

**Wie solide ist PISA?
oder
Ist die Kritik von Joachim Wuttke begründet?**

PISA, das „Programme for International Student Assessment“, findet in Deutschland große öffentliche Aufmerksamkeit. Erkenntnisse aus PISA werden bei bildungspolitischen Entscheidungen berücksichtigt. Die Institute, die PISA international und national durchführen und auswerten, stellen sich dem Anspruch, belastbare Ergebnisse zu berichten, die dem Stand der neuesten Forschung und höchsten methodischen Ansprüchen genügen. Wie für alle wissenschaftlichen Unternehmungen gilt auch für PISA, dass der Untersuchungsansatz und die Methoden weiter entwickelt werden können. Deshalb freuen sich die Institute über Anregungen, Vorschläge und Kritik, die zu einer Optimierung von PISA beitragen.

Kritische Stellungnahmen zu PISA wurden in Deutschland vor kurzem in einem von Thomas Jahnke herausgegebenen Band veröffentlicht („PISA & Co“, Verlag Franzbecker). Neben bekannten kulturkritischen Reflexionen einer internationalen Testbewegung enthält der Band einen Beitrag, der sich mit statistischen Fragen befasst. Dieses Kapitel thematisiert „Fehler, Verzerrungen und Unsicherheiten in der PISA-Auswertung“. Der Verfasser des Beitrag, Joachim Wuttke, formuliert dort eine grundlegende Kritik. Die auf internationaler Ebene für die Datenaufbereitung Verantwortlichen (der Australian Council for Educational Research) seien „spektakulär gescheitert“ (S. 146). Diese Aussage begründet Joachim Wuttke mit eigenen Berechnungen zur Skalierung des PISA-Mathematiktests (S. 119-121).

Die Berechnungen von Joachim Wuttke, mit denen er einen Programmierfehler und fehlende Kompetenz des internationalen Projektleiters nachweisen möchte, sind falsch. Wuttke kommt zu inkonsistenten Ergebnissen, weil die von ihm verwendete Gleichung (Gleichung 4 auf S. 121) einen Fehler enthält. Die Gleichung berücksichtigt nicht die bei der Skalierung errechnete Standardabweichung. Setzt man in diese Gleichung den korrekten Wert ein, stimmt das Modell mit den Daten überein (vgl. Anlage). Das international zur Skalierung verwendete Programm arbeitet modellgerecht; die Berechnungen sind korrekt. Sie machen deshalb auch nicht die von Wuttke behauptete nachträgliche Umskalierung erforderlich.

Neben diesem Hauptvorwurf enthält das Kapitel eine Reihe von Kritikpunkten oder Ratschlägen, die nicht hilfreich sind. So behauptet Wuttke (S. 119, 123, 128), die offiziellen Auswertungen seien „nicht iterativ“. Tatsächlich sind die bei PISA verwendeten Schätzverfahren, wie in der Literatur dargelegt und im Auswertungsprogramm ConQuest umgesetzt, iterativ. Das von Wuttke seinerseits vorgeschlagene alternative Schätzverfahren (S. 119) ist in der Literatur seit langem bekannt und wird aufgrund seiner bekannten suboptimalen Eigenschaften bei PISA nicht angewendet (z. B. Fischer & Molenaar, 1995).

Bemerkenswert sind schließlich die Hochrechnungen von Joachim Wuttke, wie die Testergebnisse bei unterschiedlichen Testzeiten ausgefallen wären. Würde man die international vorgegebene Testzeit von 120 Minuten bei dem vorhandenen Aufgabenmaterial verkürzen, ergäben sich selbstverständlich andere Leistungsverteilungen. Auch leistungsstarke Schülerinnen und Schüler hätten dann keine Möglichkeit, anspruchsvollere Aufgaben erfolgreich zu bearbeiten.

Insgesamt zeigen diese Beispiele, dass die von Joachim Wuttke verfasste Fundamentalkritik falsch ist. Wir bedauern es sehr, dass eine Gelegenheit verpasst wurde, in einem wissenschaftlichen Dialog über Möglichkeiten und Grenzen von PISA zu sprechen.

Prof. Dr. Manfred Prenzel
Sprecher der deutschen Konsortien für
PISA 2003 und PISA 2006

Leibniz-Institut für die Pädagogik der
Naturwissenschaften (IPN) an der Universität Kiel
Olshausenstr. 62
24098 Kiel

E-Mail: prenzelm@ipn.uni-kiel.de

Anlage

Ein Programmierfehler in PISA? Joachim Wuttke hat falsch gerechnet!

Joachim Wuttke wirft dem internationalen PISA-Kon-sortium vor, die Schwierigkeiten der Mathematikaufgaben aufgrund eines Programmierfehlers im Auswertungsprogramm (ConQuest) falsch berechnet zu haben (Wuttke, 2006, S. 122). Er stützt seine Behauptung darauf, dass die Häufigkeiten, mit denen Schülerinnen und Schüler die Aufgaben lösen, nicht auf einer Kurve liegen, die mit einer bestimmten Gleichung

$$p(\xi) = \int \frac{1}{1 + \exp(-(\theta - \xi) / 100 - \ln(62/38))} h(\theta) d\theta$$

(Wuttke, 2006, Gleichung (4), S. 121) berechnet wird (siehe Abbildung unten und Abbildung 4 in Wuttke, 2006, S. 120).

Allerdings ist Wuttke in dieser Gleichung ein Fehler unterlaufen. Er hat, wie schon in seiner Gleichung (1) (S. 118), fehlerhaft den Wert 100 eingesetzt. Mit der Verwendung des Wertes 100 nimmt Wuttke an, dass die Skala, in der das Auswertungsprogramm rechnet, eine Standardabweichung von 1 hat (siehe Wuttke, 2006, S. 118, Fußnote 9). Der Wert 100 steht für den Umrechnungsfaktor von der Skala, mit dem das Auswertungsprogramm arbeitet, in die Punkte-Skala von PISA. In PISA 2003 beträgt die Standardabweichung der Skala, in der Auswertungsprogramm arbeitet, für Mathematik jedoch nicht 1, sondern 1,2838. Statt der 100 muss ein Wert von $100/1,2838 = 77,89$ in die Gleichung eingesetzt werden. Dann beschreibt die Gleichung die Lösungshäufigkeiten richtig (siehe Abbildung unten). Das Auswertungsprogramm von PISA hat also richtig gerechnet und Wuttke einen falschen Umrechnungsfaktor verwendet.

Der Fehler wäre zu vermeiden gewesen: Da Herr Wuttke offensichtlich die Seiten 412/413 des Technischen Berichts der OECD zu PISA 2003 inspiziert hat – er zitiert sie auf S. 120 – hätte er seinen Irrtum bemerken müssen. Zudem ist auf S. 215 in Gleichung (3.6) des Technischen Berichts die Umrechnungsformel von der Skala, mit der das Auswertungsprogramm arbeitet, in die Punkte-Skala von PISA 2003 explizit für Mathematik angegeben.

Prof. Dr. Manfred Prenzel / Dr. Oliver Walter
Leibniz-Institut für die Pädagogik der
Naturwissenschaften (IPN)
Olshausenstr. 62
24098 Kiel

