

Die Autoren dieser Replik, Manfred Prenzel, Jürgen Baumert und Eckhard Klieme, waren bzw. sind die Koordinatoren für die PISA–Erhebungen in den Jahren 2000, 2003, 2006 und 2009 in Deutschland und zugleich die Sprecher der Konsortien, die für die nationale Auswertung und Berichterstattung verantwortlich zeichnen.

Steuerungswissen, Erkenntnisse und Wahlkampfmunition: Was liefert die empirische Bildungsforschung? – Eine Antwort auf Klaus Klemm

Die empirische Bildungsforschung sei in Gefahr, nach einem kurzen Höhenflug abzustürzen, weil sie das von ihr erwartete politische Steuerungswissen nicht in der erforderlichen Klarheit bereit stelle. Diese These vertritt Klaus Klemm in seinem am 15. Mai 2008 in der „ZEIT“ veröffentlichten Beitrag und versucht, sie mit Beispielen aus PISA zu belegen. Klaus Klemm, der jahrelang im wissenschaftlichen Beirat von PISA mitwirkte, ist mit der Sache vertraut. Er ist auch ein von politischer Seite nachgefragter Berater. In seinem ZEIT-Artikel äußert er unter anderem Zweifel, ob es im deutschen versäulten Schulsystem sozial gerechter zugehen könne oder Verbesserungen gerade im unteren Leistungsbereich nachweisbar seien, und er macht sich Sorgen, dass eine zu breit und inkonsistent definierte Risikogruppe Arbeitgebern eine Legitimation für unzureichende Ausbildungsleistungen liefern könnte. Klemms Beitrag thematisiert damit nicht nur methodische Probleme der PISA-Studien, sondern auch das Verhältnis von Bildungsforschung und Bildungspolitik. Doch welche Ansprüche stellen die Bildungspolitik einerseits und die empirische Bildungsforschung andererseits an Wissen für die politische Steuerung?

Steuerungswissen

Es ist das Versprechen der OECD, mit PISA den Ländern Wissen zur *Steuerung* ihrer Bildungssysteme zu liefern. Dieses Versprechen macht die Teilnahme an PISA attraktiv. Welches Steuerungswissen kann von internationalen Vergleichsstudien erwartet werden?

Wer steuern möchte, muss zunächst einmal wissen, wo er oder sie sich befindet. Eine erste Steuerungsinformation betrifft die Kompetenzen von Fünfzehnjährigen, die – gewissermaßen als Bildungsergebnisse – in den verschiedenen Staaten *erreicht werden*. Hier können sich die Staaten untereinander vergleichen oder ihre Ergebnisse anhand eigener Zielsetzungen beurteilen. Der internationale Vergleich führt aber auch zu einer zweiten grundlegenden Information: Man erfährt, welches Kompetenzniveau derzeit weltweit erreicht werden *kann*. Der Bildungsstand, der von den Fünfzehnjährigen in Finnland erreicht wird, löst in Deutschland nach wie vor Erstaunen aus. Die OECD

verspricht eine dritte Art der Information: Aus Trendanalysen über mehrere Erhebungszyklen hinweg – bislang also für die PISA-Studien 2000, 2003 und 2006 – sollen die Staaten ablesen können, wie sich der Kompetenzstand über die Jahre hinweg *verändert* hat.

Für andere Politikfelder ist ein solches *Systemmonitoring* nicht neu. Wir haben uns daran gewöhnt, dass Veränderungen des Bruttonationalprodukts und der Kaufkraft sehr präzise gemessen und international vergleichend berichtet werden. Dasselbe gilt beispielsweise für die Zahl der Jugendlichen, die ohne Schulabschluss bleiben. Nie zuvor wurde jedoch versucht, zentrale Dimensionen der Grundbildung für Staaten in allen Kontinenten einheitlich zu definieren und Tests zu entwickeln, die fair und interkulturell vergleichbar über mehrere Zyklen hinweg verlässliche Daten liefern. Dieses letzte Ziel bedeutet für die Sozialwissenschaften eine große methodische Herausforderung, die – in der Tat – bis jetzt noch nicht vollständig gemeistert ist.

Nimmt man die Informationen über die Lage, das Erreichbare und die Trends zusammen, dann liegt es nahe zu fragen, wie ein Staat so gut werden kann wie Finnland. Die Erwartungen von Bildungsadministration und -politik richten sich damit auf einen grundlegend anderen Typ von Steuerungswissen. Statt Kompetenzen und deren Trends „nur“ zu beschreiben, wird nach politisch beeinflussbaren Bedingungen gefragt: Welche Merkmale zeichnen ein erfolgreiches Bildungssystem aus? Oder noch anspruchsvoller: Welche Faktoren sind dafür verantwortlich, dass ein zunächst weniger erfolgreiches Bildungssystem besser wird? PISA liefert viele Daten über Schulsysteme und Schulmerkmale, die Vergleiche erlauben. Einfach gedacht, könnte man ja versuchen, alle attraktiven Merkmale des finnischen Schulsystems auf Deutschland zu übertragen. Das Problem besteht nur darin, dass viele Merkmale des finnischen Schulsystems (z.B. die Gemeinschaftsschule) auch in anderen Staaten anzutreffen sind, die genauso oder weniger erfolgreich sind als Deutschland (z.B. Norwegen). Neben dem Schulsystem selbst spielen kulturelle Normen und Werte, soziale und ökonomische Faktoren, der politische Umgang mit Migration, aber auch Unterrichtstraditionen oder die Qualität der Lehrerbildung eine Rolle. Auch in den PISA-Daten findet man häufig keine klare Beziehung zwischen Schulsystemmerkmalen und dem erreichten Kompetenzniveau. Und wenn man Beziehungen findet, kann man sie nicht als Ursache-Wirkungs-Zusammenhänge interpretieren, denn PISA ist eine nichtexperimentelle Querschnittstudie. Aus den dort beschriebenen einfachen oder multiplen Korrelationen sind in der Regel keine Kausalaussagen ableitbar.

Die OECD hat – anders als die für Deutschland verantwortlichen PISA-Forscher – immer wieder versucht, Korrelationen mit Plausibilitätsargumenten als Kausalzusammenhänge zu interpretieren, oder solchen Interpretationen Vorschub geleistet und damit die Grenzen zwischen Politik und Wissenschaft verwischt. Beispielsweise sagt der OECD-Bericht für PISA 2006, dass Schulen, die Ergebnisse von Lernstandserhebungen oder anderen Evaluationen öffentlich bekannt geben, bessere

Ergebnisse im PISA-Test erreichen. Aus diesem eher trivialen Zusammenhang (wer würde erwarten, dass Schulen mit problematischen Leistungsergebnissen werben!) leitet die OECD die Behauptung ab, dass eine Veröffentlichung von Evaluationsbefunden die Schulqualität befördert. Die öffentliche Diskussion, auch in Deutschland, beruft sich gerne auf solche Ergebnisse, vergisst aber, dass es sich nur um Korrelationen innerhalb einer Testerhebung handelt. Die untersuchten Bedingungen müssen keineswegs kausal bedeutsam für den Bildungserfolg sein.

Für empirische Bildungsforscher heißt Steuerungswissen an diesem Punkt, laut und deutlich zu sagen, dass allein aus den beschreibenden Daten und Korrelationen des Systemmonitorings politisch gewünschte und plausible Strukturveränderungen *nicht* abgeleitet werden können. Natürlich sprechen solche Daten auch nicht *gegen* bestimmte Veränderungen, doch wären diese politisch zu begründen, und nicht durch Verweis auf PISA.

Diese Grenzen der Aussagefähigkeit machen internationale Vergleiche keineswegs überflüssig. Systemmonitoring ist zur Vergewisserung über den Stand des Bildungswesens, über Trends und Zusammenhänge unerlässlich. Wenn man aber herausfinden will, welche Folgen eine spezielle politische oder pädagogische Maßnahme hat, muss man die Entwicklung der betroffenen Schulen und Schüler systematisch und über einen ausreichend langen Zeitraum mit der Entwicklung anderer Schulen und nicht betroffener Schüler vergleichen. Wenn Grenzen und Nutzen der empirischen Bildungsforschung realistisch gesehen werden, dann ist sie dort angekommen, wo andere empirische Disziplinen mit längerer Forschungstradition längst sind: im Bereich der normalen Wissenschaft.

Die deutschen PISA-Experten haben dies immer betont und waren deshalb mit Handlungsempfehlungen zurückhaltender als die OECD. Sie haben zugleich unterstrichen, dass eine internationale Vergleichsstudie wie PISA ein Pendant solider Bildungsforschung in den einzelnen Staaten braucht. Die für Deutschland wichtigsten Ergebnisse, die über die Ländertabellen der OECD hinausgehen, sind aus den an PISA angelagerten deutschen Forschungsprogrammen hervorgegangen. Dazu gehören Themen wie die Überprüfung der Lehrplanvalidität der Tests, innovative Testerweiterungen, Ländervergleiche, Erweiterung der Migrantentstichprobe, ergänzende Untersuchungen zu Bildungsgerechtigkeit, Unterrichts- und Schulqualität, Leistungsentwicklung und Leistungsstagnation oder Untersuchungen zur professionellen Kompetenz von Lehrkräften. Forschungsergebnisse zu diesen Themen haben den Gedanken einer stärker evidenzbasierten Bildungspolitik überhaupt erst entstehen lassen.

Was sind Risikogruppen?

Nach Klemm hat kaum ein Begriff aus dem Vokabular der PISA-Studien eine vergleichbare Karriere gemacht. Recht so! PISA 2000 hat mit diesem Begriff zum ersten Mal überhaupt auf ein – wenn man

so will – skandalöses Problem des deutschen Schulsystems aufmerksam gemacht: Ein quantitativ erheblicher Anteil der 15-Jährigen erreicht in den Basisqualifikationen Lesen und Mathematik nicht das Bildungsminimum, das für selbständiges Weiterlernen und das Erlernen eines zukunftsfähigen Berufs vorausgesetzt wird. Damit erfüllt die Schule eine elementare Qualifikationsaufgabe nur unzureichend. Klemm bemängelt nun, dass die Abgrenzung der Risikogruppen unbegründet sei und überdies von Zyklus zu Zyklus wechsele. Hier hat Klemm unsere Berichte nicht genau genug gelesen.

Die Risikogruppen werden durch die Zugehörigkeit zu Kompetenzstufen eindeutig festgelegt. Die Bedeutung der Kompetenzstufen wird operational definiert, indem anhand typischer Beispielaufgaben für jedermann nachvollziehbar gezeigt wird, was Jugendliche können und was sie noch nicht beherrschen. Betrachten wir zum Beispiel die Mathematik: Zur Risikogruppe gehören Personen, die alle Lehrplananforderungen massiv unterschreiten und weit davon entfernt sind, typische Aufgaben von Einstellungstests lösen zu können. Im ersten Bericht von PISA 2000 (S. 168) heißt es zur Kompetenzstufe I: "Sie können arithmetisches und geometrisches Wissen, wie es in der Grundschule (sic!) vermittelt wird, in vertrauten Aufgabenstellungen anwenden, also zum Beispiel durch Dividieren und Multiplizieren einen Preis ermitteln oder den Flächeninhalt eines Rechtecks mit vorgegebenen Kantenlängen berechnen. Anforderungen, die über solche elementare Standardaufgaben hinausgehen, können sie nicht bewältigen. Damit sind sie ...nicht in der Lage, typische mathematische Aufgaben für Ausbildungsplatzbewerber zu lösen. Die Schülerinnen und Schüler auf Kompetenzstufe I (und a fortiori auf unterer Stufe) müssen demnach im Hinblick auf ihre weiteren Bildungs- und Berufschancen als Risikogruppe angesehen werden." Diese Definition wird konsistent über alle Berichte durchgehalten. Ab 2003 gibt es auch ein internationales Kompetenzstufenmodell, das den internationalen Vergleich ermöglicht. Zur Risikogruppe gehören in Deutschland bei PISA 2003 21,6 und bei PISA 2006 19,9 Prozent der 15-Jährigen.

Im Hinblick auf die Lesekompetenz argumentiert PISA-2000 vorsichtiger, weil es schwieriger war, vergleichbare externe Validitätskriterien heranzuziehen. Außer Frage steht die Unzulänglichkeit der Lesekompetenz für jede Form des selbständigen Weiterlernens bei Jugendlichen, die keine der PISA-Aufgaben sicher lösen können und damit unter der Stufe I bleiben. Diese Jugendlichen identifiziert PISA-2000 als eindeutige Risikogruppe. Ihr Anteil beträgt 9,9 Prozent an der Altersgruppe. Aber auch für Stufe I gibt es genügend Anhaltspunkte, dass typische Leseanforderungen, die in allen zukunftsfähigen Berufen anfallen und auch in PISA geprüft werden, nicht mit der nötigen Sicherheit und Schnelligkeit bewältigt werden können. Deshalb fasst PISA-2000 Jugendliche der Kompetenzstufen I und unter I als *potentielle Risikogruppe* zusammen. Diese Abgrenzung wird konsistent durchgehalten und die Gruppe wird genau beschrieben. Zu dieser Gruppe gehören etwa 22,6 % der Altersgruppe. Besonders gefährdet sind Jugendliche aus sozial schwachen und zugewanderten

Familien. Mittlerweile gibt es Befunde aus einer an PISA angekoppelten kanadischen Längsschnittstudie (YITS), die zeigen, dass 15-Jährige, welche die Kompetenzstufe I (und sogar II) im Lesen nicht überschreiten, in hohem Maße der Gefahr ausgesetzt sind, überhaupt keinen High-School-Abschluss zu erreichen. Danach sollte auch die Gruppe, die im Lesen Kompetenzstufe I nicht überschreitet ohne Einschränkung als Risikogruppe bezeichnet werden – wie dies in PISA-2003 geschehen ist. Zur dieser Risikogruppe gehören in Deutschland 2000 22,6 Prozent, 2003 22,3 Prozent und 2006 20,0 Prozent.

Auch im Hinblick auf die Naturwissenschaften ist PISA konsistent. Der Begriff der Risikogruppe wurde hier nicht verwendet, weil die Evidenz für ein Scheitern im gesellschaftlichen Alltag nicht so leicht beizubringen wäre. Doch auch hier gilt, dass Jugendliche auf oder unter der ersten Kompetenzstufe sehr ungünstige Voraussetzungen für jede weitere Beschäftigung mit den Naturwissenschaften mitbringen.

Klaus Klemm vermag sich mit Aussagen zu Risikogruppen nicht anzufreunden, weil sie seine häufig vorgetragene Kritik an der Ausbildungsunwilligkeit von Betrieben schwächen könnten. Mit seinem Argument aber, dass unsere Definition der Risikogruppen Betrieben „...auf bequeme Weise eine Legitimation für ihre Weigerung, Ausbildungsplätze in bedarfsgerechter Anzahl bereit zu stellen“ liefere, kommt er jenen nahe, die PISA-Befunde je nach Passung für eigene politische Interessen benutzen oder abwehren.

Soziale Herkunft und Kompetenzerwerb: Geht es in Deutschland neuerdings gerechter zu?

Die soziale Lage von Personen und Familien kann man vernünftigerweise nicht eindimensional beschreiben. PISA verwendet zur Kennzeichnung sozialer Lagen – und das ist gegenüber allen anderen internationalen Vergleichsuntersuchungen ein immenser Fortschritt – eine Reihe von Indikatoren: die durch das internationale Arbeitsamt in Genf (ILO) erarbeitete Klassifikation von Berufen und den daraus abgeleiteten internationalen Sozialschichtindex (ISEI) sowie komplexe Indikatoren für ökonomisches, soziales und kulturelles Kapital. PISA-Deutschland erlaubt darüber hinaus noch das für Sozialstrukturanalysen theoretisch am besten begründete Modell sozialer Klassen (EGP) abzubilden. Diese Merkmale werden zu deskriptiven Zwecken jeweils einzeln und mit analytischer Zielsetzung multivariat mit dem Kompetenzerwerb in Beziehung gesetzt. Ziel ist es, in einem theoretischen Modell den Zusammenhang zwischen sozialer Herkunft und Bildungsprozessen zu erklären und dieses Modell empirisch zu überprüfen.

Dies alles kann den Ökonomen Klemm nicht ernsthaft verwirren, zumal die Ergebnisse völlig konsistent sind. Das ist wissenschaftliches Alltagsgeschäft. Den Politikberater Klemm irritiert aber der Befund, dass sich die Kopplung von sozialer Herkunft und Kompetenzerwerb in Deutschland im

Zeitvergleich und im Vergleich zu Nachbarstaaten zwischen 2000 und 2006 abgeschwächt haben soll. Zu dieser Irritation hat die OECD beigetragen, als sie 2003 und nochmals verändert 2006 einzelne Indikatoren zu einem komplexen Index zusammenfasst und damit Vergleichbarkeit erschwert hat. Das PISA-Konsortium hat 2003 neben dem ISEI auch den neuen OECD-Index berichtet und damit möglicherweise zur Verwirrung beigetragen. Insofern ist Klemms Kritik tatsächlich berechtigt. Das deutsche Konsortium hat aber in seinem jüngsten Bericht zu PISA 2006 den Sachverhalt transparent dargestellt, die verschiedenen Perspektiven erklärt und Vergleiche von PISA 2000 bis PISA 2006 auf der Basis des international am besten geeigneten Indikators durchgeführt. Der Zusammenhang zwischen sozialer Herkunft und Kompetenz hat sich in Deutschland seit 2000 – auch im internationalen Vergleich - geringfügig, aber signifikant verringert.

Eigentlich haben wir hier ein schönes Beispiel dafür, wie PISA Steuerungswissen beitragen kann, indem eine für Deutschland erfreuliche Rückmeldung gegeben wird, ohne deswegen Entwarnung zu signalisieren. Der Zusammenhang zwischen Herkunft einerseits sowie Kompetenz und Bildungsbeteiligung andererseits ist schwächer geworden, aber immer noch alarmierend, so dass das Thema wohl auf der Agenda der Bildungspolitik bleiben wird. Vor diesem Hintergrund glauben wir, dass Deutschland gut beraten ist, weiterhin eigene Analysen durchzuführen, nach Bedarf die Erhebungen und Instrumente zu ergänzen und eigene Berichte anzufertigen. Andernfalls besteht die Gefahr, dass für Deutschland wichtiges Steuerungswissen verborgen bleibt. Wer könnte ein Interesse daran haben, nicht wissen zu wollen, dass ein Bildungssystem auch ohne strukturelle Veränderung etwas gerechter werden kann?

Überschätzung der Leistungen durch Unterrepräsentation von Zuwanderern in PISA 2006?

Klemm würdigt den Beitrag von PISA zur Sozialstrukturforschung und die Leistung, ein modernes Migrationskonzept in die Surveyforschung eingeführt zu haben, das es zum ersten Mal erlaubt, die Zuwanderungsproblematik im ganzen Ausmaß darzustellen. Dann jedoch vermutet Klemm, dass durch eine Unterrepräsentation von Migranten in der PISA-Stichprobe 2006 die Leistungen insgesamt und vor allem im unteren Leistungsbereich überschätzt würden und so die berichteten Leistungsverbesserungen möglicherweise ein Artefakt seien. Als Referenz zieht er den Mikrozensus 2006 heran, der erstmalig das PISA-Migrationskonzept verwendet.

Auf Klemms Vermutung gibt es zwei klare Antworten. Die erste betrifft die behauptete stichprobenbedingte Überschätzung des Trends. Die deutschen PISA-Stichproben sind exzeptionell gut und über die Zeit immer besser geworden. Die Ausschöpfungsquoten lagen im Jahr 2000 auf Schulebene bei 95 und auf Schülerebene bei 86 Prozent und stiegen bis 2006 auf 100 bzw. 92,3 Prozent. Von solchen Ergebnissen können alle anderen Surveys nur träumen. Und es gibt keinen stichhaltigen Hinweis,

dass sich die Beteiligung von Jugendlichen aus Zuwandererfamilien am Test von 2000 bis 2006 geändert hätte. Der Anteil von 15-Jährigen mit Migrationshintergrund in der PISA-Stichprobe beträgt 21,7 Prozent im Jahr 2000, 20,6 Prozent im Jahr 2003 und 19,4 Prozent im Jahr 2006. Diese Veränderungen liegen im Bereich des Vertrauensintervalls und beschreiben keine signifikante Abnahme des Migrationsanteils. Sie sind im Wesentlichen auf einen leichten Anstieg fehlender Angaben zum Geburtsland der Eltern im Fragebogen zurückzuführen. Schätzt man die fehlenden Werte unter Nutzung von verfügbaren Zusatzinformationen, liegt der Migrantenanteil in der PISA-Stichprobe von 2006 sogar bei rund 23 Prozent. Klemms Vermutung eines stichprobenbedingt überschätzten Trends lässt sich nicht aufrecht halten.

Die zweite Antwort betrifft die Frage, wie gut Zuwanderer in den PISA-Stichproben überhaupt repräsentiert sind und ob nicht die deutschen Leistungsergebnisse bei einer Unterrepräsentation dieser Gruppe in allen drei Zyklen besser ausfallen, als sie sind. Da wir über keine Populationsdaten verfügen, ist Klemms Gedanke, den Mikrozensus als Referenz zu nutzen, richtig. Denn die Stichprobenfehler sind im Mikrozensus klein und für eine Reihe von Merkmalen besteht Auskunftspflicht. Eine Reanalyse des Mikrozensus von 2006 weist für die 15-, 18- und 21-Jährigen – also die Geburtsjahrgänge, die in den PISA-Zyklen untersucht wurden – schwankende Migrantenanteile von 23 Prozent für die 2000er, von 22,6 Prozent für die 2003er und 24,4 Prozent für die 2006er PISA-Kohorte aus. Die Werte liegen über den PISA-Anteilen. Klemm übersieht jedoch, dass die Mikrozensus-Werte im Hinblick auf die PISA-Grundgesamtheit systematisch überschätzt sind. Zur PISA-Grundgesamtheit gehören alle 15-Jährigen, die tatsächlich noch eine Schule besuchen und körperlich, geistig und *sprachlich* in der Lage sind, am Test teilzunehmen. Alle Zuwanderer mit einer Verweildauer in Deutschland von unter einem Jahr werden automatisch ausgeschlossen. Insgesamt gehören 5 Prozent der 15-Jährigen nicht zur Grundgesamtheit von PISA. Diese Gruppe ist aber im Mikrozensus eingeschlossen und kann, da im Mikrozensus die Bildungsbeteiligung seit Jahren nicht mehr erhoben wird, auch nicht annähernd bestimmt werden. In dieser Gruppe sind Zuwanderer schon per definitionem überrepräsentiert. Berücksichtigt man dies, liegen die Migrantenquoten in PISA in verblüffender Nähe zum Mikrozensus. Wenn es in PISA überhaupt eine Unterschätzung der Migrantenanteile gibt, ist sie gering und in ihrem Einfluss auf die Schätzung von Kompetenzmittelwerten zu vernachlässigen.

Trendaussagen: Statistik und Politik

Für die Konsortien, die in Deutschland für die PISA-Berichte in der Verantwortung stehen, war die Sicherung der Vergleichbarkeit über die Erhebungsrunden und mit den OECD-Berichten eine der wichtigsten Verpflichtungen. Die Abstimmung mit der OECD ist 2006 nur unbefriedigend gelungen,

da die OECD ohne Information der Länder das Berichtskonzept in letzter Minute aus politischen Gründen geändert hat. Dies ist vermutlich der Aspekt der jüngsten Studie, der den stärksten politischen Wirbel verursacht hat und am wenigsten verstanden worden ist.

Tatsächlich ist es, wie oben gesagt, eine extreme und für die Forschung völlig neue Herausforderung, Trends über viele Staaten hinweg konsistent zu messen. Das internationale PISA-Konsortium unter Federführung australischer Experten ist nach bestem Ermessen zu dem Schluss gekommen, dass sich die Naturwissenschaftsaufgaben, die in allen bisherigen PISA-Zyklen eingesetzt wurden und als „Anker“ für Trendaussagen dienen, in den beteiligten Staaten keine einheitlichen Messeigenschaften besitzen: Ein Item mag in einem Staat etwas leichter und zugleich in einem anderen schwerer geworden sein. Daraufhin wurde entschieden, auf international vergleichende Trendaussagen völlig zu verzichten. Genauso korrekt war es jedoch, dass Wissenschaftler am IPN in Kiel Trends nur für Deutschland berechnet haben. Denn wenn man sich auf ein Land konzentriert und die Stabilität der Messeigenschaften der Ankeraufgaben kontrolliert, sind nationale Trendschätzungen gerechtfertigt. Und selbst wenn internationale Vergleiche möglich sind, sind nationale Trendberechnungen zu Kontrollzwecken absolut notwendig. Dies wurde im deutschen Bericht zu PISA 2006 offengelegt, aber offensichtlich nicht immer verstanden.

Maßgebliche Ursache der Verwirrung ist aber die Undiszipliniertheit des Sekretariats der OECD. Zum einen hat das Sekretariat auf politischen Druck einiger Teilnehmerstaaten, aber gegen den Rat der Wissenschaftler, in letzter Minute seinem Bericht eine unkommentierte Tabelle angefügt, die nun doch vergleichende Trends für die Naturwissenschaften berichtete. Auf diese Tabelle bezieht sich Klaus Klemm. Allerdings hat die OECD bis heute nicht offengelegt, wie diese Werte berechnet wurden, so dass sie nicht replizierbar sind. Zum zweiten haben OECD-Mitarbeiter die deutschen PISA-Ergebnisse und Trends in den Naturwissenschaften öffentlich kommentiert, bevor die Berichte überhaupt publiziert waren – so dass die Autoren des IPN keine Chance hatten, sich öffentlich zu äußern, wenn sie nicht gegen die Embargoregelung verstoßen wollten.

Sicher geben wir Klaus Klemm Recht, dass Kontinuität und Übersichtlichkeit der Auswertungen die Akzeptanz der PISA-Befunde in der Politik befördern. Das Beispiel der Trendaussagen zu den Naturwissenschaften belegt jedoch, wie schnell technisch-statistische Fragen und politische Interessen verwoben werden. Gerade deshalb werden wir uns auch in Zukunft mit aller Kraft um Präzision und Transparenz bemühen.

Allerdings werden wir auch damit leben müssen, dass in der Bildungspolitik, nicht nur in Wahlkampfzeiten, selektiv auf Befunde aus PISA zurückgegriffen wird. Um eine evidenzgesteuerte Bildungspolitik unterstützen zu können, brauchen wir eine Gemeinschaft empirischer Bildungs-

forscher, die sich überall deutlich gegen einen politisch selektiven und vor allem auch vereinfachenden Umgang mit PISA-Daten wendet. Der Diskurs – sei es in wissenschaftlichem Kreis oder in der interessierten Öffentlichkeit – kann da nur nützlich sein.