



## Die Daten der Vergleichsuntersuchungen und ihre Nutzungsmöglichkeiten für erziehungswissenschaftliche Forschung

*Knut Schwippert*

*Ruhr-Universität Bochum*

*Wilfried Bos*

*Universität Dortmund*

### *Abstract*

The article describes the research tradition in which the findings of large-scale surveys on school achievement are discussed at present. From a contemporary point of view educational science in Germany – strongly influenced by the humanities – missed to participate in international large-scale surveys for two decades. Now educational science in Germany has to catch up with the results of studies like TIMSS, PISA, and PIRLS, which meanwhile have gained pronounced attention among experts, the public and the media. This could be of great profit for the empirical educational research, but on the contrary, the extensive reception of the findings of the PISA study has turned to a drop of bitterness. Several discussants, referring to the TIMSS and PISA results, read the findings carefully, others simply methodically misinterpreted the data. The authors give examples of some of the most popular misinterpretations of the findings of large-scale surveys and present starting points for the German educational science to attend and to further develop the processes of discussion. (The article is based on a conference paper held by the authors and the oral notation was kept for this journal).

Im Folgenden wird kursorisch dargestellt, in welcher (jüngeren) Forschungstradition die Befunde von groß angelegten Schulvergleichsuntersuchungen derzeit diskutiert werden. Aus heutiger Sicht zeigt sich, dass die deutsche geisteswissenschaftlich geprägte Erziehungswissenschaft für einige Jahre versäumt hat, an internationalen Studien teilzunehmen. Umso mehr holen sie derzeit die Befunde aus Studien wie TIMSS, PISA und IGLU / PIRLS ein, die ein bisher ungekanntes Maß an Aufmerksamkeit

nicht nur in Fachkreisen, sondern auch in der Öffentlichkeit und den Medien erfahren. Was Balsam für die lange Zeit nicht akzeptierte empirische Bildungsforschung sein könnte, entwickelt sich jedoch mit zunehmender – fast lawinenartiger Rezeption der PISA-Befunde zu einem Wermutstropfen. Nicht alle Diskutierenden, die sich auf TIMSS oder PISA berufen, haben diese Quelle auch sorgfältig gelesen oder sie unterliegen methodisch bedingten Interpretationsfehlern. Daher werden im zweiten Abschnitt des Beitrags einige der populärsten Fehlinterpretationen von Befunden aus groß angelegten Schulvergleichsuntersuchungen vorgestellt. Der Abschluss des Beitrags bietet schließlich – das Thema der Sektionssitzung im Auge behaltend<sup>1</sup> – einen Überblick über mögliche Ansatzpunkte der Erziehungswissenschaft, wie diese vor dem Hintergrund ihrer langjährigen Expertise Diskussions- und Entwicklungsprozesse fachwissenschaftlich begleiten könnte. Eine Begleitung, die jedoch eine prinzipielle Offenheit gegenüber der empirischen Bildungsforschung und den Willen, sich mit der sich weiterentwickelnden Erziehungswissenschaft ernsthaft auseinandersetzen zu wollen und damit eine Lernwilligkeit voraussetzt. Denn auch wenn heute erwartetes empirisches Basiswissen in den letzten Dekaden nicht zum Standardrepertoire der erziehungswissenschaftlichen Ausbildung gehört hat, so darf man sicherlich ohne Zweifel davon ausgehen, dass eine Zunft, die sich mit Lernen – wenn nicht gar mit lebenslangem Lernen – auseinandersetzt, nicht Wasser predigt und selber Wein trinkt – oder mit anderen Worten sich der international etablierten empirischen Bildungsforschung verschließt.

## 1. Das deutsche Bildungssystem – eine Defizitanalyse

Der inzwischen legendäre TIMSS-Schock rüttelte Mitte der 1990er Jahre die deutsche pädagogische Fachöffentlichkeit auf. In der *Third International Mathematics and Science Study* (TIMSS) wurden den deutschen Siebt- und Achtklässlern nur mittelmäßige Leistungen bei mathematischen und naturwissenschaftlichen Kenntnissen bescheinigt (Baumert et al., 1997). Den älteren Jugendlichen, die das deutsche Regelschulsystem am Ende der Sekundarstufe II verließen, konnte im Rahmen der Untersuchung auch kein besseres Zeugnis ausgestellt werden (Baumert, Bos & Lehmann, 2000a, b). Immerhin bewirkte das unerwartete Abschneiden, dass die Kultusministerkonferenz entschied, Deutschland solle sich an den folgenden internationalen Schulvergleichsstudien beteiligen. In der im Jahre 2000 durchgeführten PISA-Studie (*Programme for International Student Assessment*) erzielten die untersuchten fünfzehnjährigen Schülerinnen und Schüler in den getesteten Kompetenzen Lesen, Mathematik und Naturwissenschaften nur ein schwaches mittleres Niveau. Wie auch bei TIMSS wurden einige der veröffentlichten PISA-Befunde als alarmierend wahrgenommen. In der öffentlichen Diskussion über Bildungsfragen war nun der ‚PISA-Schock‘ zum geflügelten Wort geworden. Doch auch aus wissenschaftlicher Sicht erscheinen die in der Öffentlichkeit plakativ herausgestellten Befunde nicht weniger dramatisch. Inter-

pretiert man nicht nur das einfache Länderranking, sondern misst die Leistungen der Schülerinnen und Schüler bzw. des Bildungssystems mit begründeten Leistungsnormen, wird das Ausmaß der Misere noch sichtbarer (vgl. Baumert et al., 2001). Die PISA 2000-Studie macht deutlich, dass gut jeder Fünfte der getesteten Jugendlichen nicht über eine Lesekompetenz verfügt, um ein selbständiges und selbstbestimmtes Leben in unserer Gesellschaft zu führen. Die in PISA getestete Mathematik beherrschen gut 40 % von Hauptschülern lediglich auf dem Grundschulniveau. Und schließlich gehen bei nicht mal der Hälfte der Jugendlichen ihre Kenntnisse in den Naturwissenschaften über ein normales Alltagsverständnis hinaus. Selbst vor dem Hintergrund, dass es Industrienationen gibt, die noch schlechtere Befunde aufweisen, ist es doch eine Frage des Prinzips, welchen Populationsanteil mit unzureichenden Kenntnissen man in einer Gesellschaft akzeptieren möchte.

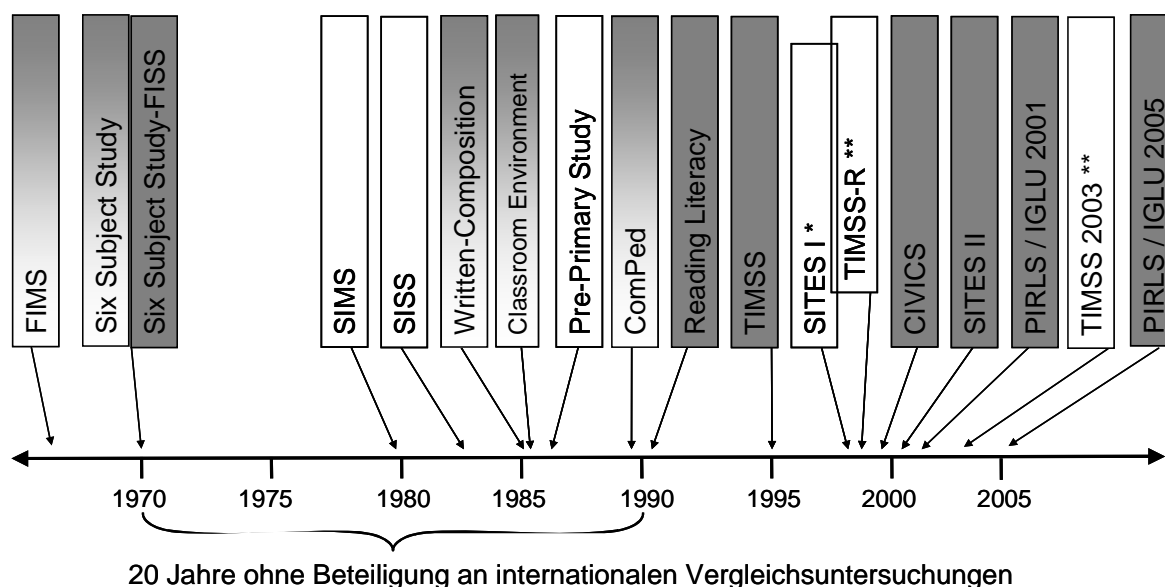
Im Folgenden sollen zwei Erklärungsansätze vorgestellt werden, wie es zu der zuvor skizzierten Situation kommen konnte. Zunächst scheint die über Jahre bestehende Informationslücke über den Zustand des Bildungssystems ein wichtiger Aspekt zu sein. Darüber hinaus scheint auch das falsche Vertrauen auf sich selbst steuernde Bildungsprozesse mit zu der Fehlentwicklung beigetragen zu haben. Am Beispiel von Ländern, die in der PISA-Untersuchung erfolgreicher als Deutschland abgeschnitten haben, können Perspektiven für die Weiterentwicklung des deutschen Bildungssystems abgelesen werden. In diesen Ländern hat sich gezeigt, dass die Formulierung von nationalen Bildungsstandards ein Instrument zur effektiven Steuerung von schulischen Bildungsprozessen sein kann. Welche Konsequenzen eine Orientierung an Bildungsstandards für die Wirksamkeit von Schule hat, aber auch welche Perspektive sich für die Ausbildung von Lehrkräften hieraus ergibt, wird zu diskutieren sein.

## 2. Das Defizit an empirischer Bildungsforschung

Die deutsche Erziehungswissenschaft ist viele Jahre durch eine geisteswissenschaftliche Tradition geprägt gewesen und zeigt daher eine Stärke in der theoretischen Reflexion von Bildungsprozessen. ‚Bildung‘ umfasste in diesem Verständnis weit mehr als lediglich singulär abgeprüfte Kompetenzen. Im Sinne einer ‚Menschenbildung‘ wird beansprucht, junge Menschen umfassend in allen ihren Fähigkeiten zu fördern, so dass sie ein individuell bestimmtes und gesellschaftlich verantwortetes Leben führen können. In dieser traditionellen Denkweise festigte sich die Überzeugung, dass man ‚Bildung‘ nicht messen könne. In der Tat entzieht sich eine so verstandene Bildung einer (quantitativ) empirischen Erfassung, stellt aber einen hervorragenden Gegenstand für philosophische Betrachtungen dar. Schließlich führte eine ideologisch fundierte Skepsis gegenüber empirischen Untersuchungen von Lernprozessen zur oben bereits skizzierten Abstinenz bei der Teilnahme an international etablierter (empirischer) Bildungsforschung, was sich exemplarisch an der westdeutschen Bildungsreformdebatte in den 1970er und 1980er Jahre zeigen lässt. ‚Bildung für alle‘ oder

„gleiche Bildungschancen für alle“ waren in dieser Zeit zentrale bildungspolitische Ziele. Nach Peisert (1967) wurde das „katholische Arbeitermädchen vom Lande“ als Synonym für im Bildungssystem mehrfach Benachteiligte bekannt. Abhilfe für solche Bildungsbenachteiligung sollte damals die Einführung einer Gesamtschule schaffen, da von dieser Schulform von Bildungspolitikern und Wissenschaftlern erwartet wurde, dass sie soziale Disparitäten in Bezug auf Bildung kompensieren könnte. Hierbei ist aus heutiger Sicht bemerkenswert, dass viele der in den vergangenen Dekaden durchgeführten Reformen – so auch die Einführung der Gesamtschule – ohne empirische Überprüfung implementiert und tradiert wurden.

Abbildung 1: Deutsche Beteiligung an internationalen Vergleichsstudien der *International Association for the Evaluation of Educational Achievement* (IEA) bis 2005



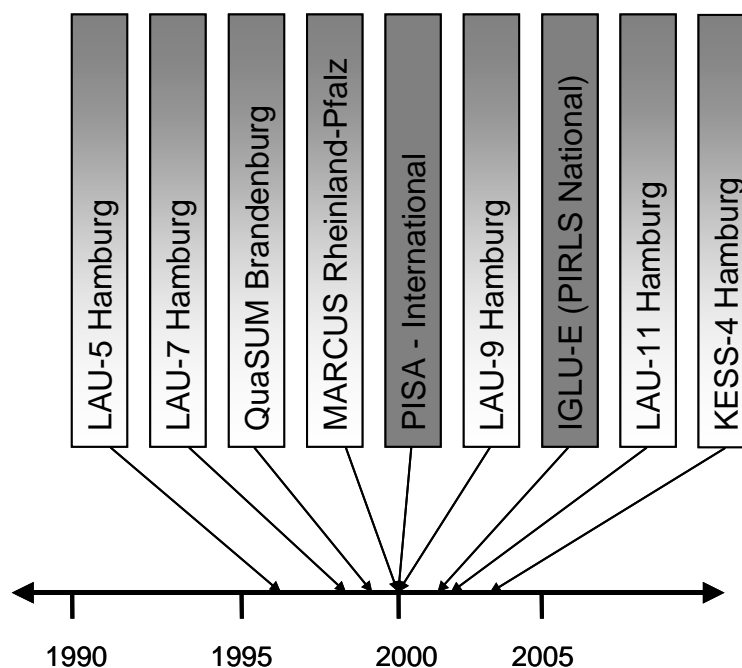
\* Deutschland nahm an dem ersten Modul der SITE-Studie nicht teil, da zur gleichen Zeit bereits das inhaltsgleiche Projekt ‚Schulen ans Netz‘ etabliert war.

\*\* Auf die Teilnahme an der TIMSS-R (TIMSS 1999) Studie verzichtete Deutschland zugunsten der Teilnahme an der PISA-Untersuchung.

Wie kam es gegenüber der empirischen Bildungsforschung zu dieser Zurückhaltung? In den frühen 1970er Jahren beteiligte sich Deutschland (West) an der ersten internationalen Naturwissenschaftsstudie (*First International Science Study*, FISS). In dieser Untersuchung wurde schon damals festgestellt, dass deutsche Schülerinnen und Schüler im internationalen Vergleich entgegen den Erwartungen lediglich unterdurchschnittliche Ergebnisse erzielten. Es folgte eine kurze inhaltliche Auseinandersetzung mit den Befunden, aber eine kontinuierliche empirische Erforschung des Dilemmas im Rahmen von internationalen Vergleichsstudien blieb aus. Abbildung 1 ist Deutschlands Teilnahme an von der *International Association for the Evaluation of Educatio-*

nal Achievement (IEA) organisierten Systemmonitoring Studien zu entnehmen. Die durchgängig grau hinterlegten Studien kennzeichnen eine für Deutschland repräsentative Beteiligung. Nur zum Teil schattierte Studien weisen auf eine partielle Teilnahme von Teilgruppen aus Deutschland hin. Die übrigen Studien wurden ganz ohne deutsche Beteiligung durchgeführt. Wie der Darstellung zu entnehmen ist, hat sich Deutschland für rund 20 Jahre nicht mit repräsentativen Stichproben an internationalen Vergleichsuntersuchungen beteiligt – 20 Jahre, in denen es bei vielen aus dem Blickfeld geriet zu schauen, auf welchem Niveau sich das deutsche Bildungssystem bei der Vermittlung von zentralen Kulturtechniken moderner Gesellschaften befand. In der geisteswissenschaftlich geprägten Tradition wurde davon ausgegangen, dass ein internationaler Vergleich keine grundlegenden Defizite aufzuzeigen in der Lage sei. Erst als 1991 Rainer H. Lehmann sich entschied, an der Internationalen Leseverständnisuntersuchung mit einer für West- und Ostdeutschland repräsentativen Stichprobe teilzunehmen, wurden erste Hinweise auf Mängel im Leistungsniveau deutscher Schülerinnen und Schüler virulent (Lehmann, Peek, Pieper & Stritzky, 1991). Entgegen der verbreiteten Auffassung, dass Bildung nicht messbar sei, entschieden sich einige Bildungsforscher, die zutage getretenen Forschungsdesiderate aufzugreifen und in den folgenden Jahren regelmäßig wieder an internationalen Schulvergleichsstudien teilzunehmen. Schließlich brauchte es aber noch weitere Jahre bis der TIMSS-Schock ein Umdenken in der forschenden Erziehungswissenschaft initiierte und die Perspektive der empirischen Überprüfung die Wahrnehmung des deutschen Bildungssystems neu strukturierte.

Abbildung 2: Nationale Schulvergleichsuntersuchungen und Ergänzungsstudien zu Internationalen Studien bis 2005



Aber nicht nur durch die Teilnahme an international angelegten Schulvergleichsuntersuchungen manifestierte eine reale empirische Wende. Durch eine ganze Reihe von bundeslandspezifischen Untersuchungen wurde seit Mitte der 1990er Jahre vertieft nach Entwicklungsmöglichkeiten des Bildungssystems geforscht (vgl. Abb. 2). Studien, die nachhaltig auf Defizite der jeweiligen Bildungssysteme hinwiesen, konnten empirische Evidenz für Fehlentscheidungen und Missstände in der aktuellen Diskussion liefern und somit zu einer objektivierten Entscheidungsfindung beitragen.

Ein empirisch belegtes Argument für die erfolgreiche Etablierung von systematischen Untersuchungen im Bildungssystem lässt sich am Beispiel Hamburgs, zumindest für die Grundschule, zeigen. Hier wurden – ebenfalls durch Rainer Lehmann initiiert – 1996 erstmals flächendeckend in einem Bundesland die Kompetenzen von Schülerinnen und Schülern in der 5. Klasse zur Erfassung der Lernausgangslage (LAU 5) erfasst und alle zwei Jahre mit denselben Jugendlichen wiederholt. Rund sieben Jahre später wurde im Rahmen der LAU-Folgestudie ‚Kompetenzen und Einstellungen von Schülerinnen und Schülern‘ (KESS 4) festgestellt, dass sich die Kompetenzen der Viertklässler in allen erfassten Bereichen (Lesen, Mathematik und Naturwissenschaften) signifikant verbessert haben (vgl. Bos, Pietsch et al., 2005). Maßgeblich zu diesem Erfolg haben vermutlich die konsequenten Rückmeldungen und Auseinandersetzungen mit den Befunden der LAU-Studie beigetragen, aber auch die systematische Bearbeitung der festgestellten Defizite im Primarschulbereich. So wurden Mitte der 1990er Jahre mehr als 150 Lehrkräfte geschult, besonders leseschwache Kinder zu fördern. Diese Lehrkräfte konnten als ‚Zusatzlehrer‘ bei Bedarf angefordert werden. Darüber hinaus wurde in den Grundschulen die Unterrichtszeit erhöht.

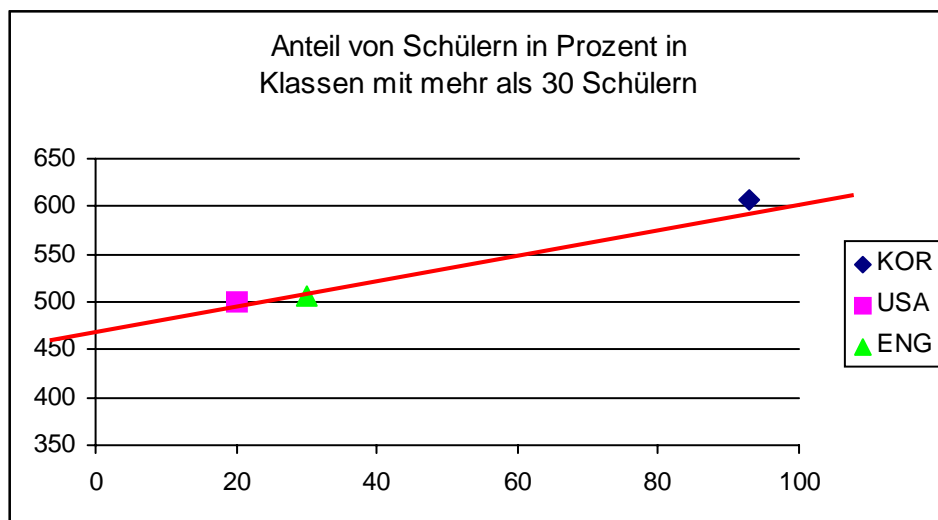
Insgesamt ist in Deutschland aber nach wie vor ein evaluationshemmendes Klima weit verbreitet, so dass die konstruktive Nutzung von Leistungsstudien erschwert wird. Bei vielen Lehrkräften besteht ein latentes Unbehagen gegen die Fremdbewertung des eigenen Unterrichts. Die Möglichkeit, sich auf der Basis empirisch gewonnener und ausgewerteter Informationen weiter entwickeln zu können und somit auch die eigenen Stärken aber auch Schwächen auszuloten, wird vielerorts noch nicht genutzt.

Es zeichnet sich ab, dass es noch eines längeren Lernprozesses bedarf, um Evaluationen primär als Weiterentwicklungsmöglichkeit für die eigene pädagogische Arbeit zu verstehen. Flankierend dürften die Schulverwaltungen die Evaluation nur in einem fördernden Sinne, nicht aber als Ausgangspunkt für Kontrollen und Sanktionen nutzen. Wenn dies gelingt, kann man von Lehrkräften eine positive Haltung gegenüber der Evaluation des eigenen Unterrichts erwarten. In Ländern wie den Niederlanden oder Kanada werden Untersuchungen bereits selbstverständlich genutzt, um Reformen auf ihre Zielerreichung hin zu überprüfen. Hier werden Reformen nicht nur auf Basis von plausiblen Annahmen durchgeführt, sondern ihre tatsächliche Wirkung anhand von geeigneten Tests empirisch überprüft und so schließlich auch weiterentwickelt.

### 3. Populäre Fehlinterpretationen

Kurz nach der Veröffentlichung der internationalen TIMSS-Datenbank wurde diese von vielen Wissenschaftlern für Sekundäranalysen genutzt. Aufgrund des komplexen Stichproben- und Studiendesigns ist es bei der Auswertung jedoch notwendig, über übliche Standardverfahren der Statistik hinweg spezifische Annahmen zu machen und bei den Auswertungen auch zu berücksichtigen. Neben diesen methodischen Fallstricken ergaben sich jedoch auch für viele Autoren Fehlinterpretationen, da sie schlicht das gesamte Untersuchungsdesign – trotz richtiger empirischer Auswertung – nicht angemessen berücksichtigten. Kenneth Ross (1997) griff diese populären Fehlinterpretationen auf, um auf die impliziten Fallstricke aufmerksam zu machen. So hat er aus dem TIMSS-Datensatz drei Länder herausgegriffen und bei diesen die mathematischen Kompetenzen einem Indikator für die Klassengrößen gegenüber gestellt.

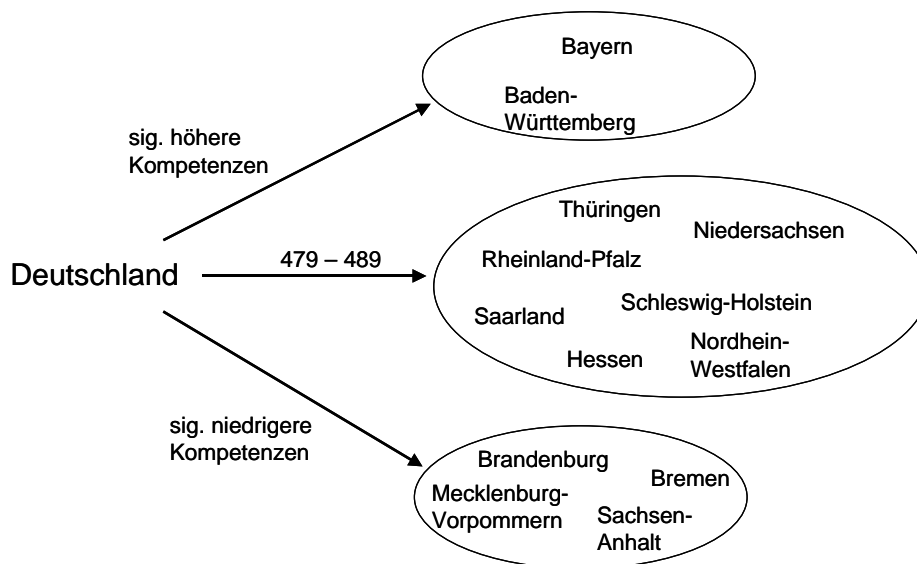
Abbildung 3: Exemplarische Gegenüberstellung mathematischer Kompetenzen und Klassengröße nach Kenneth Ross



Diese Darstellung ist in Abbildung 3 repliziert. Wie dieser zu entnehmen ist, scheint ein positiver Zusammenhang zwischen Klassengröße und Kompetenzen zu bestehen. Hieraus könnte man den Schluss ziehen, die Klassen größer zu machen, um bessere mathematische Kompetenzen zu erzielen. Einen entsprechenden Zusammenhang – nur mit negativem Vorzeichen – hat Kenneth Ross zwischen Lehrergehältern und mathematischen Kompetenzen nachgewiesen und hieraus schließlich – ebenso augenzwinkernd wie beim ersten Befund – gefordert, den Lehrkräften weniger zu bezahlen, um den Schülerinnen und Schülern bessere Lernmöglichkeiten zu geben. Beide Schlüsse basieren selbstverständlich auf einem Interpretationsdilemma, da bei der rein numerischen Betrachtung die unterschiedlichen Kulturen der berücksichtigten Länder außer Acht gelassen wurden. Es ist, führt man sich die willkürliche Auswahl von Korea, den

Vereinigten Staaten und England einmal vor Augen, selbstverständlich nicht möglich, solche (folgenreichen) Schlussfolgerungen zu ziehen, ohne sich weitere Informationen über die berücksichtigten Länder zu vergegenwärtigen. Auch wenn in Korea die Klassen deutlich größer und die Kinder besser in Mathematik sind als in den anderen aufgeführten Ländern, so ist doch klar, dass man nicht einfach das koreanische Bildungssystem auf die anderen Nationen übertragen kann, um so nachhaltige Verbesserungen zu erzielen. Gerade die hohen Klassenfrequenzen in asiatischen Staaten eignen sich gut, um auf entsprechende Fehlinterpretationen hinzuweisen, da niemand substantielle Unterschiede in den Gesellschaften Asiens, Europas und Nordamerikas leugnen wird. Entsprechendes gilt selbstverständlich auch für die Lehrergehälter.

Abbildung 4: Beispiel einer angemessenen Abbildung für populäre Journale des PISA 2000-Bundesländervergleichs (unter Berücksichtigung der Konfidenzintervalle)



Um entsprechende Fehlinterpretationen aufzudecken, braucht man aber nicht das internationale Parkett zu betreten. Populäre Darstellungsfehler lassen sich auch innerhalb von Deutschland finden. Bei der Veröffentlichung der bundeslandspezifischen Ergebnisse aus PISA 2000 wurden z.B. in der Zeitschrift Focus (Adler et al., 2002) die Befunde der Bundesländer in die internationale Rangreihe hineingesteckt und so der Eindruck erweckt, dass Bayern immerhin noch gut zwischen Schweden und Österreich rangiert, während die anderen Bundesländer sich weit abgeschlagen in der unteren Tabellenhälfte befinden – jedoch angeführt durch den Spitzenreiter Baden-Württemberg. Diese Darstellung ist aus empirischer Sicht nicht nur fragwürdig, sie führt den geneigten – aber doch in Bezug auf die Interpretation von Statistiken in der Regel unversierten – Leser in die Irre. Würde man jedoch, wie von Seiten der ausführenden Wissenschaftler ermahnt, das Stichprobendesign mit den dazugehörigen Unsicherheiten



berücksichtigen, ergäbe sich ein anderes Bild – ein Bild, das in Abbildung 4 visualisiert ist. Hier ist, bezogen auf den bundesdeutschen Durchschnitt berechnet worden, welche Länder auf einem signifikant besseren, gleichen oder schlechteren Niveau liegen. Hierbei zeigt sich nun, dass es im PISA 2000-Bundesländervergleich zwei Spitzenreiter – nämlich Bayern und Baden-Württemberg – gibt und dass mit nur vier Ausnahmen die restlichen Bundesländer sich nicht signifikant vom gesamtdeutschen Mittel unterscheiden. Die hier dargestellte gruppierte Ergebnisdarstellung stellt mit Sicherheit weniger eine Basis für polemische Interpretationen dar, ist dafür jedoch empirisch haltbar und dient damit auch für erste weiterreichende fundierte Interpretationen, die schließlich in die Reform des deutschen Bildungswesens einfließen können.

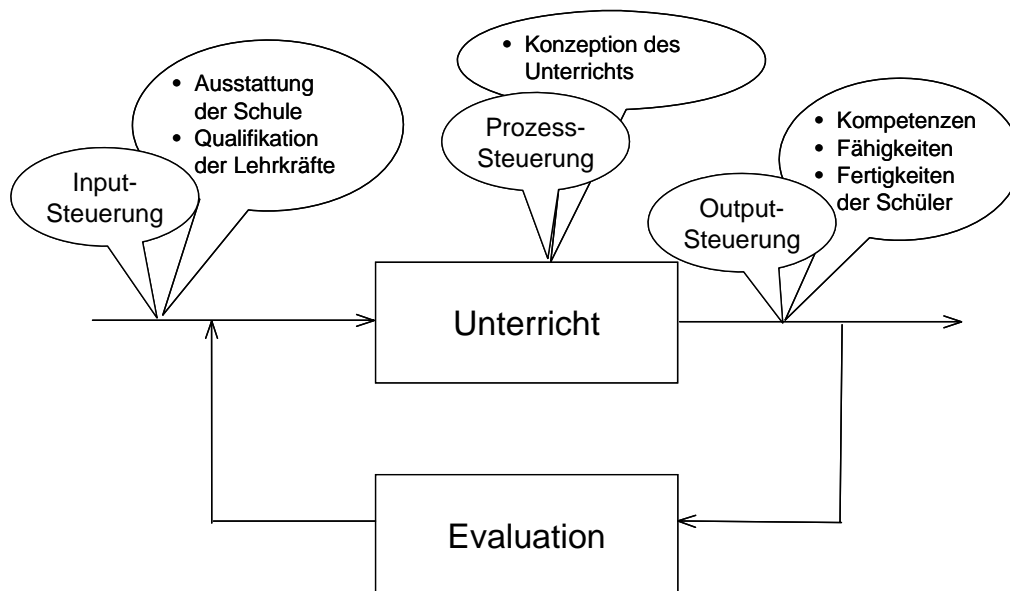
#### 4. Wandel in der Systemsteuerung

Die verstärkte Partizipation an internationalen Vergleichsstudien macht sich in einem tendenziellen Paradigmenwechsel bei der Diskussion um die Steuerung von Bildungsprozessen bemerkbar. Wie in Abbildung 5 gezeigt, lassen sich prinzipiell drei Steuerungsmodelle im Bildungssystem unterscheiden, die hier kursorisch in der Chronologie des Wandels skizziert werden sollen: Folgt man dem Prinzip der *Inputsteuerung*, geht man implizit davon aus, dass der Lern- bzw. der Bildungserfolg vornehmlich durch ein ausreichend gutes Angebot von Lerngelegenheiten determinierbar ist. Diese über lange Zeit auch in Deutschland praktizierte Steuerung macht sich bis heute durch die exzellente Ausbildung von Lehrkräften und eine umfassende Ausstattung der Bildungseinrichtungen bemerkbar. Ausgangspunkt für diese Überlegung ist die Auffassung, dass bei fachwissenschaftlich und erziehungswissenschaftlich professionalisierten Lehrkräften der Unterrichtsgegenstand bei den Bildungsempfängern ankommen muss. Wenn nicht, war davon auszugehen, dass es an dem Unvermögen der Bildungsempfänger, die gebotenen Lerngelegenheiten zu antizipieren, aber nicht an dem guten Angebot gelegen haben muss.

Im Sinne der *Prozesssteuerung* stand vor dem Hintergrund wahrgenommener differenzierter Rezeption von Lerngelegenheiten die Unterrichtsqualität im Zentrum der Betrachtung. Merkmale wie ‚Klarheit des Unterrichts‘, ‚effektive Zeitnutzung‘, ‚Anspruchsniveau‘, ‚Übungsintensität‘ oder ‚Instruktionstempo‘ lenkten den Blick von den lehrenden Akteuren auf ein optimiertes Lehrangebot (vgl. Einsiedler, 1997). Als Weiterentwicklung der *Inputsteuerung* wird bei der *Prozesssteuerung* nicht primär die Qualität des Angebots an Lerngelegenheiten, sondern auch die Qualität der Vermittlung in den Blick genommen. Aber auch dieses modifizierte Steuerungsmodell ist als ein Angebotsmodell angelegt, bei dem ein Lern(miss)erfolg den individuell Lernenden zugeschrieben wurde. Ob die optimierten Lernangebote tatsächlich von den Lernenden nutzbar sind, blieb außer Acht. Die Verantwortung für das Lernergebnis konnte somit immer noch den Lernenden zugeschrieben werden.

Erst der Einbezug outputorientierter Steuerungselemente ließ die konkrete Frage, was bei den Lernenden ankommt, zu. Wie in Abbildung 5 dargestellt, lässt erst die empirisch fundierte Diagnose der Kompetenzen, Fähigkeiten und Fertigkeiten der Schülerinnen und Schüler eine wirksame Modifizierung der Lerngelegenheiten zu. Diese *Outputsteuerung* ist jedoch nur dann sinnvoll, wenn rechtzeitig vor Abschlüssen auf den Lernstand geschaut wird, um noch intervenieren und somit die Schülerinnen und Schüler auf optimierten Wegen möglichst zum Erfolg führen zu können. Eine Überprüfung des Lernstandes im Rahmen zentraler Abschlussprüfungen entspricht nicht der Idee dieses Steuerungsprinzips. Nur wenn Raum für Korrekturen bleibt, ist der Blick auf den individuellen Lernerfolg zielführend.

Abbildung 5: Exemplarische Darstellung von Steuerungsmodellen am Beispiel von Unterricht

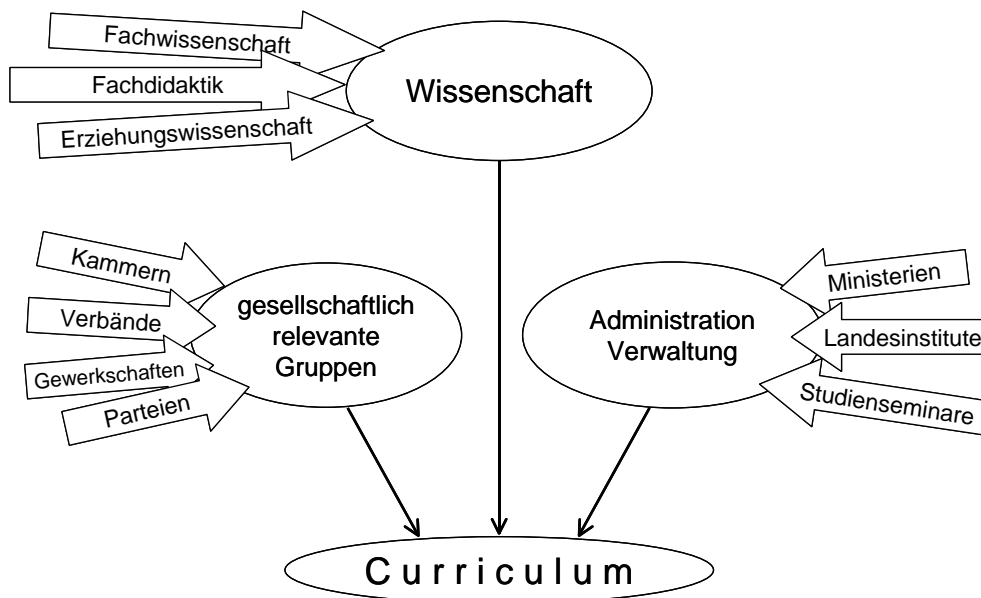


## 5. Was erwartet die deutsche Erziehungswissenschaft?

TIMSS wie auch PISA haben gezeigt, dass Untersuchungen aus groß angelegten Schulvergleichsuntersuchungen genügend Informationen liefern, um nachhaltig empirisch gestützte Veränderungen im Bildungssystem vornehmen zu können. Jedoch ergibt sich hierbei auch die Gefahr populärer Fehlinterpretationen, wenn ohne entsprechende Sachkenntnis an die oberflächliche Ursachenforschung gegangen wird. Gerade der öffentliche Druck, der durch die Veröffentlichung der PISA-Ergebnisse aufgebaut wurde, hat hierbei eine kontraproduktive Kraft entwickelt. Jeder konnte auf einmal alles mit PISA belegen – auch wenn die Befunde in der Regel von den mitarbeitenden Wissenschaftlern mit der notwendigen Methodenkenntnis und unter Vorsicht publiziert wurden, haben sich verschiedene Beteiligte punktuell nur die Informationen

herausgegriffen, die ihnen zur Absicherung ihrer Vorstellungen von Nutzen erschienen. Was lange Zeit – praktisch schleichend – seine Wirkung tat, war die normative Kraft der Untersuchung. Nachdem sich der erste Schrecken gelegt hatte, wurde vieles mit der Studie begründet, und noch mehr orientierte man sich an den in PISA erfassten Kompetenzen so, wie sie von der internationalen Forschergruppe im Rahmen des Literacy-Ansatzes operationalisiert wurden. Die Frage, die sich hier jedoch stellt ist, ob die enggeführte Orientierung im deutschen Bildungssystem angemessen ist – und vielleicht noch wichtiger, wo die Vorgaben der internationalen PISA-Autoren schon in innerdeutsche Entscheidungsprozesse einwirken, ohne dass ein entsprechender fachwissenschaftlicher Diskurs hierüber geführt wurde. Abbildung 6 gibt einen Eindruck, in welchen Wirkungsketten letztendlich Wissenschaft, Administration und Verwaltung und schließlich noch andere gesellschaftlich relevante Gruppen normativ auf das wirken, was in Deutschland gelehrt und schließlich auch gelernt werden soll. Es ist eine der größten Herausforderungen der Erziehungswissenschaft, diese impliziten Entscheidungsfindungen zu begleiten und – so wie sie es auch in den vergangenen Jahre stets gemacht hat – vor unserem eigenen bildungskulturellen Hintergrund kritisch zu hinterfragen und gegebenenfalls bei Fehlentwicklungen zu intervenieren.

Abbildung 6: Schematische Darstellung curriculabeeinflussender Institutionen



## 6. Ausblick

Entscheidungen aufgrund empirischer Befunde zu treffen liegt traditionell nicht unbedingt im Zentrum deutscher erziehungswissenschaftlicher und fachdidaktischer Forschung und administrativem und bildungspolitischem Handeln. Dies scheint sich in Folge der ‚empirischen Wende‘ langsam zu wandeln. Bei immer mehr Maßnahmen im

Bildungsbereich spielt die Evaluation, also die Überprüfung der Zielerreichung eine immer größere Rolle (vgl. z.B. <http://www.blk-foermig.uni-hamburg.de> (22.1.2006); Gogolin, Neumann & Roth, 2005). Im Bereich erziehungswissenschaftlicher Forschung ist zu beobachten, dass Fördergelder – z.B. durch die DFG – stärker an klares methodisches Vorgehen in den entsprechenden Projekten gebunden werden. In fast allen Universitäten wird bei Neuberufungen vermehrt darauf geachtet, dass in erziehungswissenschaftlichen Instituten bzw. Fakultäten auch der Aspekt der empirischen Pädagogik sichtbar wird.

In der Steuerung des Bildungswesens hat sich in fast allen Bundesländern durchgesetzt, Aspekte der Outputsteuerung, wie z.B. die Schulinspektion oder die Einführung von regelmäßigen Lernstandserhebungen zu implementieren. International vergleichende Forschung gibt hier Hinweise, dass dies zum Erfolg führen kann (vgl. z.B. Döbert, Klieme & Sroka, 2004; Döbert & Sroka, 2004). Wenig Forschung und wenig Wissen liegen aber darüber vor, wie die Schulinspektion am wirksamsten arbeitet, wie mit den Daten aus Leistungsuntersuchungen am besten umgegangen wird und wie diese zur positiven Schulentwicklung herangezogen werden können. Nahezu unbearbeitet ist das Feld der unterstützenden Systeme für erwartungswidrig schlecht abschneidende Schulen.

Für diese Bereiche sind in erziehungswissenschaftlicher Forschung – insbesondere auch in der international vergleichenden Forschung – große Anstrengungen zu unternehmen, um die Versäumnisse der letzten Jahrzehnte aufzuarbeiten. Damit dies reflektiert und nicht nur oberflächlich am Phänomen verhaftet geschieht, muss die Kompetenz verschiedener wissenschaftlicher Traditionen gebündelt werden.

#### *Anmerkung*

1. Der vorliegende Beitrag ist die Verschriftlichung des Eröffnungsvortrags auf der Jahrestagung der Sektion ‚International und Interkulturell Vergleichende Erziehungswissenschaft‘ vom 2.–4. März 2005 in Münster. Der Sprachduktus des Vortrags wurde beibehalten.

#### *Literatur*

- Adler, K., Behrend, T., Esser, B., Hilbig, M., Müller, V., Pannenberg, R., Plewnia, U., Reinke-Nobber, H., Röhl, T., Schattauer, G. & Zorn, T. (2002, 24. Juni). Ländervergleich. Gute Noten, schlechte Noten. *Focus*, S. 38–46.
- Baumert, J., Bos, W. & Lehmann, R. H. (Hrsg.). (2000a). *TIMSS/III. Dritte Internationale Mathematik- und Naturwissenschaftsstudie. Mathematische und naturwissenschaftliche Bildung am Ende der Schullaufbahn, Bd. 1: Mathematische und naturwissenschaftliche Grundbildung am Ende der Pflichtschulzeit*. Opladen: Leske + Budrich.
- Baumert, J., Bos, W. & Lehmann, R. H. (Hrsg.). (2000b). *TIMSS/III. Dritte Internationale Mathematik- und Naturwissenschaftsstudie. Mathematische und naturwissenschaftliche Bildung am Ende der Schullaufbahn, Bd. 2: Mathematische und physikalische Kompetenzen am Ende der gymnasialen Oberstufe*. Opladen: Leske + Budrich.

- Baumert, J., Klieme, E., Neubrand, M., Prenzel, M., Schiefele, U., Schneider, W., Stanat, P., Tillmann, K.-J. & Weiß, M. (Hrsg.). (2001). *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich*. Opladen: Leske + Budrich.
- Baumert, J., Lehmann, R. H., Lehrke, M., Schmitz, B., Clausen, M., Hosenfeld, I., Köller, O. & Neubrand, J. (1997). *TIMSS – Mathematisch-naturwissenschaftlicher Unterricht im internationalen Vergleich. Deskriptive Befunde*. Opladen: Leske + Budrich.
- Bos, W., Pietsch, M. et al. (2005). *Kompetenzen und Einstellungen von Schülerinnen und Schülern – Jahrgangsstufe 4* (hrsg. Freie und Hansestadt Hamburg, Behörde für Bildung und Sport). Hamburg: Bergmann.
- Döbert, H., Klieme, E. & Sroka, W. (Hrsg.). (2004). *Conditions of school performance in seven countries*. Münster: Waxmann.
- Döbert, H. & Sroka, W. (Hrsg.). (2004). *Features of successful school systems*. Münster: Waxmann.
- Einsiedler, W. (1997). Unterrichtsqualität und Leistungsentwicklung: Literaturüberblick. In F. E. Weinert & A. Helmke (Hrsg.), *Entwicklung im Grundschulalter* (S. 225–240). Weinheim: Beltz.
- Gogolin, I., Neumann, U. & Roth, H.-J. (Hrsg.). (2005). *Sprachdiagnostik bei Kindern und Jugendlichen mit Migrationshintergrund* (FörMig Edition, Bd. 1). Münster: Waxmann.
- Lehmann, R. H., Peek, R., Pieper, I. & Stritzky, R. v. (1991). Lesefähigkeit und Lesegewohnheiten von Achtkläßlern. *Deutschunterricht*, 44, 410–417.
- Peisert, H. (1967). *Soziale Lage und Bildungschancen in Deutschland*. München: Piper.
- Ross, K. N. (1997). Research and policy: A complex mix. *IIEP Newsletter*, 15 (1), 1–4.