

Notation

x_i	Meßwert
\bar{x}	arithmetischer Mittelwert der Meßgröße
σ_x	Standardabweichung der Messung
μ	Lageparameter
$n(x)$	Häufigkeit des Merkmals x
H_0	Ausgangshypothese
H_1	Gegenhypothese
p	Fehler erster Ordnung (Signifikanz)
α_i	unbekannter Effekt des i-ten Treatments
π	Testanteil
π_i	$P(A_i)$, Wahrscheinlichkeit für Ausprägung A_i
π_{ij}	$P(A_i, B_j)$, Wahrscheinlichkeit für Ausprägungspaar (A_i, B_j)
F, G	Verteilungsfunktionen F, G
$E(X)$	Erwartungswert
$Var(X)$	Varianz von X, (σ_x^2)
$x_{0,25}, x_{0,5},$	
$x_{0,75}$	Quartile
β_i	Beta-Koeffizient im regressionsanalytischen Modell
r_p	Korrelationskoeffizient nach Pearson
r_s	Korrelationskoeffizient nach Spearman
ρ	Korrelation (rho)
\hat{y}	regressionsanalytisch bestimmter Funktionswert
K, N	Zahl der Beobachtungswerte, Stichprobenumfang
J	Zahl der Regressoren
R^2	Bestimmtheitsmaß
x_{\min}	kleinster Wert eines Merkmals
x_{\max}	grösster Wert eines Merkmals
$R(x_i)$	Rang des i-ten Meßwertes von X
d_i^2	Quadrat der i-ten Rangplatzdifferenz $d_i = R(x_i) - R(y_i)$
z_{Fisher}	transformierter Wert von r_p als Fisher-Z-Wert
Z	standardnormalverteilte ZV mit $Z \sim N(0,1)$
ζ	Erwartungswert der Fisher-z-transformierten Zufallsvariable ρ für kleine Stichproben
$\frac{\mu_3}{\sigma^3}$	Schiefe der Verteilung

Δ	Differenz
df	Freiheitsgrade (Degree of Freedom)
χ_{emp}^2	Wert des empirisch ermittelten Chi-Quadrat Wertes
$\chi_{df;1-\alpha}^2$	theoretischer Wert der durch Anzahl der Freiheitsgrade df definierten Chi-Quadrat Verteilung zum Niveau $1 - \alpha$
\tilde{n}_{ij}	erwartete Zellbesetzung in Zeile i, Spalte j
n_{ij}	Zellbesetzung in Zeile i und Spalte j
$t_{df;1-\alpha}$	Wert der durch Anzahl der Freiheitsgrade (df) bestimmten t-Verteilung zum Niveau $1 - \alpha$
u_{ij}	standardisiertes Residuum
t_0	Zeitpunkt 0, Ausgangszeit
t_1	Zeitpunkt 1
L	Levene Statistik
W_k	Summe der Fallgewichtungen in Gruppe k
n_k	Anzahl der Fälle in Gruppe k
x_{ki}	Wert des i-ten Falls in Gruppe k
w_i	Gewichtung von Fall i
w_{ki}	Gewichtung des i-ten Falls in Gruppe k
W	Summe der Gewichte

Abkürzungen

ACER	Australian Council for Educational Research
BC	Boston College, Lynch School of Education
BPC	Board of Participating Countries
IEA	International Association for the Evaluation of Educational Achievement
IGLU	Internationale Grundschul-Lese-Untersuchung
INES	Indicators of Education Systems
KMK	Kultusministerkonferenz der Länder
MPIB	Max-Planck-Institut für Bildungsforschung, Berlin
OECD	Organisation for Economic Co-operation and Development
PISA	Programme for International Students Assessment
TIMSS	Third International Mathematics and Science Study
WESTAT	Research Corporation, Rockville, Maryland

Variablen

AL_ERZ	Anteil Alleinerziehende
ANT_GYM	Gymnasialbesuch anteilig
B_GANG	Bildungsgang
BIP_JEWP	Bruttoinlandsprodukt zu jeweiligen Preisen
BIP_KONP	Bruttoinlandsprodukt zu konstanten Preisen
BSP_JEWP	Bruttosozialprodukt zu jeweiligen Preisen
BSP_KONP	Bruttosozialprodukt zu konstanten Preisen
CENTRAL	Anteil zentral geprüfter Schüler (PISA 2003)
COUNTRY – ISO CODE	Teilnehmerstaat
DIFF	Differenz der Mittelwerte (vgl. <i>Disparität</i>)
DIFF_9	Differenz der Mittelwerte Klassenstufe 9 (vgl. <i>Disparität</i>)
EXTRA	Ergänzungsunterricht in Prozent der Nennungen
G_DIFF	Differenz der durchschnittlichen Klassenstufen zwischen Schülern mit und ohne Migrationshintergrund
GRADE – Q2	Klassenstufe
GYM_Z	Durchschnittlicher relativer höchster sozioökonomischer Index der Gymnasiasten eines Bundeslandes, z-standardisiert
H_DIFF	Differenz der Mittelwerte des sozioökonomischen Index zwischen Schülern mit und ohne Migrationshintergrund
HISEI	Höchster sozioökonomischer Index (internationaler Datensatz)
ID-Stud	Identitätsnummer des Probanden (TIMSS)
ISEI	International Socio-Economic Index

ISEI_F	Höchster SES in der Familie
KL_GROES	Durchschnittliche Anzahl der Schüler pro Klasse
MEAN_TIM	Landesmittel mathematisch-naturwissenschaftliches Grundbildungsniveau in der TIMSS
Migration 2 Gruppen (4,8)	Migrationshintergrund Gruppen 4,8 (<i>MIG/Non_MIG 2 Gruppen</i>)
MIXED	Anteil <i>MIXED FAMILY</i> (neue Partnerschaften)
N_LESER	<i>NICHTLESER</i>
NUCLEAR	Anteil Kernfamilie
PV(1)READ	Lesetestwert PISA2000 International (Plausible value in reading) Wert (1) des mehrdimensionalen Rasch-Modells
PV10READ	Lesetestwert PISA2000 International (Plausible value in reading) Wert (1) des mehrdimensionalen Rasch-Modells, Auswahl 10. Klassenstufe
QUANT1	Arithmetisches Mittel der Testwerte über den Bereich $x_{\min} - x_{0,25}$ einer Verteilung
RETENT	Retentivität/Rückhaltequote
SCHIEFE, SKEWNESS	Schiefe der Verteilung der Testwerte eines Teilnehmerstaates
SES	Socio-Economic Status
SINGLE	Anteil Alleinerziehende
STD_G_N	Standardabweichung über Variable GRADE (Klassenstufe)
STD_G_W	Standardabweichung über Variable GRADE (Klassenstufe), Fälle gewichtet über Variable W_FSTUWT (Fallgewichtung)
TCI	TIMSS Coverage Index
VOLKEINK	Volkseinkommen
W_FSTUWT	Gewichtungsvariable

Hinweise

Die vorliegende Evaluation basiert auf den Public-Use Files als Download vom Server der OECD und der KMK. In Hinblick auf das im Test eingesetzte probabilistische Modell und der daraus resultierenden Probleme bezüglich der Auswertung mit in der Wissenschaft genutzten Standardverfahren und Programmen folgt die Evaluation den Empfehlungen des *Technical Report*, der Dokumentation der Erhebungsinstrumente, dem *Manual for the PISA Database*, der Vorgehensweise der PISA-Studie sowie der Rücksprache mit C. Artelt (MPI Berlin) und R. Adams (ACER Australien).

Auf die ausführliche Darstellung der statistischen Modelle und Verfahren wurde der Übersichtlichkeit und leichten Lesbarkeit wegen nur jeweils einmal eingegangen.

Als *t-Test bei Ungleichheit der Varianzen* wird der Welch-Test eingesetzt.

Die Codierungen der Variablen wurden unverändert aus den Public-Use-Files der Studie übernommen.

Abbildungen wurden unter Angabe der Quelle ohne Berücksichtigung eventueller rechtlicher Beschränkungen in dieser Arbeit veröffentlicht. Die mangelhafte graphische Qualität einzelner Abbildungen folgt aus dem Kopierverfahren, insofern keine Alternative zur Verfügung stand.

Die Auswahl der in der vorliegenden Evaluation eingesetzten statistischen Verfahren orientiert sich an der PISA-Studie sowie weiteren thematisch verbundenen Vergleichsuntersuchungen (IGLU, TIMSS). Sämtliche nicht gesondert gekennzeichnete Tabellen und Abbildungen basieren auf eigenen Berechnungen.

Zum mathematischen Hintergrund der Berechnung geschätzter Randmittel siehe: SPSS Inc. Chicago: SPSS Statistical Algorithms GLM UNIVARIATE AND MULTI-VARIATE S.5 ff (<http://support.spss.com/tech/default.asp>).

Zur Messung des sozioökonomischen Index siehe: *PISA 2000 Dokumentation der Erhebungsinstrumente*, S.226.

1 Einleitung

1.1 Gegenstand der Evaluation

Die Kernthesen der PISA 2000-Studie, sowohl des internationalen Vergleichs als auch der nationalen Ergänzungsstudie, werden anhand der Public Use Files der TIMSS, PISA-I und PISA-E evaluiert. Der Fokus der vorliegenden Arbeit liegt ausschließlich auf den Statistiken, Verfahren und Methoden in Hinblick auf die zentralen Thesen. Der Fokus ist bewusst eng gewählt und beschränkt sich allein auf die Frage, ob mit dem erhobenen Datenmaterial die Kernthesen der beiden PISA-Studien zu belegen sind.

Die Kernthesen betreffen

- a. die geringe Performanz der Schülerinnen und Schüler der Bundesrepublik Deutschland,
- b. deren hohe soziale Disparitäten im Lesen,
- c. die über die Bundesländer systematisch differierenden Testergebnisse,
- d. die so genannte Diagnosefähigkeit der Lehrkräfte,
- e. den Zusammenhang von sozioökonomischem Hintergrund der Familie und Testleistung in der PISA-Studie im Vergleich der Schüler des integrierten und des gegliederten Systems,
- f. den rückläufigen Anteil der Kernfamilie an der Gesamtbevölkerung der Bundesrepublik Deutschland und dessen Bedeutung für die gemessenen Testergebnisse,
- g. den positiven Vorbildcharakter der Schulsysteme der Teilnehmerstaaten Japan und Korea,
- h. die einzelnen Studien (TIMSS, PISA und IGLU) als längsschnittanalytischer Vergleich zur Messung von Leistungsfortschritten,
- i. die Entwicklung einer international sinnvollen Stichprobendefinition.

1.2 Studie

PISA steht für *Programme for International Student Assessment*. Die Studie ist Teil des Indikatorenprogramms INES der Organisation für wirtschaftliche Zusammenarbeit und Entwicklung (OECD). Die OECD startete die PISA-Studien im Jahr 1997.

PISA represents a desire by governments to monitor the outcomes of education systems in terms of student achievement on a regular basis and within an internationally accepted common framework. An international consortium, led by the Australian Council for Educational Research (ACER), manages the design and implementation of PISA. Other consortium partners are the National Institute for Educational Measurement (CITO) in the Netherlands, Westat and the Educational Testing Service (ETS) in the United States, and the National Institute for Educational Policy Research (NIER) in Japan. ¹

1.2.1 Untersuchungsgegenstand und Durchführung der PISA-Studie

Im Mittelpunkt des Interesses stehen in PISA so genannte *Basiskompetenzen*, die in modernen Staaten für eine Teilhabe am gesellschaftlichen, wirtschaftlichen und politischen Leben als notwendig erachtet werden. Neben der Frage, inwieweit Jugendliche diese Kompetenzen erworben haben und soziale Ungleichheiten im Bildungserfolg bestehen, interessieren Aspekte schulischer und außerschulischer Lern- und Lebensbedingungen.

1.2.2 Ziele von PISA

PISA ist ein langfristig angelegtes Projekt in drei Erhebungszyklen, in denen drei inhaltlich abgegrenzte Bereiche, so genannte Kompetenzbereiche, *Lesekompetenz* (reading literacy), *mathematische Grundbildung* (mathematical literacy) und *naturwissenschaftliche Grundbildung* (scientific literacy) mit wechselndem Schwerpunkt untersucht werden.

Der erste Zyklus im Jahr 2000 umfasst die *Lesekompetenz* als Schwerpunkt, der zweite Zyklus (Erhebung im Jahr 2003) die *mathematische* und der dritte Zyklus (Erhebung im Jahr 2006) die *naturwissenschaftliche Grundbildung*.

¹ OECD (Hrsg.) Thomson, Cresswell, De Bortoli: Facing the Future PISA 2003 S.27

1.2.3 Testkonzeption

PISA zielt nicht auf reproduzierendes Faktenwissen ab. Mit der Erhebung soll vielmehr untersucht werden, inwieweit die Jugendlichen in der Lage sind, ihre Kenntnisse und Fähigkeiten in realistischen Situationen anzuwenden und zur Bewältigung von Alltagsproblemen zu nutzen. Der Test zielt auf ein tieferes Verständnis für zentrale Konzepte, indem er aufzeigen soll, ob Schüler Prozesse wie das *Modellieren von Situationen* beherrschen, das *Kommunizieren von Ergebnissen* oder das kritische Beurteilen von Informationen verstehen und ob sie in der Lage sind, *Konzept- und Prozesswissen* in unterschiedlichen Kontexten anzuwenden.

Die Rahmenkonzeption wurde in enger Zusammenarbeit zwischen internationalen und nationalen Expertengruppen entwickelt. Im Lesetest wurde eine breite Palette verschiedener Arten von Texten eingesetzt, die neben kontinuierlichen Texten wie Erzählungen, Beschreibungen oder Anweisungen auch nichtkontinuierliches Material wie Tabellen, Diagramme oder Formulare umfasst.

Mathematische Grundbildung bedeutet die Fähigkeit, mathematische Begriffe als Werkzeuge in einer Vielfalt von Kontexten einzusetzen. Naturwissenschaftliche Grundbildung wird als Verständnis grundlegender naturwissenschaftlicher Konzepte, wie etwa Energieerhalt, *Vertrautheit* mit naturwissenschaftlichen Denk- und Arbeitsweisen sowie als Fähigkeit, dieses *Konzept- und Prozesswissen* vor allem bei der Beurteilung naturwissenschaftlich-technischer Sachverhalte anzuwenden, verstanden.

PISA besteht aus einer Mischung von Multiple-Choice Aufgaben und Aufgaben mit offenem Format. Insgesamt umfasst der Test Items für eine Gesamtdauer von sieben Stunden, wovon die Schüler jeweils unterschiedliche Kombinationen in einer zweistündigen Testsitzung bearbeiten. Die Testleistungen sämtlicher Schüler wurden auf eine Skala mit $\mu = 500$ und $\sigma = 100$ standardisiert.

Hintergrundmerkmale von Schülern zur sozialen Herkunft, Aspekte der Beziehung der Jugendlichen zu ihren Eltern, Einstellungen der Schüler zum Lesen sowie ihre privaten Lesegewohnheiten wurden zusätzlich erhoben. Den Schulen wurden Fragen zur finanziellen und personellen Ausstattung, Größe von Lerngruppen, Organisationsstrukturen und Entscheidungsprozessen gestellt. Eine Befragung der Eltern diente schließlich zur Verifikation der Schülerangaben.

1.2.4 Projekt

Weltweit nahmen im Frühsommer 2000 rund 180.000 Jugendliche aus 28 OECD-Mitgliedsstaaten sowie aus Brasilien, Lettland, Liechtenstein und der Russischen Föderation an der Erhebung teil. In Deutschland bestand die Stichprobe für den internationalen Vergleich aus 5.073 Schülern in 219 Schulen, wobei im Durchschnitt 23 15-Jährige pro Schule getestet wurden.

Auf Beschluss der Kultusminister der Länder in der Bundesrepublik Deutschland wurde die PISA-Stichprobe zusätzlich erweitert, so dass ein Vergleich der Länder möglich wurde. Insgesamt wurden 45.899 Schülerinnen und Schüler (zwei überlappende Stichproben von 33.809 15-Jährigen und 33.766 Schülern der 9. Klasse) in 1.466 Schulen untersucht. Die Tests fanden von Mai bis Juni 2000 in den Schulen statt.²

Die Beteiligungsquoten der ausgewählten Schulen wurden vorab festgelegt, so dass in jedem Teilnehmerstaat mindestens 85 Prozent der für die Stichprobe gezogenen Schulen und mindestens 80 Prozent der ausgewählten Schüler an der Untersuchung teilnehmen mussten. Die Niederlande verfehlten diese Mindestbeteiligungsquoten und wurden vom internationalen Vergleich ausgeschlossen. In Deutschland wurde auf Schülerebene eine Teilnahmequote von 86 Prozent erreicht.

Die internationalen Kriterien für die Beteiligungsquoten wurden auch auf den nationalen Ländervergleich [PISA-E] übertragen. Diese Mindeststandards wurden in Berlin und Hamburg deutlich unterschritten. Ein Vergleich der Ergebnisse an Gymnasien war jedoch möglich.

1.2.5 Koordination

PISA versteht sich als kooperatives Unternehmen, das wissenschaftliche Expertise aus allen Teilnehmerstaaten zusammenführt und von den Regierungen der Teilnehmerstaaten gelenkt wird. Die wesentlichen Entscheidungen trifft das Board of Participating Countries (BPC) der OECD. Mit der Planung und wissenschaftlichen Koordination der Studie wurde vom BPC ein internationales Konsortium unter Federführung des Australian Council for Educational Research (ACER) beauftragt. In Deutschland wurde die PISA-Studie von der Kultusministerkonferenz in Auftrag gegeben. Verantwortlich für ihre Durchführung und Erweiterung war ein nationales Konsortium unter der Leitung des Max-Planck-Instituts für Bildungsforschung (MPIB) in Berlin.

² Angaben des MPIB. Die Stichprobenumfänge der Public-Use-Files weichen geringfügig ab: 34.561 15-Jährige, 31.751 Schüler der 9. Klasse.