

Moser, Urs

**Karl-Heinz Arnold: Fairneß bei Schulsystemvergleichen. Diagnostische Konsequenzen von Schulleistungsstudien für die unterrichtliche Leistungsbewertung und binnenschulische Evaluation. (Pädagogische Psychologie und Entwicklungspsychologie. Bd. 9.) Münster: Waxmann 1999.**

**[Rezension]**

*Zeitschrift für Pädagogik 46 (2000) 6, S. 942-944*



Empfohlene Zitierung/ Suggested Citation:

Moser, Urs: Karl-Heinz Arnold: Fairneß bei Schulsystemvergleichen. Diagnostische Konsequenzen von Schulleistungsstudien für die unterrichtliche Leistungsbewertung und binnenschulische Evaluation. (Pädagogische Psychologie und Entwicklungspsychologie. Bd. 9.) Münster: Waxmann 1999. [Rezension]  
- In: Zeitschrift für Pädagogik 46 (2000) 6, S. 942-944 - URN: urn:nbn:de:01111-pedocs-112508

in Kooperation mit / in cooperation with:

**BELTZ JUVENTA**

<http://www.juventa.de>

#### Nutzungsbedingungen

Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Die Nutzung stellt keine Übertragung des Eigentumsrechts an diesem Dokument dar und gilt vorbehaltlich der folgenden Einschränkungen: Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen. Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

#### Terms of use

We grant a non-exclusive, non-transferable, individual and limited right to using this document.  
This document is solely intended for your personal, non-commercial use. Use of this document does not include any transfer of property rights and it is conditional to the following limitations: All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

#### Kontakt / Contact:

peDOCS  
Deutsches Institut für Internationale Pädagogische Forschung (DIPF)  
Informationszentrum (IZ) Bildung  
E-Mail: [pedocs@dipf.de](mailto:pedocs@dipf.de)  
Internet: [www.pedocs.de](http://www.pedocs.de)

Digitalisiert

# Zeitschrift für Pädagogik

Jahrgang 46 – Heft 6 – November/Dezember 2000

## *Essay*

- 795 ROLAND REICHENBACH  
„Es gibt Dinge, über die man sich einigen kann, und wichtige Dinge.“  
Zur pädagogischen Bedeutung des Dissenses

## *Thema: Qualitätsdiskussion im Bildungsbereich*

- 809 EWALD TERHART  
Qualität und Qualitätssicherung im Schulsystem.  
Hintergründe – Konzepte – Probleme
- 831 WERNER NÜSSLE  
Qualität für wen? Zur Angemessenheit des Kundenbegriffs in  
der Sozialen Arbeit
- 851 JOCHEN WISSINGER  
Rolle und Aufgaben der Schulleitung bei der Qualitätssicherung  
und -entwicklung von Schulen

## *Diskussion*

- 867 ELIYAHU ROSENOW  
„Nietzsche als Erzieher“ kontra „Nietzsche in der Pädagogik?“  
Ein Vergleich der anglo-amerikanischen und der deutschen Nietzsche-  
Interpretationen am Vorabend des 21. Jahrhunderts
- 881 IRINA MCHITARJAN  
John Dewey und die pädagogische Entwicklung in Rußland vor 1930.  
Bericht über eine vergessene Rezeption
- 905 HEINZ-ELMAR TENORTH  
„Vom Menschen“ – Historische, pädagogische und andere  
Perspektiven einer „Anthropologie“ der Erziehung.  
Eine Sammelbesprechung neuerer Literatur

## *Besprechungen*

- 927 ROLAND REICHENBACH  
*Anton Hügli: Philosophie und Pädagogik*
- 931 HENNING RÖHR  
*Norbert Ricken: Subjektivität und Kontingenz.  
Markierungen im pädagogischen Diskurs*
- 934 WOLFGANG GRÄBER  
*Werner Kutschmann: Naturwissenschaft und Bildung.  
Der Streit der „Zwei Kulturen“*
- 937 PETER DUDEK  
*Wolfgang Schröer: Sozialpädagogik und die soziale Frage.  
Der Mensch im Zeitalter des Kapitalismus um 1900*
- 939 HANNS PETILLON  
*Georg Stöckli: Eltern, Kinder und das andere Geschlecht.  
Selbstwerdung in sozialen Beziehungen*
- 942 URS MOSER  
*Karl-Heinz Arnold: Fairneß bei Schulsystemvergleichen. Diagnostische  
Konsequenzen von Schulleistungsstudien für die unterrichtliche  
Leistungsbewertung und binnenschulische Evaluation*

## *Dokumentation*

- 945 Pädagogische Neuerscheinungen

## *Content*

### *Essay*

- 795 ROLAND REICHENBACH  
“There are things that are open to discussion and important things.”  
On the pedagogical significance of disagreement

### *Topic: The debate on quality in the educational sector*

- 809 EWALD TERHART  
Quality and Quality Maintenance within the School System:  
Background – Concepts – Problems
- 831 WERNER NÜSSLE  
Quality for Whom? On the adequacy of the concept of the customer  
in social work
- 851 JOCHEN WISSINGER  
Roles and Tasks of Principals in the Maintenance and Development  
of the Quality of Schools

### *Discussion*

- 867 ELIYAHU ROSENOW  
“Nietzsche as Educator” versus “Nietzsche in Pedagogics?”  
A comparison of Anglo-American and German readings of Nietzsche  
at the dawning of the 21st century
- 881 IRINA MCHITARJAN  
John Dewey and the Pedagogical Development in Russia before 1930.  
A report on a forgotten reception
- 905 HEINZ-ELMAR TENORTH  
“Of Man” – Historical, Pedagogical and other Perspectives of an “An-  
thropology” of Education.  
A comprehensive review of recent literature
- 927 BOOK REVIEWS
- 945 NEW BOOKS

Buch einen wichtigen Beitrag zur Etablierung einer in der Entwicklung begriffenen Beziehungswissenschaft.

Prof. Dr. HANNS PETILLON  
Adolf-Kessler-Str. 48, 76829 Landau

**Karl-Heinz Arnold:** *Fairneß bei Schulsystemvergleichen*. Diagnostische Konsequenzen von Schulleistungsstudien für die unterrichtliche Leistungsbewertung und binnenschulische Evaluation. (Pädagogische Psychologie und Entwicklungspsychologie. Bd. 9.) Münster: Waxmann 1999. 214 S., DM 49,90.

Das Interesse an internationalen Vergleichen von Schulleistungen ist in den letzten Jahren stark gestiegen. Ergebnisse von Leistungsuntersuchungen werden sowohl von der breiten Öffentlichkeit als auch von der Bildungspolitik vermehrt beachtet. Die bildungspolitische Bedeutung solcher Ergebnisse drückt sich nicht zuletzt im großen Engagement der OECD für internationale Schulleistungsvergleiche aus. Die Publikationen der Ergebnisse der „Third International Mathematics and Science Study (TIMSS)“ haben im deutschen Sprachraum ein für Themen der Bildungsforschung selten großes Echo ausgelöst. Die Leistungsmessung als ein Instrument der Evaluation im Bildungswesen hat zwar an Aktualität gewonnen, ihre methodisch korrekte Anwendbarkeit und die Aussagekraft ihrer Ergebnisse werden jedoch entsprechend häufig in Frage gestellt. Wer allerdings aufgrund der Fragestellung der Untersuchung erwartet, daß die längst bekannten pauschalen Kritikpunkte gegenüber internationalen Schulleistungsvergleichen (z.B. H. BRÜGELMANN: *Was leisten unsere Schulen? Zur Qualität und Evaluation von Unterricht*. Seelze-Velber 1999) erneut aufgerollt werden, liegt zum Glück falsch. In dieser Untersuchung werden TIMSS und andere Schulleistungsstudien nach wissen-

schaftlichen Kriterien beurteilt und gewürdigt.

ARNOLD geht von der bemerkenswerten Konstellation aus, daß die „mit hohem wissenschaftlichen Aufwand entwickelten und durchgeführten Schulleistungsstudien entweder als fragwürdig eingeschätzt werden bzw. deren Ergebnisverwendung im Rahmen von Bildungspolitik als ‚unfair‘ kritisiert wird“, während „die unterrichtliche Praxis der Ziffernbewertung im Zustand allgemeiner öffentlicher Duldung und Fortführung“ verbleibt, d.h. eine Übertragung der gewonnenen kritischen Positionen auf diese unterrichtliche Beurteilungspraxis kaum stattfindet (S. 5). Eines der zentralen Anliegen seiner Untersuchung richtet sich darauf, die einseitige Kritik an internationalen Schulleistungsvergleichen zu relativieren und in einem größeren pädagogischen Kontext zu betrachten. Die internationalen Schulleistungsstudien sollen deshalb auf die intendierten und praktischen Formen von Fairneß hin geprüft werden, gleiches soll aber auch in Übertragung der Untersuchungsergebnisse für die Praxis unterrichtlicher Leistungsbewertung und binnenschulischer Evaluation geschehen.

Die Fairneßprüfung bei Schulsystemvergleichen geschieht anhand exemplarisch ausgewählter Modelle, wobei Bezüge zwischen den Ansätzen der Leistungsmessung auf den Ebenen von nationalen Schulsystemen, Einzelschulen und Individuen hergestellt werden. Entsprechend den drei Ebenen, werden zu Beginn des Buchs unfaire Vergleiche durch Beispiele illustriert, die allesamt plausibel sind. Am Beispiel von TIMSS wird die Problematik des Vergleichs der Schulleistungen von Ländern, die sich in ihren Stichproben bzw. in den definierten Populationen stark unterscheiden, diskutiert. Ein unfairer Vergleich der Effektivität von Schulen wird am Beispiel der „school performance tables“ in England dargestellt. Schulvergleiche werden vor allem dann als unfair bezeichnet, wenn sie unabhängig von der Lernausgangslage vorgenommen werden.

Die Schülerinnen und Schüler können von unfairen Vergleichen betroffen sein, wenn sich die Bewertung ihrer Leistungen nur im Bezugsrahmen der jeweiligen Schulklasse interpretieren läßt.

Der Begriff „Fairneß“, der im alltäglichen Sprachgebrauch vor allem im sportlichen und weniger im pädagogischen Kontext verwendet wird, wird von ARNOLD mit unterschiedlichen theoretischen Ansätzen in Verbindung gebracht. Als „pädagogischer Leitbegriff“ (S. 17) wird Fairneß mit Gerechtigkeit oder Chancengleichheit umschrieben. Bezugnehmend auf RAWLS' Verständnis von Gerechtigkeit, nach welchem faire Verteilungskriterien dann vorherrschen, wenn Ungleichheiten an Positionen und Ämter gebunden sind, die allein durch Faktoren wie Begabung, Vorbildung und Leistungen vergeben werden, wird Fairneß im Sinne von Chancengleichheit im Bildungssystem abgehandelt. Auf eindrückliche Weise wird anschließend Fairneß als Qualitätsmerkmal pädagogischer Diagnostik erläutert. ARNOLD prüft in diesem Zusammenhang verschiedene methodisch-formale Qualitätsmerkmale der Leistungsmessung hinsichtlich ihrer Bedeutung für faire Leistungsvergleiche. Dabei berücksichtigt er in gleichen Maßen Kriterien der klassischen Testtheorie und der Item-Response-Theorie. Weniger relevant für die Untersuchung ist die darauf folgende Betrachtung der Fairneß als eines methodisch-inhaltlichen Qualitätsmerkmals. Zwar ist unbestritten, daß beispielsweise individualisierte Leistungsbeschreibung und Lernentwicklungsberichte, Schüler-selbstbeurteilung oder Förderdiagnostik ebenfalls als mehr oder weniger fair bezeichnet werden können, doch sind es – sofern es um Urteile oder Messungen geht – vor allem methodisch-formale Kriterien, die für das Ausmaß der Fairneß entscheidend sind. Im weiteren wird Fairneß mit der Evaluation in Verbindung gebracht und ein Evaluationsmodell zur Qualitätsbeurteilung im Schulsystem als strukturierender Rahmen der Fairneßprü-

fung zugrunde gelegt. Weil in diesem Evaluationsmodell alle Ebenen (Schulsystem, Einzelschule und Individuum) berücksichtigt werden, können die Fairneßaspekte internationaler Schulleistungsvergleiche auf die unterrichtliche Leistungsbeurteilung und die binnenschulische Evaluation von Einzelschulen übertragen werden (S. 68).

ARNOLD beschränkt sich bei der Fairneßprüfung nicht auf TIMSS, sondern bezieht sich auch auf das nationale Evaluationssystem der USA („National Assessment of Educational Progress“ – NAEP), das in seiner methodischen Ausrichtung TIMSS sehr ähnlich ist. Die Studien werden als Modelle externaler Schulleistungsuntersuchungen bezeichnet. Die Ergebnisse der Fairneßprüfung werden auf den unterrichts- und schulbezogenen Anwendungsbereich übertragen, um den praktischen Nutzen aufzeigen zu können. Dabei dienen das in England umgesetzte Evaluationskonzept „National Curriculum Assessment“ (NCA) und BECKERS Modell für die Entwicklung, Durchführung, Auswertung und Interpretation von schriftlichen Leistungsüberprüfungen im Unterricht als Konkretisierungsbeispiele (G. BECKER: *Auswertung und Beurteilung von Unterricht. Handlungsorientierte Didaktik. Teil II.* Weinheim 1986).

Die eigentliche Fairneßprüfung betrifft die drei Phasen „Planung“, „Auswertung“ und „Ergebnismitteilung“ eines internationalen Leistungsvergleichs. ARNOLD umschreibt jeweils kurz die Problemlage und referiert primär die in TIMSS angewendeten Problemlösungsstrategien. Dabei gelingt es, eine Fülle von Problemen internationaler Schulleistungsvergleiche differenziert aufzuzeigen und inhaltliche wie methodische Aspekte der Fairneß sachlich zu beurteilen. ARNOLD zeigt, wie minutiös die Untersuchung TIMSS geplant wurde und mit welcher großer Sorgfalt international valide Testaufgaben entwickelt werden. Die verschiedenen Strukturierungshilfen zur Aufgabenentwicklung sowie die angewen-

deten methodischen Kriterien zur Aufgabenselektion werden detailliert dargestellt. Wie ARNOLD richtig feststellt, verlieren sie – trotz ihrer Ausrichtung auf die internationale Vergleichbarkeit von Schulleistungen – ihre Bedeutung für die unterrichtliche Schülerbeurteilung und die binnenschulische Evaluation in keiner Weise.

Auch die Erkenntnisse über die Auswertung internationaler Leistungsmessungen lassen sich nach ARNOLD problemlos auf die Evaluation einer Einzelschule oder die Beurteilung einer Schülerleistung übertragen. So setzt beispielsweise eine zuverlässige summative Leistungsbeurteilung eines Schulsystems wie einer Schule oder eines Individuums voraus, daß das Ausmaß der Übereinstimmung zwischen „Lehrplan“ und „Stoffplan des erteilten Unterrichts“ ausgewiesen wird und die im Unterricht nicht behandelten Bereiche kenntlich gemacht werden (S. 133).

Selbst die Erfahrungen mit der – nicht immer geglückten – Darstellung und Verbreitung der TIMSS-Ergebnisse bleiben nicht ohne Folgerungen für die Schülerbeurteilung und die Evaluation von Einzelschulen. Während unfaire Vergleiche als Folge ungenügender Stichprobenqualität oft aus politischen Gründen publiziert werden, schließlich erfordern diese Studien einen hohen finanziellen Einsatz (S. 144), besteht das Problem der Ergebnisdarstellung vor allem darin, daß übermäßige Vereinfachungen nicht leicht zu vermeiden sind und Interpretationseinschränkungen als Folge der Datenqualität nicht genügend vermittelt werden. Dem ist nicht zu widersprechen, doch wie ARNOLD überzeugend darstellt (S. 147), gibt

es auch keinen einsichtigen Grund, die Meßfehlerproblematik in der Schülerbeurteilung im Alltag vollständig auszuklammern und so zu tun, als ob sich die Frage der Zuverlässigkeit von Leistungsmessungen nur in größeren Evaluationsvorhaben stellt. Die Qualität der Schülerbeurteilung kann nach ARNOLD vor allem auch dadurch verbessert werden, daß eine Verbindung von Lehrerurteil und Schulleistungsmessung ermöglicht wird. Auf ähnliche Weise könnte es sinnvoll sein, die internationale Leistungsmessung mit nationalen Bewertungssystemen von Schulen zu verbinden.

Das Ergebnis von ARNOLDS' Fairneßprüfung ist eine Fülle von Informationen über methodische Aspekte international vergleichender Leistungsmessung, wie sie im deutschen Sprachraum zum ersten Mal vorzufinden ist. Durch die Verbindung der Fairneßprüfung von TIMSS mit der unterrichtlichen Schülerbeurteilung und der binnenschulischen Evaluation gelingt es ARNOLD, die Erfahrungen aus der Anwendung der Leistungsmessung im internationalen Kontext für Wissenschaft und Praxis nutzbar zu machen und die Relevanz dieser Studien anzudeuten. Diese Verbindung geschieht allerdings etwas bruchstückhaft und wird immer wieder neu aufgerollt, ohne dass sie an einer Stelle eingehend behandelt würde. Dies sollte in naher Zukunft vor allem deshalb geleistet werden, weil es für die Akzeptanz kommender internationaler Schulleistungsvergleiche von besonderer Bedeutung ist.

Dr. URS MOSER

Kompetenzzentrum, Seilergraben 53, CH-8001 Zürich