Schriften des Vereins für Socialpolitik

Band 302

Ökonomie der Evaluation von Schulen und Hochschulen

Von

Steffen Blaga, Felix Büchel, Maja Coradi Vellacott,
Günter Fandel, Erich Gundlach, Ulrike L. Heregger, Hendrik Jürges,
Martin G. Kocher, Kerstin Schneider, Matthias Sutter,
Gert G. Wagner, Stefan C. Wolter, Ludger Wößmann

Herausgegeben von

Uschi Backes-Gellner
Petra Moog

Duncker & Humblot · Berlin

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

Schriften des Vereins für Socialpolitik

Gesellschaft für Wirtschafts- und Sozialwissenschaften

Neue Folge Band 302

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

SCHRIFTEN DES VEREINS FÜR SOCIALPOLITIK
Gesellschaft für Wirtschafts- und Sozialwissenschaften
Neue Folge Band 302

Ökonomie der Evaluation von Schulen und Hochschulen

Duncker & Humblot · Berlin

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

Ökonomie der Evaluation von Schulen und Hochschulen

Von

Steffen Blaga, Felix Büchel, Maja Coradi Vellacott,
Günter Fandel, Erich Gundlach, Ulrike L. Heregger, Hendrik Jürges,
Martin G. Kocher, Kerstin Schneider, Matthias Sutter,
Gert G. Wagner, Stefan C. Wolter, Ludger Wößmann

Herausgegeben von

Uschi Backes-Gellner
Petra Moog

Duncker & Humblot · Berlin

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

Bibliografische Information Der Deutschen Bibliothek

Die Deutsche Bibliothek verzeichnet diese Publikation in
der Deutschen Nationalbibliografie; detaillierte bibliografische
Daten sind im Internet über <http://dnb.ddb.de> abrufbar.

Alle Rechte, auch die des auszugsweisen Nachdrucks, der fotomechanischen
Wiedergabe und der Übersetzung, für sämtliche Beiträge vorbehalten
© 2004 Duncker & Humblot GmbH, Berlin
Fremddatenübernahme und Druck:
Berliner Buchdruckerei Union GmbH, Berlin
Printed in Germany

ISSN 0505-2777
ISBN 3-428-11569-4

Gedruckt auf alterungsbeständigem (säurefreiem) Papier
entsprechend ISO 9706 ♾

Internet: http://www.duncker-humblot.de

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

Vorwort

Der vorliegende Sammelband dokumentiert die Beiträge der Jahrestagung 2003 des Bildungsökonomischen Ausschusses des Vereins für Socialpolitik. Die Tagung fand auf Einladung von Frau Prof. Dr. Andrea Schenker-Wicki (organisatorische Vorbereitung) und Frau Prof. Dr. Uschi Backes-Gellner (inhaltliche Vorbereitung) am 27. / 28. März 2003 an der Universität Zürich statt. Thema der Jahrestagung war die Evaluation von Bildungsinstitutionen. Angeregt wurde es durch die Ergebnisse der PISA-Studie, insbesondere das vergleichsweise schlechte Resultat für Deutschland und auch die Schweiz und die Diskussion über die daraus zu ziehenden organisatorischen und bildungspolitischen Konsequenzen. In zehn Vorträgen widmete sich der Ausschuss Fragen zur Messung von Schul- und Hochschulleistungen, der organisatorischen Determinanten variierender Bildungsleistungen und den sozioökonomischen Einflussfaktoren sowie deren markroökonomischen Konsequenzen. Alle hier erschienenen Texte wurden, wie in den vorhergehenden Jahren auch, einer doppelten Begutachtung unterzogen und aufgrund dessen teils erheblich überarbeitet. Die ersten drei Beiträge beschäftigen sich mit der Evaluation von Schulleistungen, die letzten drei mit der Messung von Lehr- und Forschungsleistungen von Universitäten.

Erich Gundlach und *Ludger Wößmann* untersuchen in ihrem Beitrag einerseits die Bildungsqualität als treibende Kraft der gesamtwirtschaftlichen Entwicklung und andererseits mikroökonomische Determinanten der Bildungsqualität. Ausgangspunkt ist die Beobachtung, dass zwar in öffentlichen Diskussionen stets die große Bedeutung des Faktors Humankapital für die Produktivität und internationale Wettbewerbsfähigkeit einer Volkswirtschaft hervorgehoben wird, dass aber neuere Studien immer wieder zu dem Ergebnis kommen, dass Differenzen in der Humankapitalausstattung die großen internationalen Unterschiede in der Produktivität des Faktors Arbeit nicht erklären konnten. Ein Mangel bisheriger Studien ist wohl, dass sie das Humankapital nur anhand des durchschnittlichen schulischen Ausbildungsstandes der Arbeitskräfte erfassen. Dabei wird z. B. die Qualität der schulischen Ausbildung nicht berücksichtigt, die aber je nach Land deutlich unterschiedlich ausfallen kann, was nicht erst seit PISA bekannt ist. *Gundlach* und *Wößmann* gewichten deshalb in ihrer Analyse die Anzahl der Schuljahre mit einem Qualitätsindikator, der auf den Ergebnissen internationaler Vergleichstests in den Fächern Mathematik und Naturwissenschaften beruht. Es zeigt sich, dass unter Verwendung eines solchen qualitätsbereinigten Humankapitalkonzepts rund die Hälfte der internationalen Produktivitätsunterschiede erklärt werden können. Beschränkt man sich in den Analysen auf solche Länder, für die verlässliche

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

Datenquellen vorliegen (z. B. nur auf OECD-Staaten), zeigt sich sogar, dass bereits allein mit den Unterschieden im Humankapitalbestand die gesamte Variation in der Arbeitsproduktivität rein statistisch gesehen erklärt werden kann. Darüber hinaus zeigt sich, dass ein höherer Bestand an qualitätsbereinigtem Humankapitel eindeutig auch mit einem deutlich höheren durchschnittlichen Einkommen des ärmsten Teils der Bevölkerung einhergeht. Vor diesem Hintergrund bekommt die Frage nach der Erhöhung der Bildungsqualität einen weiteren wirtschafts- und bildungspolitisch äußerst prominenten Stellenwert. Vor dem Hintergrund der großen Bedeutung der Bildungsqualität untersuchen *Gundlach* und *Wößmann* im zweiten Teil Einflussgrößen auf die Bildungsqualität bzw. genau genommen den relativen Leistungsstand der Schüler im internationalen Vergleich. Dabei wird besonderer Wert darauf gelegt, möglichen Endogenitätsproblemen Rechnung zu tragen, um tatsächlich den Effekt unterschiedlicher (bildungs-)organisatorischer Designs bzw. den so genannten Maßnahmeeffekt unterschiedlicher Bildungsorganisationsformen herausarbeiten zu können. So zeigt sich dann beispielsweise, dass die Klassengröße für die überwiegende Anzahl der betrachteten Länder keinen statistisch signifikanten Effekt auf die Schülerleistungen ausübt. Nur in vier Ländern haben größere Klassen einen negativen Effekt auf die Schülerleistungen. Dies sind allerdings alles Länder, die sich gleichzeitig durch niedrige Lehrergehälter auszeichnen. Das kann dahin gehend interpretiert werden, dass große Klassen nur dann ein Problem darstellen, wenn sie zusammen mit einer unterdurchschnittlich qualifizierten Lehrerschaft auftreten. Während also die Klassengröße lange nicht den Effekt hat, den man aufgrund der Bedeutung in öffentlichen Diskussionen erwarten würde, haben zentrale Leistungskriterien dagegen einen in der öffentlichen Diskussion eher unterschätzten, deutlich positiven Effekt auf die Qualität der Schülerleistungen. So weisen Schüler, die in Ländern mit zentralen Leistungskriterien zur Schule gehen, einen beachtlichen Leistungsvorsprung auf, der im Effekt ungefähr dem eines ganzen Schuljahres entspricht. Außerdem belegen die Analysen von *Gundlach* und *Wößmann,* dass zentrale Prüfungen auch den Einfluss des elterlichen Bildungsstandes nivellieren, so dass sie im Ergebnis auch zu mehr Chancengleichheit im Bildungswesen führen in dem sie die Leistungen von Schülern mit nachteiligem familiären Hintergrund positiv beeinflussen. Entgegen mancher bildungspolitischen Diskussion stehen zentrale Prüfungen auch nicht einer dezentralen Organisation von Schulsystemen im Wege, sondern sind vielmehr als Voraussetzung für deren effizientes Funktionieren anzusehen. Nur die Kombination von autonomen Schulentscheidungen mit zentralen Leistungsüberprüfungen garantiert positive Effekte auf die Schülerleistungen, da allein in dieser Kombination geeignete Anreizwirkungen für alle Beteiligten (Schüler, Lehrer, Schulen und Schulbehörden) gewährleistet zu sein scheinen, was wiederum auf ein zentrales bildungspolitisch nicht zu vernachlässigendes Problem hinweist: die Bedeutung zentraler Leistungsstandards bzw. Abschlussprüfungen für die Effizienz eines Schulsystems.

Letztgenanntes Problem, das bereits 1997 im bildungsökonomischen Ausschuss auf Basis eines Referats von *Backes-Gellner* und *Weckmüller* diskutiert wurde,

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

greifen auch *Felix Büchel*[1], *Hendrik Jürges* und *Kerstin Schneider* in ihrem Beitrag auf. Sie untersuchen den Einfluss zentraler Abschlussprüfungen auf schulische Leistungen basierend auf einer Auswertung von TIMSS-Daten bei deutschen Haupt- und Realschülern. Sie nutzen dabei die institutionelle Variation im deutschen Schulsystem und versuchen, den Effekt der nur in machen Bundesländern existierenden zentralen Prüfungen zu schätzen und ihn insbesondere von anderen bundesländer- oder schülerspezifischen Einflussfaktoren zu isolieren. Hierfür wird ein Differenzen-in-Differenzen-Schätzer verwendet. Es zeigt sich, dass nur ein kleiner Teil der ursprünglich großen bundesländerspezifischen Leistungsdifferenzen auf die Wirkung zentraler Abschlussprüfungen zurückzuführen ist. Während der unkontrollierte Leistungseffekt ungefähr zwei Schuljahren entspricht, reduziert sich der kontrollierte Effekt zentraler Prüfungen auf die Schulleistung auf durchschnittlich ein halbes Schuljahr. Dabei ist der Effekt von zentralen Prüfungen in dieser Studie vom sozialen Hintergrund der Schüler statistisch nicht signifikant beeinflusst. Darüber hinaus weist die Studie aber nach, dass in Ländern, in denen zentrale Prüfungen existieren, die Noten besser mit den tatsächlichen Leistungen der Schüler korrespondieren. Noten aus Zentralprüfungen stellen dementsprechend auch ein besseres Arbeitsmarktsignal dar, was wiederum einen positiven Effekt auf die Lernanstrengungen der Schüler ausüben dürfte (vgl. hierzu auch *Backes-Gellner / Weckmüller* 1998).

Im dritten Beitrag zur Qualität von Schulleistungen untersuchen *Stefan Wolter* und *Maja Vellacott* schließlich für die Schweiz, welchen Effekt ein spezieller Aspekt des familiären Hintergrundes, nämlich die Geschwisterzahl, auf die schulischen Leistungen ausübt. Sie zeigen, dass die Zahl der Geschwister zwar einen Einfluss auf die schulischen Leistungen hat, dass dieser aber nur zu Tage tritt, wenn er in Kombination mit anderen bildungsfernen, sozioökonomischen Faktoren zusammen auftritt. Nur in den Teilpopulationen, die einen geringeren Bildungsstand, eine größere kulturelle Distanz zum Aufenthaltsland und einen niedrigeren sozioökonomischen Status aufweisen, lassen sich signifikante familiengrößenspezifische Effekte nachweisen. Es wird vermutet, dass dies daran liegt, dass die Eltern in solchen Familien ihren Kindern erstens nur ein geringes Maß an Ressourcen für schulische Bildung zukommen lassen können und dass diese wenigen Ressourcen dann noch mit mehreren Geschwistern geteilt werden müssen. Aus diesem Grund wird die Frage, ob der Leistungseffekt der Geschwisterzahl tatsächlich über eine Reduktion der je Kind verfügbaren Ressourcen zu Stande kommt, einem direkteren Test unterzogen. Die materielle Ausstattung eines Kindes wurde daran bemessen, ob diesem ein eigenes Zimmer zur Verfügung steht. Dabei wird unterstellt, dass die Verfügbarkeit über ein eigenes Zimmer sowohl direkt lernförderlich

[1] Leider ist *Felix Büchel* am 12. 7. 2004 verstorben – nach kurzer, aber schwerer Krankheit und viel zu früh. Mit großer Trauer haben wir miterlebt, wie er trotz seines schweren Krebsleidens bis wenige Tage vor dem Tode seine bildungsökonomischen Arbeiten vorangetrieben und am bildungspolitischen Diskurs aktiv teilgenommen hat. Im bildungsökonomischen Ausschuss werden wir ihn als engagiertes Mitglied sehr vermissen.

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

sein kann, als auch als indirekter Indikator für die restliche materielle Ausstattung der Kinder gewertet werden kann. Wie erwartet zeigt sich, dass die Geschwisterzahl tatsächlich einen signifikant negativen Effekt auf die Ausstattung mit einem eigenen Zimmer ausübt. Es deutet sich also an, dass zumindest für eine bestimmte Teilpopulation der Bevölkerung, d. h. die ökonomisch eher schlechter ausgestatten Familien, die Zahl der Kinder aufgrund von Budgetbeschränkungen tatsächlich einen negativen Einfluss auf die Schulleistungen nach sich zieht. Demnach könnten gezielte Ausgleichsmaßnahmen für solche Familien einen signifikant positiven Effekt auf die Schulleistungen bewirken. Dabei ist es aber aufgrund der eingeschränkten Aussagemöglichkeiten des vorliegenden Datensatzes nicht eindeutig zu entscheiden, ob Geld alleine der entscheidende Faktor ist, oder ob es nicht andere unbeobachtbare Faktoren, wie ein lernförderliches oder lernhinderliches soziales Umfeld sind, die zufällig mit der materiellen Ausstattung korrelieren, letztlich aber entscheidender als diese sind. Unter diesen Umständen wäre es möglicherweise hilfreicher, das Geld für eine ausgedehnte Kinderbetreuung außer Hauses zu verwenden und dort eine optimale lernförderliche Atmosphäre sicherzustellen. Zudem ist offen, ob solche gegensteuernden Maßnahmen während der Schulzeit – und damit in einem relativ späten Stadium – nicht zu spät greifen. Zu überdenken ist daher, ob es nicht effizienter wäre, in einer sehr viel früheren Phase etwa durch ausgeweitete Kindergarten- oder Krabbelstundenangebote mit einer Förderung zu beginnen.

Im zweiten Teil des Sammelbandes, in dem die Evaluation von Hochschulen im Mittelpunkt steht, beschäftigen sich *Martin G. Kocher, Matthias Sutter* und *Ulrike Heregger* zunächst mit der Frage, welche Bedeutung die Auswahl der Zeitschriften für eine Evaluation von Forschung spielen könnte. Sie untersuchen dies ganz konkret für die Evaluation von Forschung im Fach Ökonomie. Dabei besteht in der Profession weitgehende Einigkeit darüber, dass Zeitschriftenpublikationen grundsätzlich ein geeignetes Instrument zur Messung von Forschungsleistungen darstellen. Allerdings scheiden sich die Geister, wenn es darum geht, welche Zeitschriften konkret berücksichtigt werden sollten. *Kocher, Sutter* und *Heregger* untersuchen vor diesem Hintergrund, ob die Auswahl der Zeitschriften überhaupt einen Unterschied macht bei einer Evaluation von volkswirtschaftlichen Fachbereichen im deutschsprachigen Raum und in Kanada. Dabei zeigt sich, dass es entgegen der in öffentlichen Diskussionen oft behaupteten Probleme – aber in Übereinstimmung mit vorhergehenden scientometrischen Befunden – keinen Trade-off zwischen dem Publikationsoutput in internationalen und in nationalen Spitzenzeitschriften gibt. Nebenbei zeigt sich aber auch, dass der Konzentrationsgrad in der Gruppe der amerikanischen Spitzenzeitschriften bei weitem am höchsten ist. Fast ein Drittel aller Einträge in amerikanischen Spitzenzeitschriften gehen auf nur 10 Top-Institutionen in den USA zurück, was möglicherweise einen Hinweis auf nicht zu vernachlässigende Oligopoltendenzen darstellt, deren langfristige Probleme *Brinkmann* bereits auf der Sitzung des Bildungsökonomischen Ausschusses 2002 zur Diskussion gestellt hatte (*Brinkmann* 2003). Dennoch gibt es eine sehr hohe Kor-

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

relation zwischen dem Output von führenden deutschsprachigen Institutionen in amerikanischen bzw. anderen internationalen Spitzenzeitschriften und in Spitzenzeitschriften aus dem deutschen Sprachraum. Institutionen, die international erfolgreich sind, zählen auch national zu den besten Institutionen und umgekehrt.

Im zweiten Beitrag zur Evaluation von Hochschulen beschäftigen sich *Günter Fandel* und *Steffen Blaga* mit der Frage, in welchem Maße der Erfolg von Hochschulen von der Qualität der Studierenden – gemessen anhand des Durchschnitts der Noten der ersten vier Klausuren des Grundstudiums – abhängt. Dabei zeigt sich, dass erhebliche Notenabweichungen zwischen Eingangsnoten und Vordiplom bzw. zwischen Vordiplom und Diplom kaum vorkommen. Studierende erzielen demnach auf den verschiedenen Qualifikationsstufen im Prinzip die gleichen Noten, wobei weibliche Studierende noch konstanter sind als männliche Studierende. Außerdem zeigt sich, dass eine verlängerte Studiendauer nicht mit einer Verbesserung der Noten einhergeht, genau wie eine verkürzte Studiendauer nicht mit schlechteren Noten erkauft wird. Der Output von Hochschulen gemessen an Noten hängt also offensichtlich von der Eingangsqualität der Studierenden ab. Offen bleibt allerdings, wie die genannten Befunde erklärt und hochschulpolitisch interpretiert werden können.

Im dritten Aufsatz zur Evaluation von Bildungseinrichtungen schließlich untersucht *Gert G. Wagner* anhand einer explorativen Studie die Frage, inwieweit es einen Trade-off zwischen den klassischen Aufgaben in Forschung und Lehre und der wissenschaftlichen Politikberatung als einer weiteren Aufgabe von Hochschullehrern gibt. Er geht dieser Frage anhand einer Untersuchung der volkswirtschaftlichen Politikberatung im angelsächsischen Raum auf der Basis einer Befragung eines Netzwerkes an ausgewiesenen Forschern, dem CEPR, nach. Dabei zeigt sich zunächst, dass ein sehr großer Teil der Befragten durchaus den Wunsch nach Politikberatung hat, dass aber nur ein sehr kleiner Teil darunter eine direkte und kurzfristige Beeinflussung der Politik versteht. Allgemein wird der kurzfristige Erfolg volkswirtschaftlicher Politikberatung sehr skeptisch gesehen, während die langfristigen Einflussmöglichkeiten (über die universitäre Lehre, die zukünftige Entscheidungsträger und Experten prägt) sehr viel optimistischer eingeschätzt werden. Eine ausführliche deskriptive Untersuchung des Zusammenhangs zwischen der Qualität der Forscher und der Politikberatung zeigt, dass es zumindest für ein Sample von Spitzenforschern offensichtlich keinen Trade-off zwischen akademischen Publikationserfolgen und dem Grad der Politikberatung zu geben scheint. Top-Publizierende mögen zwar weniger in populären Medien vertreten sein. Sie werden aber möglicherweise stärker gehört aufgrund publikumswirksamer Vorträge und öffentlicher Anhörungen und werden häufiger von Entscheidungsträgern unmittelbar um Rat gebeten. Vor allem auf lange Sicht scheint demnach Spitzenforschung auch einen wesentlichen Beitrag zur Politikberatung zu leisten. Es zeigt sich aber auch, dass Spitzenforschung mit besonders forschungsförderlichen Rahmenbedingungen einhergeht. Vor diesem Hintergrund alarmierend sind die länderspezifischen Unterschiede in den materiellen Bedingungen von Spitzenforschern.

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

Liegt das Lehrdeputat mit 3,6 Veranstaltungen pro Jahr in den kontinentaleuropäischen Ländern schon deutlich höher als in den angelsächsischen Ländern mit 2,6 Kursen, so übersteigen die Lehrbelastungen in Deutschland mit sechs und mehr Veranstaltungen pro Jahr vollkommen den internationalen Rahmen. Gleichzeitig wird nur in Deutschland angegeben, dass pro Jahr mehr als sechs Diplom- oder Magisterarbeiten betreut werden. D.h. im Ausland ist die Belastung mit Studierenden offensichtlich merklich geringer. Im Gegensatz dazu wird im Ausland eine leicht höhere Zahl an Doktorarbeiten betreut, die gleichzeitig mit einem höheren Anteil an Doktorandenkursen einhergeht. Es bestätigt sich also die in hochschulpolitischen Diskussionen oft von Hochschullehrern vorgetragene Vermutung, dass an guten Universitäten außerhalb Deutschlands (insbesondere in angelsächsischen Ländern) die gesamte Lehrbelastung deutlich geringer ausfällt und dass gleichzeitig der Anteil der Zeit, der für forschungsnahe Lehre (Doktorandenausbildung) zur Verfügung steht, merklich höher ist. Damit scheint es also – wie ansatzweise auch schon von *Backes-Gellner* auf der Jahrestagung 1988 gezeigt – weniger einen Trade-off zwischen Spitzenforschung und Politikberatung als vielmehr einen Trade-off zwischen forschungsferner Lehre einerseits und Spitzenforschung sowie Politikberatung andererseits zu geben. Hierüber sollte in hochschulpolitischen Diskussionen verschärft nachgedacht werden, insbesondere dann, wenn tatsächlich und ernsthaft die Herausbildung von Elitehochschulen weiterverfolgt wird.

Insgesamt wurden damit auf der Jahrestagung 2003 eine Vielzahl an schul- und hochschulpolitisch wichtigen Fragestellungen aufgegriffen und neue Erklärungszusammenhänge bzw. -vermutungen aufgedeckt. Diese können für bildungspolitische Entscheidungen mehr als wichtig sein, da sie Rückschlüsse auf die (mangelnde) Effektivität verschiedenartiger Maßnahmen zulassen. Gleichzeitig dokumentiert die diesjährige Tagung auch eine große Kontinuität der Arbeit des Bildungsökonomischen Ausschusses einerseits, da viele Fragen und Problemstellungen vergangener Jahrestagungen neu aufgegriffen wurden, und einen deutlichen Erkenntnisfortschritt andererseits, der aufgrund neuer Datensätze und fortgeschrittener ökonometrischer Methoden im Laufe der Jahre erzielt werden konnte. Mit der Jahrestagung 2003 wurde meine Amtszeit als Vorsitzende des bildungsökonomischen Ausschusses, die mit meiner Wahl 1999 und der 25-jährigen Jubiläumsfeier 2000 in Köln ihren Auftakt hatte, nach vier Jahren turnusgemäß und – wie ich hoffe – ertragreich abgeschlossen. Natürlich hat eine Vielzahl von Personen an diesem Ertrag mitgewirkt, denen hier abschließend noch einmal gedankt sei. Neben den Mitgliedern des Bildungsökonomischen Ausschusses, die durch ihr großes Vertrauen in die Arbeit ihrer neuen Vorsitzenden und ihre Diskussionsfreudigkeit maßgeblich am Erfolg der Tagungen beteiligt waren, sind dies die Referenten und Autoren, die mit hochwertigen, innovativen oder einfach nur diskussionsfreudigen Beiträgen für eine äußerst fruchtbare und lebendige Diskussion gesorgt haben. Und last but not least sind es die vielen helfenden Hände im Hintergrund, die maßgeblich zum organisatorischen Gelingen der Tagungen und

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

der Sammelbände beigetragen haben. Stellvertretend seien hierfür meiner Sekretärin Frau Christa Pálffy und meiner Mitherausgeberin Frau Dr. Petra Moog gedankt.

Zürich, im März 2004 *Uschi Backes-Gellner*

Literatur

Brinkmann, Gerhard (2003): Englisch: Die Hoffnung der deutschsprachigen Nationalökonomen. Ein Diskussionsbeitrag. In: Backes-Gellner, U. / Schmidtke, C. (Hrsg.): Hochschulökonomie – Analysen interner Steuerungsprobleme und gesamtwirtschaftliche Effekte. Berlin: Duncker & Humblot 2003. S. 51 – 61.

Backes-Gellner, Uschi / *Weckmüller,* Heiko (1998): Ist das Ende der Hauptschule aufzuhalten? Ein institutionsökonomischer Beitrag zur Wirkung alternativer Schulregulierungsstrategien auf Schulnachfrageverhalten. In: v. Weizsäcker, R.K. (Hrsg.): Deregulierung und Finanzierung des Bildungswesens. Berlin: Duncker & Humblot 1998. S. 49 – 77.

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

Inhaltsverzeichnis

Bildungsressourcen, Bildungsinstitutionen und Bildungsqualität: Makroökonomische Relevanz und mikroökonomische Evidenz

Von *Erich Gundlach,* Kiel und *Ludger Wößmann,* München 15

Leistungs- und Signaleffekte zentraler Abschlussprüfungen – Eine TIMSS-Auswertung bei deutschen Haupt- und Realschülern

Von *Felix Büchel,* Berlin, *Hendrik Jürges,* Mannheim und *Kerstin Schneider,* Dortmund 53

Sibling Rivalry in Education – an Empirical Investigation for Switzerland

By *Stefan C. Wolter* and *Maja Coradi Vellacott,* Aarau 75

Die Bedeutung der Zeitschriftenauswahl für die Evaluation ökonomischer Forschung

Von *Martin G. Kocher, Matthias Sutter* und *Ulrike L. Heregger,* Innsbruck 101

Dienstleistungsproduktion in Hochschulen. Zur Beschreibung der Qualitäten der externen Faktoren „Studierende“

Von *Günter Fandel* und *Steffen Blaga,* Hagen 119

Verhindert Politikberatung gute Lehre und Forschung? Eine explorative Evaluierung volkswirtschaftlicher Politikberatung mit Hilfe eines Expertensurveys

Von *Gert G. Wagner,* Berlin 139

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

Bildungsressourcen, Bildungsinstitutionen und Bildungsqualität: Makroökonomische Relevanz und mikroökonomische Evidenz

Von *Erich Gundlach,* Kiel[1] und *Ludger Wößmann,* München[2]

A. Einleitung

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

[1] Institut für Weltwirtschaft, Postfach 43 09, 24100 Kiel, Tel.: (04 31) 88 14 – 2 84, E-mail: egundlach@ifw.uni-kiel-de

[2] Ifo Institut für Wirtschaftsforschung, Poschingerstraße 5, 81679 München, Tel.: (0 89) 92 24 – 16 99, E-mail: woessmann@ifo.de

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

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 mikroökonometrische 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

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

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 mikroökonometrische 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 familiären 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)

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

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 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.

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 beschrieben 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:

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

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.

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

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:

$$H = L \cdot e^{r \cdot S} \; . \tag{2}$$

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

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

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 Qualitätsindexes gewichtet werden würde.[3]

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

Ausgangspunkt für die Berechnung des qualitätsgewichteten Humankapitalindexes 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 Werte für Q, die auf Basis der jeweiligen Durchschnittswerte für Ländergruppen nach Regionen sowie nach Einkommensklassen ermittelt wurden.

Die von uns gewählten 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.

---

[3] Zu den Einzelheiten der Berechnung des Qualitätsindexes vgl. *Gundlach* et al. (2002) sowie *Wößmann* (2003a).

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

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.

Tabelle 1

Humankapital als Komponente der Arbeitsproduktivität[a)]

	Anteiliger Beitrag der Faktoren . . .		
	Humankapital	Sachkapital	Totale Faktor-produktivität
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).

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

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

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 sich 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.

Tabelle 2

Humankapital als Komponente der Arbeitsproduktivität in OECD Ländern[a)]

	Anteiliger Beitrag der Faktoren . . .		
	Humankapital	Sachkapital	Totale Faktor-produktivität
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).

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

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*

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

(2002) reproduzieren, die eine Stichprobe von 269 gepoolten Querschnitts- und Zeitreihenbeobachtungen benutzen. Die OLS-Schätzgleichung mit Daten für das Jahr 1990

$$\ln y_p = a_0 + 1,06 \ln y_c \qquad \overline{R}^2 = 0,86, \quad n = 111 \tag{4}$$

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 der 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 für 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

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

wobei X für sonstige Einflussfaktoren auf das Einkommen der Armen steht.[4] 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.

[4] Zu den Einzelheiten der im Folgenden diskutierten Schätzungen vgl. *Gundlach* et al. (2004).

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

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	$n = 101$	$n = 101$	$n = 99$	$n = 91$	$n = 89$
$\overline{R}^2$	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).

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

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

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 $\left(\sigma(y_p)\right)$ 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:

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

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

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.[5]

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.[6] Im Durchschnitt 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

---

[5] 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 geographische Variablen bei Aufnahme in die Schätzgleichung als statistisch 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.

[6] 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. Der Test auf Überidentifikation bietet insofern keinen umfassenden Schutz vor Fehlspezifikationen, er kann aber durchaus Fehlspezifikationen offen legen.

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

durchschnittliche Einkommen der Armen hat. Wie ein Vergleich mit den Ergebnissen von *Dollar* und *Kraay* (2002) 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.

Tabelle 4

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

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

Instrument-Variablen	MEANTEMP	DISTANCE	*DISTANCE, MEANTEMP*
a_0	–1,69 (0,10)	–1,79 (0,12)	–1,69 (0,10)
$\ln(H/L)$	0,3 (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 Überidentifikation

Standardfehler in Klammern.

Quelle: *Gundlach* et al. (2004).

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.

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

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

ist. Dabei ist zu beachten, dass die in der mikroökonometrischen 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 positiven externen 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

$$\ln y_S = y_0 + rS \,. \tag{7}$$

Wenn diese Beziehung als Schätzgleichung auf die gesamtwirtschaftliche Ebene übertragen wird, wo y für das Bruttoinlandsprodukt pro Person und S für die Anzahl der Schuljahre pro Person stehen, dann reflektiert der Regressionskoeffizient r nicht nur die individuellen Erträge der schulischen Ausbildung wie bei mikroökonometrischen 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 sozioökonomische 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

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

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

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

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

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.[7] 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).

[7] 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).

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

Solange die Kapitalintensität nicht als erklärende Variable in die Schätzgleichung aufgenommen 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 mikroökonometrischen 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.

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)
$LIFE$	–	–	0,04 (0,01)	0,00 (0,01)
$\ln k$	–	–	–	0,43 (0,06)
Stichprobe	$n = 80$	$n = 71$	$n = 71$	$n = 71$
$\overline{R}^2$	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).

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

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

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

$$r_S = \frac{r}{1-\alpha} = 0,21 \ . \tag{10}$$

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 mikroökonometrischen 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.

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

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

Tabelle 6

Alternative Schätzungen der makroökonomischen Ertragsrate des Humankapitals

	$\ln y_c$		$\ln y_c - 0.3 \ln k$	
Abhängige Variable:	(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)
LIFE	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	$n = 72$	$n = 71$	$n = 72$	$n = 71$
$\overline{R}^2$	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)
Überidentifikation *p*-Wert	–	–	–	0,22
Testergebnis	–	–	–	keine Überidentifikation

Standardfehler in Klammern.

Quelle: *Gundlach* et al. (in Vorbereitung).

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

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

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.[8]

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 positiv auf die Durchschnittseinkommen wirkt, möglicherweise einen eigenen positiven Effekt auf das Einkommen der Armen besitzt sowie positive Einkommensexternalitäten hervorrufen könnte. Damit stellt

[8] 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.

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

sich unmittelbar die Frage, wie eine hohe Bildungsqualität mit Hilfe wirtschaftspolitischer Maßnahmen erreicht werden 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 ökonometrischer 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.

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

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.[9] Das

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

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

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 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

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

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 Kleinstquadrate-Schätzung folgender Schätzgleichung ergibt:

$$T_{icgs} = \alpha_1 S_c + Ctrl_{icgs}\beta + \gamma G_g + \upsilon_c + \varepsilon_{icgs} \ . \tag{11}$$

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, υ eine klassenspezifische Komponente des Schätzfehlers und ε eine schülerspezifische Komponente des Schätzfehlers.[10] 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. 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:

$$T_{icgs} = \alpha_2 S_c + Ctrl_{icgs}\beta + \gamma G_g + D_s\delta + \upsilon_c + \varepsilon_{icgs} \ . \tag{12}$$

[10] 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).

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

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

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

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:

$$S_c = \phi A_{gs} + Ctrl_{icgs}\beta + \gamma G_g + D_s\delta + \upsilon_c + \varepsilon_{icgs} \ , \tag{13}$$

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

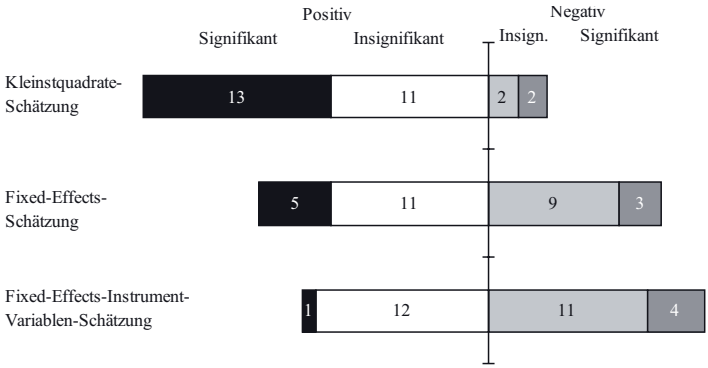
$$T_{icgs} = \alpha_3\hat{S}_c + Ctrl_{icgs}\beta + \gamma G_g + D_s\delta + \upsilon_c + \varepsilon_{icgs} \ , \tag{14}$$

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 erhalten.

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.

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/



Quelle: *Wößmann* und *West* (2002).

Abbildung 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.

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.[11] Auf die TIMSS-Skala 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

[11] *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.

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

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

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

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 Entscheidungszentralität, 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:

$$T_{icgso} = \lambda Z_o + Ctrl_{icgso}\beta + \mu_o + \upsilon_c + \varepsilon_{icgs} \ , \tag{15}$$

wobei Z_o eine Dummy-Variable ist, die angibt, ob im Land o Zentralprüfungen gegeben sind oder nicht.[12] Der Kontrollvariablenvektor umfasst in dieser Spezifi-

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

kation 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;[13] μ_o ist eine zusätzliche länderspezifische Komponente des Schätzfehlers.[14]

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 die zuvor ausschließlich anhand von TIMSS erzielten Ergebnisse (*Bishop* 1997; *Wößmann* 2002a), wonach Schülerleistungen in Zentralprüfungssystemen höher sind.[15] 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 Exis-

---

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

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

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

15 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 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.

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

tenz 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.

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).

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 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

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

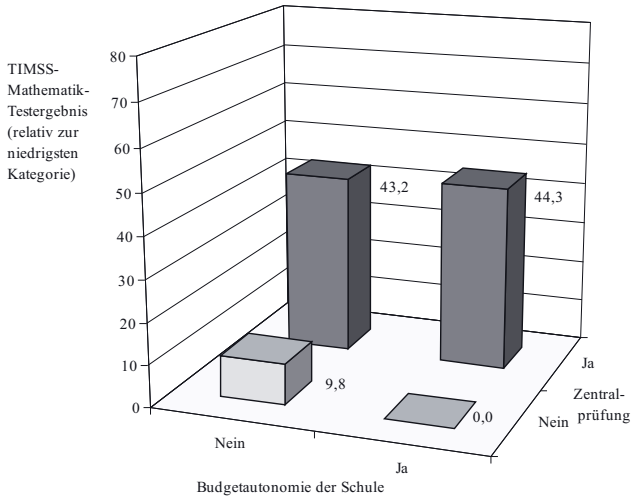
Tabelle 8 berichtet, ergeben die Schätzungen sogar dann statistisch signifikante Zentralprüfungseffekte, wenn die Variation zwischen den neun Regionen Westeuropa, Osteuropa, Nordamerika, Südamerika, Ozeanien, Asien, Mittlerer Osten, Nordafrika und Südafrika unberücksichtigt bleibt. Der geschätzte Zentralprüfungseffekt 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 familiären 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 Immigrantenkindern 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 familiären Hintergrund beitragen.

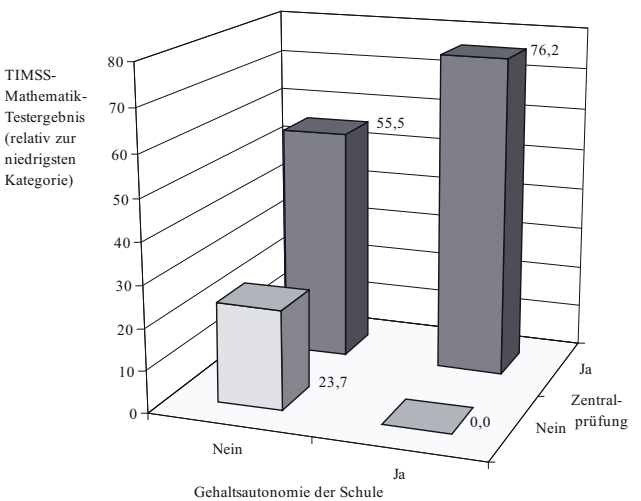
Aus weiteren empirischen Analysen wird deutlich, dass sich durch externe Prüfungen nicht nur die Anreize und damit das Verhalten der Schüler ändern, sondern dass sich auch Anreize und Verhalten weiterer Akteure des Schulsystems ändern (*Wößmann* 2002b). Als Beispiel sei hier auf die Interaktion von Zentralprüfungen 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

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/



(a) Budgetautonomie



(b) Gehaltsautonomie

Quelle: *Wößmann* (2002b).

Abbildung 2: Zentralprüfungen, Schulautonomie und Schülerleistungen

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

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

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.

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

Literatur

Acemoglu, Daron / *Johnson,* Simon / *Robinson,* James A. (2001): The Colonial Origins of Comparative Development: An Empirical Investigation. American Economic Review 91 (5): 1369 – 1401.

Angrist, Joshua D. / *Lavy,* Victor (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. / *Lee,* Jong-Wha (1996): International Measures of Schooling Years and Schooling Quality. American Economic Review 86 (2): 218 – 223.

Barro, Robert J. / *Lee,* Jong-Wha (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. / *Wößmann,* Ludger (2003): Institutional Effects in a Simple Model of Educational Production. Erscheint in: Education Economics.

Bloom, David E. / *Sachs,* Jeffrey D. (1998): Geography, Demography, and Economic Growth in Africa. Brookings Papers on Economic Activity (2): 207 – 295.

Case, Anne / *Deaton,* Angus (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 / *Squire,* Lyn (1996): A New Data Set Measuring Income Inequality. World Bank Economic Review 10 (3): 565 – 591.

Dollar, David / *Kraay,* Aart (2002): Growth Is Good for the Poor. Journal of Economic Growth 7 (3): 195 – 225.

DuMouchel, William H. / *Duncan,* Greg J. (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 / *Henseler,* Joana / *Knips,* Vivien (in Vorbereitung): Macroeconomic Mincering: The Social Rate of Return to Schooling. Institut für Weltwirtschaft, Kiel (mimeo).

Gundlach, Erich / *Navarro de Pablo,* Jose / *Weisert,* Natascha (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 / *Rudman,* Desmond / *Wößmann,* Ludger (2002): Second Thoughts on Development Accounting. Applied Economics 34 (11): 1359 – 1369.

Gundlach, Erich / *Wößmann,* Ludger / *Gmelin,* Jens (2001): The Decline of Schooling Productivity in OECD Countries. Economic Journal 111 (471): C135-C147.

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

Hall, Robert E. / *Jones,* Charles I. (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. / *Kimko,* Dennis (2000): Schooling, Labor Force Quality, and the Growth of Nations. American Economic Review 90 (5): 1184 – 1208.

Heckman, James J. / *Klenow,* Peter J. (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 / *Schneider,* Kerstin / *Büchel,* Felix (2002): The Effect of Central Exit Examinations on Student Achievement: Quasi-Experimental Evidence from TIMSS Germany. Mimeo.

Klenow, Peter J. / *Rodriguez-Clare,* Andrés (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. / *Whitmore,* Diane M. (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 / *Renelt,* David (1992): A Sensivity Analysis of Cross-Country Growth Regressions. American Economic Review 82 (4): 942 – 963.

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

McArthur, John W. / *Sachs,* Jeffrey D. (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. / *Howell,* William G. / *Wolf,* Patrick J. / *Campbell,* David E. (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 (1994): Returns to Investment in Education: A Global Update. World Development 22: 1325 – 1343.

Psacharopoulos, George / *Patrinos,* Harry A. (2002): Returns to Investment in Education: A Further Update. World Bank, Policy Research Working Paper, 2881, September.

Rodriguez, Francisco / *Sachs,* Jeffrey D. (1999): Why Do Resource-Abundant Economies Grow More Slowly? Journal of Economic Growth 4 (3): 277 – 303.

Rosenzweig, Mark R. / *Wolpin,* Kenneth I. (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. / *Wößmann,* Ludger (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.

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

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 / *West,* Martin R. (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.

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

Leistungs- und Signaleffekte zentraler Abschlussprüfungen – Eine TIMSS-Auswertung bei deutschen Haupt- und Realschülern[1]

Von *Felix Büchel*[2], Berlin, *Hendrik Jürges,* Mannheim
und *Kerstin Schneider,* Dortmund

A. Einleitung

In der aktuellen Diskussion über die Reform des deutschen Schulsystems wird häufig die Ansicht vertreten, zentrale Abschlussprüfungen verbesserten die durchschnittliche schulische Leistung. Zudem sei zu erwarten, dass die Schüler von zentralen Standards profitierten, da Noten in Systemen mit zentral definierten Standards auf dem Arbeitsmarkt verlässlichere Signale für die Fähigkeit von Bewerbern aussenden würden als Ergebnisse nicht-zentraler Prüfungen. Ein überzeugender empirischer Nachweis dieser unterstellten Zusammenhänge steht jedoch bislang aus. Dies soll in diesem Beitrag geleistet werden.

Nicht zuletzt als Reaktion auf das schlechte Abschneiden der deutschen Schüler in der PISA-Studie planen die Bundesländer Niedersachsen, Brandenburg und Berlin die Einführung des Zentralabiturs. Auch in anderen Bundesländern ist eine heftige Debatte über das Für und Wider zentraler Standards als Teil einer umfassenden Reform des deutschen Bildungssystems entbrannt.

Reformvorschläge für das Bildungssystem lassen sich in zwei Kategorien einteilen. Zum einen gibt es Vorschläge, die auf eine bessere finanzielle Ausstattung von Schulen zielen. Zum anderen stehen institutionelle Rahmenbedingungen des Schulsystems zur Diskussion. Deren Veränderungen, mit denen leistungsfördernde Anreize geschaffen werden, könnten mehr bewegen als die bloße Erhöhung der dem Bildungssektor zugewiesenen Ressourcen. Tatsächlich scheinen institutionelle Rahmenbedingungen des Schulsystems mehr zur Erklärung der Schulleistung beizutragen als die finanzielle Ausstattung der Schulen (vgl. hierzu ausführlicher

[1] Wir danken Dr. Rainer Watermann, Max-Planck-Institut für Bildungsforschung, Berlin, für die Ermöglichung des administrativen und technischen Zugangs zu den dieser Studie zugrunde liegenden regionalen Daten sowie Prof. Dr. Jürgen Baumert für seine Sondergenehmigung zur Verwendung dieser Daten. Die regionalen TIMSS-Daten unterliegen derzeit noch qualifizierten Datenschutzbestimmungen und werden interessierten Forschern erst nach ihrer faktischen Anonymisierung zugänglich sein. Zwei anonymen Gutachtern danken wir für hilfreiche Anregungen und konstruktive Kritik.

[2] † 12. 7. 2004.

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

Jürges und *Schneider,* 2004). Eine der Fragen, die in diesem Zusammenhang diskutiert werden, ist die Schaffung von mehr Wettbewerb zwischen Schulen (Hoxby, 2000). Ein weiterer Aspekt, und dies ist das Thema dieses Beitrags, sind zentral festgelegte Standards in Form zentraler Abschlussprüfungen.

Die nicht sehr umfangreiche theoretische Literatur zum Thema der zentralen Abschlussprüfungen kommt überwiegend zu dem Schluss, dass zentrale Abschlussprüfungen die Schulleistung erhöhen und sogar die allgemeine Wohlfahrt steigern können (*Costrell,* 1994, 1997; *Effinger* und *Polborn,* 1999). Zentrale Prüfungen setzen eher als dezentrale die richtigen Anreize für Schülerinnen und Schüler[3], Lehrkräfte und Schulen (z. B. *Bishop,* 1997, 1999). Die Schüler zum Beispiel profitieren, weil Ergebnisse von zentralen Abschlussprüfungen auf dem Arbeitsmarkt bessere Signale darstellen als Ergebnisse von nicht zentralen Prüfungen. Außerdem entfällt, wenn zum Abschluss der Schullaufbahn externe Standards erfüllt werden müssen, der Anreiz zu einer Kartellbildung in der Klasse, bei der sich die Schüler, womöglich mit stillschweigendem Einverständnis der Lehrkräfte, auf ein niedrigeres Leistungsniveau einstellen. Testergebnisse könnten auch zur regelmäßigen Kontrolle der Unterrichtsqualität und damit auch der Lehrkräfte herangezogen werden. Schließlich könnte gar der Ruf ganzer Schulen von den Leistungen ihrer Schüler abhängig gemacht werden, so dass – vorausgesetzt, die aggregierten Ergebnisse der zentralen Abschlussprüfungen sind öffentlich zugänglich – gute Schulen bessere Schüler anziehen würden.

International vergleichende Studien zum Leistungseffekt zentraler Abschlussprüfungen kommen durch Auswertung der TIMSS-Daten (Third International Mathematics and Science Study) zu dem Schluss, dass zentrale Prüfungen die Schulleistung signifikant in der Größenordnung etwa eines Schuljahrs erhöhen (*Bishop,* 1997, 1999). Allerdings ist der Befund eines positiven Leistungseffekt in internationalen Vergleichen auf Grundlage von TIMSS nicht sehr robust (*Jürges* und *Schneider,* 2004). Studien von Leistungsunterschieden in einzelnen Ländern mit regionaler institutioneller Variation erscheinen deutlich leistungsfähiger. *Bishop* (1997, 1999) legt neben internationalen Vergleichen auch Ergebnisse, die auf kanadischen Mikrodaten basieren, vor. Ähnlich wie in Deutschland gibt es in Kanada ein gemischtes System, bei dem in manchen Provinzen die Highschool mit zentralen Prüfungen abgeschlossen wird und in anderen nicht. Bishops Schätzungen zufolge läßt sich der positive Leistungseffekt zentraler Abschlussprüfungen in Kanada mit etwa zwei Dritteln des Äquivalents einer Jahrgangsstufe angeben. Die internationalen Mikrodaten der TIMSS werden in *Wößmann* (2002, 2003) ausgewertet. Wößmann findet in seinen Schätzungen ebenfalls starke positive Leistungseffekte zentraler Prüfungen in der Größenordnung eines Schuljahres.

Allerdings ist es für die Beurteilung der Vorteilhaftigkeit zentraler Standards wichtig, zwischen einfacher Korrelation und Kausalität zu unterscheiden. So könn-

[3] Wir verwenden im folgenden im Interesse der besseren Lesbarkeit nur die männliche Form.

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

ten in Ländern oder Regionen mit zentralen Abschlussprüfungen Bildung und Leistung ganz allgemein höhere Priorität haben. In diesem Falle wären sowohl die hohe durchschnittliche Schulleistung als auch die zentralen Abschlussprüfungen lediglich Ausdruck einer Präferenz der Bevölkerung für gute Schulbildung. *Bishop* (1997) versucht, diesem Problem Rechnung zu tragen, indem er der Frage nachgeht, ob sich Länder mit zentralen Abschlussprüfungen von Ländern ohne solche Prüfungen auch noch in anderen Dimensionen als nur der mittleren Schulleistung unterscheiden. Die verwendeten Daten lassen jedoch letztlich keine überzeugende Identifikationsstrategie zu, so dass wichtige Fragen offen bleiben.

Die föderale Struktur des deutschen Schulsystems stellt demgegenüber eine besonders geeignete Quelle institutioneller Variation zur Identifikation des Kausaleffekts zentraler Abschlussprüfungen auf die Schulleistung dar. So konnten frühere Studien bei Schülern in Mathematikgrundkursen der Sekundarstufe II Leistungsunterschiede nach Art der Abiturprüfung nachweisen (*Baumert* und *Watermann,* 2000). Die Sekundarstufe II des deutschen Schulsystems ist jedoch nach unserer Auffassung aufgrund der vielfältigen Wahlmöglichkeiten der Schüler und dem damit verbundenen Selektivitätsproblem für eine Analyse des Effekts von zentralen Abschlussprüfungen nur eingeschränkt geeignet. Wir konzentrieren uns deshalb auf den Effekt von Abschlussprüfungen am Ende der Sekundarstufe I, die interessanterweise in der öffentlichen Debatte zum Thema Abschlussprüfungen eine deutlich geringere Beachtung finden. In *Jürges, Schneider* und *Büchel* (2003) werden die deutschen TIMSS-Daten genutzt, um den kausalen Effekt zentraler Abschlussprüfungen zu schätzen. In diesem Beitrag greifen wir die Ergebnisse auf, modifizieren das Schätzverfahren und erweitern den Katalog der empirischen Fragestellungen.

Unser Beitrag ist wie folgt gegliedert: Abschnitt B beschreibt die Praxis zentraler Abschlussprüfungen in den Bundesländern. Danach behandelt Abschnitt C zunächst die Datenbasis und Fragen der Schätzung und Identifikation; im Anschluss werden die empirischen Ergebnisse dargestellt. Abschnitt D fasst die wichtigsten Ergebnisse kurz zusammen.

B. Zentrale Abschlussprüfungen in Deutschland

Bevor die Wirkung zentraler Abschlussprüfungen empirisch getestet wird, geben wir einen Überblick über die Praxis zentraler und dezentraler Abschlussprüfungen in Deutschland. Die aktuelle Diskussion zentraler Abschlussprüfungen in Deutschland konzentriert sich im Wesentlichen auf das Zentralabitur, also den Abschluss der Sekundarstufe II. Nur selten wird die Form der Abschlüsse der Sekundarstufe I thematisiert, obwohl immerhin etwa zwei Drittel der Schüler der Sekundarstufe I Haupt- und Realschulen besuchen. Bei Abschluss der Sekundarstufe I erhalten die Schüler der Haupt- oder Realschule ein Abschlusszeugnis, sofern sie die 9. oder 10. Klasse erfolgreich abgeschlossen haben. In einigen Bundesländern müssen die Schüler eine zentrale Abschlussprüfung ablegen. Tabelle 1 stellt den Status quo für 1995 dar, dem Erhebungsjahr der von uns ausgewerteten TIMSS-Daten.

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

Tabelle 1: **Zentrale Abschlussprüfungen nach Bundesland und Art des Abschlusses; Anteil Schüler nach Schulform / Art des Abschlusses (Stand von 1995)**

	Hauptschulabschluss	Realschulabschluss	Abschluss Sekundarstufe II (Abitur)	Schüler in nicht zum Abitur führenden Bildungsgängen**	Schüler in der 8. Klasse Haupt- / Realschule
Baden-Württemberg (BW)	D / M / F / A	D / M / F	J	71.8	71.6
Bayern (BY)	D / M / F / N* / A	D / M / F / N* / A	J	73.2	73.1
Berlin (BE)				67.2	57.2
Brandenburg (BB)				71.7	54.8
Bremen (HB)				69.1	63.9
Hamburg (HH)				65.6	56.7
Hessen (HE)				67.2	61.4
Mecklenburg-Vorpommern (MV)		D / M / F	J	72.2	70.7
Niedersachsen (NI)				73.2	72.0
Nordrhein-Westfalen (NW)				69.6	64.7
Rheinland-Pfalz (RP)				72.0	70.5
Saarland (SA)				71.8	66.0
Sachsen (SN)	D / M / F	D / M / N	J	69.9	69.9
Sachsen-Anhalt (ST)		D / M	J	67.6	67.2
Schleswig-Holstein (SH)				73.7	71.5
Thüringen (TH)	D / M	D / M / F	J	67.7	67.4

D = Deutsch; M = Mathematik; F = Fremdsprache (meist Englisch); N = Naturwissenschaften; A = Andere; J = Jedes für die schriftlichen Prüfungen gewählte Fach; * Wahlfach; ** Schüler in der 8. Klasse der Haupt- und Realschule plus zwei Drittel aller Schüler in der 8. Klasse der Gesamtschule.

Quelle: Eigene Recherchen bei den Kultusministerien der Länder.

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

Gymnasialschüler erhalten am Ende der Sekundarstufe I kein Abschlusszeugnis, erwerben aber die Qualifikation zum Besuch der gymnasialen Oberstufe. Am häufigsten sind zentrale Abschlussprüfungen am Ende der Sekundarstufe II. So gab es 1995 in sieben der sechzehn deutschen Bundesländer ein zentrales Abitur (Abschluss der Sekundarstufe II mit Berechtigung zum Hochschulstudium) auf Landesebene. Diese Länder konzentrieren sich im Süden (Baden-Württemberg, Bayern, Saarland) und Osten (Mecklenburg-Vorpommern, Sachsen, Sachsen-Anhalt, Thüringen) Deutschlands. In den anderen Bundesländern gab es dezentrale Systeme, bei denen die Aufgaben für die Abschlussprüfung von den Lehrkräften ausgearbeitet und von der Schulaufsichtsbehörde einzeln genehmigt wurden. In sechs Ländern gab es zentrale Prüfungen zum Realschulabschluss, und in vier Ländern auch zum Hauptschulabschluss.

Abschlussprüfungen umfassen nicht immer alle Unterrichtsfächer. Beim Abitur können die Schüler innerhalb eines bestimmten Rahmens drei oder vier Prüfungsfächer wählen. Bei Prüfungen zum Real- und Hauptschulabschluss sind Deutsch und Mathematik stets Pflichtfächer. Zur Bewertung des Leistungseffekts von zentralen Abschlussprüfungen an der Haupt- und Realschule bildet die Mathematikleistung unsere wichtigste Ergebnisvariable: Die von uns ausgewerteten TIMSS-Daten enthalten Maße für die Mathematikleistung. Andere Fächer bei zentralen Abschlussprüfungen sind Fremdsprachen (meist Englisch) oder – weniger häufig – naturwissenschaftliche Fächer.

C. Untersuchungsansatz und empirische Ergebnisse

Es sollen im Folgenden zwei Hypothesen empirisch getestet werden. Zum einen gilt es zu prüfen, ob Bundesländer mit zentralen Abschlussprüfungen bessere Leistungen in der TIMSS-Studie vorweisen konnten. Um den kausalen Effekt zentraler Abschlussprüfungen zu schätzen, ist zunächst die Frage nach einer angemessenen Identifikationsstrategie zu beantworten. Elementare Datenrestriktionen früherer Studien machten es bislang unmöglich, den kausalen Effekt zentraler Abschlussprüfungen zu identifizieren. Die quasi-experimentelle Natur zentraler Abschlussprüfungen in Deutschland erlaubt uns dagegen eine konsistente Differenzen-in-Differenzen-Schätzung des Effekts. Neben der Frage nach mittleren Effekten zentraler Standards ist aber auch die Wirkung zentraler Standards auf die Leistungsverteilung von politischem Interesse. Ein viel beachtetes Ergebnis international vergleichender Studien ist, dass die Leistungsverteilung in Deutschland eine überdurchschnittlich große Varianz aufweist und dass soziale Unterschiede stärker als in anderen Ländern im deutschen Schulsystem festgeschrieben werden. Deshalb unterscheiden wir bei der Schätzung des Effektes zentraler Standards auch nach dem sozialen Hintergrund.

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

I. Datenbasis: TIMSS-Deutschland

Die TIMSS II-Stichprobe für Deutschland enthält Daten zu fast 6.000 Schülern. Diese wurden im Schuljahr 1994 / 95 in den 7. und 8. Klassen von 137 Schulen erhoben. Die Schüler kamen aus 14 der 16 Bundesländer; Baden-Württemberg und Bremen nahmen nicht teil.[4] Wir verwenden hier eine Unterstichprobe, die nur Schüler in Haupt- und Realschulen (bzw. Mittelschulen) enthält. Diese umfasst insgesamt 3.197 Schüler mit vollständigen Daten aus 168 Klassen in 84 Schulen. Außer den Testergebnissen für Mathematik und Naturwissenschaften enthalten die TIMSS-Daten ein breites Spektrum von Variablen zu sozialem Hintergrund und Einstellungen der Schüler wie auch zu Lehrkräften und Schulen. Trotz der Fülle des vorliegenden Materials werden wir einen eher eng gefassten Ansatz verfolgen und nur eine begrenzte Anzahl Kontrollvariablen zum Hintergrund von Schüler und Schulen auswählen. Dabei beschränken wir uns auf Merkmale mit einem bekannt hohen Zusammenhang mit schulischen Leistungen.

Tabelle 2 enthält Variablendefinitionen und deskriptive Statistiken, aufgeschlüsselt nach der Art der Abschlussprüfung. Im Gegensatz zu Veröffentlichungen, bei denen der internationale Vergleich von Schulleistungen im Mittelpunkt steht, verwenden wir als abhängige Variablen die separat für Deutschland berechneten und von uns zusätzlich auf Mittelwert Null und Standardabweichung Eins standardisierten Rasch-Scores. Die Höhe der geschätzten Regressionskoeffizienten ist damit direkt als Standardabweichung interpretierbar.

Der auffälligste Unterschied zwischen Schülern mit und ohne zentrale Abschlussprüfungen findet sich bei den Leistungen in Mathematik und den naturwissenschaftlichen Fächern (Tabelle 2). Die Werte von Schülern, die zentral geprüft werden, liegen im Durchschnitt um etwa eine Drittel Standardabweichung über denen von Schülern ohne zentrale Prüfungen. Diese Unterschiede spiegeln sich nicht in allen Hintergrundvariablen wider. So ist der Bildungshintergrund, gemessen an der Zahl der im Elternhaus vorhandenen Bücher, von Schülern aus Bundesländern mit und ohne zentralen Abschlussprüfungen sehr ähnlich.[5] In der Gruppe der Schüler in Ländern ohne zentrale Abschlussprüfungen finden sich jedoch weit mehr Kinder mit Migrationshintergrund als in derjenigen, die zentrale Abschlussprüfungen bestehen müssen. Dies erklärt sich vor allem dadurch, dass in Ostdeutschland, wo die meisten Bundesländer zentrale Abschlussprüfungen haben (ein Erbe des früheren DDR-Bildungssystems), nur wenige Ausländer leben.

---

[4] Weitere Informationen zu TIMSS II finden sich unter http: // www.mpib-berlin.mpg.de / TIMSSII-Germany / index.htm (Mai 2003).

[5] Die Verwendung der Anzahl der Bücher im Elternhaus als Indikator für den Bildungshintergrund von Schülern mag auf den ersten Blick merkwürdig anmuten. Diese Variable hat jedoch einen mehr als doppelt so hohen Erklärungsgehalt für die Leistungen der Schüler als der häufig genutzte höchste formale Bildungsabschluss der Eltern.

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

Tabelle 2: **Deskriptive Statistik (gewichtet)**

Variable	Schüler in Ländern ohne zentrale Abschlussprüfungen		Schüler in Ländern mit zentralen Abschlussprüfungen	
	Mittel	StdAbw	Mittel	StdAbw
Leistung in Mathematik (Rasch-Score)	–0.134	0.943	0.213	0.971
Leistung in Naturwissenschaften	–0.114	0.966	0.216	0.969
Geschlecht (1 = Mädchen)	0.486		0.459	
Bücher zu Hause: 0 – 10	0.183		0.188	
Bücher zu Hause: 11 – 25	0.291		0.333	
Bücher zu Hause: 26 – 100	0.168		0.175	
Bücher zu Hause: 101 – 200	0.234		0.209	
Bücher zu Hause: 200+	0.123		0.096	
Immigrantenkind (beide Eltern im Ausland geboren)	0.191		0.077	
Schulform (1 = Realschule)	0.522		0.541	
Klasse (1 = 8. Klasse)	0.498		0.478	
mindestens einmal Klasse wiederholt	0.359		0.246	
Gewöhnlich gut in Mathe: stimme voll zu	0.400		0.335	
Gewöhnlich gut in Biologie: stimme voll zu	0.315		0.358	
Ostdeutschland	0.042		0.455	
Kumulierte Mathematikstunden (in 1000)[a)]	1.1988	0.121	1.292	0.105
N Beobachtungen	1.834		1.363	

Quelle: Eigene Auswertungen von TIMSS-Deutschland – [a)] Quelle: *Frenck* (2001).

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

II. Identifikation und Schätzverfahren

Anhand der deutschen TIMSS-Daten, die neben individueller Schulleistung in einzelnen Fächern auch Informationen zu Bundesland und Schulform enthalten, ist eine Identifikation des Kausaleffekts von zentralen Abschlussprüfungen auf die Schulleistung auf verschiedene Weise möglich. Diese beruhen jeweils auf unterschiedlichen Grundannahmen und weisen entsprechend unterschiedliche Stärken und Schwächen auf.[6] Im folgenden stellen wir einen Differenzen-in-Differenzen-Schätzer vor, der den Einfluss zentraler Abschlussprüfungen über die gesamte Leistungsverteilung misst.

Ein naiver Schätzer für den Effekt zentraler Abschlussprüfungen ist die Berechnung *einfacher Differenzen* zwischen der Durchschnittsleistung in Ländern mit und in Ländern ohne zentrale Abschlussprüfungen (bei Kontrolle des Schülerhintergrunds und anderer einschlägiger Variablen, z. B. der kumulierten Unterrichtszeit, die bis zum Erhebungszeitpunkt auf Mathematik und Naturwissenschaften verwendet wurde). Einfache Differenzen sind allerdings nur von deskriptivem Wert, weil zwei potentiell konfundierende Effekte außer Betracht bleiben. Der eine ist ein Zusammensetzungseffekt: Wie Tabelle 1 zeigt, besuchen in Ländern mit zentralen Abschlussprüfungen im Durchschnitt mehr Schüler die Haupt- und Realschule und weniger Schüler das Gymnasium als in Ländern ohne zentrale Abschlussprüfungen. Unter der Annahme, dass die Zuweisung der Schüler zu den Sekundarschulen hauptsächlich aufgrund ihrer Leistungen in der Primarschule erfolgt, ergibt sich der bemerkenswerte Effekt, dass für jede Schulform die durchschnittliche Schulleistung in Ländern mit zentralen Abschlussprüfungen schon allein deshalb höher ist, weil sich dort in allen Schulformen relativ mehr gute Schüler befinden.[7] Unsere Stichprobe umfasst nur Schüler in Haupt- und Realschulen. In einer früheren Studie haben wir den Zusammensetzungseffekt statistisch relativ grob kontrolliert, indem wir den Haupt- und Realschüleranteil als Kovariaten in die Regression einbezogen haben (*Jürges, Schneider* und *Büchel* 2003). Der Zusammensetzungseffekt erweist sich in den Schätzungen als äußerst wichtige Kontrollvariable. Wir folgen hier einem alternativen Ansatz, indem wir für jeden Schüler in Ländern ohne zentrale Prüfungen die (kontrafaktische) Wahrscheinlichkeit berechnen, in einem Land mit zentralen Prüfungen Schüler der Haupt- oder Realschule zu sein.[8] Dafür schätzen wir jeweils ein Probit-Modell für Schüler in Bundesländern mit zentralen und ohne zentrale Abschlussprüfungen. Aus diesen Schätzungen lässt sich für jeden Schüler in Bundesländern ohne zentrale Abschlussprüfungen, der eine Real- oder Hauptschule besucht, die Wahrscheinlichkeit P

---

6 Eine ausführliche Beschreibung alternativer Schätzverfahren findet sich in *Jürges, Schneider* und *Büchel* (2003).

7 Kritiker dieser Annahme bezweifeln, dass der Wechsel auf eine weiterführende Schule rational erfolgt, und führen an, dass weder die elterliche Entscheidung noch die Grundschulempfehlung sozialschichtenneutral sind.

8 Das Vorgehen folgt *DiNardo, Fortin* und *Lemieux* (1996).

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

berechnen, dass er, gegeben seine beobachtbaren Eigenschaften *x,* in der Teil-Stichprobe der Haupt- und Realschüler enthalten ist.[9] Wir bezeichnen die Wahrscheinlichkeit, dass ein Schüler aus einem Bundesland ohne zentrale Abschlussprüfungen in der Stichprobe ist, als

$$P_1 = P\,(\text{Haupt- oder Realschule} \mid x, \text{keine zentrale Abschlussprüfung}).$$

Ähnlich lässt sich die Wahrscheinlichkeit, dass dieser Schüler in der Stichprobe ist, lebte er in einem Bundesland mit zentralen Abschlussprüfungen, als

$$P_2 = P\,(\text{Haupt- oder Realschule} \mid x, \text{zentrale Abschlussprüfung})$$

prognostizieren. *DiNardo, Fortin* und *Lemieux* (1996) folgend werden die Gewichte der TIMSS-Daten bei Schülern aus Bundesländern ohne zentrale Abschlussprüfungen mit dem Faktor P_2/P_1 multipliziert. Diese prognostizierten Wahrscheinlichkeiten werden als Gewichte in den Schätzungen berücksichtigt. Damit wird die unterschiedliche Zusammensetzung in den Schulformen kontrolliert.

Abbildung 1 zeigt die geschätzten Dichtefunktionen der Mathematik-Testergebnisse. Die Unterschiede in der Verteilung der Testergebnisse sind deutlich. Schüler aus Bundesländern mit zentralen Abschlussprüfungen schneiden besser ab als ihre Altersgenossen aus Ländern ohne Zentralprüfungen. Kontrolliert man den Zusammensetzungseffekt, so erhält man eine kontrafaktische Verteilung, die zwar noch rechts von der Verteilung der Länder ohne Zentralprüfungen liegt, die Verteilung der Länder mit Zentralprüfungen allerdings nicht erreicht. Schüler im oberen Bereich der Leistungsverteilung sind dabei tendenziell stärker betroffen als Schüler im mittleren Bereich.

Neben der genannten unterschiedlichen Zusammensetzung der Schüler in den einzelnen Schulformen kommt es zu einem weiteren Konfundierungseffekt, wenn sich Länder mit zentralen Abschlussprüfungen in nicht beobachteten Aspekten von Ländern ohne zentrale Abschlussprüfungen unterscheiden. So könnte sich beispielsweise die Einstellung der Eltern zu Schulbildung und Schulleistung in den Bundesländern systematisch unterscheiden. Eine zentrale Abschlussprüfung könnte somit auch einfach ein Indikator für die hohe Priorität sein, die Bildung in der Bevölkerung genießt. Insofern wären zentrale Abschlußprüfungen als im statistischen Sinne (zumindest langfristig) endogene Variable zu betrachten. Einfache Differenzen werden daher immer dem Einwand unterliegen, dass sie lediglich eine Korrelation und keinen Kausalzusammenhang abbilden.

Wenden wir uns nun unserer Identifikationsstrategie für den kausalen Effekt zentraler Abschlussprüfungen zu. Tabelle 1 zeigt nicht nur, welche Bundesländer

[9] Kovariaten in der Probit-Schätzung sind: Jahrgangsstufe, Anzahl der Bücher zu Hause, Geschlecht, Sprache im Elternhaus der Schülers, Schüler benötigt gute Mathematiknote, um auf bevorzugte Schule zu wechseln, Schüler lebt mit leiblicher Mutter, Schüler hat mindestens eine Klasse wiederholt, Ostdeutschland.

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

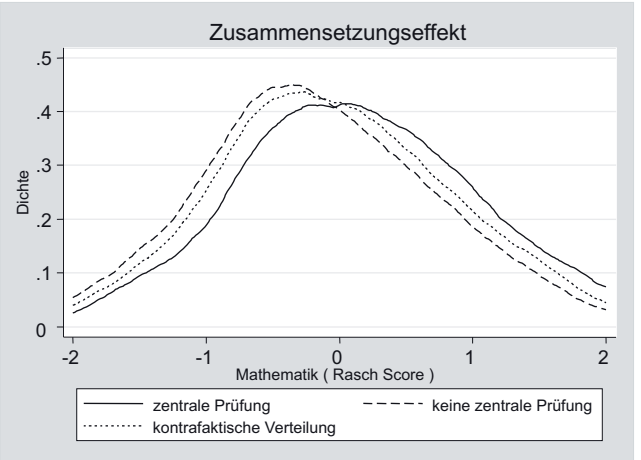


Abbildung 1: Zentrale Abschlussprüfungen nach Bundesland und Schulform

zentrale Abschlussprüfungen in welcher Schulform haben, sondern auch, welche Fächer geprüft werden. Zu beachten ist, dass es zentrale Abschlussprüfungen an Haupt- und Realschulen (falls überhaupt) nur in Deutsch, Mathematik und einer Fremdsprache – meist Englisch – gibt. Die naturwissenschaftlichen Fächer sind bei zentralen Prüfungen nicht vertreten – mit zwei Ausnahmen: Sachsen und, mit Einschränkung, Bayern. Ist Mathematik ein Prüfungsfach bei zentralen Abschlussprüfungen, die Naturwissenschaften aber nicht, so sollte sich der kausale Effekt von zentralen Abschlussprüfungen auf Leistungen in Mathematik beschränken. Da in TIMSS Testergebnisse sowohl für Mathematik als auch für die naturwissenschaftlichen Fächer vorliegen, können wir die Differenzen-in-Differenzen zwischen Fächern schätzen. Der größte Vorteil dieses Schätzers ist, dass jeder einzelne Schüler gewissermaßen seine eigene Kontrollgruppe bildet, das heißt, dass wir in der Lage sind, Heterogenität – zum Beispiel Befähigung allgemein – auf individueller Ebene zu kontrollieren.

Wichtig für die Konsistenz unseres Schätzers ist, dass es keine Spillover-Effekte zwischen den Unterrichtsfächern gibt. So muss zum Beispiel ausgeschlossen sein, dass gute Mathematikkenntnisse Voraussetzung für gutes Abschneiden bei den naturwissenschaftlichen Fragen sind. Ein Blick auf die Testfragen zeigt, dass in der Regel keine oder nur rudimentäre Mathematikkenntnisse nötig sind, um die Fragen aus dem Bereich der Naturwissenschaften zu beantworten. Eine ausführliche Darstellung des Argumentes und die Diskussion weiterer möglicher Einwände gegen die identifizierende Annahme unseres Schätzers findet sich in *Jürges, Schneider* und *Büchel* (2003).

Der Differenzen-in-Differenzen-Schätzer nach Fächern erfordert, dass Sachsen aus der Analyse herausgenommen wird und die bayerische Teilstichprobe beson-

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

ders berücksichtigt werden muss. In Sachsen sind die Naturwissenschaften Pflichtfach für den Realschulabschluss, deshalb schließen wir Sachsen aus diesem Teil unserer Analyse aus.[10] In Bayern dagegen sind die Naturwissenschaften Wahlfach. Dort werden jedes Jahr rund 40 Prozent aller Schüler, die einen Hauptschulabschluss anstreben, in Biologie, Chemie und Physik geprüft, das heißt, in allen Fächern, die auch Gegenstand des TIMSS-Tests für die naturwissenschaftlichen Fächer sind. 20 bis 25 Prozent der Schüler, die einen Realschulabschluss anstreben, legen eine schriftliche Prüfung allerdings nur in Physik ab (etwa ein Drittel der naturwissenschaftlichen Items in TIMSS stammen aus der Physik). Robustheitsprüfungen in *Jürges, Schneider* und *Büchel* (2003) legen jedoch nahe, dass die Ergebnisse durch diese Eigenheit der bayerischen Teilstichprobe nicht verzerrt werden.[11]

Eine weitere Besonderheit der Leistungen deutscher Schüler in internationalen vergleichenden Studien ist die große Varianz in den Ergebnissen. Insbesondere die schwachen Schüler werden im deutschen Schulsystem nicht so intensiv gefördert wie in anderen Ländern. Soziale Unterschiede werden in Deutschland durch das Schulsystem zementiert; insbesondere stellt die mangelnde Integration von Ausländerkindern ein großes Problem dar. Theoretisch ist die Wirkung zentraler Abschlüsse auf die Breite des Leistungsspektrums nicht eindeutig. Eine immer wieder vorgebrachte Kritik an Zentralstandards ist, dass gleiche Maßstäbe für alle den individuellen Fähigkeiten und Bedürfnissen der einzelnen Schüler nicht angemessen Rechnung tragen und damit für einige Gruppen von Schülern zu suboptimalen Ergebnissen führen. Zu hohe Standards könnten einen demotivierenden Effekt auf lernschwächere Schüler haben, da sich die Ministerien nicht an der Leistungsintensität der weniger leistungsstarken Schüler orientieren, sondern am durchschnittlichen Schüler. Mit Blick auf die Gruppe der leistungsstärksten Schüler könnte vermutet werden, dass die Zentralstandards zu gering sind, um den Schülern optimale Lernanreize zu bieten. Zur Überprüfung dieser Hypothese werden wir Differenzen-in-Differenzen-Schätzer nach Schülergruppen vorstellen, und zwar differenziert nach Bildungshintergrund, Migrationshintergrund und Geschlecht.

Nun zur Frage des Signalwerts von Noten in Schulsystemen mit und ohne Zentralstandards. Zwar haben Schüler mit guten Leistungen im Schnitt bessere Noten als Schüler mit schlechten Leistungen, doch kann sich die mit einer bestimmten Schulnote verbundene Leistungsverteilung innerhalb einer Schulform deutlich überlappen, d. h. das Signal „Schulnote“ wird durch starkes „Rauschen“ beeinträchtigt. Dies konnten zuletzt *Baumert, Trautwein* und *Artelt* (2003) als Ergebnis

---

10 Ein alternatives Vorgehen wäre, alle sächsischen Schüler mit einer Dummy-Variable zu versehen, was nur zu geringfügigen Veränderungen unserer Ergebnisse führt; die Ergebnisse dieser Variante werden nicht tabellarisch dokumentiert.

11 Da Bayern das einzige westdeutsche Bundesland mit zentralen Prüfungen in Haupt- und Realschulen in der Stichprobe ist, ist es problematisch, Bayern komplett aus der Stichprobe zu nehmen, wenn mit dem Zentralprüfungseffekt auch ein Ostdeutschlandeffekt geschätzt werden soll.

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

der PISA-E-Vertiefungsstudie zeigen. Wir überprüfen im folgenden die These, dass in Ländern mit zentralen Abschlussprüfungen Noten eher schulische Leistung widerspiegeln als in Ländern ohne zentrale Abschlussprüfungen; dies ist eine entscheidende Erweiterung dieses Beitrags gegenüber der Vorläuferstudie *Büchel, Jürges* und *Schneider* (2003). Wir schließen damit zugleich an frühere Studien an, die gezeigt haben, dass dezentrale Abschlüsse im Vergleich zu zentralen am deutschen Arbeitsmarkt stark an Wert verloren haben (*Backes-Gellner* und *Weckmüller,* 1998).

Zwar sind Noten in der TIMSS-Studie nicht direkt erhoben worden, die Schüler wurden jedoch gefragt, ob sie gewöhnlich gute Leistungen in der Mathematik erbringen. Da die Selbsteinschätzung der Schüler im Wesentlichen von der Benotung ihrer Leistung abhängen wird, verwenden wir die Selbsteinschätzung der Schüler als Indikator für das Signal, das von der Benotung ausgeht. Ein Vergleich geschätzter Dichtefunktionen nach Selbsteinschätzung sowie die Berechnung der statistischen Erklärungskraft der Selbsteinschätzung für die tatsächliche Leistung geben Aufschluss darüber, ob Noten in Ländern mit zentralen Standards einen höheren Signalwert haben als in anderen Ländern.

III. Empirische Ergebnisse

Unsere Schätzergebnisse sind in Tabelle 3 zusammengefasst. Spalte (1) enthält einen Schätzwert für die einfache Differenz in der Mathematikleistung von Schülern mit und Schülern ohne zentrale Abschlussprüfung in Mathematik. Die Differenz beträgt immerhin 0,573 Standardabweichungen, d. h. mehr als das Äquivalent zweier Schuljahre (0,243 Standardabweichungen). Man beachte, dass dabei Herkunftseffekte bereits kontrolliert werden. Bei *Bishops* (1997) Vergleich der kanadischen Provinzen mit und ohne zentrale Abschlussprüfungen beträgt der Effekt zentraler Abschlussprüfungen, wie bereits berichtet, nur rund zwei Drittel des Äquivalents einer Jahrgangsstufe.

Alle Hintergrundvariablen haben die erwarteten Effekte auf die Mathematikleistungen der Schüler. Die Leistungen der Realschüler liegen weit über denen der Hauptschüler. Die Zahl der zu Hause vorhandenen Bücher wird als Stellvertretervariable für den Bildungshintergrund der Eltern benutzt. Tatsächlich beträgt die Leistungsdifferenz zwischen Schülern aus Haushalten mit weniger als zehn Büchern und Schülern aus Haushalten mit über 200 Büchern mehr als ein Schuljahr. Immigrantenkinder schneiden etwas schlechter ab als andere, ebenso Kinder, die schon einmal eine Klasse wiederholt haben. Mädchen erbringen im Durchschnitt schlechtere Leistungen in Mathematik als Jungen,[12] westdeutsche Schüler bessere als ostdeutsche. In Spalte (2) kontrollieren wir den Zusammensetzungs-

[12] Zur detaillierten Analyse geschlechtstypischer Unterschiede in der Lösung von TIMSS-Aufgaben vgl. *Mullis* et al. (2000).

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

Tabelle 3

Effekte zentraler Abschlussprüfungen auf Schulleistung in Mathematik

	Einfache Differenzen[a)]	Einfache Differenzen, Zusammensetzungseffekt berücksichtigt[a)]	Diff-in-Diff zwischen Fächern[b)]	Diff-in-Diff zwischen Fächern, Zusammensetzungseffekt berücksichtigt[b)]
	(1)	(2)	(3)	(4)
Zentrale Abschlussprüfung	0.573 (8.08)**	0.544 (7.54)**	0.123 (2.93)**	0.107 (2.53)*
Realschule	0.773 (12.31)**	0.758 (11.69)**	0.138 (3.27)**	0.124 (2.86)**
Klasse	0.243 (2.90)**	0.221 (2.62)**	–0.005 (0.14)	–0.015 (0.37)
11 – 25 Bücher	0.112 (1.82)	0.083 (1.38)	–0.116 (2.10)*	–0.119 (2.16)*
26 – 100 Bücher	0.270 (4.38)**	0.252 (4.20)**	–0.132 (2.25)*	–0.134 (2.28)*
101 – 200 Bücher	0.387 (5.59)**	0.380 (5.52)**	–0.132 (2.00)*	–0.123 (1.82)
200+ Bücher	0.437 (7.07)**	0.420 (6.79)**	–0.160 (2.72)**	–0.153 (2.51)*
Mädchen	–0.279 (7.12)**	–0.275 (6.99)**	0.090 (2.58)*	0.083 (2.44)*
Immigrantenkind	–0.201 (3.08)**	–0.172 (2.63)**	0.275 (4.92)**	0.299 (5.26)**
Klasse wiederholt	–0.110 (2.87)**	–0.100 (2.57)*	–0.066 (1.68)	–0.070 (1.80)
Osten	–0.803 (8.73)**	–0.789 (8.67)**	–0.294 (3.68)**	-0.272 (3.46)**
Kum. Mathematikstunden	0.274 (0.68)	0.388 (0.96)		
Rel. kum. Mathematikstunden			0.590 (0.60)	0.744 (0.76)
Konstante	–0.639 (1.74)	–1.099 (2.34)*	–0.176 (0.71)	–0.194 (0.78)
Beobachtungen	2995	2928	2995	2928
R-squared	0.34	0.34	0.04	0.04

Robuste *z*-Werte (korrigiert um den Effekt der Stichprobenklumpung auf Klassenebene) in Klammern; * signifikant auf 5%-Niveau; ** signifikant auf 1%-Niveau; [a)] abhängige Variable: Leistung in Mathematik; [b)] abhängige Variable: Differenz der Leistungen in Mathematik minus Naturwissenschaften.

Quelle: Eigene Auswertungen von TIMSS-Deutschland.

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

effekt durch Gewichtung der Schüler mit den oben beschriebenen kontrafaktischen Wahrscheinlichkeiten. Dadurch sinkt der Koeffizient der Zentralprüfungsvariable leicht auf 0,544. Die anderen Koeffizienten verändern sich ebenfalls nur leicht.

Die in den Spalten (1) und (2) wiedergegebene Effekte zwischen zentraler Abschlussprüfung und Schulleistung könnten auch auf nicht beobachtbare Variablen zurückzuführen sein, die mit Prüfungsform und Leistung korrelieren. Um diese Korrelation von einem Kausalzusammenhang zu unterscheiden, wenden wir uns nun der von uns bevorzugten Schätzfunktion zu, nämlich Differenzen-in-Differenzen zwischen Fächern. Wie bereits erwähnt, hat diese Schätzfunktion den großen Vorteil, dass jeder Schüler quasi als seine eigene Kontrollgruppe fungiert, weil die zentrale Prüfung nur in Mathematik, nicht aber in den Naturwissenschaften abgelegt wird. Wir haben diese Schätzfunktion implementiert, indem wir für jeden Schüler die Differenz zwischen seinen Testergebnissen in Mathematik und Naturwissenschaften berechnet und diese Differenz einer Regressionsanalyse mit den üblichen erklärenden Variablen unterzogen haben. Die einzige Änderung in den Kovariaten liegt darin, dass wir – da wir nun relative Leistungen vergleichen – statt der absoluten Zahl der Mathematikstunden das Verhältnis zwischen den kumulierten Mathematikstunden und einer Stellvertretervariablen für das Volumen erteilter Stunden in naturwissenschaftlichen Fächern benutzen.[13] Die Schätzwerte sind in den Spalten (3) und (4) von Tabelle 3 aufgeführt. Dabei fällt als erstes auf, dass der Koeffizient für zentrale Abschlussprüfungen positiv und signifikant bleibt. Er fällt jedoch deutlich auf 0,123 Standardabweichungen bzw. auf die Hälfte eines Schuljahres. Die Schüler in Ländern mit zentralen Abschlussprüfungen weisen in Mathematik relativ bessere Leistungen auf als in den naturwissenschaftlichen Fächern. Dies ist mit der Hypothese vereinbar, dass zentrale Prüfungen die Schulleistung verbessern. Kontrolliert man den Zusammensetzungseffekt in Spalte (4), so verringert sich der Schätzwert für den Effekt von zentralen Abschlussprüfungen auf 0,107, also auf etwas weniger als ein halbes Schuljahr. Damit bestätigen sich die qualitativen Ergebnisse der Analyse internationaler Datensätze, der Zentralprüfungseffekt ist positiv und signifikant. In unserer Studie fällt er jedoch geringer aus.

Obwohl unser Interesse vorrangig der Variable „Zentrale Abschlussprüfung“ gilt, sind auch die Schätzwerte für andere Parameter in den Spalten (3) und (4) bemerkenswert. Die Dummyvariable „Realschule (vs. Hauptschule)“ ist positiv und signifikant, also weisen Realschüler relativ zu Hauptschülern in Mathematik relativ bessere Leistungen auf als in den Naturwissenschaften. Die „Anzahl Bücher“-Koeffizienten sind negativ und nehmen mit der Zahl der im Elternhaus vorhandenen Bücher leicht an Effektstärke zu. Somit erbringen Schüler aus Haushalten mit höherem Bildungsstand in den Naturwissenschaften relativ bessere Leistungen als andere. Dies überrascht kaum, da vieles von dem, was TIMSS im Be-

[13] Diese Stellvertretervariable besteht aus der Gesamtzahl der Unterrichtsstunden minus Mathematikstunden minus Deutschstunden. Informationen nur zu den Unterrichtsstunden in den naturwissenschaftlichen Fächern lagen nicht vor.

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

reich Naturwissenschaften abfragt, auch als Allgemeinwissen bezeichnet werden kann. Kinder aus Immigrantenfamilien sind in Mathematik relativ besser als deutsche, was auf eine häufig unzureichende Beherrschung der deutschen Sprache zurückzuführen sein dürfte, die sie in der Mathematik weniger einschränkt als in den Naturwissenschaften.

In Tabelle 4 differenzieren wir den Zentralprüfungseffekt nach Bildungshintergrund, Migrationsstatus und Geschlecht. Bei der Differenzierung nach Bildungshintergrund fällt auf, dass der Schätzwert der Variable „Zentrale Abschlussprüfung" stark variiert. Der stärkste positive Effekt ist für die Gruppe der Schüler zu vermerken, in deren Haushalt weniger als 10 Bücher vorhanden sind. Für die Gruppe derer, die 11 – 25 Bücher haben, verschwindet der Effekt allerdings vollständig. Der Schätzwert steigt dann mit der Zahl der Bücher, fällt aber für Haushalte mit mehr als 200 Bücher wieder ab. Ein Test auf Gleichheit der Effekte über alle Gruppen hinweg lehnt indes die Nullhypothese nicht ab. Von einer Begünstigung von Schülern einer bestimmten sozialen Herkunft kann also nicht gesprochen werden. Zwar lässt Spalte (2) vermuten, dass Ausländerkinder im Gegensatz zu deutschen Kindern nicht von zentralen Abschlussprüfungen profitieren können, doch ist der Anteil an Ausländerkindern in Ländern mit zentralen Prüfungen mit weniger als 4 Prozent sehr gering, so dass auch dieser Unterschied nicht statistisch signifikant und mit Vorsicht zu interpretieren ist. In Spalte (3) unterscheiden wir zwischen Jungen und Mädchen. Auch hier zeigt sich ein deutlicher Unterschied in der Effektstärke, der jedoch nicht statistisch signifikant ist.

Abschließend vergleichen wir die Korrespondenz von Leistung in der TIMSS-Mathematik-Studie mit der Selbsteinschätzung der Schüler. Dazu haben wir erneut Kerndichten geschätzt. Die Testergebnisse wurden dabei, um unmittelbare Vergleichbarkeit herzustellen, innerhalb jeder Gruppe standardisiert. Abbildung 2 zeigt oben die Kerndichten der Mathematik-Testergebnisse von Schülern aus Bundesländern ohne zentrale Abschlussprüfungen und unten die entsprechenden Dichten für Schüler aus Ländern mit zentralen Abschlussprüfungen. Die Abbildung bestätigt den Befund von *Baumert, Trautwein* und *Artelt* (2003). Die Überlappung der Dichtefunktionen ist deutlich in beiden Abbildungen zu sehen. Noten bzw. die Selbsteinschätzung der Schüler sagen nur in der Tendenz etwas über die Leistungsfähigkeit von Schülern aus. Dies trifft für alle Bundesländer zu. Vergleicht man jedoch die obere und die untere Abbildung, so wird deutlich, dass insbesondere in Ländern ohne zentrale Abschlussprüfungen eine Differenzierung von Leistung nach Selbsteinschätzung, abgesehen von sehr schlechten Schülern, kaum möglich ist. In Ländern mit zentralen Prüfungen entspricht das Selbstbild der Schüler in stärkerem Maße ihrer tatsächlichen Leistungsstärke. Unter der plausiblen Annahme, dass die Selbsteinschätzung stark mit den Schulnoten korreliert ist – worauf sonst sollte sich die Selbsteinschätzung der Schüler stützen? –, legt dies den Schluss nahe, dass Noten dann ein zuverlässigeres Signal auf dem Arbeitsmarkt darstellen, wenn sie unter Zentralstandard-Regimes erteilt werden, da sie stärker mit dem tatsächlichen Leistungsvermögen der Schüler korrespondieren.

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

Tabelle 4

Interaktionseffekte zentraler Prüfungen mit verschiedenen Kovariaten

	Bildungs-hintergrund	Migranten-status	Geschlecht
	(1)[b),c)]	(2)[b),c)]	(3)[b),c)]
< 10 Bücher*ZA[d)]	0.291 (3.26)**		
11 – 25 Bücher*ZA	–0.004 (0.05)		
26 – 100 Bücher*ZA	0.082 (1.26)		
101 – 200 Bücher*ZA	0.202 (2.54)*		
> 200 Bücher*ZA	0.079 (1.08)		
Immigrantenkind*ZA		–0.028 (0.23)	
Nicht-Immigrantenkind*ZA		0.128 (2.95)**	
Junge*ZA			0.154 (2.83)**
Mädchen*ZA			0.047 (0.84)
F-Test auf Gleichheit der Koeffizienten	n.s.	n.s.	n.s.
Beobachtungen	2928	2928	2928
R-Quadrat	0.05	0.04	0.05

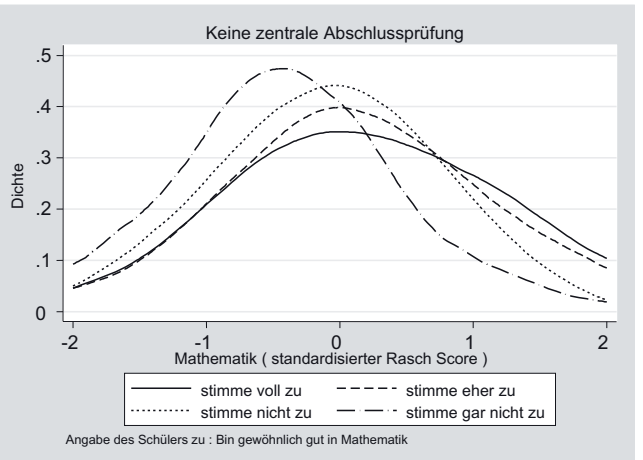
Robuste *z*-Werte (korrigiert um den Effekt der Stichprobenklumpung auf Klassenebene) in Klammern; * signifikant auf 5 %-Niveau; ** signifikant auf 1 %-Niveau; [b)] abhängige Variable: Differenz der Leistungen in Mathematik minus Naturwissenschaften; [c)] Zusammensetzungseffekte sind in der Gewichtung berücksichtigt; [d)] ZA: Zentrale Abschlussprüfungen. Kontrollvariablen: Schulform, Jahrgangsstufe, Anzahl Bücher, Geschlecht, Migrationsstatus, Klasse wiederholt, Neue Bundesländer, rel. kumulierte Mathematikstunden.

Quelle: Eigene Auswertungen von TIMSS-Deutschland.

Tabelle 5 schließlich quantifiziert den Zusammenhang von Leistungsfähigkeit und Selbsteinschätzung in zentralen und dezentralen Systemen. Die aufgeführten Werte geben die statistische Erklärungskraft von Regressionen der Mathematikleistung in TIMSS auf die Selbsteinschätzung der Schüler, getrennt nach Prüfungsform, wieder. Es zeigt sich, dass die Selbsteinschätzung der Schüler in Bundesländern mit zentralen Prüfungen die mittlere Leistungsstärke besser vorhersagt als in Bundesländern ohne zentrale Prüfungen. Betrachtet man erneut das gesamte Leistungsspektrum, so ergibt sich der interessante Befund, dass die dezentralen Prüfungen an dessen unterem Ende so gut wie keine Rückschlüsse vom Selbstbild

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

auf die Leistung ermöglichen, d. h. dass insbesondere schlechte Schüler über ihre tatsächliche Leistung im Unklaren sind. Umgekehrt gilt für sehr gute Schüler, dass sie ihre Leistung in dezentralen Prüfungssystemen präziser einschätzen können als in zentralen.[14]



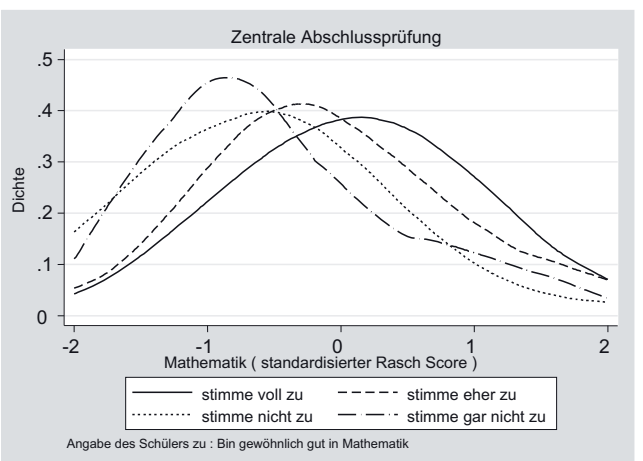


Abbildung 2: Leistungsverteilung in Mathematik nach Selbsteinschätzung und Prüfungsform

[14] Zum Erklärungswert von Noten für die Leistung in der PISA-Studie vgl. *Baumert, Trautwein* und *Artelt* (2003).

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

Tabelle 5

Vorhersagekraft der Selbsteinschätzung für die tatsächliche Leistung

	ohne zentrale Abschlussprüfung	mit zentraler Abschlussprüfung
R-Quadrat	0,034	0,051
Pseudo-*R*-Quadrat 10-Prozent-Quantil	0,008	0,045
Pseudo-*R*-Quadrat 25-Prozent-Quantil	0,009	0,035
Pseudo-*R*-Quadrat 50-Prozent-Quantil	0,016	0,029
Pseudo-*R*-Quadrat 75-Prozent-Quantil	0,029	0,023
Pseudo-*R*-Quadrat 90-Prozent-Quantil	0,035	0,016

Quelle: Eigene Auswertungen von TIMSS-Deutschland.

D. Resümee

Dieser Beitrag geht zwei Fragen nach, die in der aktuellen Diskussion um zentrale Prüfungen im deutschen Schulsystem relevant sind. Die erste Frage lautet, ob Standards in Form zentraler Abschlussprüfungen positive Auswirkungen auf die Schulleistung haben und wie sich die Effekte für verschiedene Untergruppen der Stichprobe unterscheiden. Eine zweite Frage, die sich in diesem Zusammenhang stellt, ist die, ob Noten in einem System mit zentralen Prüfungen einen höheren Signalwert haben als in einem System ohne solche Prüfungen. Da Noten in der TIMSS-Studie nicht erhoben wurden, müssen wir auf eine Proxy-Variable, nämlich die Selbsteinschätzung der Schüler zurückgreifen. Mit dieser zweiten Fragestellung erweitern wir den Fokus der Basisstudie von Büchel, Jürges und Schneider (2003).

Theoretisch sind positive Leistungseffekte von zentralen Prüfungen zu erwarten, und der Signalwert von Noten sollte höher sein, wenn Leistungen zentral geprüft werden. Empirisch jedoch ist die Identifikation eines Kausaleffekts zentraler Prüfungen nicht trivial. Anders als frühere Studien nutzen wir die institutionelle Variation zwischen den deutschen Bundesländern, um eine Identifikationsstrategie für die Schätzung des Kausaleffekts von zentralen Abschlussprüfungen auf die Schulleistung zu entwickeln. Im deutschen Schulsystem haben nur einige Länder zentrale Abschlussprüfungen, meist in den Kernfächern Deutsch und Mathematik. Wir verwenden Daten aus der deutschen TIMSS-Studie von Haupt- und Realschülern in der Sekundarstufe I (Klassenstufen 7 und 8). Beim einfachen Vergleich der Testergebnisse in Mathematik schneiden Schüler in Bundesländern mit zentralen Abschlussprüfungen deutlich besser ab als in Bundesländern ohne entsprechende Prüfungen. Der Unterschied beträgt selbst bei Kontrolle wichtiger Hintergrundvariablen etwa eine halbe Standardabweichung bzw. umgerechnet rund zwei Schuljahre. Ein solcher Unterschied besteht jedoch in geringerem Maße auch für die Testergeb-

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

nisse in naturwissenschaftlichen Fächern, die meist keine Bestandteile zentraler Abschlussprüfungen sind. Unsere bevorzugte Schätzfunktion (Differenzen-in-Differenzen zwischen Fächern) interpretiert die in TIMSS beobachtbaren Unterschiede zwischen Ländern mit und ohne Zentralprüfung in der Differenz zwischen Schülerleistungen in Mathematik und Naturwissenschaften als den Kausaleffekt der fachspezifischen Abschlussprüfungen auf die Schulleistung.

Bei der Schätzung ist außerdem zu berücksichtigen, dass in Ländern mit zentralen Abschlussprüfungen im Durchschnitt mehr Schüler die Haupt- und Realschule und weniger Schüler das Gymnasium besuchen als in Ländern ohne zentrale Abschlussprüfungen. Unsere Stichprobe umfasst nur Schüler in Haupt- und Realschulen. Deshalb muss der Zusammensetzungseffekt angemessen kontrolliert werden. Anders als in *Jürges, Schneider* und *Büchel* (2003) schätzen wir hier (kontrafaktische) Wahrscheinlichkeiten, in einem Land mit zentralen Prüfungen Schüler der Haupt- oder Realschule zu sein und gewichten die Beobachtungen mit diesen Wahrscheinlichkeiten.

Unter der Annahme, dass es keine externen Effekte zwischen den Leistungen in Mathematik und den naturwissenschaftlichen Fächern gibt, wird ein durchschnittlicher Kausaleffekt von zentralen Abschlussprüfungen auf die Mathematikleistung von etwa 0,1 Standardabweichungen oder etwas weniger als das Äquivalent eines halben Schuljahres ermittelt. Differenziert man die Effekte weiter nach sozioökonomischem Hintergrund der Schüler, so zeigt sich, dass Mädchen tendenziell weniger von zentralen Prüfungen profitieren als Jungen, Ausländer weniger als Deutsche, und Kinder mit mittlerem Bildungshintergrund weniger als solche mit entweder sehr schwachem oder sehr gutem. Statistisch signifikant ist indes keiner dieser Unterschiede.

Die Abweichung zwischen den zunächst ermittelten Leistungsunterschieden zwischen Ländern mit und Ländern ohne zentrale Abschlussprüfungen und dem, was wir als den Kausaleffekt solcher Abschlussprüfungen identifizieren, ist recht groß. Eine Interpretation der beobachteten Differenzen zwischen Ländern mit und ohne zentrale Abschlussprüfungen als Effekt allein der Prüfungsform erscheint entsprechend problematisch. Die Korrelation zwischen zentralen Abschlussprüfungen und Schulleistung dürfte zu einem erheblichen Teil (wenn auch nicht vollständig) auf unbeobachtete Unterschiede zwischen den Bundesländern zurückgeführt werden.

Unsere Studie bestätigt zudem, dass Noten besser mit Leistungen korrespondieren, wenn Standards zentral definiert werden. Sind Noten ein besseres Signal auf dem Arbeitsmarkt, so lässt dies positive Anreizeffekte auf die Lernanstrengung von Schülern und somit auf den Lernerfolg erwarten. Insgesamt lassen unsere Ergebnisse erwarten, dass die Einführung zentraler Abschlussprüfungen zu einem signifikanten, nennenswerten Anstieg der Schulleistung führen würde, auch wenn damit die beobachtete Differenz zwischen Ländern mit und Ländern ohne zentrale Abschlussprüfungen nicht vollständig aufgehoben werden dürfte. Im Vergleich zu

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

anderen diskutierten Politikmaßnahmen ist die Einführung zentraler Prüfungen mit geringeren Kosten verbunden. Die Bildungspolitik sollte deshalb zentrale Abschlussprüfungen in allen Bundesländern oder besser noch zentrale Prüfungen zum Ende bestimmter Schuljahre einführen. Sie würde damit sowohl Schülern als auch Lehrkräften stärkere Anreize geben, ein höheres Leistungsniveau zu erreichen.

Literatur

Backes-Gellner, U. / *Weckmüller,* H. (1999): Ist das Ende der Hauptschule aufzuhalten? In Weizsäcker, R. v. (Hrsg.), Deregulierung und Finanzierung des Bildungswesens. Berlin: Duncker & Humblot, 49 – 77.

Baumert, J. / *Trautwein* / *Artelt,* C. (2003): Schulumwelten – institutionelle Bedingungen des Lehrens und des Lernens. In Baumert, J. / Artelt, C. / Klieme, E. / Neubrand, M. / Prenzel, M. / Schiefele, U. / Schneider, W. / Tillmann, K.-J. / Weiß, M. (Hrsg.), PISA 2000 – Ein differenzierter Blick auf die Länder der Bundesrepublik Deutschland. Opladen: Leske + Budrich.

Baumert, J. / *Watermann,* R. (2000): Institutionelle und regionale Variabilität und die Sicherung gemeinsamer Standards in der gymnasialen Oberstufe. Kap. 8 in Baumert, J. / Bos, W. / Lehmann, R. (Hrsg.), TIMSS – Mathematische und physikalische Kompetenzen am Ende der gymnasialen Oberstufe. Opladen: Leske+Budrich.

Bishop, J. H. (1997): The effect of national standards and curriculum-based exams on achievement. American Economic Review 87, 260 – 264.

Bishop, J. H. (1999): Are national exit examinations important for educational efficiency? Swedish Economic Policy Review 6, 349 – 401.

Büchel, F. / *Jürges,* H. / *Schneider,* K. (2003): Die Auswirkungen zentraler Abschlussprüfungen auf die Schulleistung – Quasi-experimentelle Befunde aus der deutschen TIMSS-Stichprobe. Vierteljahrshefte zur Wirtschaftsforschung 72, 238 – 251.

Costrell, R. M. (1994): A simple model of educational standards, American Economic Review 84, 956 – 971.

Costrell, R. M. (1997): Can educational standards raise welfare?, Journal of Public Economics 65, 271 – 293.

DiNardo, J. / *Fortin,* N. M. / *Lemieux,* T. (1996): Labor market institutions and the distribution of wages, 1973 – 1992: A semiparametric approach, Econometrica 64, 1001 – 1044.

Effinger, M. R. / *Polborn,* M. K. (1999): A model of vertically differentiated education, Journal of Economics 69, 53 – 69.

Frenck, I. (2001): Stundentafeln der Primar- und Sekundarstufe I im Ländervergleich – eine empirische Studie am Beispiel der Fächer Deutsch und Mathematik. Unveröffentlichte Magisterarbeit, Fachbereich Erziehungswissenschaft, Universität Essen.

Hoxby, C. M. (2000): Does competition among public schools benefit students and taxpayers, American Economic Review 90, 1209 – 1239.

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

Jürges, H. / *Schneider,* K. (2004): International differences in student achievement: An economic perspective. German Economic Review 5, 105 – 128.

Jürges, H. / *Schneider,* K. / *Büchel,* F. (2003): The effect of central exit examinations on student achievement: Quasi-experimental evidence from TIMSS Germany. CESifo Working Paper No 939.

Mullis, I. V. S. / *Martin,* M. O. / *Fierros,* E. G. / *Goldberg,* A. L. / *Stemler,* S. E. (2000): Gender Differences in Achievement. International Study Center, Lynch School of Education, Boston College.

Wößmann, L. (2002): Central exams improve educational performance: International Evidence. Kieler Diskussionsbeiträge 397, Institut für Weltwirtschaft, Kiel.

Wößmann, L. (2003): Zentrale Prüfungen als „Währung“ des Bildungssystems: Zur Komplementarität von Schulautonomie und Zentralprüfungen. Vierteljahrshefte zur Wirtschaftsforschung 72, 220 – 237.

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

Sibling Rivalry in Education – an Empirical Investigation for Switzerland

By *Stefan C. Wolter* and *Maja Coradi Vellacott,* Aarau

A. Introduction[1]

In the psychological, sociological and economic literature of the past four decades it has been argued, that the educational achievement of children and thus their well being in their adulthood depends to a great degree on the interactions within the family, in which the child is raised.[2] These interactions concern the ones between the child and parents as well as with brothers and sisters. Thus the size of the family, the birth-order, the sibling-gender-composition as well as the child spacing (because of the sequential nature of births) have to be taken into account when analysing the effects of these interactions within families on the educational achievement.

One of the most discussed interactions is the potential competition of siblings for the financial and time budget of their parents. In the case of budget constraints in time and money, everything else equal, more children will dilute the limited resources of the family. Hence, the larger the family the lower the parental investments in each child and the lower the educational outcome.

Recent studies have shown, with the exception of the USA, that family size effects are small or even insignificant, whereas older studies showed significant size effects. This development is explained partially by the fact that families became much smaller over time, education is free in most of the industrialised countries and that remaining budget constraints are fought with distributional policies that target larger families with child allowances, free child care and other measures.

---

[1] The authors would like to thank the Swiss Federal Office for Statistics and the OECD for the provision of data. We also thank Rainer Winkelmann for helpful suggestions on the first draft of this paper and John Bennett and Greg Wurzburg from the OECD for valuable comments and discussions. The paper circulated previously under the title "Sibling rivalry: A look at Switzerland with PISA Data" as an IZA discussion paper (No. 594). We also thank discussants at the International PISA Conference in Berlin and the TEAM research seminary at the Sorbonne University, both in November 2002. The authors gratefully acknowledge helpful advices from two anonymous referees and the editors. The usual disclaimer holds.

[2] See e.g. *Haveman & Wolfe* (1993).

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

For educational policy, a remaining negative effect of the size of a family on the educational outcomes of their children would therefore be a serious challenge. It would mean that some parents still face binding budget constraints and that their children do not have the same chances in the educational system as children from smaller families. Public investment in the human capital of children in families subject to capital constraints could therefore be needed to restore equity.

This paper attempts to shed light on the family size effect on educational achievement using Swiss PISA data. The paper consists of six sections. After the introduction, section B gives a short overview on the past literature and situates the present work, section C describes the data used and section D presents the empirical results. Section E revisits the sibling rivalry theory and tries to present an extension to the classical model that could better explain the results found in this paper. Section F concludes.

B. Literature and Hypothesis

Although not only economists have looked at the effects of family size on children's' outcomes, the approach of *Becker* (1960) to the economics of the family has influenced a great number of empirical studies in the past four decades. In his initial work on fertility decisions he introduced the notion of "quality of children" and presented the decision of families on the number of children they would have as a joint decision about quantity and quality of their descendants. Whereas this initial work was more interested in explaining the pattern of fertility in the twentieth century, Becker also discussed the implications for the investment (private and public) in human capital and the potential need for redistributional policies (see e.g. *Becker & Tomes,* 1986).

The size of families was only one argument within the literature on the sibling effect. The gender composition of the sibship, the birth-order, and the spacing of births were also looked at.

In the case of gender composition it was assumed, that if the marginal rates of return to investments of parents in their children were not equal, they would not invest the same amount of money and time in each of their children. As most labour markets showed a certain degree of wage discrimination against women, it was assumed, that on the basis of this, one could expect that girls would be treated differently than boys. In this case, girls growing up with brothers would – everything else equal – have benefited less from parental investment than girls growing up with sisters only.[3] However the direction of these effects is theoretically un-

[3] Gender composition effects were found in the studies of *Butcher & Case* (1994) or *Garg & Morduch* (1998), while *Kaestner* (1997), *Hauser & Kuo* (1998) found no significant gender composition effects in their studies. *Bauer & Gang* (2000) found no composition effects for households with two children and some effects for households with more than two

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

clear. If parents pursue achievement maximisation, they would unevenly distribute their wealth and time, whereas if they were lead by fairness, they would tend to invest more in the disadvantaged children in order to equalise outcomes.

Spacing and birth-order are two additional aspects of sibling configuration that could make a difference in parental investment. Close spacing of children reduces the moments in which parents can devote all their time to a single child, whereas regarding the birth-order the same happens for the so-called "middle" child(ren). The oldest and the youngest child are both likely (depending on the spacing) to profit from a period where they are the single child, at least in the pre-school period or later in life[4]. As the investment of parents in their children in the pre-school period proves to be crucial, this could make a significant difference in the impact on the educational outcome.

Regarding the educational outcome – the dependent variable in all the empirical analysis – three different measures are commonly used.[5] The most straightforward measure is educational achievement, measured in school tests[6] or like in this study in comparative tests of competencies. The second measure, widely used in the empirical literature, is educational attainment, usually measured by grades or completed school years or levels.[7] Thirdly and evidently for economists, the impact on wages as an educational outcome can also be tested.[8] In the ideal case, the three outcomes would be linked with almost complete correlation and the choice of the dependent variable would not make any difference. However, we are well aware of the fact, that due to many exogenous (and endogenous) factors, high achievers in school tests are not attaining automatically higher school levels or stay longer in school and educational attainment is not always reflected in higher salaries. In order to test the effect of family size and configuration on educational outcomes we would expect to see these – if any – in the school and test performance of pupils.

---

children. Another strand of literature tests the differences of parental influence on their children when taking the gender effect into account. They look specifically at mother-daughter or father-son relationships (e.g. *Mare & Chen,* 1986a, b) (and vice versa), generally in the framework of intergenerational income mobility (e.g. *Chadwik & Solon,* 2002).

4 The youngest child could additionally profit from the fact that parents have more resources because of rising income over the life cycle of parents. *Ejrnaes & Pörtner* (2002) develop a model that predicts higher educational achievement for the first and the last-born child, without assumptions about preferences for specific birth-order. On birth-order see also *Lindert* (1977, p. 214) and *Hanushek* (1992, pp. 101 – 105).

5 Besides the three outcomes mentioned here, other dependent variables, like health outcome (e.g. *Garg & Morduch*) have been used as well.

6 *Stafford* (1976) uses teacher ratings of cognitive skills of pupils. *Willms* (1986) uses the number of siblings as one of his independent variables in explaining differences in exam results in Scotland. *Hanushek* (1992) uses results from the Gary Income Maintenance Experiment (which at the same time limits his sample to black, low-income families).

7 e.g. *Lindert* (1977), *Mare & Chen* (1986), *Hauser & Kuo* (1998), *Bauer & Gang* (2000).

8 e.g. *Kessler* (1991), *Björklund & Jäntti* (1994).

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

Many but not all of the empirical studies so far have made use of information that permitted the analysis of intrafamily sibling differences. The availability of information about all siblings of a family presents a clear advantage in so far as it avoids the problem of comparing siblings from different backgrounds. Interfamily differences are vulnerable to the influence of omitted variables or unobserved parental characteristics being correlated with the family size or the family configuration and can therefore provoke misleading estimates. It could well be, that more able parents have less children and invest more in the education of their children and vice versa. In this case smaller sibships would produce higher educational achievements but not because of competition for parental resources. However, intrafamily differences are not free of problems either. They too underlie the problem of unobserved characteristics of the parents. Even when taking into account the spacing between the siblings, there is the additional problem, that not all the siblings were raised in the same socio-economic living conditions. The first-born child is likely to have parents that earn less and live in an environment that is not as good as the one his siblings are born into later. The direction of the effect, however, is not clear, because parental income may grow over time, while the family gets bigger and they move to better school environments, but at the same time more siblings are competing for the parental resources, so that the net effect on achievement differentials might be insignificant.

Almost all studies, with the notable exception of the study of *Hanushek* (1992), measure the impact of contemporaneous or cumulated inputs in a cross-sectional analysis of achievement levels, a fact that can create problems. Hanushek had the advantage of a data set with several achievement observations over time. Therefore he was able to regress changes in inputs on changes in achievement ("value-added" specification). Unfortunately, most of the data sets at hand do not allow comparing the change in achievement for individuals over time.

Besides the problem, that family resources can differ in quantity and quality, another distinction should be made. Some resources are *divisible* and others are not. *Indivisible* resources, like the socio-economic status of the parents or the location of living have the same impact on the achievement of children independent of the size of the family. Other resources, monetary and non-monetary, are divisible, and an additional child diminishes the resources available for the other siblings. Some resources can change their nature in order to be better adapted to the size of the family. *Hanushek* (1992, p. 86) discusses in this respect the concept of "public time" versus "private time". Public time has the nature of a public good and all children can share it without lowering the amount available for the others. Private time is the time parents spend with a single child and that therefore cannot be spent on the other children. The decision how parents divide their time into private and public time will probably depend on the size of the family. When the family gets larger, parents can substitute private time with public time. But as public time most probably has a lesser educational value than private time, overall achievement will still be affected negatively. In any case, due to the fact that not all the

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

family resources are indivisible and substitution of resources has limits, theory would predict that the size of the family has a negative effect on all siblings but the reduction in achievement should not be linear. As parents will also differ in the quality and quantity of indivisible resources, we can predict that the negative impact of the family size will be different over families and smaller for families with a bigger share of indivisible resources[9].

The present study will focus on the family size. The data used here contains no specific data on gender composition and spacing and as a consequence, these family configuration aspects are not analysed further. Two other restrictions have to be accepted when using the PISA data. First, we only have one observation in time and therefore can only test the impact on attainment levels and second, because the data is restricted to 15-year-old pupils we will have to carry out interfamily comparisons. We hope, however, that despite the shortcomings of the data set, compared to the optimal situation, the large-scale assessment with its rich background data will provide some new insights that will be complimentary to the ones gathered in previous studies.

C. Data

In this paper we use the national PISA[10] data set 2000 for Switzerland. The full sample of the data has roughly 8'000 observations.[11] Table 1 gives an overview of the variables used in the empirical part of the paper.

Reading literacy in PISA is standardised to 500 points (OECD average) with a standard error of 100 points (see *OECD,* 2001). Switzerland had an average score of 494 points, close to the OECD average. Most variables are taken from the PISA data set with some exceptions. The "ISEI" variable is constructed on the basis of the professions of parents independent of the fact whether they are employed or not. Therefore the variable "parents employment situation" was created in order to have additional information about the current income of parents. Regarding the education of parents, pupils apparently had problems with the questionnaire in Switzerland (for more explanations see Coradi Vellacott & Wolter, 2002). Never-

[9] According to the theories of *Bourdieu* (1983) or *Coleman* (1988) on social capital, parents differ in their possession of economic, cultural and social capital. Especially the latter two but also parts of the economic capital (prestige, power) are mostly indivisible. In this case we would expect that children from families from a higher social class should suffer the least from a dilution of parental resources due to the family size. Regarding private time of parents, which is perhaps the most divisible resource, richer parents have also the possibility to substitute their own time with purchased child-care (on the decisions to demand for nonrelative child care see e.g. *Joesch & Hiedemann,* 2002 or *Lundholm & Ohlsson,* 2002).

[10] Programme for International Student Assessment of the OECD (see *OECD,* 2001).

[11] Only public schools are used in our investigation. Although there are some private schools (< 5 % of all schools) in the sample, these observations are not representative for the private school sector as a whole.

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

theless, the variable used for parents' education corresponds to the one found in the original data set. We also use some information about the closeness of parents to education and culture. These variables are the number of books at home and the possession of cultural goods. We used both variables because they showed a high degree of independence from each other.

We further use the variable "other language than official language" as a proxy for the distinction of Swiss nationals and foreign residents. Again, as with the question on the highest educational level of parents, the average number of pupils saying that they did not possess the Swiss nationality was far too low to be correct. We think that the variable on the language spoken at home comes closer to reality and serves our purposes well. We are aware of the fact that by using this variable we treat some Swiss pupils like foreigners (because they come from another language region) and some foreigners (who speak a Swiss national language at home, like German, French or Italian pupils) like Swiss. Knowing from the literature on migration that migrates coming from neighbouring countries assimilate faster and differently than foreigners from countries that are geographically and culturally more distant to Switzerland, the use of the language as a proxy for the degree of assimilation seems to make sense.[12]

Table 1

Variable Definitions

Variable	*Description*
Reading	Achievement in reading literacy as defined in PISA
ISEI	Socio-economic index of occupational *status* (values range from 16 to 90)[13]
Parent education	Years of schooling of the parent with the highest educational attainment
Gender	Dummy: 1 for girls, 0 for boys
Italian	Dummy for pupils living in the Italian speaking part of Switzerland; not for all pupils speaking Italian at home
French	Dummy for pupils living in the French speaking part of Switzerland; not for all pupils speaking French at home

[12] On the differences in the educational behaviour among different groups of migrants, see e.g. *Gang & Zimmermann* (2000). Although their study treats the German case, it should come close to the Swiss situation.

[13] This is an internationally comparable and standardised method of ranking the parent's profession according to their (socio-economic) status (see *Ganzeboom* et al., 1992). The Index is used as a proxy for income and wealth. Direct information on income and wealth could not be obtained, because students filled in the background survey. This might have an influence on the results. *Björklund & Jäntti* (1994) report in their paper that when using direct income measures instead of proxies like education and occupation, the sibling effect is either substantially reduced or disappears fully.

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

Youngest child	Dummy; value 1 if the pupil is the last-born in the family[14]
Middle child	Dummy; value 1 if the pupil is neither the first nor the last-born in the family
Single headed family	Dummy; value 1 if the family has only one adult person
Mixed family	Dummy; value 1 if at least one of the two adult persons in the family are not the father or the mother
Other family	Dummy: value 1 if none of the adult persons is considered to be father or mother
Parents employment situation	Dummy: value 1 if at least one adult in the family is full time employed and 0 otherwise
Other language than official language	Child speaks most of the time a language at home that is different from the language of assessment, from other official languages or from national dialects
Few books	Dummy; value 1 if the family has between 11 and 250 books at home
Many books	Dummy; value 1 if the family has more than 250 books at home
Cultural goods	Variable that indicates the amount of cultural goods at home
Room	Variable that indicates, whether the pupil has his own room at home
NSIB	Number of siblings (discrete cardinal scale). The variable is also used as a dummy, with a dummy for every size of the sibship

The PISA questionnaire contains a large set of questions about the living conditions of the tested pupils. We use the variable "room" of which we know from other research (see *Goux & Maurin,* 2002) that it has a direct, causal influence on reading literacy and of which we presume that the variable also depends on the number of siblings in a family, thus indicating whether there is rivalry for parental resources that matter for educational success.

D. Results

I. General findings

In a first step, we analyse the impact of the number of siblings on individual test scores in reading literacy.[15] Regression 1 in table 2 shows the coefficient of the

[14] Note that the PISA assessment is taken at the age of 15 and that therefore it is possible, that the family size measured at that point is not completed.

[15] All regressions were run on Stata 6.0. The complete national data set contains 7997 observations. Reductions in the number of observations can be explained with missing

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

linear specification of the sibling variable. The coefficient is highly significant but its magnitude is not overly impressive. In regression 2 we add some control variables that represent structural differences (like language region in Switzerland or gender), reflect the different (observable) abilities of parents[16] (formal education, closeness to education, language spoken at home), control for the parent's budget constraints (like ISEI or parents employment situation) and control for aspects of the family configuration (birth-order). As expected, the coefficient for the number of siblings is reduced substantially but remains highly significant.[17] The birth-order also plays a significant role, with the oldest sibling always being better in achievement than the youngest of the family.[18]

In regression 3 we add another set of independent variables, the configuration of adults in the family; variables that could also indicate constraints in time that can and will be devoted to the children. The inclusion of these variables does not change anything, except for lowering the influence of the variable of the employment situation of parents. There seems to be some correlation between the employment situation and the fact that the family is not composed of the two parents. The effect of the number of siblings on reading achievement, however, is not directly affected by the inclusion of these variables and remains significant. Additionally we introduced the variable "room". Although we could expect that this variable depends itself on the wealth of families and the number of siblings, it has a significant independent effect on the literacy achievement. We also tested whether this variable could be one of the channels through which the number of siblings influences reading literacy (see table 6).

---

values. No imputation for missing values was carried out. Cook-Weisberg Test was used to test for heteroskedasticity. The hypothesis of constant variance could not be rejected for the regressions in table 2. In the sub-group specification used for tables 4 & 5, we had a problem with heteroskedasticity; therefore we used a log-linear specification in all the regressions. After that the hypothesis of homoskedasticity could not be rejected anymore. Ramsey regression specification error test was used for the regression in table 2. The hypothesis of omitted variables was rejected. Because of the stratified sample all regressions have robust standard errors, using (233) schools as clusters. Specifying thereby that the observations are independent across groups (schools) but not necessarily within groups.

16 We are well aware, that even when controlling for observable abilities, we might miss significant ability differences between parents and that this could bias our results. The ways to correct for this potential bias would be experiments, instrumental variables or twin studies. *Behrman & Rosenzweig* (2002) find a substantial bias in the correlation of schooling of mothers and daughters due to heritable ability and assortative mating. Their study used a (very small) sample of identical twins.

17 We also included the additional variables step-wise but it doesn't change the qualitative results.

18 *Hanushek* (1992) showed in his paper that this effect might be artificial because when merging all family sizes together, the reference group (only child) could just capture some of the small family effect.

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

Table 2

Regression on reading literacy (number of siblings)

Regressions	1	2	3	4
Independent Variables	Coefficient	Coefficient	Coefficient	Coefficient
Constant	6.22* (0.009)	6.03* (0.016)	6.01* (0.017)	5.99* (0.018)
ISEI		0.002* (0.000)	0.002* (0.000)	0.002* (0.000)
Gender (Girl=1)		0.041* (0.005)	0.042* (0.005)	0.042* (0.005)
Italian		–0.022 (0.013)	–0.019 (0.013)	–0.018 (0.013)
French		0.012 (0.012)	0.012 (0.012)	0.012 (0.012)
Youngest child		–0.012** (0.005)	–0.013** (0.005)	–0.016* (0.005)
Middle child		–0.005 (0.008)	–0.004 (0.008)	–0.007 (0.007)
Single headed family			–0.021* (0.007)	–0.019* (0.007)
Mixed family			–0.029** (0.011)	–0.029* (0.011)
Other family			–0.073* (0.022)	-0.071* (0.022)
Parents employment situation		0.034* (0.008)	0.027* (0.008)	0.026* (0.008)
Other language than official language		–0.116* (0.009)	–0.112* (0.009)	–0.111* (0.009)
Few books		0.092* (0.012)	0.085* (0.011)	0.085* (0.011)
Many books		0.141* (0.013)	0.133* (0.013)	0.132* (0.013)
Cultural goods		0.014* (0.002)	0.014* (0.003)	0.015* (0.003)
Room			0.037* (0.010)	0.037* (0.010)
Parent education		0.005* (0.001)	0.005* (0.001)	0.005* (0.001)
NSIB	–0.023* (0.003)	–0.016* (0.003)	–0.014* (0.004)	0.006 (0.007)
NSIB squared				–0.004* (0.001)
Mean dependent Var.	6.19			
SD dependent Var.	0.202			
Adjusted *R*-squared	0.01	0.26	0.26	0.27
Number of observations	7901	6949	6916	6916

*, ** Significance at the 1% level and the 5% level respectively; the reference person is male, lives in the German speaking part of Switzerland, in a classical family with both parents (for regressions 3 & 4), his mother tongue is the local language and he is the oldest child in his family. All observations are weighted with the students' weight. Standard errors in brackets.

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

Regression 4 is a first test of the assumption that the family size effect is linear. The inclusion of the squared term of the variable changes the results in so far, that now the linear effect turns insignificant and the squared term is significant. Although we cannot deduce from this the exact form of the sibling effect, we have evidence, that the impact is not linear and that the children coming from the largest families must drive the effect. In the following regressions we therefore replace the linear specification with dummies for each size of the sibship (as done e.g. in *Björklund & Jäntti,* 1994).

The overall measure of the family size effect seems to indicate that there is a negative effect of the number of siblings on the individual educational achievement, although not high in magnitude[19] and not linear. To give some idea of the differences between different family sizes regarding literacy achievement, we show in table 3 the average scores.

These results encourage us to analyse the question further, using sub-samples of the data set, to see, whether the effect is the same for different groups of the population or differs within and between groups. The group building follows the two variables that seem to have the highest single effect on achievement, the socio-economic background of children and the language spoken at home.

Table 3

Average literacy achievement per family size[20]

Family size	Literacy score	Significant differences
a. 1 Child	497.81	a > e & f
b. 2 Children	505.14	b > c & d & e & f
c. 3 Children	498.61	c > e & f
d. 4 Children	488.17	d > f
e. 5 Children	462.46	
f. > 5 Children	444.30	

II. "Foreigners" versus "Natives"; "Rich" versus "Poor"

We divided the full sample into eight sub-groups[21]. First the sample was divided according the four quartiles of the ISEI, the socio-economic status, with quartile 1

[19] One additional child lowers the average achievement by 0.069 of one standard deviation in reading literacy.

[20] Average literacy achievement is calculated using plausible values and replicate weights. The specific descriptions of these techniques are in the technical report of PISA (OECD, 2002).

[21] Instead of building sub-groups, we also used interaction terms of language, isei-quartile and number of siblings. This allows us to regress everything in one regression using all

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

being the lowest quartile and quartile 4 the highest. Then we divided each quartile further into those pupils who spoke the official school language at home and those who did not. By dividing the sample first into the ISEI quartiles, we guarantee that "foreigners" and "natives" in each quartile come from an almost[22] comparable socio-economic background.

The results for the native speakers can be seen in table 4. We use the "full" model as in regression 3 in table 2 with all the independent variables and with the family size represented by dummies. The reference pupil in all cases is a boy with no siblings. The grouping of observations according to the socio-economic status of parents takes away some of the explanatory effect of variables reflecting differences in budget constraints between families but not everything. Furthermore the variable that describes the closeness of parents to education remains highly significant in all quartiles. The variables reflecting the configuration of adults in the family have a strong and significant negative effect in the upper half of the ISEI distribution, with the exception of single raising adults. The birth-order seems to play a role in the upper quartiles, in the sense that the oldest child fares better than his brothers and sisters.

Contrary to the overall effect in table 2, the sibling effect is at first sight rather blurred and only two dummies remain significant, interestingly positive and not negative. For the "richer" parents, small sibship seems to produce better achievements than families with just one child. However, for pupils that were tested in their mother tongue we cannot find any negative effects of siblings on their educational achievement.

Analysing the regressions for the pupils who speak a different language at home than the one in which they were tested presents different results – at least for the lower two quartiles of the socio-economic distribution.

observations. The results are not presented here, but show qualitatively the same picture as the sub-sample regressions. An F-test on all interaction terms showed that the hypothesis that they are not different from 0 could be rejected. Due to the large number of interaction terms (40) we preferred to show the regressions by sub-sample, as they are more straightforward in their interpretation.

22 The mean of the ISEI in the upper quartiles is slightly but not significantly higher for the native speakers and slightly lower for natives in the lowest quartile.

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

Table 4

Regression on reading literacy per ISEI quartile (mother tongue speakers only)

ISEI quartiles	1	2	3	4
Independent Variables	Coefficient	Coefficient	Coefficient	Coefficient
Constant	5.95*	5.90*	6.07*	5.98*
	(0.057)	(0.074)	(0.063)	(0.061)
ISEI	–0.001	0.007*	0.000	0.002**
	(0.002)	(0.001)	(0.001)	(0.001)
Gender (Girl=1)	0.072*	0.051*	0.022**	0.030*
	(0.012)	(0.009)	(0.10)	(0.009)
Italian speaking part	–0.003	–0.019	–0.044*	–0.038**
	(0.021)	(0.019)	(0.016)	(0.015)
French speaking part	0.037**	0.015	0.002	–0.014
	(0.016)	(0.017)	(0.015)	(0.014)
Youngest child	–0.009	–0.007	–0.015	0.026*
	(0.012)	(0.013)	(0.010)	(0.010)
Middle child	0.003	–0.010	–0.004	–0.023
	(0.020)	(0.017)	(0.018)	(0.014)
Single headed family	–0.008	–0.019	–0.026	–0.015
	(0.019)	(0.017)	(0.014)	(0.014)
Mixed family	–0.032	–0.013	–0.045**	-0.144**
	(0.039)	(0.022)	(0.018)	(0.185)
Other family	–0.012	–0.040	–0.120*	–0.063
	(0.036)	(0.027)	(0.030)	(0.033)
Parents employment situation	0.018	0.041**	0.030**	–0.016
	(0.020)	(0.019)	(0.015)	(0.015)
Few books	0.065*	0.138**	0.072**	0.063**
	(0.021)	(0.019)	(0.031)	(0.029)
Many books	0.114*	0.092*	0.113*	0.108*
	(0.026)	(0.023)	(0.033)	(0.031)
Cultural goods	0.006	0.005	0.024*	0.015*
	(0.007)	(0.006)	(0.005)	(0.006)
Room	0.075*	0.007	0.068	0.053**
	(0.006)	(0.020)	(0.018)	(0.023)
Parent education	0.007**	0.005	0.006*	0.001
	(0.003)	(0.003)	(0.002)	(0.002)
1 sibling	0.012	–0.002	0.019	***0.060****
	(0.024)	(0.029)	(0.015)	(0.002)
2 siblings	0.048	–0.015	0.018	***0.046*****
	(0.026)	(0.030)	(0.019)	(0.022)
3 siblings	0.038	0.001	0.006	0.033
	(0.030)	(0.031)	(0.024)	(0.024)
4 siblings	–0.017	–0.012	–0.045	0.059
	(0.038)	(0.044)	(0.034)	(0.032)
More than 4 siblings	–0.064	–0.024	–0.017	–0.008
	(0.056)	(0.056)	(0.049)	(0.046)
Mean dependent Var.	6.17	6.20	6.23	6.29
SD dependent Var.	0.185	0.175	0.165	0.154
Adjusted *R*-squared	0.14	0.09	0.12	0.13
Number of observations	1244	1397	1622	1431

*, ** Significance at the 1 % level and the 5 % level respectively; the reference person is male, lives in the German speaking part of Switzerland, in a classical family with both parents, his mother tongue is the local language and he is the oldest child in his family. All observations are weighted with the students' weight. Robust standard errors in brackets.

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

Table 5

Regression on reading literacy per ISEI quartile (does not speak the official school language)

ISEI quartiles	1	2	3	4
Independent Variables	Coefficient	Coefficient	Coefficient	Coefficient
Constant	6.05*	5.90*	5.83*	5.82*
	(0.083)	(0.159)	(0.187)	(0.153)
ISEI	–0.002	0.002	0.004	0.006*
	(0.002)	(0.005)	(0.003)	(0.002)
Gender (Girl=1)	0.017	0.063**	0.006	0.037
	(0.023)	(0.026)	(0.027)	(0.034)
Italian speaking part	0.088*	0.094**	0.025	–0.028
	(0.027)	(0.047)	(0.051)	(0.042)
French speaking part	0.080*	0.070**	–0.007	–0.086**
	(0.031)	(0.035)	(0.035)	(0.041)
Youngest child	–0.060	0.008	–0.020	0.005
	(0.031)	(0.024)	(0.032)	(0.033)
Middle child	0.010	0.029	–0.055	–0.077
	(0.037)	(0.045)	(0.040)	(0.058)
Single headed family	–0.057	–0.048	–0.017	–0.064
	(0.041)	(0.060)	(0.042)	(0.056)
Mixed family	0.015	0.034	–0.147	–0.155
	(0.064)	(0.045)	(0.076)	(0.110)
Other family	–0.064	–0.109	–0.041	–0.026
	(0.072)	(0.086)	(0.104)	(0.033)
Parents employment situation	0.058	0.039	0.009	–0.012
	(0.030)	(0.040)	(0.032)	(0.053)
Few books	0.092*	0.152*	0.175*	0.111
	(0.026)	(0.039)	(0.047)	(0.100)
Many books	0.253*	0.220*	0.221*	0.181
	(0.053)	(0.058)	(0.055)	(0.108)
Cultural goods	0.009	0.026	0.038**	0.048**
	(0.014)	(0.014)	(0.017)	(0.023)
Room	–0.015	0.011	0.017	-0.122
	(0.026)	(0.033)	(0.034)	(0.072)
Parent education	0.004	0.003	0.009	0.005
	(0.003)	(0.003)	(0.005)	(0.006)
1 sibling	0.003	-0.045	-0.060	-0.058
	(0.044)	(0.040)	(0.034)	(0.067)
2 siblings	–0.050	***–0.090*** **	–0.064	–0.043
	(0.042)	(0.042)	(0.043)	(0.066)
3 siblings	***–0.112*** **	***–0.114*** **	–0.079	–0.114
	(0.051)	(0.052)	(0.070)	(0.078)
4 siblings	***–0.140*** **	***–0.200*** *	*–0.042*	***–0.238*** **
	(0.061)	(0.066)	*(0.071)*	(0.119)
More than 4 siblings	–0.127	***–0.198*** **	*0.036*	–0.242
	(0.107)	(0.088)	*(0.073)*	(0.140)
Mean dependent Var.	6.01	6.06	6.13	6.21
SD dependent Var.	0.236	0.218	0.220	0.199
Adjusted *R*-squared	0.17	0.27	0.36	0.37
Number of observations	506	313	193	157

*, ** Significance at the 1% level and the 5% level respectively; the reference person is male, lives in the German speaking part of Switzerland, in a classical family with both parents, his mother tongue is *not* the local language and he is the oldest child in his family. All observations are weighted with the students' weight. Robust standard errors in brackets.

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

Again, as in table 4, standard errors have risen, but nevertheless, dummies for large families (3 sibs and more) in the lower half of the socio-economic distribution show high[23] and negative effects. Contrary to the native speakers, however, almost all the signs are negative even in the top quartile and there does not seem to be a negative single child effect as observed for native speakers.

III. Siblings rivalry for space at home

The calculations presented so far analysed the impact of the family size on educational achievement. PISA allows us to go one step further and test directly some hypotheses on the potential channels through which the rivalry for parental resources may work. We take the variable "room" as a proxy for (dividable) physical capital for which rivalry exists. Table 6 presents three different regressions to test whether the probability of a child to have his / her own room at home depends negatively and significantly on the number of brothers and sisters and not just the overall wealth of the family.

Regression 1 just shows the "raw" effect of the number of siblings on the probability to have ones own room. Familiy size is significant and negative. When we add control variables that control for time restrictions that could be due to the family configuration, the employment situation or the wealth of the family, the coefficient of the variable of interest is not affected significantly. The negative effect of the family size is not completely linear therefore we also present (regression 4) the results with the dummy specification, in order to have a better picture of the magnitude of the effect. The coefficients are marginal effects and show that the probability of a child in a family with more than five children to have its own room is about half of the probability of a single child.

[23] Between 0.5 and 1 standard deviation of the dependent variable.

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

Table 6

Probit-regression on the variable "room" (marginal effects)

Regressions	1	2	3	4
Independent Variables	Coefficient	Coefficient	Coefficient	Coefficient
ISEI		0.002*	0.001*	0.001*
		(0.000)	(0.000)	(0.000)
Gender (Girl=1)			–0.007	–0.007
			(0.007)	(0.007)
Single headed family			0.014	0.011
			(0.013)	(0.013)
Mixed family			0.037**	0.035**
			(0.013)	(0.013)
Other family			0.011	0.008
			(0.022)	(0.022)
Parents employment situation			0.022	0.022**
			(0.012)	(0.011)
Other language than official language			–0.105*	–0.105*
			(0.015)	(0.015)
Parents' education			0.006*	0.006*
			(0.002)	(0.015)
NSIB	***–0.049****	***–0.043****	***–0.041****	
	(0.004)	(0.003)	(0.004)	
1 sibling				***–0.066****
				(0.026)
2 siblings				***–0.162****
				(0.036)
3 siblings				***–0.288****
				(0.060)
4 siblings				***–0.321****
				(0.076)
More than 4 siblings				***–0.455****
				(0.087)
Pseudo *R*-squared	0.06	0.008	0.13	0.13
Number of observations	7869	7623	7030	7030

*, ** Significance at the 1 % level and the 5 % level respectively; the reference person is male, only child, lives in a classical family with both parents, his mother tongue is the official school language. All observations are weighted. Robust standard errors in brackets.

E. The sibling rivalry model revisited

In its basic version, the model that predicts a negative family size effect starts from the idea that limited parental resources have to be divided by the number of siblings and therefore any increase in the family size will diminish the beneficial effect family resources can have for their children. Although this rule applies to all levels of income and endowments, it is clear that parents with different budgets of money and time face different constraints. Richer parents can purchase resources, poor parents cannot always do it and they do not always have access to credit. Besides the possibility of using outside, non-relative, sources to counterbalance the

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

family size effect, the proportion of indivisible resources is also likely to depend on the parents' education, wealth and status.[24]

Outside factors, like the provision of free education, free child-care or generous child allowances also have to be taken into account when predicting the size of a sibling effect.[25]

Credit constraints, the quality of parental resources and the amount (quantity and quality) of indivisible resources therefore lead to a reinterpretation of the model. We would expect that children from better off parents (both in terms of income and education) would not suffer significantly from the presence of siblings. Those parents would be almost completely unconstrained. At the same time children in poor families with the same size of sibship would be significantly affected by the presence of siblings, as their parents face binding budget constraints.

We try to propose an extension of the classical family size model, as simple as possible, that tries to take into account the various research findings and at the same time fits the results found in this paper.

The family size effect on the educational achievement (E) of student i in family j depends on the size of the parental resources (R) and the number of children (C).

$$E_{ij} = \frac{R_j}{C_j} \tag{1}$$

An improved model explaining the effect of the size of the sibship on educational outcome should make a distinction between indivisible (iR) and divisible resources (dR). As explained earlier, both the magnitude and proportion (α) of the indivisible resources will probably depend on the socio-economic status (SeS) of the parents. We therefore reformulate equation 1 as follows:

$$R_j = \alpha iR_j + (1-\alpha)dR_j \tag{2}$$

$$E_{ij} = \alpha iR_j + \frac{(1-\alpha)dR_j}{C_j} \tag{3}$$

[24] *Behrman & Rosenzweig* (2002, p. 334) refer in their paper to information intensive versus time intensive resources that impact outcomes. If it is the first effect that matters most, education of parents leads to better schooling outcomes of their children, independent of the time parents spend with each child.

[25] *Björklund & Jäntti* (1994) have explained the differences in the sibling size effect between Sweden, Finland and the USA with the relatively more generous child allowances in the Nordic countries and the provision of free education of the same quality level for all children. *Bauer & Gang* (2000) have argued in the same manner, when explaining the absence of any substantial sibling effect in Germany: "It may well be that as incomes increase and education is more freely and cheaply available, gender effects and size effects disappear." (p. 248)

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

where

$$\alpha_j = \varphi(SeS_j)\ , \qquad \varphi > 0 \tag{4}$$

$$iR_j = \kappa(SeS_j)\ , \qquad \kappa > 0 \tag{5}$$

$$dR_j = \lambda(SeS_j)\ , \qquad \lambda > 0 \tag{6}$$

For the sake of simplicity we don't make an additional assumption about the linearity of the equations 4 – 6. What becomes evident from equation 3 and 4 is that with an increasing socio-economic status of the parents the negative impact of one additional child in a rich family is *ceteris paribus* lower than for a poor family. The higher socio-economic status translates into a smaller proportion of the resources being divisible.

We can differentiate the model further to take into account more aspects of the quality of the indivisible resources by introducing a variable that measures the cultural distance or proximity of the family to the host country (L) and reformulate equation 5. For parents originating from the host country L takes the value of 1.

$$iR_j = \kappa(SeS_j) * L_j\ , \qquad \kappa > 0;\ 0 < L < 1 \tag{7}$$

Having these equations, we can rewrite equation 3 in a stylised reduced form, which provides a shortcut to the main intuitions behind the model.

$$E_{ij} = f_1\left(\overset{+}{SeS_j}, \overset{-}{C_j}, \overset{+}{L_j}\right) \tag{8}$$

Equation 8 tells us that the educational achievement for children from families with a high socio-economic background and a higher proximity to the local culture, everything else equal, is higher but that the family size effect is negative for all families.

This result does not fit well our findings in table 3, where we have seen, that at least in the upper quartiles of the socio-economic distribution single children had a significantly lower educational achievement than children from small families. There must be another effect in the equation we have missed so far, because equation 8 is unambiguous, regarding the effect of additional siblings.

The missing link in our model is a peer effect.[26] We assume, that in a family brothers and sisters can act as co-educators, as role models and child carer, so that

[26] In introducing a peer effect in our model, we follow the example of the theory on the effects of class size on educational achievement. Empirical research confirmed many times that although the time a teacher can spend per pupil falls steadily when class size increases, we can measure that the average achievement in a class follows an inverse U shape. This means that there must be a counterbalancing factor, leading to an optimum in the class size. This effect is commonly called the peer effect and accounts for the positive interaction

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

some of the time and resources that parents can not spend on their children, is compensated by the interaction between children. Educational achievement is therefore a positive function of the number of siblings (S)[27]. We further assume that this effect is non-linear, because with an increasing number of siblings the interaction can also become negative, e.g. when siblings disturb each other in their homework. Finally we assume that the interaction between siblings depends on the quality of siblings[28], the better brothers and sisters are educated and have incorporated the cultural values of their parents, the more beneficial will their interaction with an additional child be. The quality of the children (Q) will depend on the socio-economic background and the proximity to the local culture of their parents. Equation 9 shows the peer-effect in family j on the educational achievement of student i.

$$E_{ij} = \left(S_j - \mu S_j^2\right) * Q_j \ , \qquad 0 < \mu < 1 \tag{9}$$

where

$$Q_j = \upsilon\left(SeS_j * L_j\right) \ , \qquad \nu > 0 \tag{10}$$

As in equation 8, a shortcut of the peer-effect shows, that educational achievement is a positive function of the socio-economic background and the cultural proximity and in this case of the number of siblings. The latter could turn negative, depending on the size of μ but in general not.

$$E_{ij} = f_2\left(\overset{+}{SeS_j}, \overset{+}{S_j}, \overset{+}{L_j}\right) \tag{11}$$

Bringing both effects (the family-size effect (f_1) and the peer effect (f_2)) together in one equation (f_3), we can see that the structural effect of the socio-economic background and the cultural proximity is reinforced. For small family sizes and high quality in the interaction between siblings, the effect of an additional child, however, is not clear anymore.

$$E_{ij} = f_1\left(\overset{+}{SeS_j}, \overset{-}{C_j}, \overset{+}{L_j}\right) + f_2\left(\overset{+}{SeS_j}, \overset{+}{S_j}, \overset{+}{L_j}\right) \tag{12}$$

$$E_{ij} = f_3\left(\overset{+}{SeS_j}, \left[\overset{+}{S_j} / \overset{-}{C_j}\right], \overset{+}{L_j}\right) \tag{13}$$

between pupils in a class. In the literature on educational production, pupils in a class are therefore also regarded as co-producers of education.

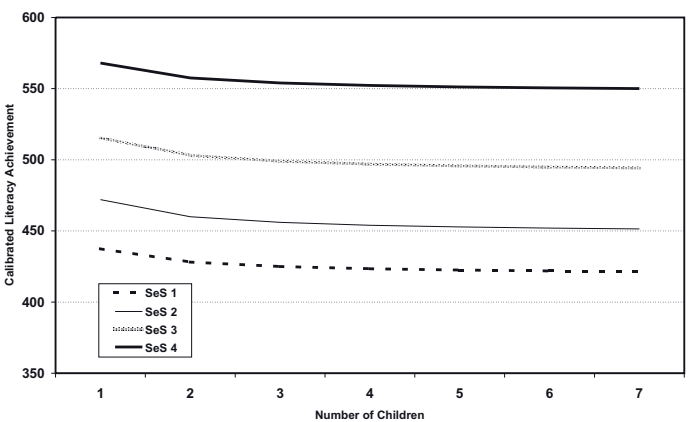
27 Note that now we don't speak of the number of children but the number of siblings instead.

28 Note that as in Becker, the world quality has no moral meaning.

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

In the following part we will show results of calibrated model that shows in graph 1 the family size effect, in graph 2 the peer effect and in graph 3 the combined effects of the number of children in a family on the average educational achievement within a given family size, depending on the socio-economic status of the family. Graph 4 finally presents the calibrated results for the highest and the lowest socio-economic quartile, divided into native and foreign families. The only difference between the foreign and the native family result from differences in the cultural proximity (*L*).

As one can easily see in graph 1, the average educational achievement in a family, conditional on its size decreases with every additional child in the family. The reduction is not linear. The first additional child takes half of the resources of the first child whereas the third child would "only" take a third of the resources from each of the first two children. We show a non-linear reduction in our model that *Hanushek* (1992) also stipulated.

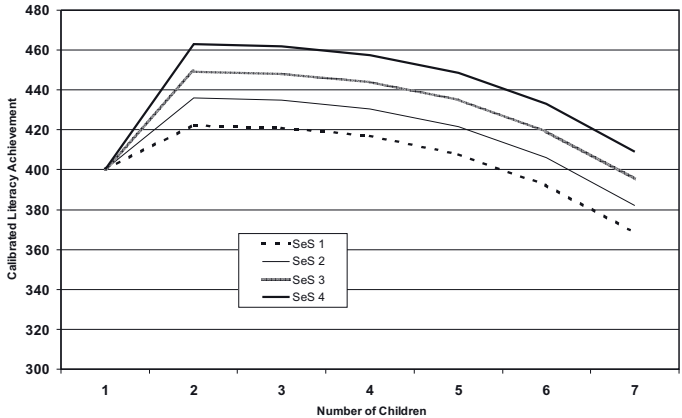


Graph 1: Calibrated family-size effect[29]

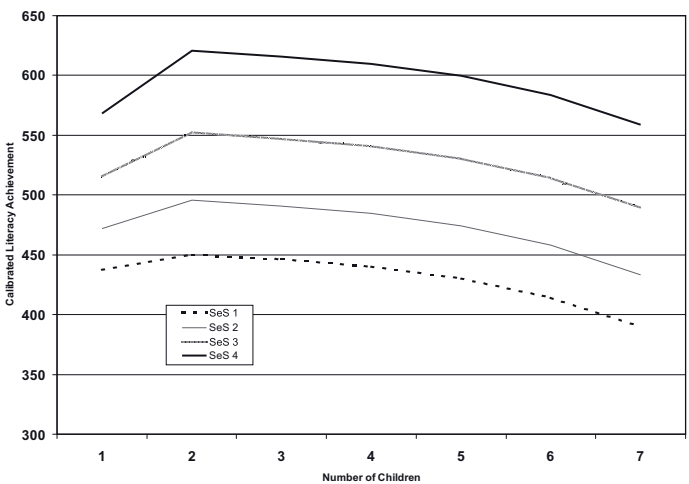
In graph 2 we see the peer effect. The marginal effect is the greatest between the first and the second child, where we have a 0 / 1 situation, and declining afterwards. It even turns negative for very large families, due to the squared term in equation 9. In graph 3 we make the sum of both effects and we can see that very large families will have significantly lower average achievements than families with only one child. Families in the highest socio-economic quartile are better off, if they have at least two children. For the lower socio-economic quartiles the "only-child" effect is visible but small (as in table 4).

[29] *SeS* 1 stands for the lowest quartile of socio-economic status and 4 for the highest quartile respectively.

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/



Graph 2: Calibrated peer-effect



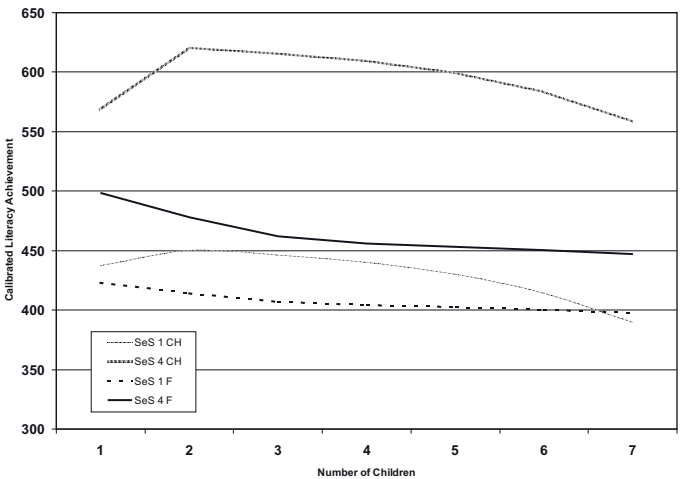
Graph 3: Calibrated total family size effect

Finally in graph 4 we show the model results for natives and foreigners. Due to the effect of the cultural proximity in both, the peer and the family-size equation, we can calibrate results that show a decrease in the average achievement from the second child onwards. The gap between natives and foreigners widens first and closes when the family size gets very large as in the empirical results.

Thus with this model that takes into account two effects instead of only looking at the detrimental effect of an additional child, we can explain positive as well as negative total effects of the family size on the average educational achievement in

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

a family. We can conclude from this that a non-existing family size effect at the macro level might well hide significant and substantial sibling effects for subgroups of the population. A possibility educational policy should be aware of.



Graph 4: Calibrated total family size effect with natives and foreigners[30]

F. Conclusions

Looking at the intergenerational patterns of educational attainment, many highly developed and industrialised countries still show a disturbingly high degree of transmission from one generation to the next (e.g. *Acemoglu & Piscke,* 2001 or *Dustmann,* 2001 or *Ermisch & Francesconi,* 2001). Although the sources for this low intergenerational mobility in education and consequently in income status and wealth are not clear yet, some results of the PISA study 2000 have clearly indicated an urgent need for a better understanding of the way the educational system deals with social differentiation. Contrary to the political statements, many indicators in the PISA data let us think that the educational system does not fulfil one of its functions, namely to reduce the impact social differences can have on educational achievement.

In this paper we look at a particular source of social differentiation in the educational system that handicaps children coming from a different social background. We know that educational achievement is only partly the result of schooling and to an important degree also the product of interactions between children and their parents. In this context the size of the family can have a detrimental effect on chil-

30 CH denotes Swiss natives and F stands for foreigners.

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

dren's school performance. The most recent empirical literature has, however, found smaller and even non-existing family size effects and explained this observation with the free provision of education in most countries, lower opportunity costs of education and the providence of free child care or child allowances. All these developments should have built up a framework in which parents would no longer face binding budget constraints.

The empirical results in this paper show, at least for the case of Switzerland, that the real magnitude of the family size effect is only apparent when looking at sub-samples of the population. The results point in the direction that those families that already suffer from a lower socio-economic status, less education or a greater cultural distance to their host-country exhibit an additional and significant family size effect. Therefore children from these families are triple handicapped, firstly because their parents cannot give them a lot of resources and secondly because the low endowments have to be split on more siblings and thirdly because as foreigners integration into the host countries culture seems to be difficult. In the empirical results, the overall family size effect is diluted by the fact, that there is also a "single-child" effect observable that works in the opposite direction. This effect masks the real magnitude of the family size effect in the sub-groups of the population that suffer from it.

Regressing the number of siblings on the achievement, we use outcome data for our analysis, which can be problematic in the sense that it still poses an identification problem. Our hypothesis, that the number of siblings reduces the resources parents can spend on their children was therefore directly tested with a variable that tells something about the physical resources (own room) available for the child. The results confirm our hypothesis insofar we can observe a significant and negative impact of the family size on this variable.

It is fair to say, that due to the various limitations of the data set, we cannot rule out the possibility of any bias in these results completely. It is still possible that some unobserved characteristics of the parents influence simultaneously the "quality" of children and the decision to have more children. In this case the location of the problem of underachievement would still be valid but it would influence the choice of the remedy.

If the results found reflect really the consequence of budgetary constraints of a sub-sample of families, financial aid (e.g. targeted[31] child allowances) should help to overcome these handicaps by providing those parents with access to the same resources as other families.[32]

---

[31] Targeting groups of families and targeting the type of measures would still present a difficult task to solve (see for an overview of what we know and what we do not know of interventions in early childhood, *Waldfogel,* 2002).

[32] Recent studies show, that with financial incentives, the number of parents that make use of early child-care and education can be influenced positively (e.g. *Del Boca,* 2002 or *Michalopoulos & Robins*, 2002).

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

If, however, the concentration of low achievers in big families would have to be explained by unobservable traits of their parents, money alone would not be sufficient to improve the situation. In this case the social and educational system should provide these children with a beneficial environment to compensate for the negative influences in their parental home.[33] The same type of measure would be necessary to combat the single child effect, in order to stimulate early interaction of those children with other children (peers).

Another unsolved question is the time of intervention. It is not because we measure the effect at the age of fifteen that it has to be cured at that age. On the contrary, budget constraints parents can face in the pre-school-time of their children can be much more important than the constraints during school-time.[34]

In any case, whatever the causality behind the family size effect is, PISA data allow us to localise sibling rivalry among specific groups of parents. Targeted interventions for these families should therefore help to reduce the social differentiation within the society and enhance economic mobility across generations.

Literature

Acemoglu, D. / *Pischke,* J.-S. (2001): Changes in the Wage Structure, Family Income, and Children's Education, European Economic Review, Vol. 45, pp. 890 – 904.

Bauer, T. / *Gang,* I. N. (2001): Sibling Rivalry in Educational Attainment: The German Case, Labour, Vol. 15(2), pp. 237 – 255.

Becker, G. S. (1960): An Economic Analysis of Fertility, in: Coale et al. Demographic and Economic Change in Developed Countries, Princeton: National Bureau of Economic Research, pp. 209 – 240.

Becker, G. S. / *Tomes,* N. (1986): Human Capital and the Rise and Fall of Families, Journal of Labor Economics, Vol. 4(3), pp. 1 – 39.

Behrman, J. R. / *Rosenzweig,* M. R. (2002): Does Increasing Women's Schooling Raise the Schooling of the Next Generation?, American Economic Review, Vol. 92 (1), pp. 323 – 334.

33 Those interventions should of course be compulsory because we would expect that even with free access to these measures, those who would benefit the most from them would not automatically choose to use them. At the same time we did not discuss further the effect free and compulsory child care would have on the employment behaviour (on which a large literature already exists, see e.g. *Gelbach,* 2002) of women and how this could again have an influence on the time parents spend with their children.

34 This would also correspond with the findings of empirical research in early childhood and pre-school child-care. *Hanushek* (1992) found a strong effect of the family configuration in the pre-school time on educational achievements at the beginning of the school career. Other studies, like a recent paper by *Jenkins & Schluter* (2002) show contrasting results, indicating that it is later income that has a bigger effect than income during early childhood, although modestly, when analysing the educational attainment of 14-year-old Germans.

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

Björklund, A. / *Jäntti,* M. (1994): The Impact of the Number of Siblings on Men's Adult Earnings: Evidence from Finland, Sweden and the United States, Paper presented at the EALE-conference in Warzaw.

Bourdieu, P. (1983): Oekonomisches Kapital, kulturelles Kapital, soziales Kapital, in: R. Krekel (Ed.), Soziale Ungleichheiten. Soziale Welt, Sonderband 2, Göttingen: Schwartz, pp. 183 – 198.

Butcher, K. F. / *Case,* A. (1994): The Effect of Sibling Sex Composition on Women's Education and Earnings, Quarterly Journal of Economics, Vol. CIX (3), pp. 531 – 563.

Chadwick, L. / *Solon,* G. (2002): Intergenerational Income Mobility Among Daughters, American Economic Review, Vol. 92 (1), pp. 335 – 344.

Coleman, J. S. (1988): Social Capital in the Creation of Human Capital, American Journal of Sociology, Vol. 94, pp. 95 – 120.

Coradi Vellacott, M. / *Wolter,* S. C. (2002): Origine Social et Egalité des Chances, in: OFS+CDIP (Eds.): Préparés pour la vie?, Neuchâtel: Office Fédéral de la Statistique, pp. 88 – 109.

Del Boca, D. (2002): The Effect of Child Care and Part Time Opportunities on Participation and Fertility Decisions in Italy, Journal of Population Economics, Vol. 15(2), pp. 549 – 573.

Dustmann, C. (2001): Parental Background, Primary to Secondary School Transitions, and Wages, IZA Discussion Paper No. 367.

Ejrnaes, M. / *Pörtner,* C. C. (2002): Birth Order and the Intrahoushold Allocation of Time and Education, Paper presented at the ESPE Conference 2002, Bilbao.

Ermisch, J. / *Francesconi,* M. (2001): Family Matters: Impacts of Family Background on Educational Achievements, Economica, pp. 137 – 156.

Gang, I. N. / *Zimmermann,* K. F. (2000): Is Child like Parent?, Journal of Human Resources, Vol. XXXV (3), pp. 550 – 569.

Ganzeboom, H. B. G. / *De Graaf,* P. / *Treiman,* D. J. (1992): A Standard International Socioeconomic Index of Occupational Status, Social Science Research, Vol. 21(1), pp. 1 – 56.

Garg, A. / *Morduch,* J. (1998): Sibling Rivalry and the Gender Gap: Evidence from Child Health Outcomes in Ghana, Journal of Population Economics, Vol. 11, pp. 471 – 493.

Goux, D. / *Maurin,* E. (2003), The Effects of Overcrowded Housing on Children's Performance at School, CEPR Discussion Paper, No. 3818, March 2003.

Gelbach, J. B. (2002): Public Schooling for Young Children and Maternal Labor Supply, American Economic Review, Vol. 92 (1), pp. 307 – 322.

Hanushek, E. A. (1992): The Trade-off between Child Quantity and Quality, Journal of Political Economy, Vol. 100 (1), pp. 84 – 117.

Hauser, R. M. / *Kuo,* H.-H. D. (1998): Does the Gender Composition of Sibships Affect Women's Educational Attainment?, Journal of Human Resources, Vol. XXXIII (3), pp. 644 – 657.

Haveman, R. / *Wolfe,* B. (1993): Children's Prospects and Children's Policy, Journal of Economic Perspectives, Vol. 7 (4), pp. 153 – 174.

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

Jenkins, S. P. / *Schluter,* C. (2002): The effect of family income during childhood on later-life attainment: evidence from Germany, mimeo, University of Essex.

Joesch, J. M. / *Hiedemann,* B. G. (2002): The Demand for Nonrelative Child Care among Families with Infants and Toddlers: A Double-Hurdle Approach, Journal of Population Economics, Vol. 15(2), pp. 495 – 526.

Kaestner, R. (1997): Are Brothers Really Better? Sibling Sex Composition and Educational Achievement Revisited, Journal of Human Resources, Vol. XXXII (2), pp. 250 – 284.

Kessler, D. (1991): Birth Order, Family Size, and Achievement: Family Structure and Wage Determination, Journal of Labor Economics, Vol. 9 (4), pp. 413 – 426.

Lindert, P. H. (1977): Sibling Position and Achievement, Journal of Human Resources, Vol. XII (2), pp. 198 – 219.

Lundholm, M. / *Ohlsson,* H. (2002): Who takes Care of the Children? The Quantity-Quality Model Revisited, Journal of Population Economics, Vol. 15(3), pp. 455 – 461.

Mare, R. D. / *Chen,* M. D. (1986a): Further Evidence on Number of Siblings and Educational Stratification, American Sociological Review, Vol. 51, pp. 403 – 12.

Mare, R. D. / *Chen,* M. D. (1986b): Sibship Size and Educational Stratification. Rejoinder to Blake, American Sociological Review, Vol. 51, pp. 418 – 420.

Michalopoulos, C. / *Robins,* P. K. (2002): Employment and Child-Care Choices of Single-Parent Families and the United States, Journal of Population Economics, Vol. 15(2), pp. 465 – 493.

OECD (2001): Knowledge and Skills for Life, First Results from PISA 2000, Paris: OECD.

OECD (2002): PISA 2000 Technical Report, Paris: OECD.

Stafford, F. P. (1987): Women's Work, Sibling Competition, and Children's School Performance, The American Economic Review, Vol. 77 (5), pp. 972 – 980.

Waldfogel, J. (2002): Child Care, Women's Employment, and Child Outcomes, Journal of Population Economics, Vol. 15(2), pp. 527 – 548.

Willms, J. D. (1986): Social Class Segregation and its Relationship to Pupils Examination Results in Scotland, American Sociological Review, Vol. 51, pp. 224 – 241.

Wolter, S. C. (2003): Sibling Rivalry: A Six Country Comparison, IZA Discussion Paper No. 734.

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

Die Bedeutung der Zeitschriftenauswahl für die Evaluation ökonomischer Forschung*

Von *Martin G. Kocher*[1], *Matthias Sutter*
und *Ulrike L. Heregger,* Innsbruck

A. Einleitung

Enger werdende Budgetrestriktionen (in vielen EU-Staaten) führen dazu, dass öffentliche Institutionen verstärkt Rechenschaft über die Verwendung öffentlicher Finanzmittel abgeben müssen. Auch die Neugestaltung des Verhältnisses zwischen dem Staat und einigen öffentlichen Einrichtungen, die oft unter dem Stichwort „Autonomie" firmiert, bedingt ein höheres Maß an Evaluation als Mittel zur strategischen Steuerung. Dies betrifft insbesondere auch den Hochschulsektor, für den in den letzten Jahren die Evaluation von Forschung und Lehre stark an Bedeutung gewonnen hat. In Deutschland etwa sieht die Besoldungsreform vor, dass ein Teil des Gehaltes von Wissenschaftlern abhängig von den erbrachten Leistungen in Lehre und Forschung ausbezahlt werden soll. In Österreich beispielsweise ist die regelmäßige Evaluation von Lehre und Forschung gesetzlich festgeschrieben worden[2], verbunden mit der Option, die Ressourcenzuteilung im akademischen Bereich von Evaluationsergebnissen abhängig zu machen. Dabei geht es allerdings nicht primär um eine Evaluation individueller Leistungen, wie dies etwa die Besoldungsreform in Deutschland impliziert, sondern um eine Evaluation von akademischen Einheiten (Institutionen, Fachbereiche etc.). Das bedeutet, dass institutionelle Einheiten innerhalb einer Universität in Abhängigkeit von ihren Leistungen größer oder kleiner (im Sinne der Verfügbarkeit von Ressourcen) im Verhältnis zu anderen Einheiten werden können. Aufgrund der offensichtlichen Relevanz und der weitreichenden Konsequenzen, die mit Evaluationen einhergehen bzw. in nicht

* Wir danken Heidrun Brugger-Sutter und Robert Mrsic für wertvolle Hilfe bei der Aufbereitung der Daten. Außerdem gebührt den Teilnehmerinnen und Teilnehmern am Bildungsökonomischen Ausschuss 2003 in Zürich, insbesondere Uschi Backes-Gellner, sowie zwei anonymen Gutachtern Dank für wertvolle Kommentare und stimulierende Diskussionen.

[1] Korrespondierender Autor, Addresse: Institut für Finanzwissenschaft, Universität Innsbruck, Universitätsstraße 15 / 4, A-6020 Innsbruck, mail: martin.kocher@uibk.ac.at, Tel.: 0043 / 512 507 7172, Fax: 0043 / 512 507 2970

[2] Verordnung des Bundesministers für Wissenschaft und Verkehr über Grundsätze für die Durchführung von Evaluierungen in Forschung und Lehre der Universitäten (Evaluierungsverordnung – EvalVO); BGBl. II Nr. 224 / 1997.

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

allzu ferner Zukunft möglicherweise einhergehen werden, gibt es erhebliche Auffassungsunterschiede, auf welcher Grundlage Evaluationen durchgeführt werden und welche konkreten Konsequenzen sich daraus ergeben sollten.

Bei der Evaluierung von Lehre können etwa das Betreuungsverhältnis von wissenschaftlichem Personal zu Studierenden, die Anzahl der Absolventen bzw. Promoventen oder die subjektive Wahrnehmung der Lehrveranstaltungsleiter durch die Studierenden eine wichtige Rolle spielen. Die Evaluation der Forschung[3] konzentriert sich in der Regel auf wissenschaftliche Publikationen.[4] Hinsichtlich der unterschiedlichen Publikationskategorien (Bücher, Herausgeberschaften, Buchbeiträge, Rezensionen, Aufsätze in Zeitschriften) spielen innerhalb der Ökonomie Publikationen in der Form von Aufsätzen in international angesehenen Zeitschriften eine immer wichtigere Rolle.[5] Damit hält auch im wirtschaftswissenschaftlichen Bereich eine ‚Kultur' Einzug, die bisher im Wesentlichen auf die Naturwissenschaften und die Medizin beschränkt war. Die zunehmende Internationalisierung der ökonomischen Forschung und die Akzeptanz des Englischen als Verständigungs- und Publikationssprache haben den Trend zu Publikationen in international anerkannten und referierten Fachzeitschriften auch in den nicht-englischsprachigen europäischen und asiatischen Ländern in den letzten Jahrzehnten deutlich verstärkt.

Daher muss sich folgerichtig auch die Forschungsevaluation auf wissenschaftliche Aufsätze in Zeitschriften konzentrieren, wobei gleichzeitig Bücher und Buchbeiträge quasi automatisch an Bedeutung verlieren. Jedoch bleibt dabei die Auswahl und Bewertung der Zeitschriften für die Evaluation von Forschungseinrichtungen auch unter Wissenschaftlern höchst umstritten. Eine Gesamterhebung bzw. die vollständige Übernahme von Datenbankinhalten ist angesichts der Vielfalt und Menge an wissenschaftlichen Zeitschriften oft nicht möglich und birgt zudem das Problem von adäquaten Qualitätsgewichtungen. Deshalb gewinnen Rankings von Institutionen und einzelnen Forschern hinsichtlich des Forschungsoutputs in weltweiten Spitzenzeitschriften (‚Top'-Journals) zunehmend an Popularität (vgl. *Elliott, Greenaway* und *Sapsford,* 1998; *Kalaitzidakis, Mamuneas* und *Stengos,*

---

3 Selbstverständlich bestehen auch Interdependenzen zwischen Lehre und Forschung, so dass man etwa von einer Verbundproduktion (*Backes-Gellner* und *Zanders,* 1989) sprechen könnte. Im Folgenden werden wir diesen Aspekt aber nicht berücksichtigen.

4 Natürlich gibt es eine Reihe von anderen Forschungsleistungen (Vorträge bei internationalen Konferenzen, Refereeing etc.), die aber alle für die Forschungsevaluation in der Regel eine untergeordnete Rolle spielen. Außerdem sei an dieser Stelle darauf hingewiesen, dass neben Lehre und Forschung, die zweifelsohne zum Kerngeschäft von Universitäten gehören, in Abhängigkeit von den Zielsetzungen einer Institution noch eine Reihe anderer Outputkategorien relevant sein können.

5 *Schlinghoff* (2002) hat in seiner Untersuchung der Berufungspolitik deutscher wirtschaftswissenschaftlicher Fakultäten gefunden, dass für die erste Berufung insbesondere Aufsätze in deutschen Top-Zeitschriften einen signifikanten Einfluss haben. Es wäre eine interessante Untersuchung, ob dies im Wesentlichen auf die Berufungspolitik für BWL-Lehrstühle zurückzuführen ist oder ob dieses Ergebnis auch für VWL-Lehrstühle zutrifft.

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

1999, 2003; *Hodgson* und *Rothman,* 1999; *Coupé,* 2003). Ein häufig gehörter – und auf Tagungen immer wieder geäußerter – Vorwurf gegen Forschungsarbeiten in dieser Tradition besteht jedoch darin, dass dabei der ‚nationale' oder ‚regionale' Markt von Forschern nicht adäquat berücksichtigt und deshalb ein verzerrtes Bild des Forschungsoutputs einzelner Institutionen wiedergegeben wird. Hinter diesem Einwand steht die Hypothese, dass es Forschungsinstitutionen gibt, die sich auf die internationalen (eher ‚theoretischen') Zeitschriften in ihrer Disziplin spezialisieren, aber ansonsten kaum Forschungsoutput für den ‚nationalen' Markt produzieren, während andere Institutionen im nationalen Publikationsmarkt mit eher ‚angewandten', ‚politikberatenden' Arbeiten sehr präsent, dafür aber international nicht vertreten sind, weil politikorientierte und angewandte Arbeiten mit nationalem (Daten-)Hintergrund kaum in weltweiten Spitzenzeitschriften publizierbar wären.

In der folgenden Arbeit überprüfen wir diese Hypothese, um den oben präzisierten Vorwurf entweder verwerfen oder bestätigen zu können. Eine Bestätigung würde bedeuten, dass jede Evaluation, die sich ausschließlich auf internationale Top-Zeitschriften beschränkt, notwendigerweise zu verzerrten Ergebnissen kommt und damit für eine Ressourcenzuteilung innerhalb von Universitäten bzw. zwischen Universitäten bestenfalls sehr eingeschränkt geeignet ist.[6] Ein Verwerfen des Vorwurfs würde implizieren, dass die Auswahl der Zeitschriften als Basis für Evaluierungen deutlich weniger problematisch ist als vielfach angenommen und dass keine systematischen Unterschiede bzgl. des Forschungsoutputs in weltweiten Spitzenzeitschriften und ‚lokalen' Zeitschriften bestehen, solange man sich auf wissenschaftliche Zeitschriften beschränkt.

Zur Überprüfung der Hypothese führen wir zwei unabhängige empirische Untersuchungen durch:

1. Deutsche Studie: Auf Basis des Publikationsoutputs von Forschungsinstitutionen aus dem deutschsprachigen Raum in drei verschiedenen Gruppen von ökonomischen Zeitschriften (amerikanische, britische bzw. kontinentaleuropäische Zeitschriften) im Zeitraum von 1997 bis 2001 wird die Korrelation zwischen den Outputs in diesen drei Gruppen analysiert.
2. Kanadische Studie: Wir untersuchen den Publikationsoutput kanadischer Forschungsinstitutionen von 1980 bis 2000 in *dem* ‚lokalen' Journal (*Canadian Journal of Economics*) und den ‚Top 10'-Zeitschriften gemäß Impact-Faktoren. Kanada hat den großen Vorteil, dass tatsächlich nur eine zentrale ökonomische Zeitschrift (im weiteren auch als ‚home journal' bezeichnet) existiert, obwohl die Anzahl der wissenschaftlichen Institutionen durchaus mit jener in mittelgroßen, europäischen Ländern verglichen werden kann.

---

[6] Wir beschränken uns im Folgenden auf die Evaluation von Institutionen, obwohl einige Ergebnisse auch auf individuelle Evaluationen anwendbar sind. Da individuelle Publikationsmuster aber bei Weitem diverser sind als diejenigen, die auf institutioneller Ebene nach einer Aggregation zustande kommen, sind unsere Schlussfolgerungen nur bedingt auf individuelle Evaluationen umzulegen.

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

Der Rest des Aufsatzes ist folgendermaßen gegliedert: Abschnitt B stellt unsere Methode ausführlich dar. Die Abschnitte C und D präsentieren die Ergebnisse der deutschen bzw. kanadischen Studie. Abschnitt E zieht ein Resümee.

B. Methode

I. Reihung von Zeitschriften

Die Fülle an ökonomischen Fachzeitschriften nimmt ständig zu. *Hodgson* und *Rothman* (1999) erwähnen, dass von 1431 Zeitschriften, die den Sozialwissenschaften zugeordnet werden können, allein 231 Zeitschriften den Wortstamm ‚econ-' (wie ‚economics', ‚economy' oder ‚econometric') enthalten. Grob gesagt lassen sich drei unterschiedliche Methoden zur Reihung und Auswahl von Zeitschriften unterscheiden:

– Man schreibt bestimmten Zeitschriften mehr oder weniger willkürlich die Attribute ‚top', ‚core', ‚blue ribbon' zu, ohne ein objektiv nachvollziehbares Kriterium dafür anzugeben (vgl. *Elliott, Greenaway* und *Sapsford,* 1998; *Kalaitzidakis, Mamuneas* und *Stengos,* 1999).

– Man befragt Ökonomen nach den besten Zeitschriften (vgl. *Bräuninger* und *Haucap,* 2001). Mit dieser Methode, die naturgemäß sehr stabile Rankings generiert, kann das Renommee einer Zeitschrift erfasst werden. Problematisch dabei ist allerdings, dass es sehr starke und lang andauernde ‚Renommee-Resistenzen' geben kann, die auf der Einschätzung der Güte einer Zeitschrift aus einer relativ weit zurückliegenden Vergangenheit beruhen können.

– Man reiht Zeitschriften nach der Zitationshäufigkeit ihrer Artikel (vgl. *Hodgson* und *Rothman,* 1999; *Kocher* und *Sutter,* 2001), basierend auf der Annahme, dass ‚citations are the scientific community's version of dollar voting by consumers for goods and services' (*Laband* und *Piette,* 1994, S. 641). Die relative Zitationshäufigkeit von Artikeln in einer Zeitschrift kann als Maß für die Verbreitung und den Einfluss einer Zeitschrift auf ein Wissenschaftsgebiet gesehen werden. Das Institute for Scientific Information (ISI) publiziert jährlich die Journal Citation Reports (JCR), in denen die Zeitschriften aus dem Social Science Citation Index (SSCI) anhand ihrer Impact-Faktoren gereiht werden können. Auf diesem Ansatz basiert die folgende Arbeit.

II. Der Impact-Faktor in den Journal Citation Reports (JCR)

Der Impact-Faktor ist eines der Maße der relativen Zitationshäufigkeit der Artikel in einer bestimmten Zeitschrift. Er misst, wie oft im Jahr *t* Artikel aus einer Zeitschrift, die in den Jahren *t-1* und *t-2* publiziert worden sind, zitiert werden und

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

dividiert diese Zahl durch die Anzahl der in den Jahren *t-1* und *t-2* publizierten Artikel. Der Impact-Faktor ist damit ein Maß für die relative Zitationshäufigkeit von Artikeln. Betrachten wir beispielsweise den höchsten Impact-Faktor einer ökonomischen Zeitschrift, den des *Journal of Economic Literature (JEL)* für das Jahr 2001. Der Wert von 7,92 ergibt sich folgendermaßen: Es sei N die Anzahl der Zitate in sämtlichen Zeitschriften des SSCI im Jahr 2001, die sich auf einen Artikel im *JEL* aus den Jahren 1999 (N_{99}) oder 2000 (N_{00}) beziehen. Mit A bezeichnen wir die Anzahl der im *JEL* publizierten Artikel in den Jahren 1999 (A_{99}) und 2000 (A_{00}). Der Impact-Faktor für 2001 ergibt sich dann als: $Impact\text{-}Faktor_{01} = (N_{99}+N_{00}) / (A_{99}+A_{00})$. Beide im Folgenden berichteten Teilstudien basieren auf einer Reihung von Zeitschriften entsprechend ihren Impact-Faktoren.[7]

C. Die deutsche Studie

Für unsere deutsche Studie haben wir drei Gruppen von Zeitschriften gebildet: amerikanische (Gruppe A), britische (Gruppe B) und kontinentaleuropäische (Gruppe C) Zeitschriften. In Betracht gezogen haben wir nur ‚general economics'-Zeitschriften (mit Refereesystem), aber keine ‚field journals'. In jeder Gruppe haben wir jene vier Zeitschriften ausgewählt, die im Zeitraum 1997 bis 2001 die höchsten durchschnittlichen Impact-Faktoren aufwiesen. Für die Auswertung sind alle Jahrgänge von 1997 bis 2001 berücksichtigt. Tabelle 1 listet die 12 Zeitschriften, deren durchschnittlichen Impact-Faktor und die Anzahl der publizierten Artikel im Beobachtungszeitraum auf.

Insgesamt wurden 3.230 Artikel mit insgesamt 5.331 Einträgen für Autoren erfasst. Die Anzahl der Artikel verteilt sich fast gleichmäßig auf die drei Gruppen (1.138 in Gruppe A, 1.020 in Gruppe B, 1.086 in Gruppe C). Die durchschnittlichen Impact-Faktoren sind am höchsten in Gruppe A, gefolgt von Gruppe B und dann Gruppe C. Unsere Gruppeneinteilung lässt also eine deutliche Abstufung der Zitationshäufigkeit und damit des Einflusses der jeweiligen Zeitschriften auf nachfolgende Publikationen erkennen. Innerhalb der Gruppen sind die Impact-Faktoren der Zeitschriften relativ homogen.

Tabelle 2 weist jeweils die Institutionen mit den meisten Einträgen und den sich daraus ergebenden Anteil der ‚Top 10'-Institutionen an allen Publikationen innerhalb einer der drei Gruppen aus. Es zeigt sich dabei, dass in allen drei Zeitschriftengruppen die Institutionen aus dem jeweiligen Herkunftsbereich der Zeitschriften dominieren. Besonders auffällig ist diese Eigenschaft in Gruppe A, in der alle 10

[7] Die Impact-Faktoren der Zeitschriften im SSCI werden seit 1977 jährlich vom Institute for Scientific Infomation publiziert. Seit 1979 werden die über 1000 im SSCI erfassten Zeitschriften einzelnen Fachdisziplinen in den Journal Citation Reports zugeordnet, um eine Reihung innerhalb einer Fachdisziplin erstellen zu können. *Sutter* und *Kocher* (2001) zeigen, dass zitationsbasierte Reihungen von Zeitschriften intertemporal sehr stabil sind.

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

Top-Institutionen ausschließlich in den USA liegen. Bemerkenswert ist auch, dass diese 10 Institutionen für 30.9 % *aller* Einträge verantwortlich zeichnen. Dies ist ein sehr hoher Konzentrationsgrad – deutlich höher als in den anderen beiden Zeitschriftengruppen – der zweifelsohne auf Oligopoltendenzen in diesen Kernzeitschriften unserer Profession hinweist.[8] In den anderen Zeitschriftengruppen sind US-Institutionen deutlich weniger stark vertreten, was auf eine klare Spezialisierung dieser Institutionen auf die heimischen (US) Zeitschriften schließen lässt. Publikationen in Randzeitschriften (und als solche sind britische und europäische Zeitschriften aus der Sicht dieser Institutionen einzustufen) sind nur die zweite Wahl.

Tabelle 1

Zeitschriftenauswahl für deutsche Studie

	Impact-Faktor[a)]	Artikel (1997 – 2001)
Gruppe A: Amerikanische Zeitschriften		
Quarterly Journal of Economics	3.75	206
Journal of Political Economy	2.46	253
Econometrica	1.96	271
American Economic Review[b)]	1.85	408
Gruppe B: Britische Zeitschriften		
Review of Economic Studies	1.49	183
Economic Journal	1.31	420
Cambridge Journal of Economics	0.93	225
Oxford Economic Papers	0.71	192
Gruppe C: Kontinentaleuropäische Zeitschriften		
European Economic Review[c)]	0.86	488
Journal of Institutional and Theoretical Economics	0.49	266
Scandinavian Journal of Economics	0.46	202
Kyklos	0.37	130

[a)] Durchschnittlicher Impact-Faktor von 1997 bis 2001; Quelle: Journal Citation Reports.
[b)] Die *Papers and Proceedings* sind nicht berücksichtigt, weil darin keine referierten Artikel abgedruckt werden.
[c)] Die Beiträge der Jahrestagung sind berücksichtigt, weil sie referiert sind.

Während keine britischen oder europäischen Institutionen in die Top 10 der US-Journals kommen, gelingt es einigen renommierten Institutionen innerhalb der USA dennoch, in den Top 10 der Zeitschriftengruppen B und C aufzuscheinen. Die

[8] In dieser Richtung äußern sich etwa explizit *Hodgson* und *Rothman* (1999) oder *Brinkmann* (2003).

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

britische Gruppe B wird allerdings ganz klar von britischen Institutionen dominiert. Sieben der Top 10-Institutionen liegen in Großbritannien. Der Konzentrationsgrad der Top 10 liegt hier bei 18.9 %.

Erst in Gruppe C (den europäischen Zeitschriften) sind die ersten kontinentaleuropäischen Institutionen unter den Top 10 zu finden. Institutionen aus 5 kontinentaleuropäischen Staaten scheinen unter den Top 10 auf, wobei mit der Universität Zürich die einzige Institution aus dem deutschsprachigen Bereich vertreten ist. Auf den Rängen 12, 13, und 16 liegen die Universitäten München, Bonn bzw. Humboldt Berlin. Der Konzentrationsgrad in den kontinentaleuropäischen Zeitschriften ist mit 14.7 % am niedrigsten von allen 3 Gruppen, was zum einen auf die stärkere Heterogenität (bzgl. der geografischen Herkunft) der berücksichtigten Zeitschriften, aber zum anderen auch auf eine grundsätzlich stärkere Fragmentierung der ökonomischen Spitzenforschung auf dem europäischen Sub-Kontinent zurückzuführen ist. Im Folgenden werden wir die Publikationsaktivitäten der deutschsprachigen Institutionen im Detail analysieren.

Tabelle 3 listet alle Institutionen aus dem deutschsprachigen Raum auf, die zumindest zwei Mal in einer der 12 Zeitschriften genannt wurden.[9] Neben den 48 aufgezählten Institutionen gibt es noch 61 weitere Institutionen mit jeweils einem Eintrag. Insgesamt machen die Nennungen von Institutionen aus dem deutschsprachigen Raum 8.0 % der gesamten Nennungen in allen 12 Zeitschriften aus. Der Anteil der deutschsprachigen Institutionen schwankt natürlich erheblich in den einzelnen Zeitschriftengruppen, wie aus Tabelle 4 ersichtlich wird. In den amerikanischen Zeitschriften zeichnen Institutionen aus dem deutschsprachigen Raum für lediglich 2.6 % der Einträge verantwortlich. Dies zeigt, dass die deutschsprachigen Ökonomen in den Zeitschriften mit der höchsten Reputation lediglich marginal vertreten sind. Nicht viel besser ist die Situation in den britischen Zeitschriften mit einem Anteil von 4.6 %. Erwartungsgemäß am besten schneiden die deutschsprachigen Institutionen in den kontinentaleuropäischen Zeitschriften mit 17.2 % ‚Marktanteil' ab. Dieser Anteil ist allerdings wiederum hohen Schwankungen unterworfen, wie ein Blick auf Tabelle 5 zeigt, wo die besten deutschsprachigen Institutionen separat für die 4 Zeitschriften der Gruppe C aufgeführt werden, um besser zwischen *European Economic Review* und *Scandinavian Journal of Economics* auf der einen Seite und den im deutschen Sprachraum herausgegebenen Zeitschriften *JITE* und *Kyklos* differenzieren zu können. Während die relativen Anteile deutschsprachiger Institutionen in *EER* und *Scandinavian Journal of Economics* unter 10 % liegen, stammen 30 % bis 40 % der Arbeiten in *JITE* und

[9] Es ist zu beachten, dass wir keine Produktivitätsmaße berechnen und dass dadurch kleinere Universitäten unmöglich auf den ersten Rängen landen können, auch wenn ihr Forschungsoutput relativ gesehen sehr hoch ist. Uns geht es hier allerdings – wie erwähnt – nicht um das Abschneiden einzelner Institutionen im Ranking, sondern um die Stabilität der Rankings hinsichtlich unterschiedlicher Zeitschriften, die als Datenbasis herangezogen werden können.

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

Tabelle 2: **Top 10-Institutionen in Zeitschriftengruppen (1997 – 2001)**

Gruppe A (USA)			Gruppe B (UK)			Gruppe C (EU)		
Institution	Staat	Einträge	Institution	Staat	Einträge	Institution	Staat	Einträge
Harvard U	USA	107	U Cambridge	UK	49	U Stockholm	SWE	41
U Chicago	USA	98	LSE	UK	39	LSE	UK	37
MIT	USA	88	U Warwick	UK	38	Harvard U	USA	27
Princeton U	USA	65	Oxford U	UK	29	U Coll London	UK	26
Stanford U	USA	49	U Coll London	UK	28	Free U Brussels	BEL	24
U CA Berkeley	USA	48	U Nottingham	UK	28	Norw. School Ec	NOR	19
Federal Reserve	USA	46	Harvard U	USA	27	U Oslo	NOR	19
Northwestern U	USA	42	U Essex	UK	27	U Zürich	CH	19
U MI	USA	40	Princeton U	USA	21	Stanford U	USA	17
Columbia U	USA	33	U Chicago	USA	20	Tilburg U	NL	17
Top 10-Anteil an Gesamtnennungen			Top 10-Anteil an Gesamtnennungen			Top 10-Anteil an Gesamtnennungen		
Gruppe A 30.3 %			Gruppe B 18.9 %			Gruppe C 14.7 %		

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

Kyklos[10] von Autoren, die an Institutionen im deutschsprachigen Raum angestellt sind. 4 Institutionen (Universitäten Freiburg, München, Wien, Zürich) sind in allen vier kontinentaleuropäischen Zeitschriften vertreten; 13 weitere in drei der vier Zeitschriften, d. h. in mindestens einer der beiden Zeitschriften, die nicht im deutschsprachigen Raum herausgegeben werden (Universitäten Würzburg, Osnabrück, Mannheim, Lausanne, Konstanz, Kiel, Heidelberg, Hamburg, Dortmund, Bonn, Basel, TU Dresden, HU Berlin, FU Berlin). Diese 17 Institutionen sind auch in anderen Rankings deutschsprachiger Institutionen im vorderen Feld vertreten (siehe etwa *Bommer* und *Ursprung,* 1998 und *Ursprung,* 2003). Mithin wird der Schluss nahegelegt, dass diese Universitäten keine Spezialisierung auf Publikationsorgane im deutschsprachigen Raum aufweisen, sondern auch in anderen führenden europäischen Zeitschriften präsent sind und dort zu den besten Institutionen aus dem deutschen Sprachraum zählen.[11] Es kann also nicht davon ausgegangen werden, dass die führenden deutschsprachigen Institutionen (allen voran die Universitäten in Zürich, Bonn, München oder Mannheim) eine einseitige Orientierung an den (weltweiten und europäischen) Spitzenzeitschriften aufweisen und den eigenen, ‚lokalen' Markt vernachlässigen.[12] Vielmehr zeigt sich, dass jene Institutionen, die international erfolgreich sind, auch national zu den besten Institutionen zählen, was natürlich auch vice versa gilt. Im folgenden Abschnitt stellen wir überblicksartig eine Studie dar, die für den kanadischen Markt untersucht, ob es eine Segregation von Forschungsinstitutionen in Kanada hinsichtlich der Publikationen in weltweiten Top-Zeitschriften oder in *dem* heimischen Journal, dem *Canadian Journal of Economics,* gibt.

---

10 *JITE* und *Kyklos* wurden ausgewählt, weil sie über den untersuchten Zeitraum die höchsten durchschnittlichen Impact-Faktoren aufweisen. Eine logische Erweiterung unserer Studie mit einem noch stärkeren Focus auf europäische Zeitschriften müsste das Sample der in deutschsprachigen Staaten herausgegebenen Zeitschriften um einige prominente Zeitschriften erhöhen.

11 Die Korrelationskoeffizienten zwischen den Einträgen in den vier verschiedenen Zeitschriften der Gruppe C sind in der Regel positiv und signifikant (die geringe Zahl an Beobachtungen führt mitunter zu insignifikanten Ergebnissen).

12 Natürlich sprechen wir hier nur vom Forschungsoutput im engeren Sinne (d. h. mit klarem Focus auf die scientific community) und beispielsweise nicht von Transferleistungen, Beratungstätigkeit und ähnlichen universitären Aufgaben.

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

Tabelle 3: **Institutionen aus dem deutschsprachigen Raum in den 12 ausgewählten Zeitschriften (1997 – 2001)***

Universität	Einträge	Universität	Einträge	Universität	Einträge
U Zürich	39	U Basel	7	Studienzentrum Gerzensee	3
U Bonn	29	U Bern	7	U Hagen	3
U München	27	U Bielefeld	7	U Hannover	3
HU Berlin	21	U Regensburg	6	U Ulm	3
U Mannheim	21	U St Gallen	6	U Würzburg	3
FU Berlin	11	WZB Berlin	6	WU Wien	3
U Lausanne	11	U Frankfurt am Main	6	Goldman, Sachs Frankfurt am Main	2
U Dortmund	10	Swiss National Bank	5	Inst. Arbeit Technik Gelsenkirchen	2
U Heidelberg	10	U Tübingen	5	U Halle-Wittenberg	2
U Linz	10	U Kiel	5	Max-Planck-Institut Jena	2
U Saarland	10	Europa U Frankfurt / Oder	4	U Dresden	2
U Wien	10	U Innsbruck	4	U Duisburg	2
U Köln	9	U Münster	4	U Essen	2
U Freiburg	8	U Osnabrück	4	U Giessen	2
U Hamburg	8	U Genf	3	U Göttingen	2
U Konstanz	8	IHS Wien	3	U Koblenz-Landau	2

* aufgenommen wurden alle Institutionen mit mehr als einem Eintrag.

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

Tabelle 4: **Institutionen aus dem deutschsprachigen Raum in den 3 Zeitschriftengruppen (1997 – 2001)**

Gruppe A		Gruppe B		Gruppe C	
Universität	Einträge	Universität	Einträge	Universität	Einträge
U Zürich	14	HU Berlin	6	U Zürich	19
U Bonn	8	U Mannheim	6	U München	16
U Mannheim	4	U Zürich	5	U Bonn	15
U München	3	FU Berlin	4	HU Berlin	14
FU Berlin	2	U Bonn	3	U Mannheim	11
U Bielefeld	2	U Linz	3	U Saarland	7
U Regensburg	2	13 Institutionen	2	U Freiburg	8
U Wien	2			U Heidelberg	8
WZB Berlin	2			U Lausanne	8
15 Institutionen	1			U Wien	8
				U Bern, Dortmund, Hamburg, Köln	7
Anzahl Institutionen deutscher Sprachraum	24	Anzahl Institutionen deutscher Sprachraum	40	Anzahl Institutionen deutscher Sprachraum	90
Anteil deutscher Sprachraum an Gesamtnennungen	2,6 %	Anteil deutscher Sprachraum an Gesamtnennungen	4,6 %	Anteil deutscher Sprachraum an Gesamtnennungen	17.2 %

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

Tabelle 5: **Institutionen aus dem deutschsprachigen Raum in den 4 Zeitschriften der Gruppe C (1997 – 2001)**

EER		Scan J Econ		JITE		Kyklos	
Universität	Einträge	Universität	Einträge	Universität	Einträge	Universität	Einträge
U München	7	U Dortmund	3	HU Berlin	11	U Bern	6
U Zürich	7	U München	3	U Bonn	9	U Heidelberg	4
U Lausanne	6	U Mannheim	2	U Zürich	9	U Linz	4
U Bonn	5	U Freiburg	2	U Saarland	8	U St Gallen	4
BIZ Basel	3	U Ulm	2	U Mannheim	7	U Hannover	3
DIW Berlin	3	U Zürich	1	U München	7	WZB Berlin	3
U Wien	3	FU Berlin	1	U Köln	6	U Basel	2
FU Berlin	2	U Wien	1	U Frankfurt	5	U Essen	2
HU Berlin	2	U Bonn	1	U Münster	4	U Freiburg	2
Gerzensee	2	U Basel	1	U Regensburg	4	U Hamburg	2
U Bielefeld	2	U Konstanz	1	U Tübingen	4	U München	2
U Dortmund	2	U Innsbruck	1	U Freiburg	3	U Zürich	2
U Hamburg	2	TU Wien	1	U Hamburg	3	U Wien	1
U Konstanz	2	U Kiel	1	U Heidelberg	3	HU Berlin	1
U Mannheim	2	8 weitere	1	U Wien	3	U Kiel	1
19 weitere	1			U Konstanz	3	14 weitere	1
				10 weitere	2		
				30 weitere	1		
# Institutionen	34	# Institutionen	22	# Institutionen	56	# Institutionen	29
Relativer Anteil	8.1 %	Relativer Anteil	9.0 %	Relativer Anteil	40.4 %	Relativer Anteil	30.5 %

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

D. Die kanadische Studie

Davies, Kocher und *Sutter* (2003) haben für Kanada erhoben, inwieweit kanadische Institutionen, die erfolgreich in den führenden Zeitschriften der Welt publizieren, auch in ihrem lokalen Journal, dem *Canadian Journal of Economics,* vertreten sind. Dahinter steht die Frage, ob es so etwas wie eine Spezialisierung einzelner Institutionen auf den lokalen Publikationsmarkt bzw. auf die Spitzenzeitschriften gibt; mit anderen Worten, ob die Publikationsmuster kanadischer Institutionen stark segregiert sind oder nicht.

Als Spitzenzeitschriften werden die zehn Zeitschriften mit den höchsten durchschnittlichen Impact-Faktoren in den Jahren von 1980 bis 2000 bezeichnet. Zur Wahl standen dabei ausschließlich Zeitschriften, die in der Kategorie ‚Economics' der JCR aufscheinen und seit 1980 ohne Unterbrechung publiziert worden sind. Tabelle 6 listet die ausgewählten Zeitschriften und deren durchschnittliche Impact-Faktoren von 1980 bis 2000 auf.[13]

Die Publikationen (mit allen aufgeführten Autoren und aufgelisteten Institutionen) wurden direkt aus den jeweiligen Zeitschriften erhoben.[14] Aufgrund des enormen Aufwandes haben wir nur die geraden Jahre von 1980 bis 2000 erfasst. Für die ‚Top 10' ökonomischen Fachzeitschriften ergibt dies eine Gesamtzahl von 5.384 Artikeln; im *Canadian Journal of Economics* waren es im selben Zeitraum 599 Artikel.

Aufgrund des deutlich größeren Datensatzes als in der deutschen Studie haben wir für die kanadische Studie eine Gewichtung der Einträge für einzelne Institutionen nach der Anzahl der Autoren vorgenommen. Ist N_i die Anzahl der Autoren eines Artikels *i,* dann ist das einem Autor und seiner Institution (seinen Institutionen) zugewiesene Gewicht $1/N_i$.[15]

In den ‚Top 10'-Zeitschriften stammen 77 % aller Artikel aus der Feder von Autoren, die an einer amerikanischen Universität angestellt sind. Dieser Anteil schwankt zwischen 70.5 % im Jahr 2000 und 81.9 % im Jahr 1996. Das Vereinigte Königreich (im Wesentlichen England) und Kanada sind auf dem zweiten bzw. dritten Platz mit 5.1 % bzw. 4.8 % der Artikel. Der Anteil Kanadas in den ‚Top

[13] Sieben der zehn von uns ausgewählten Zeitschriften scheinen auch in einer Befragung von deutschsprachigen Ökonomen unter den zehn mit dem höchsten Renommee auf. Die restlichen drei folgen knapp dahinter (siehe *Bräuninger* und *Haucap,* 2001).

[14] Die elektronische Datenbank ‚EconLit' z. B., weist vor 1990 keine Institutionen der Autoren aus und führt im Allgemeinen nicht mehr als 3 Autoren pro Artikel an.

[15] *Kocher* und *Sutter* (2001) zeigen, dass die Ergebnisse bzgl. des Forschungsoutputs von Institutionen nicht sensitiv auf die Art der Gewichtung nach der Anzahl der Autoren reagieren. Ob jeder Autor einer Forschungsarbeit, die von mehreren Koautoren erstellt wurde, gleich stark gewichtet wird wie ein Einzelautor oder nicht, spielt auf institutioneller Ebene praktisch keine Rolle. Die geringe Sensitivität von Rankings in Bezug auf Veränderungen der Gewichtungsregeln scheint übrigens ein Phänomen zu sein, das sich durchgängig zeigt (z. B. in *Coupé,* 2003, und *Ursprung,* 2003).

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

10‘-Zeitschriften reicht von 3.1 % in 1996 bis 7.2 % in 1988. Alle anderen Staaten der Welt zusammen tragen weniger als 13 % aller Publikationen bei.

Tabelle 6

Zeitschriftenauswahl für kanadische Studie

Zeitschrift	IF[a)]	Zeitschrift	IF[a)]
Journal of Economic Literature	5.01	Quarterly Journal of Economics	2.12
Journal of Financial Economics	2.65	American Economic Review	1.68
Brookings Papers on Economic Activity	2.53	Journal of Law and Economics	1.63
Journal of Political Economy	2.38	Review of Economic Studies	1.34
Econometrica	2.17	Journal of Monetary Economics	1.33
Canadian Journal of Economics (42.)	0.43		

[a)] IF: Durchschnittlicher Impact-Faktor von 1980 bis 2000; Quelle: Journal Citation Reports.

Im *Canadian Journal of Economics* sind 55 % der Publikationen von kanadischen Autoren geschrieben. Allerdings gibt es einen stetigen Rückgang des Anteils kanadischer Autoren von 70 % in den frühen 80er Jahren auf 42 % im Jahr 2000. Dieser Rückgang spiegelt die zunehmende Internationalität der Zeitschrift wider. Die USA sind im CJE am zweitstärksten vertreten mit 27 % der Aufsätze. Danach folgen das Vereinigte Königreich (2.3 %), Japan (1.5 %) und Australien (1.3 %).

In Tabelle 7 sind die (gewichteten) Einträge für alle in den betreffenden Zeitschriften zumindest ein Mal aufscheinenden kanadischen Institutionen aufgeführt. Die besten fünf Institutionen sind in beiden Kategorien (‚Top 10‘ vs. *CJE*) die Universitäten von Western Ontario, Toronto und British Columbia, bzw. die Queen’s und die McMaster Universität. Insgesamt scheinen 33 verschiedene kanadische Institutionen im *CJE* auf bzw. 26 Institutionen in den ‚Top 10‘-Zeitschriften. Die Korrelation zwischen den (gewichteten) Einträgen im *CJE* bzw. den ‚Top 10‘-Zeitschriften ist bemerkenswert hoch ($r = 0.9$; $p < 0.01$). Dies zeigt eindeutig, dass jene Institutionen, die erfolgreich in den weltbesten Zeitschriften publizieren, auch am lokalen Markt dominieren. Allerdings ist die Konzentration der Publikationen einzelner kanadischer Institutionen im *CJE* deutlich niedriger (Herfindahlindex = 0.064) als in den ‚Top 10‘-Zeitschriften (Herfindahlindex = 0.121). Letzteres Resultat impliziert, dass die Publikationen im *CJE* gleichmäßiger auf verschiedene kanadische Institutionen verteilt sind als die Publikationen in den ‚Top 10‘-Zeitschriften. Dies ist ein leichtes Indiz dafür, dass sich einige Institutionen in ihrem Publikationsoutput[16] stärker auf die führenden Zeitschriften konzentrieren – ohne allerdings den lokalen Markt dabei zu vernachlässigen.

[16] Und damit auch in ihrer Berufungspolitik.

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

Tabelle 7

Institutionelle Zugehörigkeit der Autoren in ‚Top 10'-Zeitschriften und im Canadian Journal of Economics (1980 – 2000)

Kanadische Institutionen	CJE[a)]	Top 10[a)]	Ausgewählte andere Institutionen	CJE[a)]	Top 10[a)]
Acadia U	1.0	1.0	Australian National U	3.0	15.6
Bank of Canada	2.8	4.2	Hebrew U	2.5	37.2
Brock U	5.3	0.0	Tel Aviv U	0.7	53.2
Carleton U	12.8	12.6			
Concordia U	9.0	1.5	LSE	3.0	74.5
Dalhousie U	4.6	1.0	U Cambridge	0.5	19.8
Lakehead U	4.0	0.0	U Oxford	0.0	25.2
McGill	5.3	2.8			
McMaster	18.0	14.7	Top 10 USA		
Queen's U	30.5	21.3	Harvard U	0.5	263.4
Simon Fraser	4.2	10.2	U Chicago	1.0	225.4
Trent U	3.0	0.0	MIT	3.0	209.2
U Alberta	15.2	7.3	Stanford U	4.0	165.8
U British Columbia	36.1	37.5	Princeton U	1.0	149.5
U Calgary	11.3	1.3	U Pennsylvania	2.3	139.0
U Guelph	6.5	4.0	Northwestern U	0.8	127.5
U Laval	6.5	3.7	Yale U	0.0	120.4
U Lethbridge	0.5	0.0	UC Berkeley	3.0	101.8
U Manitoba	3.0	0.0	Columbia U	3.0	91.0
U Montreal	17.3	9.2			
U New Brunswick	2.0	1.0	U Linz		1.8
U Ottawa	4.5	3.3	IHS Wien		1.3
U Quebec	6.0	2.2	TU Wien		1.8
U Regina	2.0	0.0	U Wien		0.8
U Saskatchewan	5.5	1.7			
U Toronto	22.5	46.2	WZB Berlin		2.8
U Victoria	5.0	0.5	U Bonn		10.2
U Waterloo	6.0	3.7	U Mannheim		6.2
U Western Ontario	45.8	55.7			
U Windsor	6.3	2.5	U Genf	2.33	6.3
U Winnipeg	1.5	0.0	U Zürich		6.7
Wilfried Laurier U	8.0	0.5			
York U	11.0	4.8	Anzahl der Artikel	599	5384

[a)] Gesamteinträge für gerade Jahre von 1980 bis 2000. Gewichtung mit $1/N_i$, wobei N_i die Anzahl der Autoren eines Artikels bezeichnet.

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

Zusätzlich zu den kanadischen Institutionen sind die führenden Zeitschriften aus anderen Staaten zum Vergleich aufgelistet. Institutionen aus dem deutschen Sprachraum sind erwartungsgemäß nur sehr selten im *Canadian Journal of Economics* vertreten (in Summe 7.3 gewichtete Einträge). Auch die Anzahl der Einträge in den ‚Top 10'-Zeitschriften ist gering. Hier sind allerdings wiederum jene Institutionen am stärksten vertreten, die auch in unserer deutschen Studie insgesamt (d. h. in allen drei Zeitschriftengruppen) am stärksten vertreten waren (U Bonn, U Mannheim, U Zürich).

E. Zusammenfassung

Unsere Ergebnisse legen den Schluss nahe, dass es keinen Trade-off zwischen dem Forschungsoutput in weltweiten Spitzenzeitschriften und dem ‚nationalen' Markt gibt. Sobald wissenschaftliche Forschung als ein Ziel bzw. als eine Aufgabe einer Institution definiert worden ist, hat die Auswahl der Zeitschriften, auf deren Basis Evaluationen durchgeführt werden, nur sehr bedingte Auswirkungen auf das Ergebnis, solange man sich auf eindeutig wissenschaftliche Zeitschriften mit gewissen Qualitätsstandards beschränkt. Dies gilt sowohl prinzipiell für den deutschen Sprachraum als auch für Kanada. Letzteres Beispiel zeigt etwa, dass es eine außergewöhnlich hohe Korrelation ($r > 0.9$, $p < 0.01$) zwischen dem Forschungsoutputs einzelner Institutionen im *Canadian Journal of Economics* und den ‚Top 10'-Zeitschriften gibt. Die Korrelation zwischen den Publikationen von Universitäten aus dem deutschen Sprachraum im *JITE,* der renommiertesten deutschen Fachzeitschrift, und den US- und britischen Zeitschriften (in Summe) ist ebenfalls hoch signifikant und positiv ($r = 0.62$, $p < 0.01$). Diese Ergebnisse legen den Schluss nahe, dass der Einwand, Evaluationen auf der Grundlage von Publikationen in weltweiten Spitzenzeitschriften würden zu systematischen Verzerrungen führen, nicht gerechtfertigt ist.

Die Auswahl der Zeitschriften (oder Datenbanken, wie etwa WISO oder SSCI) für die Forschungsevaluation spielt also keine zentrale Rolle für deren Ergebnisse. Sogar die Beschränkung auf Zeitschriftenaufsätze scheint weniger problematisch als vielfach von Verfechtern der Bedeutung von Monographien angenommen: Eine kürzlich durchgeführte Forschungsevaluation an der Heimatuniversität der Autoren (siehe *Egger* et al., 2003) hat gezeigt, dass jene Institutionen und sogar einzelne Forscherinnen und Forscher, die in SSCI-Zeitschriften vergleichsweise gut publizieren, auch in anderen Publikationstypen (wie Bücher, Buchbeiträge, Buchherausgeberschaften, Rezensionen, Nicht-SSCI-Aufsätze) die Nase vorne haben.[17] Wenn überhaupt, findet also eine ‚Spezialisierung' von Forschern weniger in der

[17] Es gibt sogar einige, wenn auch noch nicht endgültige Evidenz, dass eine positive Korrelation zwischen Forschungsoutput und volkswirtschaftlicher Politikberatung besteht. D. h., jene Wissenschaftler, die in der scientific community einen guten Ruf genießen, werden vermehrt auch als Berater herangezogen und geschätzt (*Wagner,* 2003).

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

Auswahl von Publikationsorganen oder -typen ihren Niederschlag, sondern in der Grundsatzentscheidung ‚Publizieren!' versus ‚Nicht publizieren!'.[18] Dass letztere ‚Spezialisierung' durch Evaluationen aufgedeckt werden sollte, versteht sich unserer Ansicht nach von selbst.

Literatur

Backes-Gellner, U. (1989): Ökonomie der Hochschulforschung. Organisationstheoretische Überlegungen und betriebswirtschaftliche Befunde. Wiesbaden, Gabler.

Backes-Gellner, U. / *Zanders,* E. (1989): Lehre und Forschung als Verbundproduktion. Zeitschrift für Betriebswirtschaft, vol. 59(3), pp. 271 – 290.

Bommer, R. / *Ursprung,* H. W. (1998): Spieglein, Spieglein an der Wand. Eine publikationsanalytische Erfassung der Forschungsleistungen volkswirtschaftlicher Fachbereiche in Deutschland, Österreich und der Schweiz. Zeitschrift für Wirtschafts- und Sozialwissenschaften, vol. 118(1), pp. 1 – 28.

Bräuninger, M. / *Haucap,* J. (2001): Was Ökonomen lesen und schätzen: Ergebnisse einer Umfrage. Perspektiven der Wirtschaftspolitik, vol. 2(1), pp. 185 – 210.

Brinkmann (2003): Englisch: Die Hoffnung der deutschsprachigen Nationalökonomen. Ein Diskussionsbeitrag. in: Backes-Gellner, U. / Schmidtke, C. (Hrsg): Hochschulökonomie – Analysen interner Steuerungsprobleme und gesamtwirtschaftlicher Effekte, Berlin, Duncker & Humblot, pp. 51 – 61.

Coupé, T. (2003): Revealed performances. Worldwide rankings of economists and economics departments. Journal of the European Economic Association, vol. 1(6), pp. 1309 – 1345.

Davies, J. B. / *Kocher,* M. G. / *Sutter,* M. (2003): Economics research in Canada. A long-run assessment of journal publications. University of Western Ontario, University of Innsbruck, mimeo.

Egger, P. / *Kocher,* M. G. / *Pfaffermayer,* M. / *Sutter,* M. / *Winner,* H. (2003): Nabelschau – Eine Evaluation der betriebswirtschaftlichen und volkswirtschaflichen Forschungsleistung an der SoWi Innsbruck. University of Innsbruck, mimeo.

Elliott, C. / *Greenaway,* D. / *Sapsford,* D. (1998): Who's publishing who? The national composition of contributors to some core US and European journals. European Economic Review, vol. 42(1), pp. 201 – 206.

Hodgson, G. M. / *Rothman,* H. (1999): The editors and authors of economics journals: A case of institutional oligopoly?. Economic Journal, vol. 109(2), pp. F165 – 186.

Kalaitzidakis, P. / *Mamuneas,* T. P. / *Stengos,* T. (1999): European economics: An analysis based on publications in the core journals. European Economic Review, vol. 43(4 – 6), pp. 1150 – 1168.

Kalaitzidakis, P. / *Mamuneas,* T. P. / *Stengos,* T. (2003): Ranking of academic journals and institutions in economics. Journal of the European Economic Association, vol. 1(6), pp. 1346 – 1366.

[18] Einen ersten Befund in dieser Richtung lieferte bereits *Backes-Gellner* (1989).

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

Kocher, M. G. / *Sutter,* M. (2001): The institutional concentration of authors in top journals of economics during the last two decades. Economic Journal, vol. 111(5), F405 – 421.

Laband, D. N. / *Piette,* M. J. (1994): The relative impacts of economics journals: 1970 –1990. Journal of Economic Literature, vol. XXXII (2), pp. 640 – 666.

Schlinghoff, A. (2002): Personalauswahl an Universitäten – die Berufungspraxis deutscher wirtschaftswissenschaftlicher Fakultäten in den neunziger Jahren. Zeitschrift für Betriebswirtschaft – Ergänzungsheft 2 / 2002, vol. 72, pp. 139 – 147.

Sutter, M. / *Kocher,* M. G. (2001): Tools for evaluating research output: Are citation-based rankings of economics journals stable?. Evaluation Review, vol. 25(5), pp. 555 – 66.

Ursprung, H. W. (2003): Schneewittchen im Land der Klapperschlangen: Evaluation eines Evaluators. Perspektiven der Wirtschaftspolitik, vol. 4(2), pp. 177 – 189.

Wagner, G. G. (2003): Qualität der Volkswirtschaftslehre und der volkswirtschaftlichen Politikberatung – Eine Evaluation auf Basis einer Expertenbefragung. Vortrag.

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

Dienstleistungsproduktion in Hochschulen

Zur Beschreibung der Qualitäten der externen Faktoren „Studierende“

Von *Günter Fandel* und *Steffen Blaga,* Hagen*[1]

A. Vorbemerkungen

In der Literatur gibt es zahlreiche Ansätze, die Anstrengungen von Hochschulen in Lehre und Forschung als Dienstleistungsproduktion zu verstehen und durch produktionstheoretische Modelle zu beschreiben (*Albach et al.* 1978; *Stieger* 1980; *Fandel / Paff* 2000). Im Vordergrund der Analysen stehen dabei Überlegungen zur Effizienz bzw. zu Kosten und Erfolg von Hochschulen (*Fandel* 1995, 2003; *Fandel / Paff* 2000). Die Ergebnisse derartiger Untersuchungen bilden mitunter die Basis für Umverteilungen von Mitteln zwischen den Hochschulen (*Fandel / Gal* 2001).

In der Hochschulausbildung werden die Studierenden üblicherweise als Inputs und die Absolventen als Outputs dieser Dienstleistungsproduktion aufgefasst. Aus outputorientierter Perspektive bemisst man den Erfolg der Hochschulen in der Lehre oft danach, wie hoch die Abbrecherquote bzw. drop-out-Rate ist (*Smith / Naylor* 2001), wie hoch die durchschnittliche Studiendauer in bestimmten Fachrichtungen ausfällt (*Hackl / Sedlacek* 2002a, 2002b) und in welchem Maße die Absolventen am Arbeitsmarkt der Akademiker untergebracht werden können (*Lohmar et al.* 1982; BfA 2000). Letzteres wird wesentlich durch die Diplomnote bestimmt; sie dient als Anhaltspunkt für die Einschätzung der Qualität des Absolventen einer bestimmten Hochschule. Aus inputorientierter Sicht beurteilt man

---

* Prof. Dr. Günter Fandel und Dipl.-Kfm. Steffen Blaga, Lehrstuhl für Betriebswirtschaft, insbes. Produktions- und Investitionstheorie, Fachbereich Wirtschaftswissenschaft, FernUniversität in Hagen, Universitätsstraße 41, 58084 Hagen.

[1] Wir danken dem Geschäftsführer des Prüfungsamtes des Fachbereichs Wirtschaftswissenschaft der FernUniversität in Hagen, Herrn Dr. Heinz Hoffmann, für die tatkräftige Unterstützung bei der Datenerhebung und -auswertung sowie Herrn Dipl.-Kfm. Dipl.-Volksw. Thomas Mazzoni, Lehrstuhl für Statistik, FernUniversität in Hagen, für seine inhaltlichen Anregungen und Diskussionsbereitschaft. Ebenso bedanken wir uns bei zwei anonymen Gutachtern für ihre konstruktiven Anregungen und Verbesserungsvorschläge.

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

dagegen die Leistungsfähigkeit einer Hochschule danach, wie viele Studierende sie bei den gegebenen Ressourcen aufnimmt und im Grundstudium betreut (*Fandel / Gal* 2001). Schenkt man der input- und outputorientierten Sicht gleichermaßen Beachtung, so wird bildungspolitisch erwartet, dass eine Hochschule bei den vorhandenen Kapazitäten möglichst viele Studierende aufnimmt und davon möglichst viele erfolgreich zum Abschluss bringt. Dabei wird unterstellt, dass die ausbildende Hochschule durch die Qualität ihrer Lehre die Qualität der Absolventen beeinflusst.

Ein Aspekt ist in diesen Erörterungen kaum untersucht worden: In welchem Maße der Erfolg von Hochschulen in der Lehre, gemessen an der Qualität ihrer Absolventen, von der Qualität der Studierenden als externe Faktoren (*Maleri* 1973) dieser Dienstleistungsproduktion selbst abhängt und wie sich diese Qualität im Verlauf des Studiums bis hin zum Abschluss im Produktionsprozess der Lehre entwickelt. Damit ist das bekannte, aber ungelöste produktionstheoretische Problem der Hochschulausbildung angesprochen, ob es sich bei der Qualität der Studierenden um eine Eigenschaft des Inputs oder des Outputs von Universitäten handelt (*Albach et al.* 1978). Dabei könnte man die Qualität der Studierenden als Inputs nach ihren Abiturnoten oder nach Noten bemessen, die sie in verpflichtenden Fächern im Verlauf ihres Studiums erreichen. Eine neuere Studie von *Fries* (2002) hat gezeigt, dass die Abiturnote für das untersuchte Datenmaterial den mit Abstand besten Indikator für den Studienerfolg darstellt.

B. Gegenstand der Untersuchungen

Im Folgenden wird für 272 Absolventen der Fachrichtung Wirtschaftswissenschaft an der FernUniversität in Hagen, die ihr Vollzeitstudium in den Jahren 2000 – 2002 abgeschlossen haben, beschrieben, wie ihre Qualität als Studierende bei Eingang ins Studium mit ihrer Qualität im Vordiplom bzw. ihre Qualität im Vordiplom mit ihrer Qualität als Absolventen zusammenhängt. Als Qualitätsindikatoren der einzelnen Ausbildungsstufen dienen hierbei der Notendurchschnitt der ersten vier Pflichtklausuren des Grundstudiums als Eingangsqualifikation, die Vordiplomnote als Zwischenqualifikation sowie die Diplomnote als Endqualifikation der Studierenden. Dabei wird unterstellt, dass sich die Qualität des Ausbildungsprozesses des Fachbereichs Wirtschaftswissenschaft der FernUniversität in Hagen in den Jahren ihres Studiums nicht verändert hat. Auf der Grundlage von Wanderungen der Studierenden von den Eingangsqualifikationen zu den Vordiplomnoten und von den Vordiplomnoten zu den Diplomnoten werden Migrationsmatrizen hergeleitet, die produktionstheoretisch gemäß der Aktivitätsanalyse erklären, wie aus Studierenden mit unterschiedlichen Eingangsqualifikationen bzw. Vordiplomnoten (Vorprodukte unterschiedlicher Qualität) Studierende mit bestimmten Vordiplomnoten bzw. Absolventen mit bestimmten Diplomnoten (Zwischenprodukte bzw. Endprodukte bestimmter Qualität) entstehen. Formal handelt es sich

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

dabei um eine mehrstufige Mehrproduktfertigung. Die Produktionsfunktion ist im Prinzip substitutional, da sich Studierendenmengen mit bestimmten Vordiplom- bzw. Diplomnoten aus solchen unterschiedlicher Eingangsqualifikationen bzw. Vordiplomnoten zusammensetzen können. Die konkreten Migrationsmatrizen stellen dann aktivitätsanalytisch bestimmte Produktionsprozesse dar, die auf der Grundlage des Produktionszusammenhangs realisiert worden sind.

C. Rahmenbedingungen

Für die 272 Absolventen, die einen erfolgreichen Studienverlauf genommen haben, waren die Informationen über ihre Abiturnoten als Eingangsqualifikationen für das Studium nicht verfügbar. So werden hier die Qualitäten der Absolventen bei Eingang ins Studium (Eingangsqualifikationen) durch den Durchschnitt der Noten gemessen, die die Studierenden in den ersten vier nach Studienplan für das erste Semester empfohlenen Teilklausuren im Grundstudium erzielen. Diese Vorgehensweise hat gegenüber der sonst üblichen Heranziehung der Abiturnote den Vorteil einer homogenen Qualitätsbeurteilung, da der den Teilklausuren zugrunde liegende Prüfungsstoff durch die entsprechenden Studienbriefe weitgehend standardisiert und über die Zeit im Wesentlichen für alle Studierenden gleich geblieben ist. Die vier Teilklausuren beziehen sich auf die Fächer „Grundzüge der Betriebswirtschaftslehre I“, „Mikroökonomik“, „Mathematik für Wirtschaftswissenschaftler“ sowie „Wirtschaftsinformatik“. Zur Beurteilung der Qualitätseigenschaften der Absolventen in der Vordiplom-Prüfung und in der Diplom-Prüfung dienen die jeweiligen Durchschnittsnoten aus allen Teilklausuren, deren Bestehen zur erfolgreichen Ablegung dieser Prüfungen erforderlich ist. Diese Durchschnittsnoten sind zugleich die Zeugnisnoten. In den Urdaten sind die Durchschnittsnoten der vier Grundstudiumsklausuren, aus denen die Eingangsqualifikation berechnet wird, sowie die Vordiplom- und Diplomnoten mit einer Dezimalstelle nach dem Komma angegeben. Zur Charakterisierung von Qualitätsklassen werden hier aber nur ganze Noten auf der Notenskala von 1 – 4 herangezogen. Unter den 272 betrachteten Studierenden befindet sich keiner mit einer Eingangsqualifikation schlechter als „ausreichend“. Ab sechs Zehnteln nach dem Komma sind die gebrochenen Noten der Urdaten auf ganze Noten aufgerundet, sonst abgerundet worden. Während die Durchschnittsnote der vier Grundstudiumsklausuren als Qualität des Inputs „Studierende“ und die Note des Diploms als Qualität des Outputs „Absolventen“ interpretiert werden können, stellt die Note des Vordiploms sowohl eine Outputqualität wie auch eine Inputqualität dar.

Zuvor ist bereits unterstellt worden, dass sich die Qualität des Ausbildungsprozesses des Fachbereichs Wirtschaftswissenschaft der FernUniversität in den Jahren nicht verändert hat, in denen die 272 Absolventen ihr Studium durchgeführt haben. Implizit wird damit also von statischen Produktionsbedingungen ausgegangen. Das mag über den Verlauf von 11 Jahren, in denen über 90 % der Absolventen

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

ihr Studium beendet haben, als kritisch angesehen werden; dennoch gibt es dafür plausible Anhaltspunkte:

- Die Inhalte der Studienbriefe zu den für die Klausuren maßgeblichen Fächern sind in diesem Zeitraum weitgehend gleich geblieben. Das gilt insbesondere für alle Fächer des Grundstudiums und ebenso auch für alle Fächer des Hauptstudiums, die für alle Studenten verpflichtend sind.
- Die Betreuung der Studierenden hat sich in der angegebenen Zeitspanne kaum verändert, da die die Pflichtfächer betreuenden Hochschullehrer weitgehend dieselben geblieben sind und auch das Betreuungssystem durch die Mentoren in den Studienzentren der FernUniversität keine nennenswerten strukturellen Veränderungen erfahren hat.
- Die prüfenden Universitätsprofessoren sind dieselben geblieben; die von ihnen zusammengestellten Teilklausuren und die darin geforderten Prüfungsinhalte sind weitgehend ähnlich gestaltet worden.
- Die Notenvergabe in den für die Eingangsqualifikationen, Vordiplom- und Diplomnoten herangezogenen Klausuren folgte einem festen Schema der Punkteverteilung, das den Studierenden bekannt war und von dem im betrachteten Zeitraum nicht abgewichen worden ist.

Die Vordiplomnote setzt sich aus 10 Teilklausuren des Grundstudiums zusammen, zu denen auch die Teilklausuren gehören, aus denen die Eingangsqualifikation der Studierenden ermittelt wird. In die Diplomnote gehen die Noten der 5 Prüfungsklausuren ein; als weitere Prüfungsleistung kommt die Anfertigung der Diplomarbeit hinzu, deren Note bei der Ermittlung der Diplomnote doppelt gewichtet wird. Alle Teilklausuren können nach einem Credit-Point-System konsekutiv während des jeweiligen Studienabschnitts erbracht werden und werden bei Bestehen kumulativ zur Gesamtnote zusammengezählt. Dadurch werden zusätzlich Zufallsstörungen in der Notenvergabe vermieden, die durch inhaltlich und umfänglich zu unterschiedliche Gesamtklausuren entstehen könnten. So verwundert es nicht, dass die in Zehnteln ausgedrückten Noten der Eingangsqualifikationen der Studierenden ebenso wie ihre Vordiplom- und Diplomnoten annähernd normalverteilt sind. Diese Normalverteilungen sind nicht – wie man vermuten könnte – das Ergebnis von Notenanpassungen eines bestimmten Prüfungstermins. Vielmehr kommen bei den 272 erfolgreichen Studierenden die Noten aus Teilklausuren zusammen, die zu ganz unterschiedlichen Prüfungsterminen geschrieben worden sind. Die Diplomnoten fallen im Durchschnitt etwas besser aus als die Vordiplomnoten, was daran liegt, dass bei der Ermittlung der Diplomnote die Note der Diplomarbeit doppelt zählt. Die entsprechend aufbereiteten Daten können der Tabelle 1 entnommen werden.

Korrelationsrechnungen im Hinblick auf die Ränge, welche die 272 Studierenden untereinander in den Eingangsnoten, den Vordiplom- und Diplomnoten einnehmen, zeigen, dass ein Studierender im Vordiplom bzw. im Diplom einen umso

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

Tabelle 1

Verteilung der Eingangsqualifikationen, Vordiplom- und Diplomnoten der 272 Absolventen des Fachbereichs Wirtschaftswissenschaft der FernUniversität in Hagen

Notenstufe	Häufigkeit					
	Eingangsqualifikation		Vordiplomnote		Diplomnote	
	absolut	relativ	absolut	relativ	absolut	relativ
1,0	0	0,00%	0	0,00%	0	0,00%
1,1	2	0,74%	0	0,00%	0	0,00%
1,2	1	0,37%	1	0,37%	1	0,37%
1,3	1	0,37%	0	0,00%	0	0,00%
1,4	2	0,74%	2	0,74%	1	0,37%
1,5	5	1,84%	2	0,74%	2	0,74%
1,6	4	1,47%	3	1,10%	0	0,00%
1,7	2	0,74%	1	0,37%	5	1,84%
1,8	1	0,37%	6	2,21%	4	1,47%
1,9	3	1,10%	6	2,21%	6	2,21%
2,0	17	6,25%	5	1,84%	9	3,31%
2,1	10	3,68%	4	1,47%	6	2,21%
2,2	11	4,04%	7	2,57%	12	4,41%
2,3	13	4,78%	13	4,78%	11	4,04%
2,4	11	4,04%	9	3,31%	13	4,78%
2,5	12	4,41%	15	5,51%	17	6,25%
2,6	16	5,88%	12	4,41%	26	9,56%
2,7	17	6,25%	23	8,46%	18	6,62%
2,8	13	4,78%	15	5,51%	15	5,51%
2,9	7	2,57%	21	7,72%	18	6,62%
3,0	22	8,09%	21	7,72%	16	5,88%
3,1	19	6,99%	16	5,88%	18	6,62%
3,2	10	3,68%	23	8,46%	14	5,15%
3,3	13	4,78%	17	6,25%	14	5,15%
3,4	11	4,04%	18	6,62%	18	6,62%
3,5	9	3,31%	8	2,94%	12	4,41%
3,6	10	3,68%	10	3,68%	6	2,21%
3,7	8	2,94%	7	2,57%	3	1,10%
3,8	8	2,94%	5	1,84%	4	1,47%
3,9	6	2,21%	1	0,37%	3	1,10%
4,0	8	2,94%	1	0,37%	0	0,00%
Summe	**272**	**100,00%**	**272**	**100,00%**	**272**	**100,00%**

höheren Qualitätsrang einnimmt, je höher sein Qualitätsrang bei der Eingangsqualifikation bzw. in der Vordiplomnote war. Auch hier ist zu beachten, dass es bei den Rängen nicht um eine Platzierung der Absolventen geht, die sie jeweils innerhalb ihrer Prüfungskohorte eingenommen haben, sondern dass vielmehr die Ränge untereinander bei recht unterschiedlichen Prüfungsterminen verglichen werden. Auch das ist sicherlich kritikwürdig; allerdings legt dieser Tatbestand die Vermutung nahe, dass der Qualifikationsfortschritt der Studierenden vom Eingang ins Studium bis zum Vordiplom bzw. vom Vordiplom bis zum Diplom in gewissen Grenzen beurteilungsneutral ausfällt, d. h. einmal gute Studierende gut bleiben und

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

umgekehrt. Tatsächlich haben sich die Studierenden bis auf zwei von der Eingangsqualität bis hin zum Vordiplom und vom Vordiplom bis hin zum Diplom nicht um mehr als eine Notenstufe verbessert oder verschlechtert. Von der Eingangsqualifikation bis zur Diplomnote sind lediglich zwölf Studierende (weniger als 5 % der Stichprobe) maximal um zwei Notenstufen gesprungen.

Die Studiendauern der 272 Absolventen der FernUniversität sind interessanterweise fast ähnlich verteilt wie die der Absolventen der Wirtschaftsuniversität Wien (*Hackl / Sedlacek* 2002a, 2002b). Tabelle 2 enthält hierzu wesentliche statistische Kenngrößen. Die grafische Darstellung der Verteilungen veranschaulicht Abbildung 1.

Die Urdaten bestätigen die Hypothese nicht, die Güte der Noten im Vordiplom bzw. im Diplom hinge von der Studiendauer der entsprechenden Studienabschnitte ab.

Auf der Grundlage der dargelegten Befunde wird im Folgenden davon ausgegangen, dass die Entwicklungen zwischen den Eingangsqualifikationen, Vordiplomnoten und Diplomnoten als Qualitäten der 272 Studierenden, die externe Faktoren der Hochschulausbildung darstellen, durch ein statisches Produktionsmodell der mehrstufigen Mehrproduktfertigung beschrieben werden können. Die Modellierung erfolgt im nächsten Abschnitt.

Tabelle 2

Vergleich statistischer Kenngrößen bezüglich der Studiendauern der 272 Absolventen der FernUniversität in Hagen und der Absolventen der Wirtschaftsuniversität Wien (*Hackl / Sedlacek* 2002b)

	FernUniversität in Hagen	Wirtschaftsuniversität Wien
Mittelwert	15,8 Semester	15,4 Semester
0% (Min)	8,0 Semester	6,7 Semester
10%	10,2 Semester	10,0 Semester
25% (Q1)	11,7 Semester	11,6 Semester
50% (Median)	14,5 Semester	14,2 Semester
75% (Q3)	18,1 Semester	17,5 Semester
90%	21,7 Semester	22,4 Semester
100% (Max)	36,0 Semester	39,2 Semester

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

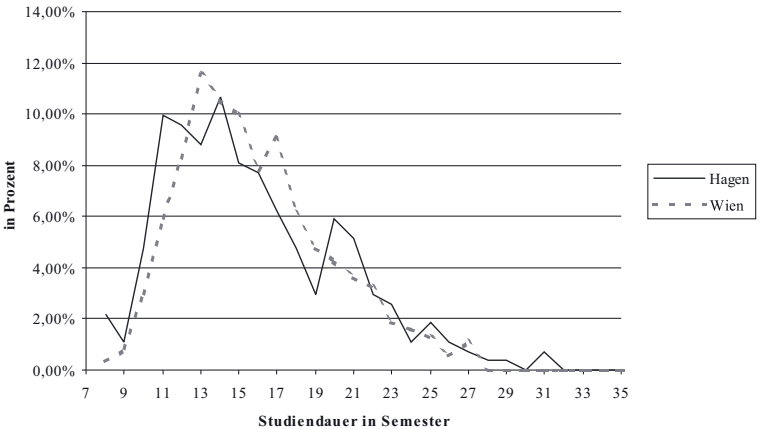


Abbildung 1: Vergleich der Studiendauern der 272 Absolventen der FernUniversität in Hagen und der Absolventen der Wirtschaftsuniversität Wien (*Hackl / Sedlacek* 2002b)

D. Das produktionstheoretische Beschreibungsmodell

Zur produktionstheoretischen Beschreibung, wie aus der Gruppe der 272 Absolventen Studienanfänger mit einer bestimmten Eingangsqualifikation zu Studierenden mit einer bestimmten Vordiplomnote und aus diesen wiederum Absolventen mit einer bestimmten Diplomnote werden, wird auf die mikroökonomisch übliche Vorgehensweise zurückgegriffen, Güter durch ihre Eigenschaften zu definieren. Es bezeichne

x_i	die Anzahl der Studienanfänger, die in ihrer Eingangsqualifikation die Note $i,\ i \in \{1;\ 2;\ 3;\ 4\}$, aufweisen,
y_j	die Anzahl der Studierenden mit der Vordiplomnote $j,\ j \in \{1;\ 2;\ 3;\ 4\}$,
z_k	die Anzahl der Absolventen mit der Diplomnote $k,\ k \in \{1;\ 2;\ 3;\ 4\}$,
r_{ij} bzw. v_{jk}	die Anzahl der Studierenden, die von der Gruppe x_i mit der Eingangsqualifikationsnote i bzw. von der Gruppe y_j mit der Vordiplomnote j zur Gruppe y_j mit der Vordiplomnote j bzw. zur Gruppe z_k mit der Diplomnote k gelangen,
$a_{ij} = r_{ij}/x_i$ bzw. $b_{jk} = v_{jk}/y_j$	die entsprechenden Anteile von x_i bzw. y_j, die zur Gruppe y_j bzw. z_k gelangen,
$R = (r_{ij})$, $V = (v_{jk})$, $A = (a_{ij})$ und $B = (b_{jk})$	die Migrationsmatrizen in Anzahlen und Anteilen der Studierenden,

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

$\bar{P}$ das über den Betrachtungszeitraum unterstellte statische System der Dienstleistungsproduktion „Lehre“ des Fachbereichs Wirtschaftswissenschaft der FernUniversität in Hagen

weiterhin sei

$$x = (x_1, \ldots, x_4)^T,\ y = (y_1, \ldots, y_4)^T,\ z = (z_1, \ldots, z_4)^T\ ,$$

dann lässt sich die zweistufige Mehrproduktfertigung der Erzeugung von Absolventen mit bestimmten Diplomnoten aus Studienanfängern unterschiedlicher Eingangsqualifikation durch die implizite Produktionsfunktion (*Koopmans* 1951)

$$f(x,\ y,\ z,\ \bar{P}) = 0$$

beschreiben mit $\boldsymbol{x}$ als Inputvektor, $\boldsymbol{y}$ als Zwischenproduktvektor und $\boldsymbol{z}$ als Endproduktvektor. Die zugrunde liegende Produktionsstruktur bzw. ihre Verfahren kann man durch Studentenflussgleichungen bzw. Studentenbilanzgleichungen der Form

$$z = B^T y \text{ und } y = A^T x$$

darstellen mit

$$\sum_{i=1}^{4} x_i = \sum_{j=1}^{4} y_j = \sum_{k=1}^{4} z_k = N = 272\ .$$

Da für jedes z_k, $k \in \{1;\ 2;\ 3;\ 4\}$, bzw. für jedes y_j, $j \in \{1;\ 2;\ 3;\ 4\}$ gilt

$$z_k = B_k^T y = \sum_{j=1}^{4} b_{jk} y_j = \sum_{j=1}^{4} v_{jk} \text{ bzw.}$$

$$y_j = A_j^T x = \sum_{i=1}^{4} a_{ij} x_i = \sum_{i=1}^{4} r_{ij}$$

könnten die Produktionsbedingungen prinzipiell durch totale Substitution gekennzeichnet sein (siehe Abbildung 2).

Je geringer jedoch die Rekrutierungen von Absolventen bzw. Studierenden mit Vordiplom einer festen Notenklasse bei größer werdender Notendifferenz zur vorherigen Qualifikation (Note des Vordiploms bzw. Note der Eingangsqualifikation) ausfallen, desto eher darf man vermuten, dass diese Substitutionalität eingeschränkt ist.

Für die explizite Formulierung der Produktionsfunktion müßte man daher in diesem Fall von den Hypothesen ausgehen, dass die Ausbeutekoeffizienten zwischen den Qualitätsstufen weitgehend starr sind und dass sie umso kleiner ausfallen, je stärker sich die Qualitätsstufen aufeinander folgender Qualifikationsstufen unter-

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

scheiden. Dieser Sachverhalt würde sich in den Migrationsmatrizen durch den Umstand aufdrängen, wenn die entsprechenden Migrationsrelationen über die Zeit konstant blieben und die Nebendiagonalen der Migrationsmatrizen nur schwach oder überhaupt nicht besetzt wären. Starre Ausbeutekoeffizienten entsprächen festen Übergangswahrscheinlichkeiten zwischen den Qualitätsstufen, die sich mit Hilfe der Migrationsmatrizen schätzen ließen; sie würden limitationale Produktionsbeziehungen zwischen den Qualitätsstufen bedingen.

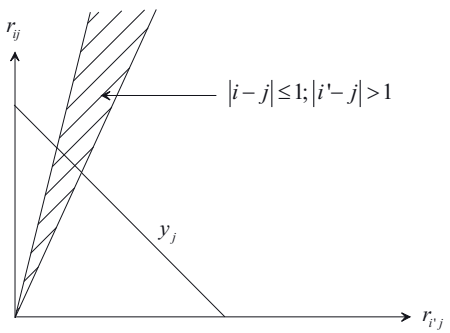


Abbildung 2: Rekrutierung der Studierenden mit der Vordiplomnote j aus Studienanfängern mit den Eingangsqualifikationsnoten i bzw. i'

Auf der anderen Seite bedeuten aber schwach besetzte Nebendiagonalen der Migrationsmatrizen, dass im Ausbildungssystem des Fachbereichs Wirtschaftswissenschaft der FernUniversität die Qualifizierung vom Studienanfang bis zum Vordiplom und vom Vordiplom bis zum Diplom mehr oder weniger notenneutral erfolgt, d. h. die Notenqualität eines Absolventen schon sehr stark von seiner Notenqualität im Vordiplom und diese wiederum von der zu seinem Studienbeginn beeinflusst würde. Das hieße allerdings, dass die Qualitätseigenschaften des externen Faktors „Studierende“ für den letztendlichen Erfolg im Studium maßgeblicher sind als die Qualität des Ausbildungsprozesses. Es wäre also nicht verwunderlich, dass Elitehochschulen mit der Auswahl der besten Studienanfänger nach einer Eingangsprüfung auch die besten Absolventen produzieren.

E. Ergebnisse und Kommentierungen

Die Migrationsmatrizen R von den Gruppen der Eingangsqualifikationsnoten zu den Gruppen der Vordiplomnoten und die Migrationsmatrizen V von den Gruppen der Vordiplomnoten zu den Gruppen der Diplomnoten sind in Tabelle 3 sowie im Anhang zusammengestellt. Sie sind so arrangiert, dass links jeweils die Matrizen R und rechts daneben die entsprechenden Matrizen V stehen. Die Matrizen R^G, R^0,

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

R^1 und R^2 enthalten die Migrationen aller Absolventen (R^G) bzw. nur derjenigen, die ihr Studium im Studienjahr 2000 (Wintersemester 1999 / 2000 und Sommersemester 2000) (R^0) oder 2001 (R^1) oder 2002 (R^2) erfolgreich abgeschlossen haben. Die Matrizen $R^{\bullet m}$ bzw. $R^{\bullet w}$ im Anhang geben davon nur jeweils die Migration männlicher bzw. weiblicher Studierender wieder. Analog gilt dies für die Matrizen V bzw. deren Teilmatrizen. Die Zeilensummen der Matrizen R geben die Mengen x_i, ihre Spaltensummen die Mengen y_j an. Die y_j sind zugleich die Zeilensummen der Matrizen V, deren Spaltensummen die Mengen z_k ergeben. Die zugehörigen Anteilsmatrizen A und B sind hier nicht aufgeführt, da man sie aus den Migrationsmatrizen einfach herleiten kann.

Eine nähere Betrachtung der Migrationsmatrizen zeigt, dass es bis auf zwei Migrationen der Absolventen des Jahres 2000 von der Gruppe mit der Vordiplomnote 4 (y_4^0) zur Gruppe mit der Diplomnote 2 (z_2^0) keine Rekrutierungen in den Gruppen der Vordiplomnoten bzw. der Diplomnoten gibt, die aus Gruppen der Noten der Eingangsqualifikation bzw. der Vordiplomnoten kommen würden, die mehr als eine Notenstufe schlechter oder besser liegen (man beachte, dass die beiden Ausnahmen – je einmal weiblich und männlich – nur in den Matrizen V^0 bzw. $V^{0\bullet}$ auftreten und in den Summationen der Matrizen V^G bzw. $V^{G\bullet}$ bereits enthalten sind). Bis auf die beiden genannten Ausnahmen sind alle Matrixelemente außerhalb der Hauptdiagonalen und der unmittelbaren Nebendiagonalen gleich Null. Dies bedeutet, dass die Substitutionalitäten zwischen den Qualitätsstufen tatsächlich erheblich eingeschränkt sind, d. h. gute Absolventen sich in der Regel aus guten Studierenden rekrutieren.

Noch deutlicher wird dieses Bild, wenn man die Notendriften untersucht, die sich aus den Migrationsmatrizen ergeben. Dabei sei die Notendrift der Ausprägung d durch die Anzahl n_d der Migrationen der Studierenden definiert, die gegenüber ihrer Note in der vorherigen Qualifikationsstufe in der unmittelbar folgenden Ausbildungsstufe eine Note erreichen, die um d schlechter ist; entsprechend bedeutet eine negative Notendriftausprägung eine Notenverbesserung. Es gilt also:

$$n_d = \sum_{i=1}^{4} r_{i,i+d} \text{ bzw. } n_d = \sum_{j=1}^{4} v_{j,j+d} \ .$$

Für $d = -1$ erhält man aus der Matrix V^{0m} beispielsweise

$$n_{-1} = v_{10}^{0m} + v_{21}^{0m} + v_{32}^{0m} + v_{43}^{0m} = 0 + 1 + 6 + 3 = 10 \ .$$

Für $d = 0$ ergibt sich jeweils die Summe der Matrixelemente in der Hauptdiagonalen (keine Notenveränderung zwischen jetziger und vorheriger Qualifikationsstufe). Die Notendriften von bzw. zu den Randnoten 1 und 4 fallen naturgemäß (prozentual) kleiner aus, da keine Veränderungen zu bzw. von den Noten 0 und 5 stattfinden können. Eine Notendriftausprägung von Null bedeutet allerdings kei-

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

neswegs, dass keine weitere Qualifikation erfolgt ist – vielmehr besagt sie nur, dass das Vordiplom bzw. Diplom mit derselben Note erreicht worden ist wie in der Eingangsqualifikation bzw. im Vordiplom. Tabelle 4 zeigt die Notendriften für die Matrizen der Tabelle 3 an. Dort, wo die Notendrift von Null verschieden ist, gibt die obere Zahl die Anzahl und die untere Zahl die Prozentzahl – gemessen an der jeweiligen Spaltensumme – an.

Aus Tabelle 4 ist ersichtlich, dass erhebliche Notenverbesserungen oder -verschlechterungen von der Eingangsqualifikation zum Vordiplom und vom Vordiplom zum Diplom (fast) nicht vorkommen. In großer Mehrheit werden auf den verschiedenen Qualifikationsstufen von den Studierenden dieselben Noten erzielt, wobei die weiblichen Studierenden im Grundstudium etwas leistungskonstanter ($d = 0$) als die männlichen Studierenden sind (69 % in R^{Gw} gegen 55 % in R^{Gm}). Die Notenveränderungen verteilen sich nahezu gleich auf die Notenverbesserung ($d = -1$) und Notenverschlechterung ($d = +1$) um jeweils eine Note. Auffällig sind lediglich die Notenverschlechterungen von 29 % der männlichen Studierenden im Vordiplom, die im Studienjahr 2001 ihr Diplom abgelegt haben, sowie die Notenverbesserungen von 33% der weiblichen Absolventen im Diplom des Studienjahres 2002 (hervorgehobene Werte in Tabelle 4).

Die Daten über die 272 Absolventen der FernUniversität liefern keine Anhaltspunkte dafür, dass Studierende mit längerer Studiendauer ihre Diplomnote gegenüber der Vordiplomnote bzw. ihre Vordiplomnote gegenüber der Eingangsqualität haben verbessern können (Zeit gegen Noten). Ebenso wenig ist feststellbar, dass kürzere Studiendauern nur durch die Inkaufnahme schlechterer Noten verwirklicht werden können (Note gegen Zeit). Bemerkenswerte Erscheinungen offenbaren die Daten in diesem Kontext:

1. Von den 137 Studierenden, die mit der Note 3 als Eingangsqualität ins Ausbildungssystem gingen, wiesen die mit den besseren Qualitäten in der Diplomnote die kürzeren Studiendauern auf. Für diese Studierenden war es also besser, schneller zu studieren. Diejenigen, die länger studierten, schnitten schlechter ab.
2. Von den 170 Studierenden, die mit der Note 3 ihre Diplomprüfung bestanden, weisen diejenigen mit der kürzeren Studiendauer eine bessere Eingangsqualität auf. Diejenigen, bei denen eine schwächere Eingangsqualität vorlag, brauchten länger, um diese Diplomnote zu erreichen.

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

Tabelle 3: **Migrationsmatrizen (Fortsetzung im Anhang)**

R^G

Eingangs-qualifikation	Vordiplomnote 1	2	3	4	Summe	
1	3	8			**11**	x_1
2	2	42	40		**84**	x_2
3		19	107	11	**137**	x_3
4			27	13	**40**	x_4
Summe	**5**	**69**	**174**	**24**	**272**	
	y_1	y_2	y_3	y_4		

V^G

Vordiplomnote	Diplomnote 1	2	3	4	Summe	
1	1	4			**5**	y_1
2	3	43	23		**69**	y_2
3		33	129	12	**174**	y_3
4		2	18	4	**24**	y_4
Summe	**4**	**82**	**170**	**16**	**272**	
	z_1	z_2	z_3	z_4		

R^0

Eingangs-qualifikation	Vordiplomnote 1	2	3	4	Summe
1	1	4			**5**
2	1	16	14		**31**
3		6	32	2	**40**
4			11	7	**18**
Summe	**2**	**26**	**57**	**9**	**94**

V^0

Vordiplomnote	Diplomnote 1	2	3	4	Summe
1		2			**2**
2	1	17	8		**26**
3		9	44	4	**57**
4		2	7		**9**
Summe	**1**	**30**	**59**	**4**	**94**

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

V^1

Vordiplomnote	Diplomnote 1	2	3	4	Summe
1	1				**1**
2	2	12	12		**26**
3		8	43	4	**55**
4			3	3	**6**
Summe	**3**	**20**	**58**	**7**	**88**

R^1

Eingangs-qualifikation	Vordiplomnote 1	2	3	4	Summe
1	1	4			**5**
2		14	11		**25**
3		8	37	6	**51**
4			7		**7**
Summe	**1**	**26**	**55**	**6**	**88**

V^2

Vordiplomnote	Diplomnote 1	2	3	4	Summe
1		2			**2**
2		14	3		**17**
3		16	42	4	**62**
4			8	1	**9**
Summe	**0**	**32**	**53**	**5**	**90**

R^2

Eingangs-qualifikation	Vordiplomnote 1	2	3	4	Summe
1	1				**1**
2	1	12	15		**28**
3		5	38	3	**46**
4			9	6	**15**
Summe	**2**	**17**	**62**	**9**	**90**

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

Tabelle 4

Notendriften n_d zu den Matrizen in Tabelle 3

Matrix / Driftausprägung d	insgesamt			2000			2001			2002		
	R^G	R^{Gm}	R^{Gw}	R^0	R^{0m}	R^{0w}	R^1	R^{1m}	R^{1w}	R^2	R^{2m}	R^{2w}
+3	0	0	0	0	0	0	0	0	0	0	0	0
+2	0	0	0	0	0	0	0	0	0	0	0	0
+1	59/22	40/25	19/17	20/21	13/23	7/19	21/24	$\boxed{17/29}$	4/13	18/20	10/22	8/18
0	165/61	88/55	77/69	56/60	31/54	25/67	52/59	31/54	21/70	57/63	26/58	31/69
-1	48/17	32/20	16/14	18/19	13/23	5/14	15/17	10/17	5/17	15/17	9/20	6/13
-2	0	0	0	0	0	0	0	0	0	0	0	0
-3	0	0	0	0	0	0	0	0	0	0	0	0

Matrix / Driftausprägung d	insgesamt			2000			2001			2002		
	V^G	V^{Gm}	V^{Gw}	V^0	V^{0m}	V^{0w}	V^1	V^{1m}	V^{1w}	V^2	V^{2m}	V^{2w}
+3	0	0	0	0	0	0	0	0	0	0	0	0
+2	0	0	0	0	0	0	0	0	0	0	0	0
+1	39/14	26/16	13/12	14/15	8/14	6/16	16/18	10/17	6/20	9/10	8/18	1/2
0	177/65	106/66	71/63	61/65	38/66	23/62	59/67	40/69	19/63	57/63	28/62	29/65
-1	54/20	27/17	27/24	17/18	10/18	7/19	13/15	8/14	5/17	24/27	9/20	$\boxed{15/33}$
-2	2/1	1/1	1/1	2/2	1/2	1/3	0	0	0	0	0	0
-3	0	0	0	0	0	0	0	0	0	0	0	0

F. Grenzen und Erweiterungen

Das vorgetragene Beschreibungsmodell zur Erfassung von Qualitäten der externen Faktoren „Studierende“ in der Dienstleistungsproduktion „Lehre“ von Hochschulen hat aus der Sicht der betriebswirtschaftlichen Theorie einige Einschränkungen, die unbefriedigend sind. Sie sind zugleich aber die Ansatzpunkte für Weiterentwicklungen.

Die im Modell untersuchten Migrationen können Eigenheiten über die Entwicklungen offenbaren, wie aus Studierenden mit bestimmten Eingangsqualitäten über das Vordiplom in einem zweistufigen Ausbildungsprozess Absolventen mit bestimmten Diplomnoten werden. Die entsprechenden Anteilsmatrizen liefern dabei Informationen über die Ausbeutekoeffizienten, die für die Bestimmung von Studentenflussgleichungen dienlich sind. Durch diese für die Input-Output-Analyse (*Koopmans* 1951) übliche Vorgehensweise kann der Beitrag des externen Faktors „Studierender“ zum Erfolg der Dienstleistungsproduktion der Hochschule explizit modelliert werden. Zeitreihenmodelle mit größerem Stichprobenumfang mögen Auskunft darüber geben, ob die Ausbeutekoeffizienten bei sonst gleichen

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

Produktionsbedingungen des Fachbereichs über größere Zeiträume hinweg konstant bleiben oder sich verändern, d. h. die Qualität der Studierenden im Zeitablauf variiert.

Ein Nachteil des Beschreibungsmodells liegt darin, Qualitäten der Absolventen nur durch Eingangsqualitäten bzw. Vordiplomnoten der Studierenden zu beschreiben, ohne die Studiendauer als Erklärungsvariable explizit einzubeziehen. Das vorliegende Datenmaterial liefert dafür noch keinen Anhaltspunkt. Größere Stichprobenumfänge würden es aber nahe legen, die Studiendauer bei gegebener Eingangsqualifikation ebenso für die Erklärung der Absolventennoten heranzuziehen und dabei auch nach dem Geschlecht der Studierenden zu differenzieren. Nimmt man dazu die Urdaten in Zehntelnoten als Basis, so ließe sich die Verteilung der Notendrift noch feiner in Abhängigkeit der Studiendauer bestimmen. Das würde zu einer so genannten Diffusionsgleichung führen, die aus dem hier vorgetragenen einfachen Beschreibungsmodell ein stochastisches Erklärungsmodell der Qualitätsveränderungen erzeugen würde.

Geht man von den Befunden aus, dass höhere Qualifikationen im Studium (Anfänger, Vordiplom, Diplom) zu höheren Einstiegsgehältern in der Wirtschaft führen (*Bartels* 1993) und bessere Abschlussnoten mit höheren Gehältern einhergehen[2], dann kann das aus dem Beschreibungsmodell hergeleitete Erklärungsmodell als Grundlage eines Entscheidungsmodells fungieren. Der Fachbereich könnte auf Studierende mit schlechteren Eingangsqualitäten mit intensiveren Betreuungsaktivitäten in der Lehre reagieren, um die Ausbeutekoeffizienten schlechterer Notenklassen für bessere Notenklassen zu erhöhen und so zu einer größeren Wertschöpfung in der Bildungsökonomie beizutragen. Das akquisitorische Potenzial solcher Hochschulen wäre entsprechend hoch. Sofern darunter die Qualität der Lehre keine Einbußen erlitte, entstünde ein großer Reputationsgewinn.

Literatur

Albach, H. / *Fandel,* G. / *Schüler,* W. (1978): Hochschulplanung, Baden-Baden 1978.

Bartels, J. (1993): Absolventen des Fachbereichs Wirtschaftswissenschaft – Ergebnisse einer Repräsentativbefragung, Zentrum für Fernstudienentwicklung – FernUniversität – Gesamthochschule – in Hagen, 1993.

BfA (2000): Analysen zum Arbeitsmarkt / Strukturen der Arbeitslosigkeit: Akademiker / innen, Bundesanstalt für Arbeit, Landesarbeitsamt Nordrhein-Westfalen, Düsseldorf 2000.

Fandel, G. (1995): Wirtschaftlichkeitsaspekte einer Hochschule. Inhalte einer gutachterlichen Stellungnahme über die FernUniversität, Weinheim 1995.

[2] Große Anwaltskanzleien bestimmen die anfänglichen Jahresgehälter ihrer Juristen im Schnitt so, dass jeder Punkt des Zweiten Juristischen Staatsexamens mit 500 € bewertet wird.

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

Fandel, G. (2003): Zur Leistung nordrhein-westfälischer Universitäten – Gegenüberstellung einer Verteilungslösung und der Effizienzmaße einer Data Envelopment Analysis, in: Backes-Gellner, U. / Schmidtke, C. (Hrsg.): Hochschulkökonomie – Analysen interner Steuerungsprobleme und gesamtwirtschaftlicher Effekte, Schriften des Vereins für Sozialpolitik, Neue Folge, Band 296, Berlin 2003, S. 33 – 50..

Fandel, G. / *Gal,* T. (2001): Redistribution of funds for teaching and research among universities: The case of North Rhine-Westphalia, in: European Journal of Operational Research, Volume 130 (2001), S. 111 – 120.

Fandel, G. / *Paff,* A. (2000): Eine produktionstheoretisch fundierte Kostenrechnung für Hochschulen – dargestellt am Beispiel der FernUniversität Hagen, in: Zeitschrift für Betriebswirtschaft, Ergänzungsheft 3 / 2000, S. 191 – 204.

Fries, M. (2002): Abitur und Studienerfolg. Welchen „Wert" hat das Abitur für ein erfolgreiches Studium?, in: Beiträge zur Hochschulforschung, Heft 1, 24. Jahrgang, 2002, S. 30 – 51.

Hackl, P. / *Sedlacek,* G. (2002a): Analyse der Studiendauer am Beispiel der Wirtschaftsuniversität Wien, in: Dutter, R. (Hrsg.): Festschrift 50 Jahre Österreichische Statistische Gesellschaft, Wien 2002, S. 41 – 59.

Hackl, P. / *Sedlacek,* G. (2002b): Forschungsbericht Studienverlaufsanalyse, Forschungsbericht der Wirtschaftsuniversität Wien, 2002.

Koopmans, T. C. (1951): Analysis of production as an efficient combination of activities, in: Koopmans, T. C. (Hrsg.): Activity Analysis of Production and Allocation, New York 1951, S. 33 – 97.

Lohmar, U. / *Ortner,* G. E. / *Bayer,* M. (Hrsg.) (1982): Die deutsche Hochschule zwischen Numerus clausus und Akademikerarbeitslosigkeit. Der doppelte Flaschenhals, Hannover 1982.

Maleri, R. (1973): Grundzüge der Dienstleistungsproduktion, Berlin et al. 1973.

Smith, J. / *Naylor,* R. A. (2001): Dropping out of university: A statistical analysis of the probability of withdrawal for UK university students, in: Journal of the Royal Statistical Society, C, 164; S. 389 – 405.

Stieger, H. (1980): Zur Ökonomie der Hochschule, Gießen 1980.

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

Anhang: Fortsetzung Tabelle 3 – geschlechtsspezifische Migrationsmatrizen

R^{Gm}

	Vordiplomnote				
Eingangs-qualifikation	**1**	**2**	**3**	**4**	**Summe**
1	2	6			**8**
2	2	27	28		**57**
3		15	54	6	**75**
4			15	5	**20**
Summe	**4**	**48**	**97**	**11**	**160**

V^{Gm}

	Diplomnote				
Vordiplomnote	**1**	**2**	**3**	**4**	**Summe**
1	1	3			**4**
2	3	30	15		**48**
3		18	71	8	**97**
4		1	6	4	**11**
Summe	**4**	**52**	**92**	**12**	**160**

R^{Gw}

	Vordiplomnote				
Eingangs-qualifikation	**1**	**2**	**3**	**4**	**Summe**
1	1	2			**3**
2		15	12		**27**
3		4	53	5	**62**
4			12	8	**20**
Summe	**1**	**21**	**77**	**13**	**112**

V^{Gw}

	Diplomnote				
Vordiplomnote	**1**	**2**	**3**	**4**	**Summe**
1		1			**1**
2		13	8		**21**
3		15	58	4	**77**
4		1	12		**13**
Summe	**0**	**30**	**78**	**4**	**112**

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

Fortsetzung Anhang (Tabelle 3):

R^{0m}

	Vordiplomnote				
Eingangs-qualifikation	**1**	**2**	**3**	**4**	**Summe**
1		2			**2**
2	1	10	10		**21**
3		6	18	1	**25**
4			6	3	**9**
Summe	**1**	**18**	**34**	**4**	**57**

V^{0m}

	Diplomnote				
Vordiplomnote	**1**	**2**	**3**	**4**	**Summe**
1		1			**1**
2	1	13	4		**18**
3		6	25	3	**34**
4		1	3		**4**
Summe	**1**	**21**	**32**	**3**	**57**

R^{0w}

	Vordiplomnote				
Eingangs-qualifikation	**1**	**2**	**3**	**4**	**Summe**
1	1	2			**3**
2		6	4		**10**
3			14	1	**15**
4			5	4	**9**
Summe	**1**	**8**	**23**	**5**	**37**

V^{0w}

	Diplomnote				
Vordiplomnote	**1**	**2**	**3**	**4**	**Summe**
1		1			**1**
2		4	4		**8**
3		3	19	1	**23**
4		1	4		**5**
Summe	**0**	**9**	**27**	**1**	**37**

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

R^{1m}

	Vordiplomnote				
Eingangs-qualifikation	**1**	**2**	**3**	**4**	**Summe**
1	1	4			**5**
2		12	9		**21**
3		5	18	4	**27**
4			5		**5**
Summe	**1**	**21**	**32**	**4**	**58**

V^{1m}

	Diplomnote				
Vordiplomnote	**1**	**2**	**3**	**4**	**Summe**
1	1				**1**
2	2	11	8		**21**
3		5	25	2	**32**
4			1	3	**4**
Summe	**3**	**16**	**34**	**5**	**58**

R^{1w}

	Vordiplomnote				
Eingangs-qualifikation	**1**	**2**	**3**	**4**	**Summe**
1					**0**
2		2	2		**4**
3		3	19	2	**24**
4			2		**2**
Summe	**0**	**5**	**23**	**2**	**30**

V^{1w}

	Diplomnote				
Vordiplomnote	**1**	**2**	**3**	**4**	**Summe**
1					**0**
2		1	4		**5**
3		3	18	2	**23**
4			2		**2**
Summe	**0**	**4**	**24**	**2**	**30**

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

Fortsetzung Anhang (Tabelle 3):

R^{2m}

	Vordiplomnote				
Eingangs-qualifikation	**1**	**2**	**3**	**4**	**Summe**
1	1				**1**
2	1	5	9		**15**
3		4	18	1	**23**
4			4	2	**6**
Summe	**2**	**9**	**31**	**3**	**45**

V^{2m}

	Diplomnote				
Vordiplomnote	**1**	**2**	**3**	**4**	**Summe**
1		2			**2**
2		6	3		**9**
3		7	21	3	**31**
4			2	1	**3**
Summe	**0**	**15**	**26**	**4**	**45**

R^{2w}

	Vordiplomnote				
Eingangs-qualifikation	**1**	**2**	**3**	**4**	**Summe**
1					**0**
2		7	6		**13**
3		1	20	2	**23**
4			5	4	**9**
Summe	**0**	**8**	**31**	**6**	**45**

V^{2w}

	Diplomnote				
Vordiplomnote	**1**	**2**	**3**	**4**	**Summe**
1					**0**
2		8			**8**
3		9	21	1	**31**
4			6		**6**
Summe	**0**	**17**	**27**	**1**	**45**

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

Verhindert Politikberatung gute Lehre und Forschung?

Eine explorative Evaluierung volkswirtschaftlicher Politikberatung mit Hilfe eines Expertensurveys

Von *Gert G. Wagner,* Berlin

A. Zur Ausgangslage*

Die Öffentlichkeit erwartet von Hochschullehrern nicht nur gute Lehre und Forschung, sondern auch – so formulierte Bundeskanzler Helmut Schmidt vor über 20 Jahren – als „Bringschuld" die Information der Öffentlichkeit und der Politik über wissenschaftliche Ergebnisse.[1] Insbesondere von Volkswirten wird „Politikberatung" erwartet, da sie sich von der Natur ihres Forschungsgegenstandes her mit „öffentlichen Belangen" beschäftigen. Umso bedenklicher wiegt eine offenbar zunehmende Unzufriedenheit der breiten Öffentlichkeit mit der „ökonomischen Zunft".[2] Durchaus typisch ist die folgende Einschätzung, die dem Wirtschaftsteil einer großen deutschen Tageszeitung, der Frankfurter Allgemeinen Zeitung (FAZ), vom 13. 12. 1999 entnommen ist:

„Die meisten Ökonomen sind nicht nur langweilig, sie enttäuschen auch all jene, die sich Rat suchend an sie wenden. Wer sich von ihnen Antworten auf wichtige wirtschaftliche Fragen erhofft, erntet zumeist nur Gleichungen, Diagramme und unverständlichen Fachjargon. Das Bedürfnis nach klaren Antworten oder Ratschlägen bleibt ungestillt".

Aus Sicht der modernen Volkswirtschaftslehre machen jedoch „Gleichungen" und „Diagramme" die wissenschaftliche Qualität guter Volkswirtschaftslehre aus.

---

* Das empirische Material dieses Beitrags entstand im Rahmen des vom Bundesministerium für Wirtschaft in Auftrag gegebenen Gutachtens „Neuere Entwicklungen in der Wirtschaftswissenschaft" (vgl. für den ersten Teil des Abschlußberichtes *Zimmermann* 2002). Der Autor dankt seinen Kollegen in diesem Projekt. Für kritische und konstruktive Kommentare zu der auf der Züricher Arbeitstagung des Bildungsökonomischen Ausschusses vorgetragenen Fassung des vorliegenden Papiers dankt der Autor den Mitgliedern des Ausschusses, zwei anonymen Gutachtern für ihre Geduld und insbesondere der seinerzeitigen Vorsitzenden des Ausschusses und Herausgeberin dieses Bandes, Uschi Backes-Gellner.

1 Vgl. zuletzt *Schmidt* (1998, 102 ff.).

2 Vgl. auch *Wagner* und *Wiegard* (2001).

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

Deswegen stellt sich die Frage, ob nur diejenigen Volkswirte einflussreiche Politikberater sein können, die vereinfachen, um gehört zu werden, und deswegen schlechte Forschung (und damit auch Lehre) betreiben?

Diese Fragestellung knüpft an ältere Arbeiten von Backes-Gellner an[3], die untersuchte, ob gute und viele Publikationen die Qualität der Lehre durch publikationsstarke Hochschullehrer mindert. Der Frage nach dem Zusammenhang zwischen Lehre und Forschung einerseits und Politikberatung andererseits wird in dem vorliegenden Aufsatz mit Hilfe einer Befragung bei einschlägigen Volkswirten nachgegangen. Die Ergebnisse sind aufgrund der Subjektivität der Angaben nicht gut gesichert; aber immerhin werfen sie Licht in einen Bereich, über den bislang weitestgehend nur spekuliert und philosophiert wird.

Kenner der Universitäten wird das Ergebnis der hier vorgelegten empirischen Untersuchung nicht überraschen: Es zeichnet sich, für den im folgenden untersuchten Kreis von Hochschullehrern ab, dass gute Qualität in Forschung und Lehre guter Politikberatung nicht im Wege stehen.

B. Empirische Analyse

I. Stand der Forschung

Über den Zusammenhang zwischen volkswirtschaftlicher Lehre und Forschung einerseits und Politikberatung andererseits gibt es kaum empirische Literatur. Vor dem Hintergrund der Frage nach der besten Form der Organisation der Politikberatung ist in Deutschland z. B. das Verhältnis von Universitäten und außeruniversitären Einrichtungen untersucht worden (vgl. *Wilts* 1999, 2000). Der Untersuchungsgegenstand dieser Studie sind aber Institutionen und keine Personen.

Die relativ breite Literatur über Veröffentlichungspraxis und Zitationen arbeitet im Grunde rein deskriptiv. Die einzige erklärenden Variablen sind i. d. R. das Herkunftsland und die Alters-Kohortenzugehörigkeit von Autoren.[4] Denkbar wäre eine Verknüpfung der (frei zugänglichen) Daten über die Lebensläufe von Wissenschaftlern und deren Publikationen bzw. erhaltenen Zitationen. Derartige Analysen, deren Datengewinnung sehr aufwendig wäre, stehen bislang aus. Im Hinblick auf die Messung von „Politikberatung“ kommt das Problem hinzu, dass diese Tätigkeit mit bibliometrischen Methoden nicht gut messbar ist, denn sie geht über das Publizieren (z. B. in Tageszeitungen etc.) hinaus.

Es gibt inzwischen Messinstrumente für die Erfassung von Publikationen und deren Zitationen im akademischen Bereich, jedoch keine (standardisierten) Messinstrumente für die Erfassung des Outputs im Bereich der Politikberatung (und der

[3] Vgl. u. a. *Backes-Gellner* (1989).

[4] Vgl. z. B. *Eichenberger et al*. (2000).

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

Lehre), die weltweit zugänglich und auswertbar wären. Deswegen liegt der Versuch nahe, die Aktivitäten und Erfolge von Volkswirten im Bereich der Politikberatung durch eine Befragung von einschlägigen Volkswirten zu erheben und zu analysieren.

Zwar nur vereinzelt, aber immerhin in letzter Zeit zunehmend, findet man Befragungen von Wissenschaftlern zu ihrem Verhalten sowie zu ihren Präferenzen und Werturteilen. Derartige Surveys beziehen sich für den Bereich von Wirtschaftswissenschaftlern / Volkswirten bislang vorwiegend auf die Einschätzung der Qualität von Zeitschriften und von Basis-Werturteilen (vgl. dazu *Frey* und *Eichenberger* 1993)[5], weniger auf die Erhebung der tatsächlichen Verhaltensweisen in Forschung, Lehre und Politikberatung.[6]

Eine Ausnahme stellt die Expertenbefragung von *Wyplosz* (1999) dar, der 32 persönliche Freunde aus 16 Ländern in Nordamerika und Europa ansprach, von denen 27 an der Befragung teilnahmen. Wyplosz hat im Wesentlichen einige Angaben zur Struktur der akademischen Volkswirtschaftslehre und der institutionellen Politikberatung erhoben. Er hat jedoch keine Individualinformation über die persönliche Performance der befragten Experten erfragt.

II. Konzept und Durchführung einer Expertenbefragung bei den Fellows des CEPR

Die im Folgenden dargestellte Befragung basiert auf der Idee, dass (hochrangige) Volkswirte nicht nur in der Lage sind, ihre Arbeitsbedingungen und ihren akademischen Erfolg (der relativ leicht messbar ist) in einer Befragung anzugeben, sondern auch ihr Interesse und ihren Erfolg in der Politikberatung. Man kann derartige subjektive Angaben natürlich mit guten Gründen anzweifeln, aber gemessen an der bislang auf Introspektion und Spekulation beruhenden Analysen der hier verfolgten Fragestellung – meist von Kollegen mittleren Alters in den Festschriften für ihre Doktorväter und Mentoren angestellt – ist der gewählte Ansatz zumindest als explorative Analyse vertretbar.

Die Grundgesamtheit der Befragung stellen die Fellows des CEPR (Centre für Economic Policy Analysis, London) dar, die ein Netzwerk hochrangiger Volkswirte bilden. Diese sind eine stark selektive Auswahl der akademisch arbeitenden Volkswirte in Europa und Nordamerika. CEPR-Fellows werden in erster Linie

---

[5] Vgl. auch *Rotte* und *Zimmermann* (1998) zu einer Befragung von Volkswirten zu deren Einschätzung der Europäischen Währungsunion (EMU). Genannt sei auch eine Umfrage der Hochschule der Bundeswehr (Hamburg) zur Bewertung von Fachzeitschriften durch die Mitglieder des Vereins für Socialpolitik (vgl. *Bräuninger* und *Haucap* 2001).

[6] Im Hinblick auf die Nachfrage und die Verarbeitung volkswirtschaftlichen Wissens durch Entscheidungsträger, ihre Helfer und die breite Öffentlichkeit herrscht ebenso ein Mangel an empirischer Evidenz; jedoch gibt es zumindest einige Vorbildstudien in der Literatur (vgl. insbesondere die Arbeit von *Gellner* 1995).

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

nach Maßgabe ihrer Publikationen in hochangesehenen internationalen Fachzeitschriften für Economics ausgewählt. Ein substantielles Interesse für Politikanalysen ist faktisch eher Nebensache. Insofern stellen die CEPR-Fellows eine Auswahl angesehener akademischer Volkswirte dar, die sowohl auf den Gebieten der Theorie als auch der Ökonometrie und der wirtschaftspolitischen Analyse arbeiten.

Da es sich bei den Experten um eine Auswahl international reputierlicher Volkswirte handelt, können die gefundenen Ergebnisse nicht ohne weiteres verallgemeinert werden. Aus den Strukturen der Antworten in der Stichprobe sind jedoch durchaus vorsichtige Kausalaussagen ableitbar.

Bei hochkarätigen akademischen Volkswirten ist es inzwischen üblich, dass sie das Internet aktiv benutzen. Die Expertenbefragung wurde deswegen als Internet-Survey durchgeführt. D. h., dass der Fragebogen via Internet mit jedem gängigen Browser ausgefüllt werden konnte. Zuvor wurden die Befragten CEPR-Fellows mit einer E-Mail, die individuell adressiert war, über die Befragung, ihren Zweck und die technischen Einzelheiten informiert. Diese E-Mail ging an 310 CEPR-Fellows; 18 Adressen erwiesen sich als nicht „zustellbar". Die bereinigte Brutto-Stichprobe enthält also 292 Fälle.[7]

Nach drei Wochen hatten schließlich 47 Fellows den Fragebogen online ausgefüllt, womit sich eine Rücklaufquote von 16 % ergab (wobei die auswertbare Fallzahl für die meisten Fragen bei etwa 35 Fällen liegt, was einer „vollständigen Antwortrate" von etwa 11 % entspricht). Ein Rücklauf von knapp 20 % ist wenig, jedoch entspricht er dem erwartbaren Ergebnis einer nicht durch Interviewer gestützten Erhebung.

III. Ergebnisse der Expertenbefragung

Im Folgenden beziehen sich Prozentangaben auf die jeweils auswertbaren Fälle, die ja nach Frage (aufgrund fehlender Antworten) variieren. Da es sich durchweg um sehr kleine Fallzahlen handelt, wird die Fallzahl stets angegeben (wenn Ergebnisse auch in Tabellenform dargestellt sind, werden die Fallzahlen meist nur dort ausgewiesen). Differenzen sind in der Regel nicht signifikant, d. h. die Expertenbefragung kann nur Tendenzen aufzeigen, jedoch selbst für den Expertenkreis keine statistisch gesicherten Ergebnisse liefern.

Die Stichprobe ist weit überwiegend männlich (5 von 45 Angaben entsprechen 85 %) und mit 60 % von der Altersklasse der 40 bis unter 50 Jährigen dominiert. 20 % sind jünger als 40 Jahre und 15 % sind 50 Jahre und älter. Die Experten sind

[7] Nach der ersten Aussendung wurde an einige Fellows noch eine persönliche Notiz geschickt; auf eine zweite Notiz wurde verzichtet, da – wie gesagt – den Auswertenden der Daten nicht bekannt ist, wer an der Befragung teilgenommen hat bzw. nicht teilgenommen hat.

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

in 14 verschiedenen Ländern geboren (10 in deutschsprachigen Ländern und 12 in Großbritannien oder den USA). Die beiden letztgenannten Länder werden im Folgenden auch als „Angelsachsen“ bezeichnet. Die Experten arbeiten in 12 Ländern, wobei auffällt, dass 12 in deutschsprachigen Ländern und 18 in Großbritannien oder den USA arbeiten, d. h. jeweils mehr als dort geboren sind.

35 (von 38) der Experten sind „Full Professors“ und 11 (von 35) geben an, dass sie außerhalb der Universität noch eine andere „permanent position“ haben. Ein derartiger Anteil von etwa 30 % zeigt, dass es sich bei den Experten um beruflich ungewöhnlich erfolgreiche Hochschullehrer handeln dürfte (wenngleich repräsentative Vergleichszahlen fehlen).

1. Relevanz verschiedener Formen der volkswirtschaftlichen Politikberatung

Von den Experten geben 85% an, dass sie Politikberatung betreiben wollen (vgl. auch Tabelle 1). Dabei gibt aber nur ein gutes Drittel der Experten an, dass sie Politiker unmittelbar beeinflussen wollen. Dies ist legitim, da es viele unterschiedliche Wege der Politikberatung gibt. Man kann die Öffentlichkeit – und damit auch Entscheidungsträger – durch Publikationen und Vorträge unterrichten und beeinflussen; Politikwissenschaftler unterscheiden dabei zwischen der Information zu bekannten Themen und der Themensetzung, d. h. des Setzens neuartiger Fragestellungen auf die politische Agenda (*Agenda Setting*) (vgl. z. B. *Gellner* 1995). Wissenschaftler können jedoch auch – und dies stellen sich die meisten unter „Politikberatung“ vor – Entscheidungsträger direkt beraten. Direkte Beratung kann man wiederum in öffentliche und interne Beratung unterscheiden. Eine öffentliche Form der Beratung erfolgt z. B. durch Gutachten, wie sie etwa der Sachverständigenrat für die Beurteilung der gesamtwirtschaftlichen Entwicklung veröffentlicht. Interne Beratung findet wiederum formalisiert in Beiräten, aber auch informell durch persönliche Gespräche statt.

Tabelle 2 zeigt, dass der kurzfristige Erfolg volkswirtschaftlicher Politikberatung insgesamt sehr skeptisch eingeschätzt wird. Kurzfristiger Einfluss (vgl. den oberen Block der Tabelle) wird am ehesten Regierungsbeiräten, seien sie politisch unabhängig oder abhängig, zugerechnet. Aber immerhin 44% der Experten geben an, dass ein unabhängiger Regierungsbeirat keinen Einfluss hat. Und selbst von politiknahen Beiräten (etwa das US-amerikanische Council of Economic Advisers), die der Regierung internen Rat geben, glauben ein Viertel, dass sie ohne Einfluss seien. Als einflussreich werden Volkswirte in Zentralbankbeiräten eingeschätzt. Nicht überraschend ist, dass Publikationen in Fachzeitschriften praktisch kein kurzfristiger Einfluss auf die Wirtschaftspolitik zugesprochen wird. Auch Publikationen in populären Medien und Fernsehauftritte werden von fast der Hälfte der Experten – im Hinblick auf kurzfristige Politikberatung – als nutzlos angesehen.

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

Tabelle 1

Präferenzen der CEPR-Fellows bezüglich Politikberatung (in % von *N*)

Wunsch nach Politikberatung	Zustimmung[1)] Ablehnung[2)] Weiß nicht	85 15 –
	N	34
Wunsch nach direkter Beeinflussung von Politikern	Zustimmung Ablehnung Weiß nicht	36 40 24
	N	33

[1)] Vorgaben „I strongly agree“ und „I agree“.
[2)] Vorgaben „I strongly disagree“ und „I disagree“.

Quelle: CEPR Expertensurvey; eigene Berechnungen.

Tabelle 2

Einschätzung der CEPR-Experten des kurz- und langfristigen Einflusses verschiedener Publikationen und Beiratstätigkeiten auf die Wirtschaftspolitik

Kurzfristiger Einfluss (Zeilenprozent von *N*)						
	Keiner	Gering	Fair	Groß	Sehr groß	*N*
Publikationen in Fachzeitschriften	44	44	12	–	–	34
Publikationen in populären Medien	–	44	44	9	3	34
Fernsehauftritte	3	42	36	15	3	33
Beiratstätigkeit						
– Wertautonomer Regierungsbeirat	6	38	35	21	–	34
– Wertgebundener Regierungsberater	–	24	30	33	12	33
– Zentralbankbeirat	3	19	50	22	6	32
– Arbeitgeberbeirat	13	33	47	7	–	30
– Gewerkschaftsbeirat	17	35	41	7	–	29
Langfristiger Einfluss (Zeilenprozent von *N)*						
Schulunterricht	6	32	30	27	6	34
Universitätslehre	–	9	29	47	15	34
Publikationen in Fachzeitschriften	9	41	27	18	6	34
Publikationen in populären Medien	9	36	39	15	–	33

Quelle: CEPR Experten-Survey; eigene Berechnungen.

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

Bezüglich des langfristigen Einflusses von Volkswirten auf die Wirtschaftspolitik schneidet – schlicht und einfach – die universitäre Lehre im Urteil der Experten am besten ab. Über 60 % halten den Einfluss für „groß“ bzw. sogar „sehr groß“. Hingegen schneiden populäre Medien schlecht ab; selbst dem Schulunterricht wird im Vergleich hierzu größerer Einfluss zugeschrieben. Hinter diesem Ergebnis steht offenbar die Überlegung, dass Entscheidungsträger sich stark an dem Basis-Wissen orientieren, dass sie in jungen Jahren gelernt haben. Darauf wird im abschließenden Abschnitt nochmals Bezug genommen werden.

Bezüglich des eigenen Erfolgs in der Politikberatung sind die befragten Experten zwar skeptisch, aber keineswegs völlig desillusioniert. Immerhin ein Drittel der 36 Experten gibt als Antwort auf eine „offene Frage“ an, dass ihre Arbeiten einen nachweisbaren Einfluss auf die Gesetzgebung eines Landes hatten. Als Beispiele nennen die Experten folgende Bereiche: Arbeitsmarktstatistik, Deregulierung von Elektrizitätswirtschaft, diverse Politikbereiche im „zentralgelenkten“ Frankreich, Erziehungsurlaub, Gestaltung der Europäischen Währungsunion, Privatisierung, Steuerpolitik.

2. Trade-off zwischen Forschung und Politikberatung?

Der Publikationserfolg der Experten ist von zentraler Bedeutung, weil er zeigt, ob Volkswirte, die Interesse an Politikberatung haben, in angesehenen Fachzeitschriften publizieren (können). Dahinter steht die These, dass gute Forschung (und Lehre) nicht genug Zeit lässt, um sich in der Politikberatung zu engagieren. Diese Aktivität wird in Form von Publikationen in populären Medien, Fernsehauftritten und Vorträgen gemessen, sowie mithilfe der Selbsteinschätzung der befragten Experten.

Tabelle 3 weist den Publikationserfolg in Form der arithmetischen Mittelwerte der Publikationen in verschiedenen Medien aus. Im Text wird auch der Median angegeben, da dieser weniger stark von „vielpublizierenden Ausreißern“ beeinflusst wird. Unterschieden wird in „Angelsachsen“, d. h. Experten, die in Großbritannien und den USA tätig sind, und „Nicht-Angelsachsen“, die in Kontinentaleuropa forschen und lehren. Diese Unterscheidung ist wegen des leichteren Zugangs der englischsprachigen Angelsachsen zu internationalen Zeitschriften von Bedeutung, aber auch – wie sich zeigen wird – aufgrund unterschiedlicher Arbeitsbedingungen.

Ein inzwischen allseits anerkannter Indikator für die Qualität volkswirtschaftlicher Forschung ist die Zahl der Veröffentlichungen in *referierten* Fachzeitschriften, bei denen also – wie in den einflussreichen Naturwissenschaften – vor der Veröffentlichung die Manuskripte eine Begutachtung durch Kollegen (*peers*) durchlaufen. Der Median der von den Experten jemals publizierten Beiträge in solchen Fachzeitschriften liegt bei 20,5 Artikeln; für die letzten drei Jahre beträgt der Me-

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

dian 5 und der Mittelwert 6,7 Aufsätze[8]. Unterscheidet man nach Angelsachsen und Nicht-Angelsachsen, dann gibt es praktisch keinen Unterschied der Publikationen in den letzten drei Jahren; dies dürfte an der Selektion der CEPR-Fellows liegen, die u. a. anhand der Qualität ihrer Forschung und ihrer Veröffentlichungen ausgewählt werden. Gleichwohl findet man deutlich mehr Lebenszeitpublikationen bei den Angelsachsen (deren Median beträgt 27 im Vergleich zu 19,5 der Nicht-Angelsachsen).

Tabelle 3

Arithmetische Mittel (einschließlich Null-Werte) verschiedener Publikationen und Vertragsformen der CEPR-Experten (*N* in Klammern)

	Kontinentaleuropäer	Angelsachsen
Aufsätze in referierten Zeitschriften		
– in den letzten drei Jahren	**6,3** (22)	**7,4** (22)
– insgesamt	**22,1** (22)	**32,3** (14)
Aufsätze in nicht-referierten Zeitschriften		
– in den letzten drei Jahren	**9,9** (17)	**5,1** (12)
– insgesamt	**36,9** (17)	**14,7** (17)
Aufsätze in populären Medien		
– in den letzten drei Jahren	**13,7** (20)	**1,1** (15)
TV / Radio Auftritte		
– in den letzten drei Jahren	**3,4** (20)	**3,1** (17)
Agenda-Setting Vorträge		
– in den letzten drei Jahren	**2,3** (17)[1)]	**1,4** (14)
Öffentliche Anhörungen		
– in den letzten drei Jahren		
– national	**0,7** (15)	**0,3** (15)
– international	**1,3** (15)	**0,3** (13)

[1)] ohne Einschluss eines Extremwertes (über 500 Vorträge).

Quelle: CEPR Expertensurvey; eigene Berechnungen.

Der Median der jemals publizierten Artikel in *nicht-referierten* Fachzeitschriften liegt bei 18 Artikeln, also in der Größenordnung der für referierte Artikel ermittelten Werte. Für die letzten drei Jahre beträgt der Median – ebenso wie für referierte Artikel – 5 und der Mittelwert liegt mit 7,9 auch in der Größenordnung der referierten Artikel. Unterscheidet man nach Angelsachsen und Nicht-Angelsachsen, dann findet man deutlich mehr solcher Artikel bei den Nicht-Angelsachsen, was

[8] Diese Angaben werden nicht anhand von bibliometrischen Analysen validiert, da die Angaben offenkundig hochplausibel sind.

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

daran liegen dürfte, dass in der kontinentaleuropäischen Stichprobe mehr prominente Volkswirte enthalten sein dürften als insbesondere in der US-Teilstichprobe, da das CEPR-Netzwerk europazentriert ist. Bezogen auf alle nicht-referierten Publikationen beträgt für Nicht-Angelsachsen der Median 30 im Vergleich zu 12,5 der Angelsachsen; die Mittelwerte betragen 36,9 und 14,7. Bei den referierten Publikationen betragen die Werte 7 und 3 sowie 9,9 und 5,1. In diesen Unterschieden wird deutlich, dass Kontinentaleuropäer für ihre Publikationen in ihrer Muttersprache offensichtlich nach wie vor oft nicht-referierte Zeitschriften auswählen (müssen).

Der Median der Veröffentlichungen in *populären Medien* (in den letzten drei Jahren*)* beträgt 1, wobei sich – wahrscheinlich aufgrund der fehlenden „Prominenten" bei den Angelsachsen – der Wert für die beiden Regionen mit 2,5 für Kontinentaleuropa und Null für die Angelsachsen deutlich unterscheidet. Die Mittelwerte liegen mit 13,7 und 1,1 noch deutlicher auseinander. Die niedrigen Werte liegen daran, dass fast die Hälfte der Experten gar nicht in populären Medien veröffentlicht, wobei dieser Anteil bei den Angelsachsen am größten ist.

Noch deutlich seltener sind Auftritte im *Fernsehen / Radio.* Der Median liegt (in den letzten drei Jahren) bei Null, da der Anteil derer, die diese Aktivität gar nicht angeben, bei knapp über 50 % liegt. Der Mittelwert liegt bei 3,2 und ist für die Kontinentaleuropäer mit 3,4 praktisch nicht größer als für die Angelsachsen mit 3,1. Auch der Median für *öffentliche Reden,* die dem *Agenda-Setting* dienen liegt (in den letzten drei Jahren) ebenfalls bei Null. Der Mittelwert von 52 Reden wird durch einen Experten, der über 500 Auftritte angibt, extrem verzerrt. Aber auch ohne diesen Ausreißer ist der Mittelwert der Kontinentaleuropäer mit 2,3 etwas höher als der der Angelsachsen mit 1,4 öffentlichen Reden. Der Median für öffentliche Anhörungen liegt (in den letzten drei Jahren) wiederum bei Null, wobei dahinter ein Anteil von etwa 80 % der Experten steht, die angaben, zu keinerlei öffentlichen Anhörungen eingeladen gewesen zu sein.

Über 40 % der Experten haben in den letzten drei Jahren keinerlei Veröffentlichungen in Tageszeitungen, Magazinen etc. gehabt, obwohl es sich bei den CEPR-Fellows um Wissenschaftler handelt, die Interesse an Politikberatung haben (sollen). Der Median für derartige Veröffentlichungen beträgt 1, während der weit höhere Mittelwert von 8,3 zeigt, dass diejenigen, die derartige Publikationen haben, dann auch recht aktiv sind (wobei freilich nur ein Viertel derer, die derartige Veröffentlichungen aufweisen, 10 und mehr populäre Veröffentlichungen innerhalb von drei Jahren hatten). Die Angelsachsen hatten einen Median von Null und einen Mittelwert von 1,1, während die Nicht-Angelsachsen einen Median von 2,5 und einen Mittelwert von 13,7 aufweisen. Die geringen Werte der Angelsachsen dürften vor allem daran liegen, dass – wie gesagt – die populären Spitzenwissenschaftler der USA nicht Fellows des europazentrierten CEPR-Netzwerkes sind.

Die zentrale Frage dieser Untersuchung ist, ob es einen Trade-off zwischen Veröffentlichungen in (sehr guten) referierten Fachzeitschriften einerseits und anderer-

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

seits populären Veröffentlichungen und Auftritten gibt und ob wirtschaftspolitischer Einfluss von derartigen populären Aktivitäten abhängt. Tabelle 4 beruht auf einer Unterteilung der Experten in „Vielpublizierer in referierten Fachzeitschriften“ und „Top-Publizierer“. Als „Vielpublizierer“ werden die Experten bezeichnet, die oberhalb des Medians der Publikationen in referierten Fachzeitschriften in den letzten drei Jahren liegen (6 und mehr Publikationen weisen 17 der 36 Experten auf). Als „Top-Publizierer“ werden diejenigen Experten eingestuft, die drei und mehr ihrer fünf Spitzenpublikationen in führenden internationalen Fachzeitschriften platziert haben (das sind 15 von 36 Experten).[9]

Zuerst einmal sei festgehalten, dass Top-Publizierer diejenigen sind, die am meisten internationale Zitierungen angeben.[10] Das dürfte auch damit zusammenhängen, dass die Top-Gruppe nicht nur qualitativ, sondern auch quantitativ besser publiziert als die andere Gruppe (9 zu 5 referierte Veröffentlichungen in den letzten drei Jahren). Die Top-Gruppe der Experten hat im übrigen auch mehr nicht-referierte Veröffentlichungen als die andere Gruppe (8,3 zu 6,6). Dies deutet nicht auf einen Trade-off zwischen hoher Forschungsleistung und Politikberatung hin. Freilich deuten die Zahlen bezüglich populären Veröffentlichungen und TV / Radio auf den vermuteten Trade-off hin: Die Top-Publizierer sind in den Massenmedien deutlich weniger präsent als die anderen (2 zu 14 populäre Veröffentlichungen, 4 zu 9 TV / Radio-Auftritte in den letzten drei Jahren). Allerdings ist es keineswegs so, dass die Top-Gruppe gar nicht in den Massenmedien präsent wäre.[11] Die Publikationen sagen freilich noch keineswegs Endgültiges über die Rezeption der verschiedenen Gruppen in der Öffentlichkeit und in der Politik aus.

Bezüglich größerer öffentlicher Vorträge (Agenda-Setting) zeigt sich kein Unterschied zwischen den Top-Publizierern und der anderen Gruppe (im Durchschnitt

---

[9] Das Ausmaß von Spitzenpublikationen wird anhand der Angaben der Experten über die beste Platzierung ihrer Veröffentlichungen gemessen. Anhand der von *Eichenberger et al.* (2000, 145) benutzten Liste der „inhaltlich breiten“ und – gemessen an der Zahl der Zitierungen – führenden amerikanischen und europäischen Zeitschriften kann bestimmt werden, wie viele dieser Zeitschriften ein Fellow mit seinen – nach eigener Einschätzung – fünf besten Publikationen abdeckt. Dabei handelt es sich um: American Economic Review, Journal of Political Economy, Quarterly Journal of Economics, Econometrica, Review of Economics and Statistics, Review of Economic Studies, European Economic Review, Economic Journal und Kyklos. Fast 10 % (von 37 Fällen) haben keine einzige Publikation in den führenden Fachzeitschriften und ein einziger hat alle seiner fünf besten Publikationen in einer dieser neun Zeitschriften platziert. Jeweils etwa ein Viertel der Experten kommt auf 1, 2 und 3 Platzierungen; 10 % auf deren vier. Bemerkenswerte Unterschiede zwischen Angelsachsen und Nicht-Angelsachsen gibt es nicht.

[10] Wissenschaftssoziologisch bemerkenswert ist, dass nur ein Fünftel aller Experten in der Lage ist, eine quantitative Zahl ihres Zitiererfolges geben zu können (7 von 32). Der Median für die Zitierungen beträgt 50. Diese Angaben wurden wiederum nicht bibliometrisch validiert, da sie plausibel sind.

[11] Betrachtet man die reine Anzahl der referierten Veröffentlichungen, dann erkennt man auch, dass eine größere Zahl derartiger Arbeiten zu Lasten der Präsenz in den Massenmedien geht.

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

Tabelle 4

Indikatoren zum Trade-off von akademischen Leistungen und effektiver Politikberatung (*N* in Klammern)

	Akademische Leistung gemessen an			
	viele referierte Aufsätze[1)]		Spitzenpublikationen[2)]	
	Ja	Nein	Ja	nein
Indikatoren zur Akademischen Leistung				
Internationale Zitationen[3)] in %	**50** (12)	**33** (15)	**73** (11)	**19** (16)
In den letzten drei Jahren				
– referierte Aufsätze	**9,9** (17)	**3,8** (19)	**9,1** (15)	**5,0** (21)
– Nicht-referierte Aufsätze	**8,9** (13)	**6,0** (15)	**8,3** (12)	**6,6** (16)
Indikatoren zur Politikberatung				
In den letzten drei Jahren				
– Populäre Veröffentlichungen	**4,0** (17)	**12,9** (17)	**2,1** (15)	**13,5** (19)
– TV / Radio Auftritte	**2,5** (17)	**3,9** (17)	**4,1** (14)	**8,7** (17)
– Öffentliche Vorträge	**2,3** (15)	**1,4** (17)[5)]	**1,9** (14)	**1,9** (17)[5)]
– Öffentliche Anhörungen				
– National	**0,2** (13)	**2,1** (16)	**0,2** (14)	**0,6** (15)
– International	**1,2** (14)	**0,6** (16)	**1,1** (15)	**0,7** (15)
Selbsteinschätzung				
– Direkter Einfluss auf Gesetze	**29** (17)	**33** (18)	**40** (15)	**25** (20)
– Erfolg als Agenda-Setter[4)]	**69** (13)	**53** (17)	**75** (12)	**50** (18)
– Einfluss auf Wirtschaftspolitik	**94** (17)	**83** (18)	**100** (15)	**80** (20)

[1)] Zahl der Veröffentlichungen oberhalb des Median der Expertenstichprobe (6 und mehr Veröffentlichungen).
[2)] Drei und mehr Platzierungen der besten fünf Veröffentlichungen in den neun internationalen Spitzenzeitschriften. Es handelt sich um American Economic Review, Journal of Political Economy, Quarterly Journal of Economics, Econometrica, Review of Economics and Statistics, Review of Economic Studies, European Economic Review, Economic Journal und Kyklos.
[3)] „good deal“, „often“ und „very often“.
[4)] „average“ und mehr.
[5)] ohne Ausreißer (mehr als 500 Vorträge).

Quelle: CEPR Expertensurvey; eigene Berechnungen.

jeweils nur 2 Vorträge in den letzen drei Jahren); diejenigen, die überdurchschnittlich in referierten Fachzeitschriften publizieren, geben sogar mehr derartiger Vorträge an als die anderen (2,3 zu 1,4). Bei Anhörungen ist das Bild gemischt, wobei das insgesamt sehr niedrige Niveau von Anhörungen auffällt: Top-Publizierer geben zwar weniger auf der nationalen Ebene, dafür aber mehr international „öffentliche Auskunft“.

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

Man darf diese Unterschiede aufgrund der sehr kleinen Fallzahl natürlich nicht überbewerten. Vorsichtig interpretiert lautet das Ergebnis, dass es keine gravierenden Unterschiede in der öffentlichen Präsenz der verschiedenen Publikationsgruppen der Experten gibt. Mit anderen Worten: Aufwändiges wissenschaftliches Arbeiten hindert nicht daran in der Öffentlichkeit ebenso präsent zu sein wie diejenigen, die auf wissenschaftliches Renommee weniger Wert legen (gegeben das ohnehin hohe wissenschaftliche Niveau der CEPR-Fellows).

Auch im Hinblick auf die Selbsteinschätzung des persönlichen Einflusses auf die Wirtschaftspolitik fallen Top- und Viel-Publizierer gegenüber den anderen Gruppen nicht auf. Im Gegenteil (vgl. den untersten Block von Tabelle 4): Insbesondere die Top-Publizierer unter den Experten sind mehr als die anderen davon überzeugt, wirtschaftspolitischen Einfluss zu haben. 40 % glauben, dass ihre Arbeit direkten Einfluss auf wirtschaftspolitisches Handeln und die Gesetzgebung hatte, 75 % halten sich für überdurchschnittlich erfolgreiche Agenda-Setter und alle glauben, dass ihre Arbeit Einfluss auf die Wirtschaftspolitik hat.

Auch der in Deutschland vielfach vermutete Zusammenhang zwischen „wenig Publikationen“ und „guter Lehre“ lässt sich – in Übereinstimmung mit der Analyse von *Backes-Gellner* (1989) für Deutschland – im Expertensample nicht nachweisen (nicht tabellarisch dargestellt). Sicherlich sollte man die positive Selbsteinschätzung der Qualität der Lehre nicht überbewerten (nur einer von 24 Experten gibt an, dass seine Lehrqualität unterhalb des Durchschnitts in seinem Fachbereich liegt). Deswegen wurden für die folgende Aussage auch diejenigen (7 von 24), die ihre Lehre *durchschnittlich* einschätzen, als *keine* guten Lehrer betrachtet. Vergleicht man die Mittelwerte der referierten und nicht-referierten Publikationen der letzten drei Jahre (weiter zurückliegende Publikationen kosten ja aktuell keine Zeit mehr), dann zeigt sich für „gute Lehrer“ sogar ein leicht höherer Wert für die Zahl der Publikationen als für „schlechte Lehrer“: Die erste Gruppe weist 7,3 referierte und 6,8 nicht-referierte Publikationen auf, die zweite Gruppe 5,6 und 6,7.

Während es offenbar keinen massiven Trade-off zwischen guter Forschung und volkswirtschaftlicher Politikberatung gibt, so gibt die Expertenbefragung freilich auch Hinweise, auf welchen materiellen Bedingungen der insgesamt größere Publikations- und Beratungserfolg der Spitzenvolkswirte, die überwiegend angelsächsische Volkswirte sind, beruht (vgl. Tabelle 5).

Die wöchentliche Arbeitszeit von durchschnittlich 55 Stunden unterscheidet sich nicht nach den beiden Regionen „Angelsachsen“ und „Kontinentaleuropa“. Auch die Belastung durch „Gremienarbeit“, die insbesondere in Deutschland von Hochschullehrern immer wieder beklagt wird, unterscheidet sich kaum für die beiden untersuchten Regionen. Die angelsächsischen Experten geben sogar mit 23 % zu einem kleineren Anteil als die Kontinentaleuropäer (mit 35 %) an, durch die Gremien ihres Fachbereichs nur marginal belastet zu sein. Andere Gremien, insbesondere Universitätsgremien, sind ohnehin nur für eine Minderheit von einem Drittel der Experten eine nennenswerte Art der Zeitverwendung.

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

Tabelle 5

Indikatoren zur professionelle Belastungen kontinentaleuropäischer und angelsächsischer CEPR-Experten (*N* in Klammern)

	Kontinentaleuropa	Angelsachsen
Wöchentliche Arbeitszeit (in Stunden)	**55**	**55**
(in % von *N*)	(19)	(13)
Belastung durch Fachbereich-Ausschüsse		
– marginal	**35**	**23**
– 3 und mehr	**18**	**31**
(in % von *N*)	(17)	(13)
Belastung durch Instituts-Ausschüsse		
– marginal	**25**	**73**
– 3 und mehr	**31**	**27**
(in % von *N*)	(16)	(11)
Nicht-marginale Belastung durch Universitäts-Gremien		
– marginal	**69**	**69**
– 3 und mehr	–	**8**
(in % von *N*)	(13)	(13)
Lehrdeputat (Zahl der Kurse pro Jahr)		
– Nicht-Doktoranden-Lehre	**3,6**	**2,6**
– Doktoranden-Lehre	**1,8**	**1,8**
(in % von *N*)	(16)	(11)

Quelle: CEPR Expertensurvey; eigene Berechnungen.

Aber das Lehrdeputat ist im Bereich der Nicht-Doktoranden-Lehre mit 3,6 Veranstaltungen pro Jahr in den kontinentaleuropäischen Ländern deutlich höher als in Angelsachsen mit 2,6 Kursen. Ein Drittel der Hochschullehrer (in beiden Regionen) gibt an, keine Doktorandenveranstaltungen abzuhalten. Der Rest gibt mit 1,8 Doktorandenkursen pro Jahr in beiden Regionen etwa gleich viele Kurse an. Insgesamt bestätigt sich die vielfach geäußerte Ansicht, dass an guten Universitäten (Research Universities) in Angelsachsen die gesamte Lehrlast geringer ist als in kontinentaleuropäischen Universitäten und der Anteil des Lehrdeputats, der für die forschungsnahe Doktorandenausbildung zur Verfügung steht, höher ist. Zu berücksichtigen ist außerdem, dass etwa 15 % der Experten sich mit Hilfe von Drittmitteln teilweise aus der Lehre freigekauft haben (*teaching buy out*); in dieser Gruppe findet sich kein deutscher Experte.

Völlig aus dem Rahmen fallend ist die Lehrbelastung in Deutschland: nur hier werden sechs und mehr Veranstaltungen pro Jahr angegeben (nicht tabellarisch ausgewiesen). Streng genommen müssten aufgrund des jährlichen Lehrdeputats

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

von 16 Semesterwochenstunden alle Deutschen mindestens acht Kurse pro Jahr angeben. Allerdings gibt es auch in Deutschland – informelle und formelle – Möglichkeiten der Reduktion des Lehrdeputats. Dazu zählt z. B. die offizielle Reduktion des Deputats durch eine „gemeinsame Berufung" mit einem außeruniversitären Forschungsinstitut. Diese Form der Berufung ist freilich weniger flexibel als der *buy out;* insbesondere setzt die gemeinsame Berufung nur für sehr wenige Hochschullehrer einen Anreiz, da klar ist, dass nur ganz wenige eine solche Berufung erreichen werden.

Auch die Betreuung von sechs und mehr Diplomarbeiten bzw. Master Theses pro Jahr wird nur in Deutschland angegeben. Die durchschnittliche Betreuungszahl beträgt in Angelsachsen 2,1 und in Nicht-Angelsachsen 4,3 (lässt man Deutschland draußen, sinkt die Zahl für diese Region auf 2,2 und unterscheidet sich damit praktisch nicht von Angelsachsen). Der geringeren Belastung durch Master Theses steht in Angelsachsen eine leicht höhere Zahl betreuter Doktorarbeiten gegenüber (2,5 gegenüber 2,4 in Nicht-Angelsachsen). Da Hochschullehrer im Allgemeinen von Doktorarbeiten in der eigenen Forschung profitieren, ist die Lehr-Situation von Spitzen-Hochschullehrern, wie sie CEPR-Experten darstellen, in angelsächsischen Ländern insgesamt besser als in Kontinentaleuropa und viel besser als in Deutschland.

C. Zusammenfassung und Schlussfolgerungen

Um die Frage, ob gute volkswirtschaftliche Forschung daran hindert, Politikberatung zu betreiben, auf einer verallgemeinerbaren empirischen Grundlage zu beantworten, wurde vom Autor ein Experten-Survey bei den Fellows des in London ansässigen Netzwerks CEPR durchgeführt. Die Teilnehmer am Survey stellen herausragende Volkswirte dar, deren Beschreibung ihrer Arbeitsbedingungen keineswegs repräsentativ für alle Volkswirte sein kann. Genau deswegen kann man für die Gestaltung der volkswirtschaftlichen Lehre, Forschung und Politikberatung aus den Antworten der Experten viel lernen.

Insgesamt zeigt sich, dass für Spitzenvolkswirte, die Politikberatung machen (wollen), kein Trade-off zwischen akademischen Publikationserfolgen und dem Grad der Politikberatung erkennbar ist. Zwar sind Top-Publizierer etwas weniger in populären Medien vertreten als andere, aber es gibt Hinweise, dass Spitzen-Publizierer in Form von großen öffentlichen Vorträgen und öffentlichen Anhörungen eher mehr denn weniger gehört werden als ihre Kollegen. Auch dürften Spitzen-Forscher eher von Entscheidungsträgern intern angehört werden. Die Aussage, dass es keinen erkennbaren Trade-off zwischen sehr guter volkswirtschaftlicher Forschung und Politikberatung gibt, dürfte drei ganz unterschiedliche Gründe haben.

Zum Ersten sind Spitzenforscher in der Regel auch fleißig. Sie wenden viel Zeit für ihren Beruf auf und sind deswegen auch – wie man entgegen populistischen

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

Vorurteilen weiß – in der Lehre erfolgreich (vgl. dazu bereits *Backes-Gellner* 1989). Spitzenvolkswirte wenden offenbar auch ausreichend Zeit für Politikberatung auf.

Zum Zweiten dürfte der Erfolg in der Politikberatung zumindest zum Teil aber auch dadurch zustande kommen, dass in angelsächsischen Hochschulen – wo die Spitzenvolkswirte forschen und lehren – das Lehrdeputat deutlich niedriger als insbesondere in Deutschland ist. In Großbritannien und den USA steht im Besonderen mehr Zeit für die Doktorandenausbildung zur Verfügung, von der die eigene Forschung der Hochschullehrer profitiert. Hinzu kommt: Das „Herauskaufen" aus Lehrverpflichtungen mit Hilfe von für Forschungszwecke eingeworbenen Drittmitteln ist für Spitzenforscher in den angelsächsischen Hochschulen üblich (*teaching buy out*), während es in Deutschland faktisch nicht möglich ist.

Zum Dritten dürfte der Erfolg der kleinen Gruppe von Spitzenvolkswirten, die alle angeben, Einfluss auf die Wirtschaftspolitik zu haben, auch daran begründet liegen, dass sie dem Schul- und Universitätsunterricht eine große Bedeutung zumessen. Gute Forschung und gute Lehre tragen damit gewissermaßen simultan auch zur Politikberatung bei. Allerdings nicht zur kurzfristigen Beratung. Aber volkswirtschaftlicher Unterricht hat nach Überzeugung vieler der befragten Experten durch die Vermittlung von Wissen einen langfristigen Einfluss auf spätere Entscheidungsträger. Ein besserer Ökonomie-Unterricht in Schulen und eine ökonomische Grundausbildung *aller* Studenten könnte somit zur Verbesserung der Wirtschaftspolitik beitragen.

Die eingangs zitierte Skepsis der Öffentlichkeit bezüglich des Nutzens der Volkswirtschaftslehre bezieht sich offenkundig auf den kurzfristigen Beratungserfolg. Langfristig dürften selbst viele der Journalisten, die ihren Zweifel an der (modernen) Volkswirtschaftslehre veröffentlichen, von Professoren der Volkswirtschaftslehre und anderen Hochschullehrern stark beeinflusst sein. Deswegen sei abschließend – obwohl bei ihm viel Skepsis mitschwingt – *John Maynard Keynes* (1936, S. 323) zitiert. Der wahrscheinlich einflussreichste aller Volkswirte hat im letzten Absatz seiner „General Theory" geschrieben: „Von … zeitgenössischen Stimmungen abgesehen, sind aber die Gedanken der Ökonomen und Staatsphilosophen, sowohl wenn sie im Recht, als wenn sie im Unrecht sind, einflussreicher als gemeinhin angenommen wird. (…) Praktiker, die sich ganz frei von intellektuellen Einflüssen glauben, sind gewöhnlich die Sklaven irgendeines verblichenen Ökonomen. Wahnsinnige in hoher Stellung, die Stimmen in der Luft hören, zapfen ihren wilden Irrsinn aus dem, was irgendein akademischer Schreiber ein paar Jahre vorher verfasste."

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

Literatur

Backes-Gellner, U. (1989): Zum Verhältnis von Forschung und Lehre in sozialwissenschaftlichen Fachbereichen. In: Helberger, Ch. (Hg.), Ökonomie der Hochschule I, Berlin / München, 51 – 76.

Bräuninger, M. / *Haucap,* J. (2001): Was Ökonomen lesen und schätzen: Ergebnisse einer Umfrage, in: Perspektiven der Wirtschaftspolitik, Bd. 2, 185 – 210.

Eichenberger, R. / *Meier,* U. / *Arpagaus,* R. (2000): Ökonomen, Publikationen und Zitationen: Ein europäischer Vergleich. In: Perspektiven der Wirtschaftspolitik, 1(2), 143 – 160.

Frey, B. S. / *Eichenberger,* R. (1993): American and European Economics and Economists. In: Journal of Economic Perspectives, Vol. 7, No. 4, 185 – 193.

Gellner, W. (1995): Ideenagenturen für Politik und Öffentlichkeit – Think Tanks in den USA und Deutschland. In: Studien zur Sozialwissenschaft, Bd. 157, Opladen.

Keynes, John Maynard (1936): Allgemeine Theorie der Beschäftigung, München / Leipzig.

Rotte, R. / *Zimmermann,* K. F. (1998): Fiscal Restraint and the Political Economy of EMU. In: Public Choice, 94, 385 – 406.

Schmidt, H. (1998): Globalisierung – Politische, ökonomische und kulturelle Herausforderungen. München.

Wagner, Gert G. / *Wiegard,* Wolfgang (2001): Volkswirtschaftliche Forschung und Politikberatung, in: I. Becker, N. Ott und G. Rolf. (Hg.), Soziale Sicherung in einer dynamischen Gesellschaft – Festschrift für Richard Hauser zum 65. Geburtstag, Frankfurt am Main

Wagner, Gert G. / *Hauser,* Richard / *Zimmermann,* Klaus F. (2004): Ökonomik der Volkswirtschaftslehre – Eine Analyse unter besonderer Berücksichtigung der wirtschaftspolitischen Beratung, Berlin (in Vorbereitung).

Wilts, A. (1999): Theoretical and Practical Research Orientations in Germany. In: Science Studies, 12, 3 – 20.

Wilts, A. (2000): Forms of research organization and their responsiveness to external goal setting. In: Research Policy, 29, 767 – 781.

Wyplosz, Ch. (1999): The Culture of Economic Policy Advice: An International Comparison With Special Emphasis on Europe. In: E. Mohr (Hg.), The Transfer of Economic Knowledge, Cheltenham.

Zimmermann, K. F. (2002) (Hg.): Neuere Entwicklungen in der Wirtschaftswissenschaft, Heidelberg.

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/

Autoren- und Herausgeberverzeichnis

Baetge, Jörg; Professor (emeritus) Dr. rer. pol., Dr. h. c., Universität Münster, Institut für Revisionswesen, Universitätsstr. 14 – 16, 48143 Münster

Drukarczyk, Jochen; Professor Dr. rer. pol., Dr. h. c., Universität Regensburg, Institut für Betriebswirtschaftslehre, 93040 Regensburg

Grothe, Helmut; Professor Dr. iur., Freie Universität Berlin, Fachbereich Rechtswissenschaft, Van't-Hoff-Str. 8, 14195 Berlin

Haas, Ulrich; Prof. Dr. iur., Universität Mainz, Fachbereich Rechts- und Wirtschaftswissenschaft, Jakob-Welder-Weg 9, 55099 Mainz

Hax, Herbert; Professor (emeritus) Dr. rer. pol., Dres. h. c., Universität zu Köln, Seminar für Allgemeine Betriebswirtschaftslehre und Finanzierungslehre, Albertus-Magnus-Platz, 50923 Köln

Heintzen, Markus; Professor Dr. iur., Freie Universität Berlin, Fachbereich Rechtswissenschaft, Van't-Hoff-Str. 8, 14195 Berlin

Hoffmann-Theinert, Roland; Dr. iur., RAuN, Görg Rechtsanwälte, Klingelhöferstr. 5, 10785 Berlin

Isensee, Josef; Professor (emeritus) Dr. iur., Dr. h. c., Universität Bonn, Rechts- und Staatswissenschaftliche Fakultät, Adenauerallee 24 – 42, 53113 Bonn

Kruschwitz, Lutz; Professor Dr. rer. pol., Freie Universität Berlin, Institut für Bank- und Finanzwirtschaft, Boltzmannstr. 20, 14195 Berlin

Schreyögg, Georg; Professor Dr. rer. pol., Freie Universität Berlin, Institut für Management, Garystr. 21, 14195 Berlin

Sickmann, Eric; Dipl.-Kfm., Universität Münster, Institut für Revisionswesen, Universitätsstr. 14 – 16, 48143 Münster

Uhlenbruck, Wilhelm; Hon.-Prof. Dr. iur., RiAG i. R., Friedrich-Schmidt-Str. 39, 50931 Köln

DOI https://doi.org/10.3790/978-3-428-51569-1 | Generated on 2025-12-16 20:37:25
OPEN ACCESS | Licensed under CC BY 4.0 | https://creativecommons.org/about/cclicenses/