

Rainer Bölling

Beiträge zur Kritik an den PISA-Studien

Zur Fragwürdigkeit des PISA-Rankings (November 2017)	2
https://bildung-wissen.eu/fachbeitraege/zur-fragwuerdigkeit-des-pisa-rankings.html	
Die Mär von der sozialen Ungerechtigkeit	10
Frankfurter Allgemeine Zeitung vom 30. Mai 2018, S. 7	
PISA ist Politik (Interview)	13
Frankfurter Allgemeine Woche 1/2019, S. 25-27	
Schiefer als der schiefe Turm	17
Die PISA-Studie weist schwerwiegende Mängel auf	
Frankfurter Allgemeine Zeitung vom 28. November 2019, S. 6	
Wozu PISA?	22
Bildung aktuell 01/2020, S. 10-11	
Wie PISA Bildungsgerechtigkeit zu messen versucht	25
Februar 2020	
Bessere PISA-Ergebnisse durch selektives Testen (Februar 2021)	28
https://bildung-wissen.eu/kommentare/bessere-pisa-ergebnisse-durch-selektives-testen.html	
20 Jahre PISA – ein kritischer Rückblick (Dezember 2021)	31
https://www.schulverwaltung.de/bibliothek/recherche-im-werk/document/0/0:9985026,0	
Zweifel an der Sinnhaftigkeit der PISA-Daten	37
Frankfurter Allgemeine Zeitung vom 18. Januar 2024, S. 6	

Zur Fragwürdigkeit des PISA- Rankings

Seit im Dezember 2001 die erste PISA-Studie erschien¹, gelten die dabei erreichten Platzziffern der beteiligten Staaten der Öffentlichkeit gemeinhin als zuverlässige Belege bildungspolitischen Erfolgs bzw. Misserfolgs. Besonders heftig war die Reaktion auf die neuartigen Testergebnisse in Deutschland, das im ersten PISA-Triathlon aus Lesekompetenz, mathematischer und naturwissenschaftlicher Grundbildung nur den 20. Platz unter 31 teilnehmenden Ländern belegte. Nach dem viel beschworenen „PISA-Schock“ hat sich Deutschland in den folgenden fünf Testrunden allerdings deutlich verbessert, wogegen ehemalige Musterschüler wie Finnland und Schweden abgefallen sind. Inwieweit solche Verschiebungen im Klassement auf Veränderungen in den getesteten Schulsystemen zurückzuführen sind, ist jedoch eine offene Frage. Denn PISA selbst liefert ja keine Aussagen über die Ursachen veränderter Messergebnisse – was dazu verlaudet, sind in der Regel Interpretationen, die allzu oft auf vorgefertigten Sichtweisen beruhen. Doch nicht nur die Interpretation der PISA-Daten, sondern schon ihr Zustandekommen bzw. der Umgang der PISA-Verantwortlichen mit Problemen bei der Datenerhebung ist zuweilen äußerst fragwürdig. Das soll hier an einigen Beispielen aufgezeigt werden.

Das PISA-Ranking 2015

Das offizielle PISA-Ranking des Jahres 2015 (Anhang 1²) erinnert an Medaillenspiegel internationaler Sportereignisse, ist aber noch fragwürdiger als diese. Medaillenspiegel richten sich zunächst nach der Zahl der Goldmedaillen, berücksichtigen bei Gleichstand aber auch Silber- und Bronzemedailles. Das PISA-Ranking dagegen erfolgte 2015 allein nach der Durchschnittspunktzahl in Naturwissenschaften. Dabei erreichte beispielsweise **Deutschland** mit 509 Punkten denselben Wert wie das Vereinigte Königreich, wurde aber hinter diesem auf Platz 16 eingestuft, obwohl es im Bereich Lesen um 11, in Mathematik sogar um 14 Punkte besser war. Wenn es schon drei Testbereiche gibt, wäre es doch wohl sinnvoller, das Ranking nach dem Durchschnitt aus allen drei Bereichen vorzunehmen (Anhang 2³). Dann landet Deutschland auf Platz 13, das Vereinigte Königreich dagegen nur auf Platz 23. Das offizielle Ranking der OECD liefert also ein irreführendes Bild.

Noch krasser ist der Fall **Vietnam**, das 2015 im offiziellen Ranking wegen seiner 525 Punkte in Naturwissenschaften auf Platz 8 steht. Ginge es nach der durchschnittlichen Punktzahl, käme

¹ Lernen für das Leben. Erste Ergebnisse der internationalen Schulleistungsschule PISA 2000, OECD 2001.

² Ursprünglich http://www.oecd.org/berlin/themen/pisa-studie/PISA_Snapshot.gif (06.12.2016), jetzt unter <https://img.4plebs.org/boards/pol/image/1481/02/1481021564586.gif>. (11.11.2017).

³ Punktzahlen wie Fn. 2, Erfassungsquoten nach: OECD, PISA 2015 Ergebnisse (Band I): Exzellenz und Chancengerechtigkeit in der Bildung, W. Bertelsmann Verlag 2016, S. 315.

das Land nur auf Platz 21. Noch gravierender ist, dass in Vietnam nur 48,5 Prozent der nationalen Stichprobe vom PISA-Test erfasst wurden, während ansonsten in den überdurchschnittlich abscheidenden Ländern mehr als 90 Prozent üblich sind (Deutschland 96,1 %). Laut OECD deuten nahe bei 100 liegende Werte darauf hin, „dass die PISA-Stichprobe das gesamte Schulsystem nach Definition für PISA 2015 repräsentiert“. Wenn man annehme, „dass die ausgeschlossenen Schülerinnen und Schüler durchgehend schlechtere Ergebnisse erzielt hätten als die teilnehmenden und dass diese Korrelation mittelstark ausgeprägt ist, würde eine Ausschlussrate in einer Größenordnung von 5 % wahrscheinlich nur zu einer Überzeichnung der nationalen Durchschnittsergebnisse um weniger als 5 Punkte führen“.⁴ Das bedeutet aber, dass eine Ausschlussrate von über 50 % zu einer Überzeichnung von bis zu 50 Punkten geführt haben könnte. Damit würde Vietnam im Ranking auf Basis der Durchschnittspunktzahl auf einen Platz unterhalb der 40 abstürzen – es sei denn, dass die der Berechnung der OECD zugrundeliegende Zahl von 1,8 Millionen 15-Jährigen in Vietnam falsch ist. Das wiederum erscheint denkbar, weil deren Anteil im Verhältnis zur Gesamtbevölkerung doppelt so hoch liegt wie in vielen anderen Ländern. Auf jeden Fall sind die Daten zu Vietnam in dieser Form unbrauchbar.

PISA und China

Das Problem einer zu niedrigen Erfassungsquote stellte sich schon bei **Shanghai**, das erstmals 2009 bei PISA auftauchte und sogleich – wie auch 2012 – im Ranking mit großem Abstand an der Spitze lag. In einer Publikation der OECD wurde es 2011 zusammen mit Hongkong als Vorbild gepriesen, das ein inklusives Schulsystem mit hohem Leistungsanspruch verbinde.⁵ Der PISA-Test war in 12 chinesischen Provinzen durchgeführt worden, doch von der OECD wurden ohne Begründung nur die Ergebnisse Shanghais veröffentlicht.⁶ Schon früh zweifelten Autoren in den USA ihren Wert an. Hauptkritikpunkt war, dass offensichtlich die Kinder der armen Wanderarbeiter, die 2012 fast zehn der 24 Millionen Einwohner Shanghais stellten⁷, vom PISA-Test nicht erfasst wurden.⁸ Denn der Besuch einer höheren Schule, an deren Beginn der Test durchgeführt wurde, war den Migrantenkindern versagt. Daher ist der Anteil der bei PISA erfassten Fünfzehnjährigen (2012: 108.056) im Verhältnis zur Gesamtbevölkerung Shanghais

⁴ OECD, PISA 2015 Ergebnisse, S. 318.

⁵ OECD, Strong Performers and Successful Reformers in Education. Lessons from PISA for the United States, OECD Publishing 2011, S. 83-115.

⁶ Tom Loveless, PISA's China Problem, 09.10.2013 (<https://www.brookings.edu/research/pisas-china-problem>)

⁷ China Daily, 27.02.2014 (http://www.chinadaily.com.cn/china/2014-02/27/content_17311272.htm).

⁸ Tom Loveless, Attention OECD-PISA: Your-Silence-on-China-is-Wrong, 12.12.2013 (<http://educationnext.org/attention-oecd-pisa-your-silence-on-china-is-wrong>); Valerie Strauss, Did Shanghai cheat on PISA? 15.12.2013 (<http://www.washingtonpost.com/blogs/answer-sheet/wp/2013/12/15/did-shanghai-cheat-on-pisa>)

nur halb so hoch wie in anderen Ländern. Und selbst auf dieser Basis lag die Erfassungsquote 2012 nur bei 79 Prozent.⁹ Erst seit 2013 können Wanderarbeiter in Shanghai einen Einwohnerstatus erlangen, der ihren Kindern Zugang zum höheren Schulwesen eröffnet.¹⁰

Bei PISA 2015 erscheint Shanghai nicht mehr allein, sondern zusammen mit Peking, Jiangsu und Guangdong als Teil einer neuen Einheit (P-S-J-G) mit 223 Millionen Einwohnern, die China repräsentieren soll. Sie kommt auf Platz 10 des offiziellen Rankings wie auch eines Rankings nach Durchschnittspunktzahl. Ihre Erfassungsquote liegt aber nur bei 64 Prozent, so dass es laut OECD zu einer Überzeichnung von bis zu 35 Punkten gekommen sein könnte. Das könnte einen Absturz auf einen Platz unterhalb der 30 bedeuten.

Darüber hinaus ist das chinesische Schulwesen in einem in Europa kaum vorstellbaren Maße auf Abschlusstests fokussiert und unterwirft die Schüler einem gnadenlosen Paukstress, der im konfuzianischen Kulturkreis Tradition hat, aber auch ein Kennzeichen diktatorischer Regime ist. Darauf hat besonders der an der Universität von Oregon lehrende chinesische Bildungsforscher Yong Zhao hingewiesen, der in den USA zu den wichtigsten PISA-Kritikern gehört.¹¹

PISA und die USA

Umgekehrt kann das Ergebnis eines Landes auch negativ verzerrt werden, wenn in der nationalen PISA-Stichprobe die 15-Jährigen aus unteren sozialen Schichten, die in der Regel schlechtere Testergebnisse liefern, überrepräsentiert sind. Dies war der zentrale Kritikpunkt einer Studie, die *Martin Carnoy* und *Richard Rothstein* vom Economic Policy Institute in Washington zu den PISA-Ergebnissen der USA von 2009 vorgelegt haben. Nach ihren korrigierten Punktzahlen hätten die USA im Ranking um zehn und mehr Plätze besser abgeschnitten.¹² Über Berechtigung und Tragweite ihrer Korrekturen kam es zu einer Kontroverse mit dem PISA-Koordinator Andreas Schleicher, die hier nicht im Einzelnen wiedergegeben werden soll. Auf jeden Fall ließ sie deutlich werden, welche Probleme allein schon die korrekte Bestimmung einer nationalen Stichprobe aufwirft.

⁹ OCDE, PISA 2012 Ergebnisse: Was Schülerinnen und Schüler wissen und können. Band I (Überarbeitete Ausgabe, Februar 2014): Schülerleistungen in Lesekompetenz, Mathematik und Naturwissenschaften, W. Bertelsmann Verlag 2014, S. 290 f.

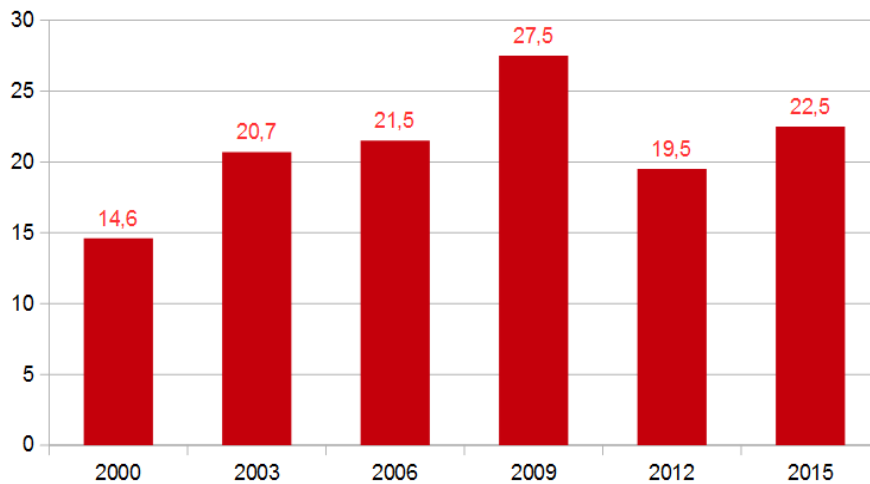
¹⁰ Vgl. Shanghai Daily, 07.06.2013 (<http://en.people.cn/203691/8275550.html>).

¹¹ Siehe Yong Zhao, Who's Afraid of the Big Bad Dragon? Why China Has the Best (and Worst) Education System in the World, 2014. Dazu die Rezension von Diane Ravitch, The Myth of Chinese Super Schools, 20.11.2014 (<http://www.nybooks.com/articles/2014/11/20/myth-chinese-super-schools>)

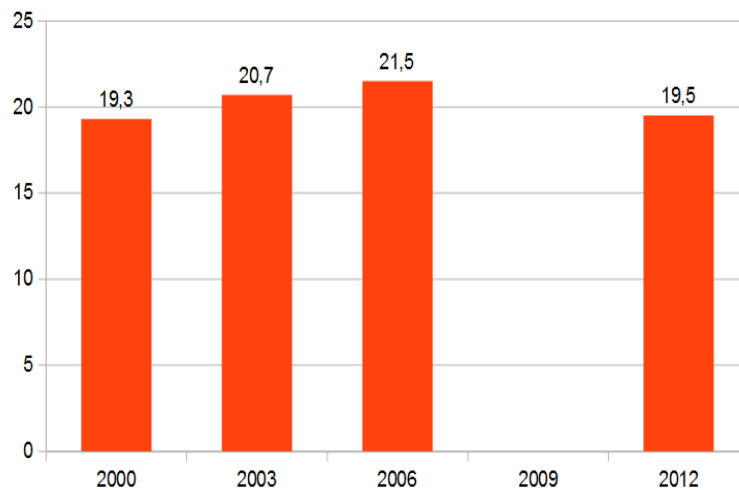
¹² Martin Carnoy/Richard Rothstein, What do international tests really show about U.S. Student performance? Washington 2013 (<http://www.epi.org/publication/us-student-performance-testing>). Dazu Response from Martin Carnoy and Richard Rothstein to OECD/PISA Comments (<http://www.epi.org/files/2013/EPI-Carnoy-Rothstein-Resp-to-Schleicher.pdf>)

PISA und Österreich

Die Medaillenspiegel großer Sportereignisse werden oft nachträglich durch den Ausschluss gedopter Sportler geändert. Ähnlich ist es auch bei PISA, wie das Beispiel Österreichs zeigt. Dort hat sich der Anteil der Schüler, die auf der Gesamtskala Lesekompetenz unter Stufe 2 blieben („Risikoschüler“), den ursprünglichen PISA-Publikationen zufolge so entwickelt:



In einer Tabelle in der PISA-Publikation 2012 erscheinen jedoch folgende Daten:¹³



Hier ist der Wert für 2000 stillschweigend korrigiert worden, während der für 2009 als *fehlend* (missing) bezeichnet wird, obwohl er drei Jahre vorher veröffentlicht wurde. Was ist der Hintergrund dieser Korrekturen?

¹³ PISA 2012 Ergebnisse (Fn. 9), S. 400.

In Österreich fiel PISA 2003 schlechter aus als drei Jahre zuvor, was wegen eines zwischenzeitlichen Regierungswechsels zu einer politischen Kontroverse führte. Eine 2006 erschienene Überprüfung durch versierte Statistiker¹⁴ förderte Ungereimtheiten bei der Stichprobenziehung und Datenauswertung zutage. Insbesondere war 2000 die Stichprobe in den Berufsschulen nicht korrekt gezogen worden, so dass die Ergebnisse zu gut ausfielen. Die OECD hatte die Auswirkungen des Stichprobenfehlers im *Technical Report* von 2000 zunächst als sehr gering bezeichnet¹⁵ und die fragwürdigen Daten ohne Vorbehalt veröffentlicht. 2007 aber übernahm sie nahezu unbeachtet die Ergebnisse der neuen Studie und korrigierte damit die österreichischen Ergebnisse aus 2000 offiziell nach unten.¹⁶

Zu den Daten von 2009 bemerkt die OECD im Anhang des grundlegenden Bandes: „In Österreich kam es aus Anlass der PISA-Erhebung 2009 zu einem Streit zwischen den Lehrgewerkschaften und der Bildungsministerin, der zur Ankündigung eines PISA-Boycott führte, welcher dann aber nach der ersten Testwoche zurückgenommen wurde. Auf Grund des Boykotts musste die OECD identifizierbare Fälle aus der Datenreihe entfernen. Wenngleich die österreichischen Daten nach der Entfernung dieser Fälle die technischen Standards von PISA 2009 erfüllten, hat die negative Stimmung gegenüber Bildungserhebungen die Testbedingungen belastet und könnte die Motivation der Schüler zur Beantwortung der PISA-Aufgaben ungünstig beeinflusst haben. Die Vergleichbarkeit der Erhebungsdaten des Jahres 2009 mit denen früherer PISA-Erhebungen kann dementsprechend nicht gewährleistet werden, weshalb die österreichischen Daten aus den Trendvergleichen ausgeschlossen wurden.“¹⁷ In die tabellarische Darstellung der PISA-Ergebnisse des Jahres 2009 wurden die fraglichen Daten aber ohne Einschränkung übernommen.

Wer in den Trendvergleichen im Anhang des PISA-Bandes 2012 (S. 400) auf das ominöse „m“ für *missing* stößt, kann es nur verstehen, wenn er zufällig auf Seite 58 f. denselben Hinweis wie zu 2009 gelesen und in Erinnerung behalten hat. Dass unklare oder fehlende Werte in Tabellen und darauf bezogene, oft folgenlose Anmerkungen sich ohne Querverweise an weit voneinander entfernten Stellen der mehrbändigen PISA-Publikationen finden, scheint Methode zu haben. So kann PISA-Koordinator Schleicher seinen Kritikern leichter vorhalten, die Publikation nicht richtig gelesen zu haben.¹⁸

¹⁴ E. Neuwirth, I. Ponocny, W. Grossmann (Hrsg.): PISA 2000 und PISA 2003: Vertiefende Analysen und Beiträge zur Methodik, Graz 2006.

¹⁵ OECD, PISA 2000. Technical Report, OECD 2002, S. 183.

¹⁶ E. Neuwirth, PISA 2000: Sample Weight Problems in Austria, OECD Education Working Papers, No. 5, OECD: Paris 2006 (http://www.oecd-ilibrary.org/pisa-2000_514rvsv9r5d4.pdf). Der Standard, 13.12.2007 (<http://derstandard.at/3126573/Stichprobenfehler-verursachte-PISA-Absturz>).

¹⁷ OECD (2010), PISA 2009 Ergebnisse: Was Schülerinnen und Schüler wissen und können. Schülerleistungen in Lesekompetenz, Mathematik und Naturwissenschaften (Band I), S. 201.

¹⁸ So am 20.05.2014 im DeutschlandradioKultur als Antwort auf einen offenen Brief (<http://bildung-wissen.eu/wp-content/uploads/2014/05/offener-brief-schleicher-autorisierte-fassung.pdf>), der von mehreren

Noch gravierender aber ist, dass Schleicher die verzerrten PISA-Resultate Österreichs von 2009 zum Anlass genommen hat, der österreichischen Bundesregierung den Rat zu geben, „doch endlich auf die gemeinsame Schule der Zehn- bis Vierzehnjährigen umzusteigen, um die Testleistungen zu verbessern“. ¹⁹ Wie sich gezeigt hat, beruhte dieser Rat auf dubiosen Testergebnissen, von denen die OECD heute nichts mehr wissen will. Umso mehr sind verantwortliche Regierungen gut beraten, den Ratschlägen dieses Statistikers, der sich als Bildungsfachmann geriert, mit größter Skepsis zu begegnen.

hundert Lehrern und Bildungsforschern in den USA und Europa unterzeichnet wurde (http://www.deutschlandradiokultur.de/schule-substanzlose-behauptungen.954.de.html?dram:article_id=286897).

¹⁹ Konrad Paul Liessmann, Was der Glaube an Statistiken bewirken kann – Eine Nachlese zu Pisa, in: Frankfurter Allgemeine Zeitung vom 23.12.2010, S. 6.

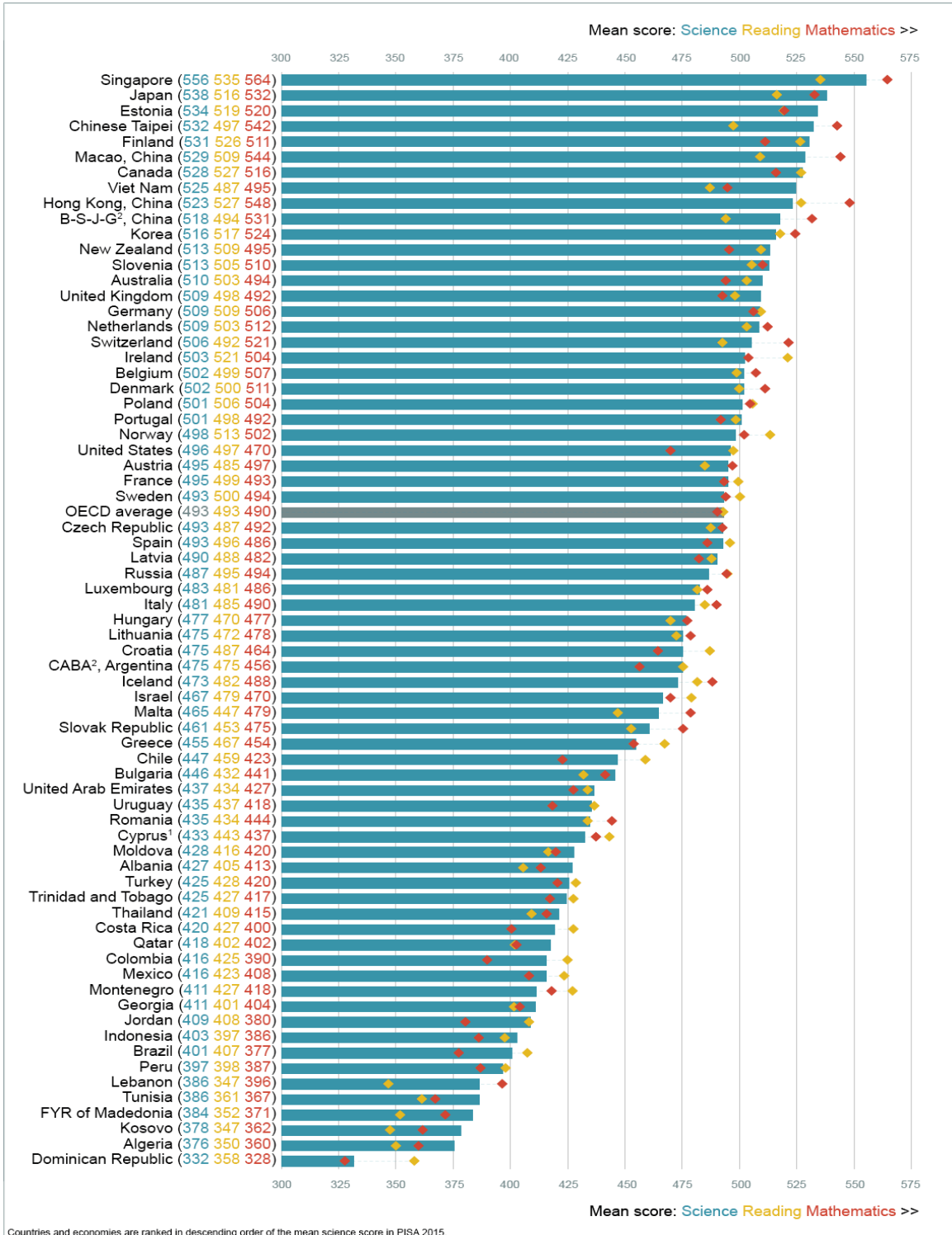
Anhang 1



PISA 2015 results



Snapshot of performance in science, reading and mathematics



Anhang 2: PISA-Ranking 2015 nach Durchschnittspunktzahlen

PISA-Rang	Land	Erfassungsquote	Naturwissenschaften	Lesen	Mathematik	Durchschnitt	neuer Rang
1	Singapur	95,9	556	535	564	551,7	1
9	Hong Kong, China	88,6	523	527	548	532,7	2
2	Japan	94,7	538	516	532	528,7	3
6	Macau, China	88,4	529	509	544	527,3	4
3	Estland	92,8	534	519	520	524,3	5
4	Chinesisch Taipeh	85,2	532	497	542	523,7	6
7	Kanada	83,5	528	527	516	523,7	6
5	Finnland	97,3	531	526	511	522,7	8
11	Korea	91,7	516	517	524	519,0	9
10	B-S-J-G, China	63,9	518	494	531	514,3	10
13	Slowenien	92,8	513	505	510	509,3	11
19	Irland	96,5	503	521	504	509,3	11
16	Deutschland	96,1	509	509	506	508,0	13
17	Niederlande	95,1	509	503	512	508,0	13
18	Schweiz	96,2	506	492	521	506,3	15
12	Neuseeland	90,2	513	509	495	505,7	16
21	Dänemark	89,0	502	500	511	504,3	17
24	Norwegen	91,3	498	513	502	504,3	17
22	Polen	90,9	501	505	504	503,3	19
20	Belgien	92,9	502	499	507	502,7	20
8	Vietnam	48,5	525	487	495	502,3	21
14	Australien	90,6	510	503	494	502,3	21
15	United Kingdom	84,0	509	498	492	499,7	23
23	Portugal	87,6	501	498	492	497,0	24
27	Frankreich	91,0	495	499	493	495,7	25
28	Schweden	93,6	493	500	494	495,7	25
26	Österreich	83,4	495	485	497	492,3	27
32	Russ. Föderation	95,3	487	495	494	492,0	28
30	Spanien	90,9	493	496	486	491,7	29
29	Tschech. Rep.	93,5	493	487	492	490,7	30
25	USA	83,5	496	497	470	487,7	31
31	Lettland	88,8	490	488	482	486,7	32
34	Italien	80,3	481	485	490	485,3	33
33	Luxemburg	87,6	483	481	486	483,3	34
39	Island	93,3	473	482	488	481,0	35
37	Kroatien	90,8	475	487	464	475,3	36
36	Litauen	90,2	475	472	478	475,0	37
35	Ungarn	89,6	477	470	477	474,7	38
40	Israel	93,7	467	479	470	472,0	39
38	CABA (Argentinien)	55,0	475	475	456	468,7	40
41	Malta	97,7	465	447	479	463,7	41
42	Slowak. Rep.	89,2	461	453	475	463,0	42
43	Griechenland	91,1	455	467	454	458,7	43
44	Chile	79,8	447	459	423	443,0	44
45	Bulgarien	80,6	446	432	441	439,7	45
48	Rumänien	93,1	435	434	444	437,7	46
49	Zypern	94,9	433	443	437	437,7	46
46	VAE	90,8	437	434	427	432,7	48
47	Uruguay	71,5	435	437	418	430,0	49
52	Türkei	69,9	425	428	420	424,3	50

Die Mär von der sozialen Ungerechtigkeit

Wer zu Hause mit Büchern aufgewachsen ist, hat es leichter im Bildungssystem.

Das ist nicht nur in Deutschland so, auch wenn die OECD das
überhaupt nicht wahrhaben will.

Wie ungerecht geht es im deutschen Bildungswesen zu? Wenn man Verlautbarungen der Organisation für wirtschaftliche Zusammenarbeit und Entwicklung (OECD) aus den letzten 15 Jahren Revue passieren lässt, scheint die Antwort klar zu sein. In kaum einem anderen Land der Welt hängt der Schulerfolg so sehr von der sozialen Herkunft ab wie in Deutschland. So jedenfalls haben es OECD-Repräsentanten wie Andreas Schleicher unermüdlich verkündet.

Anfang dieses Jahres nun erschien eine Studie der OECD, die Deutschland überraschend Fortschritte bescheinigte, und zwar bei der Resilienz. Mit diesem aus der Psychologie übernommenen Begriff bezeichnen die Autoren die „Fähigkeit von Schülerinnen und Schülern, trotz sozialer Nachteile in allen PISA-Testfeldern mindestens die Kompetenzstufe drei (von sechs) zu erreichen und damit die Voraussetzung für eine aktive gesellschaftliche Teilhabe und lebenslanges Lernen zu erwerben“.²⁰ Konnte 2006 nur jeder vierte sozial benachteiligte deutsche Schüler solche Kompetenzen vorweisen, so war es 2015 schon jeder dritte. Nach diesem Rekordzuwachs liegt Deutschland jetzt über dem Durchschnitt der OECD-Länder. Man könnte also denken, nun habe die Bildungsgerechtigkeit gesiegt.

Doch bevor Feierstimmung aufkommen kann, bekräftigt die OECD in der gemeinsam mit der Vodafone-Stiftung erstellten deutschen Ausgabe der Studie, in Deutschland sei trotz aller Fortschritte „der statistische Zusammenhang zwischen Leistung und sozialer Herkunft noch immer sehr ausgeprägt“.²¹ Das war gar kein Ergebnis dieser Studie, die auf einer Sonderauswertung der PISA-Daten von 2015 beruht. Wie der daran beteiligte Pressesprecher der OECD in Deutschland erklärte, wurde die Aussage eingefügt, um „auf die Situation zur Chancengleichheit insgesamt hinzuweisen“.²² Diese Vorlage nahm dann ein Teil der Presse dankbar auf und verarbeitete sie weiter. Im Internetportal „Spiegel online“ lautete der erste Satz des einschlägigen Artikels: „Noch immer entscheidet vor allem die soziale Herkunft darüber, wie gut Kinder in der Schule klarkommen.“²³ Und die in Düsseldorf erscheinende „Rheinische Post“ titelte zwar: „Benachteiligte Schüler holen auf“, setzte aber gleich hinzu: „Bei der Chancengleichheit liegt Deutschland dennoch hinten.“²⁴

²⁰ Agasisti, T. *et al.* (2018), „Academic resilience: What schools and countries do to help disadvantaged students succeed in PISA“, *OECD Education Working Papers*, No. 167, OECD Publishing, Paris, S. 15.

²¹ Erfolgsfaktor Resilienz, 2018, S. 6.

²² E-Mail von Matthias Rumpf an den Verf. vom 02.02.2018.

²³ <http://www.spiegel.de/lebenundlernen/schule/pisa-studie-wie-sozial-benachteiligte-schueler-erfolgreich-werden-a-1190309.html>.

²⁴ Rheinische Post vom 30.01.2018.

Auf welch dünnem Eis diese Aussagen stehen, war den Autoren der genannten Artikel wohl kaum bewusst, denn die neue Studie bietet keine Belege für die von der OECD verbreitete Deutung. Wer ihre statistischen Grundlagen nachprüfen will, muss sich schon die Mühe machen, die Publikation der Pisa-Studie 2015 heranzuziehen. Dort findet man in einem „Überblick über die Bildungsgerechtigkeit“²⁵ als Indikatoren neben der schon genannten Resilienz den „Prozentsatz der durch den sozioökonomischen Status der Schüler erklärten Leistungsvarianz“. Der sozioökonomische Status wird mittels des PISA-Index des wirtschaftlichen, sozialen und kulturellen Status (ESCS) geschätzt, der aus mehreren Variablen des familiären Hintergrunds der Schülerinnen und Schüler abgeleitet wird: dem Bildungsniveau und dem Beruf der Eltern, der Ausstattung des Elternhauses als Hilfsindikator für den materiellen Wohlstand sowie der Zahl der Bücher und anderer Bildungsressourcen, die im Elternhaus verfügbar sind.²⁶

Das komplizierte Zustandekommen des Index kann und soll hier nicht überprüft werden; es genügt, die Ergebnisse in den Blick zu nehmen. Zunächst einmal ist festzuhalten, dass die soziale Herkunft in allen untersuchten Ländern Einfluss auf den Bildungserfolg hat. Im Durchschnitt der OECD, der bei Pisa als Maß aller Dinge gilt, sind 13 Prozent der Leistungsunterschiede dadurch zu erklären. Für Werte zwischen 11 und 15 Prozent gilt, dass sie nicht statistisch signifikant vom Durchschnitt abweichen; in dieser Bandbreite liegen 26 Länder. Weitere 26 bleiben mit Werten von 1 bis 10 Prozent unter dem Durchschnitt, können also nach Pisa-Kriterien eine höhere Bildungsgerechtigkeit für sich in Anspruch nehmen. Deutschland muss demnach zu den 17 Ländern gehören, die mit Werten von 16 bis 26 Prozent über dem Durchschnitt liegen. Wenn hier laut OECD der Zusammenhang zwischen Leistung und sozialer Herkunft „sehr ausgeprägt“ ist, deutet das für die meisten Betrachter der Statistik auf einen Wert um oder über 20. Doch weit gefehlt: Deutschland liegt wie Österreich, die Schweiz und vier weitere Länder mit einem Wert von 16 Prozent nur schwach signifikant über dem Durchschnitt der OECD. Wenn das sehr ausgeprägt sein soll, wie würde die OECD dann die deutlich höheren Werte anderer Länder bezeichnen? Auf diese Frage blieb der Pressesprecher der OECD bezeichnenderweise die Antwort schuldig.

Bemerkenswert ist auch, dass der aktuelle PISA-Sieger Singapur mit einem Wert von 17 in diesem Punkt schlechter dasteht als die deutschsprachigen Länder, ganz zu schweigen von Frankreich mit 20 Prozent. Auf der anderen Seite fällt auf, dass von den 26 Ländern mit - laut PISA - hoher Bildungsgerechtigkeit 16 nur unterdurchschnittliche Leistungen in Naturwissenschaften erreichen, 13 davon sogar weniger als 440 Leistungspunkte bei einem Durchschnitt von 493. Sie gehören mit Abstand zum unteren Drittel der Leistungsskala. Über das nach Pisa-Kriterien „gerechteste“ Bildungswesen verfügt Algerien, das aber auf der

²⁵ PISA 2015 Ergebnisse (Band I): Exzellenz und Chancengerechtigkeit in der Bildung, 2016, S. 52 (Abb. I.1.3).

²⁶ Ebd., S. 222.

Leistungsskala mit 376 Punkten den vorletzten Platz belegt. Hier tun sich statistische Zusammenhänge und Abgründe auf, bei denen man gern wüsste, wie die OECD sie interpretieren würde.

Auf Basis derselben Pisa-Daten könnte man Deutschlands Abschneiden auch so formulieren: „2015 waren die Leistungsunterschiede deutscher Schülerinnen und Schüler zu 84 Prozent unabhängig von ihrer sozialen Herkunft. In puncto Bildungsgerechtigkeit schnitt Deutschland damit besser ab als PISA-Sieger Singapur.“ Das klingt ganz anders als die Formulierung der OECD, die schlicht und einfach irreführend ist. Offensichtlich geht es der Organisation bei der Interpretation ihres Zahlenmaterials weniger um seriöse wissenschaftliche Erkenntnis als um griffige Statements für die Öffentlichkeit, mit denen sie Politik macht. Die Behauptung, man gebe nur Tatsachen wieder²⁷, ist pure Verschleierung.

Neben der tendenziösen Darstellung statistischer Befunde gibt es auch Fälle, in denen selbst die Daten fehlerhaft sind. Das gilt etwa für die Ergebnisse der PISA-Erhebung 2009 in Österreich, die deutlich schlechter ausfielen als zuvor. Daraufhin empfahl PISA-Koordinator Schleicher der österreichischen Bundesregierung, „doch endlich auf die gemeinsame Schule der Zehn- bis Vierzehnjährigen umzusteigen, um die Testleistungen zu verbessern“.²⁸ Diesen Rat bekamen auch Deutschland und die Schweiz immer wieder zu hören, deren gegliedertes Schulwesen Schleicher ein Dorn im Auge ist.

Drei Jahre später war in übergreifenden Trendvergleichen der Pisa-Publikation 2012 anstelle der ungünstigen Daten für 2009 nur noch die Angabe „m“ für „missing“ zu finden – eine in OECD-Publikationen gar nicht so seltene Erscheinung. Was sich dahinter verbirgt, lässt sich nur einer Erklärung entnehmen, die 340 Druckseiten vorher zu finden war. Demnach waren die österreichischen Daten von 2009 entgegen vorhergehenden Beteuerungen infolge eines Boykotts der Lehrgewerkschaften nicht brauchbar.²⁹ Mit solchen Pannen offen umzugehen widerspricht offenbar dem Unfehlbarkeitsanspruch der PISA-Macher. Umso skeptischer sollten Politik und Öffentlichkeit deren Verlautbarungen und Empfehlungen begegnen.

²⁷ E-Mail von Matthias Rumpf an den Verf. vom 12.02.2018

²⁸ Konrad Paul Liessmann, Was der Glaube an Statistiken bewirken kann – Eine Nachlese zu Pisa, in: Frankfurter Allgemeine Zeitung vom 23.12.2010, S. 6.

²⁹ PISA 2012 Ergebnisse: Was Schülerinnen und Schüler wissen und können, Band I, überarbeitete Ausgabe, Februar 2014, S. 58 f. u. 400.

„Pisa ist Politik“

Das deutsche Schulsystem hat einen schlechten Ruf. Der Bildungsforscher Rainer Bölling hat die Schuldigen dafür ausgemacht. Besonders eine Organisation gerät in die Kritik.

Herr Bölling, steuert Deutschland angesichts des Lehrermangels auf eine Bildungskatastrophe zu?

Mit dem Begriff "Bildungskatastrophe" bin ich vorsichtig, der wird immer etwas überstrapaziert. Was den Lehrermangel betrifft - das ist wirklich das Problem der Zukunft des deutschen Bildungswesens. Denn das trifft Deutschland ziemlich unvorbereitet, obwohl es anders sein müsste. Dass die Flüchtlingswelle kommen würde, konnte man natürlich nicht vorhersehen, aber die Geburtenentwicklung haben die Kultusministerien und auch die KMK einerseits verschlafen, aber sich vorher auch schon falsche Vorstellungen darüber gemacht.

Inwiefern falsche Vorstellungen?

Es gab die Vorstellung, dass die Schülerzahlen zurückgehen und dass man den vorhandenen Schülerberg nur sozusagen untertunneln müsse. Das hieß, dass man etwas weniger Lehrer einstellte, als aktuell benötigt wurden, weil man glaubte, in der Zukunft ohnehin weniger Lehrer zu brauchen. Da ist in vielen Ländern etwas falsch gelaufen.

Viele Eltern verstehen unter dem Stichwort "Bildungskatastrophe" ja erst einmal vor allem Lehrmethoden, die es früher nicht gab, zum Beispiel das Schreibenlernen nach Gehör in den ersten beiden Grundschuljahren, also nicht von Anfang an die richtige Schreibweise zu unterrichten. Was halten Sie davon?

Ehrlich gesagt, nicht viel. Ich bin kein Grundschullehrer, sondern schließe das aus Gesprächen mit Lehrern oder eben Eltern, aber nach allem, was ich mitbekomme, ist es schon ein Problem, zu sagen: Man lernt es erst so, und dann lernt man um. Eltern wird dann gesagt, dass sie ihre Kinder nicht korrigieren sollen. Aber wie soll das laufen, wenn es zum Beispiel noch ein älteres Kind im Haushalt gibt? Wie verhindert man, dass das seinen kleinen Geschwistern sagt: ‚Seid ihr doof, das schreibt man doch ganz anders?‘ Überhaupt, die Eltern dazu zu verdonnern, Fehler ihrer Kinder nicht zu verbessern, das finde ich abwegig. Und dass die Rechtschreibfähigkeiten dadurch nachlassen, sieht man ja jetzt an aktuellen Studien.

Was sagen Sie Eltern, die ihre Kinder nicht auf die örtliche Grundschule schicken wollen, weil ihnen dort der Ausländeranteil zu hoch erscheint?

Um das zu beurteilen, müsste man die Verhältnisse vor Ort sehr genau kennen. Es gibt in vielen großen Städten ja wirklich Schulen, wo deutsche Kinder in der Minderheit sind, und da kann ich Eltern schon verstehen, wenn sie das für ihre Kinder nicht wollen. Früher war das eher eine Frage der sozialen Distanzierung gutsituierter Eltern, die nur einen ganz bestimmten Umgang für ihre Kinder wünschten. Heute geht es oft auch um pädagogische Konzepte, wenn Eltern ihr

Kind auf andere Schulen oder Privatschulen geben.

Wie durchlässig ist das Bildungssystem heute, und welche Rolle spielt der soziale Status der Eltern?

Die in Deutschland angeblich besonders hohe Abhängigkeit des Bildungserfolgs vom Elternhaus ist mittlerweile zu einem Topos geworden, der kaum noch hinterfragt wird. Die Pisa-Studien waren die Hauptquelle für diese These, aber selbst sie belegen, dass man diesen Zusammenhang in allen Ländern beobachten kann, mit nur leichten Unterschieden. Dazu muss man wissen, dass die Pisa-Erhebungen zum sozialen Hintergrund der Eltern auf Selbstangaben der Schüler beruhen. In einem Zusatzfragebogen wurden da grobe Angaben zum Beruf, der Vorbildung der Eltern, zur Zahl der Bücher im Haushalt abgefragt. Das sind also keine sehr harten Daten. Wenn das Kind zum Beispiel schreibt: ‚Mein Vater arbeitet in einem Computerladen‘, was ist der dann genau? Der Inhaber? Oder ein Angestellter? Das wissen die Kinder manchmal gar nicht. Schon diese Daten sind also mit Vorsicht zu genießen. Das Problem ist, wenn diese Angaben dann nach allen Regeln der Statistik verarbeitet werden, als handle es sich um handfeste naturwissenschaftliche Daten, und dann aus Prozenten und Prozentbruchteilen weitreichende Schlüsse gezogen werden.

Was bedeutet das für den deutschen Ranglistenplatz?

Deutschland hat bei Pisa 2003 noch unterdurchschnittlich abgeschnitten und sich dann stetig verbessert. 2012 lag es schon im OECD-Durchschnitt zur Bildungsgerechtigkeit.

Das ging ja schnell!

Ja, das war ein Sprung von 24 Prozent im Jahr 2003 auf 16,9 Prozent im Jahr 2012. Knapp 17 Prozent der Leistungsunterschiede bei Schülern wurden also 2012 auf die soziale Stellung der Eltern zurückgeführt. 2015 waren wir dann runter auf 16 Prozent, genau so wie Österreich und die Schweiz - und besser als Singapur, der Pisa-Sieger von 2015, besser als China, das von der OECD immer sehr hervorgehoben wird, und viel besser als Frankreich, wo der Wert bei 20 Prozent liegt. Man muss aber auch sehen, dass es hier nur um Werte von zehn bis 20 Prozent der Leistungsunterschiede geht, die auf das Elternhaus zurückgeführt werden, also eigentlich nicht sehr viel.

Was schließt die OECD daraus?

Das Interessante ist, wie die OECD so etwas sprachlich aufbläst. Man sagt dann, in Deutschland sei dieser Zusammenhang weiterhin sehr ausgeprägt. Das ist eine wertende Aussage, unter der man sich Unterschiedliches vorstellen kann. Der Pressesprecher der OECD behauptet zwar, seine Organisation berichte nur Tatsachen, aber das ist blanker Unsinn.

Was ist denn so schlimm am elterlichen Einfluss?

Eltern mit höherer Bildung kümmern sich in der Regel auch mehr um die Bildung ihrer Kinder, und die haben dadurch Vorteile. Natürlich sollte das Bildungssystem versuchen, das auszugleichen, nur darf man nicht erwarten, dass man den Bildungsfaktor Eltern völlig neutralisieren könnte. Genau das ist aber die dahinterstehende Vorstellung: Bildungsgerechtigkeit besteht, wenn alle Kinder die gleichen Ergebnisse erzielen.

Wie erklärt die OECD die Verbesserung in Deutschland?

Sie weist gerne darauf hin, dass bildungspolitische Maßnahmen, die nach Pisa ergriffen wurden, zu diesen Verbesserungen geführt hätten, zum Beispiel der Kita-Ausbau. Das geht aber schon rein rechnerisch nicht. 2012 wurden Fünfzehnjährige getestet, sie waren also spätestens 2003 in die Schule gekommen. Die erste Pisa-Studie erschien Ende 2001, die empfohlenen Maßnahmen konnten frühestens ein, zwei Jahre danach ergriffen werden. Das bedeutet, dass keiner der 2012 Getesteten aufgrund von Pisa mehr frühkindliche Bildung gehabt haben kann. Aber dieser Zusammenhang wird von Pisa-Forschern trotzdem so hergestellt. Und das wird dann von vielen übernommen und nachgeredet. Der Eindruck verstärkt sich mir, dass es unter Bildungsforschern eine Bereitschaft gibt, diese Mainstream-Behauptungen, die gerade von der OECD herausgebracht werden, kritiklos weiterzugeben.

Sie üben am Pisa-Rankingverfahren auch grundsätzlich Kritik?

Das Problem ist, dass sich die Rankings oft nur auf einen der drei Bereiche stützen. 2015 waren das die Naturwissenschaften, und auf dieser Basis wurde die Platzierung vorgenommen. Deutschland und Großbritannien hatten damals die gleiche Punktzahl, und nun würde man denken, dass dann nach den anderen beiden Bereichen geschaut wird, also beim Leseverständnis und in der Mathematik. Da lag Deutschland deutlich vor Großbritannien, wurde aber trotzdem dahinter eingeordnet. Ein weiterer Punkt: Pisa testet zum Beispiel keine Fremdsprachen-Kompetenzen. Ich hatte in meinen Klassen Schüler, die Englisch, Französisch, Latein und Italienisch lernten. In den Vereinigten Staaten würde ein Test auf Fremdsprachen fatal ausgehen, aber das spielt bei Pisa keine Rolle. Ob Schüler Ahnung haben von Geschichte, Politik, ob sie sich musisches oder künstlerisches Können aneignen - wird nicht getestet. Pisa testet nur den Ausschnitt, den die OECD für relevant hält, und definiert damit für sich, was Bildung ist. Über die Leistungsfähigkeit eines Bildungswesen sagt das also wenig aus.

Gibt es noch andere Lücken in den Pisa-Erhebungen?

Ein weiteres Problem sind die Erfassungsquoten der Stichproben. Es wird bei Pisa angestrebt, durch repräsentative Stichproben möglichst alle Schüler zu erfassen; 89 Prozent sind es nach Angaben der OECD weltweit im Schnitt. Deutschland liegt seit einigen Jahren immer in der Spitzengruppe, hier wurden zuletzt 96 Prozent der Schülerpopulation erfasst. Andererseits gibt es Länder, bei denen der Erfassungsgrad viel niedriger ist. Das vielleicht krasseste Beispiel ist

Schanghai, das 2012 mit großem Abstand Pisa-Sieger war. Damals wurden nur knapp 80 Prozent der vorgesehenen Stichproben getestet. In dieser Stichprobe waren aber die Kinder der armen Wanderarbeiter nicht enthalten, weil die nicht auf die höheren Schulen gehen, an denen der Test durchgeführt wurde. Das betrifft dort etwa 40 Prozent der Bevölkerung. Die Frage ist, wie es mit denen steht, die an dem Test nicht teilnahmen. Mutmaßlich waren das die schwächeren Schüler. Ein anderes Beispiel ist Vietnam, das in der gerade erschienenen OECD-Studie über Gleichheit im Bildungswesen als Vorbild erscheint. Dort repräsentierte die Pisa-Stichprobe 2015 knapp die Hälfte aller Fünfzehnjährigen. Die OECD gibt sogar selbst zu, dass Länder mit niedrigen Erfassungsraten in Vergleichen favorisiert seien. Sie handelt aber nicht danach und stellt Vietnam trotzdem als Vorbild hin. Wenn Deutschland ungestraft die Hälfte seiner Fünfzehnjährigen vom Pisa-Test ausschließen könnte, würden die Ergebnisse vermutlich auch anders ausfallen.

In Deutschland sorgen diese Tests immer wieder für große Verunsicherung.

Ja, man sagt dann: Aha, schon wieder ein asiatisches Land, das vor uns liegt, daran müssen wir uns orientieren.

Steckt dahinter eine konkrete Absicht der OECD?

Das kann ich nicht sicher beurteilen, aber es scheint so zu sein, dass dort ein negatives Bild des deutschen Bildungssystems vorherrscht und eine starke Präferenz für die Gesamtschule und das längere gemeinsame Lernen besteht. Letzten Endes wird dort Politik gemacht, die als wissenschaftlich begründet verkauft wird.

Die vielen Reformen im deutschen Bildungswesen waren demnach eine Folge durchaus angreifbarer Pisa-Tests?

Pisa hat dabei schon eine wichtige Rolle gespielt, es herrschte damals die Stimmung: Deutschland hat schlecht abgeschnitten, also ist unser Bildungssystem schlecht und muss möglichst schnell verändert werden. Ein besonders negatives Beispiel war die dilettantische G-8-Reform, bei der man, ohne viel nachzudenken, zum Teil noch im laufenden Schuljahr einfach die Schulzeit verkürzte, ohne zuvor die Lehrpläne zu reformieren. Da haben wir uns ganz unnötig ein Problem eingehandelt, das jetzt mühsam wieder rückabgewickelt wird - was übrigens auch den Lehrerberuf wieder erhöht.

<https://www.faz.net/aktuell/politik/inland/kritik-an-schulsystem-ein-bildungsforscher-ueber-pisa-15953629.html>

Schiefer als der schiefe Turm

Bei den asiatischen Ländern wirft die OECD ihre eigenen Regeln über Bord,
weshalb die Pisa-Studie schwere Mängel aufweist.

Anfang Dezember ist es so weit: Eine neue Ausgabe der PISA-Studie kommt heraus. Wird das deutsche Schulwesen im internationalen Vergleich wieder als leistungsschwach und sozial ungerecht dastehen? Schon im Januar 2018 wies der Präsident des Deutschen Lehrerverbandes, Heinz-Peter Meidinger, warnend darauf hin, dass seit 2015 durch die Zuwanderung von Flüchtlingen mit über 200.000 schulpflichtigen Kindern die Zahl der Familien zunehme, in denen zu Hause kein Deutsch gesprochen werde.³⁰ Sprachkenntnisse aber sind für das Abschneiden im Pisa-Test von zentraler Bedeutung. Doch auch methodische Probleme der Studie können sich stark auf das nationale Pisa-Ergebnis auswirken.

Das Programme for International Student Assessment (PISA) der OECD untersucht seit der Jahrtausendwende alltags- und berufsrelevante Kompetenzen fünfzehnjähriger Schülerinnen und Schüler in den Bereichen Leseverständnis, Mathematik und Naturwissenschaften. Dabei arbeitet es mit Stichproben, die weniger als ein Prozent der Zielgruppe umfassen. Das ist in der Statistik üblich, birgt aber gleichwohl seine Tücken. Denn wenn eine Stichprobe nicht korrekt gezogen wird, kommt es zu falschen Ergebnissen. So etwa in Österreich, wo der Pisa-Test 2003 deutlich schlechter ausfiel als in der ersten Runde drei Jahre zuvor. Nachdem für diesen vermeintlichen Niedergang die neue Regierung verantwortlich gemacht worden war, förderte eine Überprüfung durch versierte Statistiker Ungereimtheiten bei der Stichprobenziehung und Datenauswertung des Jahres 2000 zutage. Daher mussten die früheren Werte nach unten korrigiert werden. Auch die Stichprobe der USA von 2009 wurde von amerikanischen Forschern als unzuverlässig kritisiert. Nach ihren korrigierten Punktzahlen hätten die USA im Ranking um zehn und mehr Plätze besser abgeschnitten.³¹ Derartige Fehler sind auch für andere Länder und Testrunden nicht auszuschließen.

Ein weiteres Problem besteht darin, dass nicht alle für den Test vorgesehenen Schüler an der Erhebung teilnehmen. Zur Begrenzung des dadurch entstehenden Schätzfehlers gilt bei PISA eine Mindestbeteiligungsquote von 80 Prozent. Als Großbritannien diese Quote 2003 knapp verfehlte, wurde es in internationalen Vergleichen nicht berücksichtigt.³² Doch als Schanghai 2012 ebenfalls diese Hürde riss, setzte sich die OECD über ihre eigenen Regeln hinweg und

³⁰ <https://www.news4teachers.de/2018/01/pisa-zu-bildungsgerechtigkeit-vbe-und-lehrerverband-registrieren-fortschritte-aber>.

³¹ Rainer Bölling, Zur Fragwürdigkeit des PISA-Rankings Gesellschaft für Bildung und Wissen, 21. 11. 2017.

³² OECD: Lernen für die Welt von morgen. Erste Ergebnisse von PISA 2003, 2004, S. 371 u. 375-377.

präsentierte Schanghai als Sieger und Vorbild für die Welt.³³ Das war umso fragwürdiger, als dort die Kinder von Millionen Wanderarbeitern schon vorab außen vor blieben, weil sie gar nicht die höheren Schulen besuchen durften, an denen der PISA-Test stattfand.³⁴

Ungeachtet dessen erklärte Pisa-Koordinator Schleicher im Dezember 2013, in Shanghai gebe es in Mathematik keinen leistungsschwachen Schüler, und folgerte daraus: „Mathematik ist etwas, das jeder Schüler aus jeder Schicht lernen kann.“³⁵ Hätte er die einschlägigen PISA-Daten zur Kenntnis genommen, dann wäre ihm vielleicht aufgefallen, dass selbst in dieser unvollständigen Stichprobe 3,8 Prozent der Getesteten die Kompetenzstufe 2 verfehlt hatten, also als sogenannte Risikoschüler gelten.³⁶ Schleichers Aussage zeigt, dass seine bildungspolitischen Wunschvorstellungen und Ratschläge oft nur vorgeblich auf den von der OECD publizierten Daten beruhen.

Bei PISA 2015 erscheint Shanghai nicht mehr allein, sondern zusammen mit Peking, Jiangsu und Guangdong als Teil einer neuen Einheit (P-S-J-G) mit 223 Millionen Einwohnern, die China repräsentiert. Sie wird auf Platz 10 des offiziellen Rankings geführt, obwohl ihre Erfassungsquote nur 64 Prozent beträgt. Dennoch lag der Anteil der Risikoschüler in Mathematik hier gut viermal so hoch wie 2012 in Shanghai.³⁷ Noch krasser ist der Fall Vietnam, das 2015 im Ranking wegen seiner 525 Punkte in Naturwissenschaften auf Platz 8 steht, obwohl es mit 48,5 Prozent die Mindestbeteiligungsquote von 80 Prozent bei weitem verfehlt hat.

Länder mit niedriger Beteiligung sind im Vorteil

Die OECD räumt ein: „Wenn man annimmt, dass aus der PISA-Stichprobe ausgeschlossene Schülerinnen und Schüler wahrscheinlich schlechtere Ergebnisse erzielt hätten als die in den Stichproben vertretenen Schülerinnen und Schüler, dürften Länder mit niedrigen Erfassungsraten in Vergleichen favorisiert sein.“³⁸ Eben deshalb aber hätten beide Länder vom internationalen Vergleich ausgeschlossen werden müssen. Man fragt sich, warum die OECD bei bestimmten asiatischen Ländern ihre eigenen Regeln über Bord wirft.

³³ OECD, PISA 2012 Ergebnisse: Was Schülerinnen und Schüler wissen und können. Band I (überarbeitete Ausgabe, Februar 2014): Schülerleistungen in Lesekompetenz, Mathematik und Naturwissenschaften, 2014, S. 290 f.

³⁴ Tom Loveless, Attention OECD-PISA: Your-Silence-on-China-is-Wrong, 12.12.2013 (<http://educationnext.org/attention-oecd-pisa-your-silence-on-china-is-wrong>); Valerie Strauss, Did Shanghai cheat on PISA? 15.12.2013 (<http://www.washingtonpost.com/blogs/answer-sheet/wp/2013/12/15/did-shanghai-cheat-on-pisa>).

³⁵ <https://www.dw.com/de/ausbruch-aus-der-schockstarre/a-17267586>.

³⁶ PISA 2012, S. 20.

³⁷ OECD: PISA 2015 Ergebnisse (Band I): Exzellenz und Chancengerechtigkeit in der Bildung. 2016, S. 315 u. 414.

³⁸ Ebd., S. 229.

Kritik an diesem Vorgehen tut die OECD auf ihrer Website als Mythen ab: „Alle PISA-Proben sind repräsentativ für die 15-Jährigen, die an Schulen angemeldet sind“, heißt es dort. „Sie werden auf einer wissenschaftlich fundierten Basis ausgewählt, um sicherzustellen, dass alle Schülerinnen und Schüler gleichmäßig repräsentiert werden. Dabei werden 89 Prozent der 15-Jährigen von PISA abgedeckt.“³⁹ Doch das ist wohl gemerkt der OECD-Durchschnitt, der eben klare Erfassungsdefizite einzelner Teilnehmerländer verschleiert. Übrigens erreichen auch Länder wie Großbritannien, die Vereinigten Staaten oder Kanada nur Erfassungsquoten unter 85 Prozent, während Deutschland mit 96 Prozent in der Spitzengruppe liegt. Solche Unterschiede wirken sich durchaus auf die nationalen Durchschnittswerte aus.

Wenn die OECD Meldungen über die Defizite schwächerer Schülerinnen und Schüler veranschaulichen will, rechnet sie gern PISA-Punkte in Schuljahre um. So war im Oktober 2018 in der Studie „Equity in Education“ zu lesen, der Bildungsrückstand sozial benachteiligter Schüler betrage im OECD-Durchschnitt 88 Punkte beziehungsweise etwa drei Schuljahre.⁴⁰ Das wirkt noch dramatischer als zuvor. 2009 und 2012 galten bei PISA ca. 125 Punkte als Äquivalent für drei Schuljahre, 2015 waren es noch 100.⁴¹ Wie unsinnig all diese Umrechnungen sind, wird dann deutlich, wenn man sie auf die Unterschiede innerhalb der Länder anwendet. Zwischen den leistungsstärksten und den leistungsschwächsten Schülern eines OECD-Landes (obere und untere 5 Prozent) lagen 2015 im Durchschnitt der drei Kompetenzbereiche fast überall 300 und mehr Punkte.⁴² Nimmt man das mittlere Maß von 2015 als Umrechnungsfaktor, so würde das einen Bildungsrückstand der Leistungsschwächsten von neun bis zehn Schuljahren bedeuten. Demnach wären beispielsweise die schwächsten finnischen Fünfzehnjährigen nach acht Jahren Schulbesuch schon neun Jahre in Rückstand geraten. Offensichtlich dient dieses unbrauchbare Umrechnungsverfahren nur der Skandalisierung des Unspektakulären und gehört daher auf den Müllhaufen der empirischen Bildungsforschung.

Viele Länder haben eine selektive Einwanderungspolitik

Unabhängig von der Größe des Umrechnungsfehlers zeigt diese Betrachtung, dass es auch PISA-Musterländern nicht annähernd gelingt, „kein Kind zurückzulassen“ und alle zu guten Leistungen zu führen. Besonders groß ist der Abstand zwischen den besten und den schwächsten Schülern in den chinesischen Provinzen, aber auch in Singapur. Dennoch

³⁹ <https://www.oecd.org/berlin/themen/pisa-studie>

⁴⁰ OECD: Equity in Education. Breaking Down Barriers to Social Mobility. Paris 2018, S. 13.

⁴¹ PISA 2009, S. 13 u. ö.; PISA 2012, S. 4 u. 272; PISA 2015, S. 41 u. 162.

⁴² Berechnet nach PISA 2015, S. 351, 404 u. 417.

behauptet Schleicher, seine zahlreichen Reisen nach Asien hätten ihm gezeigt, „dass man praktisch alle Schüler zu wirklich guten Leistungen motivieren kann“. ⁴³ Die PISA-Daten sprechen für das Gegenteil.

Im Übrigen liegt die Gruppe der besten Schüler (obere 5 Prozent) im internationalen Vergleich auf einem ähnlichen Niveau. Abgesehen von asiatischen Ländern mit teils fragwürdigen oder sehr speziellen Voraussetzungen gibt es eine Spitzengruppe mit durchschnittlichen Punktwerten zwischen 667 und 652. Dazu gehören (in dieser Reihenfolge) Kanada, Finnland, Deutschland (661), die Niederlande, Schweiz, Belgien und Frankreich. Wenn die Durchschnittswerte dieser Ländergruppe stärker voneinander abweichen, so liegt das nicht zuletzt am unterschiedlichen Anteil geringer qualifizierter Zuwanderer in der jeweiligen Stichprobe. Es wäre daher geboten, die sprachlichen, kulturellen und sozialen Voraussetzungen von Migrantenkindern genauer in den Blick zu nehmen.

Den Pisa-Machern ist durchaus bekannt, dass Länder wie Kanada, Australien und Neuseeland eine selektive Einwanderungspolitik betreiben, so dass es sich bei den Zuwanderern weitgehend um Hochqualifizierte handelt, die zudem die Landessprache beherrschen. Dagegen weisen etwa Deutschland, Österreich, Belgien und Frankreich einen hohen Anteil an dauerhaft ansässigen geringqualifizierten Zuwanderern auf. Zwar heißt es bei Pisa, dass die im Ländervergleich bestehenden Unterschiede „auch unter dem Aspekt der Selektivität der Zuwanderungspolitik der Aufnahmeländer sowie der relativen kulturellen und sprachlichen Verwandtschaft der Herkunfts- und Aufnahmeländer betrachtet werden“ müssen. Doch eine solche Differenzierung findet tatsächlich nicht statt. Man erfährt lediglich, „dass der durchschnittliche Leistungsabstand zwischen Schülerinnen und Schülern mit und ohne Migrationshintergrund und gleichem sozioökonomischem Profil 31 Punkte beträgt“. ⁴⁴ Wie sich Anteil und Qualifikationsniveau von Migrantenkindern konkret auf die nationalen PISA-Ergebnisse auswirken, wird jedoch nicht thematisiert.

Offensichtlich handelt es sich hier um ein Tabu, wie die Reaktion der deutschen PISA-Koordinatorin Kristina Reiss in einem Interview mit der „Zeit“ im Dezember 2016 zeigt. Als sie mit der These konfrontiert wurde, ohne Einwandererkinder wäre Deutschland auf dem Niveau von Europas Schulprimus Finnland, entgegnete sie: „Diese Argumentation ist gefährlich. Zum einen stammt ein gutes Viertel der Risikoschüler aus deutschstämmigen

⁴³ Wie Fn 35 - „Dass Schülerinnen und Schüler unabhängig von ihrem sozioökonomischen und soziokulturellen Status sowie unabhängig von Hautfarbe und Geschlecht gleiche Lernstände erreichen“, galt auch dem geistesverwandten Chancenspiegel der Bertelsmann-Stiftung als Inbegriff von Bildungsgerechtigkeit (Bertelsmann Stiftung: Chancenspiegel. Zur Chancengerechtigkeit und Leistungsfähigkeit der deutschen Schulsysteme, Gütersloh 2012, S. 24).

⁴⁴ OECD: PISA 2015, S. 264 f.

Familien. Zum anderen schaffen es Länder wie Kanada durchaus, Kinder aus Einwandererfamilien zu guten Leistungen zu führen.“⁴⁵ Doch Kanada verdankt seine guten Pisa-Ergebnisse ja nicht zuletzt seiner selektiven Einwanderungspolitik. Dass dies hartnäckig unterschlagen wird, verstößt eklatant gegen wissenschaftliche Regeln.

Wenn hingegen die Pisa-Ergebnisse Deutschlands Fortschritte erkennen lassen, ist die OECD schnell dabei, diese für sich zu reklamieren. So führte sie das gute Abschneiden 2012 auch auf den von ihr befürworteten Ausbau der frühkindlichen Förderung zurück. Ihr stellvertretender Generalsekretär behauptete sogar, man könne „in den Pisa-Zahlen ganz deutlich messen, dass bereits ein Jahr frühkindliche Bildung nachher die Ergebnisse im Pisa-Test deutlich verbessert“.⁴⁶ Doch die Pisa-Zahlen geben nichts dergleichen her, denn die 2012 getesteten Fünfzehnjährigen wurden ja spätestens 2003 eingeschult, also vor dem Ausbau der frühkindlichen Bildung in Deutschland. Sie kann also nicht der Grund für die Verbesserung sein.

Die hier aufgezeigten methodischen Probleme und irreführenden Interpretationen der Pisa-Studie sollten Grund genug sein, der Präsentation der neuen Ausgabe nicht blind zu vertrauen, sondern ihr kritisch zu begegnen. Auf eine konstruktive Diskussion mit den Pisa-Machern zu hoffen erscheint allerdings unrealistisch, müssten diese dann doch ihren latenten Unfehlbarkeitsanspruch aufgeben.

⁴⁵ <https://www.zeit.de/2016/51/pisa-studi-bildungspolitik-kristina-reiss-olaf-koeller/komplettansicht>.

⁴⁶ https://www.deutschlandfunkkultur.de/oecd-bildungsstudie-grosse-fortschritte-beider.2165.de.html?dram:article_id=337814.

Wozu PISA?

Im Dezember 2019 erschien die siebte Ausgabe der internationalen PISA-Studie. Ihr zufolge sind Deutschlands Mittelwerte in den drei Testbereichen Lesen, Mathematik und Naturwissenschaften 2018 auf durchschnittlich 500 Punkte gesunken. Damit liegen sie aber immer noch 12 Punkte über dem ebenfalls gesunkenen OECD-Durchschnitt, der bei PISA als Maß aller Dinge gilt.

Manche Kommentatoren glaubten wenn schon nicht einen neuen PISA-Schock (wie 2001), so doch einen PISA-Knick erkannt zu haben. Wie fragwürdig das ganze Ranking ist, blieb den meisten allerdings verborgen. Vor allem fiel unter den Tisch, in welcher Disziplin Deutschland den Spitzenplatz einnimmt: beim Erfassungsgrad der nationalen Stichprobe, den selbst PISA-Koordinator Schleicher als „fantastisch“ bezeichnet.⁴⁷ Diesmal sollen die getesteten 5451 Schüler für 99,3 Prozent aller Fünfzehnjährigen repräsentativ sein, womit Deutschland zehn Prozentpunkte über dem OECD-Durchschnitt liegt. Da kann ein stets als Vorbild geltendes Land wie Kanada mit nur 86,3 Prozent nicht mithalten.⁴⁸

Die deutsche Gründlichkeit hat freilich ihren Preis, denn – so PISA – „die Einbeziehung weiterer 15-Jähriger in die Testpopulation ist häufig mit der Berücksichtigung leistungsschwächerer Schülerinnen und Schüler verbunden“.⁴⁹ Offenbar hat man in Deutschland so gut wie keinen Gebrauch von der Möglichkeit gemacht, Schüler „mit kognitiven oder körperlichen Behinderungen oder mangelnden Kenntnissen in der Testsprache“ auszuschließen.⁵⁰ Das ist gerade diesmal von Bedeutung, hat doch seit 2015 durch die Zuwanderung von Flüchtlingen mit über 200.000 schulpflichtigen Kindern die Zahl der Familien zugenommen, in denen zu Hause kein Deutsch gesprochen wird.

In Schweden dagegen hat eine Zuwanderung vergleichbaren Ausmaßes zu einem deutlichen Anstieg der Ausschlussquote und einem starken Rückgang des Erfassungsgrades geführt: von 93,6 auf 85,7 Prozent.⁵¹ So verbesserte sich Schweden im Ranking um durchschnittlich sechs Punkte, während Deutschland acht Punkte einbüßte. Anders sieht es aus, wenn man aus der deutschen Testpopulation zehn Prozent herausrechnet, und zwar mit der durchschnittlichen Punktzahl der sogenannten Risikoschüler, die die Kompetenzstufe 2 bei PISA verfehlen. Dann würde Deutschland 15 bis 20 Punkte zulegen und nicht nur an Schweden vorbeiziehen, sondern sogar zu Kanada und Finnland aufschließen.

⁴⁷ Webinar vom 27.11.2019 (<http://www.oecd.org/pisa/webinars>).

⁴⁸ OECD: PISA 2018 Ergebnisse (Band I), Was Schülerinnen und Schüler wissen und können, S. 182.

⁴⁹ Ebd., S. 308.

⁵⁰ <http://www.oecd.org/berlin/themen/pisa-studie/haeufig-gestellte-fragen.htm> (Zugriff: 18.12.2019).

⁵¹ PISA 2018 Ergebnisse (Band I), S. 58.

Von einer Vergleichbarkeit der nationalen PISA-Ergebnisse kann also nicht die Rede sein. Das liegt eben auch daran, dass gravierende demographische Unterschiede zwischen den Ländern ausgeblendet werden. Den PISA-Machern ist zwar bekannt, dass Länder wie Kanada, Australien und Neuseeland eine selektive Einwanderungspolitik betreiben, so dass es sich bei den Zuwanderern weitgehend um Hochqualifizierte handelt, wogegen etwa Deutschland, Österreich, Belgien und Frankreich einen hohen Anteil an dauerhaft ansässigen geringqualifizierten Einwanderern aufweisen.⁵² Dennoch werden die im Ländervergleich bestehenden Unterschiede letzten Endes allein auf die Bildungssysteme zurückgeführt.

Wie wenig Unterschiede von 10 oder 20 Punkten bei den PISA-Mittelwerten überhaupt aussagen, wird deutlich, wenn man die Spreizung *innerhalb* der Länder betrachtet. Überall liegen gemäß Definition der Kompetenzstufen zwischen den leistungsstärksten und leistungsschwächsten Schülern 310 (Mathematik) bis 500 (Lesekompetenz) Punkte. Was diese Differenzen konkret bedeuten, glaubt allerdings nicht einmal mehr die OECD zu wissen. Die bisher praktizierte Umrechnung von 100 PISA-Punkten in mindestens drei Schuljahre wurde in der neuesten Ausgabe aus gutem Grund als untauglich verworfen.⁵³ Denn sie hätte beispielsweise bedeutet, dass die leistungsstärksten Schüler im Lesen nach neun Jahren Schulbesuch einen Vorsprung von mehr als 15 Schuljahren auf die schwächsten erworben hätten.

Unabhängig von der unklaren Bedeutung dieser Punkte zeigt die in allen Ländern starke Spreizung, dass es auch den PISA-Musterländern nicht annähernd gelingt, „kein Kind zurückzulassen“ und alle zu guten Leistungen zu führen. Dieser Befund wird von der OECD jedoch hartnäckig ignoriert. Vielmehr fordert Schleicher in seinem neuen Buch „Weltklasse“, Schulsysteme müssten „anerkennen, dass alle Kinder hohen Lern- und Leistungsanforderungen genügen können, wenn sie sich anstrengen und entsprechend gefördert werden.“⁵⁴ Doch die PISA-Daten ergeben nun mal ein anderes Bild und können dazu beitragen, solche von Allmachtsphantasien genährten Illusionen mancher Pädagogen und Bildungspolitikern zu entlarven.

Nimmt man einmal das Niveau der leistungsstärksten Schüler in den Blick, so zeigt sich, dass Deutschland durchaus mit Ländern wie Kanada, Estland oder Finnland mithalten kann – auch auf Basis der nicht um den Effekt des Erfassungsgrades bereinigten Zahlen. Nur die ostasiatischen Metropolen und Stadtstaaten, die seit 2009 Finnland als PISA-Weltmeister verdrängt haben, erreichen Ergebnisse auf höherem Niveau. Allerdings spricht einiges dafür, dass deren Stichproben nicht repräsentativ und daher nicht vergleichbar sind.

⁵² OECD: PISA 2015 Ergebnisse (Band I): Exzellenz und Chancengerechtigkeit in der Bildung. 2016, S. 264 f.

⁵³ PISA 2018 Ergebnisse (Band I), S. 51.

⁵⁴ Andreas Schleicher, Weltklasse. Schule für das 21. Jahrhundert gestalten, Bielefeld: wbv Publikation 2019, S. 63.

Darüber hinaus stellt sich die Frage, ob Deutschland sich diese Länder, insbesondere China, zum Vorbild nehmen soll. Denn deren PISA-Erfolg beruht nicht zuletzt auf einem gigantischen Nachhilfesystem. Es ist die Kehrseite eines Schulsystems, das auf Abschlusstests in wenigen Fächern fokussiert ist, die den weiteren Lebensweg maßgeblich bestimmen. Das hat im konfuzianischen Kulturkreis Tradition, ist aber auch ein Kennzeichen autoritärer Regime. Daher warnt der in den USA lehrende chinesische Bildungsforscher Yong Zhao vor einer Übernahme dieses Modells, weil es seine Erfolge auf Kosten von Vielfalt und Kreativität erzielt.
55

Das heißt nun nicht, dass es im deutschen Bildungswesen keinen Verbesserungsbedarf gibt. Doch um den zu erkennen und Lösungen zu finden, braucht es nicht Unmengen von PISA-Daten, deren Interpretation durch die OECD großenteils irreführend ist und einer verschwiegenen politischen Agenda folgt.

⁵⁵ Yong Zhao, *Who's Afraid of the Big Bad Dragon? Why China Has the Best (and Worst) Education System in the World*, San Francisco 2014.

Wie PISA Bildungsgerechtigkeit zu messen versucht

Wenn vom deutschen Bildungswesen die Rede ist, hört man immer wieder die Behauptung, in keinem oder kaum einem anderen Land der Welt hänge der Schulerfolg so sehr von der sozialen Herkunft ab wie in Deutschland. Dieses Mantra geht auf die ersten PISA-Studien zurück und wird seitdem in vielen Medien ungeprüft nachgebetet. Dabei stellte die OECD schon vor Jahren fest, Deutschland habe 2000 im Hinblick auf die soziale Gerechtigkeit in der Bildung zu den am schlechtesten abschneidenden OECD-Ländern gehört, 2012 aber nach dieser Messgröße schon ungefähr das OECD-Durchschnittsniveau erreicht.⁵⁶ Wie diese Daten zustande kommen und wie zuverlässig sie sind, fragt sich aber kaum jemand.

Der zentrale PISA-Indikator für Bildungsgerechtigkeit ist der „Prozentsatz der durch den sozioökonomischen Status der Schüler erklärten Leistungsvarianz“. Dieser Status wird mittels des PISA-Index des wirtschaftlichen, sozialen und kulturellen Status (ESCS) geschätzt, der aus mehreren Variablen des familiären Hintergrunds der Schüler abgeleitet wird: dem Bildungsniveau und dem Beruf der Eltern, der Ausstattung des Elternhauses als Hilfsindikator für den materiellen Wohlstand sowie der Zahl der Bücher und anderer Bildungsressourcen, die im Elternhaus verfügbar sind. Dass diese Daten auf Angaben der getesteten fünfzehnjährigen Schüler beruhen, wird bei PISA vermerkt, ohne ihre Verlässlichkeit zu thematisieren.

In Deutschland wurden diese Schülerangaben schon 2000 im Rahmen der nationalen Erweiterungsstudie PISA E durch eine Befragung von Eltern überprüft. Bei den Angaben zum Schulabschluss der Eltern lag die Übereinstimmung von Schüler- und Elternangaben zwischen 65 und 70, bei der Berufstätigkeit nur bei rund 50 Prozent. Doch die Autoren der Studie zeigten sich überzeugt, dass eine prozentuale Übereinstimmung von 49,2 Prozent „für den Bereich der Berufstätigkeit als hoch betrachtet werden“ könne. So kamen sie zu dem erstaunlichen Ergebnis, dass Schüler „als zuverlässige Informanten zur Erhebung von Bildungs- und Berufsmerkmalen der Eltern betrachtet werden können“.⁵⁷

Auf der Basis solcher Daten berechnet PISA in einem für Außenstehende nicht nachprüfbar Verfahren, wieviel Prozent der im Test festgestellten Leistungsunterschiede auf die soziale Herkunft der Schüler zurückzuführen sind. Hiernach waren das 2018 im OECD-Durchschnitt 12, in Deutschland aber 17,2 Prozent.⁵⁸ Das scheint auf den ersten Blick die

⁵⁶ Vgl. OECD (2013): Deutschland – Ländernotiz – Ergebnisse aus PISA 2012, S. 5.

⁵⁷ Kai Maaz, Frauke Kreuter und Rainer Watermann; Schüler als Informanten? Die Qualität von Schülerangaben zum sozialen Hintergrund, in: Baumert, Stanat, Watermann (Hrsg.): Herkunftsbedingte Disparitäten im Bildungswesen. Vertiefende Analysen im Rahmen von PISA 2000. Wiesbaden 2006, S. 31-59, Zitate S. 48 f. und 55.

⁵⁸ OECD (2019), PISA 2018 Results (Volume II): Where All Students Can Succeed, PISA, OECD Publishing, Paris, p. 17.

gängige These zu bestätigen, wenngleich das Ausmaß der Abweichung manchem überraschend gering erscheint. Diese Zahlen spiegeln aber nicht zwangsläufig den sozialen Effekt des deutschen Schulwesens wider. Vielmehr können Abweichungen auch auf unterschiedliche demographische Voraussetzungen zurückgehen, die die jeweiligen Bildungssysteme nicht zu verantworten haben.

So gehört Deutschland mit Frankreich, Belgien und Luxemburg schon länger zur Gruppe der europäischen Länder, die einen hohen Anteil geringqualifizierter Zuwanderer aufweisen. Diese Länder verzeichnen in puncto Bildungsgerechtigkeit ähnliche, sogar noch etwas schlechtere Werte als Deutschland (zwischen 17 und 18 Prozent). Dabei weist Deutschland nach Luxemburg mit 22 Prozent den höchsten Migrantenanteil auf.⁵⁹ Diese Zahl enthält nur die Jugendlichen, deren beide Elternteile im Ausland geboren wurden.⁶⁰ Zählt man wie das deutsche PISA-Konsortium auch die mit, bei denen nur ein Elternteil im Ausland geboren wurde, so steigt dieser Anteil auf 36 Prozent.⁶¹

Die Migranten in diesen Ländern gelten zur Hälfte als sozioökonomisch benachteiligt, was zu schlechteren PISA-Resultaten beiträgt. Ihre Resultate hängen aber auch mit sprachlichen Defiziten zusammen, da viele von ihnen zu Hause nicht die Unterrichts- und damit Testsprache sprechen. In Deutschland gab 2018 nur etwas mehr als die Hälfte der Schüler mit Migrationshintergrund Deutsch als zu Hause gesprochene Sprache an.⁶²

Diesen Zusammenhang spielen PISA-Verantwortliche gern herunter, und zwar mit dem Hinweis, Länder wie Kanada schafften es durchaus, Kinder aus Einwandererfamilien zu guten Leistungen zu führen. Dabei wird verschwiegen, dass Kanada eine selektive Immigrationspolitik verfolgt, so dass es sich bei den Zuwanderern großenteils um Hochqualifizierte handelt, die auch die Landessprache besser beherrschen. Entsprechend schneidet das Land auch in puncto „Bildungsgerechtigkeit“ überdurchschnittlich ab.

Während sich der PISA-Gerechtigkeits-Indikator für Deutschland gegenüber 2015 leicht verschlechtert hat, ist bei einem anderen Land, das seitdem in vergleichbar hohem Maße Flüchtlinge aufgenommen hat, das Gegenteil der Fall: bei Schweden. Dort wurde von der Möglichkeit, Schüler „mit kognitiven oder körperlichen Behinderungen oder mangelnden Kenntnissen in der Testsprache“ auszuschließen⁶³, ausgiebig Gebrauch gemacht. Die hohe Ausschlussquote führte dazu, dass die aktuelle schwedische PISA-Stichprobe nur noch 86

⁵⁹ Ebd., p. 310.

⁶⁰ Ebd., p. 179.

⁶¹ Kristina Reiss, Mirjam Weis, Eckhard Klieme, Olaf Köller (Hrsg.): PISA 2018. Grundbildung im internationalen Vergleich, Münster 2019, S. 151.

⁶² Ebd., S. 153.

⁶³ <http://www.oecd.org/berlin/themen/pisa-studie/haeufig-gestellte-fragen.htm> (Zugriff: 18.12.2019).

Prozent der Fünfzehnjährigen repräsentiert.⁶⁴ So konnte Schweden sich nicht nur im fachlichen Ranking, sondern auch beim PISA-Gerechtigkeits-Indikator verbessern. In Deutschland dagegen führte ein Rekord-Erfassungsgrad von über 99 Prozent zu schlechteren Werten in beiden Bereichen. Aus so unterschiedlichen Stichproben vergleichende Aussagen über die Bildungssysteme abzuleiten, hat mit seriöser Wissenschaft nichts zu tun.

Bessere PISA-Ergebnisse durch selektives Testen

Ein wenig beachteter, aber gravierender Schwachpunkt der PISA-Studie liegt darin, dass Länder bzw. Volkswirtschaften miteinander verglichen werden, in denen die für den Test relevante Schülerpopulation sehr unterschiedlich erfasst wird. Welche Folgen das haben kann, sei an einem einfachen Beispiel verdeutlicht.

Wir nehmen an, dass in einem Land idealerweise 100 Prozent aller für eine repräsentative Stichprobe gezogenen Schülerinnen und Schüler am Test teilnehmen und auf einer neunstufigen Skala folgende Punktverteilung erreichen:

Punkteskala	300	350	400	450	500	550	600	650	700	Gesamt
Messwerte	5	11	12	14	16	14	12	11	5	100
Punktzahl	1500	3850	4800	6300	8000	7700	7200	7150	3500	50000
Mittelwert										500

Diese Punkteverteilung entspricht ziemlich genau der Gaußschen Normalverteilung, die PISA zugrunde liegt: der Mittelwert aller Messwerte beträgt 500, und 68 Prozent aller Werte liegen nicht mehr als eine Standardabweichung (100 PISA-Punkte) vom Mittelwert entfernt, also zwischen 400 und 600 Punkten. Ein Land mit diesen Testwerten landet im internationalen Vergleich im Mittelfeld und muss sich wohl vorhalten lassen, dass viele Länder besser abgeschnitten haben.

⁶⁴ Vgl. OECD (2019): PISA 2018 Ergebnisse (Band I), Was Schülerinnen und Schüler wissen und können, S. 58.

Nun stellen findige Leute in diesem Land fest, dass in anderen Ländern gar nicht alle Probanden getestet wurden, so dass der durchschnittliche Erfassungsgrad bei PISA nur 89 Prozent beträgt.⁶⁵ Wenn aber „der Erfassungsgrad außer Acht gelassen wird, dürften die Leistungen der 15-jährigen Kohorte eines Bildungssystems überschätzt werden, und zwar umso mehr, je niedriger der Erfassungsgrad ist“, räumt die OECD ein.⁶⁶ Offensichtlich machen andere Länder von der Möglichkeit Gebrauch, „Schüler mit kognitiven oder körperlichen Behinderungen oder mangelnden Kenntnissen in der Testsprache (Schüler, die z.B. weniger als ein Jahr lang Unterricht in der Testsprache hatten)“ vom Test auszuschließen.⁶⁷

Das geschah 2018 besonders in Schweden, wo laut OECD „die hohe Zahl der Schülerinnen und Schüler, die seit 2015 als Zuwanderer oder Flüchtlinge in das Land gekommen waren, zu einem deutlichen Anstieg der Ausschlussquote (um 5 Prozentpunkte) und einem starken Rückgang des Erfassungsindex 3 (um 8 Prozentpunkte) führte“.⁶⁸ So wurden 2018 nur 86 Prozent aller Fünfzehnjährigen vom PISA-Test erfasst, und Schweden verbesserte sich gegenüber den vorhergehenden Erhebungen auf durchschnittlich 502 Punkte.

Auch unser Land hat seit 2015 viele Migranten aufgenommen, darunter Schülerinnen und Schüler, die auch nach einem Jahr Unterricht in der Testsprache diese verständlicherweise noch nicht wirklich beherrschten. Angesichts dessen meint man nun in unserem Land, ein Erfassungsgrad von 90 Prozent sei völlig ausreichend, und streicht die schlechtesten 10 von 100 Messwerten. Das Ergebnis ist folgendes:

Punkteskala	300	350	400	450	500	550	600	650	700	Gesamt
Messwerte		6	12	14	16	14	12	11	5	90
Punktzahl		2100	4800	6300	8000	7700	7200	7100	3500	46750
Mittelwert										519

Damit liegt dieses Land nun nicht nur vor Schweden, sondern auch vor Finnland und Kanada, die immer wieder als Vorbild hingestellt werden, aber bei PISA 2018 im Durchschnitt der drei Testbereiche nur auf 516 bzw. 517 Punkte kommen.

Doch das ist noch nicht alles, denn bei PISA genügt offiziell schon eine Mindestbeteiligungsquote von 80 Prozent. Die chinesischen Provinzen Peking, Shanghai, Jiangsu und Zhejiang, die 2018 den Spitzenplatz einnahmen, erreichten nur einen Erfassungsgrad von 81

⁶⁵ <http://www.oecd.org/berlin/themen/pisa-studie> (Zugriff 22.11.2019).

⁶⁶ OECD: PISA 2018 Ergebnisse (Band I), Was Schülerinnen und Schüler wissen und können, S. 57.

⁶⁷ <http://www.oecd.org/berlin/themen/pisa-studie/haeufig-gestellte-fragen.htm> (Zugriff 10.01.2021).

⁶⁸ OECD: PISA 2018, S. 58.

Prozent.⁶⁹ Und als Shanghai 2012 selbst diese Hürde riss, wurde es trotzdem von der OECD als Sieger und Vorbild für die Welt präsentiert.⁷⁰ Außen vor blieben dort die Kinder von Millionen Wanderarbeitern, weil sie gar nicht die weiterführenden Schulen besuchen durften, an denen der PISA-Test stattfand.⁷¹ Warum also sollte unser Land nicht nach chinesischem Vorbild 20 Prozent der Schülerpopulation unberücksichtigt lassen? Schließt man gezielt weitere schwächere Schülerinnen und Schüler aus, sieht das Ergebnis so aus:

Punkteskala	300	350	400	450	500	550	600	650	700	Gesamt
Messwerte			8	14	16	14	12	11	5	80
Punktzahl			3200	6300	8000	7700	7200	7150	3500	43050
Mittelwert										538

Mit diesen 538 Punkten hätte unser Land bei PISA 2018 seinen Platz im Mittelfeld gegen eine Bronzemedaille getauscht, ohne dass auch nur ein Testergebnis besser geworden wäre. Mit einer Erfassungsquote von 99 Prozent, der höchsten aller teilnehmenden Länder, ist ein solcher Spitzenplatz jedoch nicht zu erreichen. Das ist der Nachteil deutscher Gründlichkeit. Denn tatsächlich geht es hier um Deutschland, das 2018 bei PISA auf durchschnittlich 500 Punkte kam.

Der hier modellhaft aufgezeigte Zusammenhang zwischen Erfassungsgrad und PISA-Punktzahl ist der OECD durchaus bekannt, doch geht sie darauf nur im „Kleingedruckten“ ein und spielt das Problem herunter. So findet sich in den Berichtsbänden – nicht aber in den Pressemitteilungen – seit 2003 an unterschiedlichen Stellen folgender Satz: „Selbst wenn man z.B. unterstellt, dass die ausgeschlossenen Schülerinnen und Schüler durchgehend schlechtere Ergebnisse erzielt hätten als die teilnehmenden und dass dieser Zusammenhang mittelstark ausgeprägt ist, würde eine Ausschlussquote in einer Größenordnung von 5% wahrscheinlich nur zu einer Überzeichnung der nationalen Durchschnittsergebnisse um weniger als 5 Punkte auf der PISA-Skala führen.“⁷² Unsere Beispielrechnung zeigt, dass dieser Effekt fast doppelt so stark sein kann.

⁶⁹ OECD: PISA 2018, S. 183.

⁷⁰ OECD, PISA 2012 Ergebnisse: Was Schülerinnen und Schüler wissen und können. Band I (überarbeitete Ausgabe, Februar 2014): Schülerleistungen in Lesekompetenz, Mathematik und Naturwissenschaften, 2014, S. 20.

⁷¹ Tom Loveless, Attention OECD-PISA: Your-Silence-on-China-is-Wrong, 12.12.2013 (<http://educationnext.org/attention-oecd-pisa-your-silence-on-china-is-wrong>); Valerie Strauss, Did Shanghai cheat on PISA? 15.12.2013 (<http://www.washingtonpost.com/blogs/answer-sheet/wp/2013/12/15/did-shanghai-cheat-on-pisa>)

⁷² OECD: PISA 2018, S. 179.

Das Problem unterschiedlicher Teilnahmequoten betrifft auch andere internationale Vergleichsstudien wie TIMSS (Trends in Mathematics and Science Study), deren letzte Erhebung 2019 durchgeführt wurde. TIMSS geht aber offener damit um. Hier werden Abweichungen von den vorher definierten Stichprobenumfängen nachvollzogen und dokumentiert, „damit bei vertiefenden Vergleichen von Staaten nur solche einander gegenübergestellt werden, die tatsächlich aufgrund gleicher Teilnahmebedingungen vergleichbar sind“. ⁷³ So findet sich unter 49 Tabellen mit Darstellungen der Ergebnisse bei gut 20 Teilnehmerstaaten folgende Fußnote: „Der Ausschöpfungsgrad und/oder die Ausschlüsse von der nationalen Zielpopulation erfüllen nicht die internationalen Vorgaben.“ In den meisten Fällen liegt die Ausschlussquote zwischen fünf und zehn Prozent, nur im Falle des Spitzenreiters Singapur deutlich darüber (12,8). ⁷⁴ Zudem werden in zwei Tabellen die Mittelwerte für Deutschland dargestellt, die sich ergeben würden, wenn leistungsschwächere Schülerinnen und Schüler bis zum Erreichen einer fiktiv höheren Ausschlussquote ausgeschlossen würden. ⁷⁵

Die OECD dagegen hat 2015 Vietnam in ihre Vergleichstabellen aufgenommen, ohne an gleicher Stelle darauf hinzuweisen, dass dort nur knapp 49 Prozent der Fünfzehnjährigen vom PISA-Test erfasst wurden. Erst nach Kritik an diesem Vorgehen ⁷⁶ findet sich im Bericht 2018 folgender verklausulierter Hinweis: „Die Daten zu Vietnam wurden in den meisten Tabellen in Anhang B berücksichtigt, nicht aber in Tabellen, Abbildungen und Texten zu Leistungsvergleichen mit anderen Ländern und Volkswirtschaften oder zu Leistungsvergleichen im Zeitverlauf, da eine umfassende internationale Vergleichbarkeit der Ergebnisse zum Zeitpunkt der Veröffentlichung des Berichts nicht gewährleistet werden konnte.“ ⁷⁷ Bei einem Erfassungsgrad von mittlerweile knapp 70 Prozent kann in der Tat von Vergleichbarkeit auch weiterhin nicht die Rede sein.

⁷³ Knut Schwippert u.a.: TIMSS 2019. Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich, Münster 2020, S. 39.

⁷⁴ Ebd., S. 354.

⁷⁵ Ebd., S. 40-42.

⁷⁶ Vgl. Rainer Bölling, Zur Fragwürdigkeit des PISA-Rankings, 2017 (<https://bildung-wissen.eu/fachbeitraege/zur-fragwuerdigkeit-des-pisa-rankings.html>).

⁷⁷ OECD: PISA 2018, S. 23.

20 Jahre PISA – ein kritischer Rückblick

Im Dezember 2001 löste die Veröffentlichung der ersten PISA-Studie in Deutschland einen Schock aus, schnitt doch das „Land der Dichter und Denker“ bei diesem internationalen Bildungsvergleich nur unterdurchschnittlich ab. Seitdem hat das Programme for International Student Assessment der OECD weltweit eine Resonanz erfahren wie keine Vergleichsstudie zuvor. Bis 2018 stieg die Zahl der teilnehmenden OECD-Staaten und Partnerländer bzw. -Volkswirtschaften von 32 auf 79. In diesem Artikel werden die wichtigsten Aspekte dargestellt.

Was will PISA?

„Mit PISA wollen sich die Teilnehmerstaaten regelmäßig ein Bild davon machen, wie gut es ihren Schulen gelingt, Schülerinnen und Schüler auf die Herausforderungen der Zukunft vorzubereiten. Im Mittelpunkt steht dabei weniger das Faktenwissen der Jugendlichen, sondern es werden Basiskompetenzen untersucht, die in modernen Staaten für eine Teilhabe am gesellschaftlichen, wirtschaftlichen und politischen Leben notwendig sind. Es wird gefragt, inwieweit Jugendliche diese Kompetenzen erworben haben und inwieweit soziale Ungleichheiten im Bildungserfolg bestehen.“⁷⁸ Zu diesem Zweck werden seit 2000 alle drei Jahre standardisierte Tests durchgeführt, wobei sich Lesekompetenz, mathematische und naturwissenschaftliche Grundbildung als Schwerpunkte abwechseln.⁷⁹

Abweichend von der öffentlichen Wahrnehmung untersucht die PISA-Studie genau genommen nicht die Leistungsfähigkeit von Schulsystemen. Sie berücksichtigt nämlich nicht die Lehrpläne der teilnehmenden Länder und erfasst nur einen Ausschnitt dessen, was in Sekundarschulen gelehrt wird. So macht es zum Beispiel keinen Unterschied, ob Fünfzehnjährige eine oder mehrere Fremdsprachen lernen – sie alle zählen bei PISA nicht. Zudem bleibt offen, ob die gemessenen Kompetenzen der Getesteten in der Schule erworben wurden oder aber auch außerschulischen Instanzen zu verdanken sind wie Elternhaus oder privatem Nachhilfeunterricht. Eine große Rolle spielt dieser etwa in asiatischen Staaten, deren Schulsysteme auf Abschlusstests in wenigen Fächern fokussiert sind, die den weiteren Lebensweg maßgeblich bestimmen. PISA stellt auch keine im strengen Sinne wissenschaftliche Arbeit dar, denn die Ergebnisse werden ohne externe Begutachtung im Eigenverlag der OECD veröffentlicht, und die Testaufgaben bleiben bis auf einen geringen Prozentsatz unter Verschluss.

⁷⁸ Max-Planck-Institut für Bildungsforschung, PISA 2000: Die Studie im Überblick. Grundlagen, Methoden und Ergebnisse, Berlin 2002, S. 1.

⁷⁹ Der für 2021 geplante Test wurde wegen der Covid19-Pandemie auf 2022 verschoben.

Problematisches Nationenranking

Was von PISA in der medialen Öffentlichkeit wahrgenommen wird, ist vor allem das an die Medaillenspiegel internationaler Sportereignisse erinnernde Nationen-Ranking. Die erste PISA-Studie führte zu Pilgerreisen nicht nur deutscher Bildungsdelegationen nach Finnland, das von 2000 bis 2006 dreimal den Spitzenplatz einnahm und jetzt als bewundertes Vorbild galt. Weshalb Finnland danach von Mal zu Mal schlechter abschnitt und seinen Spitzenplatz an asiatische Länder verlor, wurde hingegen kaum hinterfragt.⁸⁰ Jetzt galt die von PISA-Koordinator Andreas Schleicher befeuerte Bewunderung den PISA-Siegern Shanghai (2009 und 2012), Singapur (2015) und den chinesischen Provinzen Beijing, Shanghai, Jiangsu und Zhejiang, die 2018 zusammen auf dem ersten Platz landeten.

Doch schon vor Jahren haben Forscher davor gewarnt, aus dem Ranking lineare Schlüsse auf die Qualität des jeweiligen Schulsystems zu ziehen, da die zugrundeliegenden Datensätze nicht durchweg vergleichbar sind.⁸¹ PISA arbeitet nämlich mit Stichproben, die oft weniger als ein Prozent der Zielgruppe umfassen. Wenn eine Stichprobe nicht korrekt gezogen wird, kommt es zu falschen Ergebnissen. So etwa in Österreich, wo der PISA-Test 2003 deutlich schlechter ausfiel als in der ersten Runde drei Jahre zuvor. Nach Überprüfung durch versierte Statistiker mussten dann aber die Daten von 2000 nach unten korrigiert werden. Auch die Stichprobe der USA von 2009 wurde von amerikanischen Forschern als unzuverlässig kritisiert. Nach ihren korrigierten Punktzahlen hätten die USA im Ranking um zehn und mehr Plätze besser abgeschnitten.⁸²

Ein durchgehendes und daher gravierenderes Problem besteht darin, dass zumeist nicht alle für den Test vorgesehenen Schülerinnen und Schüler an der Erhebung teilnehmen. Zur Begrenzung des dadurch entstehenden Fehlers gilt bei PISA eine Mindestbeteiligungsquote von 80 Prozent. Als Großbritannien diese Quote 2003 knapp verfehlte, wurde es in internationalen Vergleichen nicht berücksichtigt.⁸³ Doch als Shanghai 2012 ebenfalls diese Hürde riss, setzte sich die OECD über ihre eigenen Regeln hinweg und präsentierte Shanghai als Sieger und

80 Hierzu Thelma von Freymann, PISA in Finnland. Die deutsche Bildungsdiskussion und die finnischen Realitäten, in: Seminar 3 (2002), S. 115-123; dies., Ein anderes Land, eine andere Schule. Zu den finnischen PISA-Ergebnissen, in: Neue Sammlung 43, 2003, S. 179-201; Gabriel Heller Sahlgreen, Real Finnish Lessons. The true story of an education superpower, London 2015.

81 Judith D Singer/Henry I. Braun: Testing international education assessments. Rankings get headlines, but often mislead, in: Science 360 (6384), 2018, 38-40; Svein Sjøberg, PISA as a Challenge for Science Education. Inherent Problems and Problematic Results from a Global Assessment Regime, in: Revista Brasileira de Pesquisa em Educação em Ciências 17(1), 2017, 327-363.

82 Vgl. Rainer Bölling, Zur Fragwürdigkeit des PISA-Rankings. Gesellschaft für Bildung und Wissen, 21. 11. 2017. <https://bildung-wissen.eu/fachbeitraege/zur-fragwuerdigkeit-des-pisa-rankings.html>.

83 OECD: Lernen für die Welt von morgen. Erste Ergebnisse von PISA 2003, Paris 2004, S. 371 u. 375-377.

Vorbild für die Welt. Das war umso fragwürdiger, als dort die Kinder von Millionen Wanderarbeitern außen vor blieben, weil sie gar nicht die höheren Schulen besuchen durften, an denen der PISA-Test stattfand.⁸⁴

Der Ausschluss schwächerer Schüler führt zu besseren Testergebnissen

Die OECD räumt selbst ein: Wenn „der Erfassungsgrad außer Acht gelassen wird, dürften die Leistungen der 15-jährigen Kohorte eines Bildungssystems überschätzt werden, und zwar umso mehr, je niedriger der Erfassungsgrad ist“.⁸⁵ Offensichtlich machen manche Länder mehr oder weniger ausgiebig von der Möglichkeit Gebrauch, „Schüler mit kognitiven oder körperlichen Behinderungen oder mangelnden Kenntnissen in der Testsprache (Schüler, die z.B. weniger als ein Jahr lang Unterricht in der Testsprache hatten)“ vom Test auszuschließen.⁸⁶ Das geschah 2018 besonders in Schweden, wo laut OECD „die hohe Zahl der Schülerinnen und Schüler, die seit 2015 als Zuwanderer oder Flüchtlinge in das Land gekommen waren, zu einem ... starken Rückgang des Erfassungsindex ... führte“.⁸⁷ So wurden 2018 nur 86 Prozent aller Fünfzehnjährigen vom PISA-Test erfasst, und Schweden verbesserte sich gegenüber den vorhergehenden Erhebungen deutlich. Auch in Ländern wie dem stets als Vorbild hingestellten Kanada werden regelmäßig ca. 15 Prozent eines Jahrgangs gar nicht erfasst.

Die OECD unterschätzt die Auswirkungen auf das Nationen-Ranking

Die Auswirkungen des unterschiedlichen Erfassungsgrades auf das PISA-Ranking spielt die OECD stets herunter. Seit 2003 findet sich in den Berichtsbänden – nicht aber in den Pressemitteilungen – an unterschiedlichen Stellen folgender Satz: „Selbst wenn man z. B. unterstellt, dass die ausgeschlossenen Schülerinnen und Schüler durchgehend schlechtere Ergebnisse erzielt hätten als die teilnehmenden und dass dieser Zusammenhang mittelstark ausgeprägt ist, würde eine Ausschlussquote in einer Größenordnung von 5% wahrscheinlich nur zu einer Überzeichnung der nationalen Durchschnittsergebnisse um weniger als 5 Punkte auf der PISA-Skala führen.“⁸⁸ Eine einfache Modellrechnung zeigt jedoch, dass dieser Effekt

⁸⁴ OECD: PISA 2012 Ergebnisse: Was Schülerinnen und Schüler wissen und können. Band I (überarbeitete Ausgabe): Schülerleistungen in Lesekompetenz, Mathematik und Naturwissenschaften, 2014, S. 290 f.; Tom Loveless, Attention OECD-PISA: Your-Silence-on-China-is-Wrong, 12.12.2013 (<http://educationnext.org/attention-oecd-pisa-your-silence-on-china-is-wrong>); Valerie Strauss, Did Shanghai cheat on PISA? 15.12.2013 (<http://www.washingtonpost.com/blogs/answer-sheet/wp/2013/12/15/did-shanghai-cheat-on-pisa/>).

⁸⁵ OECD: PISA 2018 Ergebnisse (Band I), Was Schülerinnen und Schüler wissen und können, 2019, S. 57.

⁸⁶ <http://www.oecd.org/berlin/themen/pisa-studie/haeufig-gestellte-fragen.htm> (10.01.2021).

⁸⁷ OECD: PISA 2018, S. 58.

⁸⁸ Zuletzt OECD: PISA 2018, S. 179.

doppelt so stark sein kann.⁸⁹ Das heißt nichts anderes, als dass das PISA-Ranking bei Berücksichtigung des Erfassungsgrades deutlich anders ausfallen würde. Hätte sich Deutschland, das 2018 mit 99,3 Prozent den höchsten Wert aller teilnehmenden Länder erreichte, mit dem durchschnittlichen OECD-Erfassungsgrad von 89 Prozent oder – wie Kanada und Schweden – mit noch weniger begnügt, wäre es wahrscheinlich in die Spitzengruppe aufgestiegen.

TIMSS macht es besser

Vor dem Problem unterschiedlicher Teilnahmequoten stehen auch andere internationale Vergleichsstudien wie TIMSS (Trends in Mathematics and Science Study), deren letzte Erhebung 2019 durchgeführt wurde. TIMSS geht aber offener damit um. Hier werden Abweichungen von den vorher definierten Stichprobenumfängen nachvollzogen und dokumentiert, „damit bei vertiefenden Vergleichen von Staaten nur solche einander gegenübergestellt werden, die tatsächlich aufgrund gleicher Teilnahmebedingungen vergleichbar sind“.⁹⁰ So findet sich unter 49 Tabellen mit Darstellungen der Ergebnisse bei gut 20 Teilnehmerstaaten folgende Fußnote: „Der Ausschöpfungsgrad und/oder die Ausschlüsse von der nationalen Zielpopulation erfüllen nicht die internationalen Vorgaben.“ In den meisten Fällen liegt die Ausschlussquote zwischen fünf und zehn Prozent, nur im Falle des Spitzenreiters Singapur deutlich darüber (12,8).⁹¹ Zudem werden in zwei Tabellen die Mittelwerte für Deutschland dargestellt, die sich ergeben würden, wenn leistungsschwächere Schülerinnen und Schüler bis zum Erreichen einer fiktiv höheren Ausschlussquote ausgeschlossen würden.⁹² Die OECD hingegen präsentiert ihre Vergleichstabellen ohne jeden Hinweis auf die angesprochenen methodischen Probleme.

Um Unterschiede in den Schülerleistungen zu veranschaulichen, hat die OECD lange Zeit PISA-Punkte in Schuljahre umgerechnet. So galten 2009 und 2012 ca. 125 PISA-Punkte als Äquivalent für drei Schuljahre, 2015 waren es noch 100.⁹³ Wie fragwürdig solche Umrechnungen sind, wird deutlich, wenn man sie auf die Leistungsspreizung innerhalb der

⁸⁹ Rainer Bölling, Bessere PISA-Ergebnisse durch selektives Testen, 2021. <https://bildung-wissen.eu/kommentare/bessere-pisa-ergebnisse-durch-selektives-testen.html>.

⁹⁰ Knut Schwippert u.a.: TIMSS 2019. Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich, Münster 2020, S. 39.

⁹¹ TIMSS 2019, S. 354.

⁹² TIMSS 2019, S. 40-42.

⁹³ OECD: PISA 2009 Ergebnisse: Was Schülerinnen und Schüler wissen und können. Schülerleistungen in Lesekompetenz, Mathematik und Naturwissenschaften (Band I), 2010, S. 13 u. ö.; PISA 2012, S. 4 u. 272; OECD: PISA 2015 Ergebnisse (Band I): Exzellenz und Chancengerechtigkeit in der Bildung, 2016, S. 41 u. 162.

Länder anwendet, die fast überall 300 und mehr Punkte beträgt.⁹⁴

Verwechslung von Korrelation und Kausalität

Ein grundsätzliches Problem im Umgang mit den PISA-Daten besteht darin, dass in öffentlichen Verlautbarungen oft nicht zwischen Korrelation und Kausalität unterschieden wird, was zu krassen Fehlinterpretationen führen kann. Zwar heißt es in einer PISA-Publikation von 2006: „Korrelationen lassen keine Rückschlüsse auf Faktoren zu, die für die Kompetenzentwicklung kausal relevant sind.“⁹⁵ Diese Erkenntnis hindert PISA-Offizielle allerdings nicht daran, in der Öffentlichkeit fragwürdige Rückschlüsse als gesicherte Erkenntnisse auszugeben. So deuteten sie die Verbesserung der deutschen PISA-Daten bis 2012 als Resultat des von ihnen befürworteten Ausbaus der frühkindlichen Erziehung. 2015 behauptete der stellvertretende Generalsekretär der OECD sogar, man könne „in den PISA-Zahlen ganz deutlich messen, dass bereits ein Jahr frühkindliche Bildung nachher die Ergebnisse im PISA-Test deutlich verbessert“.⁹⁶ Doch aus den damals aktuellen PISA-Zahlen lässt sich nichts dergleichen entnehmen, denn die 2012 getesteten Fünfzehnjährigen wurden ja spätestens 2003 eingeschult, also vor dem Ausbau der frühkindlichen Bildung in Deutschland. Worauf auch immer die Verbesserung der Testergebnisse gegenüber früheren Jahrgängen zurückzuführen sein mag – die frühkindliche Erziehung kann nicht der Grund sein.⁹⁷

Fragwürdige Interpretationen

Auch andere PISA-Erfolgsmeldungen halten einer Überprüfung nicht stand. So betont Andreas Schleicher in seinem „Weltklasse“ betitelten Buch: „Es wurden nationale Bildungsstandards für Schulen eingeführt (was in einem Land, in dem die Eigenständigkeit der Bundesländer immer sakrosankt war, bis dahin kaum vorstellbar war), und benachteiligte Schülerinnen und Schüler, darunter insbesondere jene mit Migrationshintergrund, erhielten stärkere Unterstützung. 2009 [...] schnitt Deutschland sehr viel besser ab. Es hatte sich in punkto Qualität und Chancengerechtigkeit deutlich gesteigert.“⁹⁸ Doch warum fielen die deutschen PISA-

⁹⁴ Berechnet nach PISA 2015, S. 351, 404 u, 417.

⁹⁵ Manfred Prenzel u.a.: PISA 2003. Untersuchungen zur Kompetenzentwicklung im Verlauf eines Schuljahres, Münster 2006, S. 15.

⁹⁶ Christiane Habermalz, Große Fortschritte bei der frühkindlichen Bildung (2015). https://www.deutschland-funkkultur.de/oeed-bildungsstudie-grosse-fortschritte-bei-der.2165.de.html?dram:article_id=337814 (02.10.2018).

⁹⁷ Vgl. Rainer Bölling, Das deutsche Bildungswesen – ein Hort der sozialen Ungerechtigkeit? Kritische Anmerkungen zu neueren vergleichenden Bildungsstudien, in: Pädagogische Korrespondenz. Zeitschrift für kritische Zeitdiagnostik in Pädagogik und Gesellschaft, Heft 59/2019, S. 37-55, hier S. 42.

⁹⁸ Andreas Schleicher, Weltklasse. Schule für das 21. Jahrhundert gestalten, Bielefeld 2019, S. 25.

Ergebnisse dann 2015 und 2018 wieder schlechter aus? Hierfür liefern die PISA-Verantwortlichen keine Erklärung.

Noch deutlicher wird dieses Defizit im Falle Finnlands. So liest man bei Schleicher: „Schülerinnen und Schüler, die in Rückstand geraten, werden rasch ermittelt, ihr Problem wird umgehend und präzise diagnostiziert, und es werden zügig geeignete Maßnahmen ergriffen.“⁹⁹ Folgt man den PISA-Daten, scheint das aber nichts zu bewirken, denn auch in Finnland beträgt der Abstand zwischen den oberen und unteren fünf Prozent der Schülerschaft rund 300 PISA-Punkte.¹⁰⁰ Doch nach den Ursachen für solche statistischen Befunde wird nicht gefragt, könnte doch die Antwort Zweifel am Mythos des bildungspolitischen Musterlandes aufkommen lassen.

Was hat PISA gebracht?

Nach ihrer eigenen Einschätzung „spielt die OECD eine zentrale Rolle bei der Bereitstellung von Indikatoren zu Bildungsleistungen, die staatliche Bildungspolitik nicht nur bewerten, sondern auch zu ihrer Gestaltung beitragen.“¹⁰¹ Doch 2019 musste sie einräumen, „angesichts des in den letzten zehn Jahren im OECD-Raum verzeichneten Anstiegs der Ausgaben je Schüler im Primar- und Sekundarbereich um mehr als 15%“ sei es „enttäuschend, dass sich die Schülerleistungen in den meisten OECD-Ländern seit der ersten PISA-Erhebung im Jahr 2000 praktisch nicht verbessert haben“.¹⁰² Wenn das das Ergebnis von PISA ist, stellt sich die Frage, wozu das mit hohen Ausgaben verbundene Programm eigentlich gut ist. Jedenfalls wäre zu wünschen, dass eine offene Diskussion über Ziele, Methoden und Ergebnisse der Studie nachgeholt wird.

Offizielle PISA-Beispielaufgaben: <https://www.pisa.tum.de/pisa/beispielaufgaben/>

Beispiel einer PISA-Aufgabe mit Kommentierung: Das Klon-Schaf Dolly (2003)
<https://docplayer.org/130084315-Was-hinter-pisa-steckt-die-gaukelei-der-pisa-tests.html>

<https://www.schulverwaltung.de/bibliothek/recherche-im-werk/document/0/0:9985026,0>

⁹⁹ Schleicher, S. 82.

¹⁰⁰ Berechnet nach: PISA 2018, S. 240-245.

¹⁰¹ OECD: Bildung auf einen Blick 2011. OECD-Indikatoren, BMBF 2011, S. 13.

¹⁰² PISA 2018, S. 3. Ähnlich Schleicher, Weltklasse, S. 11.

Zweifel an der Sinnhaftigkeit der PISA-Daten

Ein Blick auf die internationalen Ergebnisse, den Absturz Finnlands
und manche deutsche Kurzschluss-Reaktion

Viel entspannter als in Deutschland geht man mit schlechten Pisa-Ergebnissen in einem anderen Land um, das jüngst einen beispiellosen Absturz erlebt hat: um volle 45 Punkte ist die durchschnittliche Punktzahl in den drei Testbereichen Lesen, Mathematik und Naturwissenschaften gegenüber der ersten Ausgabe im Jahre 2000 gefallen – gegenüber fünf Punkten in Deutschland.¹⁰³ Es handelt sich um Finnland, Pisa-Sieger der Jahre 2000 bis 2006, das damals zum Ziel von Pilgerreisen deutscher Pädagogen wurde. Dort stellte das Erziehungsministerium jetzt nüchtern fest, die Pisa-Werte seien infolge der Corona-Pandemie überall gefallen, und Finnland liege immer noch über dem Durchschnitt der OECD. Und es bemerkte auch, dass die Pisa-Ergebnisse keine erschöpfende Antwort auf die Frage geben, wie viel von der Verschlechterung der Lernergebnisse auf die Pandemie und wie viel auf andere Faktoren zurückzuführen ist.¹⁰⁴ Tatsächlich handelt es sich ja – wie auch die OECD einräumt – bei den Pisa-Daten um „Querschnittsdaten, mit denen keine Kausalitäten nachgewiesen werden können“.¹⁰⁵

Zu den wenigen Ländern, die sich gegenüber Pisa 2018 nicht verschlechtert haben, gehört die Schweiz. Ihre Ergebnisse liegen jetzt deutlich über dem Durchschnitt der 37 OECD-Mitgliedstaaten, weil viele, die früher vor ihr lagen – darunter Deutschland –, bei Pisa 2022 zum Teil deutlich schwächer abgeschnitten haben. Schweizer Experten vermuten, dass dieser Trend auf die Corona-Pandemie zurückzuführen ist.¹⁰⁶ Denn in der Schweiz waren die Schulen nur 34 Unterrichtstage vollständig geschlossen, in Österreich dagegen 74 und in Deutschland 85 Tage. Andererseits schnitt Estland mit 95 Tagen Schulschließungen von allen europäischen Ländern am besten ab.¹⁰⁷ Das könnte aber auch daran liegen, dass die Digitalisierung des Unterrichts in Estland wohl am weitesten fortgeschritten ist. Jedenfalls ist der Zusammenhang zwischen corona-bedingtem Unterrichtsausfall und Leistungsentwicklung nicht so einfach zu erklären, wie man vermuten könnte.

¹⁰³ PISA 2022 Ergebnisse (Band I): Lernstände und Bildungsgerechtigkeit, OECD 2023.

¹⁰⁴ https://valtioneuvosto.fi/en/-/1410845/pisa-2022-performance-fell-both-in-finland-and-in-nearly-all-other-oecd-countries_

¹⁰⁵ PISA 2022 Ergebnisse (Band I), S. 244.

¹⁰⁶ <https://www.admin.ch/gov/de/start/dokumentation/medienmitteilungen.msg-id-99216.html>;
<https://www.nzz.ch/schweiz/pisa-erhebung-schweizer-schueler-schneiden-in-mathematik-sehr-gut-beim-lesen-hapert-es-ld.1768827>.

¹⁰⁷ OECD: Bildung auf einen Blick 2022, S. 30.

In Deutschland wurden wieder alte Erklärungsmuster bemüht, die sich längst als obsolet erwiesen haben. So beklagte die Gewerkschaft Erziehung und Wissenschaft (GEW), dass sich „die Abhängigkeit der schulischen Leistungen der Kinder und Jugendlichen vom Elternhaus seit über 20 Jahren nicht verringert“ habe.¹⁰⁸ Tatsächlich aber ist der einschlägige Pisa-Messwert in Deutschland gegenüber dem Jahr 2000 trotz Zuwanderung vieler Armutsflüchtlinge von 22,7 auf 18,7 gesunken. In der OECD insgesamt ist er dagegen von 14,3 auf 15,5 gestiegen. Dabei ist dieser Wert schon deshalb mit Vorsicht zu interpretieren, weil er allein auf nicht überprüfbaren Angaben der Getesteten beruht.

Die GEW behauptet: „Die Schülerleistungen werden sich nie in der Breite verbessern, wenn wir die Kinder weiterhin so früh auf hierarchische Schulformen aufteilen.“ Da könnte Frankreich als Vorbild dienen, das seit den Siebzigerjahren ein Einheitsschulsystem kennt. Tatsächlich wies unser Nachbarland bei den ersten beiden Pisa-Durchgängen bessere Werte auf als Deutschland. Das führte zusammen mit guten Befunden skandinavischer Länder dazu, dass Pisa-Koordinator Schleicher sich in Deutschland – und ebenso Österreich – für die Einführung eines solchen Systems stark machte und gleichgesinnte Akteure im Bildungsbereich sich nun auf Pisa berufen konnten. Doch von 2006 bis heute hat Frankreich bei den Leistungspunkten und seit 2012 auch beim Messwert für Bildungsgerechtigkeit stets schlechter abgeschnitten als Deutschland. Auch den Mythos der nordischen Bildungssysteme hat die seriöse Erziehungswissenschaft längst entzaubert.¹⁰⁹ Doch solche unliebsamen Befunde nehmen Befürworter des längeren gemeinsamen Lernens ungern zur Kenntnis.

In Frankreich wurden die jüngsten Pisa-Ergebnisse jedenfalls mit Ernüchterung aufgenommen. Die renommierte Tageszeitung „Le Monde“ konstatierte einen beispiellosen Leistungsrückgang der französischen Schüler. Das Land, dem mathematische Kompetenzen in besonderem Maße als Ausweis von Exzellenz gelten, schmerzt es besonders, dass der Pisa-Mittelwert in Mathematik, der zwischen 2006 und 2018 stabil geblieben war, jetzt um 21 Punkte auf den niedrigsten je gemessenen Wert gesunken ist.¹¹⁰ Dabei ist in Frankreich wegen der Corona-Epidemie nicht mehr Unterricht ausgefallen als in der Schweiz. Schnell kündigte der Erziehungsminister denn auch Konsequenzen an. Ab 2024 werden in den Klassen der französischen Sekundarstufe I in Französisch und Mathematik Niveaugruppen eingeführt, wobei die Gruppen mit den lernschwachen Schülern kleiner sein sollen. Zudem kommt

¹⁰⁸ <https://www.gew.de/presse/pressemitteilungen/detailseite/gew-pisa-ergebnisse-spiegeln-lehr-und-fachkraeftemangel-wider>

¹⁰⁹ Mareike Kobarg / Manfred Prenzel, Stichwort: Der Mythos der nordischen Bildungssysteme, in: Zeitschrift für Erziehungswissenschaft 13 (2009), S. 597–615.

¹¹⁰ https://www.lemonde.fr/societe/article/2023/12/05/pisa-2022-une-baisse-sans-precedent-des-resultats-des-eleves-francais-comme-partout-ailleurs-dans-le-sillage-du-covid-19_6203983_3224.html.

demnächst zu den Prüfungen in Französisch am Ende der vorletzten Klasse vor dem Abitur für alle Schüler eine neue Prüfung in Mathematik und Naturwissenschaften hinzu. Schließlich soll die Entscheidungsbefugnis der Lehrer über Klassenwiederholungen gestärkt werden, deren Zahl in den letzten 20 Jahren stark rückläufig war.¹¹¹ Solche Maßnahmen aber stoßen längst nicht überall auf Zustimmung.

Welche Länder könnten nun Deutschland und anderen als bildungspolitisches Vorbild dienen, nachdem Finnland so abgestürzt ist? Schanghai und die drei weiteren chinesischen Provinzen, die 2009, 2012 und 2018 trotz regelwidrig kleiner bzw. grenzwertiger Stichproben als Pisa-Sieger gefeiert wurden, tauchen 2022 entgegen früherer Ankündigung gar nicht mehr auf. Das mag der Corona-Pandemie geschuldet sein, wird aber von der OECD nicht erläutert. Singapur als Pisa-Sieger von 2015 und 2022 taugt als Stadtstaat mit weniger als sechs Millionen Einwohnern und der höchsten Millionärsdichte der Welt auch nur bedingt als Vorbild für einen Flächenstaat wie Deutschland. Ähnliches gilt für „Tigerstaaten“ wie Taiwan, Hongkong und Macau, die mit Abstand auf den nächsten Plätzen des Pisa-Rankings folgen.

Die ersten Flächenstaaten im aktuellen Ranking sind Japan und Korea. Sucht man nach Erklärungen für ihr gutes Abschneiden, so kommt man nicht umhin, den Anteil von Schülern mit Migrationshintergrund ins Auge zu fassen, der bei Pisa 2022 erstmals für alle Länder in einer Tabelle ausgewiesen wird.¹¹² Dabei zeigt sich, dass in diesen beiden Ländern ihr Anteil unter einem Prozent liegt. Fast alle Probanden sind demnach im Lande mit der Unterrichts- und damit Testsprache aufgewachsen. In Deutschland dagegen hat sich der Anteil der Schüler mit Migrationshintergrund seit 2012 von 13 auf fast 26 Prozent verdoppelt. Fast zwei Drittel von ihnen gaben an, dass sie zu Hause meist eine andere Sprache sprechen als die, in der sie die PISA-Tests absolviert hatten. Das trägt zu einem Leistungsrückstand von 59 Punkten gegenüber Schülern ohne Migrationshintergrund bei.¹¹³ Andere europäische Länder wie Belgien, Österreich und die Schweiz weisen ebenfalls Migrantenquoten von über 20 Prozent auf; sie schneiden bei Pisa bei der Bildungsgerechtigkeit – ebenso wie Frankreich – schlechter ab als Deutschland.

Das häufig als Musterbeispiel erfolgreicher Integration gelobte Kanada weist zwar sogar ein gutes Drittel Immigranten auf, verfolgt aber seit langem eine streng selektive Einwanderungspolitik. Folglich werden nur qualifizierte Arbeitskräfte mit entsprechender Bildung und

¹¹¹ <https://www.lefigaro.fr/actualite-france/redoublement-groupes-de-niveau-nouvelle-epreuve-du-bac-en-premiere-ce-que-va-annoncer-gabriel-attal-20231205>.

¹¹² PISA 2022, S. 368 f.

¹¹³ PISA 2022 Country Notes Deutschland, S. 6.

Sprachkenntnissen ins Land geholt. Zudem sind die Landessprachen Englisch und Französisch in den Herkunftsländern der Immigranten, von denen der größte Teil aus Indien kommt, weit verbreitet. Kanada kann daher nur bedingt als Vorbild zur Lösung von Problemen dienen, vor denen viele Länder Europas mit unkontrollierter Migration stehen. Und das betrifft in erster Linie Deutschland, das seit 2015 die meisten Asylsuchenden aufgenommen hat. Unter ihnen stellen Syrer vor Afghanen die größte Gruppe. Zum Bildungsniveau in Syrien aber hat der Bildungsökonom Ludger Wößmann 2015 desillusionierende Daten gefunden: Zwei Drittel der Schüler konnten nur sehr eingeschränkt lesen und schreiben, nur einfachste Rechenaufgaben lösen. Sein Fazit lautete, „dass diese Schüler in Deutschland, selbst wenn sie Deutsch gelernt haben, kaum dem Unterrichtsgeschehen folgen können.“¹¹⁴

Somit zeigt sich bei genauerem Hinsehen, dass der Anspruch der OECD, aus den Pisa-Daten zu lernen und Verbesserungen im Bildungswesen anzustoßen, schwieriger einzulösen ist als oft angenommen. Gegen diese Erkenntnis haben Vertreter der OECD allerdings selbst wiederholt verstoßen, indem sie sachlich fragwürdige bzw. falsche Erklärungen in Umlauf brachten. 2019 mussten sie zudem einräumen, angesichts des in den letzten zehn Jahren im OECD-Raum verzeichneten Anstiegs der Ausgaben je Schüler um mehr als 15% sei es „enttäuschend, dass sich die Schülerleistungen in den meisten OECD-Ländern seit der ersten Pisa-Erhebung im Jahr 2000 praktisch nicht verbessert haben“.¹¹⁵ Wenn darin das Ergebnis von Pisa besteht, stellt sich die Frage, wozu das mit hohen Kosten verbundene Programm eigentlich gut ist.

¹¹⁴ <http://www.zeit.de/2015/47/integration-fluechtlinge-schule-bildung-herausforderung>.

¹¹⁵ OECD: PISA 2018 Ergebnisse (Band I), Was Schülerinnen und Schüler wissen und können, 2019, S. 3. Ähnlich Andreas Schleicher, Weltklasse. Schule für das 21. Jahrhundert gestalten, Bielefeld 2019, S. 11.