

Nachteile zentraler Prüfungen: Argumente und empirische Belege

1 Zusammenfassung

Es gibt zur Zeit keine wissenschaftlich fundierten Belege dafür, dass zentrale Prüfungen zu besserer Schulbildung führen. Um Vor- und Nachteile zentral gestellter Prüfungsaufgaben wissenschaftlich fundiert bewerten zu können, wären Langzeit-Studien mit repräsentativen Schülergruppen erforderlich. Obwohl in den meisten deutschen Bundesländern die Einführung zentraler Prüfungen unmittelbar bevorsteht, fehlt also die rationale Grundlage für diese Art der Bildungsreform. Dabei lassen sich gegen die Einführung zentraler Prüfungen gewichtige Einwände vorbringen:

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 in nicht kontrollierbarer Weise mit einer Bewertung der Lehrer und der Schule verknüpft. Insbesondere werde die Schüler benachteiligt, die unter ungünstigen Schulbedingungen zu leiden hatten. Deshalb sollte man zentralgestellte Tests nur zu Diagnosezwecke einsetzen. Anstatt jungen Menschen am Ende der Schulzeit mit einer zentralen Prüfung zu bescheinigen, dass ihre Schulbedingungen in Berlin-Kreuzberg nicht so günstig waren wie in Berlin-Zehlendorf, sollten leistungsfähige Diagnose-Instrumente erprobt und entwickelt werden, die frühzeitig ein differenziertes Bild möglicher Defizite liefern.

Ein Nachteil zentraler Prüfungen ist ferner, 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.

2 Für eine Verbesserung der Qualität von Bildung fehlen die empirischen Belege

Zur Zeit ist die Meinung weit verbreitet, dass das Fehlen überregionaler Prüfungen eine wesentliche Ursache für das schlechte Abschneiden Deutschlands bei PISA gewesen sei. Dies hält man vielfach schon deshalb für plausibel, weil es in den süddeutschen Ländern, die bei PISA relativ gut abgeschnitten haben, seit langem zentrale Abschlussprüfungen gibt. Auch in Finnland sind zentral gestellte Aufgaben Bestandteil des Abiturs. Allerdings ist die reglementierende Wirkung, die vom Zentralabitur in Finnland ausgeht, vergleichsweise gering, denn im Abitur steht ein breites Angebot an Aufgaben zur Wahl. Außerdem hat man in Finnland die Möglichkeit, Prüfungen, die man bereits bestanden hat, zu wiederholen. Ferner wird die Zulassung zum Studium an allen finnischen Universitäten durch eigene Aufnah-

metests und Vorstellungsgespräche geregelt, was die Bedeutung des Abiturzeugnisses zusätzlich mindert (Abitur in Finnland, 2003).

Sowohl bei PISA wie auch bei der Vorgänger-Studie TIMSS (1996) findet man Staaten (bzw. Regionen), die gut abgeschnitten haben, obwohl es dort keine zentralen Prüfungen gibt (siehe hierzu z. B. [1]). - In dem Schulleiter-Fragebogen, der bei PISA eingesetzt worden war, war die Frage enthalten, ob die jeweilige Schule an überregionalen Leistungsvergleichen teilnimmt (bei PISA-2000 die Variable SC18Q04). Diese Variable hat in den meisten Staaten keinen signifikanten Einfluss auf die Ergebnisse im Textverständnis-Test. 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 Mittel schlechter ab als die Schulen, wo die Leistungen der 15-Jährigen nicht überregional verglichen werden.²

In einer Veröffentlichung des Deutschen Instituts für Wirtschaftsforschung (DIW Berlin) wird nachgewiesen, 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/Jürges/Schneider 2003). 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). In der DIW-Studie werden also die Testergebnisse unterschiedlich zusammengesetzter Schülergruppen miteinander verglichen. 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 aus dem vorliegenden Datenmaterial nicht ableiten. Einerseits wurden bei TIMSS einige wichtige Merkmale nicht erfasst (z. B. die berufliche Stellung der Eltern und die kulturellen Aktivitäten im Elternhaus, siehe: TIMSS Fragebögen 1999) Ferner wurden bei TIMSS nur Querschnittdaten (also nur Momentaufnahmen) erhoben. Viele Schulvariablen sind jedoch vor allem langfristig wirksam. So ist z. B. die positive Wirkung geringer Klassenfrequenzen erst nach einigen Jahren deutlich nachweisbar (siehe hierzu z. B.: Pate-Bain 1997). Um Vor- und Nachteile zentral gestellter Prüfungsaufgaben wissenschaftlich fundiert bewerten zu können, wären also Langzeit-Studien mit repräsentativen Schülergruppen erforderlich.

Auch bei TIMSS-III, deren Tests am Ende der Schulzeit eingesetzt wurden, „sind keine konsistenten Zusammenhänge zwischen Organisationsform der Abiturprüfung und Fachleistungen zu erkennen“ (Baumert u.a. 1998, S. 119). Für die Erwartung, dass mit der Einführung zentraler Prüfungen

¹ <http://volker.hagemeister.name/fileadmin/volker/Zentralexamen-und-Testergebnisse-bei-TIMSS.pdf>

² <http://volker.hagemeister.name/fileadmin/volker/Ueberregionale-Leistungsvergleiche-Lesen-2000.pdf>

das Niveau des Schulunterrichts steigt, gibt es also zur Zeit keine empirischen Belege (siehe hierzu auch: Hagemeister 2007a).

3 Zentrale Prüfungen bringen nicht mehr, sondern weniger Prüfungsgerechtigkeit

In neuerer Zeit wird immer wieder beklagt, dass PISA und andere überregionale Tests gezeigt hätten, dass die gleichen Leistungen, für die man in der einen Schule eine „Zwei“ erhält, in einer anderen Schule allenfalls mit „Vier“ benotet werden. Hier wird eine Gerechtigkeitslücke gesehen, die nach vorherrschender Meinung mit zentralen Prüfungen und landesweiten Vergleichsarbeiten beseitigt werden muss. Auch diese Erwartung werden zentrale Prüfungen, die Bestandteil der Notengebung sind, nicht erfüllen:

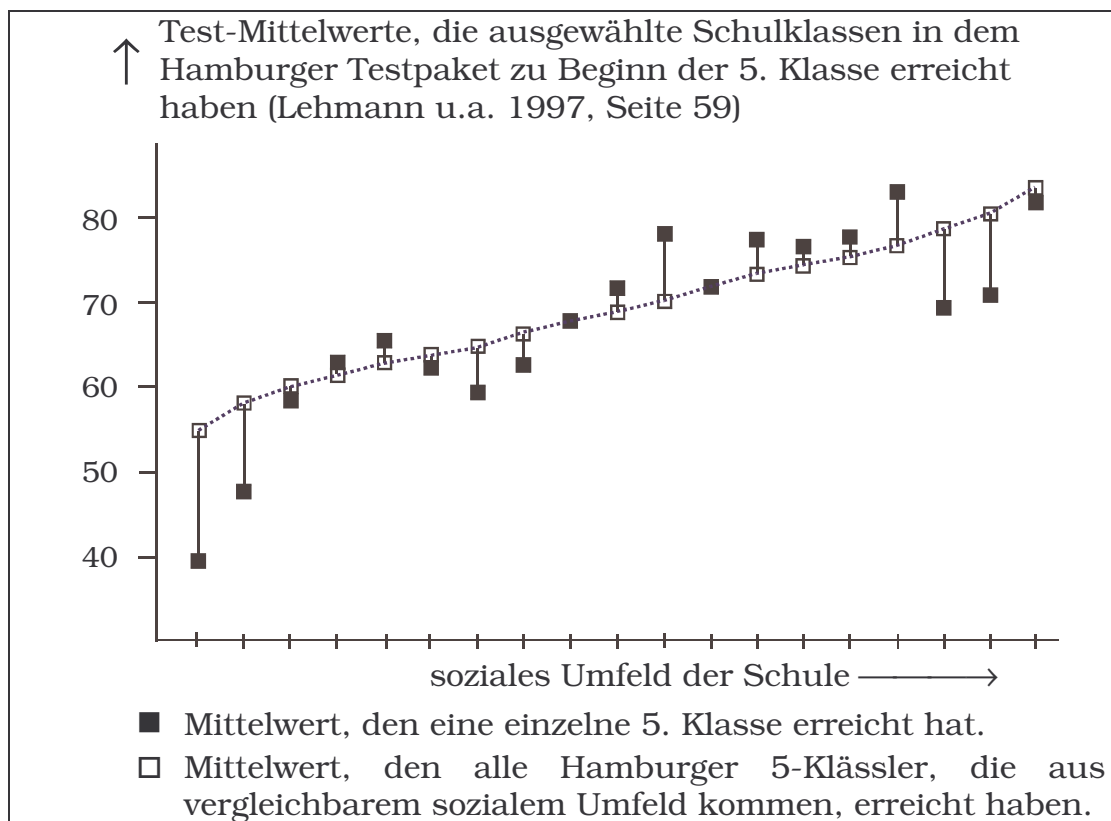
Die Erfahrungen aus Ländern, wo seit Jahrzehnten Selektion mit Hilfe überregional eingesetzter Tests betrieben wird, zeigen, dass Kinder wohlhabender Eltern in immer größerer Zahl am Nachmittag private Schulen besuchen, wo das Lösen prüfungsrelevanter Aufgaben gezielt trainiert wird. In Japan spielen die privaten Paukschulen, die Jukus, eine derart dominierende Rolle im Leben von immer mehr Schulkindern, dass die Schulverwaltung wiederholt versucht hat, den Einfluss der Jukus und der Test-Industrie zurückzudrängen. Diese Bemühungen waren bisher alle erfolglos (Toshiko Ito 1997). Nicht nur am Wochenende, auch in den Sommerferien besuchen viele japanische Schüler inzwischen private Paukschulen (Schümer 1998). Je wohlhabender die Eltern sind, um so bessere und intensivere private Zusatzförderung wird ihren Kindern zuteil. Durch den überregionalen Einsatz von Tests wird also langfristig nicht mehr Chancengleichheit realisiert, sondern es werden vorgegebene soziale Ungleichheiten verstärkt, denn Kinder aus gutbürgerlichem Elternhaus sind den sozial-benachteiligten Mitschülern bereits ohne privat finanzierte Zusatzförderung überlegen³.

Sofern die Ergebnisse von zentralen Prüfungen und überregionalen Vergleichsarbeiten in Zeugnisnoten eingehen, wird die Notengebung dem einzelnen Schüler weniger gerecht, denn Schüler, die schon unter ungünstigen Schulbedingungen zu leiden hatten, werden in einer zentralen Prüfung eher schlecht abschneiden (Hagemeister 2000). Die heute übliche Notengebung, die sich am Mittelwert der jeweiligen Lerngruppen orientiert, ist in der Regel gerechter, als wenn überregionale Standards in die Noten eingehen. Das heißt nicht, dass Zeugnisnoten, die relativ zum Niveau einer Schulklasse erteilt wurden, im Einzelfall nicht auch sehr ungerecht sein können. Diese Ungerechtigkeiten werden jedoch durch zentral gestellte Aufgaben nicht beseitigt, sondern verschärft. Die Hamburger Studie „LAU 5“

³ Siehe hierzu z. B. die mit Hilfe der OECD-Datenbank berechnete Tabelle:

„Die Abhängigkeit der Testergebnisse von der Zahl der Bücher im Elternhaus“, <http://volker.hagemeister.name/fileadmin/volker/Buecher-im-Elternhaus-Lesen-2000.pdf> - Diese Tabelle ist ein Beispiel dafür, in welchem Ausmaß die Leistungen in der Schule von Einflüsse des Elternhauses abhängen. Keine einzige Variable, die Einflüsse der Schule erfasst, beeinflusst die Testergebnisse so stark wie die Variable „Zahl der Bücher im Elternhaus“

liefert hierzu konkrete Erkenntnisse (Lehmann/Peek/Gänsfuß 1997): In Hamburg waren lehrplan-bezogene Tests aus den Bereichen Deutsch und Mathematik in allen 5. Klassen eingesetzt worden. Die Abbildung aus dem LAU-5-Bericht zeigt, dass die Testergebnisse in erheblichem Maße vom sozialen Umfeld einer Schule abhängen (die gestrichelte Linie in unserer Abbildung repräsentiert die Zunahme der Testergebnisse mit dem sozialen Umfeld der Schule). Die ausgefüllten Quadrate stellen den Test-Mittelwert einer bestimmten Schulklasse dar. Die Lage dieser ausgefüllten Quadrate zeigt, dass die Test-Ergebnisse einer einzelnen Schulklasse in erheblichem Maße von unbekanntem Einflüssen abhängt:



Die Abweichungen einer Schulklasse von der Norm (hier repräsentiert durch den Abstand des ausgefüllten Quadrates von der gestrichelten Linie) kann z. B. folgende Ursachen haben:

- Wichtige Komponenten des sozialen Umfeldes wurden durch die Schüler-, Eltern- oder Lehrer-Fragebögen nicht erfasst.
- Einflüsse, die sich aus der Heterogenität einer Klasse ergeben können, wurden unzureichend erfasst.
- Die Lehrkräfte der Schulklasse, deren Mittelwert weit unterhalb der gestrichelten Linie liegt, haben einen Unterricht abgehalten, der nicht gut zu den Anforderungen der eingesetzten Tests gepasst hat.

Es kann gute Gründe geben, den Unterricht nicht auf Inhalte und Methoden zentral-gestellter Prüfungsaufgaben auszurichten. Deshalb ist es keineswegs immer ein Zeichen für schlechten Unterricht, wenn eine Schulklasse in zentral-gestellten Prüfungsaufgaben unbefriedigende Resultate erzielt.

Obwohl immer wieder betont wird, dass zentrale Prüfungen vor allem mehr Bewertungsgerechtigkeit bringen sollen, zeigt ein Zeitungsbericht aus Berlin, dass das Gegenteil der Fall ist:

„Bildungssenator Klaus Böger (SPD) will Konsequenzen aus den schlechten Mathematik-Resultaten bei den Prüfungen zum Mittleren Schulabschluss ziehen und die Lehrer verstärkt in Fortbildungen einbinden. ... Bei der Mathematikprüfung hatten die Zehntklässler berlinweit lediglich einen Notendurchschnitt von 3,7 erreicht. ... Insgesamt haben rund 620 Schüler eine Sechs in der Matheprüfung geschrieben. ... Als Ursache für dieses schlechte Ergebnis gilt, dass viele Schüler – besonders jene mit Migrationshintergrund – die Textaufgaben nicht verstanden haben. ... Dieses Verständnis für Textaufgaben müsse mehr geübt werden.“ (Kneist 2006) Es wird also eingeräumt, dass etliche der zentral-gestellten Mathematikaufgaben nicht zu dem Unterricht gepasst haben, den viele Schüler hatten. Nun müssen Lehrer fortgebildet werden, damit sie in Zukunft einen zur zentralen Prüfung passenden Unterricht abhalten können. Dass allerdings beim ersten Durchgang „620 Schüler eine Sechs in der Matheprüfung geschrieben“ ist unabänderbar. „Trotz dieser schlechten Resultate lehnt es Böger ab, die Bewertungskriterien zu verändern“. Aus Gründen der Staatsräson kann man auf diese Schülerschicksale keine Rücksicht nehmen. Wer käme schon auf die Idee, den riesigen Apparat einer zentralen Prüfung anzuhalten und in eine andere Richtung zu lenken.

Wer alle Schüler eines Bundeslandes die gleichen Testaufgaben bearbeiten lässt, benachteiligt diejenigen Schüler, deren Schule in einem ungünstigen sozialen Umfeld liegt. Ferner hängen die Noten, die ein Schüler in einer zentralen Prüfung erhält, nicht unerheblich von den Leistungen seiner Lehrer ab. Zentrale Prüfungen, deren Resultate die Zeugnisnoten der Schüler beeinflussen, sind mit weiteren Nachteilen behaftet:

4 Zentrale Prüfungen behindern die Entwicklung leistungsfähiger Diagnose-Instrumente

Weil zu jedem Prüfungstermin neue Aufgaben benötigt werden, die niemand kennen darf, können die Aufgaben nicht erprobt und mit wissenschaftlichen Verfahren bewertet und schrittweise verbessert werden. Wesentlicher Bestandteil des wissenschaftlichen Vorgehens zur Entwicklung von Testaufgaben sind Erprobungsphasen, bei denen eine ausgewählte Schülergruppe Aufgaben-Entwürfe bearbeitet. Bei dieser Erprobung der Aufgaben zeigt sich z. B., ob Aufgaben zu leicht oder zu schwierig sind oder ob sie missverständlich formuliert wurden. Mit der Erprobung würde man allerdings die Aufgaben bekannt machen.

Aufgaben für zentrale Prüfungen dürfen keinesfalls vorab bekannt gegeben werden. Dagegen ist das Veröffentlichen der Aufgaben bei der wissenschaftlich fundierten Entwicklung von Testinstrumenten geradezu erwünscht: Ob Aufgaben, die überregional eingesetzt werden, sinnvoll (bzw. valide) sind und ob sie dem Unterrichtsgeschehen entsprechen, bedarf der öffentlichen Diskussion.

Wenn man Aufgaben, die zu Schul-Curricula passen, erprobt, erfolgen die Erprobungsschritte in der Regel in Jahresabständen. Es werden also mindestens 2 bis 3 Jahren benötigt, um Aufgaben zu entwickeln, die die Arbeit der Schule valide erfassen. So lange kann man auf die Aufgaben für eine zentrale Prüfung nicht warten. Außerdem müssen Prüfungsaufgaben im Geheimen erstellt werden. Deshalb entziehen sich Prüfungsaufgaben, die die Zeugnisnoten beeinflussen, zwangsläufig einer wissenschaftlichen Erprobung. Aus Mangel an Zeit und wegen der fehlenden wissenschaftlichen Begleitung wird es also kaum gelingen, für zentrale Prüfungen anspruchsvollere, neue Aufgabentypen zu entwickeln, durch die z. B. die Kompetenz zur Reflexion und zur Kommunikation erfasst wird. So werden durch zentral gestellte Prüfungsaufgaben Schulreform-Impulse eher verschüttet als gefördert.

Dagegen können Testaufgaben, die sorgfältig entwickelt und mit wissenschaftlichen Methoden erprobt und verbessert wurden, als Diagnose-Instrumente von erheblichem Nutzen sein. Mit Hilfe solcher Tests können Lehrkräfte Rückmeldungen darüber erhalten, wo die von ihnen unterrichteten Schüler Defizite haben. - Das von einer englischen Universität entwickelte Testprojekt „CEM“ könnte hier als Vorbild dienen. Bei CEM werden standardisierte, lehrplan-bezogene Tests eingesetzt. Die Testergebnisse werden zunächst nur den betroffenen Lehrkräften mitgeteilt. Obwohl bei CEM weder Lehrer noch Schulen an den Pranger gestellt werden, weil keine Ranking-Listen in der Presse erscheinen, hat man festgestellt, dass sich die Diagnoseberichte, die dem einzelnen Lehrer vom CEM-Team geliefert werden, positiv auf die Lernergebnisse und das Unterrichtsklima auswirken (Arnold 1999).

5 Die unvermeidbare Überschätzung zentraler Prüfungen und die Vernachlässigung sinnvoller Arbeitsstrategien

In Nr. 2 der „Vierteljahrshefte zur Wirtschaftsforschung“ aus dem Jahre 2003 wird in den Artikeln von Wößmann und von Büchel, Jürges und Schneider die Meinung vertreten, dass Abschlussprüfungen mit zentral gestellten Aufgaben fundamentale Vorteile hätten:

„So wie Geld im Wirtschaftssystem die Funktion einer »Recheneinheit« innehat, ... können einheitliche Leistungsüberprüfungen im Schulsystem eine »Recheneinheitsfunktion« übernehmen. Durch seine Funktion als Recheneinheit oder »Wertmaß« ermöglicht es das Geld, die Schätzung des Gebrauchswertes eines Gegenstandes zu präzisieren und zu quantifizieren und ihn somit mit allen Alternativen vergleichbar zu machen. ... In ähnlicher Weise können Zentralprüfungen eine entscheidende Rolle als Leistungs- und Bewertungsmaßstab im Schulsystem spielen.“ (Wößmann 2003)

„Die Schüler ... profitieren, weil Ergebnisse von zentralen Abschlussprüfungen auf dem Arbeitsmarkt schon aufgrund ihrer breiteren Vergleichbarkeit bessere Signale sind als Ergebnisse von nicht zentralen Prüfungen. ... 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.“ (Büchel u. a. 2003)

„Zentralprüfungen (verändern) die Anreize der am Bildungsprozess beteiligten Personengruppen: ... Dies erzeugt leistungsfördernde Anreize ... Dafür kommt es nicht darauf an, ob die »zentralen« Prüfungen von einer nationalen Behörde, von Behörden auf Bundesländerebene oder einheitlich von privaten Anbietern durchgeführt werden; entscheidend ist, dass sie »extern« in Bezug auf die einzelne Schule stattfinden.“ (Wößmann 2003)

Unbeabsichtigt werden hier Nachteile zentraler Prüfungen beschrieben. Weil die Aufgaben von irgendeiner zentralen Stellen kommen und im ganzen Land eingesetzt werden, erfahren sie eine unkritische Aufwertung. Es ist nicht wichtig, wer die Aufgaben konstruiert hat und ob die Aufgaben sinnvoll sind oder dem Unterricht entsprechen. Trotzdem hängen in Abschlussprüfungen mit zentral-gestellten Aufgaben Schülerschicksale und das Schicksal ganzer Schulen von solchen Aufgaben ab. - Wie unkritisch zentral gestellte Aufgaben akzeptiert werden, weil sie von einer fernen, großen Institution kommen, lässt sich an der PISA-Aufgabe „Geschwindigkeit eines Rennwagens“ ablesen: Obwohl diese Aufgabe mehrere leicht erkennbare Fehler enthält, wurde sie ungeprüft in viele Aufgaben-Sammlungen deutschsprachiger Kultusministerien und Landesinstitute übernommen (Hagemeister 2007b).

Durch die zentrale Aufgabenstellung erhält eine punktuelle, schriftliche Prüfung ein überragendes Gewicht im Schulalltag. Büchel und seine Ko-Autoren haben recht, wenn sie schreiben, dass „Zentralprüfungen ... leistungsfördernde Anreize“ erzeugen. Allerdings bestehen diese Anreize lediglich in der Vorbereitung auf einen einzigen, sehr bedeutsamen schriftlichen Test. Dieses Lernen auf punktuelle Prüfungen hin hindert viele Schüler daran, sinnvolle Arbeitsstrategien zu entwickeln. 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.

Weil die Aufgaben, die in einer zentralen Prüfung gestellt werden, von Menschen gemacht werden, werden immer wieder Aufgaben dabei sein, die zu schwierig sind oder die missverständlich oder fehlerhaft formuliert wurden, zumal die Aufgaben nicht nach bewährten wissenschaftlichen Verfahren erprobt und weiterentwickelt werden können. Aber auch, wenn eine zentrale Institution in der Lage wäre, lauter perfekte Aufgaben zu konstruieren, so sollten diese Aufgaben trotzdem nicht zum Bestandteil von Abschlussprüfungen gemacht werden, weil dadurch das punktuelle Lernen auf einen bestimmten Termin hin bestärkt wird, ohne dass zentrale Prüfungen mit einem Gewinn an Prüfungsgerechtigkeit verbunden sind.

6 Literaturverzeichnis

Abitur in Finnland (2003):

- Die Finnische Abiturprüfung,
<http://www.ylioppilastutkinto.fi/deutsch.html>
- Ministry of education, COMPOSITE NEWS BULLETIN - March 2003
http://www.minedu.fi/OPM/Tiedotteet/composite_news_bulletins/2003/march2003.html?lang=en

Arnold, Karl-Heinz (1999): „Schulen im Vergleich. Probleme des Ranking und Chancen eines Monitoring“ In: Die Deutsche Schule, 91 Jg., Heft 2, S. 218-231

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“, Max-Planck-Institut für Bildungsforschung Berlin,
<http://edoc.mpg.de/get.epl?fid=16094&did=195148&ver=0>

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

Hagemeister, Volker (2000): „Irrwege und Wege zur Testkultur“, in: Die Deutsche Schule, 92 Jg., Heft 1, Seite 87 bis 101,
<http://www.ggg-hessen.de/download/testkultur.pdf>

Hagemeister, Volker (2007a): „Zentrale Prüfungen: Für eine Verbesserung der Qualität von Bildung fehlen bislang die empirischen Belege“,
<http://volker.hagemeister.name/fileadmin/volker/Empirische-Erkenntnisse-zu-zentralen-Pruefungen.pdf>

Hagemeister, Volker (2007b): „Probleme, die der Einsatz zentral gestellter Aufgaben mit sich bringt, diskutiert an Hand einer PISA-Aufgabe“, Veröffentlichung in der Zeitschrift MNU in Vorbereitung.

Kneist, Sigrid (2006): „In Mathematik sind Berlins Schüler am schlechtesten - Mittlerer Schulabschluss - jeder Fünfte fiel durch. Schulsenator kündigt Konsequenzen an“ Der Tagesspiegel vom 25.07.2006, Seite 7,
<http://www.tagesspiegel.de/berlin/archiv/25.07.2006/2677157.asp>

Lehmann, Rainer H. / Peek, Rainer / Gänsfuß, Rüdiger (1997): „Aspekte der Lernausgangslage von Schülerinnen und Schülern der fünften Klassen an Hamburger Schulen“, Amt für Schule, Hamburg
<http://www.hamburger-bildungsserver.de/schulentwicklung/lau/lau5>

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)“, S. 22, Tab. 1.1 und S. 26, Tab. 1.2,
<http://timss.bc.edu/timss1995i/TIMSSPDF/BMathAll.pdf>

Schümer, Gundel (1998): „Mathematikunterricht in Japan – ein Überblick über den Unterricht in öffentlichen Grund- und Mittelschulen und privaten Ergänzungsschulen.“ In: Unterrichtswissenschaft, 26 Jg., Heft 3, S. 195-228

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

Toshiko Ito (1997): „Zwischen »Fassade« und wirklicher Absicht“, Zeitschrift für Pädagogik, Heft 3, Seite 449-466,
<http://www.ggg-nrw.de/Pack/welcome.html>, Abschnitt „Sonstige Texte“

Wößmann, Ludger (2003): „Zentrale Prüfungen als »Währung« des Bildungssystems: Zur Komplementarität von Schulautonomie und Zentralprüfungen“, Vierteljahrshefte zur Wirtschaftsforschung, 72 Jg., Heft 2, Seite 220-237,
http://www.diw.de/deutsch/produkte/publikationen/vierteljahrshefte/docs/papers/v_03_2_5.pdf