

Kosten- und Leistungsrechnung der Bertelsmann-Stiftung und der OECD für das deutsche Bildungssystem

Peter Bender

In letzter Zeit sind mir zwei Studien begegnet, in denen mit wirtschaftswissenschaftlichem Impetus Beträge in Milliarden- bzw. Billionen-Höhe ausgerechnet werden, die das deutsche Bildungssystem mit seinen Mängeln angeblich verursacht bzw. durch die Beseitigung dieser Mängel einsparen könnte. Beide sind im Rahmen des Schwerpunkts „Folgen unzureichender Bildung“ des Projekts „Investitionen in Bildung“ der Bertelsmann-Stiftung entstanden und wurden, zur Stützung bestimmter bildungspolitischer Absichten, mit medialem Pomp der Öffentlichkeit vorgestellt. Beide halten, jede in ihrer Art, einer vernünftigen Analyse nicht stand. Die zweite Studie wurde außerdem auf die OECD-Länder übertragen und einige Monate später von der OECD mit einem für Deutschland, mir nichts, dir nichts, *fast vervierfachen* Billionen-Betrag noch einmal unter die Leute gebracht.

- 1 *Wie teuer und wie wirksam ist das Instrument der Klassenwiederholung? Zur Studie von Klaus Klemm: Klassenwiederholungen – teuer und unwirksam*

Dieser Abschnitt ist ein überarbeiteter Auszug aus (Bender 2009).

- 1.1 *Wie ermittelt man die Kosten von Klassenwiederholungen?*

Dass das Bildungssystem viel billiger wäre, wenn es keine „schlechten“ Lernenden gäbe, ist eine Binsenweisheit. In viel größeren Klassen könnte in viel kürzerer Zeit das Nötige gelernt werden. Eigentlich bräuchte man gar keine Klassen (Schulgebäude!) mehr und kaum noch Lehrerinnen & Lehrer, weil die Kinder und Jugendlichen ja bequem zuhause an ihren Rechnern mit perfekter Software (etwa mit dem guten alten Logo?) lernen könnten. Die Realität ist aber nun einmal nicht so; es gibt „schlechte“ Lernende, und sie verursachen erhebliche Kosten, indem sie den Betrieb aufhalten, zusätzliche Maßnahmen erforderlich

machen usw., bis hin dazu, dass sie mit verantwortlich sind, wenn so manche Lehrerin, so mancher Lehrer vorzeitig den Dienst quittiert. Selbstverständlich können „schlechte“ Lernende den Unterricht auch bereichern. Das wären dann Leistungen, die von den Kosten wieder zu subtrahieren wären. Die Kostenstruktur ist jedenfalls viel komplexer, als sie wirkt, wenn man sie auf reale Zahlungen (hier: durch die öffentlichen Haushalte) reduziert. Das hat man in der Volkswirtschaftslehre schon lange erkannt, sieht sich aber mit der Bewertung nicht bezifferter Kosten und Leistungen und deren Zuordnung zu Verursacherinnen & Verursachern immer wieder vor erhebliche Probleme gestellt.

Z. B. müssten in die Berechnungen die Vorteile der abgebenden Klasse (durch die Entlastung), die der aufnehmenden Klasse (durch die Bereicherung) sowie die der Wiederholenden & Wiederholer (W&W) (durch die Chance des Neubeginns usw.) als Leistungen eingehen. Natürlich weiß man zu wenig über diese Effekte, um sie bewerten zu können. Und genau das spricht überhaupt gegen eine solche Kostenrechnung.

Wenn man sich aber einmal auf sie einlässt, dann stellt es durchaus eine wichtige Erkenntnis dar, dass die W&W nicht kostenneutral im System mitlaufen, sondern dass Klassenwiederholungen Kosten verursachen.

Diese kann man *überschlägig* leicht ermitteln: Einmal Sitzenbleiben verlängert die durchschnittliche Schulzeit von ca. 11 auf ca. 12 Jahre, also um etwa 9%. Im letzten Jahrzehnt blieben vielleicht 30% aller Schülerinnen & Schüler (S&S) irgendwann einmal sitzen (Mehrfachfälle mehrfach gerechnet). Also wird die Gesamtschulzeit aller S&S durch das Instrument des Sitzenbleibens um 2,7% erhöht. Entsprechend geringer wären also die Personal- (und verwandte) Kosten des Schulsystems, wenn es dieses Instrument nicht gäbe. – Bei der viel aufwändigeren Klemmschen Rechnung kommt ungefähr Dasselbe heraus. Sie wirkt zwar genauer und seriöser und macht mehr Eindruck, ist aber letztlich viel umständlicher und birgt

tatsächlich nicht mehr Informationsgehalt als der vorgeführte Überschlag.

Mit Recht weist Klemm (S. 13) darauf hin, dass seine Rechnung nicht nur in denjenigen Bundesländern so anzustellen ist, in denen die Zuweisung von Lehrerstellen (und von entsprechenden Mitteln) pro S&S-Kopf, sondern auch in denjenigen, in denen sie klassenweise erfolgt. Wohl ändert sich in einer Schule oft an der Klassenfrequenz eines Jahrgangs nichts, wenn W&W hinzukommen, aber manchmal, nämlich wenn die Höchstgrenze für die Klassenfrequenz überschritten wird, eben doch, und dann muss gleich eine ganze Klasse zusätzlich eingerichtet werden. Man wird diesen Fall in Zeugniskonferenzen tunlichst verhindern; aber das wird nicht immer gelingen. Jedoch auch wenn keine zusätzlichen Klassen eingerichtet werden müssen, entstehen durch W&W zusätzliche Ausgaben.

Klemm setzt zu Recht in den Bundesländern mit klassenbezogener Lehrerstellenzuweisung pro W&W pauschal nur 50 % der Kosten an, die in den Bundesländern mit kopfbezogener Zuweisung pro W&W anfallen, die er bei seinem Vorgehen wiederum sehr genau berechnen kann. – Der Prozentsatz von 50 % ist willkürlich, und insgesamt ist die ganze Rechnung voll von Annahmen, Schätzungen, Setzungen. Auch wenn statt des Werts von 50 % für diese Kosten nur 0 % angesetzt würde (was bestimmt zu niedrig wäre), käme man noch auf einen Gesamtbetrag von über 700 Millionen Euro.

In *mathematisch idealisierter* Form kann man sich die zahlenmäßige Wirkung des Sitzenbleibens so vorstellen, dass jede aufnehmende Klasse ihrerseits wieder W&W nach unten abgibt, ihr Umfang also (bei *idealisiert* unterstellter konstanter W&W-Quote!) über die gesamte Schulzeit hinweg konstant ist, und sich dann fragen, wo der W&W-Überschuss bleibt. Dieser entsteht im ersten Schuljahr, weil dort die Klassen keine W&W mehr nach unten abgeben können. Die ersten Klassen sind folglich alle um 2,6 % größer, als sie wären, wenn das Instrument der Klassenwiederholung nicht existieren würde, bzw. es gibt entsprechend 2,6 % mehr erste Klassen. Dieser Überhang zieht sich dann durch die ganze Schulzeit; jeder Jahrgang hat schließlich 2,6 % S&S mehr, und in den Abgangsklassen sind letzten Endes jedes Jahr 2,6 % der S&S ein Jahr länger in der Schule gewesen als vorgesehen.

1.2 Worauf stützt sich die Behauptung von der fehlenden Wirksamkeit von Klassenwiederholungen?

Ihre Eignung als Sensationsmeldung für die Presse zieht die Studie einerseits aus dem

scheinbar hohen Kostenbetrag von knapp 1 Mrd Euro (der viel weniger eindrucksvoll ist, wenn man ihn als etwa 2 % der Gesamtausgaben für die allgemeinbildenden Schulen angibt) und andererseits in Verbindung damit aus der Behauptung, dass Klassenwiederholungen unwirksam seien.

Diese Behauptung widerspricht den positiven Erfahrungen ganzer Heerscharen von Lehrerinnen & Lehrern (vgl. a. Tietze & Roßbach 1998, 467). Allerdings handelt es sich bei diesen Erfahrungen nur um die Anhäufung von „opinions“, die zwar von den Expertinnen & Experten für das Lehren & Lernen stammen, aber eben keine statistisch „belastbare“ Forschungsergebnisse darstellen.

Wie belegt nun Klaus Klemm die Behauptung der Unwirksamkeit? Da es zu den Wirkungen auf Klassen verständlicherweise keine brauchbaren Untersuchungen gibt, beschränkt er sich auf die Zielgruppe der W&W selbst und gründet seine Behauptung auf eine Handvoll Arbeiten aus der Literatur (S. 7, 2. und 3. Absatz). Eine davon ist die Habilitationsschrift von Karlheinz *Ingenkamp* „Zur Problematik der Jahrgangsklasse“ von 1967 in der 2., unveränderten Auflage von 1972. Diese bezieht sich auf eine Untersuchung von 1962 an Grundschulkindern im 6. Schuljahr in Berlin-Tempelhof, die wiederum mit einer entsprechenden Untersuchung von 1949 verglichen wird. Ingenkamp bedauert zwar, dass in seinem Buch nicht genug Raum für eine „ideologiekritische Analyse dieses Organisationssystems“ (der Jahrgangsklassen) vorhanden ist (ebenda, 29, s.a. 290 u. a.), aber er nimmt über weite Strecken diese Kritik doch vor und plädiert vehement für eine „Integrierte Gesamtschule“.

Dieses Plädoyer ist mehrfach widersprüchlich. Der Gegenstand seiner empirischen Forschung, die 6. Klasse, ist ja Teil einer Gesamtschule, nämlich der Berliner Grundschule von 1962, die die ersten sechs Schuljahre umfasste. Seine (negativen) Befunde legen doch gar nicht nahe, die Gesamtschule auf die nächsten Jahrgänge auszudehnen. Mit der Vereinheitlichung der Schulen wiederum fordert er zugleich eine ausgeprägte Differenzierung des Unterrichts, weil ja niemand mehr der Einheitsschule enttrinnen kann. Man muss ihm zugute halten, dass er, anders als die heutigen Protagonistinnen & Protagonisten, wenigstens noch die organisatorischen Probleme eines solchen Systems erkennt (S. 301).

Es ist klar, dass Klassenwiederholungen als Kulmination des von ihm kritisierten Systems erst recht in seine Schusslinie geraten. Zur damaligen Zeit wurden ja noch Intelligenztests durchgeführt, und es wundert nicht, dass die

W&W bei diesen und bei fachspezifischeren Tests schlechter abschnitten als die „glatt Versetzten“, erst recht, wenn die W&W mehrfach wiederholten (S. 106, als Haupt-Argument von Klemm zitiert, und vor allem S. 275). Wieso allerdings dieser Umstand gegen die Wirksamkeit von Klassenwiederholungen sprechen soll, erschließt sich mir nicht. Eine *Verringerung* des Leistungsrückstands ist doch auch eine (positive) Wirkung. Dass eine solche vorliegt, ist dermaßen plausibel, dass nicht die Frage, ob, sondern nur, in welchem Ausmaß sie vorhanden ist, interessant gewesen wäre. Dieser Frage ist Ingenkamp allerdings nicht nachgegangen. Die Klassenwiederholung wäre wohl berechtigterweise dann als nicht wirksam zu bezeichnen, wenn die Verringerung des Leistungsrückstands zu schwach ausfiel. Dass der Leistungsrückstand aber dauerhaft auf Null zurückgehen muss, ehe man, konkludent gemäß Ingenkamp und Klemm, von Wirksamkeit sprechen können soll, ist jedoch nicht einzusehen. Ganz in diesem Sinn spricht die Arbeit von Klemms nächsten Kronzeugen, *Belser & Küsel* (1976), (entgegen deren eigenen Tenor) sogar eher *für* den Erfolg von Klassenwiederholungen. Untersucht wurde die „Schullaufbahn von Volksschulabgängern an 26 [von 313] zufällig ausgewählten Schulen Hamburgs“ (S. 103) von 1963 bis 1966, also von Jugendlichen, die etwa 1948 bis 1952 geboren sind. Wohl ist „ganz allgemein zwar im Wiederholerjahr eine Leistungsverbesserung zu beobachten, aber schon im nächsten Schuljahr, in dem neue und höhere Anforderungen gestellt werden, sinken die Leistungen wieder ab“ (S. 105, zitiert von Klemm, S. 7, als Beleg für die Unwirksamkeit von Klassenwiederholungen). Wie weit die Leistungen „absinken“, ist ein paar Seiten später ausgeführt: „Insgesamt erweisen sich dabei 75 % aller zum Zeitpunkt des Sitzenbleibens ungenügenden Zensuren nach 3 Jahren als dauerhaft, mindestens auf 'ausreichend' verbessert“ (S. 111). Die Leistungen sind also vielleicht nicht mehr gut oder befriedigend wie im ersten Jahr, aber eben dauerhaft ausreichend. Dieser Erfolg schlägt sich auch in der Quote der W&W nieder, die den Abschluss erreichen. Während die Hälfte der W&W die Schule mit dem Ende der Schulpflicht, also vor dem Ende der 8. Klasse, verlässt, geht die andere Hälfte ein Jahr länger zur Schule, und davon erreichen 86 % (der nur einmal Sitzengebliebenen) den Abschluss. Dieses Faktum wird übrigens nur wenige Zeilen vor dem o. a. scheinbar kritischen Zitat auf Seite 105 mitgeteilt. Unter den von Klemm zitierten Arbeiten ist (*Belser & Küsel* 1976) die einzige, die den *langfristigen Ertrag* des Klassenwiederholens kon-

trolliert. Auch wenn die Autorin & der Autor selbst es nicht wahr haben wollen, hat sich in ihrer Untersuchung das Sitzenbleiben als eine erfolgreiche Maßnahme erwiesen, auch „schwächere“ S&S zu einem Abschluss zu führen. Beim Lexikon-Artikel von *Tietze & Roßbach* (1998) wird der Charakter des Zitats, das Klemm anführt, ebenfalls erheblich verändert, wenn man es fortsetzt. Trivialerweise schneiden die W&W bei Schulleistungstests nach einem Jahr schlechter ab als diejenigen (gleichschwachen) S&S, die nicht wiederholen und also eine Klasse höher sind. Aber direkt danach folgt: „Werden Sitzenbleiber jedoch mit (leistungsschwachen) Schülern in der gleichen Klassenstufe verglichen (die nicht-versetzten Schüler sind dann mindestens ein Jahr älter), so zeigen sich (geringe) Leistungsunterschiede zugunsten der Sitzenbleiber“ (S. 467). – Na also. Warum die W&W schlechter abschnitten als diejenigen, die nicht wiederholen, bringen *Tillmann & Meier* (2001) mit der folgenden trivialen Aussage auf den Punkt: „Zum einen sind Wiederholer im Durchschnitt mit weniger guten kognitiven Voraussetzungen ausgestattet . . . , zum zweiten wird ihnen aber auch die Befassung mit den anspruchsvolleren fachlichen Inhalten der nächsten Klassenstufe verwehrt“ (S. 475). Während bei Timss die 9. Schuljahre betrachtet werden (mit gewissen Nachteilen), untersucht Pisa die 15-Jährigen und handelt sich dadurch andere Nachteile ein: Z.B. wird, vor allem in Entwicklungsländern, nur eine nicht-repräsentative Stichprobe erfasst, nämlich die der *beschulten* 15-Jährigen. Pisa kann zwar feststellen, dass in gewissen Populationen die Quote der W&W größer ist als in anderen oder dass die W&W weniger Punkte erreichen. Um aber Aussagen zur Wirksamkeit der Klassenwiederholung machen zu können, müsste Pisa auch die 16-Jährigen untersuchen, und es müssten Kriterien festgelegt werden (vielleicht Punktedifferenzen gegenüber wohl bedachten Bezugspopulationen). Wie bei allen von Pisa festgelegten Grenzen (z. B. bei den Kompetenzstufen u.v.a.) wäre das allerdings wieder eine subjektive Angelegenheit, die nur auf „opinions“ beruhen würde. Zusammenfassend lässt sich feststellen, dass es offenbar keine empirischen Untersuchungen gibt zur entlastenden Wirkung des Sitzenbleibens für die abgebende Klasse (mit der Verwendung des Adjektivs „entlastend“ bewege ich mich zwar schon am Rande der political correctness, wie sie mit dem aktuellen Slogan vom „längeren gemeinsamen Lernen“ von interessierten Kreisen festgelegt wird; trotzdem ist diese entlastende Wirkung eine untersuchenswerte Kategorie) und zur möglichen bereichern-

den Wirkung für die aufnehmende Klasse (hier bin ich zum Glück politically correct) gibt. Die zitierten Kritikerinnen & Kritiker von Klassenwiederholungen berufen sich i. W. darauf, dass sich (früher) der IQ und (heute) die kognitive Disposition nach dem Wiederholen kaum positiv entwickeln und dass die W&W gegenüber den Versetzten in Rückstand geraten. Diese Argumentation geht jedoch an den eigentlichen Zielen, die mit dem Sitzenbleiben erreicht werden (sollen), vollständig vorbei.

Tatsächlich macht die von Klemm herangezogene Literatur, mit einer Ausnahme, keine empirisch fundierten Aussagen über die Wirksamkeit von Klassenwiederholungen, auch wenn sie, inklusive dieser Ausnahme, sich zu Klassenwiederholungen (mehr oder weniger deutlich) kritisch äußert. Die Ausnahme (Belser & Küsel 1976) hat herausgefunden, dass unter den W&W, die 1965 in Hamburg ein Jahr länger zur Volksschule gingen, 86 % den Abschluss schafften. Ob die Autorin & der Autor sich mit ihrer (sachten) Distanzierung vom „Sitzenbleiben“ (im Zuge einer recht ausgewogenen Diskussion z. B. auf S. 113) dem Zeitgeist der universitären Pädagogik (nicht nur) der 1970-er Jahre anpassen?

Man muss Klemm zugute halten (kann das aber auch kritisieren), dass er Auswahl und Interpretation seiner Literatur i.W. komplett von dem Diskussionsbeitrag (Krohne & Tillmann 2006) (Mitarbeiterin und Leiter der Bielefelder Laborschule) übernommen hat.

Natürlich gibt es viel mehr Arbeiten zu dem Thema; aber ich habe nur diejenigen für meine Analyse herangezogen, auf die Klemm seine Behauptung der fehlenden Wirksamkeit von Klassenwiederholungen stützt.

1.3 Alternativen zum Instrument der Klassenwiederholung?

Seit dem schwachen Abschneiden Deutschlands bei Pisa wittern die Anhängerinnen & Anhänger der Einheitsschule nach über dreißig Jahren wieder Morgenluft und bekämpfen in einer breiten Front von „Linken“, SPD, „Grünen“, GEW, politisch engagierten Pädagogik-Professorinnen & -Professoren, dpa, „Initiative Neue Soziale Marktwirtschaft“ (INSM), Bertelsmann-Stiftung, OECD und einem Teil der Arbeitgeberverbände mit unterschiedlichen Motiven die gegliederte Struktur des deutschen Schulsystems.

Allerdings gibt es bekanntlich keine wissenschaftlichen Belege für eine Überlegenheit der Einheitsschule, insbesondere nicht bei Pisa. Vielmehr muss der politisch motivierte, idyllisch klingende Slogan vom „längeren gemeinsamen Lernen“ als Begründung herhalten. Da

dieses Organisationsprinzip mit offensichtlichen Schwachpunkten behaftet ist, wird immer automatisch die Phrase hinzugefügt, dass natürlich *jedes Kind individuell zu fördern* sei (so sinngemäß schon Ingenkamp 1972). – Dass hier zunächst einmal ein logischer Widerspruch vorliegt, wird geflissentlich übersehen.

Ob in einem Einheits- oder in einem gegliederten Schulsystem, – so ganz harmlos ist das mit der individuellen Förderung ja nicht. Abgesehen davon, dass die *individuell geförderten* „schwächeren“ S&S mit vergleichbaren psychischen und sozialen Problemen wie die W&W konfrontiert würden (auch und gerade unter einem „Schuldach“), stoßen wir hier auf einen Grundmangel vieler Kosten- und Leistungsrechnungsmodelle, wenn sie nicht wirklich betriebswirtschaftlich umgesetzt werden (wie dies typisch für die Bildungsökonomie ist). Es wird versäumt, wirklich alle Kosten heranzuziehen; hier z. B.: Wie teuer würde denn eine individuelle Förderung aller S&S bzw. wenigstens der „schwächeren“ geraten?

Wenn wir einmal auf unser Pisa-Vorbild Finnland schauen (aber auch ohne diesen Bezug wissen wir das): Die Klassen müssten erheblich verkleinert werden, und sie müssten im Durchschnitt mit mehr als einer Lehrperson besetzt werden. Wohl könnten, besonders beim Einheitsschulprinzip, durch die Renaissance der „guten alten“ (allerdings potenziell seelenlosen) Mittelpunktschule Kosten eingespart werden, und insgesamt sind die Pläne zur individuellen Förderung aller S&S bzw. wenigstens der „schwächeren“ ja noch arg vage, aber ich schätze einmal *konservativ*, dass sie unser Schulsystem mindestens um 25 % verteuern würde. – Das klingt nicht viel, und 1 Mrd Euro hätte man ja schon, indem man das Sitzenbleiben abschaffen würde. Aber man muss sich klar machen, dass der erforderliche Betrag etwa zehnmal so hoch ist und wohl nur durch eine drastische Kürzung der Bezüge des Lehrpersonals auf das finnische Niveau zu finanzieren wäre.

2 Die Entwicklung des deutschen Bruttoinlandsprodukts (BIPs) bis 2090 und darüber hinaus – Zu den Studien von Ludger Wößmann u. a.: Was unzureichende Bildung kostet. Und: The High Cost of Low Educational Performance

Im Spätjahr 2009 erschien eine Studie der Bertelsmann-Stiftung, wonach das deutsche BIP in den nächsten 80 Jahren um 3 Billionen Euro höher ausfallen würde, wenn nur alle Jugendlichen auf wenigstens 420 Pisa-Punkte

gebracht werden könnten. Nach meiner Beobachtung wurde diese Studie von der deutschen Öffentlichkeit weniger intensiv zu Kenntnis genommen als eine ähnliche Modellrechnung der OECD, die mit deren bekannten Propagandaapparat Anfang 2010 durch die Nachrichten gejagt wurde: Da waren es auf einmal fast 13 Billionen Dollar, die Deutschland bei sogar nur 400 Pisa-Punkten gewinnen würde.

2.1 Warum beschäftigt man sich mit solchen Prognosen?

Zunächst machte ich mich über diese Prognosen, und überhaupt die Reduzierung von Bildung auf Geld, lustig und ging in einem Vortrag folgendermaßen auf sie ein: „Hätte die OECD ein paar Jahre weiter gerechnet und das Pisa-Punkte-Niveau vielleicht auf 500 angehoben, wäre sie auf noch viel höhere Beträge gekommen. Sie muss nur aufpassen, dass die anderen Länder nicht ihre Schülerleistungen *entsprechend* steigern, weil dann nämlich alle Pisa-Punkte, die ja *relative* Angaben sind, auf dem aktuellen Niveau bleiben und es gar keine zusätzliche Wirtschaftsleistung gibt.“

Wie so oft im Leben wurde auch hier die Satire von der Realität in den Schatten gestellt: Es war sogar ein Szenario mit 546 Pisa-Punkten (dem Niveau Finnlands) gerechnet und ein – zugegeben, nur qualitativer – Blick in die weitere Zukunft über das Jahr 2300 hinaus geworfen worden.

Immerhin stammt die Studie der Bertelsmann-Stiftung (ab jetzt mit „B“ abgekürzt) von unserem Bundes-Bildungsökonom Ludger Wößmann (und Mitarbeiter Marc Piopiunik), Professor an der Exzellenz-Universität LMU München, der oft in den Medien präsent ist und gerne seine ökonomisch getränkte Meinung zu allerlei Fragen unseres Bildungssystems kundtut. Seit längerem interessiert mich, woher genau er eigentlich seine Erkenntnisse hat, und so schaute ich mir trotz der Absurdität dieser Billionen-Beträge die beiden Studien genauer an. Insbesondere wollte ich wissen, wieso die eine mit viermal so vielen Billionen um sich warf wie die andere. Hier das Ergebnis:

Bei den Pressemeldungen zur OECD-Studie fiel auf, dass der leitende OECD-Angestellte Andreas Schleicher nicht namentlich erwähnt wurde, wo er doch in der Vergangenheit gerne das Volk über den Einfluss von Bildung auf die Ökonomie aufgeklärt hatte. Nachdem sein Auftreten in der deutschen bildungspolitischen Landschaft zu massiven Protesten geführt hatte, hat man ihn möglicherweise etwas aus der Schusslinie genommen. Seiner Meinung nach ist ja das dreigliedrige Schulsystem z. B. verantwortlich für das langsame Wachstum des BIPs

in Deutschland über viele Jahre hinweg, besonders im Vergleich zu dynamischen Nationalökonomien wie Irland, Spanien, Portugal und Griechenland. – Oder: die von ihm beschworene Gefahr, dass Deutschland irgendwann einmal gegenüber den expandierenden Volkswirtschaften von China und Indien ins Hintertreffen gerät (was immer damit gemeint sein soll), rührt angeblich vom deutschen gegliederten Schulsystem her, und nicht etwa von dem Umstand, dass diese Länder 15- bis 20-mal so viele Einwohner haben.

Gegen Ende des Vorworts der OECD-Studie (ab jetzt mit „O“ abgekürzt) erfährt man unauffällig, dass auch sie von Wößmann stammt. Er hat sie zusammen mit Eric A. Hanushek verfasst, Professor an der nicht minder renommierten Stanford University, mit dem er schon seit Jahren gemeinsam publiziert. Dabei wird auch die Konsultation von Andreas Schleicher (also doch!) und anderen OECD-Mitarbeitern erwähnt.

Interessant ist ferner, dass die OECD diese Studie als Teil ihres „Programme for International Student Assessment“ (PISA) deklariert. Für die Vereinnahmung durch Pisa kommt Wößmann mit seinen Arbeiten durchaus in Frage. Er beruft sich auf scheinbar genaue Pisa-Zahlen sowie „Pisa-Sonderstudien“ und behandelt wie selbstverständlich Pisa-Punkte trotz ihres relativistischen und reduktionistischen Charakters als Einheiten einer Art Größenbereich „Bildung“ (wie Kalorien für Wärmemengen oder Euro für Geldwerte), als ob das irgendwo Standard wäre. Außerdem geht es der OECD stets darum, Pisa im Gespräch zu halten, auch wenn es dort gerade nichts Neues gibt.

2.2 Prinzipielles Vorgehen in den beiden Studien

„O“ baut komplett auf „B“ auf, hat jedoch nicht wie „B“ die deutschen Bundesländer, sondern die OECD-Staaten im Blick, und weicht mit einigen Annahmen leicht ab. Weil Studie „B“ stringenter ist, lege ich sie den folgenden Ausführungen zugrunde. Die Argumentation in „B“ (und „O“) verläuft so:

Zunächst wird auf der Basis der Zahlen von 50 wichtigen Ländern „nachgewiesen“, dass die jährliche Wachstumsrate des BIPs stark, natürlich positiv, von der Pisa-Punktzahl abhängt („B“, 17ff, insbesondere 22f). Dieser Zusammenhang wird quantifiziert („B“, 28): Für eine Erhöhung der Pisa-Punktzahl um 100 gibt es eine Erhöhung der jährlichen Wirtschaftswachstumsrate um 1,265 %, und zwar liegt hier strikte Proportionalität vor. In „B“ wird das Szenario entworfen, dass durch eine Bildungsreform alle „Risiko“-Jugendlichen (die mit unter 420 Pisa-Punkten) auf 420 Pisa-Punkte gebracht

werden, und großzügigerweise nimmt man „konservativ“ nur 90 % der dadurch eigentlich entstehenden Steigerung der deutschen Pisa-Punktzahl an. Das sind 14,1 Pisa-Punkte, und die führen zu einer Erhöhung der jährlichen Wirtschaftswachstumsrate um $0,141 \cdot 1,265 \% = 0,17837 \%$.

Nun wird die Entwicklung des deutschen BIPs bis 2090 einmal ohne diese Erhöhung und einmal mit dieser Erhöhung berechnet („B“, 29ff). Dabei wird eine einheitliche sog. Potentialwachstumsrate von 1,5 % unterstellt und die Rechnung einmal mit 1,5 % und einmal mit $1,5 \% + 0,178 \% = 1,678 \%$ durchgeführt. Die jährlichen Differenzen dieser beiden Werte werden noch mit 3 % auf Preise von heute abgezinst und aufsummiert. So entsteht der ominöse Betrag von 2,808 Billionen Euro („B“, 6 u. a.), der Deutschland an seiner Wirtschaftsleistung (BIP) bis 2090 verloren geht, wenn wir nicht die Leistungen der Risiko-Jugendlichen auf 420 Pisa-Punkte steigern.

Allerdings kann man hier nicht einfach geometrische Reihen über 80 Jahre ansetzen und diese mit geschlossenen Termen ausrechnen, weil ja die erforderlichen Bildungsreformen selbst und in ihren Auswirkungen viel Zeit brauchen, so dass der volle Ertrag erst ab 2061 unterstellt werden kann. Wie die Abschlüsse bis dahin vorgenommen werden, kann man in „B“ auf S. 30 nachlesen. Es gelang mir, mit Excel die Rechnung nachzuvollziehen und mit 2,798 Billionen die Vorlage recht genau zu treffen. (Dieses iterative Vorgehen wird in „O“, 20, etwas hochtrabend als „Simulation“ bezeichnet.)

In diesem Zusammenhang kommt einem folgende alte Schulaufgabe in den Sinn: Maria und Josef legen bei Christi Geburt 1 Euro zu 3 % mit Zinseszinsen an. Wie hoch ist dieses Kapital im Jahr 2010? Mit dem Ansatz $1,03^{2010}$ ergibt sich 63 507 008 738 792 447 365 133 039,19 Euro (die 2010-fache Iteration mit Rundung ab der dritten Nachkommastelle würde leicht abweichen). Die Aufgabe hat durchaus ihren didaktischen Wert: An ihr kann man den Unterschied zwischen Zinseszinsen und einfachen Zinsen, die Wirkung exponentiellen Wachstums und die Größe großer Zahlen drastisch vor Augen führen. Mit echten Anwendungen hat sie jedoch nichts zu tun.

Solche „Zinseszinsseffekte“ ... belegen unsere folgenden Berechnungen“ („B“, 28). – Was gibt es da über die Schule hinaus noch zu belegen? – Später („B“, 44) wird festgestellt, dass „die Zuwächse ... erst im nächsten Jahrhundert abnehmen“ (genau: ab dem Jahr 2108) und „nach dem Jahr 2300 ... nicht mehr stark ins Gewicht fallen“. Diese Phänomene werden zutreffend mit dem Effekt der Abzinsung erklärt. Über

diesen braucht man sich natürlich auch nicht zu wundern; denn für ihn steht z. B. gerade die Umkehrung der obigen Aufgabe. – Ohne Abzinsung öffnet sich die Kluft zwischen dem BIP ohne und dem mit zusätzlicher Erhöhung der Wachstumsrate selbstverständlich beliebig weit, und zwar jedes Jahr schneller: Im Jahr 2090 lauten die beiden Werte 7,139 und 7,876; und im Jahr 2300 schon 162,739 und 259,612 Billionen Euro.

2.3 Mängel in den Annahmen und Interpretationen

Die Studien sind mit gravierenden Mängeln behaftet, natürlich nicht in den Rechnungen, sondern in den Annahmen und Interpretationen.

Einige will ich herausgreifen:

Das geht los mit der Korrelation zwischen Länder-Pisa-Punktzahl und Länder-BIP. Erstere wird ermittelt auf der Basis „aller 36 internationalen Vergleichstests in Mathematik und Naturwissenschaften, die zwischen 1964 und 2003 durchgeführt wurden“ („B“, 18), deren Punktzahlen auf den Pisa-Standard normiert werden. Da werden insgesamt 50 Länder einbezogen; für viele Aussagen werden später jedoch nur 23 OECD-Länder betrachtet. Bekanntlich genügen einigermaßen eigentlich erst Timss und Pisa harten statistischen Ansprüchen. Aber es wird das *einfache Mittel* („O“, 46) zwischen allen in der Zeit entstandenen Testergebnissen gebildet, worunter sich so mancher Schrott befindet.

Wohl ist im Moment die kapitalistische Wirtschaftsform mit ihrem Wachstumsfetisch weltweit vorherrschend und die Wachstumsrate des BIPs ein anerkannter Indikator für den wirtschaftlichen Erfolg eines Landes (mit allerlei Relativierungen!). Dass dies in den nächsten Jahren so bleibt, ist jedoch überhaupt nicht ausgemacht. Z.B. fehlen bei dieser Korrelationsanalyse Kuba und die alten sozialistischen Länder Europas mit vergleichsweise guten Bildungssystemen und schlechten Wachstumsraten. Auch die hohen Wachstumsraten von einigen Schwellenländern seit vielen Jahren in Verbindung mit einem relativ niedrigen Bildungsniveau bleiben unberücksichtigt, ebenso wie der Zusammenbruch bzw. Fast-Zusammenbruch Islands, Griechenlands und weiterer Länder sowie der Einbruch der Weltwirtschaft im Zuge der Finanzkrise ab 2008.

Dass das Wirtschaftswachstum mit einer (für die OECD-Länder einheitlich angenommenen; s. „O“, 51) Wachstumsrate von 1,5 % in den nächsten 80 Jahren gleichmäßig weiter geht, ist doch eine völlig unrealistische Unterstellung. Mir kommt eine solche weitreichende Prognose überhaupt sinnlos vor. Wößmann sieht diese Problematik selbst; er legitimiert sein Vorgehen u. a. mit dem Verweis auf den „langfristi-

gen Betrachtungshorizont“, wie er auch in der „Klimapolitik“ besteht (u. a. „B“, 43). Auf den Gedanken, dass gerade die seit einigen Jahren aktuelle „Diskussion um den Klimawandel“ („B“, 18) zu einer grundsätzlichen Änderung der kapitalistischen Paradigmen führen könnte, kommt er nicht. Außerdem beruhen die Prognosen zum Weltklima auf Modellen von ganz anderem Kaliber, die man nicht mit einem Excel-Programm mit 80×5 Zellen auswerten kann.

Gewiss, wenn man nichts weiß, nimmt man einfach den (bisherigen) Mittelwert an (z. B. „B“, 34, für die Bevölkerungsentwicklung). Aber es ist doch gar nicht so unwahrscheinlich, dass in den wichtigsten OECD-Ländern die Wachstumsrate das Niveau aus den Jahren 1960 bis 2000 deutlich unterschreiten wird, und dass dies sogar bewusst angestrebt wird.

Es gibt eine Fülle von Einflussgrößen für die Wirtschaftsleistungen der Länder, unter denen das Bildungsniveau gewiss auch eine ist, aber hier eine i.W. unidirektionale Kausalität zu behaupten, stellt eine grobe Simplifizierung dar. Den Autoren ist anscheinend auch nicht ganz wohl dabei, aber ihre „Hinweise auf die zugrunde liegende Kausalität“ („B“, 22f) überzeugen jedenfalls nicht.

Das Ganze erinnert mich an *Mierscheids Gesetz*, „das absonderlich klingt, aber präzisere Ergebnisse hervorbringt als viele andere demoskopische Modelle. Es lautet: Der Stimmenanteil der SPD bei Bundestagswahlen richtet sich nach dem Index der deutschen Rohstahlproduktion in den alten Ländern“ (Frankfurter Rundschau vom 03.03.2010), und zwar stimmt seit 1961 die Prozentzahl bei den Stimmen sehr genau mit der Produktion, gemessen in Millionen Tonnen, überein, z. B. 2002: 38,5 % Stimmen und 38,6 Millionen Tonnen; oder von 2005 bis 2009 Rückgang um 11,2 % Stimmen und 10,8 Millionen Tonnen. – Wenn ich an (un)heimliche Kräfte glaube, dann müsste ich mich als SPD-Funktionär bemühen, dass die Stahlproduktion angekurbelt wird, damit ich mehr Stimmen kriege, und als Stahlproduzent, dass die SPD Wahlerfolge erzielt, damit die Produktion wächst.

Vor allem ist es die Pekuniarisierung von Pisa-Punkten, der ich einen realistischen Hintergrund abspreche, und dann ebenso deren Konstanzhaltung über lange Zeiträume. Es mag ja sein, dass nur auf diese Weise solche Szenarien aufgestellt werden können. Aber wenn das so ist, dann muss man es eben lassen. Dass eine gediegene Bildung aller Mitglieder eines Staats etwas ganz Wertvolles ist, und zwar für das Individuum wie für die Gesellschaft, das durchaus auch seine pekuniären Vorteile hat, weiß

man auch so. Die Billionen, die da ausgerechnet werden, sind nichtssagend, und ich kann mir keine Planung vorstellen, wo sie irgendwie sinnvoll berücksichtigt werden könnten.

Die Autoren sprechen sich außerdem u. a. für „ein längeres gemeinsames Lernen unter einem Schuldach“ („B“, 51) aus, weil „die empirische Bildungsforschung belegt, dass davon gerade Kinder aus bildungsfernen Schichten profitieren würden, ohne dass die besseren Schüler darunter leiden würden“ („B“, 51). Diese Behauptung, die Wößmann auch gerne in seinen Interviews aufstellt, wird natürlich nicht belegt und außerdem, wie immer, mit der gegenläufigen Forderung flankiert, dass „gleichwohl jeder Schüler individuell optimal gefördert werden“ muss („B“, 51).

2.4 Die Diskrepanz zwischen der Bertelsmann- und der OECD-Studie

In „O“ hat man z. T. etwas andere Zahlen verwendet: Der Zeitraum bis zur vollen Wirksamkeit der Pisa-Punkte-Steigerung wird etwas länger angesetzt („O“, 50). Die Risiko-Jugendlichen müssen nur auf 400 Pisa-Punkte kommen („O“, 8); aber dafür unterbleibt der 90 %-Ansatz. Die Finnland-Punktzahl wird mit 546 (statt 541,5) angenommen („O“, 7). Während in „B“ ein gleichmäßiger Rückgang der deutschen Bevölkerung bis 2050 von 81,96 Millionen auf 71,35 Millionen und ab dann Konstanz unterstellt wird („B“, 34) (ebenfalls hoch-spekulativ), finden sich solche Überlegungen in „O“ nicht, da ja dann für jedes Land eine eigene Entwicklung berechnet werden müsste. Stattdessen wird, wie gesagt, für alle OECD-Länder einfach einheitlich eine Potentialwachstumsrate von 1,5 % angenommen.

Das alles sind keine rechnerisch allzu bedeutsamen Unterschiede, und auch in „O“ müsste etwas in der Größenordnung von 3 Billionen Euro als „Folgekosten durch entgangenes Wirtschaftswachstum“ für Deutschland herauskommen. Tatsächlich findet man dort die Zahl von 12,576 Billionen („O“, 26, Table 3), und diese Zahl wurde auch in der Pressekampagne Ende Januar 2010 in den Vordergrund gestellt, ohne dass die Vervielfachung gegenüber den 2,808 Billionen Euro kommentiert worden wäre.

Zunächst einmal muss man beachten, dass der Betrag von 12,576 Billionen in Dollar angegeben ist. In „O“ wird gegenüber „B“ ein Kurs von 1,21 Dollar für 1 Euro unterstellt: 12,576 Billionen Dollar sind 416 % des aktuellen deutschen BIPs, das also mit 3,023 Billionen Dollar („O“, 26, Table 3) angenommen wird, während es in „B“ 2,492 Billionen Euro lautet („B“, 74). Die „Folgekosten durch entgangenes Wirtschaftswachstum“ für Deutschland belaufen sich laut

„O“ also auf 10,393 Billionen Euro, das 3,7-Fache des in „B“ ermittelten Betrags.

Es gelang mir auch in diesem Fall, die Rechnung in Excel (mit dem Ergebnis 10,390 Billionen Euro) zu replizieren. Und zwar wird die Erhöhung der jährlichen Wirtschaftswachstumsrate durch die Verbesserung der Pisa-Leistungen im Endstadium nicht wie in „B“ mit 0,17837% , sondern mit fast dem vierfachen Satz 0,65446% angenommen, obwohl die Risiko-Jugendlichen sich bei „O“ nur auf 400 Pisa-Punkte verbessern müssen, während sie bei „B“ ja sogar auf 420 Pisa-Punkte kommen müssen.

Tatsächlich habe ich in „O“ einen arithmetischen Zusammenhang entdeckt, der auf diese höhere Wachstumsratenerhöhung führt: Im Anhang („O“, 47) wird für den „share of students reaching basic literacy ... (taken as 400 test-score points ...)“ ein einheitlicher Koeffizient (mit von mir nicht zu erschließender und sich durch das Ergebnis als offensichtlich sinnlos erweisender Bedeutung) von 0,03783 (Table B1 (5)) „geschätzt“. In Table 3 („O“, 26) wird für jedes OECD-Land dieser Koeffizient mit dem Anteil der Risiko-Jugendlichen multipliziert. Für Deutschland ergibt sich so $0,03783 \cdot 17,3\% = 0,65446\%$.

Allerdings habe ich den Eindruck, dass Table 3 (und der zugehörige Text, „O“, 26) einerseits und Table B1 (mit ihrem Text, „O“, 47) andererseits nicht aufeinander abgestimmt sind. Da geistern auch noch die Jugendlichen mit mehr als 600 Pisa-Punkten herum: Auf S. 26 sind sie „included in the growth models (see Annex Table B1, column 5)“; auf S. 47 werden sie aber explizit ausgeschlossen. Es ist auch nirgends ersichtlich, wie sie einbezogen sein sollen, und sowohl die Überschrift, als auch der einleitende Text von „Scenario III: Bring everyone up to a minimum skill level of 400 PISA points“ („O“, 25) lässt keinen Gedanken an einen Einbezug der Top-Jugendlichen aufkommen. Vermutlich handelt es sich hier um nicht beseitigte Spuren einer nebensächlichen Diskussion aus „B“ (6off), und es sind faktisch zwei Szenarien fehlerhaft vermengt.

In „B“ wurde auch noch einmal mit 400 Pisa-Punkten gerechnet („B“, 64f) und ein Gewinn von 2,036 Billionen Euro ermittelt, demgegenüber der „O“-Wert sogar über 5-mal so groß ist.

Welcher Wert gilt nun? Im Herbst 2009 lautete er 2,808 Billionen Euro, und noch am 02.04.2010, als „O“ schon lange verbreitet war, erklärte Wößmann in einem Interview mit der Wirtschaftswoche: „es geht nicht um Peanuts, sondern um gewaltige Summen – nach meiner Berechnung fast drei Billionen Euro“.

Im Januar 2010 lautete der Wert 12,576 Billionen Dollar, eine viel gewaltigere Summe, gegenüber der die 2,8 Billionen Euro von vorher doch fast wie Peanuts erscheinen. Abgesehen davon, dass man sich solche Zahlen sowieso nicht richtig vorstellen kann, verliert man spätestens dann jeden Respekt vor ihnen, wenn es offensichtlich egal ist, ob es sich um 2,8 oder 10,4 Billionen Euro handelt.

Dass in wissenschaftlichen Arbeiten Fehler gemacht werden, ist gar nicht so selten. Dass die beschriebene Diskrepanz von Wößmann, Hahnushek, Schleicher und Anderen nicht bemerkt wurde, halte ich für unmöglich. Dass dann die OECD trotzdem die Arbeit „O“ veröffentlicht und die Ergebnisse mit Getöse unter die Leute bringt, ist nicht in Ordnung, gehört aber wohl zur Philosophie der OECD. Und dass schließlich Ludger Wößmann sich nicht dagegen verwahrt, entspricht dem bekannten Pisa-Habitus, zu Pisa-eigenen Fehlern eisern zu schweigen.

2.5 Fragwürdige Parameter-Definitionen und -Schätzungen

Es ist übrigens gar nicht so selten, dass ein Parameter trotz scheinbar sinnvoller Definition oder zwei Parameter, die Ähnliches, oder sogar Dasselbe beschreiben sollen, aber begrifflich unterschiedlich definiert werden, stark divergierende, bisweilen widersprüchliche Ergebnisse nach sich ziehen; – sei es, dass sie auf weichen Daten oder eben doch allzu willkürlichen Annahmen beruhen, deren Widersprüchlichkeit zunächst nicht sichtbar wird; sei es, dass sie statistisch oder numerisch arg empfindlich sind. In (Bender 2009) habe ich für Pisa einige solche Parameter analysiert (der soziale Gradient und die Wahrscheinlichkeit des Gymnasialbesuchs).

Bei vielen Studien treten solche Mängel nicht zutage, weil es zu wenige Durchgänge (oft nur einen) gibt. Bei den „Kosten unzureichender Bildung“ haben schon zwei Versionen gereicht, und auch bei Pisa treten mehr und mehr Unzulänglichkeiten auf, nachdem jetzt immer mehr Durchgänge stattfinden.

Literatur

Belser, Helmut & Gabriele Küsel (1976): Zum Sitzenbleiber-Problem an Volksschulen. In: Rudolf Biermann (Hrsg.): Schulsche Selektion in der Diskussion. Bad Heilbrunn: Klinkhardt, 101–115

Bender, Peter (2007): Was sagen uns PISA & Co, wenn wir uns auf sie einlassen? In: Jahnke & Meyerhöfer, 281–337

Bender, Peter (2009): Problematik der Messinstrumente am Beispiel jüngerer Schulstudien. In: Jörg-Dieter Gauger & Josef Kraus (Hrsg.): Empirische Bildungsforschung – Not-

- wendigkeit und Risiko. Fachtagung 2009, Konrad-Adenauer-Stiftung, St. Augustin & Berlin, 41-69
- Ingenkamp, Karlheinz (1967 & 1972): Zur Problematik der Jahrgangsklasse. 2. Auflage. Weinheim: Beltz
- Jahnke, Thomas & Wolfram Meyerhöfer (Hrsg.) (2005 & 2007): Pisa & Co. Kritik eines Programms. 2. Auflage. Hildesheim & Berlin: Franzbecker
- Klemm, Klaus (2009): Klassenwiederholungen – teuer und unwirksam. Gütersloh: Bertelsmann-Stiftung
- Krohne, Julia & Klaus-Jürgen Tillmann (2006): „Sitzenbleiben“ – eine tradierte Praxis auf dem Prüfstand. In: Schulverwaltung Spezial 4/2006, 6–9
- Prenzel, Manfred, Jürgen Baumert, Werner Blum, Rainer Lehmann, Detlev Leutner, Michael Neubrand, Reinhard Pekrun, Jürgen Rost & Ulrich Schiefele (PISA-Konsortium Deutschland) (Hrsg.) (2005): PISA 2003. Der zweite Vergleich der Länder in Deutschland – Was wissen und können Jugendliche? Münster u. a.: Waxmann (zit. als Pisa 2005)
- Tietze, Wolfgang & Hans-Günther Roßbach (1998): Sitzenbleiben. In: Rost, Detlef H. (Hrsg.): Handwörterbuch Pädagogische Psychologie. Weinheim: Beltz, 465-469
- Tillmann, Klaus-Jürgen & Ulrich Meier (2001): Schule, Familie und Freunde – Erfahrungen von Schülerinnen und Schülern in Deutschland. In: Jürgen Baumert, Eckhard Klieme, Michael Neubrand, Manfred Prenzel, Ulrich Schiefele, Wolfgang Schneider, Petra Stanat, Klaus-Jürgen Tillmann & Manfred Weiß (Deutsches PISA-Konsortium) (Hrsg.): PISA 2000. Basiskompetenzen von Schülerinnen & Schülern im internationalen Vergleich. Opladen: Leske + Budrich, 468-509
- Wößmann, Ludger & Eric A. Hanushek (2010): The High Cost of Low Educational Performance. The Long-Run Economic Impact of Improving PISA Outcomes. Paris: OECD
- Wößmann, Ludger & Marc Piopiunik (2009): Was unzureichende Bildung kostet. Eine Berechnung der Folgekosten durch entgangenes Wirtschaftswachstum. Gütersloh: Bertelsmann-Stiftung