

1.2.6 Kontroverse

Die internationalen Studien, TIMSS und PISA sind zum Teil kontrovers zur Kenntnis genommen worden.

In Großbritannien wäre Prais³ zu nennen. In der Bundesrepublik wies Kraus bereits in der Mathematikstudie (TIMSS) auf die unterschiedliche Vorauswahl der Schüler hin.⁴ Collani verwies auf statistische Sachverhalte, die die internationalen Testergebnisse in Frage stellten.⁵ Brügelmann wies auf Inkonsistenzen zwischen den einzelnen Studien hin.⁶ Klemm stellte das relativ gute Abschneiden der Bundesländer Bayern und Baden-Württemberg in PISA-E in Frage,⁷ Fuchs sprach das Problem öffentlicher Kontrolle der OECD-Studien an.⁸

2 Evaluation

Ausgangspunkt vergleichender, empirischer Untersuchungen ist die präzise Bestimmung des Untersuchungsgegenstandes und die damit verbundene Frage nach einer geeigneten Stichprobendefinition. In der Schulvergleichsforschung sind insbesondere das Schulalter (Klassenstufe) und das Lebensalter von Bedeutung. Folglich wird die Stichprobendefinition an erster Stelle zum Gegenstand dieser Evaluation.

2.1 Stichprobendefinitionen internationaler Vergleichsstudien

Die PISA-Studien stehen in Zusammenhang vorangegangener und nachfolgender internationaler Vergleichsuntersuchungen, insbesondere der Dritten Internationalen Mathematik- und Naturwissenschaftsstudie (TIMSS) - mit unterschiedlichen Populationsdefinitionen sowie der Internationalen Grundschulstudie (IGLU).

³ S. J. Prais, Oxford Review of Education, Vol. 29, June, 2003. Prais: National Institute of Economic and Social Research (NIESR)

⁴ Kraus: Apokalyptische Inszenierung und dilettantischer Leichtsin. In: DEUTSCHER LEHRERVERBAND (DL)

⁵ Collani, E.v.: OECD PISA – An Example of Stochastic Illiteracy?, Economic Quality Control Vol 16 (2001), No.2, S. 240

⁶ Brügelmann, H.: Besserwisser und Alleskönner, Gewerkschaft Erziehung und Wissenschaft 11/12/2001

⁷ Klemm, K.: PISA E – Erste Einschätzungen (Frankfurter Rundschau, 26.7.2002)

⁸ Fuchs, H.-W.: Auf dem Weg zu einem Weltcurriculum? Zeitschrift für Pädagogik. (2003) S.161ff

2.1.1 TIMSS

Die Dritte Internationale Mathematik- und Naturwissenschaftsstudie (TIMSS) untersucht drei verschiedene Populationen.⁹

2.1.1.1 Population I

Population I entspricht der Grundschule. Ihr gehören Schülerinnen und Schüler derjenigen beiden angrenzenden Klassenstufen an, die zum Testzeitpunkt den größten Anteil der 9-Jährigen umfasst. Diese Population wurde in Deutschland nicht untersucht.

*Population 1: all students enrolled in the two adjacent grades that contain the largest proportion of students of age 9 years at the time of testing.*¹⁰

2.1.1.2 Population II

Die Population II entspricht der Sekundarstufe I. Ihr gehören Schülerinnen und Schüler der beiden angrenzenden Klassenstufen an, die zum Testzeitpunkt den größten Anteil der 13-Jährigen aufweisen. In Deutschland wurde jedoch wegen der relativ späten Einschulung der Schüler auf die Altersvorschrift verzichtet und stattdessen die Schuljahrgänge 7 und 8 gezogen.

*Population 2: all students enrolled in the two adjacent grades that contain the largest proportion of students of age 13 years at the time of testing.*¹¹

2.1.1.3 Population III

Population III entspricht der Sekundarstufe II.

Population 3: all students in their final year of secondary education, including students in vocational education programs. Population 3 has two optional subpopulations:
- Students taking advanced courses in mathematics

⁹ Baumert, Köller, Lehrke, Brockmann: Anlage und Durchführung der Dritten Internationalen Mathematik- und Naturwissenschaftsstudie zur Sekundarstufe II (TIMSS/III)- Technische Grundlagen. In: Baumert, Bos, Lehmann (Hrsg.): TIMSS/III Dritte Internationale Mathematik- und Naturwissenschaftsstudie, 2000, Band 1. S.31

¹⁰ Martin, Kelly (Hrsg.): TIMSS Technical Report 1996 Vol. 1. S. 16

¹¹ Martin, Kelly (Hrsg.): TIMSS Technical Report 1996 Vol. 1. S. 16

- *Students taking advanced courses in physics.*¹²

(Hier) *mussten für die Population III breitere Rahmendefinitionen festgelegt werden, die jeweils national angemessene Präzisierungen erlaubten. Die Schulsysteme der an TIMSS teilnehmenden Staaten sind in der Sekundarstufe II zu unterschiedlich organisiert, als dass sich eine enge Definition hätte finden lassen.*¹³

Die national angemessenen Präzisierungen umfassen ein mittleres Alter (arithmetisches Mittel des ungewichteten Datensatzes) von 17,1 Jahren/mittleres Geburtsjahr 1977,9 (Russische Föderation) bis 20,45 Jahren/mittleres Geburtsjahr 1974,55 (Island). Das Alter des jüngsten Teilnehmers betrug 14 Jahre (Geburtsjahr April 1981, ID-Stud 21581005, Russische Föderation), das des ältesten 59 Jahre (Geburtsjahr August 1936, ID-Stud 3240010, Dänemark).¹⁴

¹² Martin, Kelly (Hrsg.): TIMSS Technical Report 1996 Vol. 1. S. 16

¹³ Baumert, Köller, Lehrke, Brockmann: Anlage und Durchführung der Dritten Internationalen Mathematik- und Naturwissenschaftsstudie zur Sekundarstufe II (TIMSS/III)- Technische Grundlagen. In: Baumert, Bos, Lehmann (Hrsg.): TIMSS/III Dritte Internationale Mathematik- und Naturwissenschaftsstudie 2000, Band 1. S.32

¹⁴ Berechnung aus den Rohdaten TIMSS/III bezogen auf das Jahr 1995, ohne Berücksichtigung des für den einzelnen Schüler individuellen Geburts-, Testmonats sowie Testjahrs. Laut TIMSS Technical Report fand die Erhebung im August 1995 (südliche Hemisphäre) und Februar bis Mai 1995 (nördliche Hemisphäre) statt (TIMSS Technical Report, Vol.1. S. 16). In Island erfolgte die Untersuchung laut Datensatz im März 1996. Dies würde zu einem numerischen Durchschnittsalter von 21,4 Jahren führen. Der internationale Test fand nach Datenlage zwischen Februar 1995 und Mai 1996 statt. Der Testmonat in Österreich wird nicht genannt.

STUDENT'S DATE OF BIRTH\YEAR				
	Häufigkeit	Prozent	Gültige Prozenzte	Kumulierte Prozenzte
Gültig	36	,0	,0	,0
40	1	,0	,0	,0
41	1	,0	,0	,0
43	1	,0	,0	,0
45	2	,0	,0	,0
46	3	,0	,0	,0
47	2	,0	,0	,0
48	3	,0	,0	,0
49	3	,0	,0	,0
50	2	,0	,0	,0
51	2	,0	,0	,0
52	7	,0	,0	,0
53	6	,0	,0	,0
54	4	,0	,0	,0
55	9	,0	,0	,1
56	3	,0	,0	,1
57	12	,0	,0	,1
58	11	,0	,0	,1
59	12	,0	,0	,1
60	20	,0	,0	,1
61	15	,0	,0	,1
62	21	,0	,0	,2
63	16	,0	,0	,2
64	27	,0	,0	,2
65	26	,0	,0	,2
66	44	,0	,1	,3
67	51	,1	,1	,3
68	56	,1	,1	,4
69	87	,1	,1	,5
70	153	,2	,2	,7
71	228	,3	,3	,9
72	445	,5	,5	1,5
73	775	,9	,9	2,3
74	2359	2,6	2,7	5,0
75	8446	9,5	9,6	14,7
76	28881	32,4	33,0	47,6
77	31931	35,8	36,4	84,1
78	12659	14,2	14,4	98,5
79	1269	1,4	1,4	100,0
80	24	,0	,0	100,0
81	1	,0	,0	100,0
Gesamt	87619	98,4	100,0	
Fehlend	98	979	1,1	
System	482	,5		
Gesamt	1461	1,6		
Gesamt	89080	100,0		

Bericht			
STUDENT'S DATE OF BIRTH\YEAR			
COUNTRY ID	Mittelwert	N	Standardabweichung
Australia	77,57	3025	,695
Austria	76,11	3076	1,690
Canada	76,45	9532	,757
Cyprus	77,13	1161	,418
Czech Republic	76,72	3819	,673
Denmark	75,48	4363	2,130
France	76,32	3377	1,072
Germany	76,36	4873	1,610
Greece	78,20	803	,445
Hungary	77,01	4901	1,064
Iceland	74,55	1673	3,689
Israel	77,95	2390	,405
Italy	76,01	1974	1,167
Latvia (LSS)	76,83	708	,430
Lithuania	76,98	3620	,512
Netherlands	76,09	1454	1,202
New Zealand	77,36	1763	,608
Norway	75,32	3558	2,156
Russian Federation	77,90	4911	,382
South Africa	74,94	2629	2,726
Sweden	75,83	4572	1,071
Switzerland	75,17	5313	1,331
United States	76,67	10834	,606
Slovenia	76,14	3290	,621
Insgesamt	76,41	87619	1,532

Abbildung 1 Target Population TIMSS/III, Berechnung ungewichtet

Weiterhin unterscheiden sich die getesteten Schüler erheblich hinsichtlich ihrer Vorauswahl.¹⁵

¹⁵ Baumert, Köller, Lehrke, Brockmann: Anlage und Durchführung der Dritten Internationalen Mathematik- und Naturwissenschaftsstudie zur Sekundarstufe II (TIMSS/III) – Technische Grundlagen. In: Baumert, Bos, Lehmann (Hrsg.): TIMSS/III Dritte Internationale Mathematik- und Naturwissenschaftsstudie 2000, Band 1. S.41

2.1.2 PISA

*In PISA wurden Schülerinnen und Schüler untersucht, die zum Beginn des Testzeitraums zwischen 15 Jahren/drei Monaten und 16 Jahren/zwei Monaten alt waren – unabhängig von der besuchten Jahrgangsstufe oder Art der Bildungseinrichtung.*¹⁶

Der Technical Report der PISA-Studie definiert wie folgt:

The desired base PISA target population in each country consisted of 15-year-old students attending educational institutions located within the country. This meant that countries were to include (i) 15-year-olds enrolled full-time in educational institutions, (ii) 15-year-olds enrolled in educational institutions who attended on only a part-time basis, (iii) students in vocational training types of programmes, or any other related type of educational programmes, and (iv) students attending foreign schools within the country (as well as students from other countries attending any of the programmes in the first three categories). It was recognised that no testing of persons schooled in the home, workplace or out of the country would occur and therefore these students were not included in the International Target Population. The operational definition of an age population directly depends on the testing dates.

(...)

The 15-year-old International Target Population was slightly adapted to better fit the age structure of most of the Northern Hemisphere countries. As the majority of the testing was planned to occur in April, the international target population was consequently defined as all students aged from 15 years and 3 (completed) months to 16 years and 2 (completed) months at the beginning of the assessment period. This meant that in all countries testing in April 2000, the national target population could have been defined as all students born in 1984 who were attending a school or other educational institution. Further, a variation of up to one month in this age definition was permitted. For instance, a country testing in March or in May was still allowed to define the national target population as all

¹⁶ Baumert, Stanat, Demmrich: PISA 2000: Untersuchungsgegenstand, theoretische Grundlagen und Durchführung der Studie. In: Deutsches PISA-Konsortium (Hrsg.): PISA 2000: Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich. S. 34

*students born in 1984. If the testing was to take place at another time, the birth date definition had to be adjusted and approved by the Consortium.*¹⁷

2.1.3 IGLU

In der Internationalen Grundschul-Lese-Untersuchung (IGLU) wird die Zielgruppe wie folgt definiert:

*Die Zielgruppe der Untersuchung war „the upper of the two adjacent grades with the most 9-year-olds“ (...) In den meisten Ländern und auch in Deutschland entspricht diese Definition der vierten Jahrgangsstufe.*¹⁸

Die Studie orientiert sich an der Klassenstufe (zumeist 4. Klasse). In England, Neuseeland und Schottland wurden die Schüler jedoch am Ende des fünften Schuljahres getestet.¹⁹

Grob vereinfachend lässt sich feststellen, dass IGLU mit der Untersuchung von Schülern am Ende der vierten Jahrgangsstufe eine am Schulalter orientierte Definition vorsieht. Die PISA-Studie hingegen orientiert sich am Lebensalter der Probanden. Die Dritte Internationale Mathematik- und Naturwissenschaftsstudie (TIMSS) wiederum stellt eine Reihe unterschiedlicher Definitionen auf, in der Regel eine Kombination beider Varianten. Die Vergleichbarkeit der Studien wird später Gegenstand der Diskussion sein.

2.2 PISA – Vergleichbarkeit der Populationen innerhalb der Studie

Die im Rahmen der PISA-Studie fokussierte Gruppe der 15-Jährigen weist eine ungleiche Verteilung hinsichtlich der Variablen Schulalter über die teilnehmenden Länder auf. So findet sich in Deutschland der überwiegende Teil der 15-Jährigen in der 9. Klassenstufe.²⁰ Die folgende Abbildung gibt die Schulalterverteilung nach Teilnehmerstaat, entsprechend dem öffentlich zugänglichen Datensatz (Public Use File) wieder. Die Schulalterverteilung

¹⁷ Adams, Wu (Hrsg.): PISA 2000 Technical Report. S. 39

¹⁸ Lankes, Bos, Mohr, Plaßmeier, Schwippert, Sibberns, Voss: Anlage und Durchführung der Internationalen Grundschul-Lese-Untersuchung (IGLU) und ihrer Erweiterung um Mathematik und Naturwissenschaften (IGLU-E) Bos, Lankes, Prenzel, Schwippert, Walther, Valtin (Hrsg.): Erste Ergebnisse aus IGLU. S. 7

¹⁹ Lankes, Bos, Mohr, Plaßmeier, Schwippert: Lehr- und Lernbedingungen in den Teilnehmerländern Bos, Lankes, Prenzel, Schwippert, Walther, Valtin (Hrsg.): Erste Ergebnisse aus IGLU 2003. S. 38

der bundesdeutschen Schülerinnen und Schüler steht in Kontrast zu einer Anzahl weiterer Teilnehmerstaaten. Diese weisen einen zum Teil erheblichen Anteil in höheren Klassenstufen auf.

Country three-digit ISO code * Grade - Q2 Kreuztabelle

% von Country three-digit ISO code		Grade - Q2										Gesamt
Country three-digit ISO code		4	5	6	7	8	9	10	11	12	13	
AUSTRALIA						,1%	6,3%	76,7%	16,9%			100,0%
AUSTRIA				,0%	,3%	4,8%	42,1%	52,7%	,1%			100,0%
BELGIUM					,2%	4,2%	26,3%	68,2%	1,0%	,0%		100,0%
BRAZIL				23,8%		34,5%	34,5%	7,2%				100,0%
CANADA				,0%	,2%	1,7%	16,4%	80,4%	1,2%	,0%	,0%	100,0%
CZECH REPUBLIC				,1%	,3%	1,8%	38,2%	59,6%				100,0%
DENMARK						6,0%	91,7%	2,2%				100,0%
FINLAND					,2%	10,8%	89,0%					100,0%
FRANCE					,3%	7,5%	37,9%	51,7%	2,6%			100,0%
GERMANY				,0%	1,2%	13,2%	63,3%	22,3%	,1%			100,0%
GREECE					,1%	,8%	2,0%	79,8%	17,3%			100,0%
HUNGARY				,2%	1,0%	2,1%	59,1%	37,6%				100,0%
ICELAND								100,0%				100,0%
IRELAND					,0%	3,3%	61,6%	15,8%	19,2%	,0%		100,0%
ITALY					,5%	,6%	16,0%	77,2%	5,7%			100,0%
JAPAN								100,0%				100,0%
KOREA, REPUBLIC OF							1,2%	98,2%	,6%			100,0%
LATVIA	,0%	,1%	,1%	1,6%	7,0%	35,3%	55,4%	,5%				100,0%
LIECHTENSTEIN				1,0%	16,2%	79,2%	3,6%					100,0%
LUXEMBOURG				1,7%	16,6%	56,2%	25,6%					100,0%
MEXICO				2,8%	10,4%	27,2%	59,4%	,2%				100,0%
NETHERLANDS				,3%	4,9%	45,5%	49,2%	,1%				100,0%
NEW ZEALAND							7,0%	88,1%	4,8%	,1%		100,0%
NORWAY						,0%	,7%	98,9%	,4%			100,0%
POLAND							100,0%					100,0%
PORTUGAL			,2%	,7%	5,1%	11,9%	25,3%	56,5%	,3%			100,0%
RUSSIAN FEDERATION				,0%	,0%	1,8%	25,0%	72,6%	,6%			100,0%
SPAIN				,0%	,0%	2,0%	24,6%	73,4%	,0%			100,0%
SWEDEN					,0%	2,0%	97,8%	,2%				100,0%
SWITZERLAND				,7%	14,0%	77,0%	8,3%	,0%	,0%			100,0%
UNITED KINGDOM						,0%	33,5%	62,7%	3,8%			100,0%
UNITED STATES				,1%	1,8%	40,2%	57,4%	,4%				100,0%
Gesamt		,0%	,0%	,0%	1,1%	4,7%	32,4%	54,2%	7,2%	,3%	,0%	100,0%

Abbildung 2 Verteilung Klassenstufe

Eine der ersten Fragen wird die Vergleichbarkeit der Gruppen betreffen. Schümer kommt zu folgender Bewertung:

*Dass die in der 10. Klasse getesteten 15-Jährigen ein Jahr länger beschult worden sind als die meisten der in Deutschland erfassten Schülerinnen und Schüler ist allerdings keine hinreichende Erklärung für ihren Leistungsvorsprung, denn es gibt auch Länder, in denen die Mehrheit der 15-Jährigen ebenfalls erst auf der 9. Klassenstufe ist und trotzdem überdurchschnittliche Testleistungen erbracht hat.*²¹

²⁰ Schümer, Weiß, Steinert, Baumert, Tillmann, Meier: Lebens- und Lernbedingungen von Jugendlichen. In: Deutsches PISA-Konsortium (Hrsg.): PISA 2000: Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich. S. 413

²¹ Schümer, Weiß, Steiner, Baumert, Tillmann, Meier: Institutionelle Bedingungen schulischen Lernens im internationalen Vergleich. In: Deutsches PISA-Konsortium (Hrsg.): PISA 2000: Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich. S. 415

Dass sich die Schülerleistungen in beiden Ländern [Island, Österreich d. Verf.] nicht in der erwarteten Richtung voneinander unterscheiden, bestätigt erneut, dass die nominelle Unterrichtszeit wenig Erklärungskraft besitzt, solange die Stundentafeln, die Zeitnutzung und andere Kontextfaktoren außer Acht bleiben.²²

Die von Schümer geäußerte These widerspricht den Darstellungen von Mullis, Martin, Beaton, Gonzalez, Kelly und Smith in der Dritten Internationale Mathematik- und Naturwissenschaftsstudie (TIMSS), die in Bezug auf den Mathematiktest von substanziellen Unterschieden ausgehen.

Since for most children the opportunity to learn mathematics is anchored in the school, even one year can make a substantial difference.²³

Jürges, Schneider und Büchel berichten ebenfalls einen statistisch signifikanten Einfluss der Klassenstufe auf die in TIMSS erzielten Leistungen (student achievement in mathematics – percentile effects).²⁴

Wößmann widerspricht ebenfalls der Einschätzung von Schümer:

In der PISA-Studie, deren Zielpopulation in allen Ländern die 15-Jährigen waren, gehörten die deutschen 15-Jährigen vergleichsweise niedrigen Jahrgangsstufen an: Während zwei Drittel der 15-Jährigen aller OECD-Länder der 10. bis 13. Jahrgangsstufe angehörten, waren über drei Viertel der deutschen 15-Jährigen in der 7. bis 9. Jahrgangsstufe. Bei den PISA-Ergebnissen fällt auf, dass deutsche Schüler bei einem Vergleich innerhalb jeder Jahrgangsstufe zwischen drei und fünf Prozent über der OECD-Durchschnittsleistung der jeweiligen Jahrgangsstufe liegen. Das heißt, 15-jährige Neuntklässler aus Deutschland sind besser als der

²² Schümer, Weiß, Steinert, Baumert, Tillmann, Meier: Institutionelle Bedingungen schulischen Lernens im internationalen Vergleich. In: Deutsches PISA-Konsortium (Hrsg.): PISA 2000: Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich. S. 418

²³ Martin, M.O., Beaton, A.E., Gonzalez, E.J., Kelly, D.L., Smith, T.A.: Mathematics Achievement in the Primary School Years (TIMSS), Boston College 1997, S. 30

²⁴ Jürges, H., Schneider, K., Büchel, K.: The Effect of Central Exit Examinations on Student Achievement: Quasi-Experimental Evidence from TIMSS Germany, 2003 S. 19

OECD-Durchschnitt etc. Nur befinden sich die meisten deutschen 15-Jährigen im Vergleich zum OECD-Durchschnitt in niedrigeren Jahrgangsstufen.²⁵

Der Einfluss des Schulalters auf die Testleistungen soll nun an dieser Stelle Gegenstand der Untersuchung sein. Als inferenzstatistische Verfahren können die Varianzanalyse sowie die Regressionsanalyse eingesetzt werden. Die Autoren Bos, Lankes, Schwippert, Valtin, Voss, Badel und Plaßmeier verwenden in der Internationalen Grundschul-Lese-Untersuchung (IGLU) die Korrelationsanalyse im Rahmen einer analogen Fragestellung.²⁶ Sie untersuchen den Zusammenhang von Schuljahr und Testleistung, der sowohl mit $r=0,23$ ²⁷ als auch mit $r=0,34$ berichtet wird.^{28 29} Daher soll diese Verfahrensweise hier Berücksichtigung finden. Der sinnvolle Einsatz der Korrelation setzt die Vergleichbarkeit der Teilnehmerstaaten hinsichtlich wesentlicher struktureller Merkmale voraus. Diese Voraussetzung dürfte mit zunehmender Heterogenität der Teilnehmerstaaten, insbesondere bei einer Erweiterung der Teilnehmerstaaten wie in PISA 2003, kaum erfüllt sein.

²⁵ Wößmann: Familiärer Hintergrund, Schulsystem und Schülerleistungen im internationalen Vergleich, aus Politik und Zeitgeschichte (B 21-22/2003)

²⁶ Bos, Lankes, Schwippert, Valtin, Voss, Badel, Plaßmeier: Lesekompetenzen deutscher Grundschülerinnen und Grundschüler am Ende der vierten Jahrgangsstufe im internationalen Vergleich. In: Erste Ergebnisse aus IGLU 2003. S. 105

²⁷ Der eingesetzte Schätzer wurde im Original nicht weiter präzisiert. Die Kennwerte werden in vorliegender Studie ohne Rücksicht auf weitergehende Präzisierung des Schätzers übernommen.

²⁸ Bos, Lankes, Schwippert, Valtin, Voss, Badel, Plaßmeier: Lesekompetenzen deutscher Grundschülerinnen und Grundschüler am Ende der vierten Jahrgangsstufe im internationalen Vergleich. In: Erste Ergebnisse aus IGLU 2003. S. 105

²⁹ Bos, Valtin, Lankes, Schwippert, Voss, Badel, Plaßmeier: Lesekompetenzen am Ende der vierten Jahrgangsstufe in einigen Ländern der Bundesrepublik Deutschland im nationalen und internationalen Vergleich. In: Bos, Lankes, Prenzel, Schwippert, Valtin, Walther (Hrsg.): IGLU Einige Länder der Bundesrepublik Deutschland im nationalen und internationalen Vergleich, 2004. S. 65

2.2.1 Korrelation

Als lineares Zusammenhangsmaß kann der Korrelationskoeffizient nach Pearson beziehungsweise als Rangstatistik der Korrelationskoeffizient nach Spearman eingesetzt werden. Modell: Es seien X, Y zwei Zufallsvariablen mit den Erwartungswerten μ_x und μ_y , den Varianzen σ_x^2, σ_y^2 sowie der Kovarianz σ_{xy} mit $X \sim N(\mu_x, \sigma_x^2), Y \sim N(\mu_y, \sigma_y^2)$,

Korrelation $\rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$ und folgenden Statistiken

Korrelationskoeffizient nach Pearson

$$r_p = \frac{s_{xy}}{\sqrt{s_x^2 s_y^2}}$$

Korrelationskoeffizient nach Spearman (Rangstatistik)

$$r_s = 1 - \frac{6 \sum_{i=1}^n d_i^2}{(n-1)n(n+1)}$$

Zu testen ist nun die Hypothese $H_0 : \rho_{xy} = 0$ gegen $H_1 : \rho_{xy} \neq 0$ mit

$$\sqrt{n-2} \frac{|r|}{\sqrt{1-r^2}} > t_{n-2; 1-\alpha/2} \text{ als Ablehnungsbereich von } H_0^{30}.$$

Die zwischen Testleistung im Lesen und Schulalter über sämtliche Teilnehmerstaaten fallbasiert, über den ungewichteten Datensatz gemessene Korrelation beträgt $r_p = 0,33$. Dem entspricht ein positiver linearer Zusammenhang zwischen Schulalter und Lesetestleistung im PISA-Test³¹. Die Ausgangshypothese H_0 wird für $p < 0,01$ verworfen.

Correlations

		Plausible value in reading
Grade - Q2	Pearson Correlation	,333**
	Sig. (2-tailed)	,000
	N	172412

** . Correlation is significant at the 0.01 level (2-tailed).

Abbildung 3 Korrelation Klassenstufe Lesewert fallbasiert

Von Interesse ist nun die Differenzierung nach Teilnehmerstaat. Die Ergebnisse variieren erheblich. Das Schulalter wird in einigen Ländern, insbesondere in Korea und Japan, entweder nur sehr geringfügig oder nicht variiert.

³⁰ Bosch: Statistik 1996, S. 409

³¹ Auf die Darstellung von Teststärke (Power), Zusammenhang von Signifikanzniveau, Effektgröße und Stichprobenumfang wird verzichtet.

Mit $r_p = 0,419$ (Deutschland), $r_p = 0,429$ (Luxemburg), $r_p = 0,451$ (Niederlande), $r_p = 0,55$ (Spanien), $r_p = 0,594$ (Belgien), $r_p = 0,643$ (Frankreich) und $r_p = 0,71$ (Portugal) lässt sich ein positiver linearer statistischer Zusammenhang nachweisen. Sämtliche Korrelationen sind statistisch signifikant ($p < 0,01$). Dieser Zusammenhang lässt sich für jeden Teilnehmerstaat, in dem das Schulalter variiert wird, nachweisen ($p < 0,01$).

Correlations

Country three-digit ISO code			Plausible value in reading
AUSTRALIA	Grade - Q2	Pearson Correlation Sig. (2-tailed) N	,246** ,000 5140
AUSTRIA	Grade - Q2	Pearson Correlation Sig. (2-tailed) N	,266** ,000 4565
BELGIUM	Grade - Q2	Pearson Correlation Sig. (2-tailed) N	,594** ,000 6609
BRAZIL	Grade - Q2	Pearson Correlation Sig. (2-tailed) N	,550** ,000 4820
CANADA	Grade - Q2	Pearson Correlation Sig. (2-tailed) N	,312** ,000 29026
CZECH REPUBLIC	Grade - Q2	Pearson Correlation Sig. (2-tailed) N	,203** ,000 5183
DENMARK	Grade - Q2	Pearson Correlation Sig. (2-tailed) N	,185** ,000 4120
FINLAND	Grade - Q2	Pearson Correlation Sig. (2-tailed) N	,213** ,000 4864
FRANCE	Grade - Q2	Pearson Correlation Sig. (2-tailed) N	,643** ,000 4648
GERMANY	Grade - Q2	Pearson Correlation Sig. (2-tailed) N	,419** ,000 4997
GREECE	Grade - Q2	Pearson Correlation Sig. (2-tailed) N	,195** ,000 4586
HUNGARY	Grade - Q2	Pearson Correlation Sig. (2-tailed) N	,286** ,000 4840
ICELAND	Grade - Q2	Pearson Correlation Sig. (2-tailed) N	, ^a , 3317
IRELAND	Grade - Q2	Pearson Correlation Sig. (2-tailed) N	,233** ,000 3829
ITALY	Grade - Q2	Pearson Correlation Sig. (2-tailed) N	,354** ,000 4952
JAPAN	Grade - Q2	Pearson Correlation Sig. (2-tailed) N	, ^a , 5256
KOREA, REPUBLIC OF	Grade - Q2	Pearson Correlation Sig. (2-tailed) N	,029* ,041 4976

** . Correlation is significant at the 0.01 level (2-tailed).

* . Correlation is significant at the 0.05 level (2-tailed).

a. Cannot be computed because at least one of the variables is constant.

Abbildung 4 Korrelation Klassenstufe Lesetestwert je Teilnehmerstaat (I)

Correlations

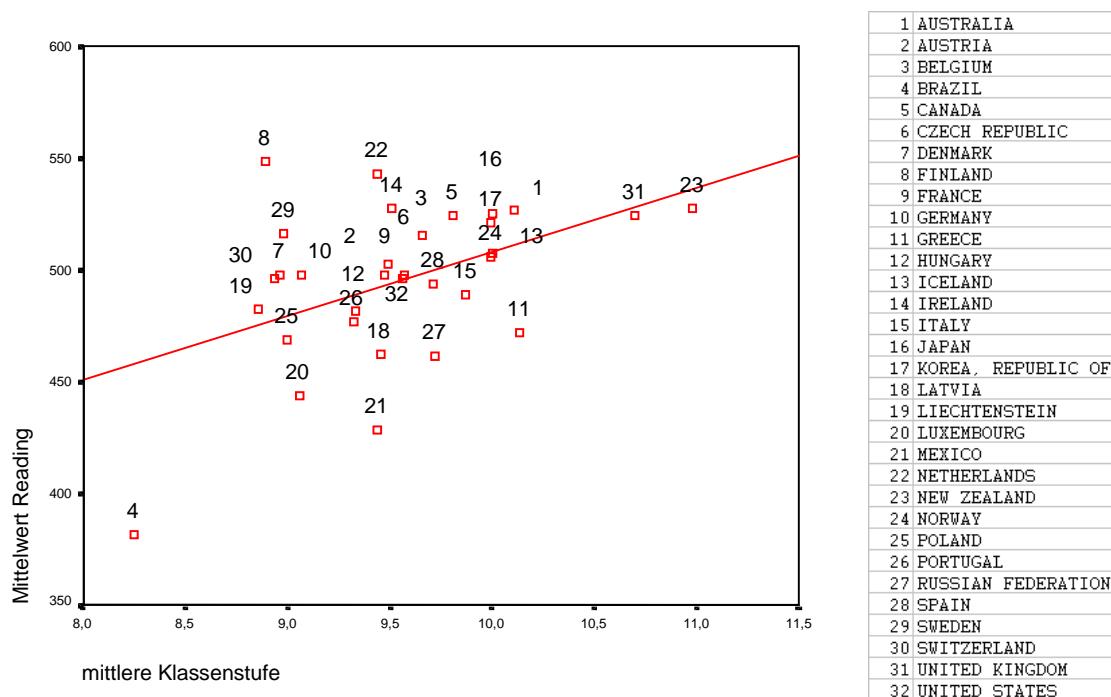
Country three-digit ISO code			Plausible value in reading
LATVIA	Grade - Q2	Pearson Correlation Sig. (2-tailed) N	,318** ,000 3846
LIECHTENSTEIN	Grade - Q2	Pearson Correlation Sig. (2-tailed) N	,403** ,000 308
LUXEMBOURG	Grade - Q2	Pearson Correlation Sig. (2-tailed) N	,429** ,000 3482
MEXICO	Grade - Q2	Pearson Correlation Sig. (2-tailed) N	,554** ,000 4297
NETHERLANDS	Grade - Q2	Pearson Correlation Sig. (2-tailed) N	,451** ,000 2470
NEW ZEALAND	Grade - Q2	Pearson Correlation Sig. (2-tailed) N	,273** ,000 3633
NORWAY	Grade - Q2	Pearson Correlation Sig. (2-tailed) N	,088** ,000 4099
POLAND	Grade - Q2	Pearson Correlation Sig. (2-tailed) N	, ^a , 3654
PORTUGAL	Grade - Q2	Pearson Correlation Sig. (2-tailed) N	,710** ,000 4504
RUSSIAN FEDERATION	Grade - Q2	Pearson Correlation Sig. (2-tailed) N	,175** ,000 6688
SPAIN	Grade - Q2	Pearson Correlation Sig. (2-tailed) N	,550** ,000 6160
SWEDEN	Grade - Q2	Pearson Correlation Sig. (2-tailed) N	,174** ,000 4376
SWITZERLAND	Grade - Q2	Pearson Correlation Sig. (2-tailed) N	,320** ,000 6017
UNITED KINGDOM	Grade - Q2	Pearson Correlation Sig. (2-tailed) N	,098** ,000 9305
UNITED STATES	Grade - Q2	Pearson Correlation Sig. (2-tailed) N	,334** ,000 3845

** . Correlation is significant at the 0.01 level (2-tailed).

a. Cannot be computed because at least one of the variables is constant.

Abbildung 5 Korrelation Klassenstufe Lesetestwert je Teilnehmerstaat (II)

Eine weitere, wenngleich wie in der fallbasierten Korrelationsrechnung unter der Einschränkung vergleichbarer Schulsysteme und vergleichbarer sozioökonomischer Niveaus sinnvolle Möglichkeit zum Nachweis eines linearen Zusammenhangs besteht auf aggregiertem Datenniveau in Form eines Vergleichs mittlerer Testleistung und mittlerer Klassenstufe (grade) eines Landes (jeweils arithmetisches Mittel). Dieser führt mit $r_p = 0,464$ zu einem statistisch signifikanten Ergebnis ($p < 0,01$).³²



Correlations

		Mittelwert Reading
mittlerer Grad	Pearson Correlation	,464**
	Sig. (2-tailed)	,007
	N	32

** . Correlation is significant at the 0.01 level

Abbildung 6 Korrelation Klassenstufe Lesetestwert (aggregiert)

Die Teilnehmerstaaten Vereinigtes Königreich, Neuseeland und Australien weisen die höchste mittlere Klassenstufe und gleichzeitig die höchsten mittleren Lesewerte auf. Die niedrigste mittlere Klassenstufe und gleichzeitig den niedrigsten Lesewert weist Brasilien auf. Es besteht ein signifikant positiver linearer statistischer Zusammenhang.

³² Korrelationsberechnungen über den gesamten Datensatz (aggregiert oder fallweise) sind prinzipiell nur über vergleichbare Teilnahmestaaten sinnvoll, was hier –insbesondere jedoch in PISA 2003- in Frage gestellt werden kann.

2.2.2 Regression

Um den Einfluss weiterer Variablen im Rahmen statistischer Zusammenhangsanalysen zu berücksichtigen, können multivariate Verfahren, wie die der Regressionsanalyse, eingesetzt werden. Das eingesetzte Modell bezieht die Variablen *höchster sozioökonomischer Index* in der Familie (*HISEI*) sowie *Geschlecht* als dummycodierte Zufallsvariable nach folgendem Modell mit ein:

$$T = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon$$

wobei sich die empirisch ermittelten Testwerte T aus der Klassenstufe X_1 , der dummycodierten Variable *Geschlecht* X_2 , dem höchsten sozioökonomischen Index der Familie X_3 dem Achsenabschnitt β_0 und dem Fehlerterm ε mit $\varepsilon \sim N(0, \sigma^2)$ ergibt. In folgender Tabelle gibt $B = (\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3)'$ die Schätzungen für $\beta_0, \beta_1, \beta_2, \beta_3$ an.

Modellzusammenfassung

Modell	R	R-Quadrat	Korrigiertes R-Quadrat	Standardfehler des Schätzers	Änderungsstatistiken				
					Änderung in R-Quadrat	Änderung in F	df1	df2	Änderung in Signifikanz von F
1	,458 ^a	,210	,210	87,68889	,210	14408,532	3	162594	,000

a. Einflussvariablen : (Konstante), Highest In. Socio-Econ. Index, Sex - Q3, Grade - Q2

Koeffizienten^a

Modell		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten	T	Signifikanz
		B	Standardfehler	Beta		
1	(Konstante)	103,947	2,921		35,589	,000
	Grade - Q2	37,164	,294	,282	126,495	,000
	Sex - Q3	-29,118	,435	-,148	-66,866	,000
	Highest In. Socio-Econ. Index	1,744	,013	,291	130,509	,000

a. Abhängige Variable: Plausible value in reading

Abbildung 7 Regression *Klassenstufe, Geschlecht, Sozioökon. Index* auf Lesetestwert

Das relativ geringe korrigierte R^2 von 0,21 muss vor dem Hintergrund der Modellvoraussetzungen bewertet werden. Der Einfluss sämtlicher Variablen erweist sich in diesem Modell als statistisch signifikant ($p < 0,01$). Unter Berücksichtigung der Codierung der Zufallsvariable *Geschlecht* lässt der negative Beta-Koeffizient die geringeren Lesetestwerte der männlichen Schüler erkennen. Das Schulalter (grade) weist auch unter Kontrolle weiterer Variablen einen statistisch signifikanten, positiven Zusammenhang zur Testleistung auf.

2.2.3 Varianzanalytischer Ansatz

Ein varianzanalytischer Ansatz ist insofern sinnvoll, als einerseits der Faktor *Klassenstufe* (grade) in wenigen Stufen vorliegt, andererseits die Lesetestwerte standardisiert wurden und annähernd eine Normalverteilung aufweisen.^{33 34}

Das Modell ist einfaktoriell mit Klassenstufe als Faktor und wie folgt beschrieben: $X_{ij} = \mu + \alpha_i + E_{ij}$, $i=1, \dots, c$, $j=1, \dots, n_i$ worin μ das (unbekannte) Gesamtmittel und α_i der (unbekannte) Effekt des i -ten Treatments bedeutet. Die Zufallsvariablen E_{ij} sind unabhängig und $N(0, \sigma^2)$ -verteilt, d.h., X_{ij} ist $N(\mu_i, \sigma^2)$ -verteilt mit $\mu_i = \mu + \alpha_i$. Die Beobachtungen x_{ij} sind also nach *einem*, so genannten α -Effekt, klassifiziert.

Die Ausgangshypothese lautet $H_0 : \mu_1 = \mu_2 = \dots = \mu_c$ gegen $H_1 : \mu_k \neq \mu_l$ für mindestens ein Paar $(k, l), k \neq l$. Die Teststatistik für den Test auf Gleichheit der c Mittelwerte (d.h. sämtliche $\alpha_i = 0$) lautet:

$$F = \frac{(N - c) \sum_{i=1}^c n_i (\bar{X}_i - \bar{X})^2}{(c - 1) \sum_{i=1}^c \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2}, \quad \sum_{i=1}^c n_i = N,$$

mit \bar{X}_i als Mittelwert der i -ten Stichprobe und \bar{X} als Mittelwert aller N Beobachtungen.

Die Teststatistik F hat unter H_0 eine F-Verteilung mit $(c - 1, N - c)$ Freiheitsgraden.³⁵

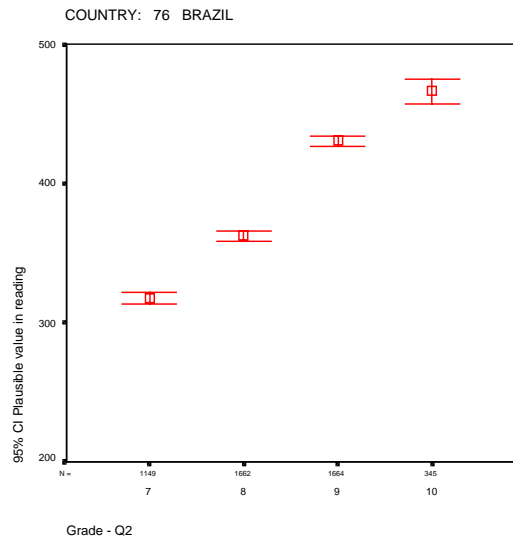
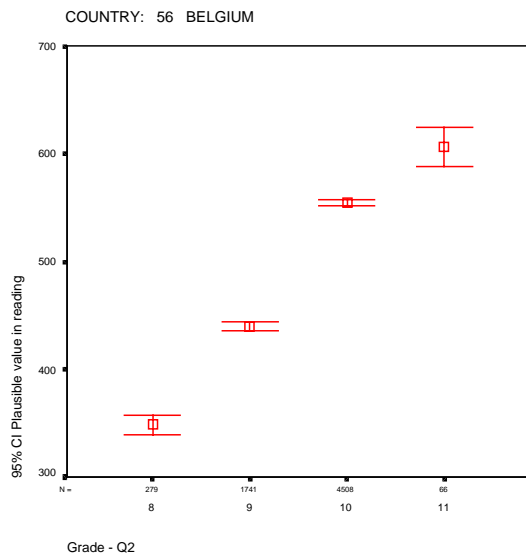
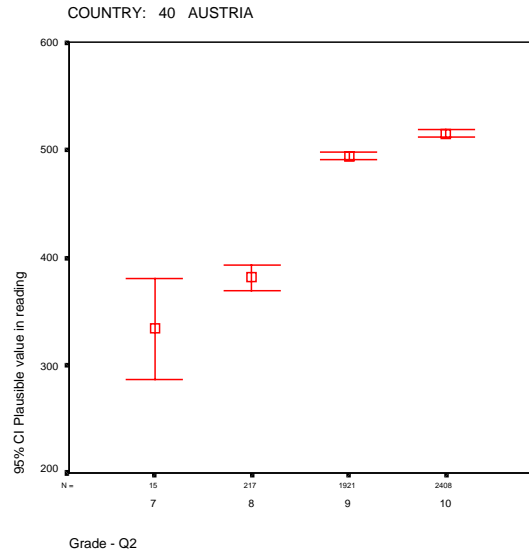
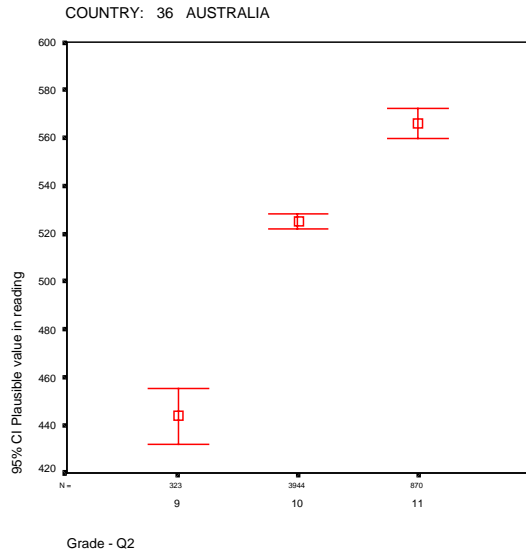
Im Folgenden sollen zunächst Fehlerbalken-Graphiken (arithmetisches Gruppenmittel mit Konfidenzintervall zum Niveau $1 - \alpha$) für jeden einzelnen Teilnehmerstaat dargestellt werden. Nur in einem Fall werden die Signifikanztests der Varianzanalyse sowie der Scheffé-Test als a posteriori Einzelvergleichstest dargestellt. Wie jedoch bereits aus den Konfidenzintervallen (95% KI) deutlich ersichtlich ist, sind die Ergebnisse jeweils statistisch signifikant ($p < 0,01$).³⁶

³³ Adams, Wu (Hrsg.): PISA 2000 Technical Report Kap. 13

³⁴ Zu Effizienzverlusten parametrischer Verfahren siehe Büning: Nichtparametrische Statistische Methoden 1994, insbesondere S. 285 ff

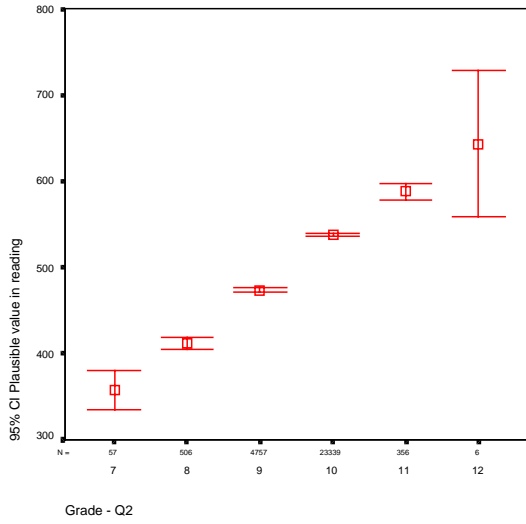
³⁵ Büning, H.: Nichtparametrische statistische Methoden 1994, S.183

³⁶ Hier werden lediglich die Berechnungen des ungewichteten Datensatzes berichtet. Berechnungen über den gewichteten Datensatz führen aufgrund der Erhöhung des Stichprobenumfangs durch die Gewichtungvariable (ohne Einsatz von Korrektur- und Replikationsvariablen) um 9525% zu progressiven Entscheidungen. N (ungew.)=174896, N (gew.) 16659562,5712. Gewichtungvariable w_{fstuw} : „The final weight variable on the

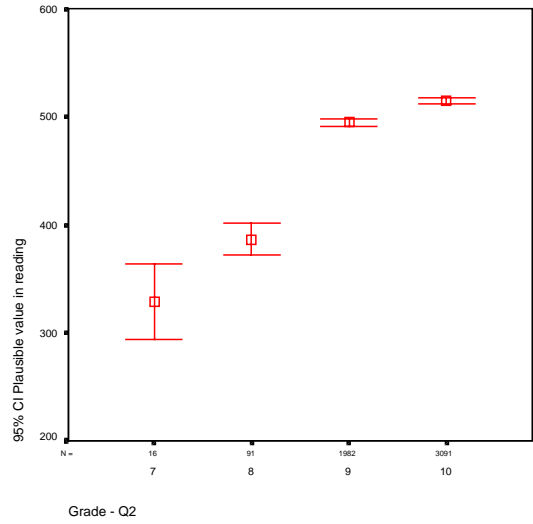


data file was called w_fstuwt, which is the final student weight that incorporates any studentlevel trimming." Vgl. Adams, Wu (Hrsg.): PISA 2000 Technical Report. S. 95

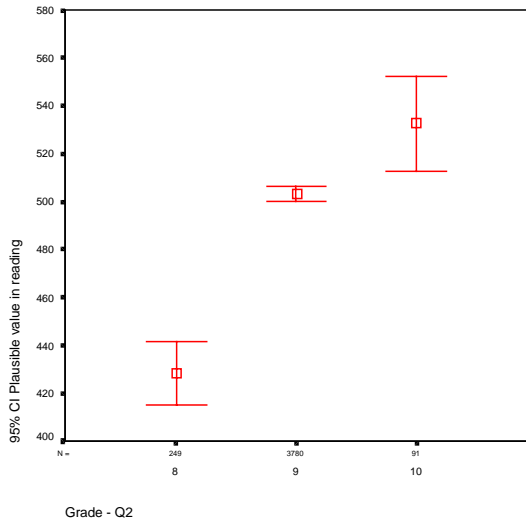
COUNTRY: 124 CANADA



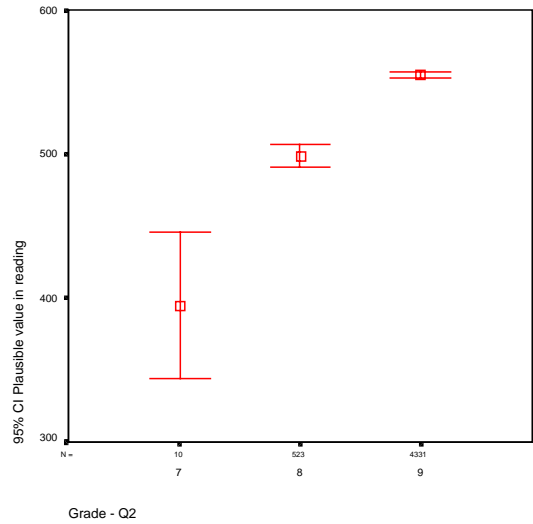
COUNTRY: 203 CZECH REPUBLIC



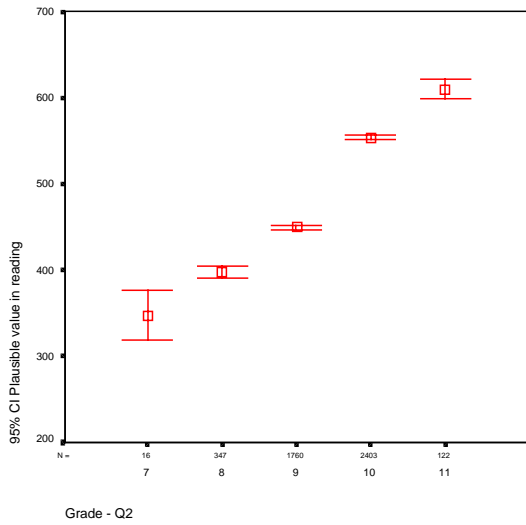
COUNTRY: 208 DENMARK



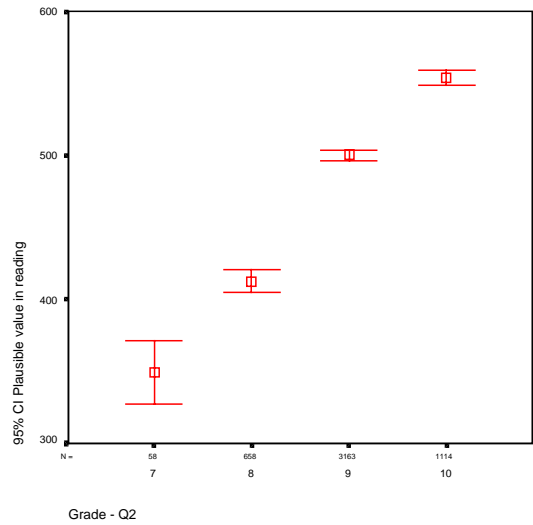
COUNTRY: 246 FINLAND

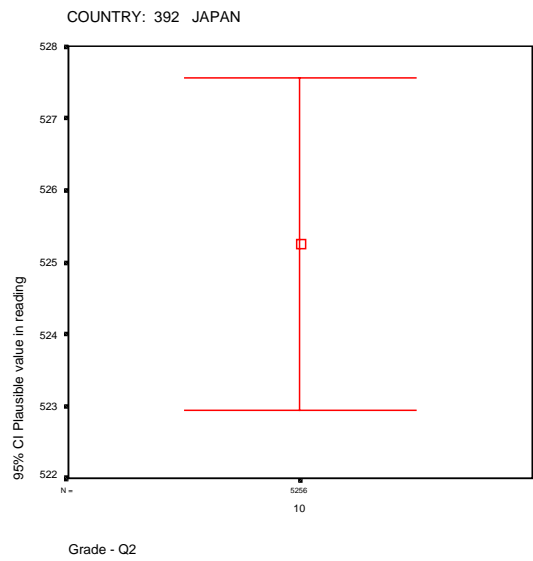
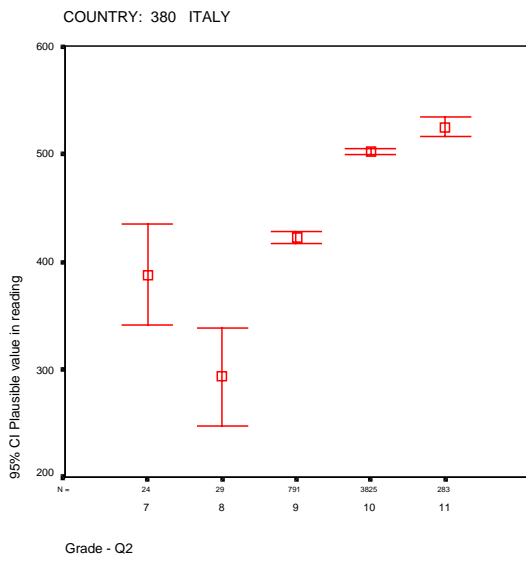
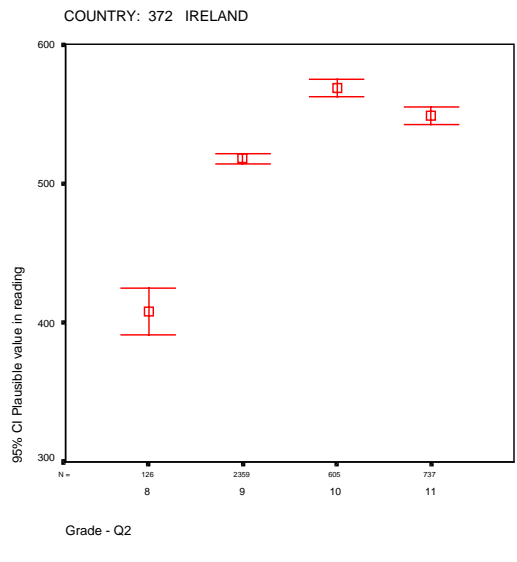
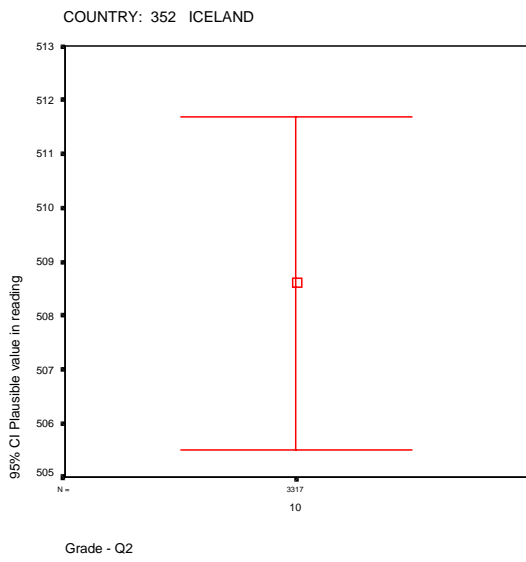
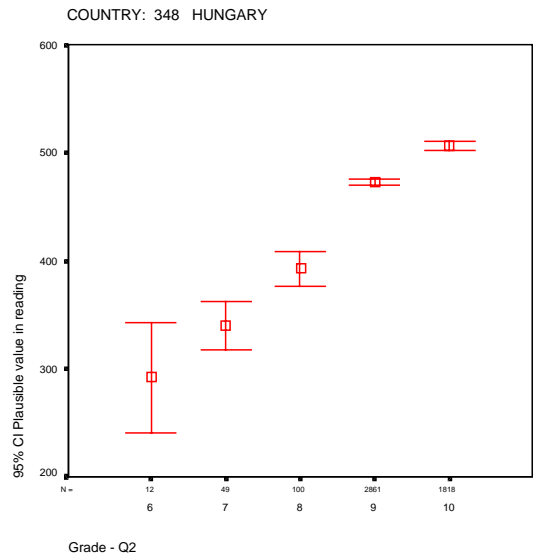
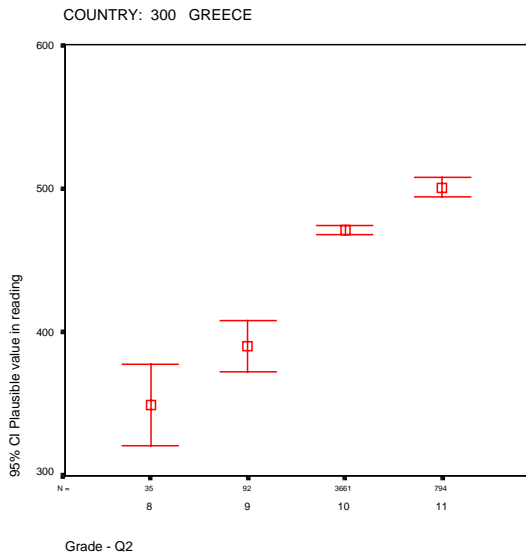


COUNTRY: 250 FRANCE

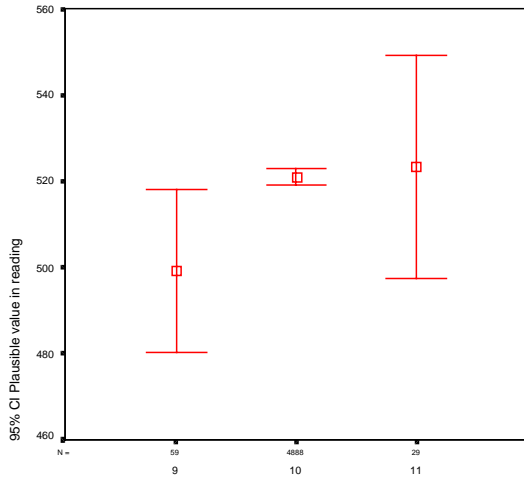


COUNTRY: 276 GERMANY



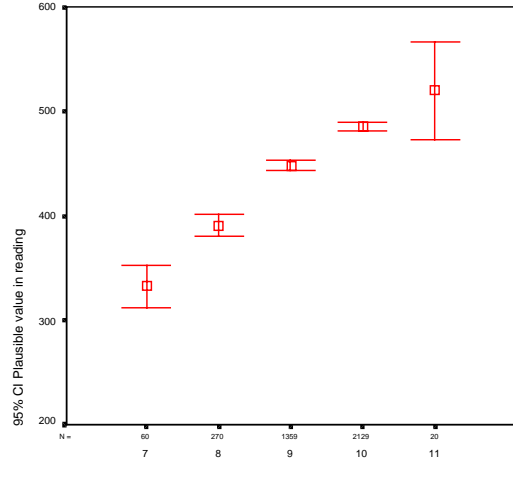


COUNTRY: 410 KOREA, REPUBLIC OF



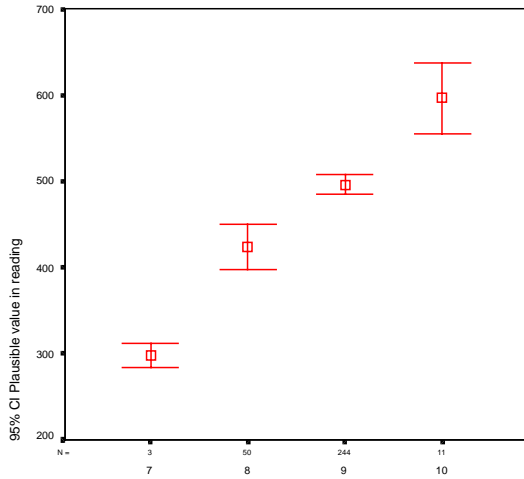
Grade - Q2

COUNTRY: 428 LATVIA



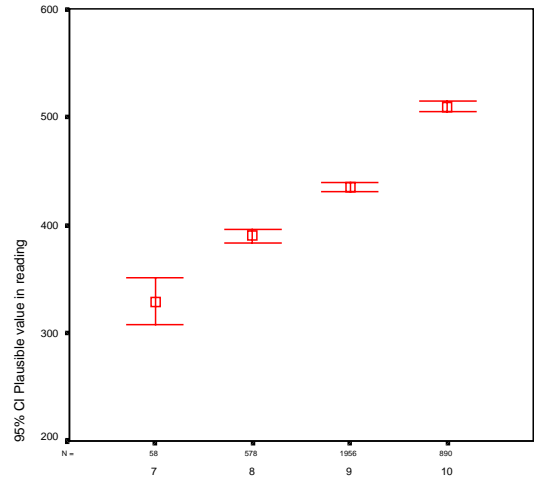
Grade - Q2

COUNTRY: 438 LIECHTENSTEIN



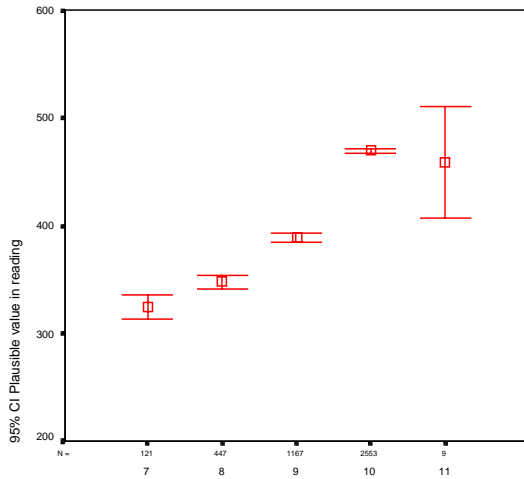
Grade - Q2

COUNTRY: 442 LUXEMBOURG



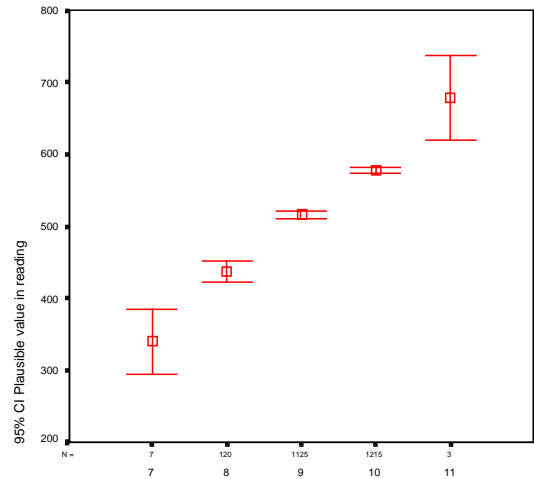
Grade - Q2

COUNTRY: 484 MEXICO

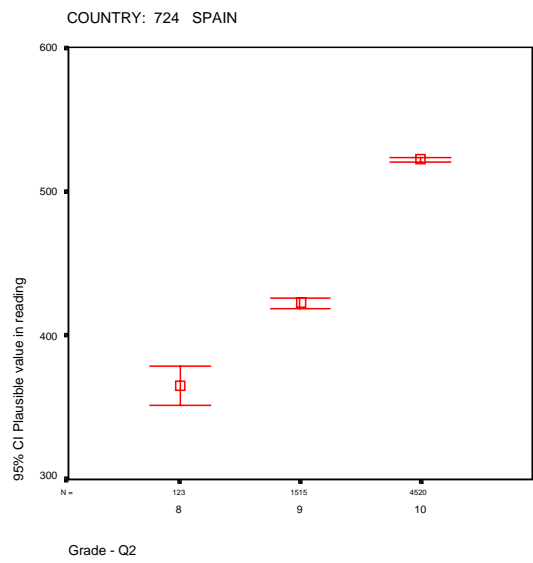
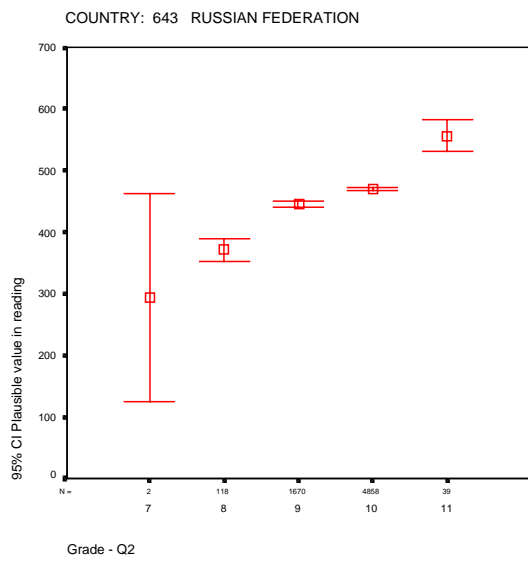
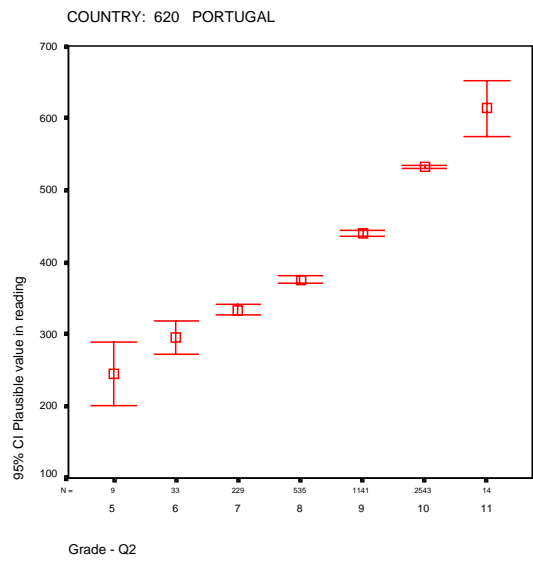
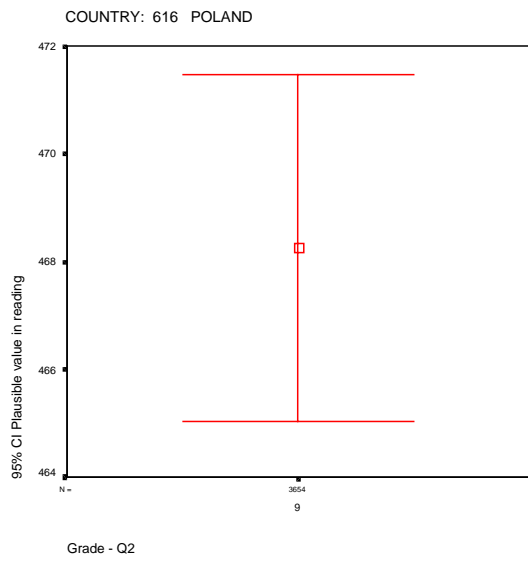
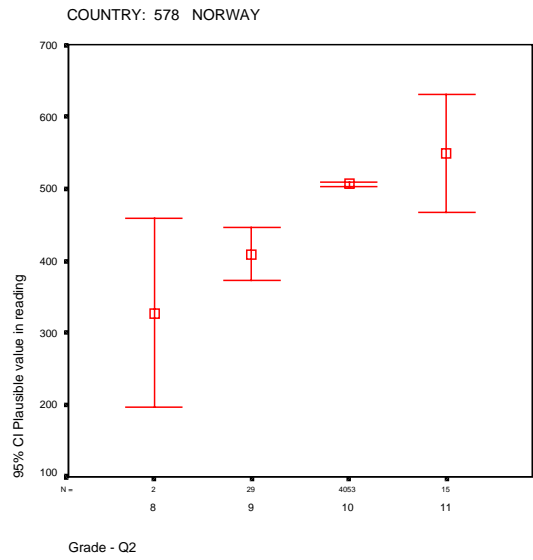
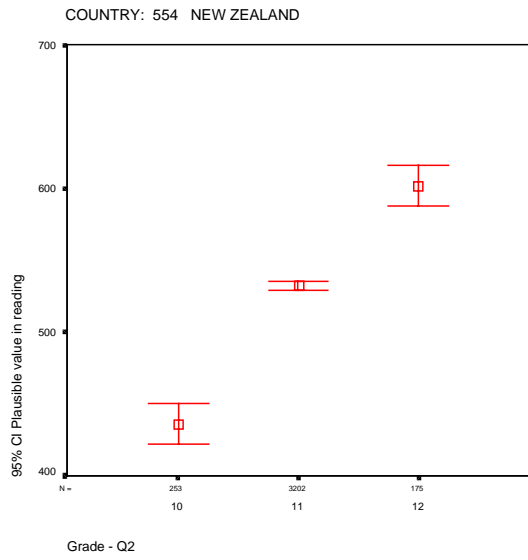


Grade - Q2

COUNTRY: 528 NETHERLANDS



Grade - Q2



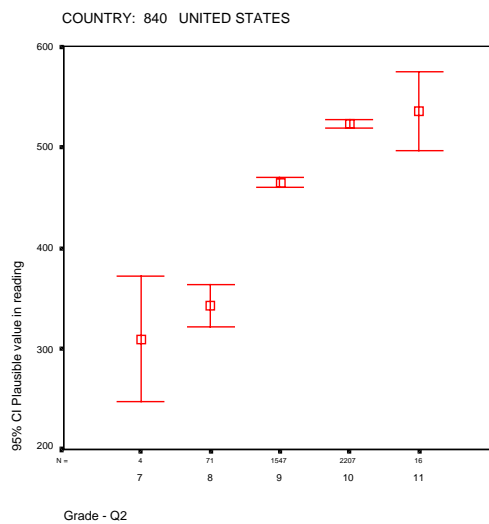
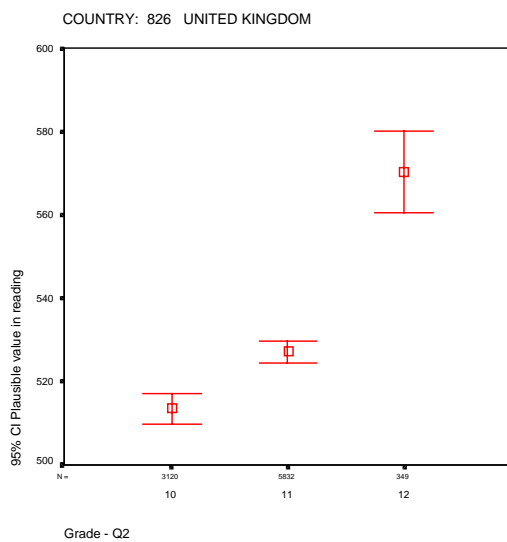
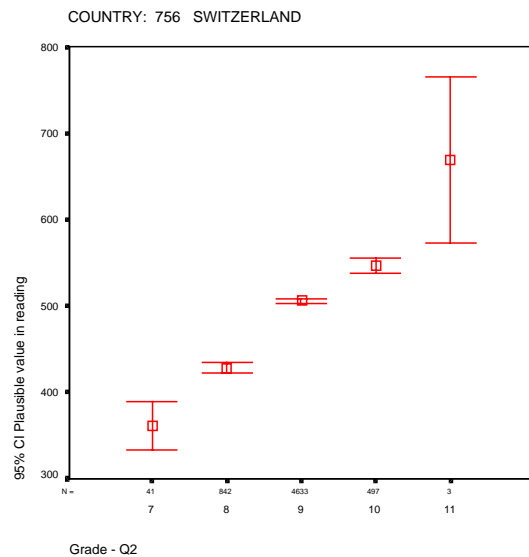
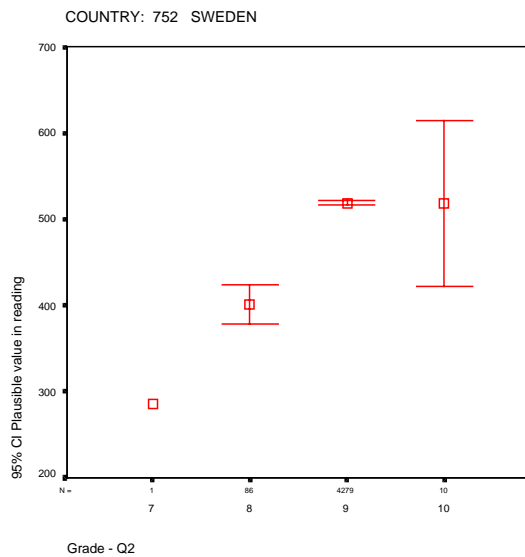


Abbildung 8 Varianzanalysen Klassenstufe und Lesetestwert

In folgender Graphik soll lediglich am Beispiel Deutschlands im internationalen Vergleich die Varianzanalyse und der Scheffé-Test dargestellt werden. Deutlich wird mit $F = 7804$ bei $df = 3$ die Differenz der Mittelwerte ($p < 0,01$). Der Scheffé-Test zeigt zwischen sämtlichen Faktorstufen statistisch signifikante Unterschiede.

Die mittleren Differenzen sind beträchtlich und zeigen mit bis zu 193 Punkten Unterschied eine auch substantiell relevante Effektgröße an. In Bezug auf den internationalen Datensatz lässt sich feststellen, dass das zwischen den Teilnehmerstaaten variierende Schulalter als starker Einfluss auf die Testleistung zu werten ist.

ONEWAY ANOVA

Plausible value in reading

	Quadratsumme	df	Mittel der Quadrate	F	Signifikanz
Zwischen den Gruppen	2,0E+08	3	67295899	7804,614	,000
Innerhalb der Gruppen	1,4E+09	159382	8622,579		
Gesamt	1,6E+09	159385			

Mehrfachvergleiche

Abhängige Variable: Plausible value in reading

Scheffé-Prozedur

(I) Grade - Q2	(J) Grade - Q2	Mittlere Differenz (I-J)	Standardfehler	Signifikanz	95%-Konfidenzintervall	
					Untergrenze	Obergrenze
7	8	-68,6247*	2,34808	,000	-75,1888	-62,0606
	9	-156,1933*	2,14450	,000	-162,1883	-150,1983
	10	-193,7751*	2,12999	,000	-199,7295	-187,8207
8	7	68,6247*	2,34808	,000	62,0606	75,1888
	9	-87,5686*	1,10594	,000	-90,6603	-84,4770
	10	-125,1504*	1,07755	,000	-128,1627	-122,1381
9	7	156,1933*	2,14450	,000	150,1983	162,1883
	8	87,5686*	1,10594	,000	84,4770	90,6603
	10	-37,5818*	,49647	,000	-38,9697	-36,1939
10	7	193,7751*	2,12999	,000	187,8207	199,7295
	8	125,1504*	1,07755	,000	122,1381	128,1627
	9	37,5818*	,49647	,000	36,1939	38,9697

*. Die mittlere Differenz ist auf der Stufe .05 signifikant.

Abbildung 9 Varianzanalyse Klassenstufe Lesetestwert/Scheffé-Test, Deutschland³⁷

Korrespondierende Daten liefert der nationale Ergänzungstest (PISA-E), der aufgrund des höheren Stichprobenumfangs auch für die Klassenstufe 7 einen genügend großen Umfang aufweist.³⁸ Die folgende Abbildung zeigt die mittleren Differenzen der in den einzelnen Klassenstufen erreichten Lesetestwerte (Itemauswahl des internationalen Tests) im bundesdeutschen nationalen Vergleich.

³⁷ Internationaler Datensatz

³⁸ Darstellung über die ungewichteten Daten.

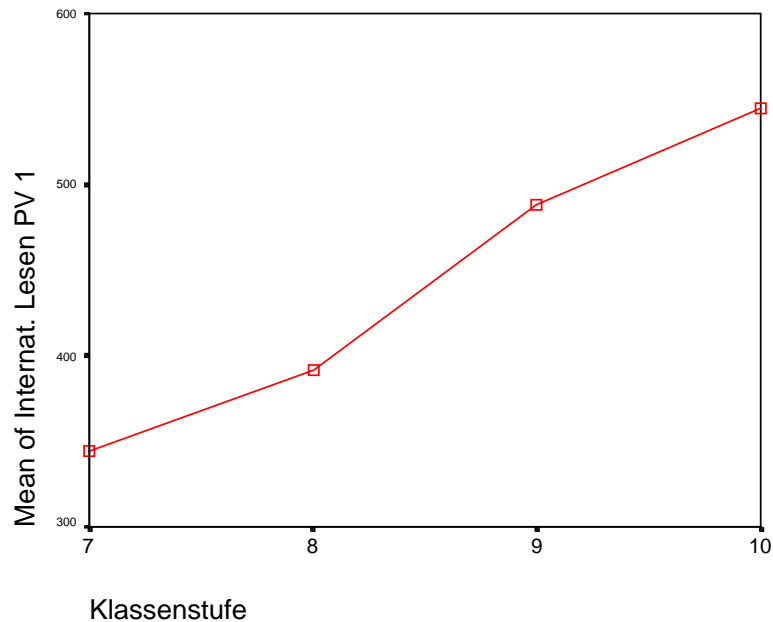
Mehrfachvergleiche

Abhängige Variable: Internat. Lesen PV 1

Scheffé-Prozedur

(I) Klassenstufe	(J) Klassenstufe	Mittlere Differenz (I-J)	Standardfehler	Signifikanz	95%-Konfidenzintervall	
					Untergrenze	Obergrenze
7	8	-46,7446*	4,68597	,000	-59,8448	-33,6444
	9	-143,2974*	4,53344	,000	-155,9712	-130,6236
	10	-199,6304*	4,62599	,000	-212,5630	-186,6979
8	7	46,7446*	4,68597	,000	33,6444	59,8448
	9	-96,5528*	1,48795	,000	-100,7125	-92,3931
	10	-152,8859*	1,74975	,000	-157,7775	-147,9942
9	7	143,2974*	4,53344	,000	130,6236	155,9712
	8	96,5528*	1,48795	,000	92,3931	100,7125
	10	-56,3331*	1,28663	,000	-59,9300	-52,7361
10	7	199,6304*	4,62599	,000	186,6979	212,5630
	8	152,8859*	1,74975	,000	147,9942	157,7775
	9	56,3331*	1,28663	,000	52,7361	59,9300

*. Die mittlere Differenz ist auf der Stufe .05 signifikant.



Abbildungen 10 Scheffé-Test/Diagramm der Mittelwerte Klassenstufe/Testwert Lesen

Das Diagramm der Mittelwerte zeigt steigende Testwerte bei höherer Klassenstufe für die deutschen Schülerinnen und Schüler im Datensatz des nationalen Ergänzungstests. Die Mittelwertdifferenzen sind jeweils statistisch signifikant ($p < 0,01$).

2.3 Bewertung

Sowohl Korrelationsverfahren (fallbasiert/aggregiert), Regressionsanalyse als auch die Varianzanalyse zeigen einen statistisch signifikanten Zusammenhang zwischen dem variierenden Schulalter der Teilnehmerstaaten und deren Lesetestwerte. Somit kann die Hypothese, wonach das zwischen den Teilnehmerstaaten variierende Schulalter keine Erklärung für den Leistungsvorsprung darstellt, verworfen werden.

Der hier nachgewiesene Einfluss von Schulalter (nomineller Lernzeit) auf die Testleistung wirft generell die Frage nach der Vergleichbarkeit verschiedener Länder, ihrer Schulsysteme und Schülerpopulationen auf.

2.4 Alternativer Ansatz

2.4.1 Vergleichbarkeit der Teilnehmerstaaten

Mit der Frage nach der Vergleichbarkeit der Schülerinnen und Schüler der Teilnehmerstaaten ist auch die Problematik der Unterschiedlichkeit der Schulsysteme verbunden. Ein adjustierter, nachträglicher Vergleich beinhaltet sowohl eine noch präzise zu konkretisierende Definition des Untersuchungsgegenstandes als auch den Einsatz einer geeigneten Methode.

Insofern die technische Frage zu klären wäre, läge die Anforderung darin, einerseits eine Vielzahl von Variablen unterschiedlichen Skalenniveaus kontrollieren und andererseits die nach der Kontrolle berechneten geschätzten Leistungstestwerte darstellen zu können. Ein solches Verfahren stellt gegebenenfalls die Kovarianzanalyse dar. Der Einsatz der Kovarianzanalyse wirft natürlich eine Vielzahl neuer Fragen auf. Vergleiche auf der Basis einzelner Klassenstufen werden ebenfalls durchgeführt. Sie berücksichtigen jedoch nicht systematische Unterschiede hinsichtlich struktureller Zusammensetzung der einzelnen Klassenstufen.

Der Einsatz kovarianzanalytischer Verfahren folgt dem Vorgehen der nationalen Ergänzungsstudie (PISA-E).³⁹

³⁹ Baumert, Artelt, Carstensen, Sibberns, Stanat: Untersuchungsgegenstand, Fragestellungen und technische Grundlagen. In: Deutsches PISA-Konsortium (Hrsg.): PISA 2000 – Die Länder der Bundesrepublik Deutschland im Vergleich. S. 35

Die gewichtete koreanische Stichprobe weist eine Differenz von 11,76 (ungewichtet: 10,74) Prozentpunkten zugunsten der männlichen Schüler auf. Dass die Variable Geschlecht überhaupt einen wesentlichen Einfluss auf die Lesetestleistung hat, zeigte sich bereits in der Regressionsanalyse.

Folgende Abbildung zeigt die Verteilung des Anteils der Schülerinnen und Schüler (Variable Geschlecht) über die Teilnehmerstaaten (Datensatz: gewichtet/ungewichtet). Es wird deutlich, dass aufgrund der disproportionalen Verteilung, die durch die Gewichtung nicht ausgeglichen wurde, die Variable Geschlecht im Vergleich der internationalen Testwerte mit in Betracht gezogen werden muss.

SEX*) COUNTRY	Female		Male	
	Female	Male	Female	Male
AUSTRALIA	48,15	51,85	47,50	52,50
AUSTRIA	49,95	50,05	52,09	47,91
BELGIUM	49,40	50,60	47,86	52,14
BRAZIL	52,38	47,62	53,96	46,04
CANADA	49,87	50,13	50,08	49,92
CZECH REPUBLIC	53,59	46,41	51,67	48,33
DENMARK	49,83	50,17	49,69	50,31
FINLAND	51,54	48,46	51,42	48,58
FRANCE	50,74	49,26	51,34	48,66
GERMANY	51,36	48,64	49,65	50,35
GREECE	49,32	50,68	49,80	50,20
HUNGARY	49,14	50,86	49,59	50,41
ICELAND	50,36	49,64	50,43	49,57
IRELAND	51,92	48,08	50,44	49,56
ITALY	51,31	48,69	49,28	50,72
JAPAN	49,54	50,46	50,52	49,48
KOREA, REPUBLIC OF	44,63	55,37	44,12	55,88
LATVIA	52,24	47,76	51,31	48,69
LIECHTENSTEIN	49,68	50,32	49,84	50,16
LUXEMBOURG	50,11	49,89	50,06	49,94
MEXICO	49,89	50,11	50,01	49,99
NETHERLANDS	50,26	49,74	49,29	50,71
NEW ZEALAND	48,96	51,04	49,74	50,26
NORWAY	49,34	50,66	48,99	51,01
POLAND	47,54	52,46	49,13	50,87
PORTUGAL	52,62	47,38	51,96	48,04
RUSSIAN FEDERATION	50,28	49,72	50,08	49,92
SPAIN	50,93	49,07	50,76	49,24
SWEDEN	49,12	50,88	49,24	50,76
SWITZERLAND	49,82	50,18	49,81	50,19
UNITED KINGDOM	49,73	50,27	50,41	49,59
UNITED STATES	52,56	47,44	51,61	48,39

*) percent

PISA 2000 database : http://pisaweb.acer.edu.au/oeed/oeed_pisa_data_s1.html
data: intstud_read.txt

Abbildung 11 Verteilung *Geschlecht*/Teilnehmerstaat (Angaben in Prozent)

2.4.2 Rückhaltequote (Retentivität)

Prinzipiell müssten weitere Variablen berücksichtigt werden. Nicht statistisch kontrolliert werden kann an dieser Stelle, mangels Kenntnis der Grundgesamtheit, die Auswirkung einer variierenden Ausschöpfungsquote der Stichprobe sowie unterschiedliche Rückhaltequoten der Schulsysteme (in der PISA-Studie als *Retentivität* bezeichnet und in der TIMSS anhand eines so genannten TIMSS Coverage Index –TCI- gemessen). Der TCI wird dem Technical Report⁴⁰ entsprechend wie folgt definiert

*The TCI was computed by forming a ratio of the size of the student population covered by the TIMSS sample, as estimated from the sample itself, to the size of the school-leaving age cohort, which was derived from official population census figures supplied by each country. The numerator in this expression is the total enrollment in the grades tested by TIMSS, estimated from the weighted sample data. This estimate corresponds to the size of the population to which the TIMSS results generalize, and makes appropriate provision for student non-response. It does not include students who are no longer attending school, or students who were excluded from the sample on grounds of physical or other disability. It also does not include students who were repeating the final grade. The denominator in the expression is an estimate of the school-leaving age cohort size. Since the age at which upper secondary students may leave school varies, TIMSS estimated the size of the school-leaving age cohort by taking the average of the size of the 1995 age cohorts for 15-, 16-, 17-, 18-, and 19-year-olds in each country. (Although the general procedure was to base the estimate on the 15-19 age group, there were exceptions. For example, in Germany, the estimate was based on the 17-19 age group.)*⁴¹

Die Rückhaltequoten der einzelnen Teilnehmerstaaten unterscheiden sich deutlich. So werden nur noch 85% des Altersjahrgangs in Korea beschult, 95% der österreichischen, 79% der liechtensteiner und sogar nur noch rund 50% der mexikanischen und brasilianischen 15-Jährigen.⁴² In diesen Fällen ist mit einer Überschätzung der Testwerte zu rechnen. Da mit sinkender Rückhaltequote (Retentivität) eines Schulsystems die Selektivität von Population und Stichprobe unter Umständen steigt, ist eine negative

⁴⁰ TIMSS Technical Report / Volume III / Implementation and Analysis S.18

⁴¹ TIMSS Technical Report / Volume III / Implementation and Analysis S.18

⁴² Baumert, Stanat, Demmrich: Untersuchungsgegenstand, theoretische Grundlagen und Durchführung der Studie. In: Deutsches PISA-Konsortium (Hrsg.): PISA 2000: Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich. S. 36

Korrelation zwischen Rückhaltequote (Retentivität) und Testergebnis nicht auszuschließen.

Baumert, Bos und Watermann stellen im Gegensatz zu dieser Überlegung die Hypothese auf, dass mit zunehmender Rückhaltequote (Retentivität) eines Schulsystems die Testleistungen eher steigen.

Umso überraschender ist der empirische Befund der TIMSS/III-Grundbildungsuntersuchung, dass sich ein straffer positiver Zusammenhang zwischen Retentivitätsrate beziehungsweise TCI einerseits und Grundbildungsniveau andererseits nachweisen lässt. Die Korrelation zwischen TCI und mittlerem nationalen Grundbildungsniveau liegt bei $r = .56$ (...) (Schließt man Südafrika als potentiellen Ausreißer aus den Berechnungen aus, ändern sich die Korrelationen praktisch nicht.) Je höher die Haltekraft eines Schulsystems in der Sekundarstufe II ist, desto besser fallen die mathematisch-naturwissenschaftlichen Grundbildungsleistungen aus - trotz geringerer Selektivität der Stichprobe.⁴³

(...)

Fasst man die Befunde der Analysen zusammen, so ergibt sich folgendes Bild: Wider Erwarten lässt sich ein straffer Zusammenhang zwischen der Haltekraft eines Bildungssystems und dem erreichten Grundbildungsniveau nachweisen. Dieser Befund ist erwartungswidrig, da mit abnehmender Retentivität des Systems die Untersuchungspopulationen zunehmend positiv selektiert sind.⁴⁴

Da die Rückhaltequote (Retentivität) über die einzelnen Teilnehmerstaaten stark variiert, ist es von Interesse, deren Einfluss auf die Testleistung zu klären. Als Vorlage liegen die Daten der Dritten Internationalen Mathematik und Naturwissenschaftsstudie, TIMSS/III, im Original der deutschen Version wie folgt vor:

⁴³ Klieme, Baumert, Bos: Mathematisch-naturwissenschaftliche Grundbildung im internationalen Vergleich. In: Baumert, Bos, Lehmann (Hrsg.): TIMSS/III Dritte Internationale Mathematik- und Naturwissenschaftsstudie 2000, Band 1. S. 138 ff

⁴⁴ Klieme, Baumert, Bos: Mathematisch-naturwissenschaftliche Grundbildung im internationalen Vergleich. In: Baumert, Bos, Lehmann (Hrsg.): TIMSS/III Dritte Internationale Mathematik- und Naturwissenschaftsstudie 2000, Band 1. S. 142

Land	Bruttosozial- produkt pro Kopf (in 1.000 US-Dollar)	Anteil der öffentlichen Ausgaben für das Schulsystem in Prozent des Bruttosozial- produkts	Retentivität im Abschlussjahr der Sekundarstufe II (in % der Alterskohorte)	TIMSS Coverage Index (TCI)	Relativer Besuch voruniversitärer Bildungswege (in % der Alterskohorte)	Mittleres Alter im Abschluss- jahr	Mathematisch- naturwissen- schaftliches Grundbildungs- niveau (Mittelwert)
Australien	18,0	3,7	72	68	37	17,7	525
Dänemark	28,1	4,8	59	58	33	19,1	528
Deutschland	25,6	2,4	84	78	25	19,5	496
Frankreich	23,5	3,6	85	84	75	18,8	505
Island	24,6	4,8	55	55	45	21,2	541
Italien	19,3	2,9	74	74	55	18,7	475
Kanada	19,6	4,6	78	70	54	18,6	526
Neuseeland	13,2	3,2	71	70	70	17,6	525
Niederlande	22,0	3,3	92	78	34	18,5	559
Norwegen	26,5	5,3	85	84	48	19,5	536
Österreich	25,0	4,2	93	76	33	19,1	519
Schweden	23,6	4,9	71	71	47	18,9	555
Schweiz	37,2	3,7	84	82	19	19,8	531
Slowenien	7,1	4,2	93	88	80	18,8	514
Südafrika	3,0	5,1	49	49	49	20,1	352
Tschechien	3,2	3,8	83	78	36	17,8	476
Ungarn	3,8	4,3	65	65	18	17,5	477
USA	25,9	4,0	65	63	35	18,1	471
Zypern	10,4	3,6	61	48	44	17,7	447

Abbildung 12 TCI-Index Lesetestwert Quelle: Abb. IV.1 TIMSS/III, S. 136⁴⁵

Die Angaben der deutschen Version stimmen nicht in jedem Fall mit denen des Technical Report der TIMSS (Vol.III)⁴⁶ überein.

Italien weist in der tabellarischen Darstellung der deutschen Version einen TCI-Wert von 74%, anstatt 52% entsprechend den Angaben des TIMSS Technical Report (Vol.III) auf. Deutschland weist einen TCI-Wert von 75% im Technical Report auf, anstatt 78% in der deutschen Ausgabe. Russland und Litauen fehlen wiederum in der tabellarischen Darstellung der deutschen Studie. Die Korrelationen berechnen sich wie folgt:

⁴⁵ Original siehe Anhang

⁴⁶ TIMSS Technical Report / Volume III / Implementation and Analysis S.19

Country	Fall- nummer	TIMSS_III	TCI	BSP*	Anteil BSP*
Australia	1	525	68	18	3,7
Austria	2	519	76	25	4,2
Canada	3	526	70	19,6	4,6
Cyprus	4	447	48	10,4	3,6
Czech Republic	5	476	78	3,2	3,8
Denmark	6	528	58	28,1	4,8
France	7	505	84	23,5	3,6
Germany	8	496	75**	25,6	2,4
Hungary	9	477	65	3,8	4,3
Iceland	10	541	55	24,6	4,8
Italy	11	475	52**	19,3	2,9
Lithuania	12	465	43**		
Netherlands	13	559	78	22	3,3
New Zealand	14	525	70	13,2	3,2
Norway	15	536	84	26,5	5,3
Russian Federation	16	476	48**		
Slovenia	17	514	88	7,1	4,2
South Africa	18	352	49	3	5,1
Sweden	19	555	71	23,6	4,9
Switzerland	20	531	82	37,2	3,7
United States	21	471	63	25,9	4

*) Angaben nach TIMSS/III deutsche Ausgabe

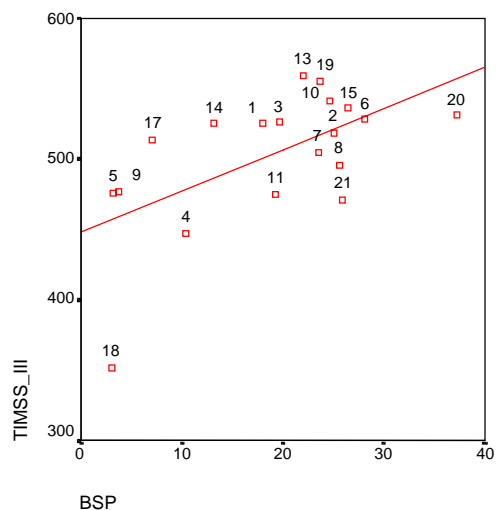
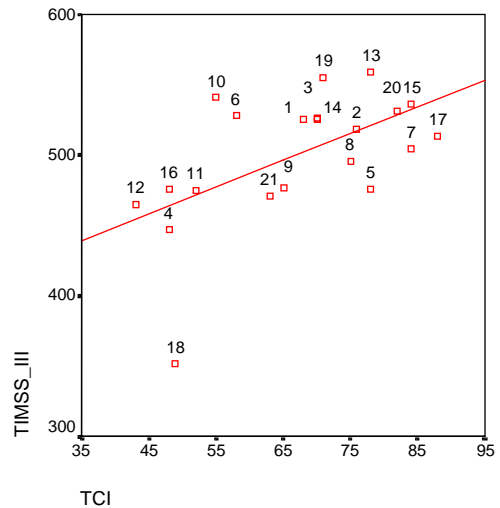
**) korrigierte Angaben nach TIMSS Technical Report Vol. III

TIMSS_III : Testergebnis der TIMSS

TCI : TIMSS Coverage Index

BSP : Bruttosozialprodukt pro Kopf in Tsd. US-Dollar

Anteil BSP: Anteil der öffentlichen Ausgaben für das Schulsystem in % des BSP



Korrelationen

	TCI
TIMSS_III Korrelation nach Pearson	,563**
Signifikanz (2-seitig)	,008
N	21

** . Die Korrelation ist auf dem Niveau von 0,01 (2-seitig) signifikant.

Korrelationen

	TCI
Spearman-Rho TIMSS_III Korrelationskoeffizient	,531*
Sign. (2-seitig)	,013
N	21

*. Die Korrelation ist auf dem 0,05 Niveau signifikant (zweiseitig).

Abbildung 13 TCI – Lesetestwert/Korrelation

Die Korrelation zwischen mathematisch-naturwissenschaftlichem Testwert (TIMSS_III) und TIMSS Coverage Index (TCI) beträgt $r_p = 0,563$ und $r_s = 0,531$. Beide Werte sind statistisch signifikant ($p < 0,01$ bzw. $p < 0,05$). Demnach besteht ein positiver linearer Zusammenhang zwischen beiden Variablen. Es wird jedoch deutlich, dass die Testwerte der Teilnehmerstaaten mit erfolgsrelevanten Indikatoren, zum Beispiel dem

Bruttosozialprodukt (BSP) der Teilnehmerstaaten, kovariieren mit $r_p = 0,592$ ($p < 0,01$) und $r_s = 0,510$ ($p < 0,05$).

	TCI	BSP
TIMSS_III Korrelation nach Pearson	,563**	,592**
Signifikanz (2-seitig)	,008	,008
N	21	19

** Die Korrelation ist auf dem Niveau von 0,01 (2-seitig) signifikant.

	TCI	BSP
Spearman-Rho TIMSS_III Korrelationskoeffizient	,531*	,510*
Sig. (2-seitig)	,013	,026
N	21	19

* Die Korrelation ist auf dem 0,05 Niveau signifikant (zweiseitig).

Abbildung 14 Korrelation TIMSS Testwert – TCI - BSP

Da weitere Variablen mit den TIMSS-Testwerten kovariieren und Angaben zum Schulalter in Bezug auf die TIMSS Studie hier nicht vorliegen sowie Fragen bezüglich der Definition des TCI-Wertes aber insbesondere dessen Datenbasis nicht geklärt werden konnten, soll der Einfluss der Rückhaltequote auf das Testergebnis internationaler Vergleichsstudien anhand der PISA-Studie geprüft werden. Es stellt sich die Frage, ob anhand der PISA-Studie der Zusammenhang von Rückhaltequote und Testwert wiederholt nachgewiesen werden kann.

Der Anteil der noch beschulten 15-Jährigen wird im Technical Report der PISA-Studie veröffentlicht.⁴⁷ Diese Kenngröße kommt der Fragestellung der so genannten Rückhaltequote (Retentivität) des Schulsystem möglicherweise näher als der TCI-Wert. Folgende Darstellung faßt zentrale Daten diesbezüglich zusammen. Die Darstellung umfaßt weitere Teilnehmerstaaten, die in der ursprünglichen Studie (in 2000) noch nicht aufgeführt wurden.

⁴⁷ Adams, Wu (Hrsg.): PISA 2000 Technical Report Tabelle 31 (S. 135-136)/Hinweis Raymond Adams

COUNTRY	PVREAD	HISEI	GRADE	SEX_male	ALL_15	Enroll_15	ENROLL_p	GNI_pc	N
ALBANIA	349,02	45,55	9,71	48,98%				1740	4652
ARGENTINA	417,41	43,35	9,59	43,61%				3810	3608
AUSTRALIA	527,79	52,25	10,10	52,50%	266878	248908	93,27%	21950	4939
AUSTRIA	507,53	49,72	9,43	47,91%	95041	90354	95,07%	26810	4635
BELGIUM	507,31	48,95	9,61	52,14%	121121	119055	98,29%	25760	6371
BRAZIL	395,20	43,93	8,50	46,04%	3464330	1841843	53,17%	2720	4469
BULGARIA	430,54	49,99	9,11	51,46%				2130	4195
CANADA	534,49	52,83	9,84	49,92%	403803	396423	98,17%	24470	28751
CHILE	409,64	39,87	9,48	46,96%				4360	4699
CZECH REPUBLIC	491,94	48,31	9,51	48,33%	134627	132508	98,43%	7150	5273
DENMARK	496,56	49,73	8,97	50,31%	53693	52161	97,15%	33570	3953
FINLAND	546,99	50,00	8,89	48,58%	66571	66561	99,98%	27060	4770
FRANCE	504,38	48,27	9,51	48,66%	788387	788387	100,00%	24730	4389
GERMANY	483,99	48,85	9,06	50,35%	927473	924549	99,68%	25270	4934
GREECE	474,46	47,76	10,05	50,20%	128175	124656	97,25%	13230	4468
HONG KONG	525,15	42,25	9,83	50,15%				25860	4224
HUNGARY	479,83	49,53	9,24	50,41%	120759	115325	95,50%	6350	4746
ICELAND	506,58	52,73	10,00	49,56%	4062	4044	99,56%	30910	3298
INDONESIA	370,76	36,38	9,51	48,86%				810	6811
IRELAND	526,68	48,43	9,50	49,56%	65339	64370	98,52%	27010	3737
ISRAEL	450,96	55,30	9,98	41,78%				16240	3531
ITALY	487,36	47,08	9,86	50,72%	584417	574864	98,37%	21570	4864
JAPAN	522,03	50,54	10,00	49,48%	1490000	1485269	99,68%	34180	2019
KOREA, REPUBLIC OF	525,03	42,80	10,00	55,88%	712812	602605	84,54%	12030	4617
LATVIA	459,14	50,15	9,38	48,69%	38000	35981	94,69%	4400	3713
LIECHTENSTEIN	482,58	47,46	8,85	50,27%	415	326	78,55%		297
LUXEMBOURG	440,40	44,79	9,05	49,94%	4556	4556	100,00%	45740	3201
MEXICO	421,74	42,48	9,36	49,99%	2127504	1098605	51,64%	6230	4222
NETHERLANDS	531,64	50,85	9,43	50,71%	178924	178924	100,00%	26230	2438
NEW ZEALAND	528,60	52,20	10,98	50,26%	54220	51464	94,92%	15530	3523
NORWAY	505,71	53,91	10,00	51,01%	52165	51587	98,89%	43400	4037
PERU	327,23	40,42	9,31	50,10%				2140	4152
POLAND	480,01	46,03	9,00	50,87%	665500	643528	96,70%	5280	3396
PORTUGAL	469,89	43,85	9,25	48,04%	132325	127165	96,10%	11800	4426
ROMANIA	427,19	45,53	8,89	47,32%				2260	4514
RUSSIAN FEDERATION	461,91	49,38	9,69	49,92%	2268566	2259985	99,62%	2610	6512
SPAIN	492,84	44,99	9,70	49,24%	462082	451685	97,75%	17040	5923
SWEDEN	515,39	50,57	8,98	50,76%	100940	100940	100,00%	28910	4313
SWITZERLAND	494,80	49,21	8,93	50,19%	81350	79232	97,40%	40680	5880
THAILAND	431,18	33,02	9,38	41,18%				2190	4798
MACEDONIA	372,96	46,79	9,25	51,16%				1980	4266
UNITED KINGDOM	523,54	51,26	10,68	49,59%	731743	705875	96,46%	28320	8843
UNITED STATES	503,71	52,40	9,53	48,39%	3876000	3836000	98,97%	37870	3242

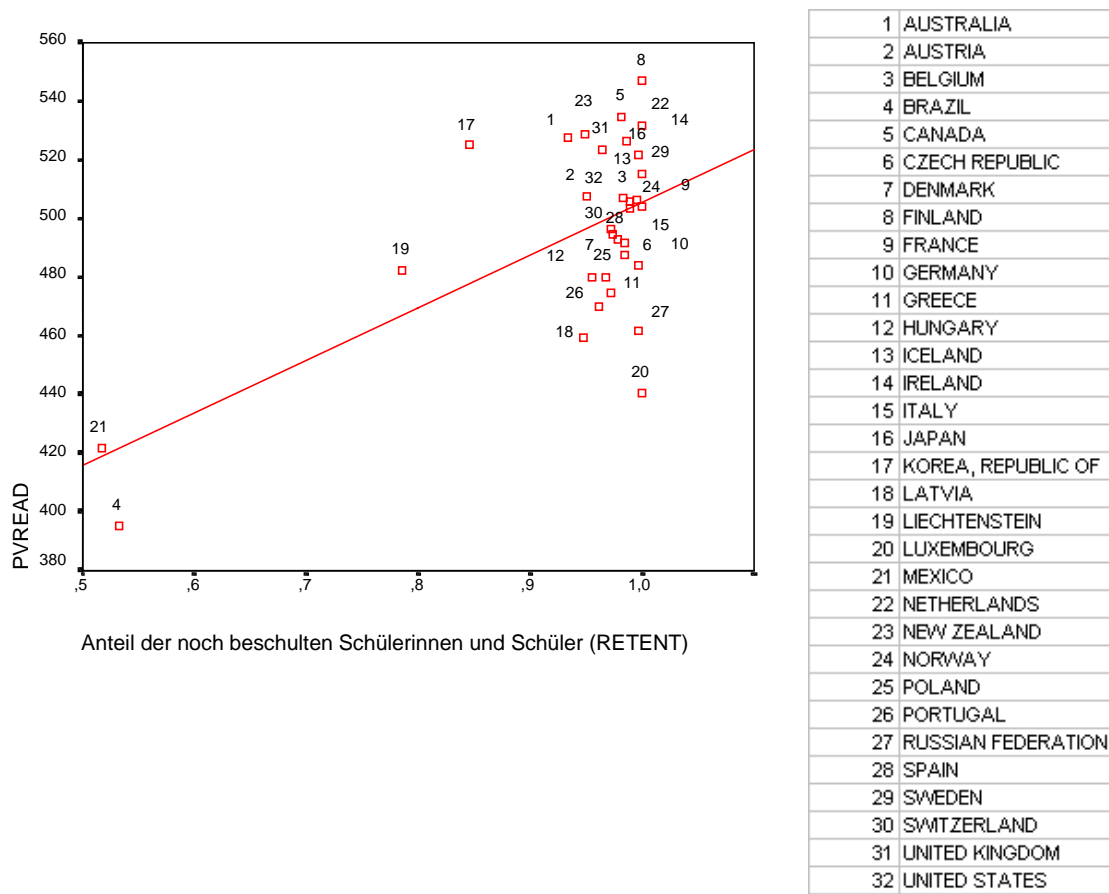
PVREAD	Mean reading score PISA2000 (pv1read)
HISEI	Highest socio-economic index of family (PISA 2000)
GRADE	Mean grade
SEX_male	Percentage of male students
ALL_15	All 15-year-olds
Enroll_15	Enrolled 15-year-olds
ENROLL_p	Percentage of enrolled 15-year-olds
GNI_pc	GNI \$ per capita, 2003 WORLDBANK Atlas method
N	(Sub-) Sample Size

Abbildung 15 Rückhaltequote unter 15-Jährigen in PISA

Insgesamt sind neun Variablen über den erweiterten Datensatz dargestellt. Gemeinsam mit dem aus dem Datensatz berechneten mittleren Lesewert wird der durchschnittliche sozioökonomische Index (HISEI), die mittlere Klassenstufe (GRADE), der Anteil

männlicher Schüler (SEX_MALE), sämtliche 15-Jährige des Teilnehmerstaates, die Anzahl der die Schule besuchenden 15-Jährigen, der daraus berechnete prozentuale Anteil der noch beschulten 15-Jährigen, das von der Weltbank publizierte Pro-Kopf-Einkommen in US\$⁴⁸ sowie die Anzahl der in der Stichprobe berücksichtigten Schülerinnen und Schüler dargestellt.

Sollte ein positiver Zusammenhang zwischen der Rückhaltequote (Retentivität) des Schulsystems und der Testleistung existieren, müsste dieser in PISA erneut nachweisbar sein. Folgende Abbildung trägt den prozentualen Anteil der noch beschulten Altersgruppe gegen die Leseleistung ab.



Correlations

		PVREAD
RETENT	Pearson Correlation	,630**
	Sig. (2-tailed)	,000
	N	32

** . Correlation is significant at the 0.01 level

Correlations

			PVREAD
Spearman's rho	RETENT	Correlation Coefficient	,252
		Sig. (2-tailed)	,164
		N	32

Abbildung 16 Rückhaltequote (Retentivität) und Testleistung in PISA (I)

⁴⁸ (Total GNI 2003, Atlas method) Download www.worldbank.org/data/databytopic/GNIPC.pdf

Zunächst lässt sich mit dem gegenüber Extremwerten empfindlichen $r_p = 0,63$ ($p < 0,01$) ein statistisch signifikanter Zusammenhang nachweisen. Jedoch der in dieser Hinsicht robuste Korrelationskoeffizient nach Spearman weist mit $r_s = 0,252$ ($p = 0,164$) keinen statistisch signifikanten Zusammenhang auf. Werden Mexiko und Brasilien entfernt, zeigt auch der parametrische Korrelationskoeffizient (Pearson) keinen signifikanten linearen Zusammenhang.

		PVREAD
RETENT	Pearson Correlation	,028
	Sig. (2-tailed)	,885
	N	30

		PVREAD	
Spearman's rho	RETENT	Correlation Coefficient	,092
		Sig. (2-tailed)	,627
		N	30

Abbildung 17 Rückhaltequote (Retentivität) und Testleistung in PISA (II)

Mit $r_p = 0,028$ ($p = 0,885$) und $r_s = 0,092$ ($p = 0,627$) muss von einem so genannten Ausreißerproblem ausgegangen werden.

Anhand der PISA-Daten konnte der positive Zusammenhang von Rückhaltequote (Retentivität) des Schulsystems eines Teilnehmerstaates und dessen Testleistungsergebnis nicht nachgewiesen werden.

2.4.3 Ergänzungsunterricht

Weitere Probleme bezüglich Vergleichbarkeit der Schulsysteme einzelner Staaten ergeben sich in Hinblick auf zusätzliche Bildungsangebote. Hier zeigen sich systematische Unterschiede. So besuchen in Japan 57,5% der 15-jährigen Schülerinnen und Schüler neben der regulären Schule regelmäßig Ergänzungsunterricht, in Deutschland lediglich 3,0%⁴⁹.

⁴⁹ Schümer, Weiß, Steinert, Baumert, Tillmann, Meier: Lebens- und Lernbedingungen von Jugendlichen. In: Deutsches PISA-Konsortium (Hrsg.): PISA 2000: Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich. S. 417

Land	Ergänzungs- unterricht	Land	Ergänzungs- unterricht
AUSTRALIA	6,1%	KOREA, REPUBLIC OF	36,0%
AUSTRIA	2,2%	LATVIA	14,8%
BELGIUM	4,4%	LIECHTENSTEIN	4,2%
BRAZIL	3,7%	LUXEMBOURG	6,7%
CANADA	3,3%	MEXICO	11,2%
CZECH REPUBLIC	6,8%	NETHERLANDS	k.A.
DENMARK	3,1%	NEW ZEALAND	3,5%
FINLAND	0,5%	NORWAY	1,1%
FRANCE	8,6%	POLAND	18,1%
GERMANY	3,0%	PORTUGAL	5,7%
GREECE	k.A.	RUSSIAN FEDERATION	15,1%
HUNGARY	25,0%	SPAIN	13,1%
ICELAND	3,9%	SWEDEN	1,3%
IRELAND	8,6%	SWITZERLAND	2,4%
ITALY	0,6%	UNITED KINGDOM	6,3%
JAPAN	57,5%	UNITED STATES	5,1%

Abbildung 18 Ergänzungsunterricht Quelle: PISA 2000

Der Anteil zusätzlichen Unterrichts variiert erheblich. Die Frage der Vergleichbarkeit wird später noch einmal Gegenstand der Diskussion.

2.4.4 Gewichtung

Die Testrohwerte wurden in der PISA-Studie nachträglich durch eine Gewichtungsvariable adjustiert. Voraussetzung für den Einsatz von Gewichtungsvariablen sind sowohl präzise Kenntnisse der für die Untersuchung relevanten Variablen als auch der strukturellen Zusammensetzung der Grundgesamtheit.

Die Gewichte wurden bei WESTAT berechnet. Ziehungswahrscheinlichkeiten sowie systematische Unterschiede zwischen Stichprobe und Grundgesamtheit können durch Gewichtungen ausgeglichen werden.⁵⁰

*Eine Überprüfung der bei WESTAT errechneten Gewichte ergab, dass die Populationsverhältnisse durch die Gewichte sehr gut abgebildet werden.*⁵¹

Diese Aussage entspricht nicht den Ausführungen des Technical Report. In Österreich beispielsweise ergeben sich vergleichbare Probleme wie in den Bundesländern Bayern

⁵⁰ Adams, Wu (Hrsg.): PISA 2000 Technical Report. Kapitel 8

⁵¹ Sibberns, Baumert: Stichprobenziehung und Stichprobengewichtung. In: Deutsches PISA-Konsortium (Hrsg.): PISA 2000: Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich. S. 515

und Baden-Württemberg, die jeweils hohe Anteile an Schülern in Berufsschulen aufweisen. In Österreich fehlen genaue Kenntnisse über das vermutlich testleistungsschwächste Segment.

*In Austria, students in vocational schools are enrolled on a part-time/part-year basis. Since the PISA assessment was conducted only at one point in time, these students are captured only partially. Thus, it is not possible to assess how well the students sampled from vocational schools represent the universe of students enrolled in vocational schools, and so those students not attending classes at the time of the PISA assessment are not represented in the PISA results.*⁵²

In Griechenland sind wesentliche Vergleichsdaten nicht verfügbar.

*The sampling frame for Greece had deficient data on the number of 15-year-olds in each school. This resulted in selecting many schools in the sample with no 15-year-olds enrolled, or that were closed (34 of a sample of 175 schools). Often the replacement school was then included, which was not the correct PISA procedure for such cases. Also quite a few sampled schools (a further 34) had relatively few 15-year-olds. (...) Rather than being assessed or excluded, these schools were also replaced, often by larger schools. This resulted in some degree of over-coverage in the PISA sample, as there were 18 replacement schools in the sample for schools that should not have been replaced, and 32 replacements for small schools where the replacement was generally larger than the school it replaced. It is not possible to evaluate what kinds of students might be over-represented (...).*⁵³

Dieses Defizit stellt ein gravierendes Problem zur Bestimmung der adäquaten Repräsentation von Schüleranteilen dar. Der Vergleich gewichteter und ungewichteter Testergebnisse zeigt den Einfluss der GewichtungsvARIABLE auf das Endresultat. Während Österreichs Ergebnis trotz Unkenntnis des vermutlich testleistungsschwächsten Segments durch die Gewichtung um eine Größenordnung von ca. 10 Punkten aufgewertet wird, reduzieren sich die Testwerte der Bundesrepublik Deutschland um ca. 14 Punkte (maximale Abwertung).

⁵² PISA 2000 International Technical Report. S. 183

⁵³ PISA 2000 International Technical Report. S. 185

PV1READ Differenz (gewichtet - ungewichtet)

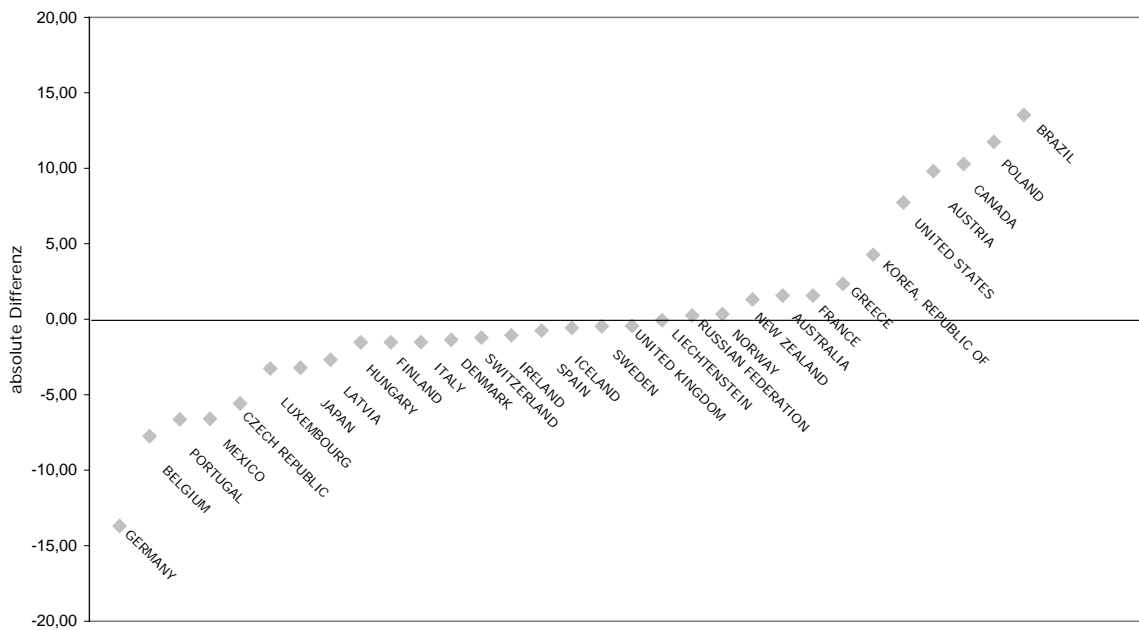


Abbildung 19 Gewichtung/Differenz der Testergebnisse

2.4.5 Darstellung der Testwerte

Aus den bisherigen Ausführungen wird deutlich, dass als Voraussetzung zur Vergleichbarkeit unterschiedlich strukturierter Systeme weitere Variablen, insbesondere Schulalter (*grade*) und *Geschlecht*, kontrolliert werden müssen. Die statistische Kontrolle unterschiedlicher Rückhaltequoten sowie außerschulischen regulären Unterrichts auf das Testergebnis der Schüler ist ein erhebliches methodisches Problem und bleibt einer späteren Betrachtung außerhalb dieser Arbeit vorbehalten.

Im Folgenden werden diejenigen Testwerte dargestellt, die sich unter Kontrolle von lediglich zwei weiteren Variablen, *Klassenstufe* und *Geschlecht*, im Rahmen einer Kovarianzanalyse als so genannte Randmittel schätzen lassen.⁵⁴ Zusätzlich werden die Testwerte nach Klassenstufe berichtet.

⁵⁴ Zur Berechnung geschätzter Randmittel siehe: SPSS Inc. Chicago: SPSS Statistical Algorithms GLM UNIVARIATE AND MULTIVARIATE S.5 ff (<http://support.spss.com/tech/default.asp>)

PISA 2000 Leseverständnis (Klassenstufe 11)

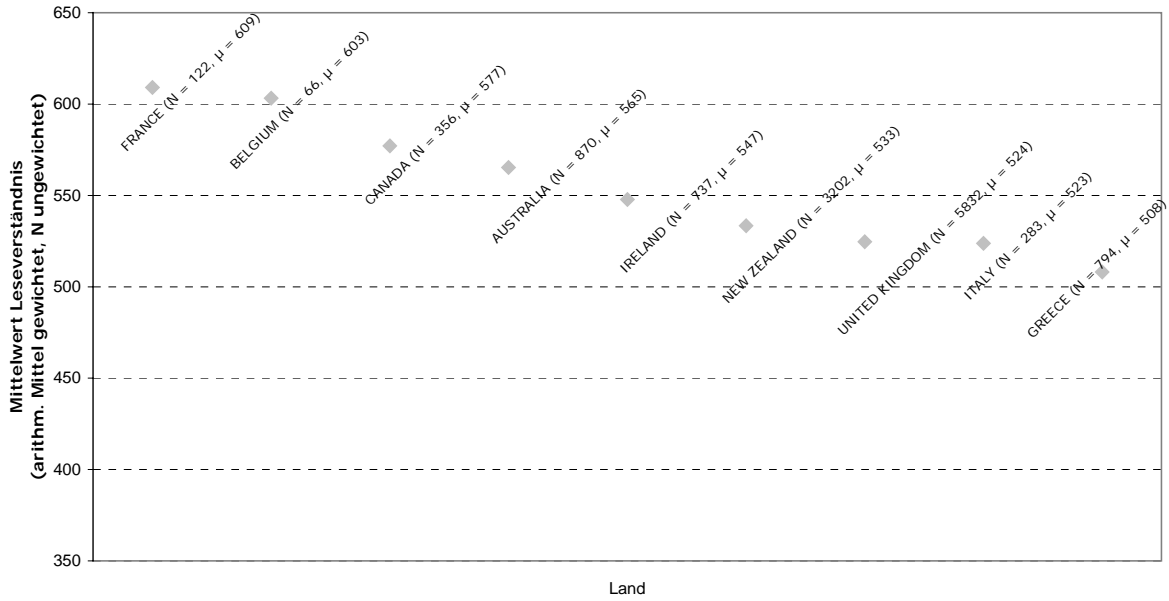


Abbildung 20 Leseverständnis Klassenstufe 11 (gewichtet)

PISA 2000 Leseverständnis (Klassenstufe 10)

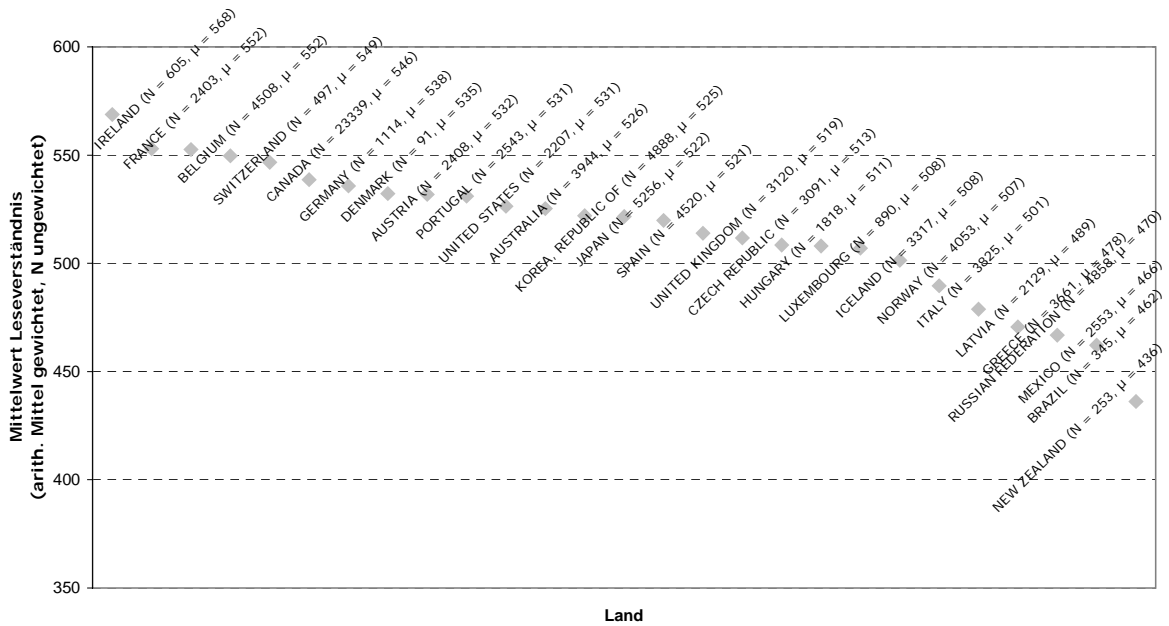


Abbildung 21 Leseverständnis Klassenstufe 10 (gewichtet)

PISA 2000 Leseverständnis (Klassenstufe 9)

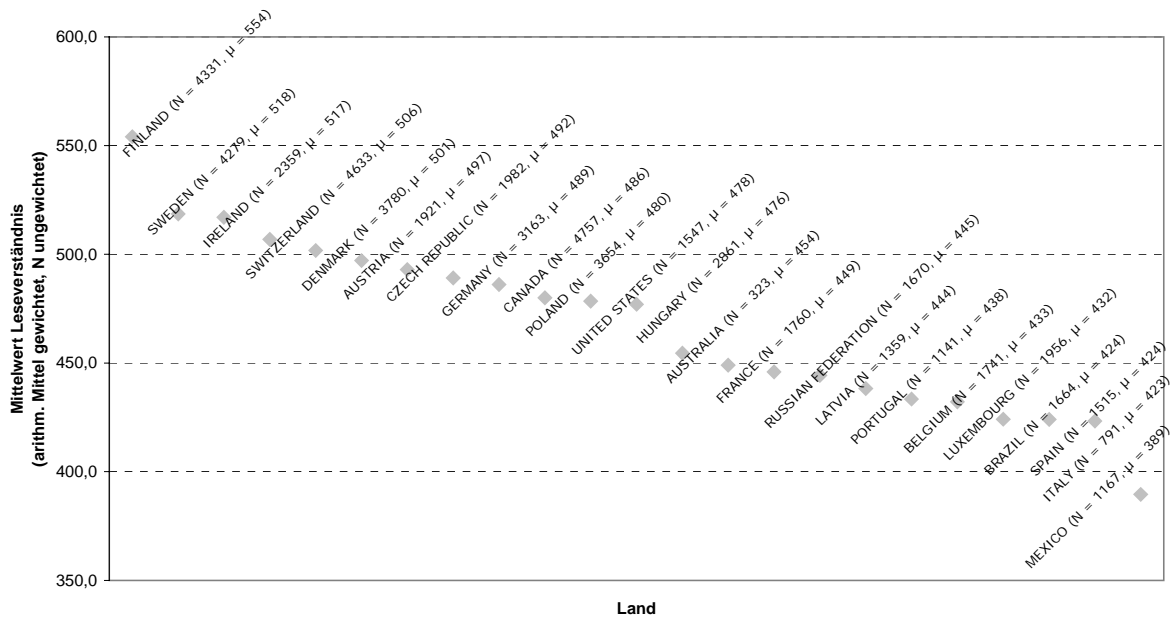


Abbildung 22 Leseverständnis Klassenstufe 9 (gewichtet)

PISA 2000 Leseverständnis (geschätzte Randmittel)

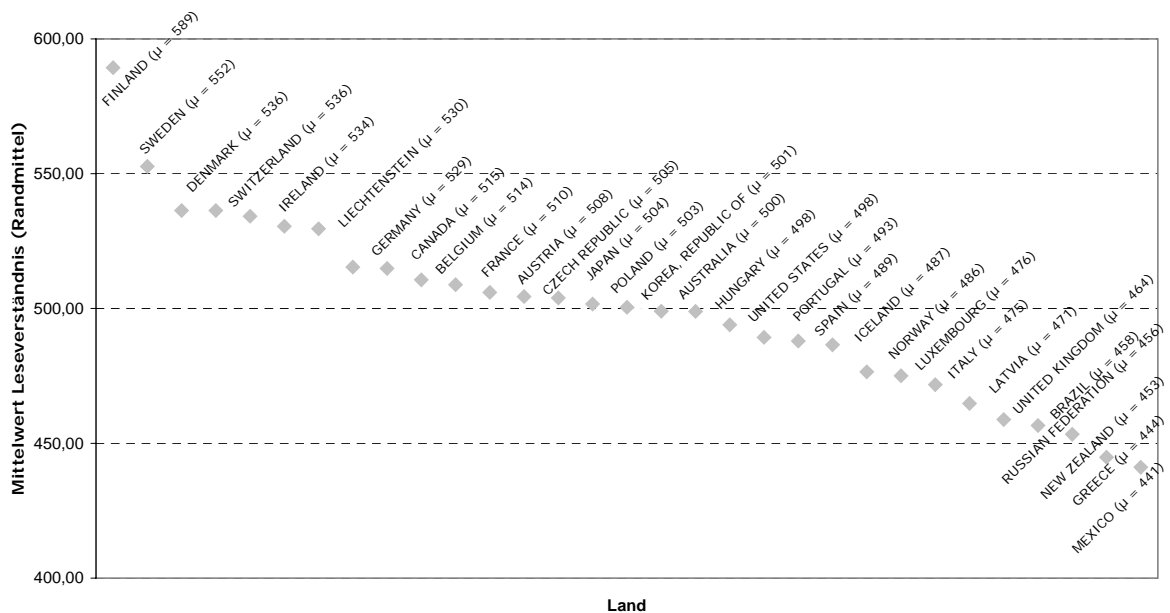


Abbildung 23 Leseverständnis (Randmittel ungewichtet)

**PISA 2000 Leseverständnis
(geschätzte Randmittel)**

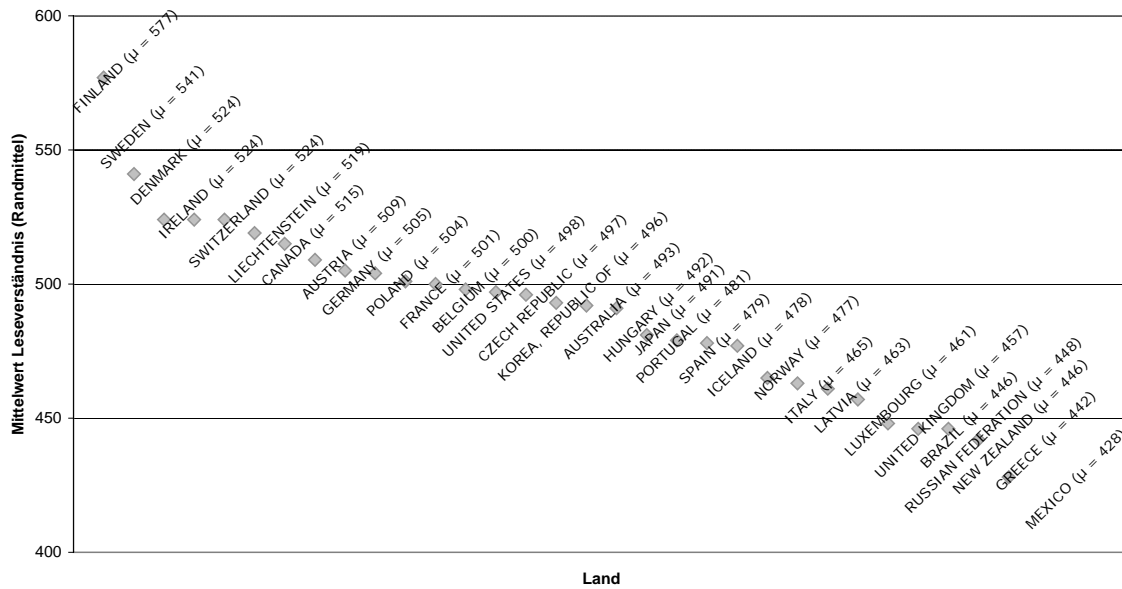


Abbildung 24 Leseverständnis (Randmittel - Datensatz gewichtet)

**PISA 2000 Science
(geschätzte Randmittel)**

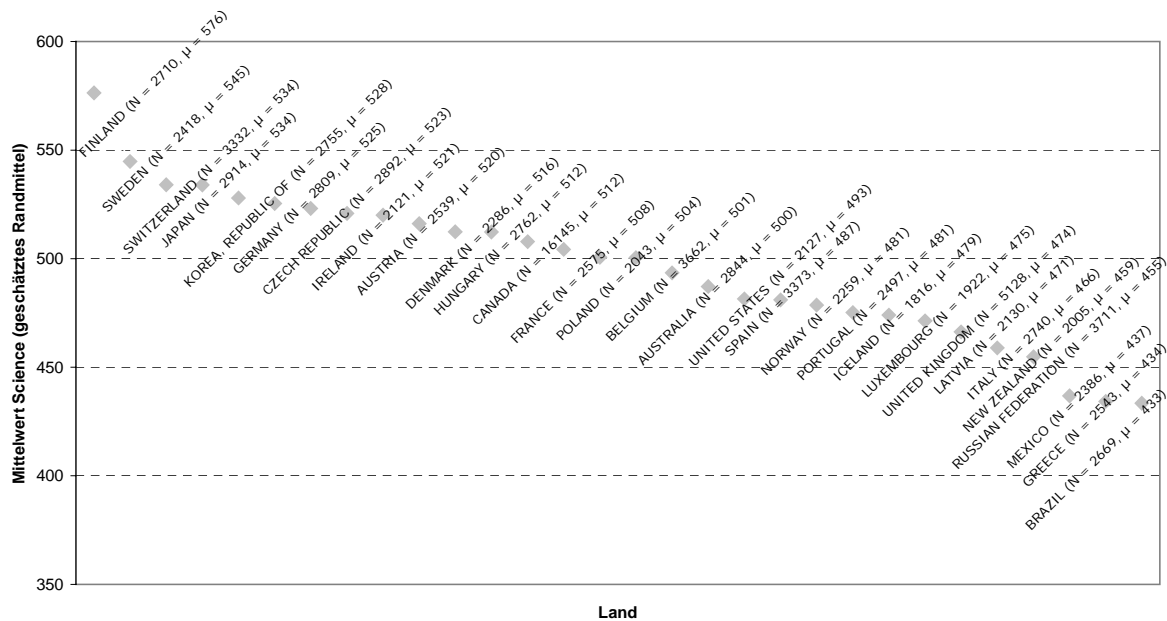


Abbildung 25 Science (Randmittel ungewichtet)

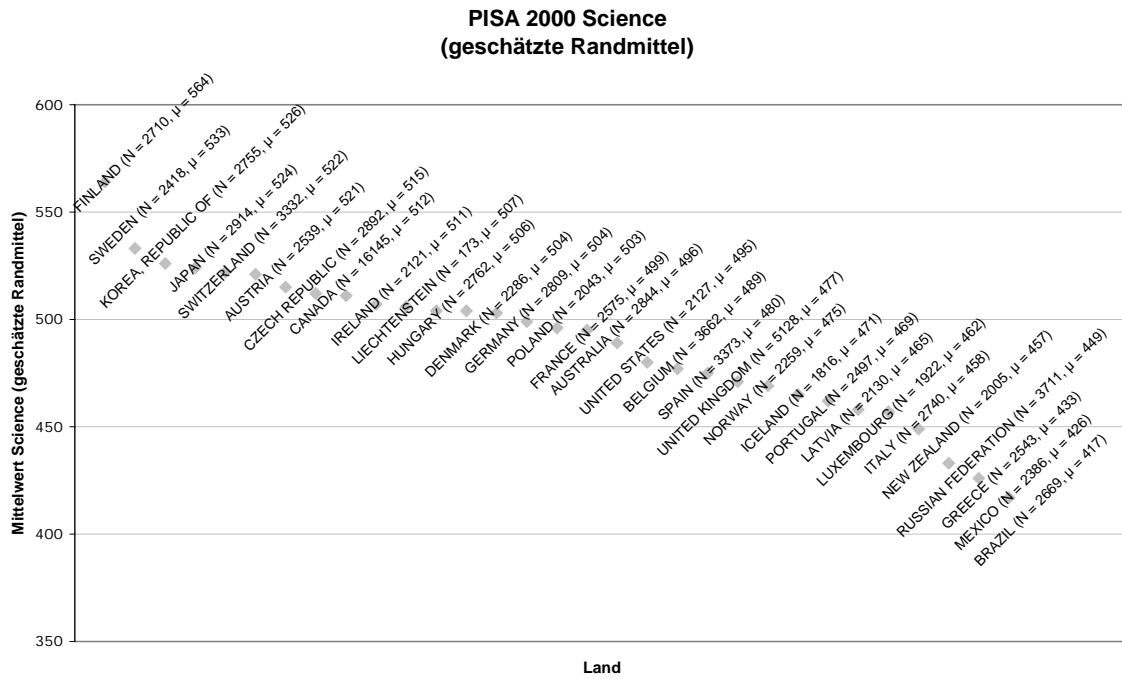


Abbildung 26 Science (Randmittel - Datensatz gewichtet)

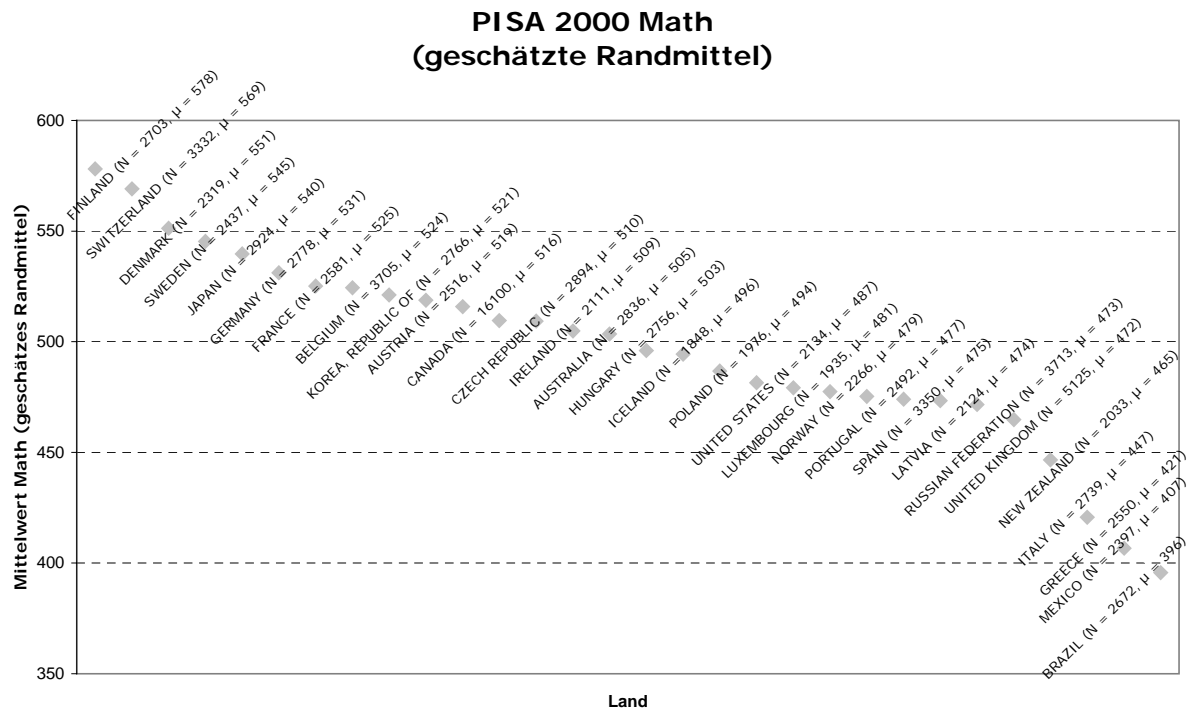


Abbildung 27 Math (Randmittel ungewichtet)

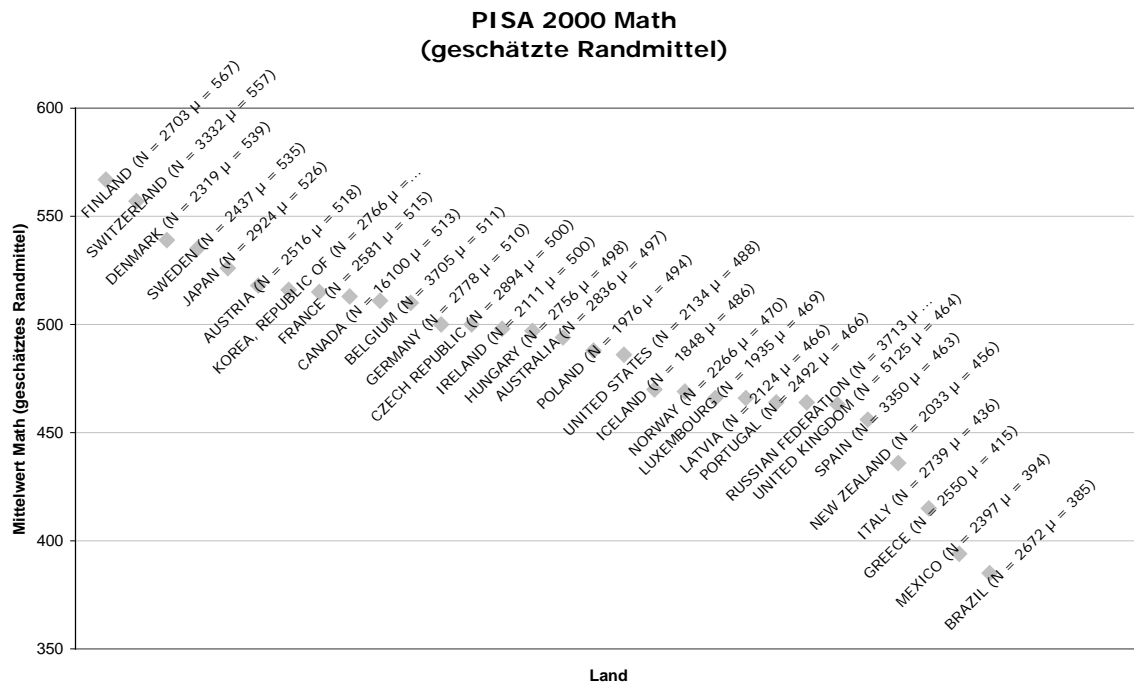


Abbildung 28 Math (Randmittel - Datensatz gewichtet)

2.4.6 Interpretation

Die Schülerinnen und Schüler im deutschen Sample erreichen über die einzelnen Darstellungen einen Testwert jeweils im oberen oder mittleren Bereich. Die Niederlande müssen wegen der niedrigen Ausschöpfungsquote auf Schulebene von nur 27 Prozent bei Mittelwertvergleichen unberücksichtigt bleiben.⁵⁵ Liechtenstein weist eine Differenz zwischen Geburtskohorte und altersgleicher Schulbevölkerung von 21% auf.⁵⁶ Dessen Testwert ist daher möglicherweise verzerrt.

Die Rangreihenfolge ändert sich nach Berücksichtigung der beiden Kontrollvariablen Schulalter und Geschlecht deutlich. Der Mittelwert der neuseeländischen und englischen Schülerinnen und Schüler reduziert sich erheblich. Unter Berücksichtigung unterschiedlicher Einschulungsmodalitäten, die sich auf das Verhältnis von Klassenstufe und tatsächlicher Besuchsdauer beziehen, ließe sich für Neuseeland und das Vereinigte Königreich die Klassenstufe um ein volles Jahr in einem Modellvergleich rein

⁵⁵ Baumert, Stanat, Demmrich: PISA 2000: Untersuchungsgegenstand, theoretische Grundlagen und Durchführung der Studie. In: Deutsches PISA-Konsortium (Hrsg.): PISA 2000: Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich. S. 37

⁵⁶ Baumert, Stanat, Demmrich: PISA 2000: Untersuchungsgegenstand, theoretische Grundlagen und Durchführung der Studie. In: Deutsches PISA-Konsortium (Hrsg.): PISA 2000: Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich. S. 36

rechnerisch reduzieren. Die geschätzten Randmittel über die korrigierte Variable zeigen für beide Länder, wie zu erwarten, höhere mittlere Werte, die jedoch insbesondere im Falle des Vereinigten Königreichs den Wert der Bundesrepublik Deutschland nicht übersteigen.

2.4.7 Zwischenbewertung

Nach Berücksichtigung der Variablen *Klassenstufe* (grade) und *Geschlecht* zeigen sich zu den ursprünglich berichteten Ergebnissen der PISA-Studie abweichende Testwerte. Ursache hierfür ist die Problematik der Vergleichbarkeit.

*Es hat sich als ausgesprochen schwierig erwiesen, Untersuchungspopulationen zu definieren, die für alle Teilnehmerstaaten eindeutig und zugleich innerhalb der Systematik der nationalen Bildungssysteme sinnvoll sind. Es ist kein Zufall, dass die IEA in ihren Schulleistungsuntersuchungen mit mehreren Populationsdefinitionen experimentiert hat. Je nach theoretischer Perspektive kann man Populationen für Schulleistungsuntersuchungen nach dem Lebensalter zu einem bestimmten Stichtag, oder nach dem Schulalter, definiert über die Zugehörigkeit zur Jahrgangsstufe zum Testzeitpunkt, festlegen. Entsprechend der leitenden Fragestellung, der untersuchten Wissensdomäne und der Abhängigkeit des Wissenserwerbs von schulischen Lerngelegenheiten wird man eher an einer Lebensalter- oder an einer Schulalterdefinition interessiert sein. Im Falle der Untersuchung des Leseverständnisses kann man bei Jugendlichen - weniger bei Grundschulern - einer Festlegung der Population über das Lebensalter den Vorzug geben, da Leseerfahrungen in dieser Altersgruppe möglicherweise stärker vom Lebensalter als von der besuchten Klassenstufe abhängen. Bei der Erhebung mathematischer Kompetenzen wird eine Populationsdefinition nach Schulalter näher liegen, da der Erwerb von mathematischem Wissen in hohem Maße von den schulischen Lerngelegenheiten abhängig ist.*⁵⁷

Die Ergebnisse zeigen, dass in der Klassenstufe ein wichtiger Faktor zur Erklärung der international variierenden Testleistungen der 15-jährigen Schülerinnen und Schüler zu sehen ist.

In Bezug auf die Messung mathematischer Kompetenzen folgen die Autoren der eigenen Analyse nicht. Im Bereich der Messung mathematischer und wissenschaftlicher

⁵⁷ Baumert, Stanat, Demmrich: PISA 2000: Untersuchungsgegenstand, theoretische Grundlagen und Durchführung der Studie. In: Deutsches PISA-Konsortium (Hrsg.): PISA 2000: Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich. S. 34

Kompetenz (Science) wurde ebenfalls eine Populationsdefinition auf der Basis des Lebensalters, nicht des Schulalters gewählt.

Dieser Befund macht die hohen Testergebnisse bei gleichzeitig geringer Streubreite der Länder Korea und Japan nachvollziehbar.

In der Darstellung sticht die Besonderheit Koreas hervor: Korea erreicht eine Spitzenleistung mit einer sehr niedrigen Leistungsstreuung. Der untere Leistungsbereich (25. Perzentil) beginnt bei einem Wert von 411; das ist etwa der Kennwert, den der untere Leistungsbereich (25. Perzentil) in Deutschland maximal erreicht. (...) Korea ist damit ein Beispiel für eine gelungene Förderung naturwissenschaftlicher Grundbildung auf einem relativ homogenen hohen Niveau.⁵⁸

Das Ausmaß der Streuung ist selbst ein wichtiger Indikator für die Fähigkeit eines Bildungssystems, die mathematische Bildung seiner Schülerpopulation zu homogenisieren.⁵⁹

In Korea und Japan befinden sich die getesteten 15-jährigen Schülerinnen und Schüler weitgehend beziehungsweise ausschließlich in der 10. Jahrgangsstufe. Eine Streuung über verschiedene Schulalter, wie in den meisten Ländern, existiert praktisch nicht. Der Befund macht ebenfalls verständlich, weshalb Schülerinnen und Schüler im Vereinigten Königreich insgesamt hohe mittlere Leistungen bei großer Streuung erreichen.

Das Vereinigte Königreich liegt in den Leistungen zwar im Mittel um etwa 20 Punkte unter Korea, doch endet hier der obere Leistungsbereich (95. Perzentil) auf einem deutlich höheren Niveau (687), im Vergleich zu Korea. Die Leistungsstreuung ist insgesamt deutlich höher; offensichtlich scheint es im Vereinigten Königreich eine relativ umfangreiche Spitzengruppe mit ausgezeichneter naturwissenschaftlicher Kompetenz zu geben.⁶⁰

⁵⁸ Prenzel, Rost, Senkbeil, Häußler, Klopp: Naturwissenschaftliche Grundbildung: Testkonzeption und Ergebnisse. In: Deutsches PISA-Konsortium (Hrsg.): PISA 2000: Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich. S. 236

⁵⁹ Klieme, Neubrand, Lüdtke: Mathematische Grundbildung: Testkonzeption und Ergebnisse. In: Deutsches PISA-Konsortium (Hrsg.): PISA 2000: Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich. S. 176

⁶⁰ Prenzel, Rost, Senkbeil, Häußler, Klopp: Naturwissenschaftliche Grundbildung: Testkonzeption und Ergebnisse. In: Deutsches PISA-Konsortium (Hrsg.): PISA 2000: Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich. S. 236ff

Der überwiegende Teil englischer Schülerinnen und Schüler befindet sich bereits in der 11. Klassenstufe, jedoch verteilt sich die Altersgruppe stärker als in Japan und Korea.

Die von Stanat und Kunter in PISA berichtete Gleichheit unter den Geschlechtern an Koreas Schulen ist ein weiterer Aspekt.

*Relativ selten kommt es allerdings in der Tat vor, dass 15-Jährige in einem Land sehr gute Ergebnisse im Lesetest erzielen und sich die Leistungen von Mädchen und Jungen kaum unterscheiden. Dies ist lediglich in Korea der Fall (...).*⁶¹

In koreanischen Schulen finden sich den Daten der PISA-Studie zufolge nur noch rund 85% der Altersgruppe der 15-Jährigen, das bedeutet 15% der Schülerinnen und Schüler haben bereits die Schule verlassen. Die frühen Schulabgänger sind zumeist Schülerinnen.⁶² Es entsteht in der verbliebenen Population ein Überhang von ca. 12 Prozentpunkten zugunsten der männlichen Schüler. Die verbliebenen Schülerinnen stellen eine Auswahl dar.

Die relativ hohen Testwerte Japans und Koreas bei geringer Streuung sind nicht nur im Falle Koreas mit einer im Verhältnis der Teilnehmerstaaten geringen Rückhaltequote des Schulsystems verbunden. Parallel neben dem öffentlichen Schulsystem finden sich in Japan und Korea private Schulen (in Japan: JUKU), die in Japan von knapp 60% und in Korea von etwa 36% der Schülerinnen und Schüler (OECD-Durchschnitt 9,4 incl. Japan und Korea) besucht werden.⁶³

Die Situation in Koreas und Japans Schulen, die im PISA-Test berücksichtigt wurden, stellt sich nach Datenlage der Studie wie folgt dar: Die öffentlichen Schulen weisen durchschnittliche Klassengrößen von 38,8 (Japan) und 37,6 (Korea) Schülern auf. Diese Größenordnung entspricht den Verhältnissen der Länder Brasilien (38,8) und Mexiko (35,0). Der berichtete Durchschnitt liegt hier inklusive Japan und Korea bei 24,6. Europäische Länder weisen durchschnittliche Klassenstärken von lediglich 19,5 (Finnland), 18,2 (Schweiz) und 17,4 (Dänemark) auf.⁶⁴ Präzise soziodemographische

⁶¹ Stanat, Kunter: Geschlechterunterschiede in Basiskompetenzen. In: Deutsches PISA-Konsortium (Hrsg.): PISA 2000: Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich. S. 251

⁶² vgl. Abbildung *Verteilung Geschlecht*

⁶³ Schümer, Weiß, Steinert, Baumert, Tillmann, Meier: Lebens- und Lernbedingungen von Jugendlichen. In: Deutsches PISA-Konsortium (Hrsg.): PISA 2000: Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich. S. 417

⁶⁴ Schümer, Weiß, Steinert, Baumert, Tillmann, Meier: Lebens- und Lernbedingungen von Jugendlichen. In: Deutsches PISA-Konsortium (Hrsg.): PISA 2000: Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich. S. 422

Angaben der koreanischen Schulabgänger konnten in der PISA-Studie nicht gefunden werden.

Nach Datenlage wird über Schulprobleme in Japan und Korea besonders wenig in den Elternhäusern gesprochen. 10% der koreanischen und 14% der japanischen Schülerinnen und Schüler geben an, nie mit ihren Eltern über Schulprobleme zu sprechen (Länderdurchschnitt: 4,1%, Finnland 1,6% und Deutschland 1,7%). Rund 43% der koreanischen Schüler geben an, die Lehrer würden sie nie in ihrer Arbeit unterstützen.⁶⁵ Der Länderdurchschnitt beträgt hier 8,6%, Deutschland 13,2%, Schweiz 6,2%, Schweden 3,4%, Finnland 3,1% und Dänemark 2%.⁶⁶

Jeweils etwa zwei Drittel der koreanischen und japanischen Lehrer beanstanden mangelhafte Bauqualität der Schulgebäude,⁶⁷ mangelnde Heizung⁶⁸ und ungenügenden Raum⁶⁹ ihrer Schulen. Zwei Drittel der koreanischen Lehrer beanstanden darüber hinaus einen Mangel an Computern,⁷⁰ eine schlecht ausgestattete Bibliothek⁷¹ und schlechte Ausstattung mit wissenschaftlichem Material.⁷²

Die folgende Abbildung zeigt den statistischen Zusammenhang zwischen Ergänzungsunterricht (supplementary), Nachhilfeunterricht (tuition/tutoring) und Klassengröße (class size). Es lässt sich mit $r_p=0,658$ ($p<0,01$) und $r_s=0,551$ ($p<0,01$) ein statistisch signifikanter Zusammenhang der Klassengröße in Bezug auf Ergänzungsunterricht in Prozent der Nennungen und mit $r_p=0,417$ ($p<0,05$) und $r_s=0,482$ ($p<0,01$) ein statistischer Zusammenhang in Bezug auf Nachhilfeunterricht nachweisen. Die Abbildung legt einen kompensatorischen Zusammenhang nahe. Finnland, Schweden und die Schweiz weisen kleine Klassengrößen bei sehr geringem zusätzlichem Unterricht auf. Japan und Korea hingegen, gemeinsam mit Mexiko und Brasilien, weisen die zahlenmäßig umfangreichsten durchschnittlichen Klassengrößen auf, wobei Mexiko und Brasilien bei niedrigen Testleistungen einen geringen Anteil an zusätzlichem Unterricht, Japan und Korea jedoch bei hohen Testergebnissen einen hohen Anteil an privatem Zusatzunterricht aufweisen.

⁶⁵ Codierung: teachers help with work Stufe: never

⁶⁶ Berechnungen aus Rohdaten der Public-Use-Files

⁶⁷ Item sc11q01 (poor buildings) Zusammenfassung: a little, some, a lot

⁶⁸ Item sc11q02 (poor heating) Zusammenfassung: a little, some, a lot

⁶⁹ Item sc11q03 (inadequate space) Zusammenfassung: a little, some, a lot

⁷⁰ Item SC11Q05 (lack of computers) Zusammenfassung: a little, some, a lot

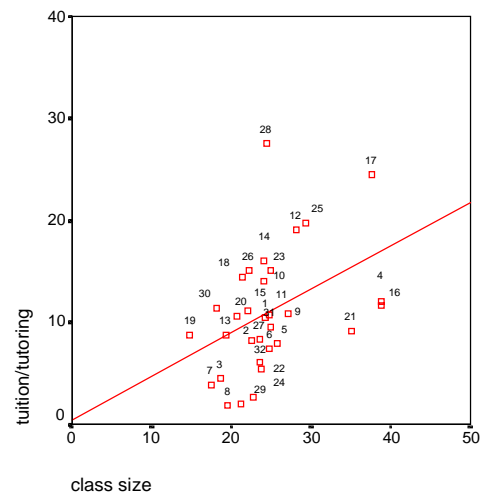
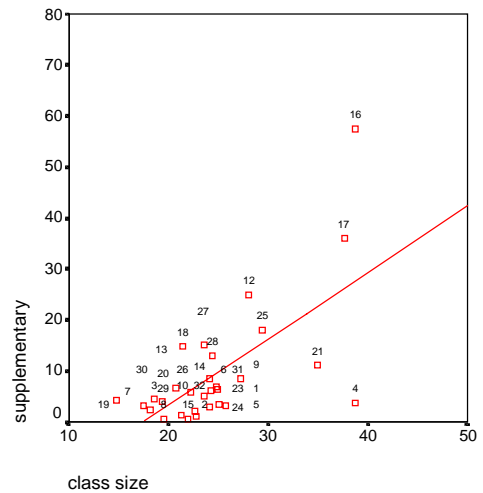
⁷¹ Item SC11Q06 (poor library) Zusammenfassung: a little, some, a lot

⁷² Item SC11Q08 (poor science equipment) Zusammenfassung: a little, some, a lot

country	case no	class size *)	supplementary *)	tuition / tutoring *)	HISEI**)
AUSTRALIA	1	24,30	6,10	10,50	52,25
AUSTRIA	2	22,60	2,20	8,20	49,72
BELGIUM	3	18,60	4,40	4,50	48,95
BRAZIL	4	38,80	3,70	12,10	43,93
CANADA	5	25,70	3,30	8,00	52,83
CZECH REPUBLIC	6	24,80	6,80	7,40	48,31
DENMARK	7	17,40	3,10	3,80	49,73
FINLAND	8	19,50	0,50	1,90	50,00
FRANCE	9	27,20	8,60	10,90	48,27
GERMANY	10	24,10	3,00	14,10	48,85
GREECE	11	24,80		10,70	47,76
HUNGARY	12	28,10	25,00	19,10	49,53
ICELAND	13	19,30	3,90	8,70	52,73
IRELAND	14	24,10	8,60	16,00	48,43
ITALY	15	22,00	0,60	11,10	47,08
JAPAN	16	38,80	57,50	11,60	50,54
KOREA, REPUBLIC OF	17	37,60	36,00	24,50	42,80
LATVIA	18	21,40	14,80	14,50	50,15
LIECHTENSTEIN	19	14,80	4,20	8,80	47,46
LUXEMBOURG	20	20,70	6,70	10,60	44,79
MEXICO	21	35,00	11,20	9,20	42,48
NETHERLANDS	22	23,80		5,40	50,85
NEW ZEALAND	23	25,00	3,50	15,10	52,20
NORWAY	24	22,70	1,10	2,70	53,91
POLAND	25	29,40	18,10	19,80	46,03
PORTUGAL	26	22,20	5,70	15,10	43,85
RUSSIAN FEDERATION	27	23,50	15,10	8,30	49,38
SPAIN	28	24,40	13,10	27,60	44,99
SWEDEN	29	21,20	1,30	2,00	50,57
SWITZERLAND	30	18,20	2,40	11,40	49,21
UNITED KINGDOM	31	24,90	6,30	9,50	51,26
UNITED STATES	32	23,50	5,10	6,10	52,40

*) Reference: PISA 2000, P.417, 422

**) DATABASE: INSTUD_READ (WITH CONSID. OF. W_FSTUWT - STUD. POP. WEIGHT -)
download <http://pisaweb.acer.edu.au/oecd>



Korrelationen

		supplementary	tuition/tutoring
class size	Korrelation nach Pearson	,658**	,417*
	Signifikanz (2-seitig)	,000	,018
	N	30	32

** Die Korrelation ist auf dem Niveau von 0,01 (2-seitig) signifikant.

* Die Korrelation ist auf dem Niveau von 0,05 (2-seitig) signifikant.

Korrelationen

		supplementary	tuition/tutoring
Spearman-Rho	class size	Korrelationskoeffizient	,551**
		Sig. (2-seitig)	,002
		N	30

** Die Korrelation ist auf dem 0,01 Niveau signifikant (zweiseitig).

Abbildung 29 Klassengröße – Extra Unterricht

Mit $r_p = -0,512$ ($p < 0,01$) $r_s = -0,498$ ($p < 0,01$) zeigt sich ein negativer statistischer Zusammenhang von durchschnittlichem sozioökonomischem Index der Schüler eines Landes und prozentualem Anteil privater Nachhilfe.

		tuition / tutoring *)
HISEI**) Korrelation nach Pearson		-,512**
Signifikanz (2-seitig)		,003
N		32

** . Die Korrelation ist auf dem Niveau von 0,01 (2-seitig) signifikant.

			tuition / tutoring *)
Spearman-Rho	HISEI**) Korrelationskoeffizient		-,498**
	Sig. (2-seitig)		,004
	N		32

** . Die Korrelation ist auf dem 0,01 Niveau signifikant (zweiseitig).

Abbildung 30 Lerngelegenheiten – Soz.Ökon. Index/Korrelation

Je geringer der durchschnittliche sozioökonomische Index der 15-jährigen Schülerpopulation eines Landes ist, desto mehr außerschulische Lerngelegenheiten werden in Anspruch genommen.

Geringer sozioökonomischer Index kovariiert positiv mit hohen Klassenfrequenzen. Japan und Korea bilden mit einem hohen Anteil außerschulischer Lerngelegenheiten hier die Ausnahme. Finnland, Schweden, Dänemark und weitere westeuropäische Teilnehmerstaaten hingegen weisen sehr geringe Anteile außerschulischer Lerngelegenheiten auf. In Finnland berichten lediglich 1,9% der Schülerinnen und Schüler die Inanspruchnahme von extra Nachhilfeunterricht und nur 0,5% von weiterem Ergänzungsunterricht.

2.5 Soziale Disparität in der Lesekompetenz

Baumert und Schümer berichten *Disparitäten* im Erwerb von Lesekompetenz.⁷³ Für jedes Land wurden die Differenzen der Mittelwerte von zwei Schülergruppen gebildet: die mittleren Testleistungen der 25% Jugendlichen aus Familien mit dem höchsten Sozialstatus innerhalb des Landes und die mittleren Testleistungen der 25% Jugendlichen aus Familien mit dem niedrigsten Sozialstatus. Die Werte in der folgenden Abbildung zeigen, welchen Testleistungsvorsprung die erste Gruppe gegenüber der zweiten aufweist.

⁷³ Baumert, Schümer: Familiäre Lebensverhältnisse, Bildungsbeteiligung und Kompetenzerwerb. In: Deutsches PISA-Konsortium (Hrsg.): PISA 2000: Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich. S. 383

*In Deutschland, gefolgt von den beiden ehemaligen sozialistischen Staaten Tschechische Republik und Ungarn sowie der Schweiz, Luxemburg und Belgien, können am Ende der Sekundarstufe I soziale Disparitäten der Lesekompetenz nachgewiesen werden, wie sie in anderen Ländern – allen voran Japan, Korea, Island oder Finnland - kaum denkbar sind.*⁷⁴

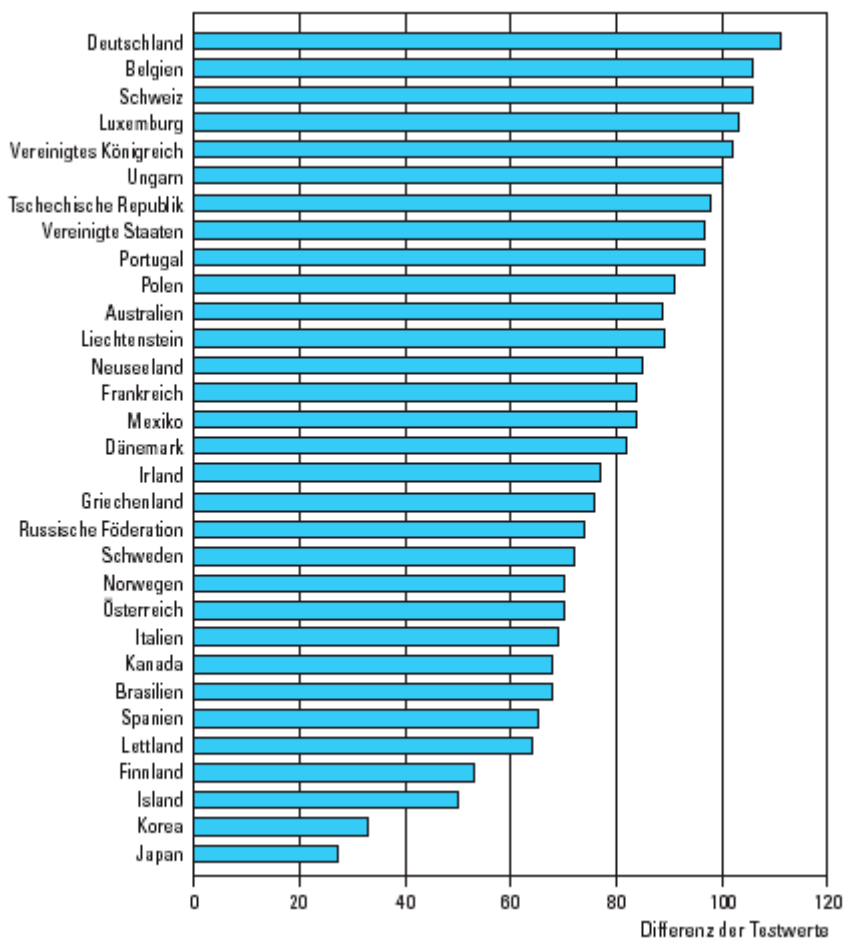


Abbildung 31 Disparität Quelle: PISA 2000

Die oben gezeigte Abbildung stellt die Teilnehmerstaaten nach der Differenz zwischen der mittleren Lesekompetenz von 15-Jährigen aus Familien des oberen und unteren Viertels der Sozialstruktur dar.

Der Befund wird für viele, auch mit der Disparitätsforschung vertraute Leser überraschend sein: Deutschland und die Schweiz sind die beiden Länder mit den größten Unterschieden in der Lesekompetenz von Jugendlichen aus höheren und

⁷⁴ Baumert, Schümer: Familiäre Lebensverhältnisse, Bildungsbeteiligung und Kompetenzerwerb. In: Deutsches PISA-Konsortium (Hrsg.): PISA 2000: Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich. S. 389

*niedrigeren Sozialschichten. Die Differenz beträgt in Deutschland mehr als eineinhalb Kompetenzstufen oder 1,2 Standardabweichungen. Selbst die Vereinigten Staaten, die immer wieder als Beispiel für große soziale Disparitäten in den Bildungschancen angeführt werden, weisen zwar immer noch beträchtliche, aber signifikant niedrigere sozial bedingte Leistungsunterschiede auf. (...) In Korea sinken die Disparitätswerte auf 36 Punkte und in Japan auf 27 Punkte, wohlgermerkt bei gleich großer sozialer Heterogenität der Elternhäuser (...).*⁷⁵

Auffallend sind nicht nur die hohen sozialen Disparitäten in der Schweiz, sondern auch die moderaten Disparitäten in der Russischen Föderation, Mexiko, insbesondere aber die außerordentlich geringen Disparitäten Brasiliens.

2.5.1 Streuung des Schulalters

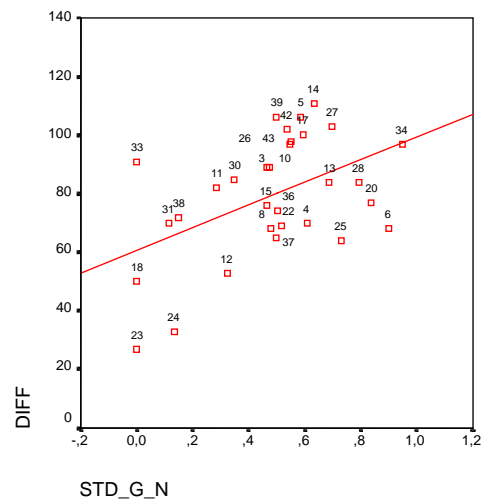
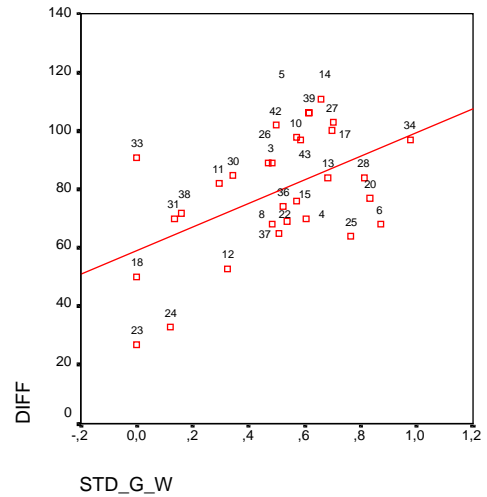
Die am Lebensalter orientierte Definition der PISA-Studie hat eine Variabilität des Schulalters und wie bereits gezeigt wurde einen Einfluss auf die Testwerte zur Folge. Es ist nun zu prüfen, in welchem Umfang diese Variabilität einen Einfluss auf die *Disparität* im Lesen (wie oben dargestellt) ausübt. Mit anderen Worten, es ist zu prüfen, ob die hohen *Disparitäten* im Lesen auf die Stichprobendefinition zurückzuführen sind. Geprüft wird der Zusammenhang von *Disparität* entsprechend den vom MPIB veröffentlichten Daten (hier als DIFF codiert)⁷⁶ und der Standardabweichung der Länderstichproben jeweils über die gewichtete (STD_G_W) und ungewichtete Variable Klassenstufe (STD_G_U).

⁷⁵ Baumert, Schümer: Familiäre Lebensverhältnisse, Bildungsbeteiligung und Kompetenzerwerb. In: Deutsches PISA-Konsortium (Hrsg.): PISA 2000: Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich. S. 383

⁷⁶ Baumert, Schümer: Familiäre Lebensverhältnisse, Bildungsbeteiligung und Kompetenzerwerb. In: Deutsches PISA-Konsortium (Hrsg.): PISA 2000: Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich. S. 384 ff

COUNTRY	CASE No	STD_G_w	STD_G_n	DIFF
ALBANIA	1	0,62	0,52	
ARGENTINA	2	0,74	0,69	
AUSTRALIA	3	0,48	0,47	89
AUSTRIA	4	0,60	0,61	70
BELGIUM	5	0,62	0,59	106
BRAZIL	6	0,87	0,90	68
BULGARIA	7	0,47	0,43	
CANADA	8	0,48	0,48	68
CHILE	9	0,76	0,70	
CZECH REPUBLIC	10	0,57	0,55	98
DENMARK	11	0,29	0,28	82
FINLAND	12	0,32	0,32	53
FRANCE	13	0,68	0,69	84
GERMANY	14	0,66	0,63	111
GREECE	15	0,57	0,46	76
HONG KONG	16	0,90	0,89	
HUNGARY	17	0,70	0,59	100
ICELAND	18	0,00	0,00	50
INDONESIA	19	0,93	0,91	
IRELAND	20	0,83	0,84	77
ISRAEL	21	0,33	0,36	
ITALY	22	0,54	0,52	69
JAPAN	23	0,00	0,00	27
KOREA, REPUBLIC OF	24	0,12	0,13	33
LATVIA	25	0,76	0,73	64
LIECHTENSTEIN	26	0,47	0,47	89
LUXEMBOURG	27	0,70	0,70	103
MEXICO	28	0,81	0,79	84
NETHERLANDS	29	0,62	0,60	
NEW ZEALAND	30	0,34	0,35	85
NORWAY	31	0,13	0,11	70
PERU	32	1,09	1,06	
POLAND	33	0,00	0,00	91
PORTUGAL	34	0,98	0,95	97
ROMANIA	35	0,60	0,47	
RUSSIAN FEDERATION	36	0,52	0,50	74
SPAIN	37	0,51	0,50	65
SWEDEN	38	0,16	0,15	72
SWITZERLAND	39	0,62	0,50	106
THAILAND	40	0,56	0,54	
MACEDONIA	41	0,52	0,48	
UNITED KINGDOM	42	0,50	0,53	102
UNITED STATES	43	0,59	0,55	97

STD_G_w standard deviation GRADE -weighted by W_FSTUWT
 STD_G_n standard deviation GRADE -unweighted
 DIFF "Disparität" (disparity)
 PISA 2000 database : http://pisaweb.acer.edu.au/oeecd/oeecd_pisa_data_s1.html
 data: intstud_read.txt



Korrelationen

	STD_G_W	STD_G_N
DIFF Korrelation nach Pearson	,514**	,483**
Signifikanz (2-seitig)	,003	,006
N	31	31

** - Die Korrelation ist auf dem Niveau von 0,01 (2-seitig) signifikant.

Korrelationen

	STD_G_W	STD_G_N
Spearman-Rho DIFF Korrelationskoeffizient	,390*	,379*
Sig. (2-seitig)	,030	,035
N	31	31

*. Die Korrelation ist auf dem 0,05 Niveau signifikant (zweiseitig).

Abbildung 32 Disparität im Lesen – Streuung des Schulalters

Mit $r_p = 0,514$ ($p < 0,01$), $r_s = 0,39$ ($p < 0,05$) (gewichteter Datensatz) und $r_p = 0,483$ ($p < 0,01$), $r_s = 0,379$ ($p < 0,05$) lässt sich ein signifikanter Zusammenhang zwischen den Disparitäten und der jeweiligen Streuung über die Schulalteranteile nachweisen.

Dieses Ergebnis zeigt einen signifikanten Einfluss der Stichprobendefinition.

Durch die am Lebensalter orientierte Stichprobendefinition variiert das Schulalter von der 4. bis zur 13. Klassenstufe. Es soll im zweiten Schritt geprüft werden, ob ein Testkonstrukt vorliegt, das auf diese hohe Variabilität abgestimmt ist.

2.5.2 Bodeneffekte/Deckeneffekte

So genannte Decken- beziehungsweise Bodeneffekte⁷⁷ treten bei ungeeignetem Verhältnis von Schwierigkeit und Leistungsniveau einer untersuchten Population auf, insbesondere bei großer Leistungsheterogenität. Die folgenden Abbildungen zeigen die Messung von Differenzen der Mittelwerte (entsprechend der Messung von so genannten *Disparitäten*) in Abhängigkeit der Schiefe einer Verteilung.

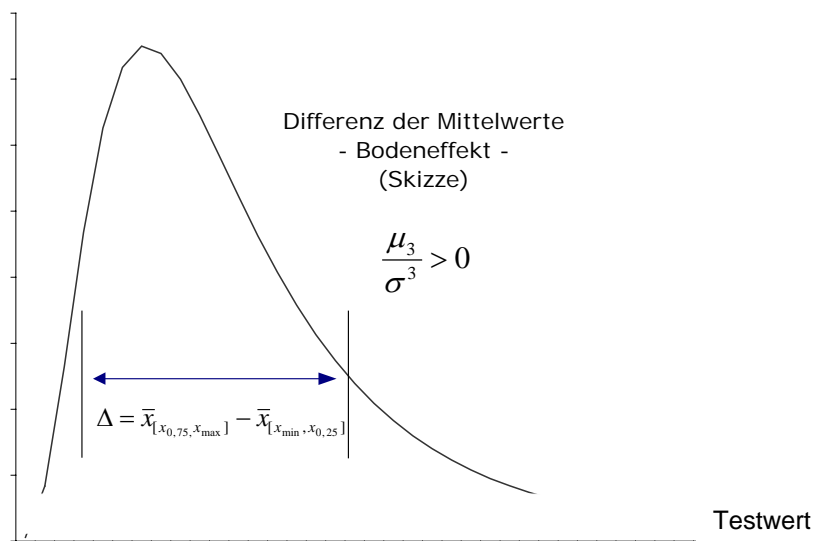


Abbildung 33 *Disparität* - schiefe Testwertverteilung (Bodeneffekt)

⁷⁷ Rost, J.: Testtheorie Testkonstruktion 1996, S.96

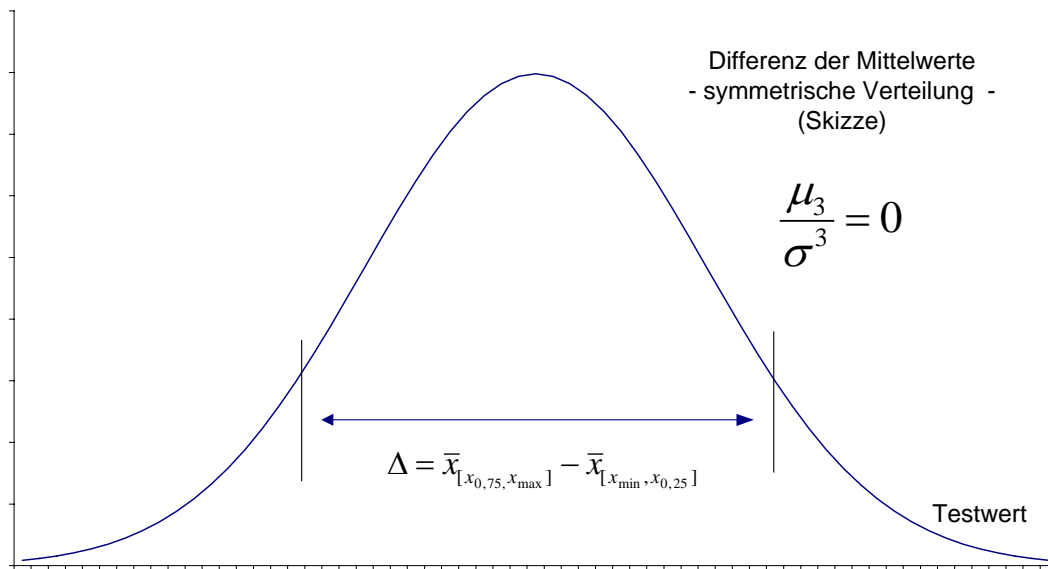


Abbildung 34 *Disparität* - symmetrische Testwertverteilung

Im Vergleich einer Verteilung mit $\frac{\mu_3}{\sigma^3} > 0$ (Bodeneffekt) gegen eine Verteilung mit $\frac{\mu_3}{\sigma^3} = 0$ (symmetrische Verteilung) wird die Ursache geringerer Differenz ($\Delta = \bar{x}_{[x_{0,75}, x_{max}]} - \bar{x}_{[x_{min}, x_{0,25}]}$) bei schiefer Verteilung deutlich. Da die Verteilung bei Vorhandensein eines Bodeneffektes eine untere Schranke aufweist, verringern sich ebenso die über die Mittelwerte $\bar{x}_{[x_{0,75}, x_{max}]}$ und $\bar{x}_{[x_{min}, x_{0,25}]}$ gemessenen Differenzen. Gleiches gilt für so genannte Deckeneffekte (obere Schranke), die hier jedoch nicht extra dargestellt werden. Decken- und Bodeneffekte können als weitere Ursache für die hohen statistischen *Disparitäten* nachgewiesen werden. Ein Deckeneffekt läge vor, wenn bei steigendem Mittelwert der Daten bis zum 1. Quartil, also das Intervall $[x_{min}, x_{0,25}]$ (hier QUANT 1 genannt), die Differenz zwischen diesem Mittelwert und dem Mittelwert des Bereichs ab dem 3. Quartil, also das Intervall $[x_{0,75}, x_{max}]$, systematisch kleiner wird sowie parallel dazu die Schiefe der Verteilung zunimmt. Die Schiefe der Verteilung wäre zunehmend linksschief (rechtssteil) ($\mu_3/\sigma^3 < 0$).⁷⁸ Für Bodeneffekte gilt entsprechend der umgekehrte Zusammenhang (siehe vorangegangene Abbildung mit $\mu_3/\sigma^3 > 0$).

⁷⁸ Notation nach Büning: Nichtparametrische Statistische Methoden 1994. S. 304

Insbesondere im Split Klassenstufe (grade) 10 lässt sich ein solcher Effekt nachweisen. Bei steigendem unterem Mittelwert im Intervall $[x_{\min}, x_{0,25}]$ (Quant1) steigt ebenfalls der mittlere Testwert eines Landes (PV1READ), jedoch bei sinkender Differenz zwischen den Mittelwerten aus den Bereichen $[x_{0,75}, x_{\max}]$ und $[x_{\min}, x_{0,25}]$ hier als DIFF bezeichnet.

Korrelationen

		QUANT1	PV1READ	DIFF
QUANT1	Korrelation nach Pearson	1	,955*	-,429*
	Signifikanz (2-seitig)	.	,000	,029
	N	26	26	26
PV1READ	Korrelation nach Pearson	,955*	1	-,164
	Signifikanz (2-seitig)	,000	.	,425
	N	26	26	26
DIFF	Korrelation nach Pearson	-,429*	-,164	1
	Signifikanz (2-seitig)	,029	,425	.
	N	26	26	26

** · Die Korrelation ist auf dem Niveau von 0,01 (2-seitig) signifikant.

* · Die Korrelation ist auf dem Niveau von 0,05 (2-seitig) signifikant.

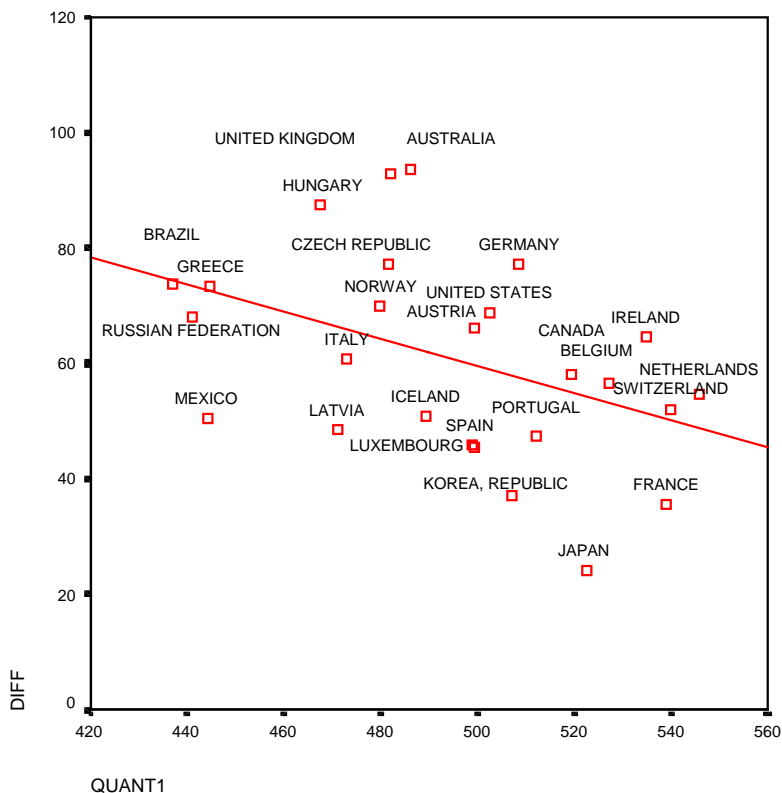


Abbildung 35 Deckeneffekt/Bodeneffekt (Abbildung I)

Der lineare Zusammenhang lässt sich mit $r_p = -0,429$ $p=0,029$ beziehungsweise $r_p = 0,955$ $p < 0,01$ nachweisen.

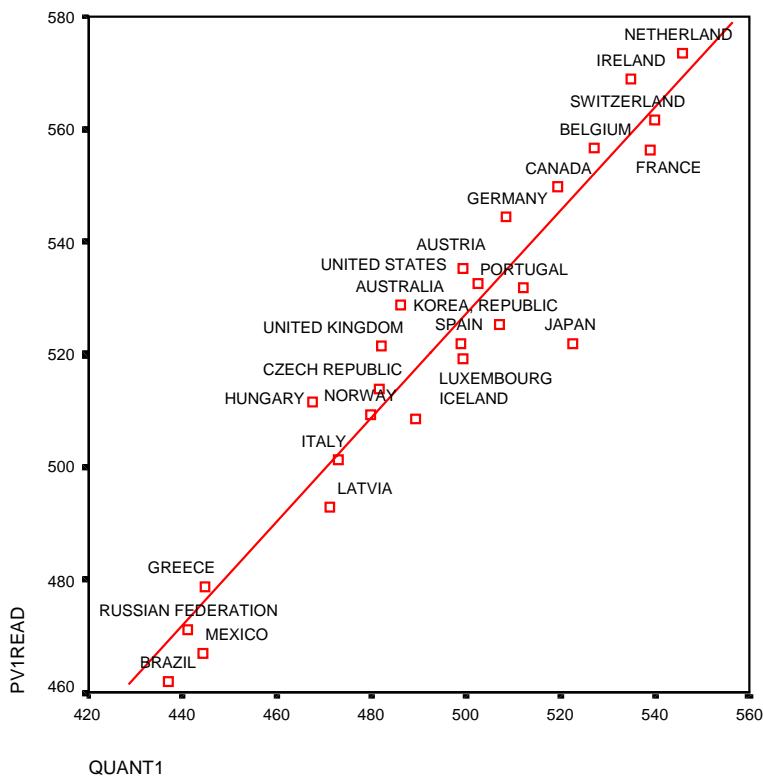


Abbildung 36 Deckeneffekt/Bodeneffekt (Abbildung II)

Für die 11. Klassenstufe lassen sich aufgrund der geringen Fallzahl keine Berechnungen sinnvoll durchführen. Die Klassenstufe 9 hingegen weist diesen Effekt nicht mehr auf, wie folgende Abbildungen zeigen. Es besteht, wie zu erwarten, ein statistisch signifikanter Zusammenhang zwischen dem unteren Mittelwert der Werte im Intervall $[x_{\min}, x_{0,25}]$ und dem mittlerem Testwert eines Landes ($r_p = 0,931$ $p < 0,01$), jedoch kein statistisch signifikanter Zusammenhang der Differenz der Mittelwerte aus den Bereichen $[x_{\min}, x_{0,25}]$ und $[x_{0,75}, x_{\max}]$.

Korrelationen

		DIF_9	QUANT1	PVREAD
DIF_9	Korrelation nach Pearson	1	-,022	,293
	Signifikanz (2-seitig)	.	,919	,155
	N	25	25	25
QUANT1	Korrelation nach Pearson	-,022	1	,931**
	Signifikanz (2-seitig)	,919	.	,000
	N	25	25	25
PVREAD	Korrelation nach Pearson	,293	,931**	1
	Signifikanz (2-seitig)	,155	,000	.
	N	25	25	25

** . Die Korrelation ist auf dem Niveau von 0,01 (2-seitig) signifikant.

Abbildung 37 Disparität/Korrelation

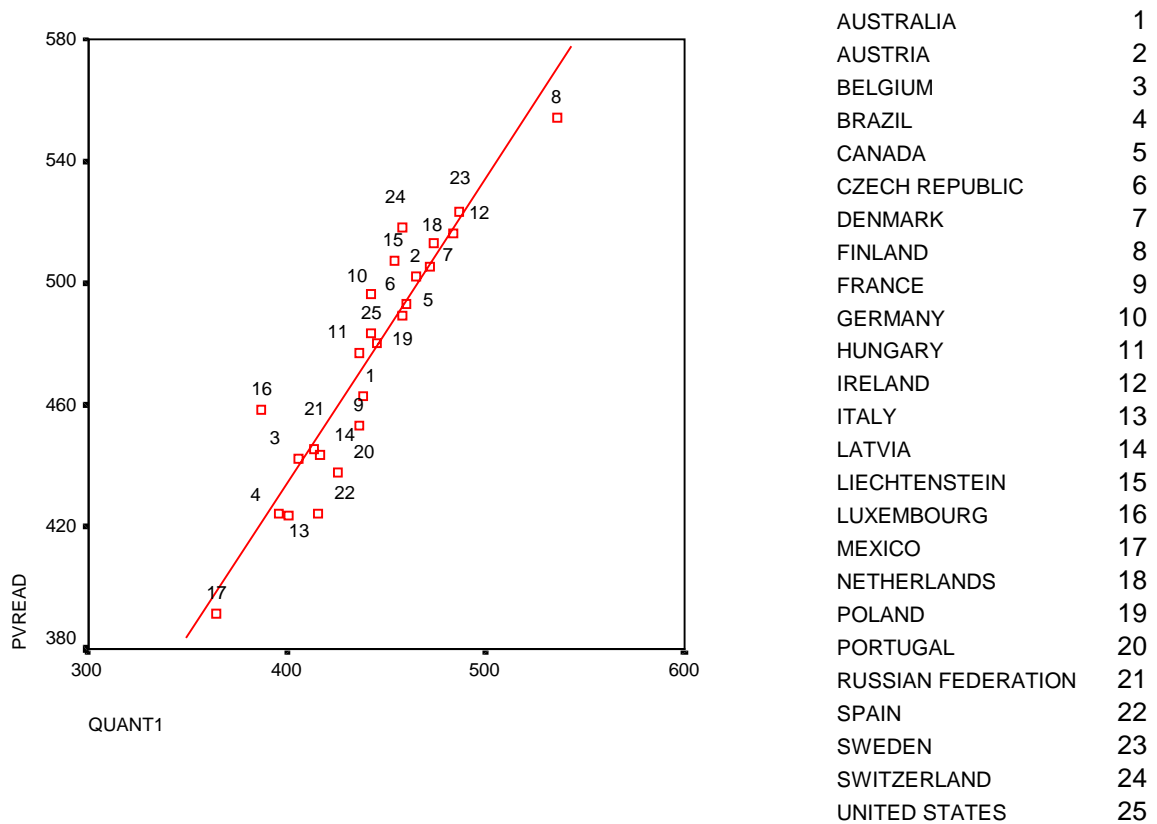


Abbildung 38 Disparität Lesetestwert Klassenstufe (GRADE) 9

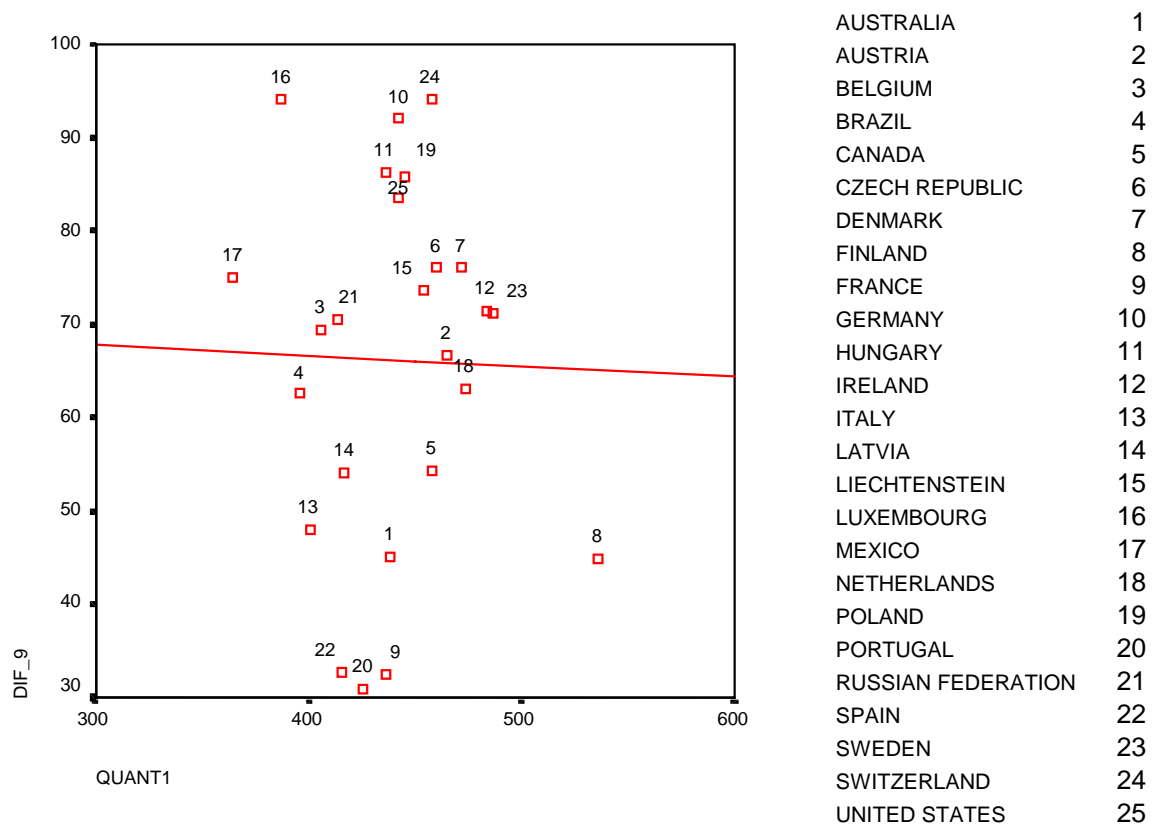
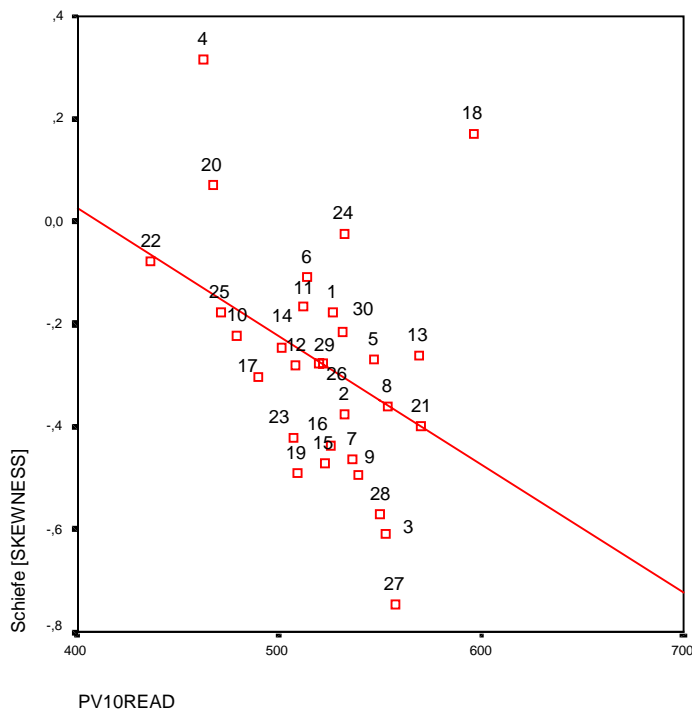


Abbildung 39 Disparität Differenz/Korrelation Klassenstufe (GRADE) 9

Prinzipiell könnte der Zusammenhang von mittlerem Testwert und Schiefe der Verteilung graphisch abgetragen werden.



- AUSTRALIA 1
- AUSTRIA 2
- BELGIUM 3
- BRAZIL 4
- CANADA 5
- CZECH REPUBLIC 6
- DENMARK 7
- FRANCE 8
- GERMANY 9
- GREECE 10
- HUNGARY 11
- ICELAND 12
- IRELAND 13
- ITALY 14
- JAPAN 15
- KOREA, REPUBLIC OF 16
- LATVIA 17
- LIECHTENSTEIN 18
- LUXEMBOURG 19
- MEXICO 20
- NETHERLANDS 21
- NEW ZEALAND 22
- NORWAY 23
- PORTUGAL 24
- RUSSIAN FEDERATION 25
- SPAIN 26
- SWEDEN 27
- SWITZERLAND 28
- UNITED KINGDOM 29
- UNITED STATES 30

Korrelationen

		SKEWNESS
PV10READ	Korrelation nach Pearson	-,386*
	Signifikanz (2-seitig)	,035
	N	30

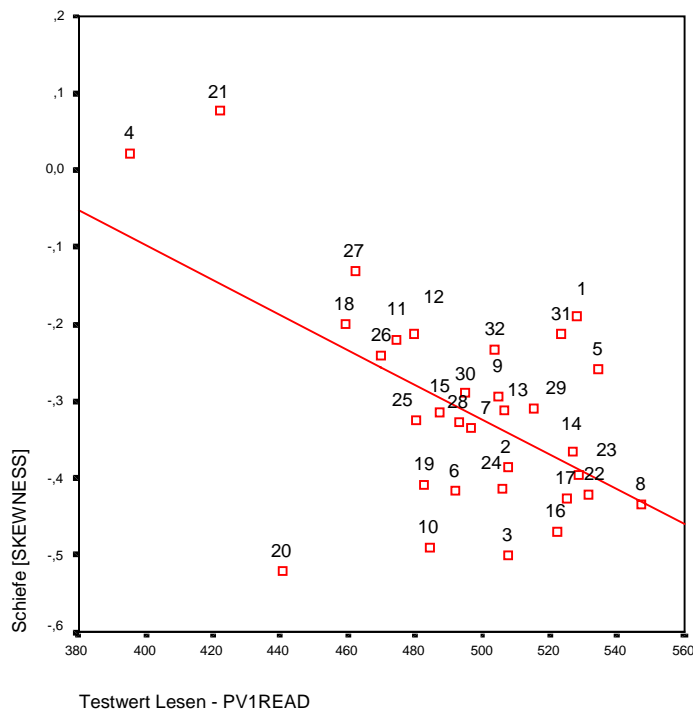
*. Die Korrelation ist auf dem Niveau von 0,05 (2-seitig) signifikant.

Korrelationen

			SKEWNESS
Spearman-Rho	PV10READ	Korrelationskoeffizient	-,422*
		Sig. (2-seitig)	,020
		N	30

*. Die Korrelation ist auf dem 0,05 Niveau signifikant (zweiseitig).

Abbildung 40 Testwert – Schiefe/Korrelation (Klassenstufe 10)



- AUSTRALIA 1
- AUSTRIA 2
- BELGIUM 3
- BRAZIL 4
- CANADA 5
- CZECH REPUBLIC 6
- DENMARK 7
- FINLAND 8
- FRANCE 9
- GERMANY 10
- GREECE 11
- HUNGARY 12
- ICELAND 13
- IRELAND 14
- ITALY 15
- JAPAN 16
- KOREA, REPUBLIC OF 17
- LATVIA 18
- LIECHTENSTEIN 19
- LUXEMBOURG 20
- MEXICO 21
- NETHERLANDS 22
- NEW ZEALAND 23
- NORWAY 24
- POLAND 25
- PORTUGAL 26
- RUSSIAN FEDERATION 27
- SPAIN 28
- SWEDEN 29
- SWITZERLAND 30
- UNITED KINGDOM 31
- UNITED STATES 32

Korrelationen

		SKEWNESS
PV1READ	Korrelation nach Pearson	-,551**
	Signifikanz (2-seitig)	,001
	N	32

** . Die Korrelation ist auf dem Niveau von 0,01 (2-seitig) signifikant.

Korrelationen

		SKEWNESS
Spearman-Rho	PV1READ Korrelationskoeffizient	-,371*
	Sig. (2-seitig)	,037
	N	32

*. Die Korrelation ist auf dem 0,05 Niveau signifikant (zweiseitig).

Abbildung 41 Lesetestwert (sämtliche Klassenstufen) Schiefe der Verteilung

Mit $r_p = -0,551$ ($p < 0,01$) und $r_s = -0,371$ ($p < 0,05$) zeigt sich die Problematik einer geeigneten Testkonstruktion angesichts einer außerordentlich heterogenen Leistungsfähigkeit der untersuchten Schüler. Als Konsequenz der am Lebensalter orientierten Stichprobendefinition umfasst die Stichprobe Schülerinnen und Schüler von der vierten bis zur dreizehnten Klassenstufe. Diese werden in einem gemeinsamen

Testverfahren bewertet.⁷⁹ Unter diesen Umständen werden Decken- und Bodeneffekte (bei so genannten J-förmigen Verteilungen) kaum vermeidbar sein.⁸⁰ Lienert und Raatz sprechen in diesem Zusammenhang von einer *anormalen Verteilung infolge mangelhafter Testkonstruktion*.⁸¹ Korrespondierende Daten liefern auch die Verteilungen der PISA Teilnehmerstaaten.

2.6 Migration/durchschnittliches Schulalter

Nachdem gezeigt wurde, dass das Schulalter (Klassenstufe) im Datensatz einen statistisch signifikanten Einfluss auf die Testleistung ausübt, ist nun die Frage von Interesse, ob ein unterschiedliches Schulalter der einzelnen Segmente der Schülerpopulationen eines Landes nachweisbar ist.

Eine Reihe europäischer Staaten weist umfangreiche Migrationsanteile in ihrer Schülerpopulation auf. Der Anteil variiert je nach Definition. Die folgend gebrauchte Definition richtet sich nicht nach der Verkehrssprache in der Familie, sondern nach dem Geburtsort des Probanden (Codierung im Datensatz: *self*), der Mutter (Codierung im Datensatz: *mother*) und des Vaters (Codierung im Datensatz: *father*). Es ergeben sich insgesamt 8 verschiedene Kombinationen.⁸²

Definiert man Schülerinnen und Schüler mit Migrationshintergrund als diejenigen, deren beide Eltern außerhalb des Ziellandes geboren wurden, unabhängig vom Geburtsland des einzelnen Schülers, was der Kombination 4 und 8 entspricht (siehe Fußnote), finden sich in Luxemburg etwa 30%, Australien, Kanada, Schweiz und Neuseeland ca. 20% und in Deutschland, den USA, Frankreich, Schweden und Österreich zwischen 10% und 15% Schüler mit Migrationshintergrund im Datensatz der PISA-Testteilnehmer.

Praktisch keine Schülerinnen und Schüler mit Migrationshintergrund nach dieser Definition weisen die Datensätze der Länder Finnland, Tschechien, Island, Brasilien, Polen, Japan und Korea auf.

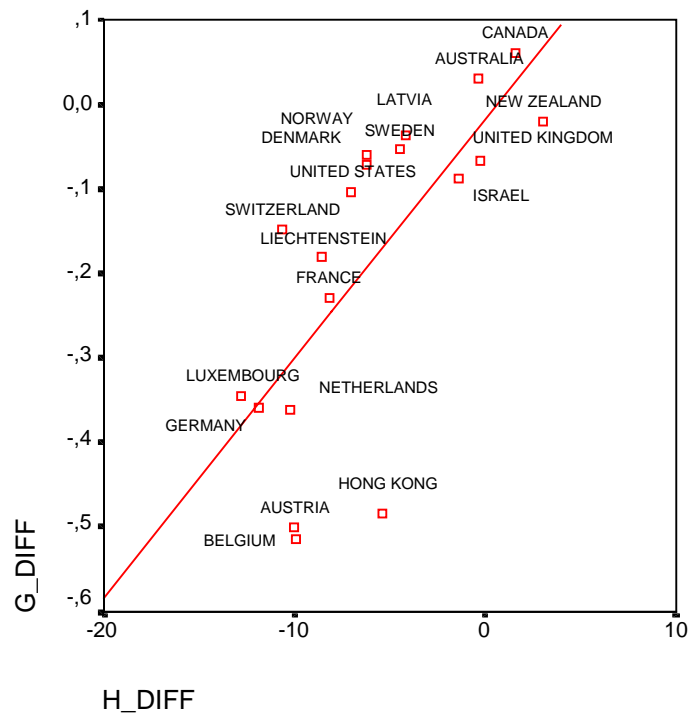
⁷⁹ Der Test umfasste zwei 60-minütige Sitzungen für die Bereiche Reading, Mathematical und Scientific Literacy (PISA-Manual. S. 9)

⁸⁰ Merkens hatte so genannte J-förmige Verteilungen bei lernzielorientierten Tests angesprochen. Merkens: Einführung in die Statistik für Pädagogen 1975. S. 68 ff

⁸¹ Lienert, Raatz: Testaufbau und Testanalyse, 1998, S. 155 ff

⁸² Die Gruppenbildung weist $2^3 = 8$ Kombinationen auf: Geburtsort (2 Ausprägungen: Inland (I)/Ausland (A)) sowie Person (3 Ausprägungen: Proband/Mutter/Vater). Folgende Gruppen werden definiert: **1** 'self (I) - mother (I) - father (I)' **2** 'self (I) - mother (I) - father (A)' **3** 'self (I) - mother (A) - father (I)' **4** 'self (I) - mother (A) - father (A)' **5** 'self (A) - mother (I) - father (I)' **6** 'self (A) - mother (I) - father (A)' **7** 'self (A) - mother (A) - father (I)' **8** 'self (A) - mother (A) - father (A)' (I: Inland, A: Ausland). Deckungsanteil: 94,1% (internationaler Datensatz) 94,3% (nationaler Datensatz) an gesamter Stichprobe.

Migration lässt sich nach sozioökonomischer Differenz zum Zielland unterscheiden. Trägt man die sozioökonomische Differenz der beiden Schülergruppen (Codierung: H_DIFF)⁸³ gegen die jeweilige Schulalterdifferenz (Codierung: G_DIFF)⁸⁴ je Teilnehmerstaat für diejenigen Länder ab, die mit 5% Anteil einen nennenswerten Migrationsanteil aufweisen, zeigt sich über die aggregierten Werte ein hochgradiger linearer Zusammenhang der Differenzen (Berechnung über den erweiterten Datensatz).



Korrelationen

		G_DIFF
H_DIFF	Korrelation nach Pearson	,715**
	Signifikanz (2-seitig)	,001
	N	19

** Die Korrelation ist auf dem Niveau von 0,01 (2-seitig) signifikant.

Korrelationen

		G_DIFF
Spearman-Rho	H_DIFF Korrelationskoeffizient	,781**
	Sign. (2-seitig)	,000
	N	19

** Die Korrelation ist auf dem 0,01 Niveau signifikant (zweiseitig).

Abbildung 42 Sozioökonomische Differenz – Schulalter Differenz/Korrelation

Zu erkennen ist der mit $r_p = 0,715$ beziehungsweise $r_s = 0,781$ ⁸⁵ hochgradige statistische Zusammenhang. Je höher also die sozioökonomische Differenz der Schülerinnen und

⁸³ Codierung orientiert sich am Anfangsbuchstaben des Variablenlabels (**H**ISEI), $H_diff = \bar{x}_{HISEI_{mitMig}} - \bar{x}_{HISEI_{ohneMig}}$

⁸⁴ Codierung orientiert sich am Anfangsbuchstaben des Variablenlabels (**G**RADE), $G_diff = \bar{x}_{GRADE_{mitMig}} - \bar{x}_{GRADE_{ohneMig}}$

⁸⁵ Berechnung des Koeffizienten berücksichtigt hier nicht die variierenden N_i der einzelnen Länder

Schüler mit Migrationshintergrund zu den Schülern des Ziellandes, desto höher deren Rückstand in Bezug auf das Schulalter.

Auffällig ist weiterhin, dass Schweden und Norwegen, die ebenfalls Aufnahmeländer für Flüchtlinge aus Krisenregionen sind und ebenfalls eine mehr oder weniger offene Arbeitszuwanderung zu verzeichnen haben, offensichtlich zu einer besseren Balance zwischen Umgangssprache in der Familie und Beherrschung der Verkehrssprache gelangen. (...) Auffallend ist die insgesamt günstigere Situation in Norwegen und Schweden. Auch wenn die Zuwandererfamilien an ihrer Herkunftssprache festhalten, gelingt es ihnen, einen höheren Sozialstatus als in Deutschland zu erreichen.⁸⁶

Schweden weist geringere Differenzen zwischen dem sozioökonomischen Niveau der Schüler mit und ohne Migrationshintergrund auf als beispielsweise die Schweiz, Österreich, Luxemburg oder die Bundesrepublik Deutschland.

Der vorliegende Befund steht ebenfalls im Gegensatz zu den berichteten Ergebnissen von Schnabel und Schwippert. Die Autoren hatten in Bezug auf das Alter der Schülerinnen und Schüler mit und ohne Migrationshintergrund in Deutschland keinen statistisch signifikant nachweisbaren Unterschied zwischen beiden Gruppen feststellen können.

Wie Tabelle VI.6 zeigt, lassen sich Vermutungen über ein höheres Durchschnittsalter der Schüler aus Familien nicht-deutscher Herkunft nicht belegen. Der Altersabstand zwischen den Jugendlichen aus deutschsprachigen Familien zu denjenigen aus partiell deutschsprachigen Familien ist mit durchschnittlich rund zwei Monaten nicht zufallskritisch abzusichern. Auch der Unterschied zwischen den Jugendlichen deutschsprachiger und nicht-deutschsprachiger Familien ist mit rund einem halben Jahr zwar etwas deutlicher ausgeprägt, aber ebenfalls statistisch nicht signifikant.⁸⁷

Nun wäre die Frage nach der Definition des Begriffes Migration zu stellen, die sich aus der Verkehrssprache ableitet sowie starker oder schwacher Abgrenzung von Vergleichspopulationen (deutschsprachig – partiell deutschsprachig). Besondere

⁸⁶ Baumert, Schümer: Familiäre Lebensverhältnisse, Bildungsbeteiligung und Kompetenzerwerb. In: Deutsches PISA-Konsortium (Hrsg.): PISA 2000: Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich. S. 394 ff

⁸⁷ Schnabel, Schwippert: Einflüsse sozialer und ethnischer Herkunft beim Übergang in die Sekundarstufe II und den Beruf. In: Baumert, Bos, Lehmann (Hrsg.): TIMSS/III Dritte Internationale Mathematik- und Naturwissenschaftsstudie 2000, Band 1. S. 288

Aufmerksamkeit sollte das eingesetzte statistische Verfahren und -angesichts der berichteten Unterschiede von im Durchschnitt eines halben Jahres- der Zusammenhang von Signifikanz, Effektgröße und Stichprobenumfang erfahren.

2.6.1 Chi-Quadrat-Test/Klassenstufe

Die Rohdaten der PISA-Studie bieten nun die Möglichkeit, Unterschiede in Bezug auf das Lebens- und Schulalter zu prüfen. Die PISA-Schülerpopulation bestand zwar lediglich aus der Altersgruppe der 15-Jährigen, jedoch wurde die Klassenstufe variiert. Es ist im Umkehrschluss anzunehmen, dass diejenigen Schülerinnen und Schüler, die im Alter von 15 Jahren niedrigere Klassenstufen besuchen, in ihren Klassen zu den älteren Schülern gehören. Es ist also der Nachweis von Interesse, ob Schüler mit Migrationshintergrund ein signifikant geringeres Schulalter aufweisen.

Der Chi-Quadrat-Unabhängigkeitstest ist angezeigt. Migration wird anhand des Geburtslandes der Eltern unabhängig von der Verkehrssprache in der Familie wie eingangs definiert. Modell: Die Ausgangshypothese H_0 geht von der Unabhängigkeit der beiden Zufallsvariablen *Migrationshintergrund* und *Klassenstufe* aus.

$$H_0 : \pi_{ij} = \pi_{i.} \pi_{.j} \text{ für alle } i = 1, \dots, k \text{ und alle } j = 1, \dots, l$$

$$H_1 : \pi_{ij} \neq \pi_{i.} \pi_{.j} \text{ für mindestens ein Paar } (i,j).$$

Getestet wird über die Prüfgröße

$$\chi^2_{emp} : \sum_{i=1}^k \sum_{j=1}^l \frac{(n_{ij} - \tilde{n}_{ij})^2}{\tilde{n}_{ij}} .$$

Die Frage nach der unter Gültigkeit von H_0 erwarteten Zellbesetzung prüfen die *standardisierten Residuen* mit $u_{ij} \sim N(0;1)$ nach folgendem Modell:

$$u_{ij} = \frac{R_{ij}}{\sqrt{\tilde{n}_{ij}}}, \text{ wobei } R_{ij} = n_{ij} - \tilde{n}_{ij} .$$

Klassenstufe * Migration 2 Gruppen (4,8) Kreuztabelle

		Migration 2 Gruppen (4,8)		Gesamt
		,00	1,00	
Klassenstufe 6	% von Migration 2 Gruppen (4,8)	,0%	,1%	,0%
	Korrigierte Residuen	-3,7	3,7	
7	% von Migration 2 Gruppen (4,8)	,9%	4,5%	1,3%
	Korrigierte Residuen	-19,5	19,5	
8	% von Migration 2 Gruppen (4,8)	12,3%	29,3%	14,4%
	Korrigierte Residuen	-29,8	29,8	
9	% von Migration 2 Gruppen (4,8)	64,9%	54,5%	63,6%
	Korrigierte Residuen	13,3	-13,3	
10	% von Migration 2 Gruppen (4,8)	21,9%	11,6%	20,6%
	Korrigierte Residuen	15,7	-15,7	
11	% von Migration 2 Gruppen (4,8)	,0%	,0%	,0%
	Korrigierte Residuen	1,1	-1,1	
Gesamt		100,0%	100,0%	100,0%

Chi-Quadrat-Tests

	Wert	df	Asymptotische Signifikanz (2-seitig)
Chi-Quadrat nach Pearson	1408,461 ^a	5	,000
Likelihood-Quotient	1168,728	5	,000
Zusammenhang linear-mit-linear	1137,396	1	,000
Anzahl der gültigen Fälle	34533		

a. 2 Zellen (16,7%) haben eine erwartete Häufigkeit kleiner 5. Die minimale erwartete Häufigkeit ist 1,00.

Abbildung 43 Chi-Quadrat-Test/Klassenstufe⁸⁸

Die Hypothese der Unabhängigkeit wird auf dem Niveau $p < 0.01$ verworfen.

Es stellt sich die Frage nach den Ursachen. Der Vergleich der Populationen in Bezug auf Klassenwiederholung zeigt, dass ein Teil dieser Schulalter-Differenz zur Vergleichsgruppe durch überproportionale Klassenwiederholung der Schülerinnen und Schüler mit Migrationshintergrund zu erklären ist. Diese weisen mit etwa 28 Prozent eine höhere Klassenwiederholung auf als die Vergleichsgruppe (nach Definition) mit etwa 16 Prozent.⁸⁹

Das mittlere Schulalter der beiden Vergleichsgruppen im Ergänzungstest derjenigen Schülerinnen und Schüler, die keine Klassenwiederholung berichten, zeigt ebenfalls eine statistisch signifikante Differenz von etwa einem viertel Schuljahr (Klassenstufe). Selbst unter Berücksichtigung von weiteren Faktoren als Kovariaten bleibt diese Größenordnung statistisch nachweisbar. Dieser Zusammenhang lässt sich für eine Reihe von Teilnehmerstaaten in unterschiedlicher Ausprägung nachweisen.⁹⁰

⁸⁸ Codierung Migration: Mit Migrationshintergrund: 1,00 [entsprechend Definition (Gruppe 4,8)]. Ohne Migrationshintergrund 0,00 [entsprechend Definition (Gruppe 1,2,3,5,6,7)].

⁸⁹ Der nationale Datensatz (PISA-OVE-15J) (gewichtet) weist insgesamt 73,14% Schülerinnen und Schüler auf, die keine Klassenwiederholung berichten. Einmal: 16,73%, Zweimal: 0,29%, Unallocated: 1,02%, missing: 1,30%, system: 7,51%, Gültige Prozente ohne Wiederholung: 81,12%, einmal: 18,55%, ein zweites Mal: 0,33%

⁹⁰ Es ergeben sich für die Länder Australien, Canada, Irland und Neuseeland signifikant höhere und für die Länder Österreich, Belgien, Brasilien, Tschechien, Deutschland, Griechenland, Ungarn und UK signifikant niedrigere Schulalter für Schülerinnen und Schüler mit Migrationshintergrund entsprechend Definition.

2.7 Vergleich IGLU/PISA

Prenzel, Geiser, Langeheine und Lobemeier stellen Vergleiche zwischen den einzelnen internationalen Studien an mit der Feststellung, dass einigen Teilnehmerstaaten eine Kompetenzentwicklung vom Ende der Grundschulzeit über die Sekundarstufe gelingt.⁹¹

*Im ersten IGLU-Bericht (Prenzel et al., 2003) war unter anderem hervorgehoben worden, dass Staaten wie die USA oder Deutschland im Grundschulbereich noch relativ gut abschneiden, dann jedoch bei den Sekundarstufenvergleichen deutliche Schwächen erkennen lassen. Demgegenüber lassen sich für andere Staaten Entwicklungen beobachten, die auf eine qualitätsvolle und beschleunigte Kompetenzentwicklung vom Ende der Grundschulzeit über die Sekundarstufe (etwa bis zum Ende der 9. Jahrgangsstufe oder bis zum 15. Lebensjahr) schließen lassen. Anhaltspunkt dafür sind die deutlichen Veränderungen in den Abständen zum deutschen Mittelwert bei IGLU-E und bei PISA. So ändern Staaten wie Kanada, Neuseeland, England oder Japan ihre relative Position zu Deutschland in einer Größenordnung von 40 Punkten oder mehr.*⁹²

Die Ausführungen zur Problematik unterschiedlicher Stichprobendefinitionen haben deutlich gemacht, dass als Ursache der Lokationsverschiebungen der Teilnehmerstaaten auf den Ratingskalen die Schwerpunktsetzung zugunsten des Schulalters beziehungsweise des Lebensalters zu sehen ist. Die von den Autoren angesprochenen Staaten Neuseeland, England (Vereinigtes Königreich) und Japan weisen Schülerpopulationen mit hohem Schulalter auf. Diese Schülerinnen und Schüler werden daher in einer am Lebensalter orientierten Stichprobe höhere Werte erreichen, in einer am Schulalter orientierten Studie hingegen aufgrund des geringen Grenzertrags früher Einschulung niedrigere. Deutlich wird dies an den sehr unterschiedlichen Testwerten Neuseelands, das im PISA-Lesetest mit 529 Punkten den dritten Platz erreichte, in IGLU mit 529 Punkten den 13. Platz. Zu bemerken ist hier, dass Neuseeland in der IGLU-Studie seine Stichprobe in der 5. Klassenstufe zog, während in den meisten übrigen

⁹¹ Prenzel, Geiser, Langeheine, Lobemeier: Naturwissenschaftliche Kompetenz am Ende der Grundschulzeit: Vergleiche zwischen einigen Ländern der Bundesrepublik Deutschland. In: Bos, Lankes, Prenzel, Schwippert, Valtin, Walther (Hrsg.): IGLU Einige Länder der Bundesrepublik Deutschland im nationalen und internationalen Vergleich (II) 2004. S. 104ff

⁹² Prenzel, Geiser, Langeheine, Lobemeier: Naturwissenschaftliche Kompetenz am Ende der Grundschulzeit: Vergleiche zwischen einigen Ländern der Bundesrepublik Deutschland. In: Bos, Lankes, Prenzel, Schwippert, Valtin, Walther (Hrsg.): IGLU Einige Länder der Bundesrepublik Deutschland im nationalen und internationalen Vergleich (II) 2004. S. 104ff

Teilnehmerstaaten die Schülerinnen und Schüler der 4. Klassenstufe bewertet wurden. Die Frage stellt sich demnach, ob die einzelnen Studien verglichen werden können.

*Für diese Staaten und Länder können die in den beiden Studien erzielten Ergebnisse zueinander in Beziehung gesetzt werden.*⁹³

Werden die Ergebnisse der einzelnen Studien im Rahmen einer Korrelationsanalyse in Beziehung gesetzt, so liegt mit $r_p = 0,193/p = 0,389$ beziehungsweise $r_s = 0,173/p = 0,44$ kein statistisch signifikanter Zusammenhang vor, d.h. die Ausgangshypothese $H_0 : \rho = 0$ kann nicht verworfen werden.⁹⁴

⁹³Walther, Geiser, Langeheine, Lobemeier: Mathematische Kompetenzen am Ende der vierten Jahrgangsstufe in einigen Ländern der Bundesrepublik Deutschland. In: Bos, Lankes, Prenzel, Schwippert, Valtin, Walther (Hrsg.): IGLU Einige Länder der Bundesrepublik Deutschland im nationalen und internationalen Vergleich (II) 2004. S. 128

⁹⁴ Korrelationsrechnung über die einzelnen Teilnehmerstaaten ohne Berücksichtigung der unterschiedlichen N_i .