

Helmke, Andreas

## **Internationale Schulleistungsvergleichsforschung. Schlüsselprobleme und Perspektiven. Einleitung in den Thementeil**

*Zeitschrift für Pädagogik 47 (2001) 2, S. 155-160*

urn:nbn:de:0111-opus-52706

in Kooperation mit / in cooperation with:

# **BELTZ**

<http://www.beltz.de>

### **Nutzungsbedingungen / conditions of use**

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.

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.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.  
By using this particular document, you accept the above-stated conditions of use.

### **Kontakt / Contact:**

**peDOCS**  
Deutsches Institut für Internationale Pädagogische Forschung (DIPF)  
Mitglied der Leibniz-Gemeinschaft  
Informationszentrum (IZ) Bildung  
Schloßstr. 29, D-60486 Frankfurt am Main  
E-Mail: [pedocs@dipf.de](mailto:pedocs@dipf.de)  
Internet: [www.pedocs.de](http://www.pedocs.de)

# Zeitschrift für Pädagogik

Jahrgang 47 – Heft 2 – März/April 2001

## *Thema: Internationale Schulleistungsvergleichsforschung*

- 155 ANDREAS HELMKE  
Internationale Schulleistungsvergleichsforschung.  
Schlüsselprobleme und Perspektiven  
*Einleitung in den Thementeil*
- 161 KARL-HEINZ ARNOLD  
Schulleistungsstudien und soziale Gerechtigkeit
- 179 ECKHARD KLIEME/JOACHIM FUNKE/DETLEV LEUTNER/PETER REIMANN/  
JOACHIM WIRTH  
Problemlösen als fächerübergreifende Kompetenz. Konzeption und erste  
Resultate aus einer Schulleistungsstudie
- 201 PETER MARTIN ROEDER  
Vergleichende ethnographische Studien zu Bildungssystemen:  
USA, Japan, Deutschland

## *Weiterer Beitrag*

- 217 MANFRED LÜDERS  
Dispositionsspielräume im Bereich der Schülerbeurteilung.  
Auch ein Beitrag zur Professions- und Organisationsforschung

## *Diskussion*

- 235 DIRK RUSTEMEYER  
Wie besonders ist das Allgemeine?
- 251 WERNER HELSPER/LEONIE HERWARTZ-EMDEN/EWALD TERHART  
Qualität qualitativer Forschung in der Erziehungswissenschaft.  
Ein Tagungsbericht
- 271 ROSWITHA LEHMANN-ROMMEL  
Neuere Veröffentlichungen über John Dewey. Eine Sammelrezension

## *Besprechungen*

- 285 KLAUS PRANGE  
*Martina Koch*: Performative Pädagogik.  
Über die welterzeugende Wirksamkeit pädagogischer Reflexivität
- 288 SABINE ANDRESEN  
*Dorle Klika*: Herman Nohl. Sein „Pädagogischer Bezug“ in Theorie,  
Biographie und Handlungspraxis
- 291 JÜRGEN ZINNECKER  
*Fritz Haselbeck*: Lebenswelt Schule. Der Schulalltag im Blickwinkel  
jugendlicher Hauptschülerinnen und Hauptschüler. Einstellungen,  
Wahrnehmungen und Deutungen
- Fritz Haselbeck*: Wie Schüler Schule sehen. Hilferufe an Lehrer,  
Eltern und Politiker? Originale, sehr aufschlußreiche Schüleraussagen  
aus Gruppendiskussionen, Einzelinterviews und Schultagebuch-  
aufzeichnungen

## *Dokumentation*

- 295 Pädagogische Neuerscheinungen

*Beilagenhinweis*: Dieser Ausgabe der Z.f.Päd. liegen Prospekte des  
VNR-Verlag für die deutsche Wirtschaft AG, Bonn, und des Verlags der  
Österreichischen Akademie, Wien, bei.

## *Content*

### *Topic: International Comparative Research On School Achievement*

- 155 ANDREAS HELMKE  
International Comparative Research On School Achievement.  
Key Issues and Perspectives  
An Introduction
- 161 KARL-HEINZ ARNOLD  
Studies On School Achievement and Social Justice
- 179 ECKHARD KLIEME/JOACHIM FUNKE/DETLEV LEUTNER/PETER REIMANN/  
JOACHIM WIRTH  
Problem Solving As Transdisciplinary Competence – Conception and  
first results of a study on school achievement
- 201 PETER MARTIN ROEDER  
Comparative Ethnographic Studies On Educational Systems:  
United States, Japan, Germany

### *Further Contributions*

- 217 MANFRED LÜDERS  
Teachers' Discretionary Powers Regarding Student Assessment –  
A contribution to research on professions and organizations

### *Discussion*

- 235 DIRK RUSTEMEYER  
How Specific Is the General?
- 251 WERNER HELSPER/LEONIE HERWARTZ-EMDEN/EWALD TERHART  
The Quality of Qualitative Research In Educational Science –  
A report on a conference
- 271 ROSWITHA LEHMANN-ROMMEL  
Recent Publications on John Dewey. A Complete Review
- 285 BOOK REVIEWS
- 295 NEW BOOKS

# Internationale Schulleistungsvergleichsforschung. Schlüsselprobleme und Perspektiven

## *Einleitung in den Thementeil*

Innerhalb nur eines Jahrzehntes hat sich in der internationalen Schulleistungsvergleichsforschung – ganz besonders in Deutschland, das sich bei etlichen vorangegangenen Schulleistungsvergleichen der IEA (International Association for the Evaluation of Educational Achievement) abstinert verhielt – ein bemerkenswerter Wandel vollzogen. Während das Interesse an solchen international angelegten Studien über die betroffenen Wissenschaftlergruppen hinaus in der Vergangenheit gering und bildungspolitisch weitgehend folgenlos war, hat sich dies in Deutschland seit der TIMSS-Studie gründlich geändert. Selten hat eine internationale Studie hierzulande einen derart großen „impact“ gehabt. Die Ergebnisse von TIMSS und ihre Rezeption waren der Motor für die „empirische Wende“ (LANGE 1999) in der deutschen Bildungspolitik und -finanzierung: In einer Zeit knapper öffentlicher Kassen wird von den Schulen in zunehmendem Maße erwartet, dass sie auch in Form von erzielten Leistungsergebnissen über den Erfolg ihrer Arbeit Rechenschaft ablegen. Für die Arbeit der Schulen heißt dies, dass der Evaluation der Wirkungen (oder Effekte, „outcomes“) des Schulunterrichts eine wachsende Bedeutung zukommt.

Dieses gesteigerte Interesse an Leistungsvergleichen und der Perspektivenwechsel in Richtung auf empirisch nachweisbare Wirkungen der Schule („accountability“) schlagen sich in mehrfacher Weise nieder, am auffälligsten sicher in Form der drei großen internationalen Vergleichsstudien, an denen Deutschland inzwischen teilnimmt. War das Interesse an der Teilnahme an TIMSS seinerzeit noch etwas zögerlich – Baden-Württemberg nahm damals an der Studie TIMSS II (Sekundarstufe I) nicht teil, und TIMSS I (Grundschule) wurde in Deutschland bedauerlicherweise gar nicht durchgeführt – so haben mittlerweile die Kultusminister aller 16 Bundesländer dem auf neun Jahre (1998 – 2006) angelegten Programm PISA zugestimmt, das in drei Zyklen die Schwerpunkte auf die Inhaltsbereiche Lesen und Textverständnis, Mathematik und Naturwissenschaften legt, sowie – quer zu diesen Bereichen – auf fachunabhängige Kompetenzen im kognitiven und sozialen Bereich. Der erste Teil von PISA begann 1998, und die erste empirische Erhebung fand im Sommer 2000 statt. Ergänzend zu PISA hat die KMK ein ebenfalls umfassendes Projekt „Vergleichsuntersuchungen zum Leistungsstand von Schülerinnen und Schülern im Englischen und in der aktiven Beherrschung der deutschen Sprache“ beschlossen, das das Akronym *DESI* erhielt (Deutsch Englisch Schülerleistungen International) und ebenfalls in allen 16 Bundesländern, darüber hinaus in einer Reihe von europäischen Ländern, durchgeführt werden wird. Drittens wird ab 2001 auch erstmals eine Leistungsstudie im Grundschulbereich stattfinden, nämlich das Projekt *IGLU* (Internationale-Grundschul-Lese-Untersu-

chung) im Rahmen der internationalen IEA-Studie PIRLS (Project International Reading Literacy Study), das in 13 der 16 Bundesländer stattfindet und eine zusätzliche Mathematikkomponente umfasst. Viertens schließlich finden in den letzten Jahren in mehreren Bundesländern umfassende Leistungsuntersuchungen und -vergleiche statt. Die wichtigsten sind die Studien LAU in Hamburg (Vollerhebung der 5. Klassenstufe in mehreren Fächern, Folgeerhebungen in der 7. und 9. Klassenstufe), QuaSUM in Brandenburg (Mathematikleistungen in Klassen der 6. und 9. Klassenstufe) und MARKUS in Rheinland-Pfalz (erste Totalerhebung in einem Flächenland: Mathematikleistungen, Unterrichtsqualität und Schulkontext bei den 8. Klassen aller Schulformen, sowie Befragung aller Lehrkräfte und Schulleitungen).

Die bildungspolitische Diskussion zu den Erwartungen, die an solche umfassende nationale und internationale Schulleistungsvergleiche gestellt werden, hat gezeigt, dass es eine Reihe konzeptueller und methodischer Schlüsselprobleme und -fragen gibt, an deren zufriedenstellender Beantwortung man nicht vorbei kommt, will man die vorliegenden Schulleistungsvergleichsstudien verstehen, sie kritisch bewerten und weitere solche Studien planen (vgl. HELMKE 2000); in diesem Zusammenhang sei auch auf das von F.E. WEINERT im Auftrag der KMK herausgegebene Buch mit dem bezeichnenden Titel „Leistungsmessungen in Schulen – Informationsquellen, Teufelswerkzeug oder Orientierungshilfe“ verwiesen (WEINERT [im Druck]). Ich konzentriere mich im folgenden auf diejenigen drei Punkte, die zugleich auch den Schwerpunkt der Beiträge in diesem Heft repräsentieren: (1) die Fairness von Vergleichen, (2) die Frage nach leistungs- und schulfachübergreifenden Kompetenzen und Schülermerkmalen, sog. „Schlüsselkompetenzen“, und (3) die Notwendigkeit, die großen „large scale“-Studien um die differenzierte und detaillierte Analyse einzelner Fälle (Schulen, Klassen) zu ergänzen.

### *Bezugsnormen und Kontext*

Will man die Qualität der in einer Studie erzielten Leistungen bewerten, bedarf es eines Maßstabs, an dem man die Güte der Leistungen messen kann. Dieses Erfordernis gilt gleichermaßen für die Leistungen einzelner Schüler, Klassen oder Schulen als auch für die Standortbestimmung eines nationalen Bildungssystems. Im Prinzip lassen sich mindestens die folgenden Arten von Bewertungsmaßstäben oder Bezugsnormen vorstellen (vgl. hierzu auch SCHRAEDER/HELMKE [im Druck]):

- a) Der *absolute Leistungswert* – etwa der Prozentsatz der Schüler/innen, die die Aufgaben eines Mathematiktests gelöst haben („percent correct“) – ist in den wenigsten Fällen aussagekräftig: In der Regel werden bei Vergleichsuntersuchungen (aus guten Gründen) Zeitlimits oder Taktungen eingebaut, sodass nicht alle Aufgaben von allen Schülern bearbeitet oder gelöst werden können und sollen. Dazu kommt, dass es sich schon aus pragmatischen Gründen – anders als bei Klassenarbeiten – bei solchen Tests ganz überwiegend um Multiple Choice-Items handelt, bei denen auch im Falle des Nichtwissens intelligente Formen des Ratens Anwendung finden können.

- b) Von zentraler Bedeutung ist der Vergleich mit einem *inhaltlichen Kriterium*, d.h. die Leistungen werden daran gemessen, ob und in welchem Umfang die Lernziele des Lehrplans (einer bestimmten Klassenstufe innerhalb eines Bildungsganges) erreicht wurden (kriteriale Bezugsnorm). Stellt man in Rechnung, dass die Lernziele unterschiedlich anspruchsvoll sind und bildet sich dies in der Struktur und Schwierigkeit der Aufgaben ab, dann lassen sich anschließend Kompetenzstufen berechnen. So wurde es beispielsweise auch bei TIMSS gehandhabt.
- c) Es sind auch *Vergleiche über die Zeit* hinweg möglich: Vergleich der Ergebnisse des Leistungstests einer Klasse am Ende eines Schuljahres mit dem Vorkenntnisniveau zu Beginn. Dies erfordert eine Zweipunktmessung und ermöglicht Aussagen über den Leistungszuwachs in diesem Zeitraum.
- d) *Bei epochalen Vergleichen* werden die erzielten Ergebnisse mit denen vorangegangener Studien, die zu einem früheren Zeitpunkt das Gleiche gemessen haben, in Beziehung gesetzt. Im Gegensatz zu den immer wiederkehrenden Erhebungen, beispielsweise im Rahmen der Jugendstudie oder zur Lage der Studierenden, gibt es im Leistungsbereich allerdings nur wenige Möglichkeiten, der damit beantwortbaren Frage nach solchen historischen und epochalen Unterschieden nachzugehen – dies würde identische Testitems und Instruktionen erfordern. Zu den wenigen – und sehr aussichtsreichen – Ausnahmen gehört die TIMS-Studie, die in einer großen Zahl von Ländern (allerdings nicht in Deutschland) wiederholt wird („TIMSS-Repeat“).
- e) Was bei der Berichtslegung über Leistungsunterschiede in der Bildungsforschung und vor allem in der öffentlichen Rezeption dominiert, sind Vergleiche der gemessenen Leistungen zwischen verschiedenen Klassen, Schulen – oder, im Falle internationaler Vergleichsstudien: von Nationen (*soziale Bezugsnorm*). Vielfach waren es überhaupt *nur* die (nach Leistungsmittelwerten geordneten) Ranglisten, die im Mittelpunkt der öffentlichen Berichterstattung standen und die Diskussion dominierten.

Hier stellt sich die schwierige Frage nach der Angemessenheit solcher Vergleiche. Dies lässt sich besonders anschaulich machen, wenn man als Analyseeinheit die Schule wählt. So liegt es auf der Hand, dass es unangemessen und „ungerecht“ wäre, von einer Schule „im sozialen Brennpunkt“ das gleiche Leistungsniveau zu erwarten wie von einer privilegierten Schule mit besonders günstigen, das Unterrichten und damit auch die Leistungsentwicklung fördernden Faktoren (z.B. hohes Vorkenntnisniveau, Leistungshomogenität, Bildungsnähe der Eltern, geringe Sprachenvielfalt innerhalb der Klasse etc.). Um Unterschieden in den Kontextbedingungen Rechnung zu tragen, sind anerkannte statistische Verfahren entwickelt worden (Stichwort: Bildung von Residualwerten; Erwartungswert), deren Logik darin besteht, benachteiligten Schulen einen „Bonus“ zu geben und bevorzugten Schulen einen „Malus“. Man kann dementsprechend zwei Arten von Leistungsprofilen unterscheiden: Eines enthält die erzielten „Rohwerte“ (unbereinigte Werte) und ein zweites Profil die „adjustierten“ Werte. Das Rohwertprofil zeigt, welche Leistungen real erzielt wurden. Das Profil der adjustierten Werte orientiert sich daran, welches Leistungsniveau man von der betreffenden Schule bzw. Klasse in Kenntnis der vorhandenen Kontextmerkmale und ihres nachweislichen Einflusses auf die Leis-

tung eigentlich erwarten würde (Erwartungswert), und korrigiert die real erzielten Leistungen dementsprechend.

Dies führt bei Schulen mit großen Standortnachteilen zu einem günstigeren Ergebnisbild. Allerdings darf die Anwendung dieses Prinzips nicht dazu führen, dass man das Kind mit dem Bade ausschüttet: Eine Minderleistung kann natürlich sehr wohl – zumindest partiell – durch einen ungünstigen Kontext erklärbar sein. Andererseits ist aber die Vergleichbarkeit von Bildungsabschlüssen und ihren Voraussetzungen ebenfalls ein Gebot der Gerechtigkeit. Mit anderen Worten: Vergibt eine standortmäßig benachteiligte Schule trotz objektiv geringerer Leistungen den gleichen formalen Abschluss bzw. die gleichen Noten wie eine privilegierte Schule, dann kann ein solches Ergebnis unter Umständen als unakzeptabel und unfair angesehen werden. Ein weiteres Problem ist, *welche Merkmale* als gegebene, nicht direkt steuer- und veränderbare Kontextmerkmale der „Bereinigung“ zugrunde gelegt werden: nur der regionale und familiale oder auch der schulische Kontext? Nur Aspekte, die mit Leistung, oder auch solche, die mit anderen Zielkriterien (z.B. Lernmotivation) zusammenhängen? Diese und viele andere Fragen (z.B. nach der Analyseinheit, auf der bereinigt wird: Schüler oder Klasse?) sind zur Zeit noch weitgehend ungeklärt.

Damit ist nur ein Teil des grundsätzlichen Problems der Fairness von Schulleistungstests angesprochen, das im Zentrum des Artikels von K.-H. ARNOLD steht. Arnold führt in seinem Artikel grundlegende Überlegungen zu Gerechtigkeit und Fairness aus schulpädagogischer Sicht mit Konzepten der Fairness in der Psychologischen Diagnostik zusammen. Auf einem integrativen Konzept von Fairness aufbauend, geht er – exemplarisch für TIMSS – der Frage nach der transnationalen Validität bei internationalen Vergleichsstudien nach – angesichts der enormen Unterschiedlichkeit der Bildungssysteme und Curricula der beteiligten Länder ein Schlüsselproblem. In diesem Zusammenhang geht er auch auf die allen neueren Leistungsstudien zugrunde liegende Probabilistische Testtheorie ein (IRT = item response theory), ohne deren Verständnis – zumindest in Grundzügen – eine kritische Bewertung der Ergebnisse schwierig erscheint.

### *Schlüsselkompetenzen*

Eine zweite Kernfrage von Leistungsvergleichen betrifft den Leistungsbegriff. Das Interesse früherer Untersuchungen richtete sich ganz überwiegend auf kognitive, durch Schulleistungstests erfasste Leistungen. Lediglich die großen Langzeitstudien und Schulsystemvergleichsprojekte von H. FEND, die die überfachlichen, erzieherischen Wirkungen der Schule zum Gegenstand hatten, sind eine Ausnahme davon (vgl. FEND 1998). Der deutlichste Beleg dafür ist das Projekt PISA, bei dem der Bereich der Schlüsselkompetenzen – neben Leistungen in Mathematik, Naturwissenschaften und Lesen – zu den tragenden Säulen gehört. Nachdem die Abnehmer schulischer Ausbildung – insbesondere die Wirtschaft und die Universitäten – seit jeher den Mangel an überfachlichen Kompetenzen beklagen (Stichwort: defizitäre Studierfähigkeit, mangelnde soziale Kompetenzen), scheint man sich inzwischen auch in der Bildungspolitik



darüber einig zu sein, dass fachliche Leistungen und Wissenserwerb allein nicht ausreichend sind und dass es deshalb kurzsichtig und unangemessen wäre, den Vergleich der Wirksamkeit von Schulsystemen und Schulen allein darauf zu beschränken. Dieser Konsens wird jedoch durch den inzwischen inflationären Gebrauch des Konzeptes der Schlüsselkompetenzen zunehmend getrübt; die Palette der Begriffe, denen eine „Schlüsselrolle“ für die Bewältigung schulischer und außerschulischer Anforderungen zukommt, ist inzwischen riesig, und nicht wenige beginnen sich von diesem Konzept frustriert abzuwenden.

Dies ist die Ausgangslage für den Artikel von E. KLIEME u.a. Gestützt auf empirische Untersuchungen im Vorfeld der PISA-Hauptstudie stecken die Autoren zunächst den konzeptuellen Rahmen ab, indem sie Facetten des Kompetenzbegriffs aus pädagogischer und psychologischer Sicht entfalten. Der weiteren Darstellung wird – in Anlehnung an eine grundlegende Arbeit von F.E. WEINERT (1999) – eine funktionale Definition von Kompetenz zugrunde gelegt: Kompetenzen werden angesehen als „funktional bestimmte, auf bestimmte Klassen von Situationen und Anforderungen bezogene kognitive Leistungsdispositionen, die sich psychologisch als Kenntnisse, Fertigkeiten, Strategien, Routinen oder auch bereichsspezifische Fähigkeiten beschreiben lassen“. Darauf aufbauend, wird vier grundlegenden Fragen nachgegangen: (1) Wie lassen sich Kompetenzen operationalisieren, (2) sind sie wirklich bereichsübergreifend, (3) lassen sie sich von kognitiven Grundfähigkeiten unterscheiden und (4) sind sie durch schulische Bildung beeinflussbar. Diese Fragen werden am Beispiel des Bereichs „Problemlösen“ diskutiert. Dabei wird klar, dass die Erfassung des Problemlösens zwar eine wichtige und fruchtbare Erweiterung der traditionellen Schulleistungsdiagnostik darstellt, wegen der starken Überlapung mit kognitiven Grundfähigkeiten („reasoning“) jedoch auch mit Schwierigkeiten verbunden ist.

### *Qualitative Fallstudien*

Es ist bedauerlicherweise viel zu wenig bekannt, dass TIMSS aus weit mehr als der vergleichenden Messung mathematischer und naturwissenschaftlicher Leistungen besteht, auf die sich oft die Rezeption der Ergebnisse in der Öffentlichkeit weitgehend beschränkt hat (und dann noch begrenzt auf die Länderrangliste von TIMSS II / Sekundarstufe). Neben einer umfassenden, international vergleichenden Analyse von Lehrplänen und Lehrbüchern umfasste TIMSS insbesondere auch eine Videostudie sowie ethnografische Fallstudien in drei Ländern: Deutschland, Japan und USA. Während die TIMSS-Videostudie inzwischen zunehmend diskutiert wird, insbesondere auch von Didaktikern (vgl. BLUM/NEUBRAND, 1998), gibt P.M. ROEDER erstmals einen differenzierten Überblick über die ethnografischen Fallstudien. Dabei werden vier Themenbereiche dargestellt: (1) die Entwicklung und Kontrolle von Leistungsstandards, (2) die Interpretation und der Umgang mit unterschiedlichen Lernvoraussetzungen auf Seiten der Schüler, (3) der relative Stellenwert der Schule, verglichen mit anderen unterrichtlichen und Sozialisationskontexten, und (4) Aspekte der Ausbildung und beruflichen Entwicklung von Lehrern.

Dabei entsteht ein äußerst facettenreiches Bild, das die Ergebnisse der „Large Scale“ Studie von TIMSS um wichtige Details ergänzt und anreichert. Besonders markant erscheint hier insbesondere das auf den Interviews und Beobachtungen in einzelnen Schulen basierende Bild der schulischen Erziehung und Leistungsförderung in Japan: Erst das Zusammenspiel der zwischen öffentlicher Schule, privater Zusatzschule und Familie zu findenden Arbeitsteilung, der ganzheitlichen – nicht auf kognitives Lernen beschränkten – Schulerziehung, des Leistungsdrucks und der Anstrengungsbereitschaft und des wesentlich engeren und zeitlich längeren Lehrer-Schüler-Kontakts erklärt die Leistungsüberlegenheit der japanischen Schüler. Aber auch zwischen den USA und Deutschland gibt es – bei annähernd gleichem Leistungsniveau, wie die TIMSS-Hauptstudie gezeigt hat – im Hinblick auf Klassenklima, Lerneinstellung und Erziehungsphilosophie bemerkenswerte Kontraste.

### Literatur

- BLUM, W./NEUBRAND, M. (Hrsg.): *TIMSS und der Mathematikunterricht: Informationen, Analysen, Konsequenzen*. Hannover 1998.
- FEND, H.: *Qualität im Bildungswesen. Schulforschung zu Systembedingungen, Schulprofilen und Lehrerleistung*. München 1998.
- HELMKE, A.: TIMSS und die Folgen. Der weite Weg von der externen Leistungsevaluation zur Verbesserung des Lehrens und Lernens. In: U.P. TRIER (Hrsg.), *Bildungswirksamkeit zwischen Forschung und Politik*. Bern 2000.
- LANGE, H.: Qualitätssicherung an Schulen. In: *Die Deutsche Schule* 2 (1999) S. 144–159.
- SCHRADER, F.-W./HELMKE, A.: Alltägliche Leistungsbeurteilungen durch Lehrer im Klassenzimmer. In F.E. WEINERT (Hrsg.): *Leistungsmessungen in Schulen – Informationsquelle, Teufelswerkzeug oder Orientierungshilfe?* Stuttgart (im Druck).
- WEINERT, F.E.: Concepts of competence. In: OECD (Hrsg.): *Definition and selection of competencies: Theoretical and conceptual foundations*. Paris (OECD) 1999.
- WEINERT, F.E. (Hrsg.): *Leistungsmessungen in Schulen – Informationsquelle, Teufelswerkzeug oder Orientierungshilfe?* Stuttgart (im Druck).

### Anschrift des Autors

Prof. Dr. Andreas Helmke, Universität Landau, FB Psychologie, Im Fort 7, 76829 Landau