

Wößmann, Ludger

## **Zentrale Abschlussprüfungen und Schülerleistungen. Individualanalysen anhand von vier internationalen Tests**

*Zeitschrift für Pädagogik 54 (2008) 6, S. 810-826*

urn:nbn:de:0111-opus-43782

in Kooperation mit / in cooperation with:

# **BELTZ**

<http://www.beltz.de>

### **Nutzungsbedingungen / conditions of use**

Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Die Nutzung stellt keine Übertragung des Eigentumsrechts an diesem Dokument dar und gilt vorbehaltlich der folgenden Einschränkungen: Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. Use of this document does not include any transfer of property rights and it is conditional to the following limitations: All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.  
By using this particular document, you accept the above-stated conditions of use.

### **Kontakt / Contact:**

**peDOCS**  
Deutsches Institut für Internationale Pädagogische Forschung (DIPF)  
Mitglied der Leibniz-Gemeinschaft  
Informationszentrum (IZ) Bildung  
Schloßstr. 29, D-60486 Frankfurt am Main  
E-Mail: [pedocs@dipf.de](mailto:pedocs@dipf.de)  
Internet: [www.pedocs.de](http://www.pedocs.de)

## Inhaltsverzeichnis

### *Thementeil: Systeme der Rechenschaftslegung und Schulentwicklung*

*Katharina Maag Merki/Knut Schwippert*

Systeme der Rechenschaftslegung und Schulentwicklung. Editorial ..... 773

*Daniel Koretz*

Test-based Educational Accountability. Research Evidence and Implications ..... 777

*Katharina Maag Merki/Eckhard Klieme/Monika Holmeier*

Unterrichtsgestaltung unter den Bedingungen zentraler Abiturprüfungen.  
Differenzielle Analysen auf Schulebene mittels Latent Class Analysen ..... 791

*Ludger Wößmann*

Zentrale Abschlussprüfungen und Schülerleistungen. Individualanalysen  
anhand von vier internationalen Tests ..... 810

*Hans Anand Pant/Miriam Vock/Claudia Pöhlmann/Olaf Köller*

Offenheit für Innovationen. Befunde aus einer Studie zur Rezeption der Bildungs-  
standards bei Lehrkräften und Zusammenhänge mit Schülerleistungen ..... 827

*Deutscher Bildungsserver*

Linktipps zum Thema „Accountability – Schulentwicklung“ ..... 846

### *Allgemeiner Teil*

*Klaus-Jürgen Tillmann*

Schulreform – und was die Erziehungswissenschaft dazu sagen kann ..... 852

*Kathrin Dederling*

Der Einfluss bildungspolitischer Maßnahmen auf die Steuerung des  
Schulsystems. Neue Erkenntnisse aus empirischen Fallstudien ..... 869

*Jürgen Reyer/Diana Franke-Meyer*  
Muss der Bildungsauftrag des Kindergartens „eigenständig“ sein? ..... 888

*Besprechungen*

*Hans-Christoph Koller*  
Heinz-Elmar Tenorth/Rudolf Tippelt (Hrsg.): Beltz Lexikon Pädagogik ..... 906

*Fritz Osterwalder*  
Holger Böning/Hanno Schmitt/Reinhart Siebert (Hrsg.): Volksaufklärung ..... 909

*Ulrich Herrmann*  
Hanno Schmitt/Anke Lindemann-Stark/Christophe Losfeld (Hrsg.): Briefe von  
und an Joachim Heinrich Campe ..... 913

*Roland Reichenbach*  
Eckart Liebau/Jörg Zirfas (Hrsg.): Ungerechtigkeit der Bildung – Bildung der  
Ungerechtigkeit  
Heiner Drerup/Werner Fölling (Hrsg.): Gleichheit und Gerechtigkeit ..... 915

*Ewald Terhart*  
Marilyn Cochran-Smith/Sharon Feiman-Nemser/D. John McIntyre/  
Kelly E. Demers (Eds.): Handbook of Research on Teacher Education  
Tony Townsend/Richard Bates (Eds.): Handbook of Teacher Education  
Marilyn Cochran-Smith/Kenneth M. Zeichner (Eds.):  
Studying Teacher Education ..... 921

*Dokumentation*

Pädagogische Neuerscheinungen ..... 928

Ludger Wößmann

# Zentrale Abschlussprüfungen und Schülerleistungen

*Individualanalysen anhand von vier internationalen Tests*

**Zusammenfassung:** Zentrale Abschlussprüfungen lassen sich im Rahmen von Prinzipal-Agenten-Modellen als „Accountability“-Maßnahme verstehen, die Schüler und Schulen für ihre erbrachten Bildungsleistungen verantwortlich macht. Umfassende Regressionsanalysen anhand der Schülerindividualdaten von vier internationalen Schüler Leistungsvergleichen – TIMSS 1995, TIMSS 1999, PISA 2000 und PISA 2003 – zeigen, dass zentrale Abschlussprüfungen im internationalen Vergleich mit wesentlich besseren Schülerleistungen einhergehen. Der Zentralprüfungseffekt findet sich ebenfalls im Vergleich der deutschen Bundesländer, und die dort gefundene Effektgröße lässt sich statistisch nicht von der internationalen Schätzung unterscheiden. Darüber hinaus belegen die internationalen Ergebnisse, dass zentrale Abschlussprüfungen ansonsten tendenziell eher negative Effekte erhöhter Schulautonomie vielfach in positive Effekte umdrehen. Dies belegt, dass zentrale Abschlussprüfungen nicht nur die Schüler, sondern auch die Entscheidungsträger in den Schulen zu leistungsfördernderem Verhalten bewegen.

## 1. Einleitung

In der schulpolitischen Debatte vieler Länder werden derzeit Maßnahmen erwogen, die die Schulen für ihren Unterrichtserfolg verstärkt in die Verantwortung nehmen sollen. Solche „Accountability“-Maßnahmen messen die Bildungsleistungen der Schülerinnen und Schüler und verbinden diese mit Konsequenzen verschiedener Art – von der bloßen Veröffentlichung der Ergebnisse über Bonusse für gutes Abschneiden oder zusätzlichen Unterstützungsmaßnahmen bis hin zur Schulschließung bei wiederholtem schlechtem Abschneiden.

In diesem Sinne können auch zentrale Abschlussprüfungen wie das Zentralabitur und zentrale Bestandteile in Prüfungen der Mittleren Reife als „Accountability“-Maßnahmen verstanden werden. Durch die externe Leistungsmessung machen sie die Schüler und damit letztlich indirekt auch Lehrer und Schulen für ihre Leistungen verantwortlich. In diesem Beitrag geht es darum, wie sich solche zentralen Abschlussprüfungen auf die erzielten Schülerleistungen auswirken.

Nach Bishop (vgl. 1995, 1997, 2006) lassen sich „Curriculum-basierte externe Abschlussprüfungssysteme“ über eine Reihe von Merkmalen definieren: 1. Curriculum-basierte externe Abschlussprüfungssysteme produzieren Signale der Schülerleistungen, die echte Konsequenzen für die Schüler haben. 2. Sie bewerten Leistungen in Relation zu einem externen Standard, nicht relativ zu anderen Schülern in derselben Klasse oder Schule. 3. Sie sind nach Fächern organisiert und auf den Inhalt spezifischer Kurssequenzen abgestimmt. 4. Sie differenzieren zwischen mehreren Leistungsniveaus in einem

Fach und geben nicht nur ein „bestanden/nicht bestanden“ Signal. 5. Ihre Durchführung erreicht flächendeckend nahezu alle Schüler der weiterführenden Schulen. 6. Sie bewerten einen großen Teil dessen, was Schüler in dem Fach wissen sollen.

Schon für Anfang der 1990er Jahre durchgeführte internationale Schülerleistungstests wurde anhand von aggregierten Länderdaten festgestellt, dass Länder mit zentralen Abschlussprüfungen signifikant bessere Durchschnittsleistungen aufweisen (vgl. Bishop 1995, 1997). Studien für Länder, in denen die Existenz zentraler Abschlussprüfungen regional variiert, bestätigen ebenfalls, dass die Leistungen in Regionen mit zentralen Abschlussprüfungen besser sind (vgl. Bishop 1995, 1997 für kanadische Provinzen; Bishop 1995 und Bishop/Moriarty/Mane 2000 für den U.S.-Bundesstaat New York im Vergleich zu den anderen Bundesstaaten und Jürges/Schneider/Büchel 2005 und Jürges/Richter/Schneider 2005 für deutsche Bundesländer).

Im vorliegenden Beitrag sollen nun die Ergebnisse verschiedener Studien zusammengetragen werden, die die Mikrodaten internationaler Vergleichstests auf Schüler-ebene nutzen, um den Zusammenhang zwischen zentralen Abschlussprüfungen und erzielten Schülerleistungen empirisch zu untersuchen. Da die Existenz zentraler Abschlussprüfungen bis auf die wenigen, gerade genannten Ausnahmen ein nationales Merkmal ist, bietet sich die internationale Analyse an. Durch die Nutzung der Mikrodaten ist es möglich, auf umfassende Weise die Effekte von individuellen Schüler-eigenschaften, familiärem Hintergrund, Ressourcenausstattung der Schulen und weiteren institutionellen Rahmenbedingungen zu kontrollieren. Darüber hinaus lässt sich anhand der Mikrodaten untersuchen, ob sich die Effekte zentraler Abschlussprüfungen für verschiedene Schüler und Schulen – etwa solche mit mehr oder weniger Entscheidungsautonomie – unterscheiden.

Im Einzelnen beziehen sich die berichteten Studien auf die Trends in International Mathematics and Science Study (TIMSS) in 1995 (vgl. Wößmann 2001, 2003a), ihre Wiederholung in 1999 (vgl. Wößmann 2003b, 2005), das Programme for International Student Assessment (PISA) in 2000 (vgl. Fuchs/Wößmann 2007; Wößmann 2007a) und seine Wiederholung in 2003 (vgl. Wößmann u.a. 2007). Darüber hinaus wird über eine Studie berichtet, die die im internationalen PISA-2003-Vergleich gefundenen Effekte zentraler Abschlussprüfungen den im deutschen Bundesländervergleich gefundenen Effekten gegenüberstellt (vgl. Wößmann 2007b). Da es im vorliegenden Beitrag um eine zusammenschauende Betrachtung der Ergebnisse zahlreicher Studien geht, wird für detaillierte Angaben zu den jeweiligen Datensätzen, Modellspezifikationen und Ergebnissen auf die zugrunde liegenden Studien verwiesen.

Der Beitrag ist folgendermaßen strukturiert. Der folgende Abschnitt stellt den theoretischen Zusammenhang zwischen zentralen Abschlussprüfungen und Schülerleistungen knapp dar. Abschnitt 3 erläutert das empirische Modell und die Daten. Abschnitt 4 berichtet die internationale Evidenz aus TIMSS 1995, TIMSS 1999, PISA 2000 und PISA 2003. Abschnitt 5 vergleicht die internationalen Befunde mit nationalen Befunden der deutschen Bundesländer. Abschnitt 6 behandelt den Zusammenhang zwischen zentralen Abschlussprüfungen und Schulautonomie. Der letzte Abschnitt schließt mit einigen zusammenfassenden Bemerkungen.

## 2. Theorie

Der Bildungsprozess lässt sich als ein Netzwerk von Prinzipal-Agenten-Beziehungen modellieren, in denen ein Prinzipal (z.B. Eltern) einen Agenten (z.B. einen Schulleiter) damit beauftragt, eine Leistung (die Bildung des Kindes) für den Prinzipal durchzuführen. Unter den beiden Bedingungen, dass der Agent eine vom Prinzipal abweichende Interessenlage hat (z.B. möglichst geringer Arbeitsaufwand) und dass keine perfekten Informationen über das Verhalten des Agenten vorliegen, führt eine solche Delegation von Aufgaben zu suboptimalen Ergebnissen (vgl. Laffont/Martimort 2002). Weichen die Interessen des Agenten nämlich von denen des Prinzipals ab und bestehen asymmetrische Informationen darüber, was der Agent tatsächlich leistet, so kann der Agent seine eigenen Interessen anstatt der Interessen des Prinzipals verfolgen, ohne dass der Prinzipal dies feststellen und sanktionieren kann.

Zentrale Prüfungen können dem Problem der unvollständigen Überprüfung („Monitoring“) der Handlungen der Agenten im Bildungssystem ein Stück weit beikommen, indem sie Informationen über die Leistungen der einzelnen Schüler relativ zur gesamten Schülerpopulation erzeugen. Indem sie die der Prinzipal-Agenten-Beziehung inhärenten Monitoring-Probleme verringern, erzeugen zentrale Prüfungen Anreize, die die Interessen des Agenten stärker mit den Interessen des Prinzipals und damit mit den Zielen des Bildungssystems in Einklang bringen.

Im Einzelnen lassen sich mindestens drei Wirkungskanäle zentraler Abschlussprüfungen modellieren (vgl. Bishop/Wößmann 2003 für eine formaltheoretische Modellierung dieser Zusammenhänge). Erstens machen zentrale Abschlussprüfungen den Leistungsstand der Schüler für Eltern, Lehrer, potentielle Arbeitgeber und weiterführende Bildungseinrichtungen sichtbar und damit vergleichbar. Die Noten des Abschlusszeugnisses haben etwa für potentielle Arbeitgeber eine wesentlich größere Signalwirkung bezüglich der tatsächlichen Bildungsleistungen eines Bewerbers, wenn die Noten durch externe Prüfung eine weitgehende Vergleichbarkeit aufweisen. Dadurch erhöhen sich die externen Belohnungen für Lernanstrengungen der Schüler. Der Anreiz zu lernen erhöht sich, wenn bessere Leistungen belohnt werden. Folglich erhöhen zentrale Prüfungen die Schülerleistungen. Zweitens machen es zentrale Abschlussprüfungen unmöglich, dass durch kollektiven Mitschülerdruck die Standards des Unterrichts gesenkt werden, ohne dass dies die durchschnittliche Benotung der Klasse senken würde. Drittens eröffnen zentrale Abschlussprüfungen den Eltern und Schulleitern, ob die Lehrer eine erfolgreiche Wissensvermittlung leisten. Damit entstehen auch für die Lehrer verstärkte Anreize, den Schülern möglichst viel des erwarteten Stoffes beizubringen. Ein Aspekt dessen besteht auch darin, dass zentrale Abschlussprüfungen es verhindern, dass ganze Wissensgebiete in einzelnen Klassen ohne Konsequenzen für die Benotung ausgelassen werden können.

### 3. Empirisches Modell und Daten

#### 3.1 Spezifikation des empirischen Modells

Empirisch lässt sich der Zusammenhang zwischen zentralen Abschlussprüfungen und Schülerleistungen schätzen, indem die Leistungen von Schülern in Schulsystemen mit Zentralprüfungen mit den Leistungen von Schülern in Schulsystemen ohne Zentralprüfungen international verglichen werden. Der Effekt  $\alpha$  von Zentralprüfungen  $Z$  wird in internationalen multiplen Regressionsanalysen anhand von Schülerindividualdaten mit Hilfe folgender Spezifikation geschätzt:

$$(1) \quad T_{iksl} = \alpha Z_l + B_{iksl} \beta + R_{ksl} \gamma + I_{ksl} \delta + a + \varepsilon_{iksl},$$

wobei  $Z_l$  der Anteil der Schüler in Land  $l$  ist, der an Zentralprüfungen teilnimmt. (Da es sich meistens um nationale Zentralprüfungssysteme handelt, handelt es sich bei  $Z$  weitgehend um eine Indikatorvariable, die die Dummywerte 0 oder 1 annimmt.)  $T_{iksl}$  ist das Testergebnis des Schülers  $i$  in Klasse  $k$  in Schule  $s$  in Land  $l$ . Neben einer Konstante  $a$  beinhaltet das Modell je nach Spezifikum des internationalen Datensatzes zwischen 40 und 60 Kontrollvariablen aus folgenden Bereichen: Variablen des persönlichen und familiären Hintergrunds der Schüler  $B$  (z.B. Bildungsstand der Eltern, Migrationshintergrund, usw.), Variablen der schulischen Ressourcenausstattung und Lehrermerkmale  $R$  (z.B. Klassengrößen, Bildungsausgaben, materielle Ausstattung, Lehrerfahrung der Lehrer, usw.) und Variablen institutioneller Merkmale der Schulsysteme  $I$  (z.B. Zentralität von Curriculum und Lehrbuchfestsetzung, Schulautonomie in verschiedenen Entscheidungsbereichen, Anteil privat geleiteter Schulen, usw.). Eine Liste der in den einzelnen Studien aufgenommenen Kontrollvariablen findet sich in den jeweils zitierten zugrunde liegenden Referenzen.

Um die hierarchische Struktur der Daten abzubilden, umfasst der Schätzfehler  $\varepsilon$  der Gleichung nicht nur eine schülerspezifische Komponente, sondern auch höhere Komponenten auf Ebene von Klassen, Schulen und Ländern. Die hierarchische Struktur des Modells wird dadurch modelliert, dass in der Varianz-Kovarianz-Matrix der Fehlerterme eine gegenseitige Abhängigkeit der Fehlerterme aller Schüler innerhalb eines Landes zugelassen wird. Durch diese sogenannte Cluster-robuste Schätzmethode (clustering-robust linear regression, CRLR) werden lediglich die Schätzfehler der verschiedenen Länder als unabhängig voneinander angenommen, was eine unverzerrte Berechnung der Standardfehler ermöglicht (vgl. dazu etwa Deaton 1997, S. 74-78).<sup>2</sup> Der stratifizier-

2 Die in der Ökonometrie üblichen regressionsanalytischen Modelle sind eng mit den in der Erziehungswissenschaft üblichen Mehrebenenmodellen verwandt. Ein Vorteil der CRLR besteht darin, nicht der strikten Annahmen der Hierarchical Linear Models (HLM) zu bedürfen, dass die Effekte zufällig und normal verteilt sind. Cohen und Baldi (vgl. 1998) belegen, dass CRLR auch dann sowohl konsistente als auch robuste Parameter und Standardfehler liefert, wenn diese Annahmen durch Ausreißer oder eine verzerrte Fehlerstruktur verletzt werden, während HLM in diesen Fällen zu signifikant unterschätzten Standardfehlern auf höheren Ebenen bzw. verzerrten Parameterschätzungen führen kann.

ten Stichprobenerhebung der internationalen Schülerleistungstests wird dadurch Rechnung getragen, dass die Observation eines jeden Schülers innerhalb seines Landes mit seiner Erhebungswahrscheinlichkeit gewichtet wird (vgl. Wooldridge 2001). In der internationalen Analyse erhält jedes Land die gleiche aggregierte Gewichtung.

### 3.2 *Mikrodaten aus vier internationalen Schülerleistungstests*

Den Analysen dieses Beitrags liegen die Schülerindividualdaten von vier internationalen Schülerleistungstests zugrunde, die jeweils repräsentative Stichproben der Schüler in jedem teilnehmenden Land erhoben haben. Die Testergebnisse sind jeweils so skaliert, dass sie einen internationalen Mittelwert von 500 und eine Standardabweichung von 100 aufweisen (bei PISA bezogen auf die OECD-Länder). Neben den Testergebnissen umfassen die Datenbanken umfangreiche Hintergrundinformationen über den individuellen und familiären Hintergrund jedes teilnehmenden Schülers, der Ressourcenausstattung der Schulen und der institutionellen Rahmenbedingungen des Schulsystems.

Die erste TIMSS-Studie wurde 1995 von der „International Association for the Evaluation of Educational Achievement“ (IEA), einer unabhängigen Kooperation nationaler Forschungsinstitute, durchgeführt. 1995 testete TIMSS Schüler der beiden Jahrgangsstufen mit dem größten Anteil 13-Jähriger (üblicherweise die 7. und 8. Klasse). Für die hier durchgeführten Analysen hat TIMSS international vergleichbare Daten für 266.545 Schüler an 6.107 Schulen in 39 Ländern geliefert (vgl. Beaton u.a. 1996 zu TIMSS; vgl. Wößmann 2003a zur Datenaufbereitung für die Sekundäranalyse, etwa zur Imputation fehlender Werte und für eine Auflistung der Kontrollvariablen). Die IEA hat 1999 eine Wiederholungsstudie des TIMSS-Tests durchgeführt, die die höhere der beiden in der ersten Studie getesteten Jahrgangsstufen testete und 180.544 Schüler in 38 Ländern umfasst (vgl. Mullis u.a. 2000; Wößmann 2003b). Da sich die Stichprobe der teilnehmenden Länder zwischen den beiden Tests unterscheidet, umfasst die zusammengefasste Datenbank der beiden TIMSS-Tests 77 Länderobservationen aus 54 unterschiedlichen Ländern (447.089 Schüler).

Die Organisation for Economic Co-operation and Development (OECD) hat 2000 die PISA-Studie durchgeführt, die 15-jährige Schüler testete. Die PISA-2000-Datenbank umfasst 31 Länder mit 175.227 Schülern, von denen 96.855 in Mathematik und 96.758 in Naturwissenschaften getestet wurden (vgl. OECD 2001; Fuchs/Wößmann 2007). Die OECD hat die PISA-Studie 2003 wiederholt, für die Daten über 219.794 Schüler aus 29 OECD-Ländern vorliegen (vgl. OECD 2004; Wößmann u.a. 2007).

Daten über zentrale Abschlussprüfungen am Ende der Sekundarstufe beziehen sich auf alle Formen der Curriculum-basierten externen Abschlussprüfungssysteme (vgl. Bishop 1997), nicht aber auf universitäre Eingangstests, da zumeist nicht alle Schüler an diesen teilnehmen und sie damit keinen integralen Bestandteil des Schulsystems darstellen. Die Informationen über Zentralprüfungen stammen aus vergleichenden Bildungsstudien, Bildungsenzyklopädien, Interviews mit Repräsentanten der nationalen Schulsysteme, Regierungsdokumenten und Hintergrundpapieren.

## 4. Internationale Evidenz

### 4.1 Ergebnisse für TIMSS 1995 und TIMSS 1999

Tabelle 1 berichtet den geschätzten Effekt von zentralen Abschlussprüfungen auf die Schülerleistungen in TIMSS 1995 und in TIMSS 1999 sowohl für die Mathematik- als auch für die Naturwissenschaftsleistungen. Die Ergebnisse belegen, dass Schüler in Schulsystemen mit zentralen Abschlussprüfungen in beiden Tests und in beiden Fächern statistisch signifikant besser abschneiden als Schüler in Schulsystemen ohne zentrale Prüfungen, wobei eine große Anzahl an Einflussfaktoren im Bereich des familiären Hintergrundes, der schulischen Ressourcenausstattung und der institutionellen Rahmenbedingungen kontrolliert wurde (vgl. Wößmann 2001, 2002, 2003a für Details der TIMSS-1995-Untersuchung und Wößmann 2003b, 2005 für Details der TIMSS-1999-Untersuchung). 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 Leistungszuwachs in einem ganzen Schuljahr. Die Größe des geschätzten Effektes ist zwischen den beiden Studien statistisch nicht signifikant voneinander zu unterscheiden.

| Tab. 1: Zentralprüfungen und Schülerleistungen in TIMSS  |             |             |                            |                             |
|--|-------------|-------------|----------------------------|-----------------------------|
|  | TIMSS 1995  | TIMSS 1999  | TIMSS 1995 +<br>TIMSS 1999 | TIMSS 1995 +<br>TIMSS 1999a |
| Mathematik   | 40,9 (13,5) | 47,0 (13,5) | 42,7 (9,8)                 | 28,6 (13,2)                 |
| Naturwissenschaften  | 39,7 (9,9)  | 35,9 (12,9) | 35,9 (8,3)                 | 41,7 (10,8)                 |
| Anzahl Schüler   | 266.545     | 180.544     | 447.089                    | 447.089                     |
| Anzahl Länder  | 39          | 38          | 77                         | 77                          |
| Schätzer des Zentralprüfungs-Koeffizienten. – Abhängige Variable: TIMSS-Testergebnis. – Kontrollvariablen: 48 Schüler-, Familien-, Ressourcen-, Lehrer- und Institutionenmerkmale. – Clusterrobuste Standardfehler (auf Länderebene) in Klammern. a Die Schätzung enthält Dummies für neun Weltregionen. (Quelle: Basierend auf Wößmann 2003b) |             |             |                            |                             |

Im Prinzip ist es denkbar, dass diese Querschnittsschätzer des Zentralprüfungseffektes durch Endogenitätsprobleme verzerrt sind (vgl. Wößmann 2003b für Details der folgenden Analysen). So könnte es unberücksichtigte Faktoren auf Länderebene geben, die mit der Existenz von Zentralprüfungen korrelieren und die die Korrelation mit den Schülerleistungen verursachen. Folgende vier Bereiche seien hier genannt: andere institutionelle Gegebenheiten des Schulsystems, allgemeiner Zentralisationsgrad eines Landes, Homogenität der Bevölkerung sowie kulturelle Unterschiede. Insoweit die Existenz von Zentralprüfungen nicht zufällig über die Länder verteilt ist, sondern in zentralisierten oder homogenen Ländern besonders häufig auftritt oder mit anderen Institutionen oder kulturellen Merkmalen einhergeht, und insoweit diese anderen Ländermerkmale

Unterschiede in den Schülerleistungen hervorrufen, würde der einfache Querschnittsschätzer des Zentralprüfungseffektes verzerrt geschätzt.

In den ersten drei Bereichen ist es möglich, das Problem durch die Berücksichtigung zusätzlicher entsprechender Kontrollvariablen anzugehen. 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. 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 (gemessen als Wahrscheinlichkeit, dass zwei zufällig gewählte Personen eines Landes nicht zur selben ethnolinguistischen Gruppe gehören, vgl. Mauro 1995) als Kontrollvariablen in die Spezifikation aufgenommen. Der geschätzte Zentralprüfungseffekt ändert sich dadurch nicht nennenswert.

Um schließlich zu testen, ob der Zentralprüfungseffekt mit sonstigen kulturellen Unterschieden zwischen Ländern konfundiert ist, können Dummies für neun Weltregionen als zusätzliche Kontrollvariablen eingefügt werden. Dadurch erfolgt die Schätzung des Zentralprüfungseffektes ausschließlich aufgrund von innerregionaler Variation, so dass er nicht mehr durch interregionale kulturelle Unterschiede, wie etwa zwischen asiatischen und europäischen Wertesystemen, beeinflusst wird. Die Schätzungen zeigen sogar dann statistisch signifikante Zentralprüfungseffekte, wenn sämtliche Variation zwischen den neun Regionen Westeuropa, Osteuropa, Nordamerika, Südamerika, Ozeanien, Asien, Mittlerer Osten, Nordafrika und Südafrika für die Schätzung des Zentralprüfungseffektes unberücksichtigt bleibt (Tabelle 1). 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 abzubilden.

#### 4.2 *Ergebnisse für PISA 2000 und PISA 2003*

Tabelle 2 berichtet vergleichbare Schätzergebnisse für die beiden PISA-Studien. Auch hier zeigt sich, dass Schüler in Ländern mit zentralen Abschlussprüfungen signifikant höhere Mathematik- und Naturwissenschaftsleistungen aufweisen als Schüler in Ländern ohne zentrale Abschlussprüfungen (vgl. Fuchs/Wößmann 2007 für Details der PISA-2000- und Wößmann u.a. 2007 für Details der PISA-2003-Untersuchung). Die Größenordnung des Zentralprüfungseffektes liegt bei 15,0 bis 24,5 Prozent einer internationalen PISA-Standardabweichung. Aufgrund der unterschiedlichen Teilnehmerländer und der dadurch unterschiedlichen Skalierung der Testpunkte zwischen TIMSS und PISA sind die geschätzten Effektgrößen nicht direkt miteinander vergleichbar. Die numerisch kleineren Werte für die PISA-Studien lassen sich zum größten Teil auf die unterschiedliche Skalierung zurückführen. Skaliert man die Tests so um, dass die jeweils in

beiden Tests teilnehmenden Länder eine einheitliche Streuung untereinander aufweisen, so sind die in den PISA-Tests geschätzten Zentralprüfungseffekte nahezu in derselben Größenordnung wie die in den TIMSS-Tests geschätzten Effektgrößen (vgl. Wößmann 2005).

| Tab. 2: Zentralprüfungen und Schülerleistungen in PISA   |                 |             |
|--|-----------------|-------------|
|  | PISA 2000       | PISA 2003   |
| Mathematik   | 19,1 (9,6)      | 24,5 (10,1) |
| Naturwissenschaften  | 15,0 (9,1)      | 21,2 (7,6)  |
| Anzahl Schüler   | 96.855 / 96.785 | 219.794     |
| Anzahl Länder  | 31              | 29          |
| Schätzer des Zentralprüfungs-Koeffizienten. – Abhängige Variable: PISA-Testergebnis. – Kontrollvariablen: 49 Schüler-, Familien-, Ressourcen-, Lehrer- und Institutionenmerkmale. – Clusterrobuste Standardfehler (auf Länderebene) in Klammern. (Quelle: Basierend auf Working-Paper-Version von Fuchs/Wößmann 2007 für PISA 2000; Wößmann u.a. 2007 für PISA 2003 Mathematik; eigene Berechnungen für PISA 2003 Naturwissenschaften) |                 |             |

Wößmann u.a. (vgl. 2007) untersuchen auch die Effekte einer Reihe weiterer „Accountability“-Maßnahmen auf die Schülerleistungen in PISA 2003. Es zeigt sich, dass weitere „Accountability“-Maßnahmen zusätzliche, additiv geschätzte positive Zusammenhänge mit den Schülerleistungen aufweisen. Dazu gehört etwa eine weitere primär auf Schüler abzielende Maßnahme, nämlich die Nutzung von Prüfungen für Versetzungsentscheidungen. Aber auch auf Lehrer abzielende Maßnahmen wie das interne Monitoring von Unterrichtsstunden durch den Schulleiter und das externe Monitoring von Unterrichtsstunden durch Schulinspektoren wirken sich positiv auf die Schülerleistungen aus. Schließlich weisen auch primär auf Schulen abzielende Maßnahmen wie die Nutzung von Prüfungen zum Schulvergleich mit regionalen oder nationalen Leistungen positive Effekte auf die Leistungen der Schüler auf.

## 5. Nationale Evidenz

In Deutschland ergibt sich aufgrund der Kultushoheit der Länder und der (in 2003 noch) klaren Einteilung in sieben Bundesländer mit zentralen Abschlussprüfungen und neun Bundesländer ohne zentrale Abschlussprüfungen die Möglichkeit, den Zusammenhang zwischen zentralen Abschlussprüfungen und Schülerleistungen anhand regionaler Variation innerhalb eines Landes zu schätzen (vgl. Wößmann 2007b für Details der folgenden Analysen). Der Vorteil des Bundesländervergleichs besteht darin, dass mögliche verbleibende Verzerrungen aufgrund unbeobachteter Heterogenitäten im internationalen Vergleich, etwa durch Unterschiede im sprachlichen oder rechtlichen Raum, noch eindeutiger umgangen werden können. Der Nachteil liegt darin, dass aufgrund der sechzehn Bundesländerobservationen nur eine geringe Anzahl statistischer

Freiheitsgrade für die Analyse bleibt. Darüber hinaus lassen sich aufgrund der Tatsache, dass die benötigten Mikrodaten unabhängigen Forschungsinstituten nicht zugänglich sind, keine Individualanalysen durchführen, die gezielter für individuelle Hintergründe kontrollieren und für Untergruppen variierende Effekte abbilden könnten. Gleichwohl erlauben die vom PISA-E-Konsortium bereitgestellten Bundesländermaße der Testergebnisse und diverse Hintergrundmaße im PISA-E-2003-Test (vgl. Prenzel u.a. 2005) eine Analyse auf Bundesländerebene, und eine gemeinsame Betrachtung der Bundesländer mit den PISA-Teilnehmerstaaten ermöglicht eine Analyse von insgesamt 54 Observationen.

Die deskriptive Darstellung in Abbildung 1 zeigt, dass überhaupt nur ein einziges Bundesland mit zentralen Abschlussprüfungen schlechter abschneidet als irgendein Bundesland ohne zentrale Abschlussprüfungen. Mit dieser Ausnahme schneidet *jedes einzelne* Bundesland mit Zentralprüfungen besser ab als *jedes einzelne* Bundesland ohne Zentralprüfungen. Die in Tabelle 3 berichteten Schätzergebnisse belegen, dass dieser Zusammenhang auch im multivariaten Regressionszusammenhang besteht. In den weitergehenden Analysen zeigt sich, dass nach Berücksichtigung der beiden im hier berichteten Basismodell enthaltenen Kontrollvariablen (ein Index des sozioökonomischen Hintergrundes und die Beschäftigungsquote der Väter) kein weiteres verfügbares Maß des sozioökonomischen Hintergrundes und der schulischen Ressourcen statistisch signifikant wird oder das berichtete Ergebnis ändert. Der Befund des Zentralprüfungseffektes erweist sich auch als robust in Bezug auf die Berücksichtigung von Indikatoren für Stadtstaaten und Ostbundesländer, Migrationshintergrund, politisches Wählerverhalten und weiterer denkbarer Einflussfaktoren (vgl. Wößmann 2007b).

Tab. 3: **Zentralprüfungen und Schülerleistungen im Bundesländer- und internationalen Vergleich**

|   | Bundesländer | OECD-Länder | Bundesländer und OECD-Länder | Bundesländer und OECD-Länder | Bundesländer und alle PISA-Teilnehmerländer |
|---|--------------|-------------|------------------------------|------------------------------|---|
| Zentralprüfungen  | 17,9 (3,7)   | 22,5 (10,6) | 22,5 (8,9)                   | 21,7 (6,0)                   | 20,0 (7,4)                                  |
| Deutschland   |              |             | -11,3 (124,6)                |                              |   |
| Zentralprüfungen x Deutschland  |              |             | -4,6 (13,5)                  |                              |   |
| Anzahl Länder   | 16           | 28          | 44                           | 44                           | 54  |
| Abhängige Variable: PISA-2003-Mathematik-Testergebnis. – Kontrollvariablen: ESCS-Index des wirtschaftlichen, sozialen und kulturellen Status, Anteil der Schüler mit vollzeitbeschäftigtem Vater (sowie in der dritten Spalte deren Interaktion mit dem Deutschland-Indikator). – Standardfehler in Klammern. (Quelle: Basierend auf Wößmann 2007b) |              |             |                              |                              |   |

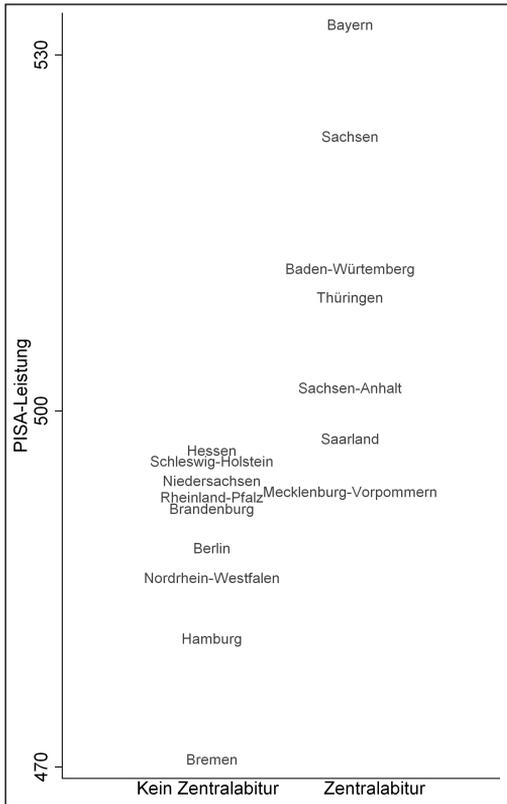


Abb. 1: Zentralprüfungen und Schülerleistungen im Bundesländervergleich  
(Quelle: Eigene Darstellung anhand der PISA-2003-E-Daten)

Der gleiche, quantitativ noch leicht größere Effekt der zentralen Abschlussprüfungen ergibt sich im identischen Modellrahmen auch für die 28 teilnehmenden OECD-Länder (unter Nicht-Berücksichtigung Deutschlands; vgl. die zweite Spalte von Tabelle 3). Die besondere Analysemöglichkeit einer gemeinsamen Berücksichtigung von Bundesländern und OECD-Staaten in einem Modell liegt darin, dass sich statistisch testen lässt, ob der geschätzte Zentralprüfungseffekt sich zwischen der Gruppe der deutschen Bundesländer und der Gruppe der OECD-Staaten unterscheidet. Wie die in der dritten Spalte von Tabelle 3 berichteten Ergebnisse zeigen, ist der geschätzte Unterschied zwischen den beiden Samples statistisch nicht signifikant (weder der schwach negative Interaktionseffekt zwischen Zentralprüfungen und einem Indikator für die deutschen Bundesländer noch der Deutschlandindikator selbst sind statistisch signifikant von null unterschiedlich). Somit lässt sich der Zentralprüfungseffekt einheitlich im gemeinsamen Sample mit einer Größenordnung von 21,7 Prozent einer internationalen Standardabweichung schätzen. Weitet man das Sample auf alle PISA-2003-Teilnehmerstaaten aus, dann beläuft sich der geschätzte Effekt auf 20,0 Prozent einer internationalen Standardabweichung unter den 54 gemeinsam betrachteten Bundesländern und Teilnehmerstaaten. Der Zentralprüfungseffekt ergibt sich also im Vergleich der deutschen Bundesländer genauso wie im internationalen Vergleich.

## 6. Zentralprüfungen und Schulautonomie

Als letzter Aspekt soll in diesem Abschnitt zunächst theoretisch und dann empirisch untersucht werden, inwiefern eine Komplementarität zentraler Abschlussprüfungen mit einer gesteigerten Selbstständigkeit der Schulen besteht.

### 6.1 Theorie

Schulautonomie oder die Dezentralisierung von Entscheidungsmacht kann als Form der im theoretischen Rahmen von Abschnitt 2 beschriebenen Delegation einer Aufgabe von einem Prinzipal, der die Wissenserzeugung im Schulsystem erreichen will, auf Agenten, nämlich die Schulen, verstanden werden. Wenn dabei unterschiedliche Interessen und asymmetrische Information zusammenkommen, bestehen für die Agenten Anreize und Möglichkeiten zu opportunistischem Verhalten, ohne dass ein solches Verhalten festgestellt und sanktioniert werden könnte. Durch die Bereitstellung von Informationen über die erzielten Leistungen können Zentralprüfungen die Informationsasymmetrie zumindest teilweise überwinden und opportunistisches Verhalten verhindern, wodurch positive Effekte von Schulautonomie aufgrund lokaler Wissensvorsprünge dezentraler Entscheidungsträger überwiegen können.

Abbildung 2 stellt die entsprechenden Leistungswirkungen von Schulautonomie für verschiedene Entscheidungsbereiche dar, die sich durch das Bestehen oder Nichtbestehen von Anreizen für opportunistisches Verhalten und durch das Bestehen oder Nichtbestehen von lokalen Wissensvorsprüngen charakterisieren lassen. Für Entscheidungsbereiche, in denen keine Anreize für opportunistisches Verhalten bestehen, da keine Interessendivergenz zwischen Agenten und Prinzipal besteht, hängt die Leistungswirkung von Schulautonomie nicht von Zentralprüfungen ab (vgl. Wößmann 2005 für Details).

In Entscheidungsbereichen, in denen aufgrund divergierender Interessen Anreize für opportunistisches Verhalten bestehen, in denen es aber keinen lokalen Wissensvorsprung und deshalb keine Vorteile dezentraler Entscheidungsfindung gibt, muss zwischen Systemen mit und ohne Zentralprüfungen unterschieden werden. Gibt es in dem Schulsystem keine Zentralprüfungen, so wirkt sich Schulautonomie in diesen Entscheidungsbereichen negativ aus, weil die dezentrale Entscheidung zu opportunistischem Handeln führt. Gibt es demgegenüber Zentralprüfungen, entstehen keine negativen Auswirkungen lokalen opportunistischen Handelns und damit keine Leistungsunterschiede zwischen Schulautonomie und zentraler Entscheidung.

In Entscheidungsbereichen, die sowohl durch Anreize für opportunistisches Verhalten als auch durch lokale Wissensvorteile gekennzeichnet sind, werden durch Zentralprüfungen wiederum die Nachteile opportunistischen Verhaltens aufgehoben, so dass die lokalen Wissensvorteile insgesamt zu einem positiven Leistungseffekt der Schulautonomie führen können. Sind Zentralprüfungen aber nicht gegeben, so stehen den lokalen Wissensvorteilen die Nachteile opportunistischen Verhaltens gegenüber, und für den Gesamteffekt der Schulautonomie kommt es auf die relative Größe dieser beiden

Teileffekte an. Empirische Ergebnisse (vgl. Wößmann 2005) legen nahe, dass der negative Opportunismuseffekt den positiven Wissenseffekt häufig zu übersteigen scheint. Dann drehen zentrale Prüfungen einen ursprünglich negativen Leistungseffekt von Schulautonomie gänzlich in einen positiven Effekt um.

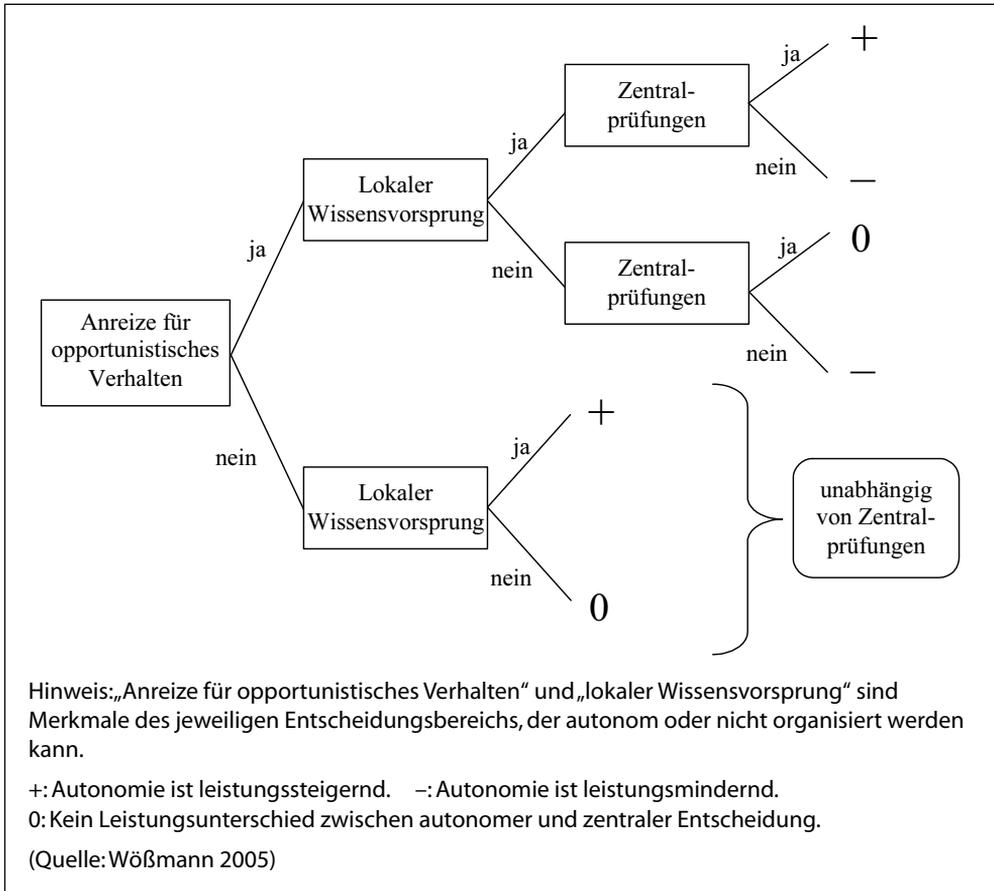


Abb. 2: Auswirkungen von Autonomie auf Schülerleistungen mit und ohne externe Prüfungen

### 6.2 Evidenz

Die Frage, ob die Existenz von Zentralprüfungen den Einfluss von Schulautonomie auf erzielte Schülerleistungen moderiert, lässt sich durch eine Modellspezifikation untersuchen, die in die Schätzgleichung (1) zusätzlich Interaktionsterme zwischen Zentralprüfungen  $Z$  und Indikatoren der Schulautonomie  $A$  (die zuvor im Vektor der institutionellen Variablen subsumiert waren) einfügt:

$$(2) T_{iksl} = \alpha Z_l + (Z_l A_{ksl}) \beta_1 + A_{ksl} \beta_2 + B_{iksl} \beta + R_{ksl} \gamma + I_{ksl} \delta + a + \varepsilon_{iksl} .$$

Die geschätzten Interaktionseffekte geben an, ob sich der Einfluss von Schulautonomie in verschiedenen Entscheidungsbereichen zwischen Systemen mit und ohne Zentralprüfungen unterscheidet.

Schätzungen von Spezifikation (2) anhand der vier genannten internationalen Schülertests bestätigen die theoretischen Überlegungen. Die Untersuchungsergebnisse zeigen nicht nur, dass externe Abschlussprüfungen die Bildungsleistungen verbessern, sondern zugleich auch, dass Schulautonomie in Systemen mit externen Prüfungen eher die erzielten Ergebnisse steigert (vgl. Wößmann 2005; Fuchs/Wößmann 2007; Wößmann u.a. 2007). In mehreren Entscheidungsbereichen verwandeln externe Prüfungen sogar negative Leistungseffekte der Schulautonomie, die sich in Ländern ohne externe Prüfungen ergeben, in positive Effekte.

Ein solcher Entscheidungsbereich, der sowohl durch Anreize für opportunistisches Verhalten als auch durch lokale Wissensvorteile gekennzeichnet sein könnte, könnte beispielsweise die Festsetzung der Lehrergehälter sein. Ohne die durch zentrale Prüfungen eingeführte Verantwortlichkeit könnten sich Schulen opportunistisch verhalten, weil ihr lokales opportunistisches Verhalten nicht extern überwacht und somit auch nicht sanktioniert werden kann. Infolgedessen könnten sich schulische Entscheidungsträger nicht dafür verantwortlich fühlen, über Lehrergehälter in einer Art und Weise zu entscheiden, die zu verbesserten Schülerleistungen beiträgt. Zentrale Prüfungen stellen aber Informationen darüber bereit, ob die Schulen gute oder schlechte Leistungen hervorbringen, so dass Eltern und Schulbehörden mögliche Konsequenzen aus leistungshemmendem Verhalten der Schulen ziehen können. Dadurch haben die Entscheidungsträger in den Schulen Anreize, ihre Autonomie bei der Festsetzung der Lehrergehälter nicht auf opportunistische Weise auszunutzen, sondern sie effektiv zur Förderung der Schülerleistungen einzusetzen. Dadurch können auch die Vorteile des überlegenen lokalen Wissens in Kraft treten, weil die schulischen Entscheidungsträger vermutlich im Allgemeinen viel besser als jede zentrale Behörde wissen, welche Lehrer es verdienen, für ihre Arbeit belohnt zu werden.

Abbildung 3 stellt die entsprechend Gleichung (2) geschätzten Mathematikleistungen der Schüler in TIMSS 1995 und TIMSS 1999 unter den vier Zuständen dar, die aus dem Vorhandensein oder Fehlen von zentralen Abschlussprüfungen und Schulautonomie über Lehrergehälter resultieren: die Schülerleistungen in Schulen ohne Gehaltsautonomie in Systemen ohne zentrale Prüfungen; mit Autonomie, aber ohne zentrale Prüfungen; ohne Autonomie, aber mit zentralen Prüfungen; und sowohl mit Autonomie als auch mit zentralen Prüfungen. Die Leistungen werden relativ zum Zustand mit der geringsten Leistung dargestellt, was in diesem Fall der Zustand mit Gehaltsautonomie und ohne zentrale Prüfungen ist.

Abbildung 3 zeigt, dass Schulautonomie über Lehrergehälter in Systemen ohne zentrale Prüfungen in der Tat einen *negativen* Effekt auf die erzielten Schülerleistungen hat (vgl. Wößmann 2003b). In Systemen mit zentralen Prüfungen sind die Leistungen im Allgemeinen sowohl mit als auch ohne Autonomie höher als in Systemen ohne zentrale Prüfungen. Darüber hinaus kehrt sich aber auch der Effekt der Autonomie in Systemen mit zentralen Prüfungen komplett um: Gehaltsautonomie der Schulen hat dort *positive*

Auswirkungen auf die Schülerleistungen. Es besteht also eine starke Komplementarität zwischen externer Überprüfung und Schulautonomie. Dieselben Effekte von Schulautonomie über Lehrergehälter mit und ohne zentrale Prüfungen finden sich auch in der Analyse der PISA-Daten (vgl. Fuchs/Wößmann 2007).

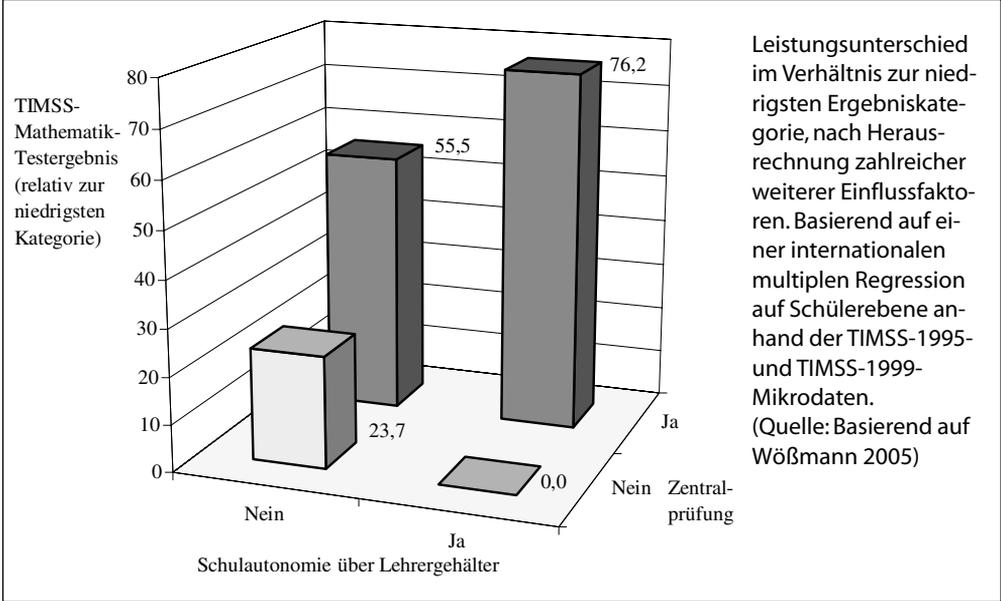


Abb. 3: Zentralprüfungen und Gehaltsautonomie in TIMSS

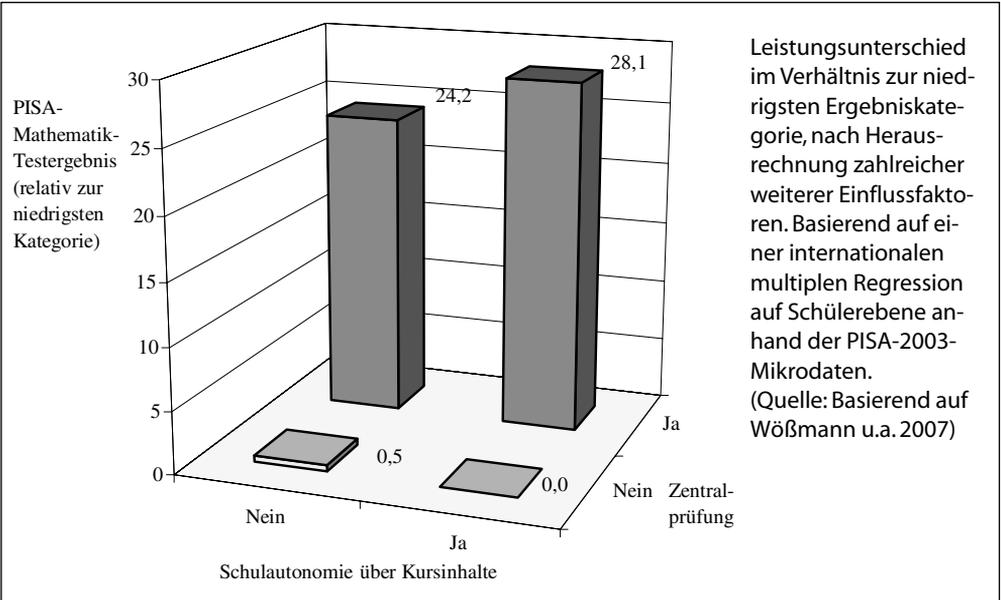


Abb. 4: Zentralprüfungen und Schulautonomie über Kursinhalte in PISA

Abbildung 4 stellt als ein zweites Beispiel einen ähnlichen Zusammenhang für Schulautonomie im Entscheidungsbereich über Kursinhalte anhand der PISA-2003-Daten dar (vgl. Wößmann u.a. 2007). In Schulsystemen ohne zentrale Abschlussprüfungen hat Schulautonomie über Kursinhalte keinen positiven Effekt (in PISA 2000 sogar einen signifikant negativen), in Systemen mit zentralen Abschlussprüfungen hat sie einen positiven Effekt. Ähnliche Fälle, in denen externe Prüfungen einen negativen Autonomieeffekt (aufgrund opportunistischen Verhaltens) in einen positiven Effekt (aufgrund Nutzung lokalen Wissens) umwandeln, konnten etwa auch für den Lehrereinfluss über die Finanzierung der Lehrmittel gezeigt werden (vgl. Wößmann 2003b). Allgemein deutet in etlichen Entscheidungsbereichen die Struktur der Evidenz darauf hin, dass sich Schulautonomie in Kombination mit externen Abschlussprüfungen eher positiv auf die Schülerleistungen auswirkt (vgl. Wößmann 2005).

## 7. Schlussbemerkungen

Im vorliegenden Beitrag wurden zentrale Abschlussprüfungen als „Accountability“-Maßnahme behandelt, die Schüler für ihre erbrachten Bildungsleistungen stärker zur Verantwortung zieht. Über vier internationale Schülerleistungsvergleiche – TIMSS 1995, TIMSS 1999, PISA 2000 und PISA 2003 – hinweg zeigen umfassende Regressionsanalysen der Schülerindividualdaten, dass zentrale Abschlussprüfungen im internationalen Vergleich mit wesentlich besseren Schülerleistungen einhergehen. Dieser Befund erweist sich als robust in Bezug auf einige Tests möglicher Verzerrungen der Querschnittsdatenschätzungen. Der gleiche Zentralprüfungseffekt findet sich in statistisch nicht zu unterscheidender Größenordnung auch im Vergleich der deutschen Bundesländer.

Darüber hinaus belegen die internationalen Ergebnisse, dass zentrale Abschlussprüfungen ansonsten tendenziell eher negative Effekte erhöhter Schulautonomie vielfach in positive Effekte umdrehen (wobei die Effekte von Abschlussprüfungen mit und ohne Autonomie positiv sind). Auch wenn die gewählte methodische Herangehensweise keine gezielte Analyse von Prozesswirkungen ermöglicht, deutet dieser Befund doch darauf hin, dass zentrale Abschlussprüfungen nicht nur die Schüler, sondern auch die Entscheidungsträger in den Schulen zu leistungsfördernderem Verhalten zu bewegen scheinen. Damit erweisen sich zentrale Abschlussprüfungen als durchaus umfassende „Accountability“-Maßnahme.

Insofern ist die Tendenz in der deutschen Schulpolitik, dass immer weitere Bundesländer zentrale Abschlussprüfungen einführen, begrüßenswert. Während bis vor kurzem lediglich sieben der sechzehn Bundesländer ein Zentralabitur (sowie zumeist gleichzeitig auch zentrale Bestandteile in den Abschlussprüfungen der anderen Schularten) durchführten, wird bald nur noch ein einziges Bundesland kein Zentralabitur haben (Rheinland-Pfalz). Jeweils zwei Bundesländer haben in 2005 bzw. 2006 zum ersten Mal ein Zentralabitur durchgeführt (Brandenburg, Hamburg; Hessen, Niedersachsen), drei Bundesländer in 2007 (Berlin, Bremen, Nordrhein-Westfalen) und Schleswig-Holstein wird in 2008 zentrale Bestandteile im Abitur einführen.

Auch die Einführung bundeseinheitlicher Standards führt im Sinne einer einheitlichen Klärung des zu erzielenden Outputs in diese Richtung. Allerdings ist es bisher unterblieben, sich auf eine bundeseinheitliche und flächendeckende bindende regelmäßige Überprüfung der Erfüllung dieser Standards zu einigen. Die hier berichtete internationale Evidenz gibt klare Hinweise für die politische Diskussion über bundeseinheitliche Abschlussprüfungen in Sekundarabschlüssen wie dem Zentralabitur.

## Literatur

- Beaton, A.E./Mullis, I.V.S./Martin, M.O./Gonzalez, E.J./Kelly, D.L./Smith, T.A. (1996): *Mathematics Achievement in the Middle School Years: IEA's Third International Mathematics and Science Study (TIMSS)*. Chestnut Hill, MA: Boston College.
- Bishop, J.H. (1995): *The Impact of Curriculum-Based External Examinations on School Priorities and Student Learning*. In: *International Journal of Educational Research* 23, S. 653–752.
- Bishop, J.H. (1997): *The Effect of National Standards and Curriculum-Based Exams on Achievement*. In: *American Economic Review* 87, H. 2, S. 260–264.
- Bishop, J.H. (2006): *Drinking from the Fountain of Knowledge: Student Incentive to Study and Learn – Externalities, Information Problems, and Peer Pressure*. In: Hanushek, E.A./Welch, F. (Hrsg.): *Handbook of the Economics of Education*. Amsterdam: North-Holland, S. 909–944.
- Bishop, J.H./Moriarty, J.Y./Mane, F. (2000): *Diplomas for Learning, Not Seat Time: The Impacts of New York Regents Examinations*. In: *Economics of Education Review* 19, S. 333–349.
- Bishop, J.H./Wößmann, L. (2003): *Institutional Effects in a Simple Model of Educational Production*. In: *Education Economics* 12, S. 17–38.
- Cohen, J./Baldi, S. (1998): *An Evaluation of the Relative Merits of HLM vs. Robust Linear Regression in Estimating Models with Multi-Level Data*. Mimeo: American Institutes for Research.
- Deaton, A. (1997): *The Analysis of Household Surveys: A Microeconomic Approach to Development Policy*. Baltimore: The Johns Hopkins University Press.
- Fuchs, T./Wößmann, L. (2007): *What Accounts for International Differences in Student Performance? A Re-examination using PISA Data*. In: *Empirical Economics* 32, S. 433–464.
- Jürges, H./Schneider, K./Büchel, F. (2005): *The Effect of Central Exit Examinations on Student Achievement: Quasi-Experimental Evidence from TIMSS Germany*. In: *Journal of the European Economic Association* 3, S. 1134–1155.
- Jürges, H./Richter, W.F./Schneider, K. (2005): *Teacher Quality and Incentives: Theoretical and Empirical Effects of Standards on Teacher Quality*. In: *FinanzArchiv* 61, S. 298–326.
- Laffont, J.-J./Martimort, D. (2002): *The Theory of Incentives: The Principal-Agent Model*. Princeton: Princeton University Press.
- Mauro, P. (1995): *Corruption and Growth*. In: *Quarterly Journal of Economics* 110, S. 681–712.
- Mullis, I.V.S./Martin, M.O./Gonzalez, E.J./Gregory, K.D./Garden, R.A./O'Connor, K.M./Chrostowski, S.J./Smith, T.A. (2000): *TIMSS 1999 International Mathematics Report: Findings from IEA's Repeat of the Third International Mathematics and Science Study at the Eighth Grade*. Chestnut Hill, MA: Boston College.
- Organisation for Economic Co-operation and Development (2001): *Knowledge and Skills for Life: First Results from the OECD Programme for International Student Assessment (PISA) 2000*. Paris: OECD.
- Organisation for Economic Co-operation and Development (2004): *Learning for Tomorrow's World: First Results from PISA 2003*. Paris: OECD.

- Prenzel, M./Baumert, J./Blum, W./Lehmann, R./Leutner, D./Neubrand, M./Pekrun, R./Rost, J./Schiefele, U. (Hrsg.) (2005): PISA 2003: Der zweite Vergleich der Länder in Deutschland – Was wissen und können Jugendliche? Münster: Waxmann.
- Wooldridge, J.M. (2001): Asymptotic Properties of Weighted M-Estimators for Standard Stratified Samples. In: *Econometric Theory* 17, S. 451–470.
- Wößmann, L. (2001): Why Students in Some Countries Do Better: International Evidence on the Importance of Education Policy. In: *Education Next* 1, H. 2, S. 67–74.
- Wößmann, L. (2002): *Schooling and the Quality of Human Capital*. Berlin: Springer.
- Wößmann, L. (2003a): Schooling Resources, Educational Institutions, and Student Performance: The International Evidence. In: *Oxford Bulletin of Economics and Statistics* 65, S. 117–170.
- Wößmann, L. (2003b): Central Exit Exams and Student Achievement: International Evidence. In: Peterson, P.E./West, M.R. (Hrsg.): *No Child Left Behind? The Politics and Practice of School Accountability*. Washington, DC: Brookings Institution Press, S. 292–323.
- Wößmann, L. (2005): The Effect Heterogeneity of Central Exams: Evidence from TIMSS, TIMSS-Repeat and PISA. In: *Education Economics* 13, S. 143–169.
- Wößmann, L. (2007a): International Evidence on School Competition, Autonomy and Accountability: A Review. In: *Peabody Journal of Education* 82, S. 473–497.
- Wößmann, L. (2007b): *Fundamental Determinants of School Efficiency and Equity: German States as a Microcosm for OECD Countries*. Program on Education Policy and Governance Research Paper PEPG 07-02. Cambridge, MA: Harvard University.
- Wößmann, L./Lüdemann, E./Schütz, G./West, M.R. (2007): *School Accountability, Autonomy, Choice, and the Level of Student Achievement: International Evidence from PISA 2003*. OECD Education Working Paper 13. Paris: Organisation for Economic Co-operation and Development (OECD).

**Abstract:** *Within the framework of principal-agent-models, central exit exams can be modelled as measures of accountability which hold students and schools responsible for their educational achievements. Comprehensive regression analyses on the basis of individual student data provided by four international comparative studies on student achievement (TIMSS 1995, TIMSS 1999, PISA 2000, and PISA 2003) show that, compared internationally, central exit exams are associated with substantially higher student performance. The effect of central exit exams is also apparent in a comparison between the German Laender; statistically, the results found here cannot be distinguished from the international estimate. Furthermore, the international results show that central exit exams tend to turn otherwise negative effects of increased school autonomy into positive effects. This, in turn, suggests that central exit exams induce not only students, but also the decision-makers at schools to act and behave in a way that is conducive to learning.*

*Anschrift des Autors:*

Prof. Dr. Ludger Wößmann, Ludwig-Maximilians-Universität und ifo Institut für Wirtschaftsforschung, Poschingerstr. 5, 81679 München, Tel.: 089/9224-1699, Fax: 089/907795-1699.  
E-Mail: woessmann@ifo.de. Internet: www.cesifo.de/woessmann.