



Ein halbes Jahrhundert internationale Schulleistungsstudien. Eine systematisierende Übersicht

Martin Goy

Technische Universität Dortmund

Isabell van Ackeren

Johannes Gutenberg-Universität Mainz

Knut Schwippert

Universität Hamburg

Abstract

This paper provides an overview over half a century of international large-scale assessments from the viewpoint of the German participation in such studies. Since the *Pilot Twelve Country Study* was launched in the late 1950s, the international interest in large-scale assessments has constantly grown so that these studies now form an integral part of educational research and policy in many countries worldwide. In Germany and its Federal States, however, this kind of research does not have such a continuous tradition. Despite participating in a number of early studies, Germany refrained from participation for most of the 1970s and 1980s. The rekindled interest since the mid-1990s reveals a change in educational approaches and reflects a certain backlog in view of empirical research. This review of five decades of international large-scale assessments retraces the German involvement in such studies and points out a number of relevant assessment results in international comparison. Further, this paper reveals how the studies have grown over time, not only in numbers of participating countries but also in terms of the subject areas they cover as well as the target groups they assess. The paper also reflects on the development of some major lines of argument that prevail in current discussions among and between educational researchers and practitioners, politicians, and the general public. The paper's summary attempts to mediate this discussion and includes a cautious outlook on a number of possible future areas of research in international large-scale assessments.

1. Einleitung

Im Rahmen von Schulleistungsuntersuchungen werden die Kenntnisse und Fertigkeiten von Schülerinnen und Schülern gemessen, um die Leistung von Schule und die

Reproduktion von außerschulischen Differenzierungen (z.B. Geschlecht, soziale und ethnische Herkunft) zu evaluieren sowie Einflussgrößen der Leistungserbringung und somit Ansatzpunkte möglichen Steuerungshandelns herauszuarbeiten. Einer breiteren Öffentlichkeit bekannt geworden sind im Verlauf des letzten Jahrzehnts insbesondere die TIMS-Studie über die mathematischen und naturwissenschaftlichen Leistungen, die PISA-Studie zur Grundbildung mit ihren unterschiedlichen Erhebungsschwerpunkten (Lesen, Mathematik, Naturwissenschaften) sowie die PIRLS/IGLU-Studie zur Lesekompetenz in der Primarstufe. Weniger bekannt ist hingegen, welche großflächig angelegten Schulleistungsstudien darüber hinaus seit rund 50 Jahren in internationaler Kooperation von Bildungsforschern durchgeführt werden sowie für die Zukunft geplant sind. An einigen dieser Studien hat sich Deutschland bzw. haben sich einzelne Bundesländer beteiligt. Viele der frühen Studien wurden von der *International Association for the Evaluation of Educational Achievement* (IEA) koordiniert; bis heute ist sie eine der zentralen Institutionen zur systematischen Bestandsaufnahme schulischer Lernerträge und ihrer Einflussfaktoren. Gleichwohl haben eine Reihe von IEA-Studien weniger öffentliche Aufmerksamkeit als TIMSS und PIRLS/IGLU erzielt. Neben der IEA ist es vor allem die OECD, die *Organisation for Economic Co-operation and Development*, die insbesondere seit der Jahrtausendwende internationale Leistungsstudien koordiniert bzw. an Erhebungen beteiligt ist. Die bekannteste, von der OECD organisierte Studie ist ohne Zweifel PISA. Weitere Organisationen, die sich an der Durchführung von internationalen Schulleistungsstudien beteiligen, sind unter anderem die Unesco, Statistics Canada und der US-amerikanische *Educational Testing Service* (für eine Übersicht über Organisationen, die international vergleichende Schulleistungsstudien durchführen vgl. Plomp, Howie & McGaw, 2003; Postlethwaite, 2004).

Vor diesem Hintergrund befasst sich die tabellarische Übersicht (siehe am Ende des Beitrags) über internationale Schulleistungsstudien mit dem Typ von Leistungsvergleichen, die im Verlauf eines Bildungsgangs für ausgewählte Alters- oder Schuljahrganggruppen, bei einer kompletten Population der ausgewählten Gruppe oder bei einer für sie repräsentativen Stichprobe durchgeführt werden. Die Quellenangaben zu den nachfolgend in ihrer Anlage und ausgewählten Befunden knapp skizzierten Studien können der zweiten Spalte der Übersicht sowie dem ausführlichen Literaturverzeichnis entnommen werden.

Der Überblick will einerseits ähnliche, bereits vorliegende Übersichten ergänzen und fortschreiben, wie sie etwa von Bos und Schwippert (2002), van Ackeren und Klemm (2002), van Ackeren (2002), Stanat und Lüdtke (2007) sowie von Schwippert und Goy (2008) erarbeitet wurden, und damit Orientierungswissen anbieten, das der Leser mit Hilfe des Quellenverzeichnisses vertiefen kann. Andererseits wird mit der Bestandsaufnahme, die auch aufzeigt, an welchen zentralen Studien die Bundesrepublik nicht beteiligt war, auch deutlich, wie spät hierzulande die empirisch angelegte Bildungsforschung Wertschätzung erfahren hat. Dies gilt insbesondere für die Mög-

lichkeiten einer evidenzbasierten Entwicklung von Schule, für die empirische Bildungsforschung eine Voraussetzung ist (vgl. Jornitz, 2008). Dabei verweist die unterschiedliche Interpretation der Befunde auf das Problem, dass Wissenschaft und Politik einer unterschiedlichen Rationalität folgen (vgl. Blömeke, 2003). Die Bereitschaft und Fähigkeit, diese je spezifische Handlungslogik zu erkennen und zu verstehen, bleibt sicherlich eine zentrale Aufgabe der Akteure an der Schnittstelle von Forschung und Steuerung, um in einen konstruktiven Dialog zu treten und Befunde einerseits anwendungsorientiert zu interpretieren, bzw. andererseits weniger interessengeleitet und rezeptologisch verkürzt nutzbar zu machen, im Sinne einer *evidence-based research and policy*. Schließlich wird mit der Übersicht deutlich, welche Bereiche bislang unzureichend mit Methoden der vergleichenden Schulleistungsforschung ausgeleuchtet sind und Forschungsdesiderate darstellen.

Large Scale Assessments, über die hier berichtet werden soll, gehen über die Leistungserhebung an der schulischen Einzelinstitution hinaus, sie beziehen sich auf Lernergebnisse, die an vielen hundert Schulen und für tausende Schülerinnen und Schüler repräsentativ und standardisiert erfasst werden. Sie haben in dieser Hinsicht ‚large scale‘- bzw. großflächigen Charakter, da sie in großem Umfang schulische Institutionen und deren Mitarbeiter (Schülerinnen und Schüler und bei verschiedenen Testdesigns auch Lehrkräfte, Schulleitungen und/oder Eltern) einbeziehen. Large Scale Assessments unterliegen einer zentralen Steuerung, da sie durch internationale Organisationen oder/und Forschungsgruppen in ihrer Entwicklung, Durchführung und Auswertung gesteuert und begleitet werden. Der Begriff ‚Assessment‘ meint schließlich – ganz allgemein gesprochen – eine Einschätzung und Bewertung einer vorgefundenen Situation oder eines Ergebnisses, womit die Funktion dieser großflächigen Leistungsmessungen angedeutet wird, die über eine reine Leistungserhebung hinausgeht (Hamilton, Stecher & Klein, 2002; van Ackeren, 2003).

2. Beginn internationaler Leistungsvergleiche Ende der 1950er Jahre

Bis zu den 1950er Jahren bezog sich die Vergleichende Erziehungswissenschaft vor allem auf die separate Beschreibung der Bildungssysteme in ausgewählten Ländern. Entwicklungspolitische Fragen führten dann in den 1950/60er Jahren zum Bewusstsein für den ‚hidden factor‘ des Humankapitals im Rahmen gesellschaftlicher und ökonomischer Entwicklungen. Diese sowie die technologische Entwicklung und die Globalisierung führten zu einer Fokussierung auf eine vorausschauende Bildungsplanung. Es vollzog sich ein Paradigmenwechsel von der Betrachtung schulischen Inputs (z.B. strukturelle und finanzielle Rahmenbedingungen) hin zur Erfassung des Lernertrags von Schule. Die *International Association for the Evaluation of Educational Research* (IEA) entwickelt seit nunmehr 50 Jahren Large Scale Assessments, in denen schulische Leistungen im Rahmen von Kontext- und Prozessvariablen erfasst werden. Dahinter steht bis heute das Ziel, Bildungspolitikern und Unterrichtspraktikern wichtige

Systemmonitoring- und Benchmark-Informationen zur Qualität des Schulsystems an die Hand zu geben und dabei zu helfen, ein Verständnis für mögliche Erklärungen der festgestellten Differenzen zwischen Bildungssystemen zu entwickeln und entsprechende Diskussionsgrundlagen zu liefern sowie Ansatzpunkte für Steuerungsmöglichkeiten zu erkennen.

Die erste in dieser Weise angelegte Studie war die *Pilot Twelve-Country Study*, mit der man in Form einer Pilotierung zwischen 1959 und 1962 in zwölf Teilnehmerstaaten herauszufinden suchte, inwieweit internationale Leistungsvergleichsstudien realisiert werden können. Die Leistungsmessungen bezogen sich in der Altersgruppe der 13-Jährigen bereits auf fünf Inhaltsbereiche, nämlich Lesen, Mathematik und Naturwissenschaften, aber auch Geographie und non-verbale Fähigkeiten. Mit der Studie konnte gezeigt werden, dass es möglich ist, vergleichend angelegte Studien dieses Umfangs und entsprechend angelegte Testinstrumente zu entwickeln. Ein zentraler Befund war der aus deutscher Perspektive im Ländervergleich große Leistungsunterschied der Geschlechter (Deutschland war jedoch nur mit dem Bundesland Hessen beteiligt). In die in den 1960er Jahren geführte Debatte zu Geschlechterdifferenzen beim Bildungszugang, die im Rahmen der Diskussion um die Ausschöpfung der Begabungsreserven geführt wurde (Dahrendorf, 1965; Peisert, 1967), flossen die empirischen Ergebnisse der *Pilot Twelve Country Study* allerdings nicht ein. Zudem wurden Geschlechterdifferenzen erst über ein Jahrzehnt später zum Gegenstand einer breiteren und systematischeren wissenschaftlichen Auseinandersetzung.

Methodische Weiterentwicklungen wurden mit der nachfolgenden *First International Mathematics Study (FIMS)* erreicht. Dazu gehörten eine optimierte Stichprobenziehung (Sampling) sowie ein umfangreicheres Testinstrumentarium, mit dem erfasst werden sollte, was Schülerinnen und Schüler zu einem bestimmten Zeitpunkt der Schulkarriere gelernt haben; dieses Prinzip der Testung an zentralen Punkten der Schullaufbahn ist bis heute bestehen geblieben. Die Hauptuntersuchung wurde 1963/64 in 12 Ländern mit einem damals enormen Stichprobenumfang von insgesamt 133.000 Jugendlichen der Sekundarstufen I (13-Jährige) und II (so genanntes voruniversitäres Jahr) durchgeführt. Inhaltlich hatte man sich mit dieser Studie für den Bereich Mathematik entschieden. Gleichwohl wurde im FIMS-Report vermerkt, dass Mathematik selbstverständlich nicht der einzige Index der Produktivität eines Systems sei, sich das Fach aber wegen der – im Vergleich zu anderen Inhalten – methodisch einfacheren Handhabung für eine erste Erhebung dieses Umfangs besonders eigne. Die Studie wurde – wie heute noch bei vergleichbaren Erhebungen üblich – durch Hintergrundbefragungen von Schülerinnen und Schülern, Lehrkräften und Schulleitung zu schulischen und außerschulischen Merkmalen ergänzt, eingebettet in den Kontext der nationalstaatlichen Kultur, Geschichte und der sozialen Strukturen.

Im Bereich der mathematischen Grundbildung schien das deutsche Schulsystem gut zu funktionieren; im Bereich der mathematischen Spitzenleistungen jedoch schon da-

mals nicht mehr: Die deutschen Abiturienten erreichten im internationalen Vergleich nur einen Platz im Mittelfeld: „Die Leistungen unserer Abiturienten im mathematisch-naturwissenschaftlichen Zweig sind bescheiden gegenüber den Leistungen, die in anderen Ländern von sehr viel jüngeren Schülern zur Zeit ihres Übergangs an die Universität erzielt werden“ (Schultze & Riemenschneider, 1967, S. 34). Auch die deutschen 13-Jährigen aus Volks-, Realschulen und Gymnasien schnitten mit Rang 5 von insgesamt 12 Systemen nur noch mittelmäßig ab hinter einer Spitzengruppe bestehend aus Israel, Japan und Belgien. Zudem konnte die Studie für alle Länder Ungleichheiten zwischen unterschiedlichen Schülergruppen aufzeigen, die den Schulsystemen Probleme bereiteten. Im Hinblick auf die Frage der Chancengerechtigkeit von Bildungssystemen zeigt die FIMS-Studie (Postlethwaite, 1968, S. 64 f.) schon damals deutlich die starke sozioökonomische Selektivität des deutschen Bildungssystems auf: Der dort berechnete Indexwert der sozialen Selektivität, der auf einem Vergleich des Anteils von Abiturienten aus der sozialen Oberschicht mit dem Anteil von Abiturienten aus den unteren sozialen Schichten beruht, war in der Bundesrepublik Deutschland im internationalen Vergleich am höchsten und dabei um ein mehrfaches höher als in allen übrigen Teilnehmerstaaten (zu den Besonderheiten der deutschen Stichprobe siehe Übersicht).

In Deutschland fielen diese Befunde in eine Zeit, in der Georg Picht (1964) den deutschen Bildungsnotstand als ökonomische Krise analysierte und Ralf Dahrendorf (1965) aus bürgerrechtlicher Perspektive aufzeigte, wie massiv das Bildungssystem an der Reproduktion sozialer Ungleichheit beteiligt ist, ohne dass die international vergleichenden Befunde aus FIMS und der nachfolgenden naturwissenschaftlichen FIS-Studie Eingang in die nationale Diskussion fanden.

Im Anschluss an FIMS ging das 1970/71 durchgeführte *Six-Subject Survey* der Frage nach, inwieweit sich identifizierte Einflussfaktoren mathematischer Bildungsprozesse auch auf andere Fächer übertragen lassen. Deshalb führte die IEA eine Reihe von Studien in unterschiedlichen Fächern durch: Leseverständnis, Naturwissenschaften, Literatur, Französisch und Englisch als Fremdsprachen sowie *Civic Education* (staatsbürgerliche bzw. politische Bildung). Neben der Sekundarstufe I und II wurde erstmals auch die Primarstufe berücksichtigt. Mit der *Six-Subject*-Studie konnten einige neue Prädiktoren schulischer Leistung herausgearbeitet werden, so etwa Interessens-, Motivations- und Einstellungsvariablen sowie Lehrmethoden. Deutschland war mit einer größeren Anzahl Bundesländer (vgl. Übersicht) an der naturwissenschaftlichen *FIS-Studie*, an der *Study of English as a Foreign Language* sowie an der *Study of Civic Education* beteiligt.

Die *FISS*-Ergebnisse belegten in den drei getesteten Kohorten ein bedenkliches ‚Abrutschen‘ von guten Leistungen der älteren Schülerinnen und Schüler über mittelmäßige Befunde für die mittlere Kohorte bis hin zu unterdurchschnittlichen Ergebnissen unterer Jahrgangsstufen: Während die älteren deutschen Schülerinnen und Schüler

(aus den Abschlussklassen der Sekundarschulen) in einem Feld von 14 Ländern – darunter Australien, England, Finnland, Schweden und die USA – nach Neuseeland den zweiten Rang erreichten, lagen die 14-Jährigen ‚nur‘ noch im oberen Mittelfeld auf Rang 5 – hinter Japan, Ungarn, Australien und Neuseeland. Die 10-Jährigen dagegen lagen auf dem drittletzten Rang: Japan, Schweden, Finnland, die Niederlande und andere lagen – zum Teil mit deutlichem Abstand – vor den Kindern aus der BRD. Unter dem Titel ‚Im internationalen Vergleich schneidet das Bildungswesen der Bundesrepublik miserabel ab: Die deutschen Schüler auf dem letzten Platz‘ schrieb Hayo Matthiesen in der ZEIT im September 1974: „Vor allem gegen die Effektivität des von vielen so gelobten traditionellen dreigliedrigen Schulwesens liefert die Studie Belege. Rigoros wie sonst kaum irgendwo in der Welt liest dieses System die Schüler aus, so dass sich bei uns in den Abschlussklassen der Sekundarstufe II nur neun Prozent eines Jahrgangs befinden; in den USA sind es 75 %, 70 in Japan, 47 in Belgien oder 29 in Australien“ (S. 17).

An der Studie zu *English as a Foreign Language* (1968–1973) war Deutschland im Bereich Leserverständnis (14-Jährige) vertreten. Ein wichtiges Ergebnis war – bezogen auf den Vergleich aller Teilnehmerstaaten – der Einfluss von Lernzeit auf die schulischen Lernerträge (Unterrichtszeit, Studienjahre, Zeitpunkt des Beginns mit Englisch in der Schullaufbahn). Mit der zeitgleich durchgeführten *Study of Civic Education* wurde untersucht, wie und in welchem Umfang die Ziele politischer Bildung in den Teilnehmerstaaten erreicht werden und welchen Einfluss außerschulische Faktoren haben (Familie, Medien, Peers u.a.). Zu dem erfreulichen Befund für Deutschland gehörte ein sehr gutes Abschneiden im Bereich des politischen Wissens in den unterschiedlichen Altersgruppen. Neben den kognitiven Tests zu Fragen aus Teilbereichen der politischen Bildung wurden über Fragebögen auch politikbezogene Einstellungen zu drei Hauptdimensionen erfasst: demokratische Werte, Loyalität gegenüber der eigenen politischen Führung sowie politisches Interesse und politische Partizipation. In der Population der 14-Jährigen zeigten sich unter den deutschen Teilnehmern besonders hoch ausgeprägte demokratische Werte und eine geringe Loyalität gegenüber der politischen Führung bei einem insgesamt relativ schwach ausgeprägten politischen Interesse. In der Gruppe der Schüler im letzten voruniversitären Schuljahr waren die Einstellungen in der deutschen Stichprobe generell ähnlich. Anders als bei den 14-Jährigen war allerdings das politische Interesse der Schüler am Ende der Sekundarschulzeit ausgesprochen stark ausgeprägt, wobei zu beachten ist, dass es sich bei der deutschen Stichprobe um ein im internationalen Vergleich hoch ausgelesenes Segment der Schülerschaft, nämlich Gymnasiasten, handelte (Torney, Oppenheim & Farnen, 1975; Avenarius et al., 2003, S. 237).

3. Deutsche Zurückhaltung in den 1970er und 1980er Jahren

An den weiteren Studien der *Six-Subject-Studie* war Deutschland nicht beteiligt; andere Länder beteiligten sich konsequenter an den Untersuchungen, so z.B. Belgien, England, Frankreich, Finnland, Neuseeland, Niederlande, Schottland, Schweden und die USA (vgl. van Ackeren, 2003). Zugleich hatten die mittelmäßigen Befunde, etwa bei FIMS und FISS, keinen Einfluss auf bildungspolitisches Handeln, obgleich sich die Bildungsforschung international mit Konsequenzen beschäftigte, so z.B. Benjamin S. Bloom mit seinem Beitrag ‚Implications of the IEA Studies for Curriculum and Instruction‘ aus dem Jahr 1974 sowie Alan C. Purves und Daniel U. Levine (1975) mit ihrem Band ‚Educational Policy and Educational Assessment. Implications of the IEA Surveys of Achievement‘.

Die *Second International Mathematics Study (SIMS)* sowie die *Second International Science Study (SISS)* wurden aus deutscher Sicht ebenfalls ausgelassen. Eine mögliche Erklärung für die Abstinenz Deutschlands kann in der bis dahin „eher geisteswissenschaftlichen Tradition deutscher Pädagogik“ (Bos & Postlethwaite, 2001, S. 253) und im Methodendiskurs der Kritischen Theorie (seit den 1920er Jahren) und später der Kritischen Erziehungswissenschaft (seit den 1960er Jahren) zu finden sein. Eine weitere Ursache der deutschen Zurückhaltung könnte die diskutierte Frage nach der vernünftigen Relation zwischen den Ergebnissen derartiger Studien gewesen sein, deren Interpretation damals noch mit methodischen Unsicherheiten behaftet war (vgl. Schultze & Riemenschneider, 1967, S. 32).

Die systematische empirische Schulleistungsforschung hat erst seit dem TIMSS-Schock (s.u.) einen anderen Stellenwert erhalten. Trotz der deutschen Zurückhaltung bei internationalen Vergleichsstudien gab es doch deutsche Arbeitsgruppen, insbesondere um Franz E. Weinert und Andreas Helmke, die auch in den 1980er Jahren an internationalen und nationalen Vergleichsstudien beteiligt waren. Die *Classroom Environment Study (CES)*, die zwischen 1981 und 1983 durchgeführt wurde (in Deutschland jedoch zwei Jahre später), wurde hierzulande als *Münchener Hauptschulstudie* bekannt. In methodischer Hinsicht wurde die Idee aufgegriffen, die schulische Leistungsentwicklung im Rahmen von Large Scale Assessments über einen längeren Zeitraum zu erfassen. Erstmals wurde der Versuch unternommen, dieselben Schülerinnen und Schüler zwei Mal zu testen. Zudem bezog sich die Analyse nun stärker auf proximale Variablen, die das Lernen beeinflussen, nämlich den Klassenkontext und den Unterricht. Im Zentrum der Studie stand die Herausarbeitung von Unterschieden der instruktionalen Praxis und ihrer Bedeutung für Lernerträge. Zu den zentralen Befunden zählt beispielsweise, dass über die Teilnehmerstaaten hinweg wenig Unterrichtszeit in kleineren Gruppen stattfand. Zu den häufigsten Aktivitäten im Klassenraum zählten z.B. der Lehrervortrag, das Stellen und Beantworten von Fragen sowie Stillarbeit.

Auch an der *Written Composition Study* war mit Hamburg nur ein deutsches Bundesland beteiligt. Die Studie wurde Mitte der 1980er Jahre durchgeführt und ergänzte Daten, die bereits im Rahmen der *Six-Subject-Studie* erhoben wurden. Im Fokus standen das Lehren und Lernen im Aufsatzunterricht hinsichtlich unterschiedlicher Formen des Schreibens (z.B. reflektierend, philosophierend, argumentierend, literarisch) einschließlich unterschiedlicher Einstellungen und Konventionen im Ländervergleich. In die Erhebung einbezogen waren Schülerinnen und Schüler am Ende der Primarstufe, zum Abschluss der Pflichtschulzeit sowie am Ende der Oberstufe. Ein zentraler Befund waren Geschlechterdifferenzen in der Bewertung schriftlicher Leistungen zugunsten der Mädchen, insbesondere in der weiblich dominierten Primarstufe und Sekundarstufe I.

Eine weitere Bildungsstufe rückte in den 1980er Jahren in den Fokus der wissenschaftlichen Aufmerksamkeit, auch aus deutscher Perspektive: Das *Pre-Primary Project (PPP)* wurde vor dem Hintergrund initiiert, dass zu dieser Zeit zunehmend mehr Kinder in den verschiedenen Bildungssystemen vorschulische Einrichtungen besuchten und man mehr über die kognitive, soziale und emotionale Entwicklung der Kinder dieses Alters in unterschiedlichen Einrichtungen mit ihren unterschiedlichen Konzepten in Erfahrung bringen wollte. Die Studie wurde Mitte der 1980er Jahre als Längsschnittuntersuchung angelegt und im Jahr 2003 abgeschlossen. Die in drei Phasen angelegte Studie konzentrierte sich zunächst auf die vergleichende Beschreibung der nationalen Strategien vorschulischer Betreuung in Form von Länderprofilen. In dieser ersten Phase, an der Deutschland mit einer repräsentativen Stichprobe beteiligt war, wurde zudem eine Bestandsaufnahme der familialen Förderstrategien in den Ländern durchgeführt. An den Phasen zwei und drei war die Bundesrepublik nicht mehr beteiligt, so dass u.a. Daten zur kognitiven und sprachlichen Entwicklung und zu entsprechenden Einflussgrößen aus dieser Studie für Deutschland fehlen.

An der *IAEP*-Untersuchung (*International Assessment of Educational Progress*) war das Land wiederum nicht beteiligt. Im Rahmen von IAEP wurden mathematische und naturwissenschaftliche Leistungen 9- bis 13-jähriger Schülerinnen und Schüler in den USA und 19 weiteren Ländern getestet. Die Erhebung, die mit Instrumenten des amerikanischen NAEP-Programms (*National Assessment of Educational Progress*) arbeitete, lud auch zum Experimentieren mit innovativen psychometrischen Methoden (wie z.B. allgemein mit probabilistischen Modellen und im Speziellen mit *Plausible Values*) ein, um die Instrumentenentwicklung voranzubringen.

4. Wiedereinstieg in den 1990er Jahren

Ungeachtet der insgesamt nicht überzeugenden deutschen Leistungen bei diesen ersten Vergleichsuntersuchungen beteiligte sich die BRD erst nach einigen Jahren zu Beginn der 1990er Jahre mit der internationalen Lesestudie IRLS wieder repräsentativ an einem solchen Projekt im schulischen Bereich. Im Rahmen der *International Reading*

Literacy Study (IRLS) wurden die Daten zu Leseleistungen, freiwilligen Leseaktivitäten und landesspezifischen Strategien der Leseförderung 1990/91 für zwei Altersgruppen (in Deutschland: 9- und 14-Jährige) erhoben. Zugleich wurde mit validen Instrumenten die Grundlage für ein Monitoring von Lesekompetenz im Zeitverlauf gelegt.

In der jüngeren IRLS-Zielgruppe (Drittklässler) und der älteren Altersgruppe (8. Jahrgangsstufe) war zum Testzeitpunkt das Leseverständnis in beiden Teilen Deutschlands (die Studie wurde vor der Wiedervereinigung begonnen) insgesamt nur durchschnittlich ausgeprägt; besonders stark fiel dies im Vergleich zu ökonomisch entwickelten Ländern auf. Ebenfalls ungünstig sahen die Ergebnisse im Bereich der Leistungsstreuung aus. Es schien folglich am Beginn der Schulkarrieren nicht zu gelingen, einen guten Mindeststandard für die Altersgruppe insgesamt zu sichern. Die Ergebnisse der IRL-Studie wurden allerdings erst rückwirkend nach der TIMSS-Rezeption Mitte der 1990er Jahre von der Fachöffentlichkeit wahrgenommen; vermutlich erschienen mathematisch-naturwissenschaftliche Kompetenzen in der ökonomisch orientierten Standortdiskussion wichtiger als Leseverständnis und -gewohnheiten. Mittlerweile gibt es ein Bewusstsein dafür, dass Leseverständnis eine grundlegende Voraussetzung für andere Kompetenzbereiche darstellt, wie es durch PISA einmal mehr bestätigt wurde.

Neben der reinen Lernstandserfassung bei über einer halben Millionen Schülerinnen und Schüler mehrerer Altersgruppen machten zusätzliche Untersuchungskomponenten zu schulischen wie außerschulischen Kontextbedingungen die Studie *TIMSS*, die *Third International Mathematics and Science Study*, Mitte der 1990er Jahre zur bis dahin umfangreichsten Schulleistungsstudie. Für Deutschland waren ihre Ergebnisse insgesamt – für alle erfassten Bildungsgänge, für die unterschiedlichen Kompetenzbereiche und für alle Bundesländer – im internationalen Vergleich derart mittelmäßig, dass sie eine heftige Debatte zur Leistungsfähigkeit des deutschen Schulsystems einleiteten. Eine Folge dieser Diskussion war es, dass die in Deutschland bis dahin verbreitete Distanz gegenüber derartigen großflächigen Vergleichsstudien innerhalb weniger Jahre aufgegeben, ja nahezu in ihr Gegenteil verkehrt wurde. Eine wesentliche Wirkung der Erhebung war damit die Zäsur, die durch TIMSS im deutschen Umgang mit Large Scale Assessments erreicht wurde.

1999 wurden im Rahmen einer weiteren TIMSS-Erhebungswelle vertiefende Analysen zum Unterricht, zum familialen Kontext und zu Charakteristika der Schulen durchgeführt. Zudem ermöglicht der 4-Jahres-Zyklus von TIMSS (1999, 2003, 2007) Trendanalysen; insofern steht das Akronym TIMSS mittlerweile für *Trends in International Mathematics and Science Study*. Aufgrund der deutschen Beteiligung an PISA (s.u.) ist Deutschland erst nach über 10 Jahren wieder an der Erhebung 2007, hier im Grundschulbereich, beteiligt. Im Hinblick auf andere Schulstufen hat sich Deutschland auf die PISA-Erhebungen konzentriert. TIMSS brachte zudem eine methodische Innovation im Bereich der Large Scale Assessments: Mit der *TIMSS 1995 Video Study*

(vgl. Stigler, Gallimore & Hiebert, 2000) wurde der Mathematikunterricht in drei Vergleichsländern – in Japan, Deutschland und den USA – aufgezeichnet, um einen systematischen transkulturellen Vergleich von fachbezogenen Unterrichtsprozessen zu ermöglichen und über die Erfassung von Unterrichtsskripts als kulturspezifische Inszenierungs- und Verlaufsmuster der Unterrichtsführung einen Beitrag zur Erklärung von gemessenen Kompetenzunterschieden zu leisten. Erhoben wurden die Daten in 231 zufällig ausgewählten Klassen der Jahrgangsstufen 7 und 8. Die auf diese Weise gewonnenen Daten ermöglichten eine qualitative Beschreibung der typischen Verlaufsmuster des Unterrichts im zeitlichen Verlauf der Unterrichtsstunde unter Berücksichtigung der eingesetzten Lehr- und Sozialformen. Am *TIMSS-R(epeat) Project* (1998–2004) war Deutschland jedoch nicht beteiligt, da sich parallel zu TIMSS die nachfolgend vorgestellte PISA-Studie etablierte, an der sich Deutschland stattdessen beteiligte.

Die mittelmäßigen TIMSS-Resultate wurden von der Kultusministerkonferenz (KMK) zum Anlass genommen, den Beschluss zu fassen, sich zukünftig an regelmäßig durchgeführten internationalen Vergleichsuntersuchungen, wie etwa an der von der OECD organisierten PISA-Studie (*Programme for International Student Assessment*), zu beteiligen. Getestet wurden 15-Jährige zu drei Erhebungszeitpunkten mit unterschiedlicher Schwerpunktsetzung in den Bereichen Leseverständnis, Mathematik und Naturwissenschaften sowie *Cross-Curricular Competencies* (Lernstrategien, Problemlösen), ergänzt durch sozioökonomische Hintergrundmerkmale und Variablen zu schulischen Lernbedingungen. Die Ergebnisse dieser Untersuchung wiederum fanden nicht nur in Fachkreisen, sondern auch in der allgemeinen Öffentlichkeit starkes Interesse. Das vergleichsweise schlechte Abschneiden der deutschen Schülerinnen und Schüler löste nach TIMSS nun auch einen ‚PISA-Schock‘ aus, und die Studie ist seitdem in der deutschen Bildungsdebatte in aller Munde. Das große Interesse gegenüber PISA im speziellen bzw. internationalen Vergleichsuntersuchungen im allgemeinen dauert – auch bis in den dritten PISA-Zyklus (2006) hinein – an, wenn auch mit leicht abnehmender Tendenz.

Besonderes Aufsehen haben Befunde aus dem internationalen Vergleich bezüglich der sozialen Abhängigkeit des Kompetenzerwerbs erregt. Für alle Teilnehmerstaaten offenbaren die PISA-Studien einen systematischen Zusammenhang zwischen sozialer Herkunft und schulischer Leistung in allen getesteten Kompetenzbereichen: Schülerinnen und Schüler aus Familien gehobener Sozialschichten erreichen am Ende der Pflichtschulzeit durchschnittlich bessere Leistungen und höhere Bildungsabschlüsse in den getesteten Domänen als Jugendliche aus sozial schwachen Familien. Allerdings ist dieser Zusammenhang in Deutschland besonders stark ausgeprägt; hierzulande geht die überdurchschnittliche Bildungsungleichheit mit nur durchschnittlichen Leistungen einher. Der Zusammenhang zwischen sozialer Lage und Kompetenzerwerb scheint zudem ein kumulativer Prozess zu sein, der vor der Grundschule beginnt und an Naht-

stellen des Bildungssystems verstärkt wird. Die in den Ländern bestehenden Unterschiede in der Relation zwischen Schülerleistungen und sozioökonomischem Status zeigen, dass die Differenzen erheblich variieren, was zugleich bedeutet, dass sie nicht unvermeidlich sind. Somit zeigen die internationalen Vergleichsuntersuchungen, dass die angestrebte Chancengerechtigkeit gerade in Deutschland bislang nicht erreicht wurde und die hierzulande beklagte Heterogenität in anderen Ländern weniger Probleme zu bereiten scheint. Es erscheint bedenkenswert, dass gerade in denjenigen Ländern, die auf eine Heterogenität im Klassenzimmer als Herausforderung bzw. Ressource setzen, die getesteten Schülerinnen und Schüler tendenziell auch bessere Leistungen verzeichnen.

Ein weiteres Augenmerk wurde mit PISA auf die Situation von Jugendlichen mit Migrationshintergrund gelegt. Es war erst der Wechsel vom Ausländer- zum Migrantenkonzept, der – angeregt durch entsprechende Analysen im Rahmen der TIMS-Studie (Schwippert & Schnabel, 2000) – durch die PISA-Studien durchgesetzt wurde und das ganze Ausmaß der Probleme in diesem Bereich sichtbar gemacht hat. Statt ausschließlich Jugendliche ohne deutsche Staatsangehörigkeit zu betrachten, berücksichtigt man nun auch Schülerinnen und Schüler, deren Väter oder Mütter zugewandert sind. Hierbei zeigte sich, dass gerade Jugendliche aus Familien, in denen beide Elternteile im Ausland geboren wurden, auffällige Defizite im schulischen Erfolg aufweisen. Auch nach Kontrolle der Sozialschichtzugehörigkeit blieben die Unterschiede zwischen den Jugendlichen ohne und mit einem bzw. beiden im Ausland geborenen Elternteilen signifikant. Die Unterschiede in der Lesekompetenz von 15-Jährigen aus Familien mit und ohne Migrationshintergrund sind in Deutschland im Vergleich mit Staaten, die wie Deutschland bedeutende Zuwanderungen aufweisen, besonders ausgeprägt.

Wenn die Schlagworte ‚Chancengleichheit‘ bzw. ‚Chancengerechtigkeit‘ im Zusammenhang mit Befunden aus der PISA 2000-Untersuchung diskutiert werden, stellt sich stets die Frage, ob der Zusammenhang von sozialem Hintergrund und Schulerfolg (sozialer Gradient) bereits ein Phänomen der Grundschule ist. Hierzu liefert PIRLS, die *Progress for International Reading Literacy Study* Daten, die in Deutschland unter dem Akronym IGLU (*Internationale Grundschul-Lese-Untersuchung*) bekannt wurde und Lesekompetenzen am Ende der Grundschulzeit als Trendstudie in einem 5-Jahres-Zyklus erhebt (2001, 2006 sowie in Planung: 2011). Die bislang vorgelegten IGLU-Befunde liefern keine Hinweise auf generelle Leistungsdefizite. Gleichwohl zeichnen sich für einige Probleme am Ende der Sekundarstufe I die Anfänge bereits in der Grundschule ab. Bei dem Vergleich des sozialen Gradienten der in IGLU 2001 getesteten vierten Jahrgangsstufe mit demjenigen der PISA-Erweiterungsstichprobe der neunten Jahrgänge zeigte sich, dass der ausgeprägte Zusammenhang auch in der Grundschule nachzuweisen ist, hier jedoch signifikant geringer ausfällt (Schwippert, Bos & Lankes, 2003). Im internationalen Vergleich ließ sich für die Grundschülerin-

nen und Grundschüler zudem feststellen, dass sie in Deutschland Schulen besuchen, in denen zwar relativ geringe Unterschiede zwischen Jungen und Mädchen bestehen, jedoch bedeutende Unterschiede zwischen Kindern aus bildungsnahen und solchen aus bildungsferneren Elternhäusern und auffällig große Unterschiede zwischen Kindern aus Elternhäusern ohne und solchen aus Elternhäusern mit Migrationsgeschichte (ebd.). Hieran lässt sich ablesen, dass die in der PISA 2000-Studie sowie in den nachfolgenden Erhebungswellen ermittelten Defizite des Bildungssystems sich bereits in der Grundschule abzeichnen. Auch bei IGLU 2006 zeigt sich für die deutschen Schülerinnen und Schüler ein im internationalen Vergleich nach wie vor deutlicher Zusammenhang zwischen ihrer sozialen Herkunft und der getesteten Lesekompetenz: Kinder aus dem oberen Quartil der sozialen Herkunft haben in Deutschland einen deutlichen Leistungsvorsprung von 67 Punkten vor Kindern aus dem unteren Quartil. Dieser Vorsprung, der in etwa dem Lernzuwachs von einem Lernjahr entspricht, fällt signifikant größer aus als im internationalen Mittel und im Mittel der IGLU-Teilnehmerstaaten aus der OECD (Bos, Schwippert & Stubbe, 2007, S. 239 ff.). Analysen, die den Leistungsvorsprung von Kindern aus bildungsnahen und bildungsfernen Schichten bei IGLU 2001 und IGLU 2006 vergleichend untersuchen, zeigen, dass sich dieser Unterschied nicht signifikant verringert hat (ebd., S. 235 f.).

5. Erweiterung von Inhaltsbereichen und Bildungsstufen

In der Bestandsaufnahme der Large Scale Assessments zeigt sich schließlich eine Ausweitung der betrachteten Inhaltsbereiche sowie eine Erweiterung hinsichtlich des Bildungsstufenbezugs über Schule hinaus; auch die Lehrerbildung wird thematisiert und in ihrer Ausgestaltung und Wirkung empirisch fundiert. *Computer Literacy* als eine für die Informationsgesellschaft notwendige Kulturtechnik hat Eingang in die Erhebungen gefunden. Ebenso werden in der Reaktion auf spezifische gesellschaftliche Probleme (Immigrationsprobleme, Rassismus, Intoleranz, Globalisierung etc.) Studien zur *Civic Education* angesichts ihrer Bedeutung für Demokratie und Zivilgesellschaft ausgebaut. Über die Bestandsaufnahme entsprechender Kompetenzen der Schulbevölkerung hinaus zeichnet sich ein zunehmendes Interesse für das ‚Humankapital‘ der Arbeitsbevölkerung ab, auf das Länder im globalen Wettbewerb zurückgreifen können. Die Evaluation von Ausbildungssystemen wird ebenso aufgegriffen wie Fragen lebenslangen Lernens. Eine Ausweitung des Forschungsinteresses gibt es schließlich auch hinsichtlich des vorschulischen Bereichs. Diese Entwicklungen werden nachfolgend skizziert.

5.1 Inhaltsbereiche

In der Verlängerung der *Study of Civic Education* (s.o.) hat sich Deutschland an der 1999 durchgeführten *Civic Education Study (CivEd)* zur Erhebung politischen

Wissens und politischer Einstellungen beteiligt. In der CivEd-Studie wurden Auskünfte über Art, Umfang und mögliche Ursachen des politischen Wissens von 14-jährigen Jugendlichen – an der Zusatzerhebung unter 16- bis 18-Jährigen nahm Deutschland nicht teil – mittels qualitativer und quantitativer Messmethoden gegeben, wofür in einer ersten Phase zwischen 1995 und 1997 Experteninterviews und curriculare Analysen durchgeführt und in einer zweiten Phase politisches Wissen, Einstellungen, Engagement sowie außerschulische, z.B. familiäre und mediale Einflüsse getestet und erfasst wurden. Die wie bei PISA im Jahr 2001 veröffentlichten Resultate der Studie im Vergleich von 28 Ländern haben im Gegensatz zu PISA wenig Aufmerksamkeit erfahren. Die deutschen Schülerinnen und Schüler nahmen im Kompetenzbereich ‚Politisches Wissen/Verstehen‘ einen mittleren Rang ein, der dem internationalen Mittelwert entspricht. Die im Jahr 2009 anstehende *International Civic and Citizenship Education Study (ICCS)* wird ohne deutsche Beteiligung stattfinden.

Eine weitere Studie blieb von der allgemeinen Öffentlichkeit vergleichsweise wenig beachtet: *CompEd*, die zweiphasige *Computers in Education Study*, mit Erhebungen in den Jahren 1989 und 1992. Erfasst wurden – auch im Zeitverlauf – u.a. der Computereinsatz im Unterricht, die Verfügbarkeit von PCs und ihr Einfluss auf das Unterrichtsgeschehen, auf Schule als Organisation und auf das Curriculum sowie funktionales Computerwissen und Kompetenzen im Umgang mit neuen Informationstechnologien. Die Resultate machten deutlich, dass es einer größeren Aufmerksamkeit für die Frage der Möglichkeiten des Computereinsatzes im Unterricht im Sinne einer qualitätssteigernden Wirkung bedurfte. Einmal mehr waren Schülerinnen und Schüler in der Primar- und den Sekundarstufen involviert. Computer waren zum Testzeitpunkt insbesondere selbst Gegenstand der Instruktion, indem ihre Anwendung und Handhabung thematisiert wurde. Sie wurden vergleichsweise selten als Medium in den anderen Fächern genutzt. Insbesondere außerschulische Lerngelegenheiten im Umgang mit Computertechnologie bildeten den Rahmen für den Wissenserwerb und die Kompetenzentwicklung von Jugendlichen in diesem Bereich. Das kontinuierliche Interesse der IEA an *Computer Literacy* setzte sich 2000/01 mit der *Second Information Technology in Education Study (SITES)* fort. Zentrale Erhebungsschwerpunkte waren die Verfügbarkeit von Computertechnologie, die Nutzung dieser Ressourcen und ihre Integration in den Unterrichtsprozess. SITES wurde 2005 fortgeführt, jedoch ohne deutsche Beteiligung.

5.2 Bildungsstufen

Im Vorschulbereich existieren nur wenige Projekte, die zum einen die Frage der Bedeutung des Elementarbereichs untersuchen, zum anderen Qualitätsmerkmale definieren und operationalisieren. So suchte das internationale *Pre-Primary Project (PPP)* der IEA nach möglichen Zusammenhängen zwischen Erziehungsverhalten der Eltern, Vorschulerfahrungen und den ersten Schulerfahrungen. Auf diese Weise wird die Be-

deutung des Vorschulunterrichts erforscht und der Frage nachgegangen, inwieweit seine Form und Qualität Einfluss auf späteren Schulerfolg nimmt. An dem Teil der Studie, der sich explizit mit der Vorschule beschäftigt, ist Deutschland allerdings nicht beteiligt.

Über die Erfassung unterschiedlicher Kompetenzbereiche im Schulalter hinaus gibt es ein zunehmendes Interesse am Bildungsstand der erwachsenen Bevölkerung. Das *International Adult Literacy Survey (IALS)* wurde im Herbst 1994 von der OECD und Statistics Canada mit dem Ziel durchgeführt, vergleichende Daten zur Partizipation an Bildung und Ausbildung im Erwachsenenalter zur Verfügung zu stellen. Bei IALS handelte es sich um die erste internationale Untersuchung von Grundqualifikationen Erwachsener. Drei Arten von Grundqualifikationen wurden gemessen: die Fähigkeit zum Umgang mit Texten (*prose literacy*), mit schematischen Darstellungen (*document literacy*) und mit Zahlen (*quantitative literacy*). Die Befunde sprachen für einen deutlichen Zusammenhang zwischen der Grundbildung und dem ökonomischen Potential eines Landes. Für Deutschland wurde deutlich, dass über 14 Prozent der Erwachsenen über 15 Jahre lediglich das niedrigste Niveau der Lesekompetenz erreichten; jedoch fanden die Befunde wenig Beachtung und zogen praktisch keine bildungspolitischen Konsequenzen nach sich. Weitere Datenerhebungen folgten 1996 und 1998. Untersuchungen, die explizit Zusammenhänge zwischen Grundkompetenzen und Aspekten der Erwerbssituation untersuchen und eine konsequente Längsschnittperspektive einnehmen, fehlen jedoch.

Eine deutsche Beteiligung am *Adult Literacy and Lifeskills (ALL) Survey*, einer international vergleichenden Studie zu den Kompetenzen der erwachsenen Bevölkerung (16- bis 65-Jährige), gab es zwar in den Erhebungsjahren 2003 und 2006 nicht. Dafür hat Deutschland jüngst seine Teilnahme am *Programme for the International Assessment of Adult Competencies (PIAAC)* der OECD zugesagt, welches das Kompetenzniveau bei Erwachsenen im Ländervergleich analysieren wird. Schwerpunkt werden dabei die Kompetenzen sein, die für eine erfolgreiche Teilnahme am Arbeitsleben Voraussetzung sind. PIAAC soll Regierungen umfassendere Informationen liefern, um ihre Bildungs- und Ausbildungspolitik zu evaluieren und zu gestalten. Dazu werden vergleichende Informationen zum Kenntnisstand der Erwachsenenbevölkerung und seiner Verteilung erhoben und untersucht, wie sie Einkommen, Beschäftigung und die Fähigkeit zu lebenslangem Lernen beeinflussen.

5.3 Lehrerbildung

Die Ausbildung von Lehrkräften wird als Einflussgröße auf ihre pädagogischen, fachwissenschaftlichen und fachdidaktischen Fähigkeiten angesehen, die wiederum die Qualität von Schule und Unterricht beeinflussen. Dieser Bereich erfährt deshalb international und auch national die Aufmerksamkeit empirischer Bildungsforschung. Im Kontext von *TEDS-M*, der *Teacher Education Study – Mathematics*, soll am Beispiel

der Lehrerausbildung für die Primarstufe und die Sekundarstufe I im Unterrichtsfach Mathematik die Wirksamkeit der Ausbildung empirisch geprüft werden. In diesem Zusammenhang werden curriculare, bildungspolitische und gesellschaftliche Rahmenbedingungen, konkrete Lehr-Lerninteraktionen sowie kognitive, affektive und handlungspraktische Effekte der Lehrerausbildung erfasst. Zentrale Fragen der 2006 bis 2008 durchgeführten Studie sind beispielsweise, was zukünftige Lehrerinnen und Lehrer in ihrer Ausbildung lernen, über welche Ausgangsvoraussetzungen sie im Vergleich zu anderen Studiengängen verfügen und wie sich die Ausbildung am Ende in beruflichem Handeln niederschlägt. Diese Fragen stellen Forschungsdesiderate dar. TEDS-M ist die erste Studie der IEA im tertiären Bereich sowie die erste international vergleichende Large Scale-Studie zur Lehrerausbildung. Der internationale Ergebnisbericht wird für Ende 2009 erwartet.

In diesem Zusammenhang sei auch das *Teaching and Learning International Survey (TALIS)* erwähnt. Damit nimmt die OECD die Lernumgebung und Arbeitsplatzbedingungen von Lehrerinnen und Lehrern in allgemein bildenden Schulen der Sekundarstufe I in international vergleichender Perspektive in den Blick. In der ersten Erhebungsphase 2007/08 sind über 20 Länder an der Untersuchung beteiligt. Weitere Länder können zu einem späteren Zeitpunkt noch einsteigen; die KMK hat eine Beteiligung Deutschlands jedoch nicht zugesagt. Zu den aktuellen Untersuchungsschwerpunkten gehören die Leitung und das Management von Schulen, die Wertschätzung der Arbeit von Lehrkräften, die Art und Weise des Feedbacks, das sie erhalten, sowie die Verwendung von Ergebnissen zur Personalentwicklung. Der erste Bericht aus diesem Projekt wird für Mitte 2009 erwartet.

6. Wirkungen, Desiderate und Perspektiven

An vielen der aufgeführten Studien hat sich Deutschland in den sich wandelnden Grenzen und mit unterschiedlichem Umfang beteiligt, wobei vor allem in den 1970er und 1980er Jahren vielfach ein Fehlen Deutschlands zu verzeichnen ist. Die Übersicht über Schulleistungstests dokumentiert die Vielfalt der Bemühungen, im Bereich der externen Evaluation verstärkt tätig zu werden und dabei unterschiedliche Bildungsstufen und Kompetenzbereiche zu berücksichtigen. Der Überblick zeigt auch, dass es im Verlauf der 1990er Jahre eine Entwicklung bei der Qualität der eingesetzten Instrumente und der dahinter stehenden Konzepte gibt. Diese Entwicklung ist dadurch gekennzeichnet, dass sich das Interesse der Leistungsstudien auf erklärende schulische und außerschulische Variablen der Leistungserbringung ausweitet und dass Rückmeldestrategien entwickelt werden, die sich nicht nur an die bildungspolitisch interessierte Öffentlichkeit wenden, sondern zunehmend an die Einzelschulen, die sich an den Erhebungen beteiligt haben.

Zugleich verweist die Übersicht über Typen und Varianten von Schulleistungstests aber auch auf eine andere Tendenz: Nach Jahren der Abstinenz im Bereich großflächig

angelegter Leistungstests beobachten wir seit Beginn der 1990er Jahre innerhalb Deutschlands eine Beschleunigung beim Einstieg in regelmäßige Vergleichsuntersuchungen. Dafür gibt es fraglos einige gute Gründe: Die Aussagekraft von Schulleistungsuntersuchungen wird durch Studiendesigns mit mindestens zwei Messungen qualitativ überproportional gesteigert. Auch ist es nachvollziehbar, dass auf die gut begründete Kritik an der Konzentration der eingesetzten Tests auf die Kulturtechniken mit einer Ausweitung auf weitere Fächer, z.B. auf geistes- und sozialwissenschaftliche Kompetenzbereiche, reagiert wird. Schließlich kann kaum bestritten werden, dass die Schulpolitik in einem föderalen Staat aus Gründen der Sicherung der Vergleichbarkeit ein starkes Interesse an regelmäßigen innerdeutschen Vergleichsuntersuchungen hat.

In internationaler Perspektive wird in der Rückschau auf die vergangenen Jahre vergleichender Schulleistungsforschung deutlich, dass zunehmend mehr Nationen an den Studien beteiligt sind. Viele der Länder, die erstmals teilnehmen, kommen aus dem arabischen und afrikanischen Raum. An TIMSS 2007 sind beispielsweise über 60 Länder beteiligt; bei FIMS und FISS waren es noch – je nach getesteter Altersstufe – 10 bis 14 Länder. Dies bedeutet zugleich einen erhöhten Bedarf an speziellen Trainings durch internationale Organisationen wie die IEA, um die Studien in allen Ländern unter vergleichbaren qualitativen Standards durchführen zu können und auch unter Berücksichtigung landesspezifischer bedeutsamer Fragestellungen auswerten und nutzbar machen zu können.

Eine weitere Forschungsperspektive betrifft die Anlage internationaler Leistungsvergleichsstudien. Baumert (2007) weist darauf hin, dass solche querschnittlichen Leistungsvergleiche nur einen begrenzten Erklärungsanspruch erheben können. Wirkungen von Einflussfaktoren im Bedingungsgefüge schulischer Leistungen sind empirisch nur mit einem längsschnittlichen Studiendesign nachzuweisen, wie es beispielsweise mit der *National Educational Panel Study (NEPS)* in Deutschland auf nationaler Ebene geplant ist. Auf internationaler Ebene kann nach Baumert (2007, S. 361) allerdings der Erklärungsanspruch von Leistungsvergleichsstudien beispielsweise dadurch gestärkt werden, dass sie in „genuine Forschungsprogramme“ mit längsschnittlichem Charakter eingebunden werden, wie beispielsweise mit der TIMSS Videostudie von 1995, die in Deutschland in einen Längsschnitt eingebettet wurde, oder mit dem PISA-Zyklus 2003, der in Deutschland ebenfalls durch eine Längsschnittstudie erweitert worden ist (PISA-I-Plus). Solche Programme schaffen ein solideres Fundament für die Dateninterpretation und sollten bei der Anlage internationaler Schulleistungsstudien, beispielsweise im Rahmen nationaler Erweiterungen, weiterhin Beachtung finden.

Kritik gibt es hierzulande hinsichtlich des Umgangs mit den Ergebnissen der unterschiedlichen Studien durch Wissenschaft und Politik. Als möglichen Erklärungsansatz für uneinheitliche Ergebnisinterpretationen sei nochmals auf Blömeke (2003, S. 27) verwiesen:

Bildungspolitik [liegt] eine andere Rationalität zugrunde ... als Wissenschaft. Politik ist notwendigerweise normativ orientiert und interessengeleitet. Wissenschaftliche Erkenntnisse sollten vor diesem Hintergrund zwar berücksichtigt werden, müssen aber nicht unbedingt handlungsleitend sein – ganz abgesehen von dem Problem, dass es ‚die‘ Wissenschaft auch nicht gibt. Die Nach-PISA-Diskussion hat gezeigt, wie unterschiedlich die empirischen Ergebnisse in der Erziehungswissenschaft interpretiert werden und wie unterschiedlich wissenschaftliche Empfehlungen und politische Entscheidungen ausfallen können.

Eine mangelnde Kontinuität in der Analyse und in der Präsentation empirischer Befunde aus Leistungsstudien konstatiert auch Klemm (2008) für die Situation *innerhalb* der wissenschaftlichen Zunft und mahnt, dass „widersprüchliche Interpretationen“ politisch folgenreich sein können:

Sie geben all denen, die bei der Konzipierung von Bildungspolitik lieber auf hergebrachte Überzeugungen als auf belastbares Wissen setzen, willkommene Argumente dafür, den gerade eingeschlagenen Weg evidenzbasierter Politik wieder zu verlassen. Die Bildungsforschung würde dann auf die Funktion zurückgedrängt, die Forschung nur zu oft für die Politik hat – nämlich die der Legitimierung einer ‚bauchbasierten‘ Steuerungspolitik. Das wäre schade (S. 69).

Hierbei sollte aber nicht vergessen werden, dass unsere wissenschaftliche Tradition auf der Diskussion von Widersprüchen und unterschiedlichen Interpretationsansätzen basiert und gerade der kontroverse Diskurs Fortschritt bedeuten kann. Bei der Bewertung des Verhältnisses von Wissenschaft und Politik bedarf es offensichtlich darüber hinaus einer Analyse der Frage, was die Akteure in Wissenschaft und Politik an wechselseitiger Beeinflussung erleben und welche Erwartungen sie an die jeweils andere Seite haben.

In inhaltlicher Hinsicht schließlich ist der Übersicht zu entnehmen, dass inzwischen zwar eine empirisch relativ gut abgesicherte Informationslage bezüglich des Primar- und Sekundarschulwesens besteht, die Informationen über den tertiären Bildungsbereich aber noch defizitär sind. Besonders interessant erscheint eine Fokussierung auf den tertiären Bereich, wenn unter dem Schlagwort des lebenslangen Lernens bzw. der *adult competencies* hinterfragt wird, ob das, was im allgemeinen Schulwesen an Wissen bzw. Kompetenzen vermittelt wird, überhaupt anschlussfähig bzw. notwendig erscheint. An dieser Stelle einmal die Abnehmerperspektive einzunehmen und zu fragen, was Kindern nach dem Übergang in die Sekundarstufe I oder Jugendlichen beim Übergang in den Beruf bzw. jungen Erwachsenen bei der Aufnahme eines Studiums an Voraussetzungen abverlangt wird, erschließt eine ganz neue Perspektive auf ein weitgehend noch unbearbeitetes Forschungsfeld. Der Gedanke, was ein späteres Studium bzw. ein später ausgeübter Beruf den Beteiligten abverlangt, legt auch eine eingehende Betrachtung der Lehrerbildung nahe. Insofern ist anzunehmen, dass internationale Schulleistungsvergleiche nach fünf Jahrzehnten auch weiterhin ein Thema bleiben werden.

Übersicht: Art und Umfang deutscher Beteiligung an für Deutschland relevanten internationalen Leistungsvergleichsstudien (Stand: Mai 2008)

Bezeichnung der Studie und internationale Koordination	Literatur (Auswahl)	Datenerhebung	Testbereiche	Altersgruppe	Anzahl beteiligter Länder	Teilnahme Deutschlands (Bundesländer)
Pilot Twelve-Country Study (IEA)	Foshay, Thorndike, Hotyat, Pidgeon & Walker, 1962	1959–60	Mathematik, Leseverständnis, Erdkunde, Naturwissenschaften, nonverbale Fähigkeiten	13-Jährige	12	Hessen (Darmstadt)
FIMS: First International Mathematics Study (IEA)	Husén, 1967; Postlethwaite, 1968	1964	Mathematik	13-Jährige; Ende Sek. II	12	nur Hessen und Schleswig-Holstein
Six-Subject Survey (IEA)	Peaker, 1975; Walker, 1976					
– Study of Reading Comprehension	Thorndike, 1973	1970–71	Leseverständnis, Leseschwindigkeit, Wortschatz	10-Jährige; 14-Jährige; Ende Sek. II	15	–
– Study of Literatur Education	Purves, 1973	1970–71	Literatur-/Textverständnis	14-Jährige; Ende Sek. II	10	–
– FISS: First International Science Study	Comber & Keeves, 1973	1970–71	Naturwissenschaften (Erdkunde, Biologie, Chemie, Physik)	10-Jährige; 14-Jährige; Ende Sek. II	19	repräsentativ
– Study of French as a Foreign Language	Carroll, 1975	1971	Französisch (Lese- und Hörverständnis, Sprech- und Schreibfähigkeiten)	14-Jährige; Ende Sek. II	8	–
– Study of English as a Foreign Language	Lewis & Massad, 1975	1971	Englisch (Lese- und Hörverständnis, Sprech- und Schreibfähigkeiten)	14-Jährige; Ende Sek. II	9	in Deutschland nur Leseverständnis in einer Auswahl von Bundesländern*
– Study of Civic Education	Torney, Oppenheim & Farnen, 1975	1971	Politische Bildung	10-Jährige; 14-Jährige; Ende Sek. II	10	10 Bundesländer

* Die genaue Zahl der teilnehmenden Bundesländer ließ sich auf Basis der Informationen zu den nationalen Stichproben in Lewis & Massad, 1975; Peaker, 1975 und Walker, 1976 nicht feststellen.

Bezeichnung der Studie und internationale Koordination	Literatur (Auswahl)	Datenerhebung	Testbereiche	Altersgruppe	Anzahl beteiligter Länder	Teilnahme Deutschlands (Bundesländer)
CES: Classroom Environment Study (IEA)	Helmke, Schneider & Weinert, 1986; Ryan, Anderson & Shapiro, 1989	1981–83 (in Deutschland: zwei Jahre später)	Mathematik, Naturwissenschaften, Geschichte	Klassenstufen 5–8	9	Bayern (nur München – Münchner Hauptschulstudie)
SIMS: Second International Mathematics Study (IEA)	Burstein, 1993; Robitaille & Garden, 1989; Travers & Westbury, 1989	1980–81	Mathematik (u.a. Arithmetik, Algebra, Geometrie)	13-Jährige; Ende Sek. II	20	–
SISS: Second International Science Study (IEA)	Keeves, 1992; Postlethwaite & Wiley, 1992; Rosier & Keeves, 1991	1983–84	Naturwissenschaften (Erdkunde, Biologie, Chemie, Physik)	10-Jährige; 14-Jährige; Ende Sek. II	23	–
PPP: Pre-Primary Project (IEA)	– Phase 1: Olmsted & Weikart, 1989, 1994 – Phase 2: Olmsted & Montie, 2001; Weikart, 1999; Weikart, Olmsted & Montie, 2003 – Phase 3: High/Scope Educational Research Foundation, 2003; Montie, Xiang & Schweinhart, 2006	Phase 1: 1986–94; Phase 2: 1989–2003; Phase 3: 1993–2003	Einflüsse der Lebens- und Lernumgebung auf die Entwicklung von Kindern im Vorschulalter	4-Jährige; 7-Jährige	Phase 1: 11; Phase 2: 15; Phase 3: 13	repräsentativ (nur Phase 1)
Written Composition Study (IEA)	Gorman, Purves & Degenhart, 1988; Purves, 1992	1985	Texte verfassen	Ende Grundschulzeit; Ende Pflichtschulzeit; Ende Sek. II	14	nur Hamburg (Hamburger Aufsatzstudie; nur Klasse 11: allg. und berufsbild. Vollzeitschulen)

Bezeichnung der Studie und internationale Koordination	Literatur (Auswahl)	Datenerhebung	Testbereiche	Altersgruppe	Anzahl beteiligter Länder	Teilnahme Deutschlands (Bundesländer)
IAEP: International Assessment of Educational Progress (ETS)	Lapointe, Askew & Mead, 1989; Semple, 1992	1988–89 (IAEP-I); 1990–91 (IAEP-II)	Mathematik; Naturwissenschaften; in IAEP-II bei den 13-Jährigen auch: Erdkunde, mathematisches und naturwissenschaftliches Problemlösen	13-Jährige; 9-Jährige; 13-Jährige	6 14 (9-Jährige); 20 (13-Jährige)	–
IRLS: International Reading Literacy Study (IEA)	Elley, 1992, 1994; Lehmann, Peek, Pieper & Stritzky, 1995; Lundberg & Linnakylä, 1993; Postlethwaite & Ross, 1992; Wagemaker, 1996	1990–91	Leseverständnis	9-Jährige; 14-Jährige	32	repräsentativ (alte und neue Bundesländer)
CompEd: Computers in Education Study (IEA)	Pelgrum & Plomp, 1991, 1993	Phase 1: 1989; Phase 2: 1992	Wissen um und Fähigkeiten der Computernutzung	10-Jährige; 13-Jährige; Ende Sek. II	21 12	9 Bundesländer Pop II und III
IALS: International Adult Literacy Survey (OECD, StatCan, ETS, UNESCO u.a.)	Murray, Kirsch & Jenkins, 1998; OECD, 1995, 1997, 2000	1994 (Welle 1); 1996 (Welle 2); 1998 (Welle 3)	Lesekompetenz, mathematische Fähigkeiten	Erwachsene (16- bis 65-Jährige)	8 (Welle 1); 5 (Welle 2); 9 (Welle 3)	Teilnahme Deutschlands (repräsentativ) in Welle 1
TIMSS: Third International Mathematics and Science Study (ab 1999: Trends in International Mathematics and Science Study) (IEA)	TIMSS 1995: – Pop I: Martin et al., 1997; Mullis et al., 1997	1994–95	Mathematik und Naturwissenschaften	9-Jährige; 13-Jährige; Ende Sek. II: – allgemein – Leistungs-/Schwerpunkte: Mathematik Physik	29 46 24 17 18	Teilnahme Deutschlands an TIMSS 1995: repräsentativ (ohne Baden-Württemberg) für 13-Jährige und Ende Sek. II; Grundschulstudie ohne dt. Teilnahme;

Bezeichnung der Studie und internationale Koordination	Literatur (Auswahl)	Datenerhebung	Testbereiche	Altersgruppe	Anzahl beteiligter Länder	Teilnahme Deutschlands (Bundesländer)
	<ul style="list-style-type: none"> - Pop II: Baumert et al., 1997; Beaton, Martin et al., 1996; Beaton, Mullis et al., 1996 - Pop III: Baumert, Bos & Lehmann, 2000a, 2000b; Mullis et al., 1998 					
TIMSS: Third International Mathematics and Science Study (ab 1999: Trends in International Mathematics and Science Study) (IEA)	<p>TIMSS 1999: Martin et al., 2000; Mullis et al., 2000</p> <p>TIMSS 2003: Mullis, Martin, Smith et al., 2003</p> <p>TIMSS 2007: Mullis et al., 2005</p> <p>TIMSS Advanced 2008: Garden et al., 2006</p>	<p>1998–99 (auch TIMSS-R(peat))</p> <p>2002–03</p> <p>2007</p> <p>2007–08</p>	<p>Mathematik und Naturwissenschaften</p>	<p>13-Jährige</p> <p>9-Jährige; 13-Jährige</p> <p>9-Jährige; 13-Jährige</p> <p>13-Jährige; Ende Sek. II: Leistung-/Schwerpunktkurse Mathematik, Physik</p>	<p>38</p> <p>25 46</p> <p>insg. 60</p> <p>insg. 10</p>	<p>Teilnahme Deutschlands an TIMSS 2007: repräsentativ (Klassenstufe 4/9-Jährige); Studie der Klassenstufe 8/13-Jährige ohne dt. Teilnahme</p>
TIMSS Video: Third International Mathematics and Science Videotape Classroom Study (IEA)	<p>TIMSS Video: Klieme, Knoll & Schümer, 1998; Stigler, Gonzales, Kawana, Knoll & Serrano, 1999</p> <p>TIMSS-R Video Project: Hiebert et al., 2003</p>	<p>1994–95,</p> <p>1998–2001</p>	<p>Mathematikunterricht</p> <p>Mathematik- und Naturwissenschaftsunterricht</p>	<p>Klassenstufe 8</p> <p>Klassenstufe 8</p>	<p>3</p> <p>7</p>	<p>Teilnahme Deutschlands an TIMSS 1995 Video Study</p>

Bezeichnung der Studie und internationale Koordination	Literatur (Auswahl)	Datenerhebung	Testbereiche	Altersgruppe	Anzahl beteiligter Länder	Teilnahme Deutschlands (Bundesländer)
CivEd: Civic Education Study (IEA)	Oesterreich, 2002; Torney-Purta, Lehmann, Oswald & Schulz, 2001	Phase 1: 1996–97; Phase 2: 1999–2000	Politisches Wissen	14-Jährige; Schüler der Sek. II (16- bis 19-Jährige)	31	repräsentativ (nur 14-Jährige)
SITES: Second Information Technology in Education Study (IEA)	SITES M1: Pelgrum & Anderson, 2001	1998–99	Nutzung von IKT in Schule und Unterricht	10-Jährige; 14-Jährige; Ende Sek. II	26	Teilnahme Deutschlands an SITES M2
	SITES M2: Kozma, 2003; Plomp, Anderson, Law & Quale, 2003	2000–01		Primarstufe; Sek. I; Sek. II (je 4 Fallstudien)	28	
ISSUS: International Survey of Schools at the Upper Secondary Level (OECD)	SITES 2006: International Association for the Evaluation of Educational Achievement, 2008b	2006	Elemente der Schulstruktur: Personelle Ausstattung, Schulmanagement, Nutzung von IKT	Klassenstufe 8 (Schulen, Mathematik- und Naturwissenschaftslehrkräfte)	ca. 20 (ca. 400 Schulen je Land)	–
	OECD, 2004a, c	2001		Schulen mit Sek. II	14	
PISA: Programme for International Student Assessment (OECD)	PISA 2000: Adams & Carstensen, 2002; Baumert et al., 2001; OECD, 1999, 2001	2000	Leseverständnis, Mathematik und Naturwissenschaften (Schwerpunkt 2000: Lesen; 2003: Mathematik; 2006: Naturwissenschaften)	15-Jährige	32	repräsentativ
	PISA 2003: OECD, 2003, 2004b, d; Prenzel, Baumert et al., 2004	2003			41	repräsentativ
	PISA 2006: OECD, 2007; Prenzel et al., 2007	2006			57	repräsentativ

Bezeichnung der Studie und internationale Koordination	Literatur (Auswahl)	Datenerhebung	Testbereiche	Altersgruppe	Anzahl beteiligter Länder	Teilnahme Deutschlands (Bundesländer)
PIRLS: Progress in International Reading Literacy Study (in Deutschland bezeichnet als Internationale Grundschul-Lese-Untersuchung – IGLU) (IEA)	PIRLS/IGLU 2001: Bos et al., 2003; Campbell, Kelly, Mullis, Martin & Sainsbury, 2001; Mullis, Martin, Gonzalez & Kennedy, 2003 PIRLS / IGLU 2006: Bos, Hornberg et al., 2007; Mullis, Kennedy, Martin & Sainsbury, 2006; Mullis, Martin, Kennedy & Foy, 2007	2001 2006	Leseverständnis	9- bis 10-Jährige	35 40	repräsentativ repräsentativ
ALL: Adult Literacy and Life Skills Survey (StatCan, ETS, OECD, UNESCO u.a.)	Statistics Canada & OECD, 2005	2003 (Welle 1); 2006 (Welle 2)	Lesekompetenz, mathematische Fähigkeiten, Problemlösen	Erwachsene (16- bis 65-Jährige)	6 7	–
TEDS-M: Teacher Education Study – Mathematics (IEA)	International Association for the Evaluation of Educational Achievement, 2008c; Tatro et al., 2008	Phase 1: 2005–06; Phase 2: 2007–08	Ausbildung von Mathematiklehrkräften für die Primar- und untere Sekundarstufe	Institutionen der Lehrerbildung; Dozenten/ Ausbilder von angehenden Lehrkräften; angehende Lehrkräfte	15	Teilnahme Deutschlands
TALIS: Teaching and Learning International Survey (OECD)	OECD, 2008a	Pretest: 2006; Haupterhebung: 2007–08	Arbeitsbedingungen und Arbeitsweisen an Schulen (Lehrer, Schulleitung)	Lehrer der Sek. I; Schulleiter	22	–

Bezeichnung der Studie und internationale Koordination	Literatur (Auswahl)	Datenerhebung	Testbereiche	Altersgruppe	Anzahl beteiligter Länder	Teilnahme Deutschlands (Bundesländer)
ICCS 2009: International Civic and Citizenship Education Study (IEA)	International Association for the Evaluation of Educational Achievement, 2007, 2008a	2008–09	Konzeptuelles Verständnis und Fähigkeiten im Bereich der politischen und staatsbürgerlichen Bildung	Klassenstufe 8	39	–
PIAAC: Programme for the International Assessment of Adult Competencies (OECD)	Gnahn, 2006; OECD, 2008b	Pretest: 2010; Haupterhebung: 2011	Lesekompetenz, mathematische Fähigkeiten, Problemlösen, Umgang mit IuK-Technologie	Erwachsene (16- bis 65-Jährige)	in Planung	Teilnahme Deutschlands

Literatur

- Ackeren, I. van (2002). Von FIMS und FISS bis TIMSS und PISA. Schulleistungen in Deutschland im historischen und internationalen Vergleich. *Die Deutsche Schule*, 94 (2), 157–175.
- Ackeren, I. van (2003). Large Scale Assessments. *Pädagogische Führung*, 14 (1), 31–33.
- Ackeren, I. van & Klemm, K. (2002). Organisation, Anlage und Umfang von Large-Scale-Assessments. *Pädagogische Führung*, 13 (1), 40–41.
- Adams, R. & Carstensen, C. (2002). Scaling outcomes. In J. R. Adams & M. Wu (Eds.), *PISA 2000 technical report* (pp. 149–162). Paris: OECD.
- Avenarius, H., Ditton, H., Döbert, H., Klemm, K., Klieme, E., Rürup, M., Tenorth, H.-E., Weishaupt, H. & Weiß, M. (2003). *Bildungsbericht für Deutschland. Erste Befunde*. Opladen: Leske + Budrich.
- Baumert, J. (2007). Internationale Schulleistungsmessungen. In H.-E. Tenorth & R. Tippelt (Hrsg.), *Beltz Lexikon Pädagogik* (S. 358–361). Weinheim: Beltz.
- Baumert, J., Bos, W. & Lehmann, R. (Hrsg.). (2000a). *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., Bos, W. & Lehmann, R. (Hrsg.). (2000b). *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., 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.
- Beaton, A. E., Martin, M. O., Mullis, I. V. S., Gonzales, E. J., Smith, T. A. & Kelly, D. L. (1996). *Science achievement in the middle school years: IEA's Third International Mathematics and Science Study*. Chestnut Hill, MA: Boston College.
- Beaton, A. E., Mullis, I. V. S., Martin, M. O., Gonzalez, E. J., Kelly, D. L. & Smith, T. A. (1996). *Mathematics achievement in the middle school years: IEA's Third International Mathematics and Science Study*. Chestnut Hill, MA: Boston College.
- Blömeke, S. (2003). Lehrerausbildung – Lehrerhandeln – Schülerleistungen. Perspektiven nationaler und internationaler empirischer Bildungsforschung. Antrittsvorlesung 10. Dezember 2003 an der Humboldt-Universität zu Berlin, Institut für Erziehungswissenschaften. Verfügbar unter: <http://edoc.hu-berlin.de/humboldt-vl/139/bloemeke-sigrid-3/PDF/bloemeke.pdf> [08.07.08].
- Bloom, B. S. (1974). Implications of the IEA Studies for Curriculum and Instruction. *The School Review*, 82 (3), 413–435.
- Bos, W., Hornberg, S., Arnold, K.-H., Faust, G., Fried, L., Lankes, E.-M., Schwippert, K. & Valtin, R. (2007). *IGLU 2006. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich*. Münster: Waxmann.
- Bos, W., Lankes, E.-M., Prenzel, M., Schwippert, K., Walther, G. & Valtin, R. (Hrsg.). (2003). *Erste Ergebnisse aus IGLU. Schülerleistungen am Ende der vierten Jahrgangsstufe im internationalen Vergleich*. Münster: Waxmann.

- Bos, W. & Postlethwaite, T. N. (2001). Internationale Schulleistungsforschung. Ihre Entwicklungen und Folgen für die deutsche Bildungslandschaft. In F. E. Weinert (Hrsg.), *Leistungsmessungen in Schulen* (S. 251–267). Weinheim: Beltz.
- Bos, W. & Schwippert, K. (2002). TIMSS, PISA, IGLU & CO. Vom Sinn und Unsinn internationaler Schulleistungsuntersuchungen. *Bildung und Erziehung*, 55 (1), 5–23.
- Bos, W., Schwippert, K. & Stubbe, T. C. (2007). Die Koppelung von sozialer Herkunft und Schülerleistung im internationalen Vergleich. In W. Bos (Hrsg.), *IGLU 2006. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 225–247). Münster: Waxmann.
- Burstein, L. (Ed.). (1993). *The IEA study of mathematics* (Vol. 3). Oxford: Pergamon Press.
- Campbell, J. R., Kelly, D. L., Mullis, I. V. S., Martin, M. O. & Sainsbury, M. (2001). *Framework and specifications for PIRLS assessment 2001* (2nd Ed.). Chestnut Hill, MA: Boston College.
- Carroll, J. B. (1975). *The teaching of French as a foreign language in eight countries*. New York: Wiley.
- Comber, L. C. & Keeves, J. P. (1973). *Science education in nineteen countries: An empirical study*. Stockholm: Almqvist & Wiksell.
- Dahrendorf, R. (1965). *Bildung ist Bürgerrecht. Plädoyer für eine aktive Bildungspolitik*. Hamburg: Nannen-Verlag.
- Elley, W. B. (1992). *How in the world do students read? IEA Study of Reading Literacy* (2nd Ed.). The Hague: International Association for the Evaluation of Educational Achievement.
- Elley, W. B. (1994). *The IEA Study of Reading Literacy*. Oxford: Pergamon Press.
- Foshay, A. W., Thorndike, R. L., Hotyat, F., Pidgeon, D. A. & Walker, D. A. (1962). *Educational achievement of thirteen-year-olds in twelve countries*. Hamburg: UNESCO Institute for Education.
- Garden, R. A., Lie, S., Robitaille, D. F., Angell, C., Martin, M. O., Mullis, I. V. S., Foy, P. & Arora, A. (2006). *TIMSS Advanced 2008 assessment frameworks*. Chestnut Hill, MA: Boston College.
- Gnahn, D. (2006). Kompetenzmessung bei Erwachsenen – zum Stand von PIAAC. In A. Grotlüschen & A. Linde (Hrsg.), *Literalität, Grundbildung oder Lesekompetenz? Literacy, basic education or reading competencies. Beiträge zu einer Theorie-Praxis-Diskussion* (S. 25–30). Münster: Waxmann.
- Gorman, T. P., Purves, A. C. & Degenhart, R. E. (Eds.). (1988). *The IEA Study of Written Composition I: The international writing tasks and scoring scales*. Oxford: Pergamon Press.
- Hamilton, L. S., Stecher, B. M. & Klein, S. P. (Eds.). (2002). *Making sense of test-based accountability in education*. Santa Monica, CA: Rand.
- Helmke, A., Schneider, W. & Weinert, F. E. (1986). Quality of instruction and classroom learning outcomes. The German contribution to the IEA Classroom Environment Study. *Teaching and Teacher Education*, 2, 1–18.
- Hiebert, J., Gallimore, R., Garnier, H., Givvin, K. B., Hollingsworth, H., Jacobs, J., Chui, A. M.-Y., Wearne, D., Smith, M., Kersting, N., Manaster, A., Tseng, E., Etterback, W., Manaster, C., Gonzales, P. & Stigler, J. (2003). Teaching mathematics in seven countries. Results from the TIMSS 1999 Video Study. *Education Statistics Quarterly*, 5 (1), 7–15.
- High/Scope Educational Research Foundation. (2003). *Sights and Sounds of Children. 14 Countries* [DVD and various videotape formats]. Ypsilanti, MI: High/Scope Press.
- Husén, T. (1967). *International study of achievement in mathematics. Vols. I and II*. Stockholm: Almqvist & Wiksell.
- International Association for the Evaluation of Educational Achievement. (2007). *International Civic and Citizenship Education Study. Assessment framework*. Amsterdam: IEA.

- International Association for the Evaluation of Educational Achievement. (2008a). *International Civic and Citizenship Education Study*. Verfügbar unter: <http://iccs.acer.edu.au/index.php?page=home> [08.02.08].
- International Association for the Evaluation of Educational Achievement. (2008b). *Second Information Technology in Education Study 2006*. Verfügbar unter: <http://www.sites2006.net/exponent/index.php> [08.02.08].
- International Association for the Evaluation of Educational Achievement. (2008c). *Teacher Education and Development Study in Mathematics*. Verfügbar unter: <http://teds.educ.msu.edu/default.asp> [14.01.08].
- Jornitz, S. (2008). Was bedeutet eigentlich „evidenzbasierte Bildungsforschung“? Über den Versuch, Wissenschaft für Praxis verfügbar zu machen, am Beispiel der Review-Erstellung. *Die Deutsche Schule*, 100, (2), 206–216.
- Keeves, J. P. (1992). *The IEA study of science III: Changes in science education and achievement: 1970 to 1984*. Oxford: Pergamon Press.
- Klemm, K. (2008, 15. Mai). Schulforscher, aufgepasst! *Die Zeit*, S. 69.
- Klieme, E., Knoll, S. & Schümer, G. (1998). *Mathematikunterricht der Sekundarstufe I in Deutschland, Japan und den USA. Multimedia-CD-Dokumentation zur TIMSS-Videostudie*. Berlin: Max-Planck-Institut für Bildungsforschung.
- Kozma, R. B. (Ed.). (2003). *Technology, innovation, and educational change: A global perspective*. Eugene, OR: ISTE.
- Lapointe, A. E., Askew, J. M. & Mead, N. A. (1989). *A world of differences. An international assessment of mathematics and science. Report of the first International Assessment of Educational Progress* [Report No. 19-CAEP-01]. Princeton, NJ: Educational Testing Service.
- Lehmann, R. H., Peek, R., Pieper, I. & Stritzky, R. v. (1995). *Leseverständnis und Lesegewohnheiten deutscher Schüler und Schülerinnen*. Weinheim: Beltz.
- Lewis, E. G. & Massad, C. E. (1975). *The teaching of English as a foreign language in ten countries*. Stockholm: Almqvist & Wiksell.
- Lundberg, I. & Linnakylä, P. (1993). *Teaching reading around the world*. The Hague: International Association for the Evaluation of Educational Achievement.
- Martin, M. O., Mullis, I. V. S., Beaton, A. E., Gonzales, E. J., Smith, T. A. & Kelly, D. L. (1997). *Science achievement in the primary school years: The IEA's Third International Mathematics and Science Study*. Chestnut Hill, MA: Boston College.
- Martin, M. O., Mullis, I. V. S., Gonzalez, E. J., Gregory, K. D., Smith, T. A., Chrostowski, S. J., Garden, R. A. & O'Connor, K. M. (2000). *TIMSS 1999 international science report. Findings from IEA's repeat of the Third International Science and Science Study at the eighth grade*. Chestnut Hill, MA: Boston College.
- Matthiesen, H. (1974, 20. September). Im internationalen Vergleich schneidet das Bildungswesen der Bundesrepublik miserabel ab. Die deutschen Schüler auf dem letzten Platz. *Die Zeit*, S. 17.
- Montie, J. E., Xiang, Z. & Schweinhart, L. J. (2006). Preschool experience in 10 countries: Cognitive and language performance at age 7. *Early Childhood Research Quarterly*, 21 (3), 313–331.
- Mullis, I. V. S., Kennedy, A. M., Martin, M. O. & Sainsbury, M. (2006). *PIRLS 2006 assessment framework and specifications. Progress in International Reading Literacy Study (2nd Ed.)*. Amsterdam: Boston College.
- Mullis, I. V. S., Martin, M. O., Beaton, A. E., Gonzalez, E. J., Kelley, D. L. & Smith, T. A. (1997). *Mathematics achievement in the primary school years: IEA's Third International Mathematics and Science Study (TIMSS)*. Chestnut Hill, MA: Boston College.

- Mullis, I. V. S., Martin, M. O., Beaton, A. E., Gonzales, E. J., Kelly, D. L. & Smith, T. A. (1998). *Mathematics and science achievement in the final year of secondary school: IEA's Third International Mathematics and Science Study (TIMSS)*. Chestnut Hill, MA: Boston College.
- Mullis, I. V. S., Martin, M. O., Gonzalez, E. J., Gregory, K. D., Garden, R. A., O'Connor, K. M., Chrostowski, S. J. & Smith, T. A. (2000). *TIMSS 1999 international mathematics report. Findings from IEA's repeat of the Third International Mathematics and Science Study at the eighth grade*. Chestnut Hill, MA: Boston College.
- Mullis, I. V. S., Martin, M. O., Gonzalez, E. J. & Kennedy, A. M. (2003). *PIRLS 2001 international report: IEA's Study of Reading Literacy achievement in primary schools in 35 countries*. Chestnut Hill, MA: Boston College.
- Mullis, I. V. S., Martin, M. O., Kennedy, A. M. & Foy, P. (2007). *PIRLS 2006 international report. IEA's Progress in International Reading Literacy Study in Primary Schools in 40 Countries*. Chestnut Hill, MA: Boston College.
- Mullis, I. V. S., Martin, M. O., Ruddock, G. J., O'Sullivan, C. Y., Arora, A. & Erberber, E. (2005). *TIMSS 2007 assessment frameworks*. Chestnut Hill, MA: Boston College.
- Mullis, I. V. S., Martin, M. O., Smith, T. A., Garden, R. A., Gregory, K. D., Gonzales, E. J., Chrostowski, S. J. & O'Connor, K. M. (2003). *TIMSS assessment frameworks and specifications 2003*. Chestnut Hill, MA: Boston College.
- Murray, S. T., Kirsch, I. S. & Jenkins, L. B. (1998). *Adult literacy in OECD countries: technical report on the first International Adult Literacy Survey*. Northford, CN: U.S. Department of Education, National Center for Educational Statistics.
- OECD. (1995). *Literacy, economy and society. Results of the first International Adult Literacy Survey*. Washington, DC: OECD Publications and Information Centre.
- OECD. (1997). *Literacy skills for the knowledge society: Further results from the International Adult Literacy Survey*. Washington, DC: OECD Washington Center.
- OECD. (1999). *Measuring student knowledge and skills: The PISA 2000 assessment of reading, mathematical and scientific literacy*. Paris: OECD.
- OECD. (2000). *Literacy in the information age: Final report of the International Adult Literacy Survey*. Washington, DC: OECD Washington Center.
- OECD. (2001). *Lernen für das Leben. Erste Ergebnisse der internationalen Schulleistungsstudie PISA 2000*. Paris: OECD.
- OECD. (2003). *PISA 2003 assessment framework: Mathematics, reading, science and problem solving knowledge and skills*. Paris: OECD.
- OECD. (2004a). *Completing the foundation for lifelong learning. An OECD survey of upper secondary schools*. Paris: OECD.
- OECD. (2004b). *Learning for tomorrow's world: First results from PISA 2003*. Paris: OECD.
- OECD. (2004c). *OECD survey of upper secondary schools. Technical report*. Paris: OECD.
- OECD. (2004d). *Problem solving for tomorrow's world – First measures of cross curricular competencies from PISA 2003*. Paris: OECD.
- OECD. (2007). *PISA 2006. Schulleistungen im internationalen Vergleich. Naturwissenschaftliche Kompetenzen für die Welt von morgen*. Paris: OECD.
- OECD. (2008a). *OECD Teaching and Learning International Survey (TALIS)*. Verfügbar unter: http://www.oecd.org/document/0/0,3343,en_2649_39263231_38052160_1_1_1_1,00.html [14.01.08].
- OECD. (2008b). *OECD Programme for the International Assessment of Adult Competencies (PIAAC)*. Verfügbar unter: http://www.oecd.org/document/35/0,3343,en_2649_201185_40277475_1_1_1_1,00.html [26.05.08].

- Oesterreich, D. (2002). *Politische Bildung von 14-Jährigen in Deutschland*. Opladen: Leske + Budrich.
- Olmsted, P. P. & Montie, J. (2001). *Early childhood settings in 15 countries: What are their structural characteristics?* Ypsilanti, MI: High/Scope Press.
- Olmsted, P. P. & Weikart, D. P. (1989). *How nations serve young children. Profiles of child care and education in 14 countries*. Ypsilanti, MI: High/Scope Press.
- Olmsted, P. P. & Weikart, D. P. (1994). *Families speak. Early care and education in 11 countries*. Ypsilanti, MI: High/Scope Press.
- Peaker, G. F. (1975). *An empirical study of education in twenty-one countries: A technical report*. (International studies in evaluation, Vol. 8). New York: Wiley.
- Peisert, H. (1967). *Soziale Lage und Bildungschancen in Deutschland*. München: Piper.
- Pelgrum, W. J. & Anderson, R. E. (Eds.). (2001). *ICT and the emerging paradigm for life long learning: An IEA educational assessment of infrastructure, goals, and practices in twenty-six countries*. Amsterdam: IEA.
- Pelgrum, W. J. & Plomp, T. (1991). *The use of computers in education worldwide. Results from the 'Computers in Education' Survey in 19 education systems*. Oxford: Pergamon.
- Pelgrum, W. J. & Plomp, T. (1993). *The IEA Study of Computers in Education I: Implementation of an innovation in twenty countries*. Oxford: Pergamon Press.
- Picht, G. (1964). *Die deutsche Bildungskatastrophe. Analyse und Dokumentation*. Olten: Walter.
- Plomp, T., Anderson, R. E., Law, N. & Quale, A. (Eds.). (2003). *Cross-national policies and practices on information and communication technology in education*. Greenwich: Information Age Publishing.
- Plomp, T., Howie, S. & McGaw, B. (2003). International studies of educational achievement. In T. Kellaghan & D. L. Stufflebeam (Eds.), *International handbook of educational evaluation. Part two: Practice* (pp. 951–978). Dordrecht: Kluwer.
- Postlethwaite, T. N. (Hrsg.). (1968). *IEA Leistungsmessung in der Schule. Eine internationale Untersuchung am Beispiel des Mathematikunterrichts*. Frankfurt a.M.: Diesterweg.
- Postlethwaite, T. N. (2004). *What do international assessment studies tell us about the quality of school systems*. Verfügbar unter: <http://unesdoc.unesco.org/images/0014/001466/146692e.pdf> [26.06.08].
- Postlethwaite, T. N. & Ross, K. N. (1992). *Effective schools in reading: Implications for educational planners*. Hamburg: International Association for the Evaluation of Educational Achievement.
- Postlethwaite, T. N. & Wiley, D. E. (1992). *The IEA Study of Science II: Science achievement in twenty-three countries*. Oxford: Pergamon Press.
- Prenzel, M., Artelt, C., Baumert, J., Blum, W., Hammann, M., Klieme, E. & Pekrun, R. (Hrsg.). (2007). *PISA 2006. Die Ergebnisse der dritten internationalen Vergleichsstudie*. Münster: Waxmann.
- Prenzel, M., Baumert, J., Blum, W., Lehmann, R., Leutner, D., Neubrand, M., Pekrun, R., Rolff, H.-G., Rost, J. & Schiefele, U. (2004). *PISA 2003. Der Bildungsstand der Jugendlichen in Deutschland – Ergebnisse des zweiten internationalen Vergleichs*. Münster: Waxmann.
- Purves, A. C. (1973). *Literature education in ten countries: An empirical study*. Stockholm: Almqvist & Wiksell.
- Purves, A. C. (Ed.). (1992). *The IEA Study of Written Composition, Vol. II: Education and performance in fourteen countries*. Oxford: Pergamon Press.
- Purves, A. C. & Levine, D. U. (1975). *Educational policy and educational assessment. Implications of the IEA surveys of achievement*. Berkeley, CA: McCutchan.
- Robitaille, D. F. & Garden, A. (Eds.). (1989). *The IEA Study of Mathematics, Vol. II: Contexts and outcomes of school mathematics*. Oxford: Pergamon Press.

- Rosier, M. J. & Keeves, J. P. (1991). *The IEA Study of Science, Vol. I: Science education and curricula in twenty-three countries*. Oxford: Pergamon Press.
- Ryan, D. W., Anderson, L. W. & Shapiro, B. J. (Eds.). (1989). *The IEA Classroom Environment Study*. Oxford: Pergamon Press.
- Schultze, W. & Riemenschneider, L. (1967). *Eine vergleichende Studie über die Ergebnisse des Mathematikunterrichts in zwölf Ländern*. Frankfurt a.M.: Deutsches Institut für Internationale Pädagogische Forschung.
- Schwippert, K., Bos, W. & Lankes, E.-M. (2003). Heterogenität und Chancengleichheit am Ende der vierten Jahrgangsstufe im internationalen Vergleich. In W. Bos, E.-M. Lankes, M. Prenzel, K. Schwippert, G. Walther & R. Valtin (Hrsg.), *Erste Ergebnisse aus IGLU. Schülerleistungen am Ende der vierten Jahrgangsstufe im internationalen Vergleich* (S. 265–302). Münster: Waxmann.
- Schwippert, K. & Goy, M. (2008). Leistungsvergleichs- und Schulqualitätsforschung. In W. Helsper & J. Böhme (Hrsg.), *Handbuch der Schulforschung* (2., durchgesehene und erweiterte Aufl.) (S. 387–421). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Schwippert, K. & Schnabel, K. U. (2000). Mathematisch-naturwissenschaftliche Grundbildung ausländischer Schulausbildungsabsolventen. In J. Baumert, W. Bos & R. H. Lehmann (Hrsg.), *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* (S. 282–300). Opladen: Leske + Budrich.
- Semple, B. M. (1992). *Performance assessment: An international experiment. Report of the International Assessment of Educational Achievement* [Report No. 22–CAEP-06]. Princeton, NJ: Educational Testing Service.
- Stanat, P. & Lüdtke, O. (2007). Internationale Schulleistungsvergleiche. In G. Trommsdorff & H.-J. Kornadt (Hrsg.), *Enzyklopädie der Psychologie: Kulturvergleichende Psychologie, Bd. 2: Kulturelle Determinanten des Erlebens und Verhaltens* (S. 279–347). Göttingen: Hogrefe.
- Statistics Canada & OECD. (2005). *Learning a living. First results of the Adult Literacy and Life Skills Survey*. Ottawa: Statistics Canada.
- Stigler, J. W., Gallimore, R. & Hiebert, J. (2000). Using video surveys to compare classrooms and teaching across cultures: Examples and lessons from the TIMSS Video studies. *Educational Psychologist*, 35, 87–100.
- Stigler, J. W., Gonzales, P., Kawanaka, T., Knoll, S. & Serrano, A. (1999). The TIMSS Videotape Classroom Study. Methods and findings from an exploratory research project on eighth-grade mathematics instruction in Germany, Japan, and the United States. *Education Statistics Quarterly*, 1 (2), 109–112.
- Tatto, M. T., Schwille, J., Senk, S., Ingvarson, L., Peck, R. & Rowley, G. (2008). *Teacher Education and Development Study in Mathematics (TEDS-M). Conceptual framework*. East Lansing, MI: Teacher Education and Development International Study Center, College of Education, Michigan State University.
- Thorndike, R. L. (1973). *Reading comprehension education in fifteen countries: An empirical study* (International studies in evaluation, Vol. 3). Stockholm: Almqvist & Wiksell.
- Torney, J. V., Oppenheim, A. N. & Farnen, R. F. (1975). *Civic education in ten countries: An empirical study*. New York: Wiley.
- Torney-Purta, J., Lehmann, R., Oswald, H. & Schulz, W. (2001). *Citizenship and education in twenty-eight countries. Civic knowledge and engagement at age fourteen*. Amsterdam: IEA.
- Travers, K. J. & Westbury, I. (Eds.). (1989). *The IEA Study of Mathematics, Vol. I: Analysis of mathematics curricula*. Oxford: Pergamon Press.

- Wagemaker, H. (Ed.). (1996). *Are girls better readers? Gender differences in reading literacy in 32 countries*. Amsterdam: The International Association for the Evaluation of Educational Achievement.
- Walker, D. A. (1976). *The IEA Six-Subject Survey: An empirical study of education in twenty-one countries* (International studies in evaluation, Vol. 9). Stockholm: Almqvist & Wiksell.
- Weikart, D. P. (1999). *What should young children learn? Teacher and parent views in 15 countries*. Ypsilanti, MI: High/Scope Press.
- Weikart, D. P., Olmsted, P. P. & Montie, J. (2003). *World of preschool experience. Observation in 15 countries*. Ypsilanti, MI: High/Scope Press.