



## Aktuelle Perspektiven und Entwicklungen in Large-Scale-Studien: Eine Hinführung

*Katrin Schulz-Heidorf & Julia Gerick*

*Universität Hamburg*

Die Teilnahme Deutschlands an internationalen Schulvergleichsstudien blickt auf eine überraschend lange Tradition zurück: Bereits 1963 beteiligten sich Hessen und Schleswig-Holstein an der *First International Mathematics Study* (FIMS), einer der ersten international vergleichenden sogenannten Large-Scale-Assessments überhaupt. Dem folgte 1970 die repräsentative Beteiligung deutscher 10- und 14-jähriger Schülerinnen und Schüler an der *First International Science Study* (FISS). Damals wie heute verfolgen Large-Scale-Assessments dabei das Ziel, schulische Kompetenzen von Kindern und Jugendlichen in einem komplexen Wirkungsgefüge von kulturellen, schulischen, curricular-strukturellen und häuslichen Lernbedingungen zu erfassen und international vergleichbar zu machen. Die Testungen an zentralen Wendepunkten der Schullaufbahn, also zum Ende der Grundschulzeit, zum Ende der allgemeinen Schulpflicht und zum Ende der voruniversitären Ausbildung (vgl. van Ackeren, 2002), haben sich dabei ebenso bewährt. Generiert wird hiermit in erster Linie bildungspolitisches und unterrichtspraktisches Steuerungswissen: über die Qualität, d.h. Lernwirksamkeit des Bildungssystems in der erfassten Domäne, über die Entwicklung der Qualität des Bildungssystems durch die wiederholte Teilnahme an entsprechenden Studien sowie über die Leistungsfähigkeit des Bildungssystems im Vergleich mit denen anderer Länder. Gerade solchen internationalen Vergleichen spricht Blömeke (2005) eine besondere Bedeutung für die Entwicklung nationaler Bildungssysteme zu:

Dass in Deutschland überhaupt wieder über das Bildungssystem diskutiert wird, dass andere Institutionalisierungsformen als die bestehenden in den Sinn kommen, ist eine Folge internationaler Vergleiche. Internationalität öffnet den Blick dafür, dass Systemstrukturen

und Lehr-Lernprozesse auch ganz anders organisiert werden können, als es vor Ort der Fall ist (S. 24).

So war das im Vergleich zu anderen Ländern unterdurchschnittliche Abschneiden deutscher Schülerinnen und Schüler in PISA 2000 als bekanntestes Beispiel ausschlaggebend für eine ganze Reihe von Bildungsreformen sowie die Initiierung der Gesamtstrategie zum Bildungsmonitoring der Kultusministerkonferenz (KMK, 2006). Allerdings zeigte die Teilnahme Deutschlands an internationalen Large-Scale-Assessments nicht von Anfang an solch eine bildungspolitische Wirkung: Anhand der FISS- und FIMS-Ergebnisse ließ sich bereits vor über 50 Jahren ablesen, dass das deutsche Schulsystem im Ländervergleich nicht den erwarteten Spitzenplatz einnimmt. Matthiesen titelt beispielsweise 1974 in DIE ZEIT „Die deutschen Schüler auf dem letzten Platz“ und „Im internationalen Vergleich schneidet das Bildungswesen der Bundesrepublik miserabel ab“ (Matthiesen, 1974, S. 17) und bezieht sich dabei auf die FISS-Ergebnisse, die „niederschmetternd“ seien. Auch in Mathematik blieben die Schülerinnen und Schüler – zumindest diejenigen, die an mathematisch-naturwissenschaftlichen Gymnasien lernten – bestenfalls im Mittelfeld; ihre Leistungen seien „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). Neben der kostspieligen, aufwendigen Auswertung der Daten (die noch auf Lochkarten per Post verschickt und auf den ersten Computern überhaupt ausgewertet wurden) ist es wohl auch auf die damals vor allem geisteswissenschaftliche Orientierung der deutschen Pädagogik und ihre wenig entwickelte empirieorientierte Denkweise (vgl. Bos, Postlethwaite & Gebauer, 2010) zurückzuführen, dass diese einschneidenden Ergebnisse in dem ‚Land der Dichter und Denker‘ kaum Aufmerksamkeit erfuhren und nahezu ohne Konsequenzen blieben (Roeder, 2003). Deutschland beteiligte sich dann auch erst (mit Ausnahme der nicht auf Kompetenzen sondern auf unterrichtliche Prozessdaten fokussierten *Classroom Environment Study* 1981–1983) knappe 20 Jahre später erneut an einer Large-Scale-Studie, der *International Reading Literacy Study* (1990–1991, ab 2001 PIRLS/IGLU) und der *Third International Mathematics and Science Study* (1994–1995, TIMSS). Während die TIMSS-Ergebnisse, die ebenfalls ein lediglich durchschnittliches Abschneiden deutscher Schülerinnen und Schüler aufzeigten, in Bildungspolitik und unter Fachlehrenden umfangreich diskutiert wurden, erreichten erst die Ergebnisse der PISA-Studie 2000 eine breite Öffentlichkeit und firmieren dort seither unter dem Terminus ‚PISA-Schock‘ (welchem eine gewisse Ironie beiwohnt, berücksichtigt man, dass PISA 2000 Ergebnisse hervorbrachte, die seit 1963 hätten aufgegriffen werden können). Der in den Medien formulierte ‚Skandal‘ führte zu einem breit angelegten bildungspolitischen Umdenken, begonnen bei der Einrichtung des Arbeitsstabs

„Forum Bildung“, welcher Empfehlungen für die strategische Umorientierung des deutschen Bildungswesens erarbeitet, über den Beschluss über die Weiterführung der deutschen Teilnahme an internationalen Schulvergleichsstudien (IGLU/PIRLS, TIMSS für die Grundschule, PISA), die Einrichtung des Instituts zur Qualitätsentwicklung im Bildungswesen (IQB Berlin), welches nationale Bildungsstandards entwickelt und ihre Umsetzung begleitet sowie Schulinspektionen bis hin zu einer gemeinsamen Bildungsberichterstattung von Bund und Ländern. Allerdings zeigt sich, nach etwas mehr als 15 Jahren ausführlichem Diskurs, dass es die oftmals erwartete einfache Kausalität ‚empirische Studie – einheitliche Ergebnisinterpretation – Ableitung und Implementierung bildungspolitischer Maßnahmen – Verbesserung‘ so nicht gibt: Zum einen, weil die ‚harten Fakten‘ immer interpretationsbedürftig sind, was zu unterschiedlichen Handlungsempfehlungen führen kann. Zum anderen, weil der Bildungspolitik eine andere Rationalität unterliegt als der Wissenschaft – erstere ist notwendigerweise normativ und geleitet durch parteipolitische Interessen (vgl. ähnlich Blömeke, 2005). So ist es nicht verwunderlich, dass die Reaktionen auf Large-Scale-Assessments den Handlungslogiken einzelner Statusgruppen folgend unterschiedlich ausfallen.

An den Hochschulen und Instituten wurden im Zuge der empirischen Wende eine ganze Reihe neuer Lehrstühle eingerichtet (Krüger, Kücker & Weishaupt, 2012), Sektionen und Gesellschaften für empirische Bildungsforschung gegründet, umfangreiche staatliche Drittmittel eingeworben (Schmidt-Hertha, 2012), Strukturen für die Erfassung, Bereitstellung und Nutzung von Forschungsdaten etabliert und internationale Netzwerke aufgebaut. Die Durchführung internationaler Schulvergleichsstudien erfolgt mittlerweile routiniert. Gleichzeitig findet um diese Tradition – und insbesondere die groß angelegten empirisch-quantitativen Studien – eine kritische Diskussion statt, die sich vor allem auf das Leistungs- bzw. Kompetenzverständnis im Vergleich zum Bildungsbegriff und auf dessen Messbarkeit bezieht. Mit dem Ziel, diesen Diskurs abzubilden und für den wissenschaftlichen Fortschritt nutzbar zu machen, veröffentlichten Baumert und Tillmann im vergangenen Jahr unter dem Titel ‚Empirische Bildungsforschung: Der kritische Blick und die Antwort auf die Kritiker‘ ein Sonderheft der Zeitschrift für Erziehungswissenschaft (vgl. Baumert & Tillmann, 2016).

Die bildungspolitischen Reaktionen auf die Teilnahme Deutschlands an Large-Scale-Assessments im Allgemeinen und auf deren Ergebnisse im Besonderen wurden oben bereits skizziert. An dem nachfolgenden Beispiel soll für einen spezifischen Kompetenzbereich ein aktueller bildungspolitischer Diskussions- und Entscheidungsprozess nachgezeichnet werden, der noch einmal besonders deutlich macht, welche Rolle empirische Befunde aus einer internationalen Schulleistungsuntersuchung für eine evidenzbasierte Steuerung spielen können. So wurden in

Deutschland – ausgehend von dem nur mittelmäßigen Abschneiden im internationalen Vergleich der computer- und informationsbezogenen Kompetenzen von Achtklässlerinnen und Achtklässlern im Rahmen der *International Computer and Information Literacy Study* (ICILS 2013) – umfangreiche bildungspolitische Maßnahmen diskutiert und beschlossen. Nach der Veröffentlichung der ICILS-2013-Ergebnisse im November 2014 wurde bereits im März 2015 ein gemeinsamer Antrag der Fraktionen der CDU/CSU und SPD zur Digitalen Bildung mit direkten Bezügen zu den Ergebnissen von ICILS 2013 vorgelegt (Deutscher Bundestag, 2015). Im April 2015 diskutierten nationale Experten, darunter ein Mitglied der wissenschaftlichen Leitung von ICILS 2013 in Deutschland, als Sachverständige im Ausschuss für Bildung, Forschung und Technikfolgenabschätzung des Deutschen Bundestags über das Thema ‚digitale Bildung‘ auf der Grundlage des oben genannten Antrags. In der im Oktober 2016 veröffentlichten Strategie ‚Bildungsoffensive für die digitale Wissensgesellschaft‘ des Bundesministeriums für Bildung und Forschung (BMBF), die auf die Förderung ‚digitaler‘ Kompetenzen in allen Bildungsbereichen abzielt, sind die Ergebnisse von ICILS 2013 ebenfalls explizit erwähnt. Im Dezember 2016 verabschiedete die Kultusministerkonferenz nach intensiven Diskussionen und Abstimmungsprozessen eine gemeinsame Strategie zur Bildung in der digitalen Welt, in die u.a. das ICILS-Kompetenzstufenmodell als Grundlage einging. Auf Bundesländerebene gilt es, diese gemeinsame Strategie mit dem Ziel ‚digitale Bildung‘ in Schulen in Deutschland zu stärken, nun umzusetzen.

Für die Wahrnehmung von Large-Scale-Assessments unter Schulpraktikern lassen sich bisher kaum empirische Untersuchungen finden – allerdings lässt sich aus Podiumsdiskussionen, Zeitungsartikeln, den Erfahrungen im Feld und Publikationen wie dem kürzlich veröffentlichten offenen Brief zu der Kompetenzorientierung im Mathematikunterricht (Abel et al., 2017) auf einen ‚Unmut‘ gegenüber der empirischen Wende und ihren Studien schließen. Hieraus ergibt sich eine gewisse Brisanz, sind es doch vor allem Lehrerinnen und Lehrer, die das empirisch gewonnene Wissen in die Klassenzimmer tragen können. Eine mögliche Ursache liegt in dem häufigen Fehlen einer empirisch-methodischen (Grund-)Ausbildung im Studium, ohne welche die Eigenheiten und mit ihnen die Geltungsbereiche vor allem groß angelegter Schulvergleichsstudien nur schwer nachvollzogen werden können. Neben solch einer Ausbildung fehle es den Lehrenden zudem an Zeit für die aufwendige Arbeit, aktuelle Erkenntnisse der empirischen Bildungsforschung für ihren Unterricht nutzbar zu machen (Jäger & Prenzel, 2005). So stellt der Transfer empirischer Evidenz in die schulische Praxis bis heute eine große Herausforderung dar (u.a. Hartmann, Decristan & Klieme, 2016), der mit verschiedenen Mitteln zu begegnen versucht wird. Zu diesen gehören beispielsweise die Berücksichtigung von Transferkonzepten bei der Begutachtung von Forschungsprojekten, die Durchfüh-

rung von besonderen Tagungsformaten oder die Veranstaltung spezifischer Fortbildungsformate, bei denen Forschung und Praxis gemeinsam diskutieren (wie z.B. auf der ‚SH-Sommeruniversität‘ am IPN in Kiel). Die im Rahmen der überarbeiteten Strategie der KMK zum Bildungsmonitoring aus dem Jahr 2015 betonte Bedeutung des Transfers von Befunden aus Schulleistungsuntersuchungen wird vor allem im Hinblick auf die besondere Rolle von Landesinstituten und Qualitätseinrichtungen der Länder formuliert, deren Aufgabe darin besteht „Forschungswissen in Kooperation mit wissenschaftlichen Einrichtungen adressatengerecht für die Schulen, die Bildungsadministration und die Bildungspolitik aufzubereiten und zu verbreiten“ (KMK, 2016, S. 15). Beispielhaft können hier die Transfer-Aktivitäten der Qualitäts- und UnterstützungsAgentur – Landesinstitut für Schule in Nordrhein-Westfalen (QUA-LiS NRW) genannt werden, z.B. in Form des Forschungsmonitors Schule ([www.forschungsmonitor-schule.de](http://www.forschungsmonitor-schule.de)), in dem Befunde aus empirischen Studien, auch aus Large-Scale-Studien, rezensiert werden, um dabei zu helfen, „die Aussagekraft und Relevanz der Forschungsergebnisse einzuschätzen“ (QUA-LiS NRW, 2017).

Dass die empirische Bildungsforschung und mit ihr die Teilnahme an internationalen Schulvergleichsstudien bei den beteiligten Akteuren unterschiedliche Reaktionen hervorruft, sollte als Chance verstanden werden: Aus einer gemeinsamen, sachlichen Diskussion können fruchtbare Konzepte entstehen, die schließlich den Lernenden in unserem Bildungswesen zugutekommen. Um solche Prozesse auch in Zukunft anregen zu können, bedarf es jedoch einer steten Anpassung der Vergleichsstudien an sich wandelnde Lebensumwelten und gesellschaftliche Bedingungen. Diese Veränderungen wahrzunehmen und in der Entwicklung der Tests und Fragebögen zu berücksichtigen, stellt eine ständige Aufgabe und Herausforderung in internationalen Large-Scale-Assessments dar. Dabei erschwert es vor allem die lange Vorbereitungszeit innerhalb eines Erhebungszyklus, solchen Entwicklungen immer ‚einen Schritt voraus‘ und zum jeweiligen Testzeitpunkt nicht ‚veraltet‘ zu sein. Dies ist vor allem bei computergestützten Testformaten der Fall, wenn es u.a. um Fragen der Gestaltung von Testoberflächen, den Einsatz (simulierter) Programme und Anwendungen oder die Darbietungsform (z.B. Laptop mit Maus vs. Tablet mit Touchpad) geht. Der Wandel von den klassischen papier- und bleistiftbasierten Erhebungsformaten hin zu einer digitalen Testung an Laptop und Tablet, wie sie 2019 zum Beispiel in TIMSS vollzogen wird, ermöglicht generell neue Formate der Kompetenzmessung: So können innovative Aufgabenformate administriert werden, die komplex und authentisch sind, Multimedia-Elemente enthalten und entsprechend einen hohen Interaktionsgrad aufweisen (z.B. die Simulation naturwissenschaftlicher Phänomene in einer Laborumgebung). Gerade für die in TIMSS mit erfasste naturwissenschaftliche Kompetenz kann durch die Nutzung

solcher *computer-based assessments* (CBA) die Passung von Testinhalten und schulischen Lerngegenständen im Sinne einer curricularen Testvalidität gestärkt und somit ein wichtiger Fortschritt hinsichtlich einer realitätsbezogenen Konstruktrepräsentation erzielt werden. Eine Herausforderung in theoretischer Hinsicht besteht allerdings bei überfachlichen Kompetenzen wie den in ICILS 2013 gemessenen computer- und informationsbezogenen Kompetenzen darin, dass diese im Zuge technologischer Entwicklungen und zunehmender Digitalisierung Veränderungen unterworfen sind und entsprechend einer stetigen Anpassung bedürfen (hierzu u.a. auch Fraillon, Schulz & Ainley, 2016).

Gerade dieser Bedarf einer konstanten Anpassung der Testinstrumente und Erhebungsverfahren an sich wandelnde Lern- und Lebensumwelten erfordert es, Perspektiven und Entwicklungsmöglichkeiten in internationalen Schulleistungsuntersuchungen aus unterschiedlichen Blickwinkeln und auf unterschiedlichen Ebenen herauszuarbeiten. Diesem Anliegen ist dieses Themenheft gewidmet.

Im ersten Beitrag von *Katrin Schulz-Heidorf* – nationale Projektleiterin von TIMSS 2019 in Deutschland – liegt der Fokus im Kontext von IGLU auf der Erfassung individualisierten Unterrichts im Rahmen von Large-Scale-Studien. Dabei werden Potenziale und Grenzen herausgearbeitet sowie alternative Möglichkeiten der Operationalisierung vorgeschlagen.

Der zweite Beitrag von *Knut Schwippert* – wissenschaftliche Leitung von TIMSS 2019 in Deutschland – und *Heike Wendt* – nationale Projektleiterin von TIMSS 2011, 2015 und PIRLS 2016 in Deutschland – beleuchtet im Kontext von TIMSS und PIRLS die Frage nach der Validität im Spannungsfeld zwischen einer Anpassung von Testinstrumenten an sich wandelnde Bildungssysteme sowie dem Bedarf an Trenduntersuchungen über mehrere Erhebungszyklen, für welche sich Testinstrumente nicht verändern sollten.

Der dritte Beitrag von *Birgit Eickelmann* – wissenschaftliche Leitung von ICILS 2018 in Deutschland – behandelt das internationale Zusatzmodul Computational Thinking zum zweiten Zyklus von ICILS im Jahr 2018 und fokussiert dabei auf die Konzeptionierung dieser programmiersprachenunabhängigen computerbezogenen Problemlösekompetenzen sowie die damit verbundenen Perspektiven für die empirische Bildungsforschung.

Der vierte Beitrag in diesem Themenheft von *Katrin Hahn-Laudenberg*, *Janina Jasper* und *Hermann Josef Abs* – wissenschaftliche Leitung und Projektteam ICCS 2016 – ist im Kontext der *International Civic and Citizenship Education Study* (ICCS) 2016 verortet und erarbeitet dabei ausgehend von einem Vergleich der Erhebungsinstrumente (u.a. zu kultureller Identität) Vorschläge für eine mögliche Veränderung der Erhebungsinstrumente.

Die beiden letzten Beiträge dieses Themenhefts sind aus der Perspektive zweier Institutionen geschrieben, die im Bereich der Large-Scale-Assessments seit Jahren eine wichtige Rolle einnehmen. Zunächst formulieren *Sabine Meinck*, *Agnes Stancel-Piqtak*, *Dirk Hastedt* und *Heiko Sibberns* aus der Perspektive der *International Association for the Evaluation of Educational Achievement* (IEA) aktuelle Entwicklungen und methodische Herausforderungen im Kontext von international vergleichenden Schulleistungsuntersuchungen. Abschließend beschäftigt sich der Beitrag von *Andreas Schleicher* aus OECD-Perspektive mit der Zukunft von PISA und bezieht sich dabei u.a. auf veränderte berufliche Anforderungen in der globalisierten Welt, zu denen vor allem Problemlöse-, aber auch soziale Kompetenzen wie Teamfähigkeit gehören. Erörtert werden Perspektiven, wie diese Kompetenzen erfasst werden können, zusammen mit den Potenzialen von ‚big data‘ im Bildungsreich.

Aus der Zusammenschau der Beiträge dieses Hefts werden somit vielschichtige Perspektiven und Entwicklungsbedarfe internationaler Large-Scale-Studien aus unterschiedlichen Blickwinkeln betrachtet, die – so hoffen wir – als eine Grundlage für ihre stete Weiterentwicklung genutzt werden können. Den Autorinnen und Autoren gilt unser herzlicher Dank, ebenso wie Heike Poppendieker für das Lektorat der Beiträge.

### Literatur

- Abel, U. et al. (2017, 17. März). *Mathematikunterricht und Kompetenzorientierung – ein offener Brief*. Frankfurt. Verfügbar unter <http://www.tagesspiegel.de/downloads/19549926/2/offener-brief.pdf> [12.04.2017].
- Baumert, J. & Tillmann, K.-J. (Hrsg.). (2016). Empirische Bildungsforschung. Der kritische Blick und die Antwort auf die Kritiker. *Zeitschrift für Erziehungswissenschaft*, 19 (Sonderheft 31). Wiesbaden: Springer Fachmedien.
- Blömeke, S. (2005). *Lehrerbildung, Lehrerhandeln, Schülerleistungen. Perspektiven nationaler und internationaler empirischer Bildungsforschung. Antrittsvorlesung 10. Dezember 2003*. Berlin: Humboldt-Universität zu Berlin.
- Bos, W., Postlethwaite, T.N. & Gebauer, M.M. (2010). Potenziale, Grenzen und Perspektiven internationaler Schulleistungsforschung. In R. Tippelt & B. Schmidt (Hrsg.), *Handbuch Bildungsforschung* (S. 275–295). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Deutscher Bundestag. (2015). *Drucksache 18/4422. Antrag der Fraktionen der CDU/CSU und SPD. Durch Stärkung der Digitalen Bildung Medienkompetenz fördern und digitale Spaltung überwinden*. Verfügbar unter <http://dip21.bundestag.de/dip21/btd/18/044/1804422.pdf> [12.06.2017].
- Fraillon, J., Schulz, W. & Ainley, J. (2016). *A global measure of digital and ICT literacy skills*. Paper commissioned for the Global Education Monitoring Report 2016, Education for people and planet: Creating sustainable futures for all. Verfügbar unter <http://unesdoc.unesco.org/images/0024/002455/245577e.pdf> [23.06.2017].

- Hartmann, U., Decristan, J. & Klieme, E. (2016). Unterricht als Feld evidenzbasierter Bildungspraxis? *Zeitschrift für Erziehungswissenschaft*, 19, 179–199.
- Jäger, M. & Prenzel, M. (2005). Erfolgreiche Bildungssysteme nutzen wissenschaftliche Erkenntnisse. Überlegungen zur Verwertung pädagogischen Wissens. In H. Heid & C. Harteis (Hrsg.), *Verwertbarkeit. Ein Qualitätskriterium (erziehungs-)wissenschaftlichen Wissens?* (S. 163–182). Wiesbaden: VS Verlag für Sozialwissenschaften.
- KMK (Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland). (2006). (Hrsg.). *Gesamtstrategie der Kultusministerkonferenz zum Bildungsmonitoring*. München: Wolters Kluwer.
- KMK (Sekretariat der Kultusministerkonferenz). (Hrsg.). (2016). *Gesamtstrategie der Kultusministerkonferenz zum Bildungsmonitoring*. Berlin: Kluwer.
- Krüger, H.-H., Kücken, C. & Weishaupt, H. (2012). Personal. In W. Thole, H. Faulstich-Wieland, K.-P. Horn, H. Weishaupt & I. Züchner (Hrsg.), *Datenreport Erziehungswissenschaft 2012* (S. 137–158). Opladen: Barbara Budrich.
- 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.
- QUA-LiS NRW (Qualitäts- und UnterstützungsAgentur – Landesinstitut für Schule). (2017). *Forschungsmonitor Schule*. Verfügbar unter <https://www.schulentwicklung.nrw.de/fms/fragen/fragen.php> [13.06.2017].
- Roeder, P.M. (2003). TIMSS und PISA – Chancen eines neuen Anfangs in Bildungspolitik, -planung, -verwaltung und Unterricht. Endlich ein Schock mit Folgen? *Zeitschrift für Pädagogik*, 49 (2), 180–197.
- Schmidt-Hertha, B. (2012). Forschung und Publikationskultur. In W. Thole, H. Faulstich-Wieland, K.-P. Horn, H. Weishaupt & I. Züchner (Hrsg.), *Datenreport Erziehungswissenschaft 2012* (S. 159–180). Opladen: Barbara Budrich.
- Schultze, W. & Riemenschneider, L. (1967). Eine vergleichende Studie über die Ergebnisse des Mathematikunterrichts in zwölf Ländern. Forschungsbericht. *Mitteilungen und Nachrichten (Deutsches Institut für Internationale Pädagogische Forschung)*, 46/47, 1–34.
- van Ackeren, I. (2002). Von FIMS und FISS bis TIMSS und PISA. Schulleistungen in Deutschland im historischen und internationalen Vergleich. *Die Deutsche Schule – Zeitschrift für Erziehungswissenschaft, Bildungspolitik und pädagogische Praxis*, 94 (2), 157–175.