl der potentiellen Jahre mit Arbeitserfahrung gemessen und basiert auf Daten zur Altersverteilung der Bevölkerung in UN (1994).

Die Ergebnisse in Tabelle 5 reproduzieren schrittweise die mit einem alternativen Datensatz durchgeführten Schätzungen von Heckman und Klenow (1997). Solange die Kapitalintensität nicht als erklärende Variable in die Schätzgleichung aufgenommen

⁵ Grundsätzlich nicht berücksichtigt werden Länder, die weniger als eine Million Einwohner haben, die vormals zur Sowjetunion oder zum Warschauer Pakt gehört haben sowie Länder, deren Wertschöpfung zu einem großen Teil aus der Förderung von Erdöl oder aus Übertragungen aus dem Ausland besteht. Zu den Einzelheiten der Daten und der im folgenden diskutierten Schätzungen vgl. Gundlach et al. (in Vorbereitung).

wird (Spalten (1)-(3)), kann der Regressionskoeffizient für die Schuljahre S nach Gleichung (7) als Ertragsrate der schulischen Ausbildung interpretiert werden. Das Schätzergebnis in Spalte (1) stimmt mit dem Befund von Heckman und Klenow (1997) überein: Wenn lediglich die Anzahl der Schuljahre als erklärende Variable benutzt wird, erhält man eine makroökonomische Ertragsrate des Humankapitals von über 30 Prozent. Demgegenüber steht die aus der mikroökonomischen Literatur bekannte durchschnittliche private Ertragsrate von rund 9 Prozent. Demnach könnte man auf enorme Humankapitalexternalitäten in einer Größenordnung von über 20 Prozent schließen. Dieser Befund ist aber keineswegs robust, wie die Ergebnisse der nachfolgenden Spezifikationen zeigen.

Nimmt man beispielsweise auch die Erfahrung der Arbeitskräfte (E) und die Lebenserwartung ($LIFE$) als Kontrollvariablen mit ins Bild (Spalten (2) und (3)), dann sinkt die geschätzte makroökonomische Ertragsrate des Humankapitals auf 20 Prozent. Nimmt man auch noch die Kapitalintensität (k) als erklärende Variable hinzu, dann sinkt der geschätzte Regressionskoeffizient der Variable Schuljahre weiter auf 12 Prozent (Spalte (4)). Heckman und Klenow (1997) schätzen mit ihren Daten für eine solche Spezifikation einen durchschnittlichen Regressionskoeffizienten von rund 9 Prozent und schließen daraus, dass es keinen nennenswerten Unterschied zwischen der makroökonomischen und der privaten Ertragsrate des Humankapitals gibt. Die öffentliche Subventionierung der schulischen Ausbildung scheint Ihnen demnach im internationalen Durchschnitt gerade hoch genug zu sein, um die vorhandenen positiven Externalitäten des Humankapitals zu internalisieren.

Bei dieser Interpretation übersehen Heckman und Klenow (1997) aber, dass der geschätzte Regressionskoeffizient der Variable Schuljahre nach Gleichung (9) nicht mehr der makroökonomischen Ertragsrate des Humankapitals entspricht. Der Regressionskoeffizient entspricht vielmehr dem Produkt aus Ertragsrate und Produktionselastizität. Da die Produktionselastizität des Sachkapitals (α) als Regressionskoeffizient der Kapitalintensität geschätzt wird, folgt die soziale Ertragsrate des Humankapitals jetzt als

$$(10) \quad r_s = \frac{r}{1-\alpha} = 0,21 \quad .$$

Der zugehörige Standardfehler, der über eine nichtlineare Schätzung ermittelt werden kann, beträgt 0,04. Damit wäre man zurück bei dem Ergebnis, dass die makroökonomische Ertragsrate deutlich, nämlich um rund 100 Prozent, über der privaten Ertragsrate des Humankapitals liegt.

Eine solche Größenordnung der geschätzten Humankapitalexternalität erscheint wenig plausibel. Alternative Spezifikationen und Schätzverfahren versprechen Aufschluss darüber, wie robust der empirische Befund wirklich ist. Die erste Spalte von Tabelle 6 enthält die Ergebnisse einer Schätzung, bei der neben der Anzahl der Schuljahre auch die Qualität (Q) der schulischen Ausbildung berücksichtigt wird. Mit der alternativen Spezifikation der Humankapital-Variable (SQ) analog zu Gleichung (3) erhält man eine soziale Ertragsrate in Höhe von 14 Prozent, die wieder näher an die Ergebnisse von Heckman und Klenow (1997) heranrückt. Dabei ist die geschätzte Ertragsrate statistisch signifikant verschieden von einer Referenz-Ertragsrate in Höhe von 9 Prozent, wie sie aufgrund einer mikroökonomischen Mincer-Spezifikation zu erwarten wäre. Dieser Befund bleibt bestehen, wenn statt der Lebenserwartung ($LIFE$) der absolute Abstand eines Landes vom Äquator ($DISTANCE$) als Näherungsgröße für mögliche internationale Technologieunterschiede benutzt wird (Spalte (2)).

Die bisher benutzte Schätzmethode (OLS) liefert allerdings nur dann unverzerrte Ergebnisse für die soziale Ertragsrate des Humankapitals, wenn die Liste der erklärenden Variablen vollständig ist und wenn es kein Endogenitätsproblem gibt. Ob die Liste der erklärenden Variablen in Gleichung (9) vollständig ist, lässt sich mit Hilfe empirischer Methoden nicht eindeutig beantworten. Aus produktionstheoretischer Sicht erscheint aber eine Spezifikation mit den Faktoren Sachkapital, Humankapital und Technologie als angemessen. Gleichzeitig dürfte aus produktionstheoretischer Sicht kein Zweifel daran bestehen, dass sowohl die Sachkapital-Variable als auch die Humankapital-Variable als endogen zu betrachten sind. Dies könnte zu einer Überschätzung der Regressionskoeffizienten von k und SQ führen. Die daraus

resultierende Richtung des Effekts auf die geschätzte makroökonomische Ertragsrate hängt, wie Gleichung (10) verdeutlicht, von der relativen Verzerrung der beiden Regressionskoeffizienten ab.

Um das potentielle Endogenitätsproblem zu lösen, kann man in einem ersten Schritt den Regressionskoeffizienten von k auf den Wert der gesamtwirtschaftlichen Gewinnquote restringieren. Spalte (3) enthält die empirischen Ergebnisse für eine Spezifikation mit einer auf 30 Prozent (Gollin 2002) restringierten Gewinnquote. Diese Spezifikation hat keinen nennenswerten Einfluss auf die Höhe der geschätzten makroökonomischen Ertragsrate des Humankapitals. Spalte (4) enthält die empirischen Ergebnisse einer Schätzung, bei der die mögliche Endogenität der Humankapital-Variable mit Hilfe von Instrument-Variablen kontrolliert wird. Als Instrumente dienen dabei vier Variablen, die von Hall und Jones (1999) zur Instrumentierung der sozialen Infrastruktur eines Landes benutzt wurden: der absolute Abstand eines Landes vom Äquator, der über geographische und klimatische Variablen geschätzte Grad der Offenheit eines Landes für Handel und Kapitalverkehr, der Anteil der Bevölkerung mit Kenntnis der englischen Sprache sowie der Anteil der Bevölkerung mit Kenntnis einer der wesentlichen europäischen Sprachen.

Alle diese Variablen (mit Ausnahme der Offenheitsvariable) sind mit der Humankapital-Variable (SQ) eng korreliert und hängen gleichzeitig *nicht* vom Pro-Kopf-Einkommen (y_c) ab. Die IV-Schätzung bestätigt abermals eine statistisch hoch signifikante makroökonomische Ertragsrate des Humankapitals, die deutlich über dem Mincerschen Referenzwert von 9 Prozent liegt. Die vormals relativ hohe Ertragsrate der Variable Erfahrung (E) erweist sich in dieser Spezifikation als statistisch insignifikant. Der Test auf Überidentifikation besagt, dass keine der Instrument-Variablen einen direkten Einfluss auf die abhängige Variable ausübt.⁶

⁶ An dieser Stelle ist wiederum kritisch anzumerken, dass der Test auf Überidentifikation keine Aussage darüber erlaubt, ob wie angenommen mindestens eine der gewählten Instrumentvariablen tatsächlich keinen direkten Einfluss auf die abhängige Variable ausübt.

Insgesamt betrachtet deuten diese empirischen Ergebnisse darauf hin, dass die soziale Ertragsrate des Humankapitals um rund 50 Prozent über der privaten liegen könnte. Ein solches Ergebnis würde implizieren, dass die Produktion von Humankapital in Schule und Universität mit erheblichen positiven Externalitäten einherzugehen scheint, die offenbar nicht vollständig durch die bereits existierende staatliche Subventionierung des Bildungssystems internalisiert werden. Für sich genommen bietet dieser empirische Befund für einen internationalen Querschnitt ein Argument dafür, die staatlichen Bildungsinvestitionen im Mittel auszuweiten, um so die private mit der gesamtwirtschaftlichen Ertragsrate des Humankapitals in Einklang zu bringen. Offen bleiben muss an dieser Stelle wie schon am Ende von Abschnitt B. II, welche Befunde sich für eine alternative Spezifikation der Humankapitalvariable sowie für alternative Instrumentvariablen ergeben würden. Wenn sich der vorliegende Befund tatsächlich als robust erweisen sollte stellt sich unmittelbar die Frage, welche speziellen staatlichen Bildungsinvestitionen denn gegebenenfalls nötig wären, um einen höheren Humankapitalbestand auch und gerade über eine höhere Qualität der Ausbildung zu erreichen.

C. Mikroökonomische Determinanten der Bildungsqualität

Die makroökonomischen Untersuchungen haben gezeigt, dass eine hohe Qualität der Bildung unter Umständen positive auf die Durchschnittseinkommen wirkt, möglicherweise einen eigenen positiven Effekt auf das Einkommen der Armen besitzt sowie eine positive Einkommensexternalitäten hervorrufen könnte. Damit stellt sich unmittelbar die Frage, wie eine hohe Bildungsqualität zu mit Hilfe wirtschaftspolitischer Maßnahmen erreichen sein könnte. Aus Sicht der Bildungspolitik lassen sich die Optionen in zwei Kategorien einteilen: Erstens kann die Bildungspolitik die Ressourcenausstattung im gegebenen Schulsystem erhöhen, und zweitens kann sie die institutionelle Struktur des Schulsystems bei einer gegebenen Ressourcenausstattung verändern. Die Einflussmöglichkeiten einer Ressourcenpolitik werden im Folgenden anhand der Bildungsressource „Klassengröße“ untersucht. Im

Anschluss daran werden die Einflussmöglichkeiten einer institutionellen Politik anhand der Bildungsinstitution „Zentralprüfung“ analysiert.

I. Klassengrößen und Schülerleistungen

Will man den Einfluss von Ressourcen auf Schülerleistungen berechnen, so sieht man sich dem Problem gegenüber, dass die Ressourcen nicht nur Ursache, sondern auch Folge der Schülerleistungen sein können. Nimmt man als Beispiel die Bildungsressource „Klassengröße“, so kommen eine ganze Reihe von Wirkungsmechanismen in Frage, die eine umgekehrte Kausalität bewirken (vgl. West und Wößmann 2003). Zum Beispiel können Wahlentscheidungen der Eltern, wo sie wohnen und zu welcher Schule sie ihre Kinder schicken wollen, mit den Leistungen der Schüler korreliert sein. Auch können die Aufteilung der Schüler auf Schulen mit unterschiedlichen Klassengrößen aufgrund von Regelungen des Schulsystems sowie die Zuweisung der Schüler in verschiedene Klassen innerhalb einer Schule auf den Leistungen dieser Schüler basieren. So hat etwa Lazear (2001) gezeigt, dass die optimale Klassengröße mit dem Verhalten der Schüler, und damit vermutlich auch mit ihren Leistungen, variieren kann.

Wann immer Wahlentscheidungen dazu führen, dass Schüler in nicht zufälliger Weise in Klassen unterschiedlicher Größe unterrichtet werden, besteht aus ökonomischer Sicht ein Endogenitäts- oder Simultanitätsproblem: Die beobachtete Variation in Klassengrößen ist teilweise verursacht von Variationen in Schülerleistungen beziehungsweise korreliert mit anderen Determinanten der Schülerleistungen. Deshalb würde eine einfache Schätzung von Klassengrößeneffekten mit der Methode der kleinsten Quadrate (KQ) zu verzerrten Ergebnissen führen (Hoxby 2000a). Auch wenn es noch immer einige Diskussion über die statistische Signifikanz der Befunde der zahlreichen KQ-Schätzungen von Klassengrößeneffekten gibt (Hanushek 2003; Krueger 2003), so vernachlässigen diese das gewichtigere Problem, dass eine kausale Interpretation dieser KQ-Schätzungen nicht möglich ist. Um zu einer kausalen Aussage kommen zu können, muss der kausale Klassengrößeneffekt aus der gesamten Korrelation zwischen Klassengröße und Schülerleistung, die zumindest teilweise auch

durch Zuweisungsentscheidungen bedingt ist, extrahiert werden. Es bedarf also einer Identifikationsstrategie, die solche Klassengrößenvariationen zu identifizieren vermag, die in Bezug auf die Schülerleistungen exogen sind.

Prinzipiell bestehen zwei Möglichkeiten zur Identifikation von zufälliger oder exogener Variation in Klassengrößen. Die erste besteht darin, ein explizites Experiment durchzuführen, das Schüler in zufälliger Weise auf Klassen unterschiedlicher Größe aufteilt. Das einzige größere bisher durchgeführte Klassengrößen-Experiment, das sogenannte „Project STAR“ im US-Bundesstaat Tennessee in den 1980er Jahren, ist von Krueger (1999) und Krueger und Whitmore (2001) im Detail analysiert worden. Leider haben explizite Experimente im Allgemeinen (Hoxby 2000a) und Project STAR im Besonderen (Hanushek 1999) einige Probleme und Defizite aufzuweisen, die die Validität und Allgemeingültigkeit der Ergebnisse in Frage stellen.

So bleibt als zweite Alternative die quasi-experimentelle Herangehensweise, die versucht, durch sogenannte „natürliche“ Experimente (vgl. Rosenzweig und Wolpin 2000) hervorgerufene exogene Variationen in Klassengrößen zu identifizieren. In solchen Fällen ist es möglich, durch die Benutzung von Instrument-Variablen (IV) die Schätzung des Klassengrößeneffektes auf einen solchen Teil der gesamten Klassengrößenvariation zu beschränken, der exogen in Bezug auf die Schülerleistungen ist. Im Prinzip vergleicht die IV-Schätzung Schülerleistungen in Klassen, deren Größe aus exogenen Gründen unterschiedlich ist, so dass der IV-Schätzer anzeigt, ob Klassengrößenunterschiede Leistungsunterschiede verursachen. Mittlerweile gibt es einige solcher quasi-experimenteller Schätzungen von Klassengrößeneffekten (Angrist und Lavy 1999; Case und Deaton 1999; Hoxby 2000a). Da diese aber jeweils notwendigerweise auf relativ spezifischen Variationen in Klassengrößen basieren, ist eine vergleichende Analyse der Ergebnisse dieser Studien nur schwer möglich.

Wößmann und West (2002) haben eine Identifikationsstrategie entwickelt, die spezifische Merkmale des Datensatzes der Third International Mathematics and Science Study (TIMSS) nutzt, um exogene Klassengrößenvariationen zu identifizieren. Da diese Identifikationsstrategie in vielen Ländern in vergleichbarer Weise implementiert werden

kann, ermöglicht sie einen direkten Vergleich der für verschiedene Länder geschätzten Klassengrößeneffekte.

Die Intuition dieser Identifikationsstrategie ist wie folgt. Da zahlreiche Wahlentscheidungen die Allokation von Schülern auf verschiedene Schulen in nicht-zufälliger Weise bestimmen, wird für die Schätzung von Klassengrößeneffekten in einem ersten Schritt die gesamte Variation zwischen Schulen herausgerechnet, indem ein kompletter Satz von Schul-Dummies („Fixed Effects“) eingeführt wird. Damit basiert der Schätzer ausschließlich auf Variationen innerhalb einzelner Schulen. Da die Zuweisung auf verschiedene Klassen innerhalb eines Jahrgangs aber auch innerhalb der Schulen in nicht-zufälliger Weise vonstatten gehen kann, wird in einem zweiten Schritt die jeweilige tatsächliche Klassengröße durch die durchschnittliche Klassengröße der entsprechenden Jahrgangsstufe in der Schule instrumentiert.⁷ Das bedeutet, dass Klassengrößenvariationen innerhalb einzelner Schulen nur insoweit berücksichtigt werden, als sie durch Variationen in der durchschnittlichen Klassengröße zwischen Jahrgangsstufen verursacht werden. Solche Variationen sind nicht durch Zuweisungsentscheidungen beeinflusst, sondern spiegeln in erster Linie zufällige Unterschiede in der Größe der Geburtenjahrgänge wider, die zu Schwankungen in der durchschnittlichen Klassengröße zwischen zwei Jahrgängen führen. Solche Schwankungen liefern also eine Klassengrößenvariation, die exogen in Bezug auf die Leistungen der Schüler ist.

Gibt es nun systematische Unterschiede in den Leistungen von Schülern in verschiedenen Klassen, die aufgrund der exogenen Variation unterschiedlich groß sind, so lassen sich diese Leistungsunterschiede als kausale Folge der Klassengrößenunterschiede interpretieren. Kurz gefasst identifiziert diese Identifikationsstrategie also kausale Klassengrößeneffekte, indem sie die relativen Leistungen der Schüler zweier Jahrgangsstufen innerhalb einzelner Schulen dem Teil

⁷ Für eine detaillierte Diskussion der Validität des Instruments, der Identifikationsstrategie insgesamt sowie der Datenbasis und der Ergebnisse siehe Wößmann und West (2002).

der Klassengrößenunterschiede gegenüberstellt, der Unterschiede in der durchschnittlichen Klassengröße zwischen den Jahrgangsstufen widerspiegelt.

Wie aus der Beschreibung der Identifikationsstrategie deutlich wird, benötigt man für ihre Umsetzung einen recht spezifischen Datensatz. Man benötigt nicht nur für die Fixed-Effects-Schätzung Daten für mehrere Klassen innerhalb einzelner Schulen, sondern auch für die IV-Schätzung Daten für Klassen aus mehreren Jahrgangsstufen innerhalb der Schulen, da die Klassen ansonsten das gleiche Instrument hätten und keine Klassengrößenvariation mehr für die Schätzung übrig bliebe. Darüber hinaus muss der Datensatz auch noch Informationen sowohl über die tatsächlichen Klassengrößen als auch über die durchschnittlichen Klassengrößen der Jahrgangsstufen beinhalten. All diese Merkmale treffen auf den TIMSS-Datensatz zu, der 1994/95 von der International Association for the Evaluation of Educational Achievement (IEA) erhoben wurde. In der TIMSS-Mittelstufenstudie wurden die beiden Jahrgangsstufen mit dem größten Anteil Dreizehnjähriger getestet, was in den meisten Ländern der siebten und achten Jahrgangsstufe entspricht. Innerhalb jedes Landes wurde eine repräsentative Stichprobe von rund 150 Schulen erhoben, innerhalb derer jeweils (zufällig ausgewählt) eine siebte Klasse und eine achte Klasse komplett getestet wurden.

Zur Implementierung der Identifikationsstrategie können nur Daten solcher Schulen verwendet werden, in denen tatsächlich sowohl eine siebte als auch eine achte Klasse getestet wurden, und in denen die Daten über tatsächliche sowie jahrgangsdurchschnittliche Klassengrößen für beide Jahrgangsstufen verfügbar sind. Dies führt in einigen Ländern zu Datenproblemen, da in einigen Schulsystemen die siebte und achte Klasse in unterschiedlichen Schulen gelehrt wird. Darüber hinaus stammen die Klassengrößendaten aus unterschiedlichen Hintergrundfragebögen – Daten über tatsächliche Klassengrößen aus Lehrerfragebögen und Daten über jahrgangsdurchschnittliche Klassengrößen aus Schulleiterfragebögen –, und aufgrund teilweise schlechter Beantwortungsraten ist die Datenbasis in einigen Ländern relativ schlecht. So wurde entschieden, die Schätzungen nur in solchen Ländern zu implementieren, in denen all diese Daten für mindestens 100 Klassen in 50 Schulen

verfügbar sind. Aufgrund der schlechten Datenlage scheidet daher auch Deutschland aus der Untersuchung aus, da hier zum Beispiel in 43 Prozent der Fälle die Angaben über die durchschnittlichen Klassengrößen in den Fragebögen der Schulleiter fehlen.

Tabelle 7 berichtet die Schätzergebnisse für alle Länder mit ausreichender Datenlage, getrennt nach Mathematik und Naturwissenschaften. Die erste Spalte berichtet jeweils den geschätzten Klassengrößenkoeffizienten α_1 , der sich aus einer Kleinstquadrat-Schätzung folgender Schätzgleichung ergibt:

$$(11) \quad T_{icgs} = \alpha_1 S_c + Ctrl_{icgs} \beta + \gamma G_g + \nu_c + \varepsilon_{icgs} .$$

In dieser Gleichung sind T_{icgs} das Testergebnis des Schülers i in Klasse c in Jahrgangsstufe g in Schule s , S die Klassengröße, $Ctrl$ ein Vektor von zwölf Kontrollvariablen für den familiären Hintergrund des Schülers (erhoben in von den Schülern ausgefüllten Hintergrundfragebögen), G die Jahrgangsstufe, ν eine klassenspezifische Komponente des Schätzfehlers und ε eine schülerspezifische Komponente des Schätzfehlers.⁸ Die TIMSS-Testergebnisse sind so skaliert, dass sie in jedem Fach einen internationalen Mittelwert von 500 und eine internationale Standardabweichung von 100 aufweisen.

In den meisten Fällen ist dieser KQ-Schätzer des Klassengrößeneffekts statistisch signifikant positiv. Das bedeutet, dass Schüler in *größeren* Klassen statistisch signifikant *besser* abschneiden. Als Klassengrößeneffekt wäre dies genau umgekehrt zur üblichen Erwartung, dass kleinere Klassen das Lernen positiv beeinflussen können.

⁸ Die Schätzungen erfolgen folglich mit Mikrodaten auf Schülerebene. Die Fehlerkomponenten werden durch sogenannte Cluster-robuste Schätzungen (clustering-robust linear regression, CRLR) implementiert, die mögliche Interdependenzen der Fehlerterme von Schülern innerhalb einzelner Klassen bei der Berechnung der Standardfehler berücksichtigen (vgl. Moulton 1986; Deaton 1997: S. 74-78). Der stratifizierten Stichprobenerhebung in TIMSS wird dadurch Rechnung getragen, dass die Observation eines jeden Schülers mit seiner Erhebungswahrscheinlichkeit gewichtet wird (vgl. DuMouchel und Duncan 1983; Wooldridge 2001).

Wie zuvor diskutiert, kann dieses KQ-Ergebnis aber auch die umgekehrte Kausalität von mit Schülerleistungen korrelierten Zuweisungsentscheidungen widerspiegeln.

Entsprechend der zuvor beschriebenen Identifikationsstrategie wird in einer zweiten Schätzung die gesamte Variation zwischen Schulen herausgerechnet, indem für einen kompletten Satz von Schul-Dummies D für jede einzelne Schule kontrolliert wird:

$$(12) \quad T_{icgs} = \alpha_2 S_c + Ctrl_{icgs} \beta + \gamma G_g + D_s \delta + \nu_c + \varepsilon_{icgs} \quad .$$

Der in dieser Fixed-Effects-Spezifikation (FE) geschätzte Klassengrößenkoeffizient α_2 ist nicht von Zuweisungsentscheidungen zwischen Schulen beeinflusst, kann aber noch durch Zuweisungsentscheidungen innerhalb der Schulen verzerrt sein. Wie die Ergebnisse in Tabelle 7 zeigen, verschwindet ein Teil der unerwarteten statistisch signifikanten positiven Schätzer bei der FE-Schätzung.

Um auch noch nicht-zufällige Zuweisung innerhalb von Schulen zu berücksichtigen, wird in der FE-IV-Schätzung zusätzlich zu den fixen Schuleffekten die tatsächliche Klassengröße S durch die durchschnittliche Klassengröße der Jahrgangsstufe A instrumentiert. Das heißt, dass in der ersten Stufe der zweistufigen Schätzung die Klassengröße auf das Instrument und alle weiteren exogenen Variablen regressiert wird:

$$(13) \quad S_c = \phi A_{gs} + Ctrl_{icgs} \beta + \gamma G_g + D_s \delta + \nu_c + \varepsilon_{icgs} \quad ,$$

um dann in der zweiten Stufe den so geschätzten Wert \hat{S} dieser ersten Stufe anstelle der tatsächlichen Klassengröße S zu verwenden:

$$(14) \quad T_{icgs} = \alpha_3 \hat{S}_c + Ctrl_{icgs} \beta + \gamma G_g + D_s \delta + \nu_c + \varepsilon_{icgs} \quad ,$$

wobei entsprechend der zuvor diskutierten Identifikationsstrategie der geschätzte Koeffizient α_3 den kausalen Effekt der Klassengröße auf die Schülerleistungen angibt.

Die FE-IV-Schätzung stellt hohe Ansprüche an die Datenbasis, da sowohl jegliche Variation zwischen Schulen als auch jegliche Variation innerhalb von Jahrgangsstufen einer Schule unberücksichtigt bleiben. Dies führt dazu, dass die FE-IV-Schätzer in einigen Ländern sehr unpräzise sind und deshalb keine Aussage über die Existenz von Klassengrößeneffekten zulassen. Dies ist der Fall in den acht FE-IV-Schätzern mit sehr großen Standardfehlern von den insgesamt 36 in Tabelle 7 berichteten FE-IV-Schätzern. Für die restlichen 28 FE-IV-Schätzer ergibt sich folgendes Bild: In vier Fällen wird ein statistisch signifikant negativer Koeffizient geschätzt; in diesen Fällen – Frankreich und Island in Mathematik und Griechenland und Spanien in Naturwissenschaften – führen kleinere Klassengrößen zu besseren Schülerleistungen. Im Großteil der Fälle ergibt sich kein statistisch signifikanter Klassengrößeneffekt, und in nur einem Fall bleibt der unerwartete statistisch signifikant positive Koeffizient.

Die unterschiedlichen Ergebnisse der drei Schätzstrategien werden in Schaubild 1 noch einmal deutlich. Dort werden für die 28 Fälle, in denen die FE-IV-Schätzung zu einigermaßen präzisen Schätzern führt, die Schätzer in statistisch signifikant positive, statistisch insignifikant positive, statistisch insignifikant negative und statistisch signifikant negative eingeteilt. Während die KQ-Schätzungen hauptsächlich zu den unerwartet statistisch signifikant positiven Koeffizienten führen, zeigen die FE-IV-Schätzungen, dass diese unerwarteten Schätzer auf nicht-zufällige Zuweisungsentscheidungen zwischen und innerhalb von Schulen zurückzuführen sind.

Es stellt sich die Frage, ob die große Anzahl statistisch insignifikanter Schätzer aufgrund von wenig präzisen Schätzungen zustande kommt, oder ob sie die Tatsache widerspiegelt, dass im Großteil der Fälle eben keine substantiellen Klassengrößeneffekte auftreten. Um diese Frage zu beantworten, wurde statistisch getestet, ob die FE-IV-Schätzer sich signifikant von der Größenordnung unterscheiden, die Krueger (1999) für das Project STAR ermittelt hat.⁹ Auf die TIMSS-Skala

⁹ Krueger (1999) zeigt, dass sich in einer groben Kosten-Nutzen-Rechnung die ökonomischen Nutzen einer Klassengrößenverringerung, die sich aufgrund der Leistungssteigerung bei zukünftigen Einkommen ergeben, bei einem solchen Klassengrößeneffekt in etwa mit den Kosten decken.

übertragen würden Krueger's Befunde einem Schätzer von -3 entsprechen. Wie in Tabelle 7 berichtet, kann ein Klassengrößeneffekt dieser Größenordnung in zehn Fällen in Mathematik und in elf Fällen in Naturwissenschaften statistisch signifikant ausgeschlossen werden. In einem weiteren Test zeigt sich, dass in insgesamt acht Fällen sogar noch kleinere Klassengrößeneffekte von -1 statistisch signifikant ausgeschlossen werden können. Insgesamt betrachtet scheint die Bildungsressource Klassengröße in den meisten Fällen also keinen kausalen Effekt auf die Schülerleistungen auszuüben. Diese internationalen Schätzungsergebnisse bestätigen frühere Ergebnisse für die USA (Hanushek 2002), Evidenz im Länderquerschnitt (Wößmann 2002a) sowie Zeitreihenevidenz für die meisten OECD-Länder (Gundlach et al. 2001), wonach eine bessere Ausstattung mit Bildungsressourcen zumeist nicht zu höheren Schülerleistungen führt.

Die internationale Evidenz für einen Querschnitt von Ländern erlaubt es weiterhin zu fragen, in welchen Ländern Klassengrößeneffekte auftreten und in welchen nicht. Wie sich aus der Untersuchung in Wößmann und West (2002) ergibt, scheint das Auftreten von Klassengrößeneffekten nicht mit der durchschnittlichen Klassengröße oder dem allgemeinen Ausgabenniveau eines Landes zusammenzuhängen; es besteht also keine Evidenz dafür, dass die Ergebnisse in erster Linie durch abnehmende Ertragsraten der Ressourcennutzung bestimmt sind. Demgegenüber scheint das Auftreten von Klassengrößeneffekten aber mit dem Gehaltsniveau der Lehrer zusammenzuhängen: Klassengrößeneffekte treten nur in Ländern mit einem relativ niedrigen Gehaltsniveau der Lehrer auf, während Länder mit relativ hohen Lehrergehältern keine Klassengrößeneffekte aufweisen.

Sieht man das durchschnittliche Gehaltsniveau der Lehrer als ein Maß für das allgemeine Fähigkeitsniveau der Lehrerschaft an, so legt dies die Interpretation nahe, dass relativ fähige Lehrer mit dem Unterricht in großen Klassen genauso gut zurecht kommen wie in kleinen Klassen, so dass sich bei ihnen kein Effekt der Klassengröße auf die Schülerleistungen ergibt. Demgegenüber scheinen weniger fähige Lehrer mit dem Unterricht in großen Klassen überfordert, während sie in kleinen Klassen einigermaßen zurechtkommen – bei ihnen ergibt sich ein Klassengrößeneffekt. Die

internationale Variation der Schätzergebnisse legt also nahe, dass Klassengrößeneffekte nur bei einem niedrigen Fähigkeitsniveau der Lehrerschaft auftreten.

Die Schätzungen haben gezeigt, dass eine überzeugende ökonometrische Identifikation von exogener Variation für die Ergebnisse von entscheidender Bedeutung ist. Für die Bildungsressource Klassengröße liegt eine solche mit der FE-IV-Schätzung vor, und die Ergebnisse legen nahe, dass kleinere Klassen nur in den wenigsten Fällen zu besseren Schülerleistungen führen – und vor allem nur in Ländern mit niedrigen Lehrergehältern, zu denen Deutschland nicht gehört. Leider liegen vergleichbare Identifikationsstrategien für andere Bildungsressourcen nicht vor, so dass eine abschließende Bewertung alternativer Ressourcenpolitiken nicht möglich ist. Allerdings legen KQ-Schätzungen – die freilich durch Endogenitätsprobleme verzerrt sein können – nahe, dass auch andere Bildungsressourcen wie die materielle Ausstattung, die Unterrichtszeit oder der Ausbildungsstand und die Lehrerfahrung der Lehrer in den meisten Ländern nicht mit höheren Schülerleistungen einhergehen (vgl. etwa Hanushek 2002; Wößmann 2003c).

II. Zentralprüfungen und Schülerleistungen

Da für die überwiegende Anzahl der betrachteten Industrie- und Schwellenländer keine statistisch signifikanten Ressourceneffekte berichtet werden können, wenden wir uns der zweiten Option der Schulpolitik zur Erhöhung der Bildungsqualität zu, nämlich der institutionellen Struktur des Schulsystems. Dies soll anhand der Bildungsinstitution „Zentralprüfung“ geschehen. In der empirischen Implementierung ist darunter jegliche Form von zentraler Abschlussprüfung zu verstehen, die die Leistungen aller Schüler auf eine für die einzelne Schule externe Weise testet. Dabei kommt es weniger darauf an, ob die Prüfung „zentral“ von einer Behörde auf der höchsten Administrationsebene eines Landes durchgeführt wird. Vielmehr geht es darum, dass nicht die einzelnen Lehrer selbst bestimmen und bewerten, was getestet wird, sondern dass den Schülern, Lehrern und Schulen die abzufragenden Wissensgebiete „extern“ von außen vorgegeben werden, und dass die Schüler nicht nur relativ zu den anderen Schülern der Klasse oder Schule, sondern relativ zu allen Schülern des von der Zentralprüfung abgedeckten Gebietes bewertet werden.

Im internationalen Vergleich haben sich Institutionen wie Entscheidungscentralität, Schulautonomie, Lehrereinfluss oder privater Wettbewerb generell als wichtige Determinanten von Unterschieden in Schülerleistungen erwiesen (Wößmann 2002a, 2003b). Aus theoretischer Sicht bestimmen solche Institutionen die Anreize, denen sich alle am Bildungsprozess beteiligten Personengruppen – Schüler, Lehrer, Schulleiter, Behörden, Eltern – gegenüber sehen (Bishop und Wößmann 2003). Je nachdem, ob die Institutionen so gestaltet sind, dass die Beteiligten positive Anreize zur Verbesserung der Schülerleistungen verspüren – dass also leistungsförderndes Verhalten belohnt und leistungsschwächendes Verhalten sanktioniert wird –, oder ob sie so gestaltet sind, dass diese Anreize kaum vorliegen, kommt es im Schulsystem zu hohen oder niedrigen Schülerleistungen. Externe Prüfungen sind eine Bildungsinstitution, die dazu beitragen kann, dass die Anreize aller Beteiligten auf die Erhöhung der Schülerleistungen ausgerichtet werden (Wößmann 2002b). Sie machen den Leistungsstand der Schüler für Eltern, Lehrer, potentielle Arbeitgeber und weiterführende Bildungseinrichtungen sichtbar und vergleichbar, so dass bessere Leistungen belohnt werden können; sie verhindern, dass ganze Wissensgebiete in einzelnen Klassen ohne Konsequenzen für die Benotung ausgelassen werden können; und sie eröffnen Eltern und Schulleitern, ob die Lehrer eine erfolgreiche Wissensvermittlung leisten.

Diese Auswirkungen von externen Prüfungen sollen empirisch getestet werden, indem die Leistungen von Schülern in Schulsystemen mit Zentralprüfungen mit den Leistungen von Schülern in Schulsystemen ohne Zentralprüfungen international verglichen werden. Um möglichst viele Observationen verschiedener Länder mit und ohne Zentralprüfungen zu haben, werden für diese Schätzungen die TIMSS-Daten mit den Daten der 1999 durchgeführten Wiederholungsstudie TIMSS-Repeat gepoolt. Während in TIMSS Daten für 266.545 Schüler aus 39 Ländern vorliegen, sind es in TIMSS-Repeat 180.544 Schüler aus 38 Ländern. Nur 23 dieser Länder haben an beiden Tests teilgenommen, so dass der gepoolte Datensatz Daten für insgesamt 54 verschiedene Länder beinhaltet.

Der Effekt λ von Zentralprüfungen Z wird mit Hilfe folgender Spezifikation geschätzt:

$$(15) \quad T_{icgso} = \lambda Z_o + Ctrl_{icgso} \beta + \mu_o + \nu_c + \varepsilon_{icgs} ,$$

wobei Z_o eine Dummy-Variable ist, die angibt, ob im Land o Zentralprüfungen gegeben sind oder nicht.¹⁰ Der Kontrollvariablenvektor umfasst in dieser Spezifikation Jahrgangsstufendummies, 15 Variablen für den familiären Hintergrund der Schüler, 13 Variablen für die schulische Ressourcenausstattung (einschließlich der Klassengröße) und Lehrermerkmale sowie 18 Variablen für weitere institutionelle Merkmale der Schulsysteme wie zum Beispiel den Grad der Schul- und Lehrerautonomie in verschiedenen Entscheidungsbereichen;¹¹ μ_o ist eine zusätzliche länderspezifische Komponente des Schätzfehlers.¹²

Die in Tabelle 8 berichteten Schätzergebnisse belegen, dass Schüler in Schulsystemen mit zentralen Prüfungen statistisch signifikant besser abschneiden als Schüler in Schulsystemen ohne zentrale Prüfungen. Dies gilt sowohl für Mathematik als auch für Naturwissenschaften, und es gilt sowohl für TIMSS als auch für TIMSS-Repeat. Im gepoolten Datensatz beträgt der Zentralprüfungseffekt 42,7 Prozent einer internationalen Standardabweichung in Mathematik und 35,9 Prozent in Naturwissenschaften. Dies entspricht in etwa dem Leistungsunterschied zwischen Schülern der siebten Klasse und Schülern der achten Klasse, also dem in einem ganzen Schuljahr erlernten Wissen. Die Ergebnisse von TIMSS-Repeat untermauern demzufolge zuvor ausschließlich anhand von TIMSS erzielte Ergebnisse (Bishop 1997; Wößmann 2002a), wonach Schülerleistungen in Zentralprüfungssystemen höher sind.¹³

¹⁰ In Ländern, in denen nur in einigen Regionen Zentralprüfungen gegeben sind, gibt Z den Anteil der Schüler an, der an Zentralprüfungen teilnimmt.

¹¹ Für detaillierte Angaben zu den Datensätzen, der Modellspezifikation und den Ergebnissen zum Einfluss von Zentralexamen siehe Wößmann (2002b).

¹² Die Schätzung erfolgt wiederum anhand schülerspezifischer Mikrodaten mit Hilfe gewichteter Cluster-robuster Schätzungen (siehe oben).

¹³ Der Leistungsvorsprung von Schülern in Zentralprüfungssystemen ist auch innerhalb Deutschlands bei einem Vergleich von Bundesländern mit und ohne Zentralprüfungen festzustellen, sowohl für TIMSS (Jürges et al. 2002) als auch für die

Auch ist die Größe des anhand der TIMSS-Repeat-Daten geschätzten Effektes statistisch nicht signifikant von dem anhand der TIMSS-Daten geschätzten Effekt zu unterscheiden.

Wie zuvor am Beispiel der Klassengrößeneffekte diskutiert, ist es auch im Falle der Bildungsinstitution Zentralprüfung denkbar, dass diese KQ-Schätzer durch Endogenitätsprobleme verzerrt sind. Dabei ist vor allem daran zu denken, dass es unberücksichtigte Faktoren („Omitted Variables“) auf Länderebene geben könnte, die mit der Existenz von Zentralprüfungssystemen korrelieren und die die Korrelation mit den Schülerleistungen verursachen. Hierzu fallen vor allem vier Bereiche ein: erstens andere institutionelle Gegebenheiten des Schulsystems; zweitens der allgemeine Zentralisationsgrad eines Landes; drittens die Homogenität der Bevölkerung; und viertens kulturelle Unterschiede zwischen Ländern. Insoweit die Existenz von Zentralprüfungen nicht zufällig über die Länder verteilt ist, sondern etwa mit anderen Institutionen einhergeht oder in zentralisierten und homogenen Ländern besonders häufig auftritt, und insoweit diese anderen Ländermerkmale Unterschiede in den Schülerleistungen hervorrufen, würde der einfache KQ-Schätzer des Zentralprüfungseffektes von solchen Einflüssen verzerrt sein.

In den ersten drei Bereichen ist es möglich, die Verzerrung zumindest zum größten Teil durch die Berücksichtigung zusätzlicher entsprechender Kontrollvariablen zu beheben. So beinhaltet die berichtete Spezifikation bereits eine große Anzahl institutioneller Kontrollvariablen (einschließlich der Zentralisierung des Curriculums und der Lehrbuchgenehmigung), und entsprechende Tests zeigen, dass ihre Berücksichtigung keinen signifikanten Einfluss auf den geschätzten Zentralprüfungseffekt hat, obwohl viele der weiteren Institutionen durchaus selbst statistisch signifikante Effekte auf die

PISA-Ergänzungsstudie von 2000, in der die deutschen Bundesländer in weit umfangreicherem Maße miteinander verglichen wurden. So zeigen eigene Berechnungen anhand der PISA-Bundesländerergebnisse, dass *jedes einzelne* der sieben Flächenländer mit Zentralprüfungen eine höhere durchschnittliche PISA-Lesekompetenz aufweist als *jedes einzelne* der sechs Flächenländer ohne Zentralprüfungen, wenn man für die Arbeitslosigkeit als Maß des sozialen und ökonomischen Hintergrunds kontrolliert.

Schülerleistungen ausüben. Zur Kontrolle für die allgemeine Zentralisierung des Systems und für die Bevölkerungshomogenität wurden zusätzlich der von der Zentralregierung kontrollierte Anteil an den Bildungsausgaben sowie ein Maß der ethnolinguistischen Fraktionalisierung der Bevölkerung als Kontrollvariablen in die Spezifikation eingeführt, ohne dass sich der geschätzte Zentralprüfungseffekt nennenswert geändert hätte.

Um schließlich zu testen, ob der Zentralprüfungseffekt sonstige kulturelle Unterschiede zwischen Ländern auffängt, können regionale (kontinentale) Dummies als zusätzliche Kontrollvariablen eingefügt werden. Dadurch würde die Schätzung des Zentralprüfungseffektes ausschließlich aufgrund von inner-regionaler Variation durchgeführt. Inter-regionale kulturelle Unterschiede wie etwa zwischen asiatischen und europäischen Wertesystemen würden den Zentralprüfungsschätzer bei einer solchen Spezifikation nicht mehr beeinflussen. Wie in der letzten Zeile von Tabelle 8 berichtet, ergeben die Schätzungen sogar dann statistisch signifikante Zentralprüfungseffekte, wenn die Variation zwischen den neun Regionen Westeuropa, Osteuropa, Nordamerika, Südamerika, Ozeanien, Asien, Mittlerer Osten, Nordafrika und Südafrika unberücksichtigt bleibt. Der geschätzte Zentralprüfungseffekt scheint also weder durch andere institutionelle Unterschiede, noch durch die allgemeine Zentralisation oder Homogenität eines Landes, noch durch kulturelle Unterschiede bedingt zu sein, sondern einen Effekt der externen Prüfungen auf die Schülerleistungen widerzugeben.

Detailliertere Schätzungen zeigen, dass der Effekt von Zentralprüfungen für Schüler mit unterschiedlichem familiärem Hintergrund unterschiedlich groß ist (Wößmann 2002b). So ist der Effekt des Ausbildungsstandes der Eltern auf die Leistungen ihrer Kinder in Systemen ohne Zentralprüfungen weit größer als in Systemen mit Zentralprüfungen. Dieser Unterschied zwischen den beiden Systemen ist, vor allem in Mathematik, statistisch hoch signifikant. Ebenso ist der Leistungsrückstand von Immigrantenkinder der ersten und zweiten Generation in Systemen mit Zentralprüfungen weit geringer als in Systemen ohne Zentralprüfungen. Diese Befunde legen nahe, dass zentrale Prüfungen zu ausgeglicheneren Bildungschancen für Schüler mit unterschiedlichem familiärem Hintergrund beitragen.

Aus weiteren empirischen Analysen wird deutlich, dass sich durch externe Prüfungen nicht nur die Anreize und damit das Verhalten der Schüler ändern, sondern dass sich auch Anreize und Verhalten weiterer Akteure des Schulsystems ändern (Wößmann 2002b). Als Beispiel sei hier auf die Interaktion von Zentralprüfungen und Schulautonomie eingegangen. Wie Schaubild 2a zeigt, hat es in Systemen ohne Zentralprüfungen einen negativen Effekt auf die Schülerleistungen, wenn die Schule Autonomie über das Schulbudget hat. Demgegenüber gibt es in Systemen mit Zentralprüfungen keine statistisch signifikanten Leistungsunterschiede zwischen Schulen mit und ohne Budgetautonomie. In Systemen ohne Zentralprüfungen hat es ebenfalls negative Auswirkungen für die erzielten Schülerleistungen, wenn Schulen Eigenverantwortung über die Gehälter der Lehrer haben (Schaubild 2b). In Systemen mit Zentralprüfungen dreht sich dieser Effekt komplett um: Die Gehaltsautonomie der Schulen hat in Zentralprüfungssystemen positive Auswirkungen auf die Schülerleistungen. Der Unterschied in den Auswirkungen der Schulautonomie zwischen den beiden Prüfungssystemen ist in beiden Fällen statistisch signifikant.

Aus ökonomischer Sicht lässt sich der diametral verschiedene Einfluss von finanzieller Schulautonomie in Systemen mit und ohne Zentralprüfungen dadurch erklären, dass die Anreizbedingungen für Schulen zwischen den beiden Systemen differieren. In Systemen ohne Zentralprüfungen kann opportunistisches Verhalten der Schulen für gewöhnlich nicht festgestellt werden und wird daher nicht sanktioniert. Demgegenüber schaffen Zentralprüfungen Informationen darüber, ob die Schulen gute Leistungsergebnisse erzielen oder nicht, so dass Aufsichtsbehörden und gegebenenfalls Eltern mögliche Konsequenzen aus leistungsminderndem Verhalten der Schulen ziehen können. Dies schafft Anreize für die Entscheidungsträger in den Schulen, finanzielle Autonomie nicht opportunistisch auszunutzen, sondern effektiv zur Leistungsförderung der Schüler einzusetzen. Somit dürften die unterschiedlichen Effekte der Schulautonomie in Systemen mit und ohne Zentralprüfungen Verhaltensänderungen der schulischen Entscheidungsträger widerspiegeln.

Insofern zeigen die Ergebnisse auch, dass zentrale Prüfungen geradezu eine Voraussetzung für ein gutes Funktionieren von ansonsten eher dezentral organisierten Schulsystemen sind, da sie leistungsfördernde Anreizwirkungen und damit Verhaltensänderungen aller Beteiligten hervorrufen. Damit eine Dezentralisierung des Schulsystems, wie sie vielfach gefordert wird (vgl. hierzu Weiß 1998), zu positiven Leistungseffekten führt, muss durch zentrale Prüfungen sichergestellt sein, dass die dezentralen Entscheidungsträger Anreize zu leistungsförderndem Verhalten haben.

Schließlich sei noch betont, dass es sich bei dem hier verwendeten Leistungsmaßstab, den TIMSS-Testergebnissen, nicht um die jeweilige zentrale Prüfung eines Landes handelt, sondern um einen unabhängigen Leistungstest, der von allen beteiligten Ländern als dem jeweiligen Mathematik- und Naturwissenschaftscurriculum entsprechend akzeptiert wurde. Das bedeutet, dass Zentralprüfungen nicht einfach nur dazu führen, dass die Lehrer ihre Schüler lediglich mögliche Fragen der jeweiligen Zentralprüfung auswendig lernen lassen und dass die Schüler den entsprechenden Test „einpauken“ – denn dies sollte ihre Leistungen in den TIMSS-Tests unberührt lassen. Statt dessen legen die hier geschätzten Effekte nahe, dass Zentralprüfungen tatsächlich zu einem besseren mathematischem und naturwissenschaftlichem Wissen der Schüler führen.

Zentralprüfungen sind nicht die einzige Bildungsinstitution, die die Schülerleistungen beeinflusst. Zahlreiche weitere Einflüsse der institutionellen Struktur des Schulsystems lassen sich im internationalen Vergleich aufzeigen (Wößmann 2003b). Auch innerhalb einzelner Länder sind in letzter Zeit verschiedene kausale Effekte institutioneller Anreizmechanismen im Schulsystem nachgewiesen worden. So hat etwa Rouse (1998) gezeigt, dass Bildungsgutscheine zum Besuch privater Schulen in Wisconsin zu besseren Schülerleistungen geführt haben, und Peterson et al. (2003) finden ähnliche Ergebnisse für farbige Schüler in Schulexperimenten in Washington, D.C., New York City und Dayton/Ohio. Hoxby (2000b) weist in den Vereinigten Staaten positive Leistungseffekte durch erhöhten Wettbewerb zwischen öffentlichen Schulen nach. Lavy (2002a, b) zeigt, dass monetäre Anreize für Schulen und insbesondere für Lehrer, die

auf bei den Schülern erzielten Leistungssteigerungen basieren, in Israel zu signifikant besseren Leistungen der Schüler geführt haben.

Die empirische Evidenz legt also nahe, dass die Schulpolitik weniger auf eine Ausweitung der Ressourcenausstattung im gegebenen System als vielmehr auf eine leistungsfördernde Veränderung der institutionellen Struktur des Schulsystems setzen sollte. Angesichts der makroökonomischen Relevanz der Bildungsqualität für Wachstum und Entwicklung scheint eine solche mikroökonomisch fundierte bildungspolitische Schwerpunktsetzung dringend erforderlich.

Literatur

- Acemoglu, Daron, Simon Johnson, James A. Robinson (2001). The Colonial Origins of Comparative Development: An Empirical Investigation. *American Economic Review* 91 (5): 1369-1401.
- Angrist, Joshua D., Victor Lavy (1999). Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement. *Quarterly Journal of Economics* 114 (2): 533-575.
- Barro, Robert J. (1999). Human Capital and Growth in Cross-Country Regressions. *Swedish Economic Policy Review* 6 (2): 237-277.
- Barro, Robert J., Jong-Wha Lee (1996). International Measures of Schooling Years and Schooling Quality. *American Economic Review* 86 (2): 218-223.
- Barro, Robert J., Jong-Wha Lee (2001). International Data on Educational Attainment: Updates and Implications. *Oxford Economic Papers* 53 (3): 541-563.
- Bishop, John H. (1997). The Effect of National Standards and Curriculum-Based Exams on Achievement. *American Economic Review* 87 (2): 260-264.
- Bishop, John H., Ludger Wößmann (2003). Institutional Effects in a Simple Model of Educational Production. Erscheint in: *Education Economics*.
- Bloom, David E., Jeffrey D. Sachs (1998). Geography, Demography, and Economic Growth in Africa. *Brookings Papers on Economic Activity* (2): 207-295.
- Case, Anne, Angus Deaton (1999). School Inputs and Educational Outcomes in South Africa. *Quarterly Journal of Economics* 114 (3): 1047-1084.
- Deaton, Angus (1997). *The Analysis of Household Surveys: A Microeconometric Approach to Development Policy*. Baltimore: The Johns Hopkins University Press.
- Deininger, Klaus, Lyn Squire (1996). A New Data Set Measuring Income Inequality. *World Bank Economic Review* 10 (3): 565-591.
- Dollar, David, Aart Kraay (2002). Growth Is Good for the Poor. *Journal of Economic Growth* 7 (3): 195-225.
- DuMouchel, William H., Greg J. Duncan (1983). Using Sample Survey Weights in Multiple Regression Analyses of Stratified Samples. *Journal of the American Statistical Association* 78 (383): 535-543.
- Gollin, Douglas (2002). Getting Income Shares Right. *Journal of Political Economy* 110 (2): 458-474.
- Gundlach, Erich, Joana Henseler, Vivien Knips (in Vorbereitung). Macroeconomic Mincering: The Social Rate of Return to Schooling. Institut für Weltwirtschaft, Kiel (mimeo).
- Gundlach, Erich, Jose Navarro de Pablo, Natascha Weisert (2004). Education Is Good for the Poor: A Note on Dollar and Kraay. In: Anthony F. Shorrocks, Rolph van der Hoeven (eds.), *Growth, Inequality and Poverty*. Oxford University Press (im Druck).
- Gundlach, Erich, Desmond Rudman, Ludger Wößmann (2002). Second Thoughts on Development Accounting. *Applied Economics* 34 (11): 1359-1369.

- Gundlach, Erich, Ludger Wößmann, Jens Gmelin (2001). The Decline of Schooling Productivity in OECD Countries. *Economic Journal* 111 (471): C135-C147.
- Hall, Robert E., Charles I. Jones (1999). Why Do Some Countries Produce So Much More Output per Worker than Others? *Quarterly Journal of Economics* 114 (1): 83-116.
- Hanushek, Eric A. (1999). Some Findings from an Independent Investigation of the Tennessee STAR Experiment and from Other Investigations of Class Size Effects. *Educational Evaluation and Policy Analysis* 21 (2): 143-163.
- Hanushek, Eric A. (2002). Publicly Provided Education. In: Alan Auerbach, Martin Feldstein (Hrsg.), *Handbook of Public Economics, Volume 4*: 2045-2141. Amsterdam: Elsevier.
- Hanushek, Eric A. (2003). The Failure of Input-Based Schooling Policies. *Economic Journal* 113 (485): F64-F98.
- Hanushek, Eric A., Dennis Kimko (2000). Schooling, Labor Force Quality, and the Growth of Nations. *American Economic Review* 90 (5): 1184-1208.
- Heckman, James J., Peter J. Klenow. (1997). Human Capital Policy. University of Chicago (mimeo).
- Hoxby, Caroline M. (2000a). The Effects of Class Size on Student Achievement: New Evidence from Population Variation. *Quarterly Journal of Economics* 115 (4): 1239-1285.
- Hoxby, Caroline M. (2000b). Does Competition among Public Schools Benefit Students and Taxpayers? *American Economic Review* 90 (5): 1209-1238.
- International Association for the Evaluation of Educational Achievement (IEA) (1998). *Third International Mathematics and Science Study. International Achievement Reports*. <http://timss.bc.edu/timss1995i/MathB.html>.
- Jürges, Hendrik, Kerstin Schneider, Felix Büchel (2002). The Effect of Central Exit Examinations on Student Achievement: Quasi-Experimental Evidence from TIMSS Germany. Mimeo.
- Klenow, Peter J., Andrés Rodríguez-Clare (1997). The Neoclassical Revival in Growth Economics: Has it Gone Too Far?. *NBER Macroeconomics Annual* 12: 73-102.
- Krueger, Alan B. (1999). Experimental Estimates of Education Production Functions. *Quarterly Journal of Economics* 114 (2): 497-532.
- Krueger, Alan B. (2003). Economic Considerations and Class Size. *Economic Journal* 113 (485): F34-F63.
- Krueger, Alan B., Diane M. Whitmore (2001). The Effect of Attending a Small Class in the Early Grades on College-Test Taking and Middle School Test Results: Evidence from Project STAR. *Economic Journal* 111 (468): 1-28.
- Lavy, Victor (2002a). Evaluating the Effect of Teachers' Group Performance Incentives on Pupil Achievement. *Journal of Political Economy* 110 (6): 1286-1317.

- Lavy, Victor (2002b). Paying for Performance: The Effect of Teachers' Financial Incentives on Students' Scholastic Outcomes. Hebrew University of Jerusalem (mimeo).
- Lazear, Edward P. (2001). Educational Production. *Quarterly Journal of Economics* 116 (3): 777–803.
- Levine, Ross, David Renelt (1992). A Sensivity Analysis of Cross-Country Growth Regressions. *American Economic Review* 82 (4): 942-963.
- McArthur, John W., Jeffrey D. Sachs (2001). Institutions and Geography: A Comment on Acemoglu, Johnson, and Robinson (2000). NBER Working Paper 8114, February.
- Mincer, Jacob (1974). *Schooling, Experience, and Earnings*. National Bureau of Economic Research. New York.
- Moulton, Brent R. (1986). Random Group Effects and the Precision of Regression Estimates. *Journal of Econometrics* 32 (3): 385-397.
- OECD (2001). Knowledge and Skills for Life: First Results from PISA 2000. Paris.
- Penn World Table (PWT) (1994). Version 5.6. NBER, Cambridge, MA. <http://www.nber.org/pwt56.html>.
- Peterson, Paul E., William G. Howell, Patrick J. Wolf, David E. Campbell (2003). School Vouchers: Results from Randomized Experiments. Erscheint in: Caroline M. Hoxby (Hrsg.), *The Economics of School Choice*. Chicago: University of Chicago Press.
- Prescott, Edward C. (1998). Needed: A Theory of Total Factor Productivity. *International Economic Review* 39 (3): 525–552.
- Psacharopoulos, George, Harry A. Patrinos (2002). Returns to Investment in Education: A Further Update. World Bank, Policy Research Working Paper, 2881, September.
- Psacharopoulos, George (1994). Returns to Investment in Education: A Global Update. *World Development* 22: 1325-1343.
- Rodriguez, Francisco, Jeffrey D. Sachs (1999). Why Do Resource-Abundant Economies Grow More Slowly? *Journal of Economic Growth* 4 (3): 277-303.
- Rosenzweig, Mark R., Kenneth I. Wolpin (2000). Natural “Natural Experiments” in Economics. *Journal of Economic Literature* 38 (4): 827–874.
- Rouse, Cecilia Elena (1998). Private School Vouchers and Student Achievement: An Evaluation of the Milwaukee Parental Choice Program. *Quarterly Journal of Economics* 113 (2): 553-602.
- United Nations (UN) (1994). *The Sex and Age Distributions of the World Populations*. New York: United Nations.
- Weiß, Manfred (1998). Schulautonomie im Licht mikroökonomischer Bildungsforschung. In: Robert K. von Weizsäcker (Hrsg.), *Deregulierung und Finanzierung des Bildungswesens*: 15-47. Schriften des Vereins für Socialpolitik, Band 262. Berlin: Duncker & Humblot.

- West, Martin R., Ludger Wößmann (2003). Which School Systems Sort Weaker Students into Smaller Classes? International Evidence. Kiel Working Paper 1145. Kiel: Institute for World Economics.
- Wooldridge, Jeffrey M. (2001). Asymptotic Properties of Weighted *M*-Estimators for Standard Stratified Samples. *Econometric Theory* 17 (2): 451-470.
- World Bank (1999). *World Development Indicators*. CD-ROM.
- Wößmann, Ludger (2002a). *Schooling and the Quality of Human Capital*. Berlin: Springer.
- Wößmann, Ludger (2002b). *Central Exams Improve Educational Performance: International Evidence*. Kieler Diskussionsbeiträge 397. Kiel: Institut für Weltwirtschaft.
- Wößmann, Ludger (2003a). Specifying Human Capital: A Review and Some Extensions. Erscheint in: *Journal of Economic Surveys* 17 (Juni).
- Wößmann, Ludger (2003b). Schooling Resources, Educational Institutions, and Student Performance: The International Evidence. *Oxford Bulletin of Economics and Statistics* 65 (2): 117-170.
- Wößmann, Ludger (2003c). European Education Production Functions. Institut für Weltwirtschaft, Kiel (mimeo).
- Wößmann Ludger, Martin R. West (2002). Class-Size Effects in School Systems Around the World: Evidence from Between-Grade Variation in TIMSS. Harvard University, Program on Education Policy and Governance Research Paper PEPG/02-02.

Tabelle 1: Humankapital als Komponente der Arbeitsproduktivität^a

	Anteiliger Beitrag der Faktoren ...		
	Humankapital	Sachkapital	Totale Faktorproduktivität
Klenow und Rodriguez-Clare (1997) ^b	0,11	0,23	0,66
Hall und Jones (1999) ^c	0,22	0,19	0,59
Länderspezifische soziale Ertragsraten der Bildung	0,20	0,19	0,60
Durchschnittliche Ertragsraten der Bildung und länderspezifische Bildungsqualität	0,46	0,19	0,35

^a Internationale Querschnittsanalyse, jeweilige Kovarianzanteile an der Gesamtvarianz der Arbeitsproduktivität; Zeile 1: 98 Länder im Jahr 1985, Zeilen 2-4: 113 Länder im Jahr 1990; zu den Einzelheiten der Berechnung und den Daten vgl. Gundlach et al. (2002). — ^b Siehe die Ergebnisse in Klenow und Rodriguez-Clare (1997), Tabelle 2, unter BK4. — ^c Basierend auf der Humankapital-Spezifikation von Hall und Jones (1999), berechnet mit einem aktualisierten Datensatz.

Quelle: Gundlach et al. (2002).

Tabelle 2: Humankapital als Komponente der Arbeitsproduktivität in OECD Ländern^a

	Anteiliger Beitrag der Faktoren ...		
	Humankapital	Sachkapital	Totale Faktorproduktivität
Hall und Jones (1999) ^b	0,44	0,15	0,41
Länderspezifische soziale Ertragsraten der Bildung	0,52	0,15	0,34
Durchschnittliche Ertragsraten der Bildung und länderspezifische Bildungsqualität	0,98	0,15	-0.13

^a Internationale Querschnittsanalyse, jeweilige Kovarianzanteile an der Gesamtvarianz des Bruttoinlandsprodukts je Person im erwerbsfähigen Alter; 23 OECD-Länder im Jahr 1990; zu den Einzelheiten der Berechnung und den Daten vgl. Gundlach et al. (2002). — ^b Basierend auf der Humankapital-Spezifikation von Hall und Jones (1999), berechnet mit einem aktualisierten Datensatz.

Quelle: Gundlach et al. (2002).

Tabelle 3: Der Effekt des Humankapitals auf das Einkommen der Armen (OLS)

Abhängige Variable: $\ln y_p$

	(1)	(2)	(3)	(4)	(5)
a_0	-0,85 (0,46)	-1,00 (0,06)	-0,88 (0,46)	-0,70 (0,49)	-1,00 (0,65)
$\ln y_c$	0,90 (0,07)	0,90 (0,07)	0,90 (0,07)	0,88 (0,07)	0,91 (0,07)
$\ln(H / L)$	0,32 (0,10)	0,34 (0,11)	0,30 (0,10)	0,32 (0,10)	0,31 (0,11)
$\ln INV$	-	-0,04 (0,09)	-	-	-0,05 (0,10)
MINING	-	-	-0,48 (0,65)	-	-0,56 (0,66)
MALARIA	-	-	-	-0,02 (0,02)	-0,01 (0,02)
Stichprobe	n=101	n=101	n=99	n=91	n=89
\bar{R}^2	0,87	0,87	0,87	0,88	0,88
s.e.e.	0,43	0,44	0,43	0,43	0,42

Standardfehler in Klammern.

Quelle: Gundlach et al. (2004).

Tabelle 4: Der Effekt des Humankapitals auf das Einkommen der Armen (IV)

Abhängige Variable: $\ln y_p - \ln y_c$

Instrument-Variablen	MEANTEMP	DISTANCE	<i>DISTANCE, MEANTEMP</i>
a_0	-1,69 (0,10)	-1,79 (0,12)	-1,69 (0,10)
$\ln(H / L)$	0,34 (0,09)	0,43 (0,11)	0,34 (0,09)
Stichprobe	n=86	n=100	n=86
s.e.e.	0,43	0,47	0,43
Überidentifikation			
p-Wert	-	-	0,63
Testergebnis	-	-	keine Überident.

Standardfehler in Klammern.

Quelle: Gundlach et al. (2004).

Tabelle 5: Die makroökonomische Ertragsrate des Humankapitals (OLS)

Abhängige Variable: $\ln y_c$

	(1)	(2)	(3)	(4)
a_0	6,14 (0,14)	4,46 (0,33)	3,18 (0,40)	2,16 (0,35)
S	0,35 (0,02)	0,31 (0,02)	0,20 (0,03)	0,12 (0,03)
E	-	0,15 (0,03)	0,11 (0,03)	0,09 (0,02)
<i>LIFE</i>	-	-	0,04 (0,01)	0,00 (0,01)
$\ln k$	-	-	-	0,43 (0,06)
Stichprobe	n=80	n=71	n=71	n=71
\bar{R}^2	0,77	0,83	0,87	0,92
s.e.e.	0,52	0,44	0,39	0,30

Standardfehler in Klammern.

Quelle: Gundlach et al. (in Vorbereitung).

Tabelle 6: Alternative Schätzungen der makroökonomischen Ertragsrate des Humankapitals

Abhängige Variable:	$\ln y_c$		$\ln y_c - 0.3 \ln k$	
	(1) OLS	(2) OLS	(3) OLS	(4) IV
a_0	2,44 (0,35)	2,41 (0,36)	2,78 (0,33)	2,81 (0,73)
SQ	0,08 (0,01)	0,07 (0,01)	0,09 (0,01)	0,09 (0,03)
E	0,06 (0,02)	0,07 (0,02)	0,06 (0,02)	0,03 (0,06)
<i>LIFE</i>	0,00 (0,01)	-	0,02 (0,01)	0,02 (0,01)
$\ln k$	0,43 (0,06)	0,46 (0,04)	-	-
DISTANCE	-	-0,02 (0,27)	-	-
Stichprobe	n=72	n=71	n=72	n=71
\bar{R}^2	0,93	0,93	0,80	-
s.e.e.	0,29	0,28	0,30	0,30
Soziale Ertragsrate	0,14 (0,02)	0,14 (0,02)	0,13 (0,02)	0,13 (0,02)
Überidentifikatio	-	-	-	0,22
n	-	-	-	keine Überid.
p-Wert	-	-	-	
Testergebnis				

Standardfehler in Klammern.

Quelle: Gundlach et al. (in Vorbereitung).

Tabelle 7: Der Effekt von Klassengrößen auf Schülerleistungen

Schätzer des Klassengrößen-Koeffizienten. – Abhängige Variable: TIMSS-Testergebnis. –
Kontrollvariablen: Jahrgangsstufe und 12 Schüler- und Familienmerkmale. –
Cluster-robuste Standardfehler (auf Klassenebene) in Klammern.

	Mathematik			Naturwissenschaften		
	KQ	FE	FE-IV	KQ	FE	FE-IV
Australien	4,33* (0,72)	4,30* (0,67)	-2,08 (3,91)	3,65* (0,64)	1,46† (0,69)	-0,70 (9,47)
Belgien (flämisch)	2,18‡ (1,16)	0,82 (1,18)	8,09 (6,67)	1,47‡ (0,86)	0,48 (1,04)	1,08 (1,44)
Belgien (französisch)	1,51‡ (0,84)	-0,53 (0,93)	0,80 (0,98)	-0,58 (0,65)	-1,70† (0,81)	-0,670 (1,10)
Frankreich	2,59* (0,80)	1,60† (0,78)	-2,73† (1,37)	0,55 (0,51)	0,10 (0,53)	0,14 (0,87)
Griechenland	0,46 (0,43)	-0,88 (0,66)	-1,53 (0,99)	0,29* (0,09)	0,05 (0,05)	-2,41† (1,26)
Hongkong	5,47* (1,07)	4,06* (0,52)	-5,22 (7,18)	5,58* (1,26)	3,50* (0,74)	-12,98 (12,82)
Island	0,16 (0,51)	-0,44 (0,58)	-2,59* (0,85)	-1,01* (0,35)	-0,47 (0,70)	-1,58 (1,52)
Japan	3,81* (0,82)	-0,29 (0,35)	0,07 (0,44)	2,59* (0,67)	-0,44 (0,31)	-0,26 (0,42)
Kanada	0,76 (0,79)	0,20 (0,43)	0,25 (0,62)	0,09 (0,09)	0,17† (0,08)	-1,23 (1,22)
Portugal	0,77* (0,27)	0,85* (0,22)	1,54† (0,70)	0,17 (0,30)	0,07 (0,30)	-0,31 (0,57)
Rumänien	2,14* (0,57)	0,30 (0,66)	-0,30 (1,71)	1,43‡ (0,83)	0,70 (0,65)	3,31 (2,19)
Schottland	2,52* (0,66)	2,92* (0,92)	-4,98 (6,32)	-0,66† (0,32)	-0,63† (0,27)	31,58 (51,88)
Singapur	4,69* (0,47)	3,10* (0,41)	0,45 (0,51)	5,03* (0,48)	3,47* (0,45)	0,52 (0,52)
Slowenien	0,52 (0,63)	-0,05 (0,74)	1,25 (1,46)	-0,39 (0,60)	0,13 (0,55)	0,29 (0,87)
Spanien	0,17 (0,17)	-0,10 (0,19)	-0,31 (0,85)	0,19 (0,17)	0,10 (0,17)	-0,70* (0,27)
Südkorea	-0,15† (0,08)	-0,21* (0,04)	-0,90 (0,57)	0,19 (0,11)	0,07 (0,10)	-0,42 (1,35)
Tschechische Republik	2,37† (1,19)	-1,10 (1,74)	2,67 (2,25)	1,43† (0,71)	-1,82† (0,86)	-1,03 (0,92)
Vereinigte Staaten	-0,16 (0,11)	0,00 (0,12)	20,26 (69,60)	0,04 (0,17)	0,15 (0,15)	-1,28 (2,76)

KQ: Kleinstquadrate-Schätzung. – FE: Fixed-Effects-Schätzung. – FE-IV: Fixed-Effects-Instrument-Variablen-Schätzung. – Signifikanzniveau (basierend auf Cluster-robusten Standardfehlern): * 1 Prozent; † 5 Prozent; ‡ 10 Prozent. – FE-IV-Schätzer statistisch signifikant (10 Prozent) unterschiedlich von: ○ –3; ● –1.

Quelle: Wößmann und West (2002).

Tabelle 8: Der Effekt von Zentralprüfungen auf Schülerleistungen

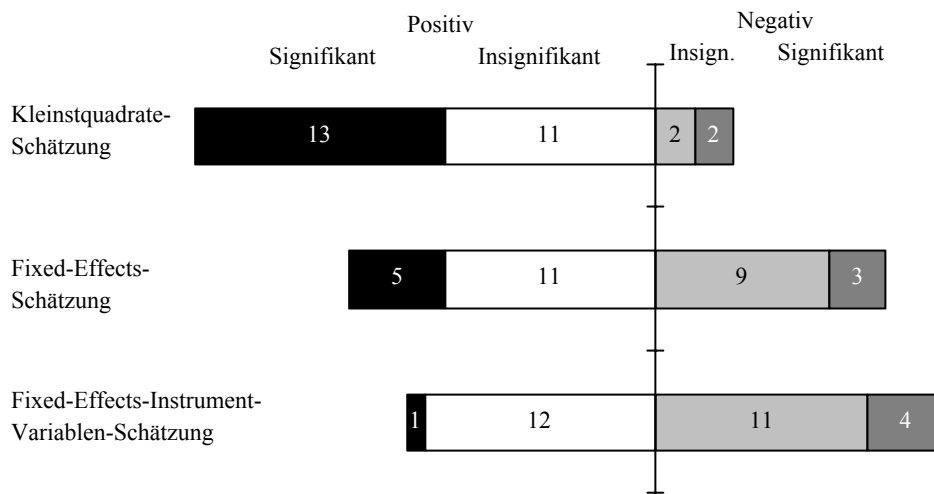
Schätzer des Zentralprüfungs-Koeffizienten. – Abhängige Variable: TIMSS-Testergebnis. – Kontrollvariablen: 48 Schüler-, Familien-, Ressourcen-, Lehrer- und Institutionenmerkmale. – Cluster-robuste Standardfehler (auf Länderebene) in Klammern.

	Mathematik	Naturwissenschaften	Schüler	Länder
TIMSS-95	40,9* (13,5)	39,7* (9,9)	266.545	39
TIMSS-Repeat	47,0* (13,5)	35,9* (12,9)	180.544	38
Beide	42,7* (9,8)	35,9* (8,3)	447.089	77
Beide, mit regionalen Dummies	28,6† (13,2)	41,7* (10,8)	447.089	77

Signifikanzniveau (basierend auf Cluster-robusten Standardfehlern): * 1 Prozent; † 5 Prozent.

Quelle: Wößmann (2002b).

Schaubild 1: Der Klassengrößen-Koeffizient^a

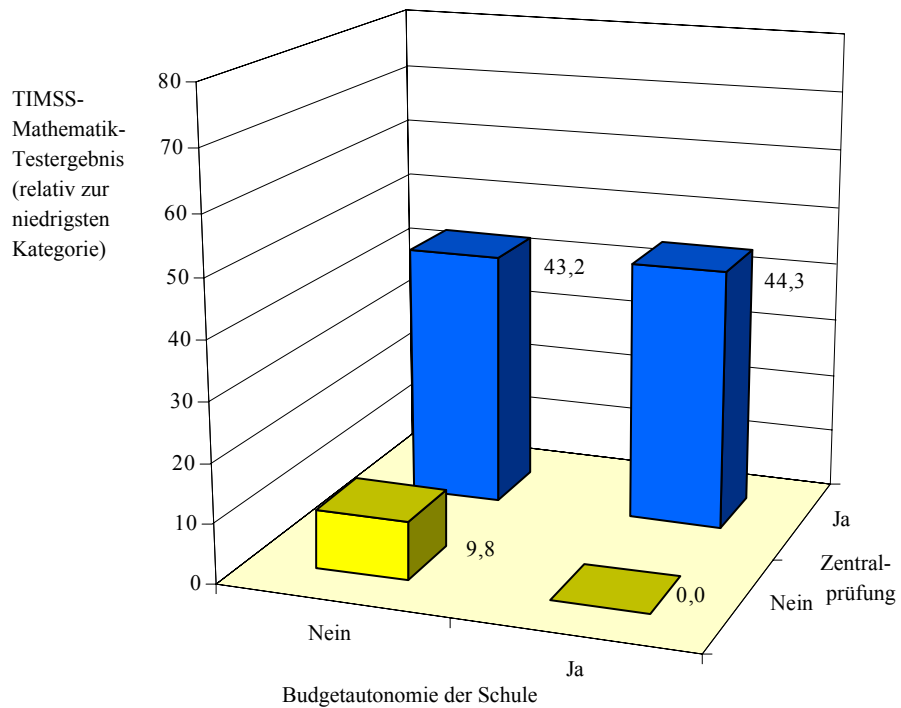


^a Anzahl statistisch signifikanter positiver (schwarz), statistisch insignifikanter positiver (weiß), statistisch insignifikanter negativer (hellgrau) sowie statistisch signifikanter negativer (dunkelgrau) Schätzer.

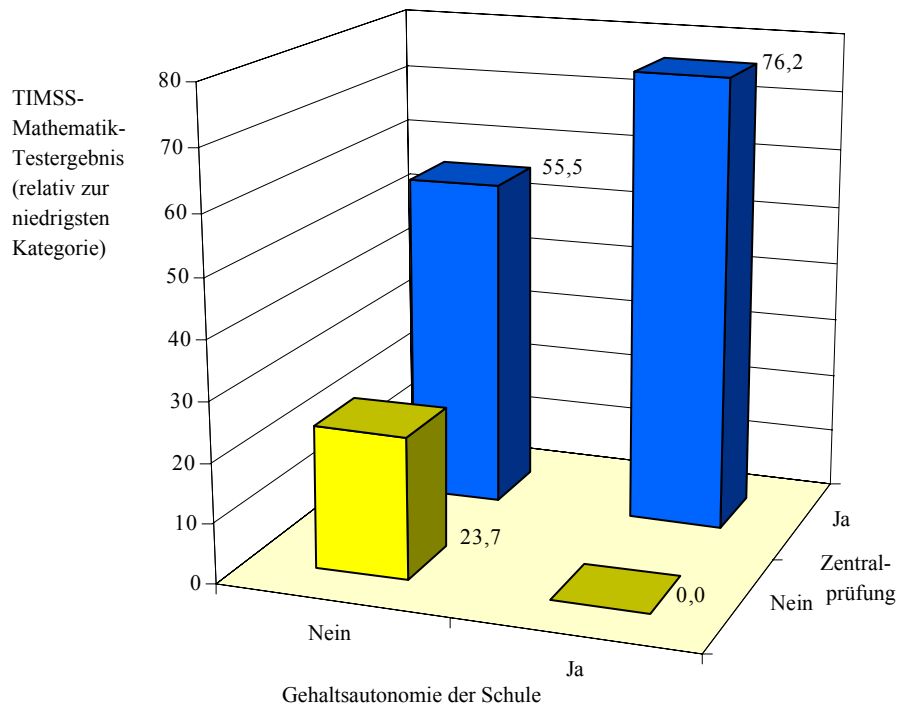
Quelle: Wößmann und West (2002).

Schaubild 2: Zentralprüfungen, Schulautonomie und Schülerleistungen

(a) Budgetautonomie



(b) Gehaltsautonomie



Quelle: Wößmann (2002b).