

Zentrale Prüfungen: Für eine Verbesserung der Qualität von Bildung fehlen bislang die empirischen Belege

1 Zusammenfassung

Weder aus TIMSS noch aus PISA lässt sich zweifelsfrei ableiten, dass zentrale Prüfungen zu besseren Leistungen führen. Um Vor- und Nachteile zentral gestellter Prüfungsaufgaben wissenschaftlich fundiert bewerten zu können, wären Langzeit-Studien mit repräsentativen Stichproben erforderlich.

Als Beleg für die vorteilhafte Wirkung zentraler Prüfungen werden in neuerer Zeit wiederholt drei Autoren bzw. Autoren-Teams zitiert (Bishop 1995, 1997; Wößmann 2002, 2003 und Büchel/Jürges/Schneider 2003). Im Folgenden wird gezeigt, dass die Arbeiten von Bishop, Wößmann, Bücher u.a. keine verlässlichen empirischen Belege für die Überlegenheit zentraler Abschlussprüfungen enthalten, weil lediglich mit Querschnittsdaten gearbeitet wird, weshalb die langfristige Wirkung wichtiger Schul-Variablen nicht erfasst werden konnte. Außerdem wurden einige bedeutsame Elternhaus-Variablen nicht kontrolliert, weil sie nicht in den TIMSS-Fragebögen enthalten waren.

2 Bishop (1995, 1997): Leistungsanreize durch nationale Standards, IAEP- und TIMSS-Ergebnisse aus Kanada

An den Anfang seiner Veröffentlichungen stellt Bishop theoretische Erörterungen über die Vorteile zentraler Prüfungen (sinngemäß übersetzt heißt es dort z.B.):

.. wenn die Zulassung zur Hochschule sich nur auf den Rang in einer Schulklasse stützt, dann entsteht bei den Schülern ein persönliches Interesse daran, sich gegenseitig davon zu überzeugen, für die Schule nichts zu tun (Bishop 1997, Seite 260¹).

Und

Sofern sich die Noten im Examen aus der Leistungskurve (der Schulklasse) herleiten, so wird das allgemeine Wohlbefinden dann maximiert, wenn niemand mehr etwas für die Schule tut. Im Laufe der Zeit wird über freundschaftliche Zuwendung einerseits und durch Gespött und Schikanen andererseits der kooperative Weg des Nichtstuns bestärkt (Bishop 1997, Seite 261²).

¹ „... basing college admissions on class rank gives students a personal interest in persuading each other not to study.“

² „When exams are graded on a curve, joint welfare is maximized when no one studies. In the repeated game that results, side payments (friendship and respect) and punishments (ridicule and harassment) enforce the cooperative »don't study« solution.“

Wären diese Aussagen zutreffend, dann müssten Länder, in denen es seit Jahrzehnten keine zentralen Prüfungen gibt, in vergleichenden Tests zunehmend schlechter abschneiden. TIMSS-III, deren Tests am Ende der Sekundarstufe II eingesetzt wurden, bestätigt die Einschätzung nicht, dass das Fehlen zentraler Prüfungen im Laufe der Schulzeit zu immer schlechteren Leistungen führt:

„Es sind keine konsistenten Zusammenhänge zwischen Organisationsform der Abiturprüfung und Fachleistungen zu erkennen.“ (Baumert u.a. 1998, S. 119)

„Auch im internationalen Vergleich lässt sich kein systematischer Zusammenhang zwischen Organisationsform der Abschlussprüfung (dezentral versus zentral) und Fachleistungen nachweisen. Die Schweiz und Frankreich sind Prototypen beider Organisationsformen. Beide Länder zeichnen sich gleichermaßen durch exzellente Leistungen (in den bei TIMSS-III eingesetzten Tests) aus.“ (Baumert u.a. 1998, S. 119)

Die Aussage, dass von zentralen Prüfungen „keine konsistenten“ Effekte ausgehen, wird durch die Testergebnisse bei PISA bestätigt. - In dem Schulleiter-Fragebogen, der bei PISA-2000 eingesetzt worden war, war die Frage enthalten, ob die jeweilige Schule an überregionalen Leistungsvergleichen teilnimmt (Variable SC18Q04). Diese Variable hat in den meisten Staaten keinen signifikanten Einfluss auf die Test-Ergebnisse bei PISA. In einigen Staaten, wo signifikante Effekte auftreten (z.B. in England und in den USA), schneiden die Schulen, die sich an überregionalen Leistungsvergleichen beteiligen, im Textverständnis-Test bei PISA-2000 im Mittel schlechter ab als die Schulen, wo die Leistungen der 15-Jährigen nicht überregional verglichen werden (Hagemeister 2005).

Aus den Daten der OECD-Studien IAEP und TIMSS-II (die 1991 bzw. 1995 durchgeführt wurden) leitet Bishop die Erkenntnis ab, dass die 8-Klässler aus den kanadischen Provinzen, in denen zentrale Abschlussprüfungen stattfinden, bessere Testergebnisse erzielen als die Testteilnehmer aus den anderen kanadischen Provinzen. Allerdings gibt es in den kanadischen Provinzen, in denen zentrale Prüfungen stattfinden, im Mittel die besser ausgebildeten Mathematiklehrer. Ferner sind dort die naturwissenschaftlichen Labore besser ausgestattet. Ein weiteres Merkmal der kanadischen Bezirke mit zentralen Prüfungen ist z.B., dass die Schüler zu Hause weniger Zeit mit Fernsehen verbringen (Bishop 1997, Seite 262, Tabelle 1 und Seite 263, Tabelle 2). Die Schülergruppen, deren Testergebnisse bei Bishop einander gegenübergestellt wurden, waren also nicht vergleichbar.

Bishop kontrolliert zwar unter Verwendung der IAEP-Daten den Einfluss verschiedener Variablen (Bishop 1995, Seite 78 bis 80 und Tabelle 7 auf Seite 81). Kontrolliert wurde z.B. die Ausstattung der Schulen mit gut ausgebildeten Lehrkräften und die Zahl der Bücher im Elternhaus. Allerdings wurden wichtige Einflüsse nicht erfasst (z.B. der Beruf der Eltern oder das Herkunftsland von Immigranten, siehe: TIMSS Fragebögen 1999). Außerdem wurden bei der IAEP-Studie (wie auch bei TIMSS und PISA) bisher nur Querschnittsdaten erhoben. Die Aussage, dass eine bestimmte 8. Schulklasse, die an dem IAEP-Test teilgenommen hat, von einem gut aus-

gebildeten Mathematiklehrer unterrichtet wird, stellt deshalb lediglich eine Momentaufnahme dar. Ob dieselben Schüler bis zum Ende der 7. Klasse womöglich nur fachfremd unterrichtet wurden, ist hier unbekannt. Daraus folgt, dass der Einfluss von Schulvariablen mit Hilfe von Querschnittsdaten nicht verlässlich kontrolliert werden kann. Unverzichtbar sind dazu Längsschnitt-Studien, denn einige sehr wichtige Schulvariablen sind vor allem langfristig wirksam. So ist z.B. die positive Wirkung geringer Klassenfrequenzen erst nach mehreren Jahren deutlich nachweisbar (siehe hierzu z.B.: Pate-Bain u.a. 1997).

Bishop schreibt den zentralen Prüfungen vielfältige Wirkungen zu. So stuft er den Fernsehkonsum zu Hause als „abhängige“ Variable ein (Bishop 1997, Seite 263, Tabelle 2). Daraus folgert er, dass 8-Klässler zu Hause weniger Zeit vor dem Fernseher verbringen, wenn am Ende der Schulzeit eine zentrale Prüfung auf sie wartet. Eine abhängige Variable ist bei Bishop auch die Ausstattung der Schule mit gut ausgebildeten Fachlehrern: Hier hätte man fragen müssen, ob die Zentral-Abitur-Provinzen in Kanada womöglich die wohlhabenderen sind, so dass sie deshalb insgesamt mehr gutausgebildete Fachlehrer einstellen können, oder ob nur für die zentral-geprüften Fächer die besser ausgebildeten Lehrer engagiert wurden. Dies hätte dann zur Folge, dass die anderen Fächer (etwa Kunst oder Musik) vernachlässigt würden, was man ohne Zweifel als negativen Effekt der zentralen Prüfung bewertet müsste.

Würden sich zentrale Prüfungen derart umfassend vorteilhaft auswirken, wie Bishop dies annimmt, dann hätten die kanadischen Schulen, die an überregionalen Leistungsvergleichen teilnehmen, auch bei PISA ganz eindeutig bessere Testergebnisse erzielen müssen. Weder bei PISA-2000 noch PISA-2003 führt jedoch der Einsatz zentraler Leistungsvergleiche an kanadischen Schulen zu besseren Resultaten (Hagemeister 2006a).

3 Wößmann (2002, 2003): Zentrale Prüfungen als Währung, Testergebnisse bei TIMSS im internationalen Vergleich

Ähnlich wie Bishop (1995, 1997) stellt auch Wößmann (2002, 2003) an den Anfang seiner Veröffentlichungen theoretische Betrachtungen, die darauf hinauslaufen, dass vor allem die extrinsische Motivation, die von zentralen Prüfungen ausgeht, zu guter Schulbildung führt. Diese einseitige Sicht entspricht glücklicherweise nicht der Realität. Offensichtlich gelingt es vielen Lehrkräften in Deutschland und in anderen Ländern, die intrinsische Motivation der Schülerinnen und Schüler so zu wecken, dass sie sich - ohne dass von zentralen Prüfungen Druck ausgeht - intensiv mit dem Unterrichtsstoff befassen. Deshalb führen zentrale Leistungsvergleiche bei TIMSS-III und bei PISA zu keinen konsistenten Effekten (wie besprochen).

„Um die theoretisch abgeleiteten Hypothesen empirisch testen zu können“ benutzt Wößmann die TIMSS-II-Daten, die im Jahre 1995 erhoben wurden, und TIMSS-Repeat-Daten aus dem Jahre 1999 (Wößmann 2003, Seite 226). Er vergleicht die Testergebnisse aller Staaten mit und ohne zentrale Prüfungen und kommt zu dem Ergebnis „dass Schüler in Schulsystemen mit zentralen Prüfungen statistisch signifikant besser abschneiden als Schüler

in Schulsystemen ohne zentrale Prüfungen“ (Wößmann 2003, Seite 228). Dieses Ergebnis ist nicht überraschend, denn in der Staatengruppe ohne zentrale Prüfungen ist der prozentuale Anteil der Entwicklungsländer, die bei TIMSS relativ schlecht abgeschnitten haben, deutlich höher als in der Vergleichsgruppe (Wößmann 2002, Seite 15, Tabelle 2). Die Ergebnisse Wößmanns hängen also von der zufälligen Zusammensetzung der betrachteten Staatengruppen ab.³ Es ist nicht erkennbar, welchen Nutzen der Vergleich derart unterschiedlicher, zufällig zusammengestellter Staaten hat. Wenn internationale Vergleiche von Schulsystemen überhaupt sinnvoll sein sollen, dann müssen Staaten mit ähnlichen Lebensbedingungen einander gegenübergestellt werden. Betrachtet man z.B. nur die wohlhabenden westeuropäischen Staaten, dann zeigt sich, dass bei TIMSS die Staaten ohne zentrale Prüfungen im Mittel deutlich besser abgeschnitten haben als die Staaten, in denen zentrale Prüfungen stattfinden. Die europäische Region, die bei TIMSS im Mittel die besten Ergebnisse erzielt hat, ist der flämische Teil Belgiens, wo es keine zentralen Examina gibt:

³ Zwar kontrolliert Wößmann (2002) eine Reihe von Variablen, allerdings gelten hier die gleichen Einschränkungen wie bei Bishop: Wichtige Variablen wurden nicht erfasst und es handelt sich wiederum lediglich um Querschnittsdaten. Deshalb kann der Einfluss vieler Variablen nur unzureichend kontrolliert werden.

Die TIMSS-II-Ergebnisse der wohlhabenden⁴
west-europäischen Industriestaaten im Mathematik-Test

Country	TIMSS-II-1995 ⁵				TIMSS-II-Repeat-1999 ⁶	
	7. Klasse zentrale Examen		8. Klasse zentrale Examen		8. Klasse zentrale Examen	
	ja	nein	ja	nein	ja	nein
Belgien (Fläm)		558		565		557
Belgien (Fr.)		507		527		
Dänemark	465		502			
Deutschland		(485)		(509) ⁷		
England	477		506		496	
Finnland					520	
Frankreich	492		538			
Irland	500		527			
Island	459		487			
Italien					479	
Niederlande	516		541		538	
Norwegen	461		503			
Österreich		509		539		
Schottland	463		499			
Schweden		477		519		
Schweiz		506		545		
Mittelwerte⁸:	479,1	511,4	512,9	539,6	508,3	557

⁴ BIP pro Einwohner über 30.000 Dollar im Jahre 2005

⁵ Aus: TIMSS (1996), S. 22, Tab. 1.1 und S. 26, Tab. 1.2

⁶ Aus: Wößmann 2002, Seite 15, Tabelle 2

⁷ Die Deutschen Ergebnisse wurden hier bei der Berechnung der Mittelwerte nicht berücksichtigt, weil mir die Zahlen, getrennt nach Bundesländern mit und ohne Zentralabitur, nicht zur Verfügung standen. Dies würde in den Spalten 2 bis 5 zu einer leichten Verschiebung der Mittelwerte zugunsten der Staaten mit zentralen Prüfungen führen. Trotzdem würden ihre Mittelwerte die der Staaten ohne Zentrale Examina nicht übertreffen.

⁸ Die Mittelwerte wurden hier ohne Gewichtung berechnet. Es sind unterschiedliche Gewichtungen denkbar, die jeweils zu anderen Mittelwerten führen. Bei PISA präsentiert die OECD zwei verschiedene Wege der Mittelwertberechnung (OECD 2004, S. 33):

OECD Average - (country average) - mean data for all OECD countries - each country contributes equally to the average.

OECD Total - (OECD as single entity) - each country contributes in proportion to the number of 15-year-olds enrolled in its schools

4 Büchel, Jürges und Schneider (2003): Die TIMSS-Ergebnisse in Deutschen Bundesländern mit und ohne zentrale Prüfungen am Ende der Sek. I

In einer Veröffentlichung von Büchel, Jürges und Schneider vom Deutschen Instituts für Wirtschaftsforschung (DIW Berlin) wird gezeigt, dass die in Deutschland getesteten 8-Klässler aus Haupt- und Realschulen bei TIMSS-II besser abgeschnitten haben, wenn beim mittleren Schulabschluss zentral-gestellte Aufgaben zum Einsatz kommen (Büchel u.a. 2003, S. 238-251). In dieser DIW-Studie wird allerdings auf wichtige Einschränkungen hingewiesen: In den Bundesländern, in denen keine zentralen Prüfungen am Ende der Sekundarstufe I stattgefunden haben, gehen „weit mehr Kinder von Immigranten“ zur Schule und außerdem ist die „Anzahl der Mathematikstunden erheblich niedriger“. Ferner „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“ (Büchel u.a. 2003, S. 243). Man muss deshalb damit rechnen, dass die besseren Ergebnisse bei TIMSS-II, die die Haupt- und Realschüler aus Bundesländern mit zentral-gestellten Aufgaben für 10-Klässler erzielt haben, nicht durch diese Prüfungsform, sondern durch andere Einflüsse (z.B. durch mehr Mathematikstunden oder den geringeren Anteil an Schülern mit Migrationshintergrund) verursacht worden sind.

Wie sehr die Unterschiedlichkeit der Schülergruppen in Ländern mit und ohne zentrale Prüfungen am Ende der Sek. I die TIMSS-Ergebnisse beeinflusst hat, lässt sich mit Hilfe des vorliegenden Datenmaterials nicht kontrollieren, da es sich (wie schon angesprochen) um Querschnitts-Daten handelt und da wichtige Einflüsse bei TIMSS nicht erfasst wurden. Außerdem muss man damit rechnen, dass mehrere Variablen (z.B. der soziale Status oder der Prozentsatz der Schüler, die die Hauptschule besuchen) nicht unabhängig von einander wirksam sind. Die Frage, wie lange verschiedene Einflüsse wirksam waren, wird damit umso bedeutsamer. Deshalb sind Längsschnittstudien notwendig, um die Wirkung zentraler Prüfungen auf den Erfolg von Unterricht zu erfassen.

5 Pädagogische Bewertung zentrale gestellter Prüfungsaufgaben

Gegen die Einführung zentraler Prüfungen lassen sich aus Sicht der Pädagogik gewichtige Einwände vorbringen:

Ein Nachteil zentraler Prüfungen ist, dass das Arbeiten auf eine punktuelle Prüfung hin in der Schule ein zu starkes Gewicht erhält, womit das Einüben sinnvoller Arbeitsstrategien vernachlässigt wird. Arbeitsstrategien, die im Nachdenken über Unterrichtsinhalte, im eigenständigen Wiederholen und Vertiefen und im Beschaffen ergänzender Informationen bestehen, sind für die Studierfähigkeit von sehr viel größerer Bedeutung als das Abschneiden in einem punktuellen schriftlichen Examen.

Zentral gestellte Prüfungsaufgaben, die in die Abschlussnoten der Schüler eingehen, bringen nicht mehr, sondern weniger Prüfungsgerechtigkeit, denn bei zentraler Aufgabenstellung wird die Bewertung der Schüler mit einer Bewertung der Lehrer und der Schule verknüpft. Insbesondere werde die Schüler benachteiligt, die unter ungünstigen Schulbedingungen zu leiden hatten.

Aufgaben für zentrale Prüfungen müssen stets im Geheimen und unter Einhaltung feststehender Fristen erstellt werden. Deshalb können solche Aufgaben nicht nach bewährten wissenschaftlichen Verfahren erprobt und weiterentwickelt werden, was zur Folge hat, dass zentral-gestellte Prüfungsaufgaben . Dieses gravierende Problem besteht nicht, wenn Tests, die überregional eingesetzt werden, nur der Diagnose dienen (siehe z.B.: Hagemeister 2006b).

6 Literaturverzeichnis

Baumert, Jürgen / Bos, Wilfried / Watermann, Rainer (1998): TIMSS/III – Schülerleistungen in Mathematik und den Naturwissenschaften am Ende der Sekundarstufe II im internationalen Vergleich - Zusammenfassung der deskriptiven Befunde. Berlin: Max-Planck-Institut für Bildungsforschung.
<http://edoc.mpg.de/get.epl?fid=16094&did=195148&ver=0>

Bishop, John H. (1995): The Impact of Curriculum-Based External Examinations on School Priorities and Student Learning. Working Paper 95 – 30. Ithaca, New York, USA: Cornell University.
<http://digitalcommons.ilr.cornell.edu/cgi/viewcontent.cgi?article=1219&context=cahrswp>

Bishop, John H. (1997): The Effect of National Standards and Curriculum-Based Exams on Achievement. In: American Economic Review 87, H. 2, S. 260-264.

Büchel, Felix / Jürges, Hendrik / Schneider Kerstin (2003): Die Auswirkungen zentraler Abschlussprüfungen auf die Schulleistung ... Befunde aus der deutschen TIMSS-Stichprobe. In: Vierteljahrshefte zur Wirtschaftsforschung 72, H. 2, S. 238-251.
http://www.diw.de/deutsch/produkte/publikationen/vierteljahrshefte/docs/papers/v_03_2_6.pdf

Hagemeister, Volker (2005), die Tabelle: PISA-2000: Ergebnisse im Test zum Textverständnis, in Abhängigkeit davon, ob sich die Schule an überregionalen Leistungsvergleichen beteiligt. Berlin.
<http://volker.hagemeister.name/fileadmin/volker/Ueberregionale-Leistungs-Vergleiche-Lesen-2000.pdf>

Hagemeister, Volker (2006), die Tabelle: PISA-2000 und -2003: Der Einfluss überregionaler Leistungsvergleiche auf die Testergebnisse in Kanada. Berlin.
<http://volker.hagemeister.name/fileadmin/volker/Ueberregionale-Leistungs-Vergl-in-Kanada--PISA-2000-und-2003.pdf>

Hagemeister, Volker (2007): Nachteile zentraler Prüfungen: Argumente und empirische Belege. Berlin

<http://volker.hagemeister.name/fileadmin/volker/Nachteile-zentraler-Pruefungen.pdf>

IAEP-Study (1995): International Assessment of Educational Progress. In: International Comparative Studies in Education: Description of Selected Large-Scale Assessments and Case Studies. Washington: The National Academy of Sciences. S 1.

http://print.nap.edu/pdf/NI000034/pdf_image/48.pdf

OECD (2004): Learning for Tomorrow's World, First Results from PISA 2003. Paris: OECD.

<http://a455.g.akamai.net/7/455/1879/v1/193.51.65.71/dataoecd/1/60/34002216.pdf>

Pate-Bain, Helen / Boyd-Zaharias, Jane / Cain, Van A. / Word, Elizabeth / Binkley, M. Edward (1997): The student/teacher achievement ratio (STAR) project: STAR follow-up studies. Lebanon, Tennessee, USA: Health & Education Research Operative Services (HEROS). <http://www.heros-inc.org/newstar.pdf>

TIMSS (1996): Mathematics Achievement in the middle School Years: IEA's Third International Mathematics and Science Study (TIMSS). Chestnut Hill, Massachusetts, USA: Center for the Study of Testing, Evaluation, and Educational Policy, Boston College.

<http://timss.bc.edu/timss1995i/TIMSSPDF/BMathAll.pdf>

TIMSS Fragebögen (1999): Alle bei TIMSS eingesetzten Fragebögen findet man unter <http://timss.bc.edu/timss1999i/questionnaires.html>

Wößmann, Ludger (2002): Central Exams Improve Educational Performance: International Evidence. Kieler Diskussionsbeiträge 397. Kiel: Institut für Weltwirtschaft.

<http://opus.zbw-kiel.de/volltexte/2006/3863/pdf/kd397.pdf>

Wößmann, Ludger (2003): Zentrale Prüfungen als „Währung“ des Bildungssystems: Zur Komplementarität von Schulautonomie und Zentralprüfungen. In: Vierteljahrshefte zur Wirtschaftsforschung 72, H. 2, Seite 220–237,

http://www.diw.de/deutsch/produkte/publikationen/vierteljahrshefte/docs/papers/v_03_2_5.pdf