

Rainer Bölling

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 verlautet, 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 Bronzemedaillen. 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.

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

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 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 abschneidenden 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 zugrunde liegende 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

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/>)

der bei PISA erfassten 15-Jährigen (2012: 108.056) im Verhältnis zur Gesamtbevölkerung Shanghais 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

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>).

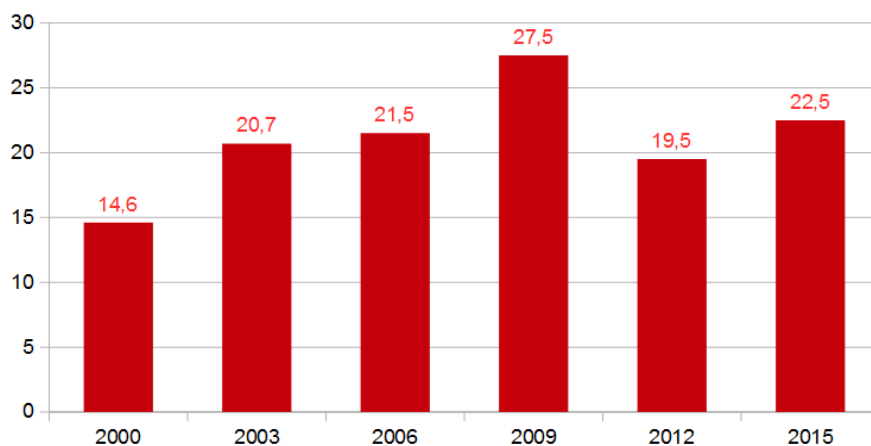
11 Siehe Yong Zhao, *Who's Afraid of the Big Bad Dragon? Why China Has the Best (and Worst) Education System in the World*, 2014. Dazu die Rezension von Diane Ravitch, *The Myth of Chinese Super Schools*, 20.11.2014 (<http://www.nybooks.com/articles/2014/11/20/myth-chinese-super-schools>)

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>)

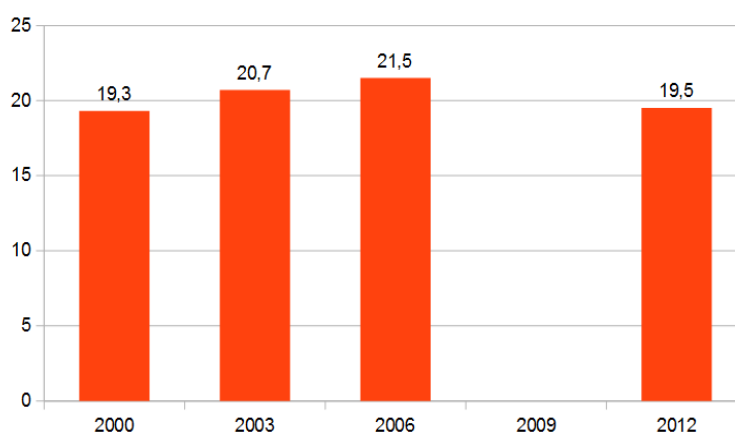
wiedergegeben werden soll. Auf jeden Fall ließ sie deutlich werden, welche Probleme allein schon die korrekte Bestimmung einer nationalen Stichprobe aufwirft.

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 (Anm. 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 Lehrergewerkschaften 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 (

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.

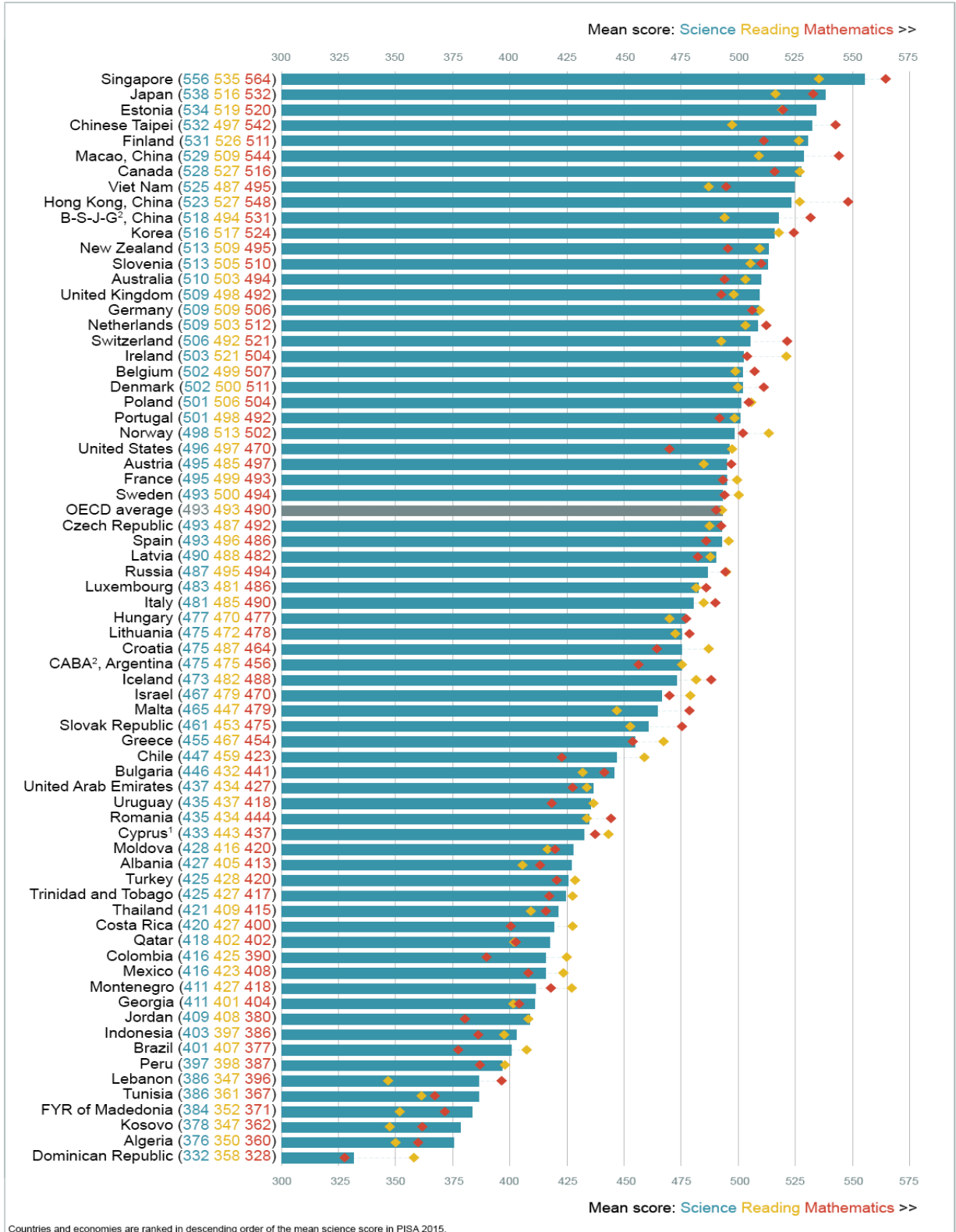
Dr. Rainer Bölling
Düsseldorf
www.rboelling.de
November 2017

wissen.eu/wp-content/uploads/2014/05/offener-brief-schleicher-autorisierte-fassung.pdf), der von mehreren 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