

Wie jedes Jahr machen sich Eltern zu Beginn des neuen Schuljahres Gedanken darüber, worauf sie achten müssen, damit ihre Kinder eine qualitativ hochwertige Schulbildung erhalten. Sollten sie dafür plädieren, dass zusätzliche Lehrer eingestellt werden? Oder muss zu Hause dringend ein Computer mit der neuesten Lernsoftware angeschafft werden? Brauchen wir bundeseinheitliche Standards? Oder muss schon viel eher, im frühkindlichen Bereich, mit spielendem Lernen angesetzt werden? Was muss an Reformen im deutschen Schulsystem geschehen? Aus ökonomischer Sicht sind dies äußerst berechtigte Fragen über gebotene Prioritäten bildungspolitischen Handelns, denn die bildungsökonomische Forschung hat gezeigt, dass Bildungsleistungen für die Zukunftschancen von Individuen und Volkswirtschaften von entscheidender Bedeutung sind. Und die bildungsökonomische Forschung hat auch einige interessante Antworten auf die gerade gestellten Fragen parat, die zum Teil recht unerwartet sein dürften.

In dieser und den kommenden drei Ausgaben des *ifo Schnelldienstes* berichtet die Artikelserie »Ökonomische Beiträge zur Schuldebatte« über die Befunde zahlreicher Forschungsarbeiten der Abteilung Humankapital und Strukturwandel des ifo Instituts zu vier bildungspolitisch zentralen Themenbereichen: Effekte der Klassengröße, Computereinsatz beim Lernen, institutionelle Reformen zur Effizienzsteigerung im Schulsystem und Bestimmungsgründe der Chancengleichheit in der Bildung. Dazu wird mit Hilfe umfangreicher empirischer Untersuchungsmethoden ermittelt, was man aus den internationalen Schüler Leistungsvergleichen wie PISA, IGLU und TIMSS für die deutsche Schulpolitik lernen kann. Der vorliegende Artikel beginnt mit dem wohl meistdiskutierten Reformvorschlag: Sollen die Landesregierungen zusätzliche Lehrer einstellen, um damit die Klassengrößen in den Schulen zu verkleinern? Aufwändige mikroökonomische Studien deuten darauf hin, dass die schulischen Leistungen dadurch kaum verbessert werden können.

Aus volkswirtschaftlicher Sicht sind kognitive Basiskompetenzen, wie sie etwa in der PISA-Studie getestet wurden, von höchster Relevanz. Anhand von PISA-Vorgängerstudien, die zwischen Mitte der 1960er und Anfang der 1990er Jahre schulische Leistungen im internationalen Vergleich testeten, konnte gezeigt werden, dass Basiskompetenzen in Mathematik und Naturwissenschaften einen starken Einfluss auf das spätere Wachstum der Volkswirtschaften und das erreichte wirtschaftliche Entwicklungsniveau haben.<sup>1</sup> Diese empirische Evidenz belegt die reale Bedeutung neuerer Wachstumstheorien, die die Bedeutung des Humankapitals der Bevölkerung für die langfristige Entwicklung von Volkswirtschaften hervorheben (vgl. etwa Jones 2002). Dass es neben dem ökonomischen auch im allgemeinen öffentlichen Interesse liegt, wie gut unsere Schulen abschneiden, zeigen

die regelmäßig mit den Veröffentlichungen der neuesten PISA-Ergebnisse ausbrechenden Debatten über die wichtigsten Prioritäten der Bildungspolitik.

Mit der Erkenntnis der makroökonomischen Relevanz der Schülerleistungen für die Gesamtwirtschaft wurde verständlicherweise auch das mikroökonomische Interesse an deren individuellen Bestimmungsfaktoren geweckt. So beschäftigt sich die mikroökonomische Forschung zunehmend damit, wie schulische Bildung am effizientesten zur Verfügung gestellt werden kann. Mit der Frage nach der institutionellen Gestaltung des Schulsystems, die die Bildungsbereitstellung am besten gewährleistet, werden sich die letzten beiden Beiträge der im *ifo Schnelldienst* veröffentlichten »Ökonomischen Beiträge zur Schuldebatte« beschäftigen.

In diesem und im nächsten Heft soll es aber zunächst darum gehen, inwiefern eine Ausweitung der Bildungsbudgets zu einer Steigerung der erlernten Basiskompetenzen der Schüler beitragen kann. Der

<sup>1</sup> Vgl. dazu Hanushek und Kimko (2000); Barro (2001); Wößmann (2003a) und Gundlach und Wößmann (2004). Für einen generellen aktuellen Überblick bildungsökonomischer Forschung vgl. Hanushek und Welch (2006).

vorliegende Beitrag befasst sich mit der wohl am meisten vorgebrachten Forderung: Sollten wir nicht mehr Geld einsetzen, um die Klassengrößen zu verkleinern, in denen die Schülerinnen und Schüler unterrichtet werden? In ökonomischer Betrachtungsweise müsste der Nutzen, den eine Verkleinerung der Klassen – zum Beispiel von durchschnittlich 25 auf 19 Schüler – möglicherweise hervorruft, ihren Kosten gegenübergestellt werden. Diese Kosten sind durchaus nicht vernachlässigbar: In diesem Beispiel entsprechen sie einem Anstieg von rund einem Viertel des gesamten Bildungsbudgets! Nur wenn Verkleinerungen der Klassengrößen kosteneffizient wären, wenn ihr Nutzen also diese Kosten überstiege, wäre es wirklich sinnvoll, die Klassen zu verkleinern. In Anbetracht der fundamentalen Bedeutung dieser Fragestellung ist es erstaunlich, dass wir in Deutschland bisher über die Effekte von Klassengrößen auf die Leistungen der Schüler eigentlich so gut wie nichts wissen. Nach einer aggregierten Betrachtung möglicher Effekte zusätzlicher Bildungsausgaben werden im Folgenden konventionelle Schätzungen von Klassengrößeneffekten in Deutschland und deren methodische Probleme sowie zwei quasi-experimentelle Methoden zur Identifikation kausaler Klassengrößeneffekte vorgestellt – die eine anhand natürlicher Kohortenschwankungen, die andere anhand regulierter Diskontinuitäten –, bevor der Beitrag mit einer kurzen Analyse des Zusammenhangs zwischen Lehrerqualität und Klassengrößeneffekten sowie einer Zusammenfassung mit Ausblick schließt.

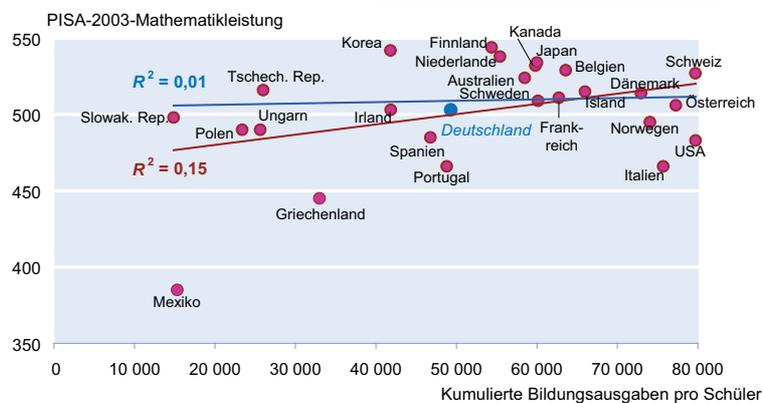
### Das aggregierte Bild: Ausgabenvergleiche zwischen Ländern und über die Zeit

Wir beginnen mit zwei aggregierten Betrachtungen, in denen die Bildungsausgaben und die Schülerleistungen zwischen Ländern und über die Zeit verglichen werden. Geben die Länder, die bei PISA und Co. die Spitzenplätze einnehmen, systematisch mehr für ihr Schulsystem aus als die Länder, die auf den hinteren Plätzen abschneiden? Abbildung 1 stellt den internationalen Zusammenhang zwischen Ausgabenniveaus und Mathematikleistungen in der neuesten PISA-Studie dar. Das Ergebnis ist aus verschiedensten internationalen Vergleichstests lange bekannt: Länder mit höheren Bildungsausgaben schneiden im Ländervergleich *nicht* systematisch besser ab.

Zwar zieht die OECD (2004, 101) aus dieser Abbildung die Schlussfolgerung, dass ein positiver Zusammenhang zwischen Ausgaben und Schülerleistungen besteht, weil sich bei Berücksichtigung aller Länder und ohne Beachtung sonstiger Effekte eine statistisch schwach signifikante positive

Abb. 1

#### Ausgaben pro Schüler und Schülerleistungen im Ländervergleich



Zusammenhang zwischen der durchschnittlichen Mathematikleistung in PISA 2003 und den kumulierten Ausgaben für Bildungsinstitutionen pro Schüler im Alter von 6 bis 15 Jahren, in US-Dollar, konvertiert mit Kaufkraftparitäten.

Quelle: OECD (2004, 102, 358) sowie eigene Berechnungen.

Korrelation ergibt (durch die rote Linie wiedergegeben). Aber schon eine einfache Betrachtung der Abbildung zeigt, wie abwegig ein solcher Schluss ist: Ganz offensichtlich basiert dieser ohnehin sehr schwache positive Zusammenhang einzig und allein darauf, dass Mexiko – und zu einem geringeren Maße Griechenland – sowohl niedrige Ausgaben als auch niedrige Leistungen aufweisen, und es gibt sicherlich noch andere Gründe dafür, dass diese Länder sehr schwach abschneiden. Ohne Berücksichtigung dieser beiden Länder ergibt sich nicht der geringste Zusammenhang zwischen Ausgaben und Leistungen (durch die blaue Linie wiedergegeben). Statistisch gesehen können Ausgabenunterschiede keine Erklärung der internationalen Varianz der Schülerleistungen liefern (das  $R^2$  dieser Regressionsgeraden beträgt 0,01). Im Durchschnitt schneiden Länder mit hohen Bildungsausgaben genauso ab wie Länder mit niedrigen Bildungsausgaben.

Dieses für eine auf Ausgabenerhöhungen fokussierte Bildungspolitik ernüchternde Bild ergab sich genauso schon in vorherigen internationalen Schülererhebungen wie etwa TIMSS (vgl. Wößmann 2002), und eine gleichzeitige Berücksichtigung weiterer Einflussfaktoren verändert an dieser Einschätzung nichts Wesentliches. Ausgaben pro Schüler sind ein umfassendes Maß des Inputs im Bildungssektor, das nicht nur Personalkosten, sondern auch materielle Kosten berücksichtigt. Allerdings sind internationale Ausgabenvergleiche immer mit dem Problem der relevanten Wechselkursumrechnung belastet (hier wurde eine Kaufkraftparitätenumrechnung gewählt), und in nahezu allen Ländern machen die Personalkosten über drei Viertel der gesamten Ausgaben aus. Damit bietet sich auch die Klassengröße als nicht-monetäres Inputmaß, das einen großen Teil der Ausgaben bestimmt, für internationale Vergleiche an. Bei einer Betrachtung der Klassengröße statt des Ausgabenniveaus ergibt sich allerdings das gleiche Bild wie in Abbildung 1.

Ist das Bild anders, wenn wir uns Veränderungen der Ausgaben über die Zeit innerhalb einzelner Länder anschauen? Wie Abbildung 2 darlegt, hat eine dezidierte Studie der zeitlichen Veränderungen von Ausgaben und Schülerleistungen gezeigt, dass sich die realen Bildungsausgaben pro Schüler in allen betrachteten OECD-Ländern zwischen Anfang der 1970er und Mitte der 1990er Jahre substantiell erhöht haben.<sup>2</sup> In Deutschland sind die realen Ausgaben pro Schüler in dem betrachteten 25-Jahres-Zeitraum im Durchschnitt jedes Jahr um 2,9% gestiegen. Über 25 Jahre macht das insgesamt sage und schreibe mehr als eine Verdoppelung der pro Schüler verfügbaren finanziellen Mittel aus! Wie hat sich diese Ausgabenexplosion auf die von den Schülern erlernten Basiskompetenzen ausgewirkt? Zwar ist ein Leistungsvergleich über einen so langen Zeitraum nicht leicht, aber Analysen internationaler Vergleichstests von 1970 und 1994/95 legen nahe, dass sich an den Durchschnittsleistungen der Schüler in allen betrachteten Ländern nichts Wesentliches getan hat. Wenn überhaupt, dann sind die Leistungen deutscher Schüler im Durchschnitt leicht gesunken. Neueste Berechnungen legen aber nahe, dass sich die durchschnittlichen Leistungen deutscher Schüler in Mathematik und Naturwissenschaften von Anfang der 1970er Jahre bis heute so gut wie gar nicht verändert haben (vgl. Hanushek und Wößmann 2005). Aus konstantem Output bei steigendem Input lässt sich auf eine »sinkende Produktivität der Schulen« (vgl. Gundlach und Wößmann 2000) schließen. Jedenfalls haben die immensen Erhöhungen der Ausgaben über die Zeit in Deutschland und anderswo nicht zu einer Verbesserung der Schülerleistungen geführt.

Tatsächlich ist die Ausgabensteigerung in Deutschland wie auch in vielen anderen Ländern zu einem wesentlichen Teil durch eine Verringerung des Schüler-Lehrer-Verhältnisses begründet. Zwischen 1960 und 1995 ist das durchschnittliche Schüler-Lehrer-Verhältnis in allgemeinbildenden Schu-

len in Deutschland nach UNESCO-Daten von 26,7 auf 16,0 gesunken.<sup>3</sup> Auch wenn diese Daten immer mit etwas Vorsicht zu betrachten sind, so zeigen sie doch sicherlich einen klaren langfristigen Trend zu durchschnittlich kleineren Klassen auf, welcher sich nicht in nennenswert besseren Schülerleistungen widerspiegelt hat.

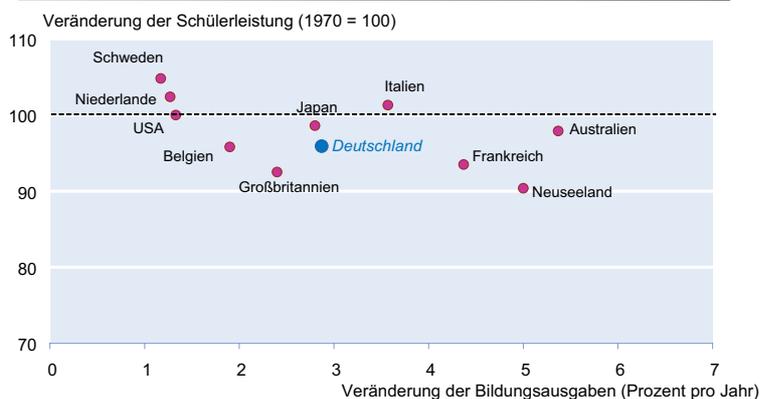
### Konventionelle Schätzung von Klassengrößeneffekten

Bessere Länder geben nicht mehr für ihre Bildung aus, und Ausgabensteigerungen haben nicht zu Leistungsverbesserungen geführt. Aber ist es trotz dieses aggregierten Bildes nicht so, dass Schüler, die in kleineren Klassen unterrichtet werden, mehr lernen? Seltsamerweise ist dies für Deutschland bisher kaum untersucht worden. In den USA gibt es dazu seit den 1960er Jahren Hunderte von empirischen Schätzungen, die den Zusammenhang zwischen Testleistungen von Schülern und verschiedenen Ausstattungsmerkmalen, insbesondere der Klassengröße, schätzen (vgl. Hanushek 2002 für einen Überblick). In der absoluten Mehrzahl kommen diese Studien zum gleichen Befund wie das aggregierte Bild: Sie finden kaum Evidenz für signifikante Leistungseffekte einer besseren Ressourcenausstattung, insbesondere nicht kleinerer Schulklassen.

Um vergleichbare Evidenz für Deutschland zu erhalten, bietet es sich an, die Daten des TIMSS-Tests (»Third International Mathematics and Science Study«) zu nutzen, der 1995 von der International Association for the Evaluation of Educational Achievement (IEA) durchgeführt wurde. Im Vergleich zur PISA-Studie, die in jeder Klasse nur einzelne Schüler getestet hat, hat TIMSS den Vorteil, dass in jeder Schule eine komplette siebte und eine komplette achte Klasse getestet wurde. Diese Datenstruktur mit zwei getesteten Jahrgangsstufen ermöglicht auch die Durchführung einer speziellen Schätzmethode, wie weiter unten näher erläutert wird. Der hier verwendete deutsche TIMSS-Datensatz umfasst 5 620 Schüler in 130 Schulen, die eine repräsentative Stichprobe der deutschen Siebt- und Achtklässler darstellen. Die durchschnittliche Klassengröße in Deutschland liegt in TIMSS bei 24,6 Schülern. Die kleinste getestete Klasse hatte 10, die größte 34 Schüler. Im Durchschnitt erreichten die deutschen Schüler im TIMSS-Mathematiktest 497 Punkte (was sehr nah am internationalen Durchschnitt von 500 liegt), mit einer Standardabweichung von 88 Punkten.

Abb. 2

#### Ausgaben pro Schüler und Schülerleistungen über die Zeit



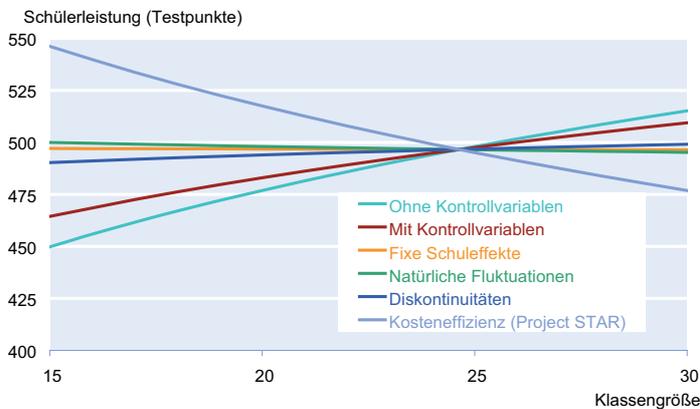
Zusammenhang zwischen der Veränderung der Schülerleistung (Schülerleistung 1994/95 relativ zu 1970) und der Veränderung der Bildungsausgaben (durchschnittliche jährliche Veränderungsrate der realen Bildungsausgaben pro Schüler zwischen 1970 und 1994/95 in Prozent).

Quelle: Wößmann (2002, 106).

<sup>2</sup> Für Details der Analyse vgl. Gundlach und Wößmann (2000); Gundlach et al. (2001) sowie Wößmann (2002).

<sup>3</sup> Leider liegen keine vergleichbaren Daten vor, um diese Zeitreihe ohne Strukturbruch bis heute fortzuführen.

Abb. 3

**Klassengröße und Schülerleistungen in Deutschland**

Geschätzter Zusammenhang zwischen den Schülerleistungen im TIMSS-Mathematiktest und der spezifischen Klassengröße im Mathematikunterricht.

Quelle: Wößmann (2005) und eigene Berechnungen anhand des TIMSS-Datensatzes.

Für eine Abschätzung der Größe der geschätzten Effekte bietet es sich an, diese mit dem durchschnittlichen Leistungsunterschied zwischen der siebten und achten Klasse zu vergleichen, welcher in Deutschland 25 Punkte beträgt. Dieses »Jahrgangsstufenäquivalent« gibt an, wie viel die Schüler im Durchschnitt in einem Schuljahr lernen.

Betrachtet man den einfachen Zusammenhang zwischen der jeweiligen Klassengröße und den Mathematikleistungen der Schüler in Deutschland, so ist dieser erstaunlicherweise positiv. Das heißt, dass bei einfacher Betrachtung die Schülerleistungen in größeren Klassen nicht schlechter, sondern sogar *besser* sind! In Abbildung 3 ist dieser Zusammenhang durch die türkisfarbene Linie dargestellt: Mit steigender Klassengröße nimmt die TIMSS-Testpunktzahl zu.

Eine solche einfache Betrachtungsweise ignoriert natürlich, dass die Schülerleistungen sehr stark vom familiären Hintergrund und möglicherweise auch von der weiteren Ausstattung der Schulen beeinflusst werden. Wenn z.B. Kinder aus hochgebildeten Elternhäusern in der Mittelstufe im Durchschnitt auf Schulen mit größeren Klassen gehen, so beeinflusst dies den einfachen Zusammenhang zwischen Klassengröße und Schülerleistung. Deshalb ist es bei konventionellen Schätzungen von Klassengrößeneffekten üblich, Einflüsse des familiären Hintergrunds mit entsprechenden ökonomischen Methoden herauszurechnen (vgl. Hanushek 2002).

Schätzen wir auf diese konventionelle Art den Klassengrößeneffekt anhand der deutschen TIMSS-Mikrodaten, indem wir für jeden Schüler individuell die Einflüsse zahlreicher persönlicher, familiärer und schulischer Merkmale konstant halten, so verringert sich zwar die Größe des positiven Zusammenhanges zwischen Klassengröße und Schülerleistung, aber er bleibt immer noch statistisch signifikant positiv, wie die rote Linie in Abbildung 3 zeigt.<sup>4,5</sup>

**Das methodische Problem: Endogenität**

Woher kommt dieser seltsame Befund, dass die Schülerleistungen in größeren Klassen im Durchschnitt besser sind? Es liegt nahe, darin keinen kausalen Effekt der Klassengröße auf die Leistungen zu sehen. Stattdessen ist zu berücksichtigen, dass die Klassengröße nicht nur Ursache, sondern auch Folge der Schülerleistungen sein kann. Eine ganze Reihe von Wirkungsmechanismen kommt in Frage, die dazu führen, dass Klassengröße und Schülerleistungen gemeinsam bestimmt werden, was zu einem signifikanten Zusammenhang zwischen den beiden führen kann, ohne dass ein kausaler Effekt der Klassengröße auf die Leistungen vorläge (vgl. dazu West und Wößmann 2005).

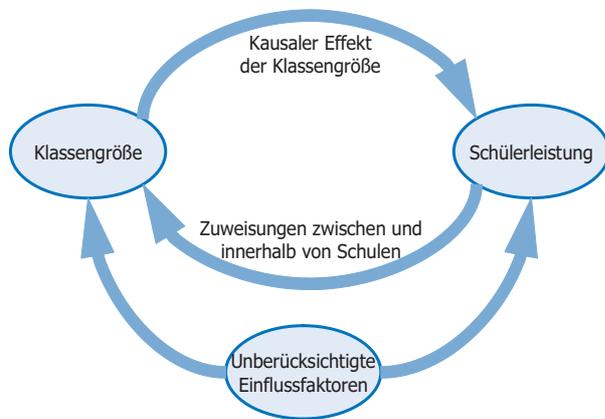
So können zum Beispiel die Aufteilung der Schüler auf Schulen mit unterschiedlichen Klassengrößen aufgrund von Regelungen des Schulsystems sowie die Zuweisung der Schüler in verschiedene Klassen innerhalb einer Schule auf den Leistungen dieser Schüler basieren. Dies ist gerade im mehrgliedrigen Schulsystem in Deutschland der Fall, wo relativ leistungsschwache Schüler Hauptschulen besuchen, die im Durchschnitt etwas kleinere Klassengrößen aufweisen. Dieser Effekt ist in Abbildung 4 durch den Pfeil von den Schülerleistungen hin zur Klassengröße dargestellt, der gewissermaßen eine umgekehrte Kausalität zu dem uns interessierenden kausalen Effekt der Klassengröße auf die Schülerleistungen darstellt.

Auch können die Wahlentscheidungen der Eltern, wo sie wohnen und zu welcher Schule (mit welchen Klassengrößen)

<sup>4</sup> Die in dieser und allen weiteren Schätzungen berücksichtigten Variablen des familiären Hintergrunds, die aus Hintergrundfragebögen der in TIMSS getesteten Schüler stammen, sind: drei Indikatorvariablen für den Bildungsstand der Eltern, vier Indikatorvariablen über die Anzahl der im Haushalt verfügbaren Bücher, Indikatorvariablen darüber, ob der Schüler mit beiden Elternteilen zusammenlebt und ob er in Deutschland geboren wurde, zwei Indikatorvariablen über die geographische Lage der Schule (dörflich bis großstädtisch), das Geschlecht und Alter des Schülers sowie die Jahrgangsstufe. Außerdem kontrolliert die Schätzung für zwei Indikatorvariablen über die materielle Ausstattung der Schule, für die jährliche Unterrichtszeit sowie das Geschlecht und die Lehrerfahrung des Mathematiklehrers. Da der benutzte Datensatz auch zugeschätzte Werte für fehlende Angaben in den Hintergrundfragebögen enthält, kontrollieren die Schätzungen zusätzlich noch für Imputationsindikatoren für jede Variable sowie Interaktionseffekte dieser Indikatoren mit der jeweiligen Variable, was sicherstellt, dass die berichteten Ergebnisse nicht von den Zuschätzungen bestimmt werden. Vgl. Wößmann (2005) für Details.

<sup>5</sup> In ähnlicher methodischer Vorgehensweise schätzt Baumgartner (2003) den Zusammenhang zwischen durchschnittlichen Klassengrößen innerhalb von deutschen Land- und Stadtkreisen und den im Sozioökonomischen Panel erhobenen Einstiegsgehältern unter Kontrolle weiterer Einflussfaktoren und findet keinen statistisch signifikanten Zusammenhang zwischen Klassengröße und Arbeitseinkommen. Allerdings besteht bei einer solchen Betrachtung von Klassengrößenmaßen auf regionaler Ebene die Gefahr von Aggregationsverzerrungen (vgl. Fertig und Wright 2005).

**Abb. 4**  
**Das Endogenitäts- und Simultanitätsproblem**



Ben) sie ihre Kinder schicken wollen, mit den Leistungen der Schüler korreliert sein. In diesem Fall könnten die Wahlentscheidungen der Eltern einen im Modell unberücksichtigten Einflussfaktor darstellen, der gleichzeitig die Klassengröße und die Schülerleistungen beeinflusst (vgl. Abb. 4).

Die Wohnort- und Schulwahlentscheidungen der Eltern, die systematische Aufteilung der Schüler auf Schulen und die Klassenaufteilung innerhalb von Schulen sind Beispiele dafür, wie die Klassengrößen innerhalb des Modells (»endogen«) zusammen (»simultan«) mit den Schülerleistungen bestimmt werden. Dieses Endogenitäts- und Simultanitätsproblem stellt das Hauptproblem der Schätzung des Einflusses von Klassengrößen (oder anderer schulischer Ressourcen) auf die Schülerleistungen, und damit der bisher beschriebenen Evidenz, dar. Wann immer Wahlentscheidungen dazu führen, dass Schüler in nicht zufälliger Weise in Klassen unterschiedlicher Größe unterrichtet werden, ist die beobachtete Variation in Klassengrößen zumindest teilweise verursacht von Variationen in Schülerleistungen beziehungsweise korreliert mit anderen Determinanten der Schülerleistungen. Deshalb kann eine einfache Schätzung von Klassengrößeneffekten mit konventionellen ökonomischen Methoden zu verzerrten Ergebnissen führen (vgl. Hoxby 2000). Eine kausale Interpretation des zuvor berichteten positiven Zusammenhangs zwischen Klassengrößen und Schülerleistungen in Deutschland ist deshalb nicht möglich.

### Identifikation von Klassengrößeneffekten durch natürliche Kohortenschwankungen

Um zu einer kausalen Aussage kommen zu können, muss der kausale Klassengrößeneffekt aus dem gesamten Zusammenhang zwischen Klassengröße und Schülerleistung, der zumindest teilweise auch durch Zuweisungsentscheidungen bedingt ist, herausgefiltert werden. Dazu bedarf es einer empirischen Identifikationsstrategie, die solche Klas-

sengrößenvariationen identifiziert, die nicht durch Endogenitäts- und Simultanitätsprobleme hervorgerufen wurden, die also in Bezug auf die Schülerleistungen »exogen« sind. In der Literatur werden drei Möglichkeiten zur Identifikation von zufälliger oder exogener Variation in Klassengrößen vorgeschlagen. Die erste besteht darin, ein explizites Experiment durchzuführen, das Schüler in zufälliger Weise auf Klassen unterschiedlicher Größe aufteilt. Ein solches explizites Klassengrößenexperiment liegt für Deutschland nicht vor. Das einzige größere bisher überhaupt durchgeführte Klassengrößenexperiment ist das so genannte »Project STAR«, das in den 1980er Jahren im US-Bundesstaat Tennessee durchgeführt wurde und dort auf die Existenz von Klassengrößeneffekten hindeutete (vgl. dazu Krueger 1999). Leider haben explizite Experimente im Allgemeinen (vgl. Hoxby 2000) und Project STAR im Besonderen (Hanushek 1999) einige Probleme und Defizite aufzuweisen, die die Validität und Allgemeingültigkeit der Ergebnisse in Frage stellen.

Für deutsche Evidenz verbleiben also nur die weiteren zwei Identifikationsstrategien, die beide eine quasi-experimentelle Herangehensweise verfolgen, um durch so genannte »natürliche« Experimente hervorgerufene exogene Variationen in Klassengrößen zu identifizieren. In solchen Fällen ist es möglich, durch entsprechende ökonomische Methoden die Schätzung des Klassengrößeneffektes auf einen solchen Teil der gesamten Klassengrößenvariation zu beschränken, der exogen in Bezug auf die Schülerleistungen ist. Im Prinzip vergleichen diese Schätzungen damit Schülerleistungen in Klassen, deren Größe aus exogenen Gründen unterschiedlich ist. Diese Vorgehensweise ermöglicht eine kausale Aussage über die Größe des Klassengrößeneffekts.

Der erste Vorschlag, solche quasi-experimentelle Evidenz zu erlangen, besteht darin, solche Unterschiede in Klassengrößen zu nutzen, die aufgrund von natürlichen Fluktuationen in der Größe aufeinander folgender Jahrgangsstufen einer Schule entstehen (vgl. Hoxby 2000). So werden regelmäßig durch eine »Laune der Natur« im Einzugsbereich einer Schule in einem Schuljahrgang ein paar Kinder mehr geboren als im nächsten – etwa indem im einen Jahr mehr Kinder Ende Juni, im nächsten Jahr mehr Kinder Anfang Juli geboren werden. Diese leichten Schwankungen in der Größe der Jahrgangsstufe kann man als »natürliches« Experiment auffassen: Sie rufen Variationen in der durchschnittlichen Klassengröße in zwei aufeinander folgenden Jahrgangsstufen einer Schule hervor, die weder aufgrund der Leistungen der Schüler noch aufgrund irgendwelcher Zuweisungsentscheidungen des Schulsystems verursacht wurden, sondern eben exogen sind in Bezug auf die Leistungen der Schüler. Damit erlauben natürliche Fluktuationen in der Jahrgangsstufengröße, lediglich einen solchen Teil der Variation der Klassengrößen über die Zeit zu nutzen, der

durch kleine Schwankungen in der Kohortengröße der Geburtenjahrgänge hervorgerufen wurde.

Wößmann und West (2005) haben eine Variante dieser Identifikationsstrategie entwickelt, die spezifische Merkmale des TIMSS-Datensatzes ausnutzt. Die Intuition dieses Ansatzes ist wie folgt. Da zahlreiche Wahlentscheidungen die Allokation von Schülern auf verschiedene Schulen in nicht-zufälliger Weise bestimmen, wird für die Schätzung von Klassengrobeneffekten in einem ersten Schritt die gesamte Variation, die zwischen Schulen besteht, durch die Berücksichtigung eines kompletten Satzes fixer Schuleffekte herausgerechnet. Damit basiert die Schätzung ausschließlich auf Variationen, die innerhalb einzelner Schulen bestehen. Da die Zuweisung auf verschiedene Klassen innerhalb eines Jahrgangs aber auch innerhalb der Schulen in nicht-zufälliger Weise vonstatten gehen kann, werden in einem zweiten Schritt Klassengroßenvariationen innerhalb einzelner Schulen nur insoweit berücksichtigt, als sie durch Variationen in der durchschnittlichen Klassengröße zwischen der siebten und achten Jahrgangsstufe der Schule verursacht werden. Solche Variationen sind nicht durch Zuweisungsentscheidungen beeinflusst, sondern spiegeln in erster Linie eben die zufälligen Unterschiede in der Größe der Geburtenjahrgänge wider, die zu Schwankungen in der durchschnittlichen Klassengröße zwischen zwei Jahrgängen führen.<sup>6</sup>

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 Klassengroßenunterschiede interpretieren. Kurz gefasst identifiziert diese Identifikationsstrategie also kausale Klassengroßeneffekte, indem sie die relativen Leistungen der Schüler zweier Jahrgangsstufen innerhalb einzelner Schulen dem Teil der Klassengroßenunterschiede 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 Daten für mehrere Klassen innerhalb einzelner Schulen zur Berücksichtigung fixer Schuleffekte, sondern auch Daten für Klassen aus mehreren Jahrgangsstufen innerhalb der Schulen, da ansonsten durch die Benutzung der durchschnittlichen Klassengröße der Jahrgangsstufe keine Klassengroßenvariation 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 Jahrgangsstu-

fen beinhalten. All diese Merkmale treffen auf den TIMSS-Datensatz zu.

Die orangefarbene Linie in Abbildung 3 gibt zunächst die Ergebnisse einer Schätzung der Klassengrobeneffekte wieder, die zusätzlich zur konventionellen Schätzmethode nur fixe Schuleffekte berücksichtigt, die also sämtliche Variationen zwischen Schulen unberücksichtigt lässt und nur Variationen innerhalb der Schulen einbezieht (vgl. Wößmann 2005 für Details zu diesen Ergebnissen für Deutschland). Der so geschätzte Klassengrobeneffekt ist nicht von Zuweisungsentscheidungen zwischen Schulen beeinflusst, kann aber noch durch Zuweisungsentscheidungen innerhalb der Schulen verzerrt sein. Wie die orangefarbene Linie zeigt, verschwindet in dieser Betrachtungsweise der positive Zusammenhang zwischen Klassengröße und Schülerleistungen, und es findet sich überhaupt kein nennenswerter Zusammenhang mehr. Dies deutet darauf hin, dass die zuvor beobachtete positive Korrelation vor allem aufgrund von Zuweisungsentscheidungen zwischen Schulen verursacht war.

Um auch noch Verzerrungen durch Zuweisungen innerhalb von Schulen auszuschließen, berichtet die grüne Linie in Abbildung 3 die komplette quasi-experimentelle Identifikationsstrategie, die zusätzlich nur solche Unterschiede zwischen der siebten und achten Klasse einer jeden Schule berücksichtigt, die aufgrund von Unterschieden in der *durchschnittlichen* Klassengröße zwischen diesen beiden Jahrgangsstufen und damit aufgrund der natürlichen Fluktuationen in den Jahrgangsstufenstärken hervorgerufen wurden. Es ergibt sich ein minimaler, statistisch nicht von null zu unterscheidender negativer Zusammenhang zwischen Klassengröße und Schülerleistungen. Weitere Spezifikationstests untermauern die Interpretation, dass die Annahmen dieser Identifikationsstrategie zutreffen, dass sich diese Schätzung also als kausaler Effekt der Klassengröße interpretieren lässt (vgl. Wößmann und West 2005 sowie Wößmann 2005).<sup>7</sup>

Heißt die Tatsache, dass der geschätzte Effekt sich statistisch nicht signifikant von null unterscheidet, dass der Effekt tatsächlich sehr klein ist, oder könnte das Ergebnis auch aufgrund einer wenig präzisen Schätzung zustande kommen? Um die statistische Präzision des geschätzten Effektes wiederzugeben, bildet Abbildung 5 nicht nur den geschätzten Effekt, sondern auch ein Signifikanzband ab. Mit 95%iger Wahrscheinlichkeit liegt der tatsächliche Effekt innerhalb dieses Signifikanzbandes, was üblicherweise der Definition für statistische Signifikanz entspricht. Das grüne Signifikanzband gibt wiederum das Ergebnis der Identifikationsstrategie durch natürliche Fluktuationen wieder. Der geschätzte Effekt – durch das Kreuz wiedergegeben – be-

<sup>6</sup> Für eine detaillierte Diskussion der Validität der Identifikationsstrategie siehe Wößmann und West (2005).

<sup>7</sup> Die in den Abbildungen 3 und 5 wiedergegebenen Ergebnisse der auf natürlichen Fluktuationen und Diskontinuitäten beruhenden Spezifikationen berücksichtigen auch mögliche Effekte der Leistungen der Mitschüler auf die Leistungen der einzelnen Schüler; vgl. Wößmann (2005) für Details.

trägt -6,9, was bedeutet, dass mit jeder Erhöhung der Klassengröße um 10% die durchschnittliche Leistung in der Klasse um 0,69 TIMSS-Testpunkte sinkt. Wie schon die grüne Linie in Abbildung 3 angedeutet hat, ist eine Größenordnung von 0,69 Punkten pro 10%iger Klassengrößenverringerung nach jeglichem Maßstab verschwindend wenig – etwa verglichen mit dem zuvor angesprochenen Jahrgangsstufenäquivalent von 25 Punkten. Die beiden Striche oberhalb und unterhalb des Kreuzes in Abbildung 5 geben nun das 95%ige Signifikanzband an. Wir können uns sehr sicher sein, dass der tatsächliche Effekt innerhalb dieses Bandes liegt – das heißt, dass er weder größer als 75,3 noch kleiner als -89,1 ist.

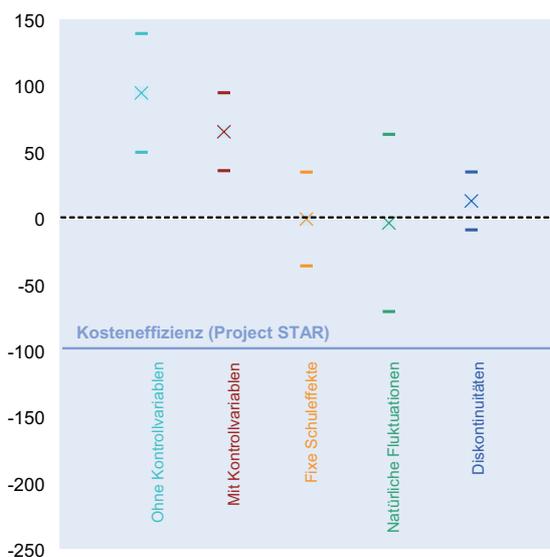
Wäre -89,1 – mit einer 10%igen Verringerung der Klassengröße steigt die Leistung um 8,9 Testpunkte – ein großer Effekt? Um diese Frage beantworten zu können, bietet es sich an, auf die eingangs angesprochene Kosten-Nutzen-Überlegung zurückzukommen. Bei welcher Effektgröße wären die durch Klassenverkleinerungen hervorgerufenen Kosten durch die von den Leistungssteigerungen hervorgerufenen zukünftigen Einkommenssteigerungen aufgewogen? Krueger (1999) hat für das bereits angesprochene Klassengrößenexperiment »Project STAR« eine Effektgröße ermittelt, zu der er eine grobe Kosten-Nutzen-Rechnung aufstellt, die die Kosten in der gleichen Größenordnung sieht wie den Nutzen. Kruegers Effektgröße würde auf die TIMSS-

Skala übertragen etwa einem Wert von -100 entsprechen. Seine Berechnungen sind allerdings recht umstritten (vgl. Carneiro und Heckman 2003), und zwei weitere Überlegungen deuten zusätzlich darauf hin, dass die kosteneffiziente Effektgröße in Deutschland absolut eher noch höher (sprich: weiter im negativen Bereich) liegen dürfte. Zum einen sind die Bildungsrenditen am deutschen Arbeitsmarkt eher geringer als in den USA, was zumindest für die Erträge auf quantitative Maße der Bildung wiederholt gezeigt wurde (vgl. etwa Trostel et al. 2002; Wößmann 2003b). Gleichzeitig sind die Lehrergehälter in Deutschland im internationalen Vergleich relativ hoch.

Nimmt man Kruegers Effektgröße trotzdem als einen ersten Richtwert für mögliche Kosteneffizienz, dann ist immer noch klar, dass der kausale Effekt der Klassengröße auf die Schülerleistungen in Deutschland weit von jeglicher kosteneffizienten Größe entfernt sein dürfte (vgl. die hellblaue Linien in Abb. 3 und 5). Wie Abbildung 5 zeigt, ist nicht nur der geschätzte Effekt weit von einer kosteneffizienten Größenordnung entfernt, sondern die Kosteneffizienz liegt nicht einmal im 95%-Konfidenzband des geschätzten Effekts: Mit weit über 95%iger Sicherheit ist der für Deutschland geschätzte Klassengrößeneffekt (absolut) kleiner als Kruegers womöglich kosteneffiziente Schätzung für das Projekt STAR. Abbildung 5 zeigt auch, dass das Konfidenzband der auf natürlichen Fluktuationen beruhenden Spezifikation größer ist als das der zuvor berichteten Spezifikationen, weil die ökonometrische Methode hohe Ansprüche an die Daten stellt. Die Schätzung, die nur die fixen Schuleffekte berücksichtigt, ist im Vergleich dazu zum Beispiel noch präziser.

Als Ergebnis dieser quasi-experimentellen Untersuchung lässt sich also festhalten, dass in der Mittelstufe in Deutschland *kein nennenswerter kausaler Effekt* der Klassengröße auf die mathematischen Schülerleistungen vorliegt.

Abb. 5  
Größe des Klassengrößeneffekts und Kosteneffizienz



Die Kreuze geben den mit den unterschiedlichen Methoden geschätzten Effekt der (logarithmierten) Klassengröße auf die TIMSS-Mathematikleistungen wieder. Die Striche oberhalb und unterhalb der Kreuze geben das Intervall von plus/minus zwei Standardfehlern an, was etwa einem statistischen Signifikanzband von 95%iger Wahrscheinlichkeit entspricht.

Quelle: Wößmann (2005) und eigene Berechnungen anhand des TIMSS-Datensatzes

### Identifikation von Klassengrößeneffekten durch regelinduzierte Diskontinuitäten

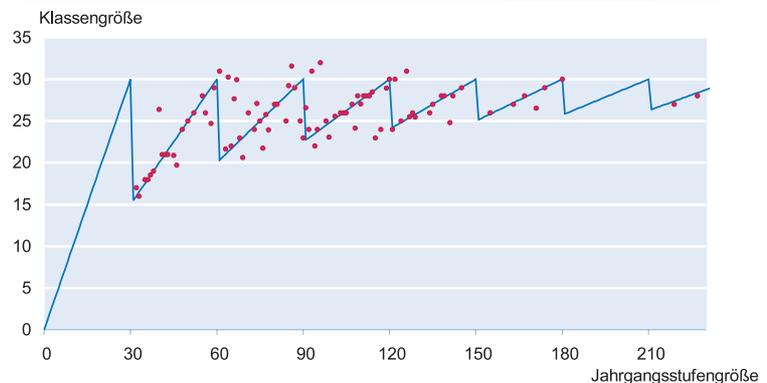
Die Robustheit dieses Befundes lässt sich mit einer zweiten quasi-experimentellen Identifikationsstrategie überprüfen. Diese von Angrist und Lavy (1999) vorgeschlagene Strategie baut darauf auf, dass es vielerorts eine maximale Klassengröße gibt, über die Klassengrößen üblicherweise nicht hinausgehen dürfen. Aufgrund dieser Regel kommt es zu einem nicht-linearen Zusammenhang zwischen der Anzahl der Schüler in einer Jahrgangsstufe und der durchschnittlichen Klassengröße, wie sie für Deutschland in Abbildung 6 abgebildet ist. Der in der TIMSS-Studie zu beobachtende Zusammenhang deutet darauf hin, dass in Deutschland in den meisten Fällen eine explizite oder implizite Regel vorliegt, die eine maximale Klassengröße von 30 Schülern vorsieht. Dies führt zu einem sägenförmigen Zusammenhang zwischen durchschnitt-

licher Klassengröße und Jahrgangsstufengröße: Die Klassengröße steigt mit der Jahrgangsstufengröße bis zu einem Wert von 30. Aber sobald 31 Schüler in der Jahrgangsstufe sind, wird eine zweite Klasse aufgemacht, und die durchschnittliche Klassengröße sinkt abrupt auf  $31/2 = 15,5$  Schüler. Dann wächst sie wieder mit der Jahrgangsstufengröße, bis letztere einen Wert von 60 und die durchschnittliche Klassengröße damit einen Wert von 30 erreicht hat. Bei einer Jahrgangsstufengröße von 61 Schülern wird eine dritte Klasse aufgemacht, so dass die durchschnittliche Klassengröße wieder abrupt abfällt, auf  $61/3 = 20,3$  Schüler, usw. Wie aus Abbildung 6 ersichtlich, trifft dieses Muster für die deutschen siebten und achten Klassen zwar nicht perfekt, aber doch einigermaßen klar erkennbar zu. Neben leichten Berichtsfehlern in den Daten gibt es also wohl auch Ausnahmen, in denen Klassengrößen durchaus einmal über 30 Schüler hinausgehen oder in denen zusätzliche Klassen aufgemacht werden, ohne dass dies aufgrund der 30er-Regel schon notwendig wäre.<sup>8</sup> Aber im Großen und Ganzen trifft dieses Muster zu.

Da die beschriebenen großen Sprünge in der durchschnittlichen Klassengröße ausschließlich aufgrund der Regel zur maximalen Klassengröße zustande kommen, kann die durch sie hervorgerufene Variation in der Klassengröße wiederum als »exogen« in Bezug auf die Schülerleistungen angesehen werden. Die Regelung führt also zu einem weiteren »natürlichen« Experiment, das Variationen in der Klassengröße hervorruft, die weder durch die Schülerleistungen noch durch Zuweisungsentscheidungen hervorgerufen wurden (für Details dieser Strategie vgl. Wößmann 2005).

In der empirischen Schätzung werden in diesem Fall also nur die in Abbildung 6 gezeigten nicht-linearen Schwankungen in der durchschnittlichen Klassengröße aufgrund der Regel zur maximalen Klassengröße genutzt. Es wird also im Prinzip betrachtet, ob Schüler, die aufgrund einer Jahrgangsstufengröße von 31 oder 61 in relativ kleinen Klassen von 15 bis 21 Schülern unterrichtet werden, im TIMSS-Test besser abschneiden als Schüler, die aufgrund einer Jahrgangsstufengröße von 30 oder 60 in relativ großen Klassen von 30 Schülern unterrichtet werden. Da die empirische Methode ausschließlich solche Variationen in den Klassengrößen nutzt, die aufgrund der Regel vorhergesagt werden können, macht es für die Validität der Schätzung nichts, dass das beschriebene Muster in Deutschland nicht ganz perfekt vorliegt. Eine »unsaubere« Anwendung der Regel wür-

**Abb. 6**  
Jahrgangsstufengröße und Klassengröße



Zusammenhang zwischen Anzahl der Schüler in einer Jahrgangsstufe einer Schule und der durchschnittlichen Klassengröße in dieser Jahrgangsstufe im deutschen TIMSS-Datensatz. Punkte: tatsächlicher Zusammenhang. Linie: von einer Regel mit maximaler Klassengröße von 30 vorhergesagter Zusammenhang.

Quelle: Wößmann (2005).

de nur dazu führen, dass der Effekt statistisch ungenauer geschätzt werden kann.

Wie die in den Abbildungen 3 und 5 blau dargestellten Ergebnisse der auf solchen regelinduzierten Diskontinuitäten in der Klassengröße basierenden quasi-experimentellen Identifikationsstrategie zeigen, erweist sich der geschätzte Effekt als genau genug, um eine sinnvolle Aussage über die Größe des Klassengrößeneffekts treffen zu können. Und diese Aussage lautet wie schon im Falle der ersten quasi-experimentellen Untersuchung: Der kausale Effekt der Klassengröße auf die Schülerleistungen lässt sich statistisch nicht signifikant von null unterscheiden und ist mit großer Gewissheit nicht so groß, dass die Kosten einer Klassengrößenreduktion ihren Nutzen aufwiegen würden. Dabei ist die statistische Präzision dieser Schätzung noch höher als die der zuvor betrachteten Identifikationsstrategie, was in dem schmaleren Signifikanzband zum Ausdruck kommt (vgl. Wößmann 2005 für eine Diskussion der verbleibenden Unzulänglichkeiten der Ergebnisse).

Beide Strategien zur Schätzung des kausalen Klassengrößeneffektes mittels quasi-experimenteller Herangehensweisen, die durch »natürliche« Experimente hervorgerufene exogene Variationen in Klassengrößen identifizieren, kommen also zu demselben Schluss: Verkleinerungen der Klassen führen in Deutschland kaum zu nennenswerten Verbesserungen der mathematischen Leistungen der Schüler, so dass eine solche Maßnahme weit von jeglicher Kosteneffizienz entfernt wäre.

### Lehrerqualität und Klassengrößeneffekte

Wößmann (2005) und Wößmann und West (2005) schätzen Klassengrößeneffekte in vergleichbarer Weise für zahlreiche weitere Länder. Es zeigt sich, dass Deutschland kein Ausnahmefall ist: In nahezu allen Ländern sind die Effekte so klein, dass eine kosteneffiziente Größenordnung ausge-

<sup>8</sup> Leider war in dieser Untersuchung eine Nutzung der offiziellen Klassenteiler nach Bundesländern und Schularten nicht möglich, weil der Forschung die Identifikation von Bundesländern und Schularten in den deutschen Schülerleistungstests im Allgemeinen nicht ermöglicht wird.

geschlossen werden kann. Ausnahmen stellen lediglich Island, und zu einem geringeren Maße Griechenland, dar. Damit stellt sich die Frage, warum einige Länder Klassengrößeneffekte aufweisen und andere nicht. Die angesprochenen Studien gehen dieser Frage nach, indem sie testen, ob die geschätzte Effektgröße systematisch mit anderen Merkmalen der Länder variiert. Sie finden systematische Unterschiede weder für Kinder mit verschiedenem familiären Hintergrund, noch für Länder mit Unterschieden in der Durchschnittsleistung, dem wirtschaftlichen Entwicklungsstand, der durchschnittlichen Klassengröße oder dem Ausgaben-niveau im Schulsystem.

Demgegenüber scheint das Auftreten von Klassengrößeneffekten aber mit dem Gehaltsniveau und Bildungsstand der Lehrer zusammenzuhängen: Klassengrößeneffekte treten nur in Ländern mit einem relativ niedrigen Gehaltsniveau und relativ geringer Ausbildung der Lehrer auf, während Länder mit relativ hohen Lehrergehältern und langer Lehrerausbildung keine Klassengrößeneffekte aufweisen. Ebenso ist der geschätzte Klassengrößeneffekt innerhalb von Ländern, in denen der Ausbildungsstand der Lehrer variiert, in solchen Klassen größer, die von Lehrern mit geringerer Ausbildung unterrichtet werden.

Sieht man das durchschnittliche Gehalts- und Ausbildungsniveau der Lehrer als Maße für das allgemeine Fähigkeitsniveau der Lehrerschaft an, so legen diese Befunde die Interpretation nahe, dass relativ fähige Lehrer mit dem Unterricht in großen Klassen genauso gut zurechtkommen wie in kleinen Klassen. Darum ergibt sich bei ihnen kein Effekt der Klassengröße auf die Schülerleistungen. Demgegenüber scheinen weniger fähige Lehrer mit dem Unterricht in großen Klassen überfordert, während sie in kleinen Klassen einigermaßen zurechtkommen. Damit ergibt sich bei ihnen ein Klassengrößeneffekt. Die Ergebnisse legen also nahe, dass Klassengrößeneffekte nur bei einer relativ niedrigen Qualität der Lehrerschaft auftreten. Kleinere Klassen führen also nur in den wenigsten Fällen zu besseren Schülerleistungen – und vor allem nur in Ländern mit niedrigen Lehrergehältern und kurzer Ausbildung der Lehrer, zu denen Deutschland nicht gehört.

### Zusammenfassung und Ausblick

Die berichteten aggregierten und individuellen Ergebnisse zum Einfluss von Ausgaben und Klassengrößen auf die Leistungen der Schüler liefern keinerlei Evidenz, dass höhere Ausgaben oder kleinere Klassen automatisch zu einem besseren schulischen Lernen führen würden. Aus ökonomischer Sicht scheint klar: Eine Kosteneffizienz von pauschalen Verkleinerungen der Klassengrößen ist in Deutschland zumindest in den weiterführenden Schulen nicht gegeben. Von einer Größenordnung, die eine ökonomische Rechtfertigung

sicherstellen könnte, sind die in der Realität beobachtbaren Effekte kleinerer Klassen auf das Erlernte der Schüler weit entfernt. Dies sagt nichts darüber aus, ob die Einstellung zusätzlicher Lehrer mit dem Ziel, Unterrichtsausfall zu verringern, nicht vielleicht ein sinnvolles bildungspolitisches Instrument sein könnte (worüber wir wiederum wenig wissen). Aber die Einstellung zusätzlicher Lehrer zur Verkleinerung der Klassen kann mit einer möglichen Steigerung der erlernten Basiskompetenzen der Schüler nicht begründet werden. Allerdings mag es andere Interessen geben, die für Verkleinerungen der Klassen sprechen. Es ist leicht vorstellbar, dass es etwa durchaus im Interesse der Lehrer sein kann, in kleineren Klassen zu unterrichten: Als Lehrer an der Universität kann ich aus eigener Erfahrung sagen, dass es in kleineren Klassen oftmals leichter ist, die Ordnung aufrecht zu erhalten, ganz zu schweigen davon, dass man unvergleichlich weniger Korrekturaufwand bei Tests und Klausuren hat, wenn man 5 statt 500 Studenten unterrichtet! Aber ob dieses Interesse von Lehrern die immensen Kosten kleinerer Klassen rechtfertigen kann, sei hier dahingestellt ...

Bedeutet die hier berichteten eher dürrtigen Befunde zu den Lerneffekten kleinerer Klassen und höherer Bildungsausgaben, dass Lehrer und Schulen für den Bildungserfolg nicht relevant sind? Weit gefehlt! So können etwa Rivkin et al. (2005) einen außergewöhnlich umfangreichen Datensatz über die jährliche Entwicklung der schulischen Leistungen aller texanischen Schüler über viele Jahre nutzen, um deren Leistungen als »Value added« von einem zum nächsten Jahr zu messen. Anhand solch umfangreicher Daten können sie zeigen, dass unterschiedliche Schulen, und unterschiedliche Lehrer innerhalb derselben Schule, denselben Schülern ganz unterschiedlich viel zusätzliches Wissen beibringen. Es gibt also durchaus große Unterschiede im Bildungserfolg verschiedener Schulen und Lehrer. Nur lässt sich diese Variation in der Qualität der Lehre so gut wie nicht auf messbare Eigenschaften der Lehrerschaft, wie etwa deren Bildungsstand und Lehrerfahrung, zurückführen.<sup>9</sup> Der Befund deutet darauf hin, dass es einerseits sehr fähiges Lehrpersonal gibt, das den Schülern viel beibringt, und andererseits auch Lehrpersonal, bei dem die Schüler sehr wenig lernen. Das ist natürlich eine Binsenweisheit, die allen Eltern aus eigener Erfahrung bekannt ist. Nur mindert dieser Befund die Möglichkeit der Schulpolitik, die Bildungschancen der Schüler anhand von Maßnahmen, die an leicht beobachtbaren Merkmalen der Lehrer ansetzen, zu verbessern. Schulen und Lehrer sind zwar sehr relevant – aber leider lässt sich diese Relevanz nicht an leicht zu messenden Merkmalen festmachen.

Von einer solch ergiebigen, weil erkenntnisgewinnenden Datenlage, die neues Wissen darüber entstehen lässt, was für

<sup>9</sup> Die einzige Ausnahme besteht darin, dass ganz neue Lehrer in ihrem ersten Jahr den Schülern weniger Wissen vermitteln.

das schulische Wohl unserer Kinder wichtig ist und was nicht, sind wir in Deutschland derzeit noch weit entfernt. Hierzulande können wir leider nur davon träumen, solch für die Entwicklung der Basiskompetenzen der Schüler fundamentales Wissen zu sammeln: Berechtigte Datenschutzbedenken werden hierzulande dazu verwendet, jeglichen nutzbringenden Zugang von Forschern zu relevanten Leistungs- und Hintergrunddaten deutscher Schülerinnen und Schüler zu verhindern. Auch in dieser Hinsicht muss sich in unserem Lande statt »mehr desselben« einiges Grundlegendes verändern, bevor wir zu wirklichen Verbesserungen der Bildungschancen unserer Kinder kommen.

## Literatur

- Angrist, J.D. und V. Lavy (1999), »Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement«, *Quarterly Journal of Economics* 114(2), 533–575.
- Barro, R.J. (2001), »Human Capital and Growth«, *American Economic Review* 91(2), 12–17.
- Baumgartner, H.J. (2003), »Kleinere Schulklassen: Wirklich nützlich? Für Westdeutschland zeigt sich kein Einfluss auf Einstiegsgehälter«, *DIW Wochenbericht* (42), 638–640.
- Carneiro, P. und J.J. Heckman (2003), »Human Capital Policy. Rejoinder«, in: J.J. Heckman und A.B. Krueger (Hrsg.), *Inequality in America: What Role for Human Capital Policies?*, MIT Press, Cambridge Mass.
- Fertig, M. und R.E. Wright (2005), »School Quality, Educational Attainment and Aggregation Bias«, *Economics Letters* 88(1), 109–114.
- Gundlach, E. und L. Wößmann (2000), »Die sinkende Produktivität der Schulen in OECD-Ländern«, in: M. Weiß und H. Weishaupt (Hrsg.), *Bildungsökonomie und Neue Steuerung*, Beiträge zur Bildungsplanung und Bildungsökonomie, Vol. 9, Lang, Frankfurt, 31–52.
- Gundlach, E. und L. Wößmann (2004), »Bildungsressourcen, Bildungsinstitutionen und Bildungsqualität: Makroökonomische Relevanz und mikroökonomische Evidenz«, in: U. Backes-Gellner und P. Moog (Hrsg.), *Ökonomie der Evaluation von Schulen und Hochschulen*. Schriften des Vereins für Socialpolitik, N.F. Band 302, Duncker & Humblot, Berlin, 15–52.
- Gundlach, E., L. Wößmann und J. Gmelin (2001), »The Decline of Schooling Productivity in OECD Countries«, *Economic Journal* 111(471), C135–C147.
- Hanushek, E.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, E.A. (2002), »Publicly Provided Education«, in: A.J. Auerbach und M. Feldstein (Hrsg.), *Handbook of Public Economics*, Vol. 4, North Holland, Amsterdam, 2045–2141.
- Hanushek, E.A. und D.D. Kimko (2000), »Schooling, Labor-Force Quality, and the Growth of Nations«, *American Economic Review* 90(5), 1184–1208.
- Hanushek, E.A. und F. Welch, Hrsg. (2006), *Handbook of the Economics of Education*, erscheint bei: North Holland, Amsterdam.
- Hanushek, E.A. und L. Wößmann (2005), »The Cross-Country and Time-Series Pattern of International Educational Performance«, mimeo, Stanford University und ifo Institut für Wirtschaftsforschung an der Universität München.
- Hoxby, C.M. (2000), »The Effects of Class Size on Student Achievement: New Evidence from Population Variation«, *Quarterly Journal of Economics* 115(4), 1239–1285.
- Jones, Ch.I. (2002), *Introduction to Economic Growth*, Second Edition, W. W. Norton & Company, New York, N.Y.
- Krueger, A.B. (1999), »Experimental Estimates of Education Production Functions«, *Quarterly Journal of Economics* 114(2), 497–532.
- Organisation for Economic Co-operation and Development, OECD (2004), *Learning for Tomorrow's World: First Results from PISA 2003*, OECD, Paris.
- Rivkin, St.G., E.A. Hanushek und J.F. Kain (2005), »Teachers, Schools, and Academic Achievement«, *Econometrica* 73(2), 417–458.
- West, M.R. und L. Wößmann (2005), »Which School Systems Sort Weaker Students into Smaller Classes? International Evidence«, erscheint in: *European Journal of Political Economy* (Working-Paper-Version verfügbar als CESifo Working Paper 1054 unter [www.cesifo.de](http://www.cesifo.de)).
- Wößmann, L. (2002), *Schooling and the Quality of Human Capital*, Springer, Berlin.
- Wößmann, L. (2003a), »Specifying Human Capital«, *Journal of Economic Surveys* 17(3), 239–270.
- Wößmann, L. (2003b), »Returns to Education in Europe (Book Review Essay)«, *Review of World Economics/Weltwirtschaftliches Archiv* 139(2), 348–376.
- Wößmann, L. (2005), »Educational Production in Europe«, *Economic Policy* 20(43), 445–504.
- Wößmann, L. und M.R. West (2005), »Class-Size Effects in School Systems Around the World: Evidence from Between-Grade Variation in TIMSS«, erscheint in: *European Economic Review*.