

Vorbemerkung

Der folgende Vorschlag ist als Gesprächsimpuls gedacht. Wir stellen ihn in das „Forum Kritische Pädagogik“ und schicken ihn an ausgewählte KollegInnen zu einer ersten Diskussion.

Reaktionen erbitten wir an unsere Mailadressen

hey mann@paedagogik.uni-siegen.de

(Hans Werner Heymann)

oase@paedagogik.uni-siegen.de

(Hans Brügelmann)

Wir überlegen, im Anschluss an diese virtuelle Diskussion bei einer Stiftung eine kleine Tagung mit interessierten 12-15 KollegInnen zu beantragen, auf der die Idee inhaltlich ausgearbeitet und auch im Blick auf konkrete Kooperationsmöglichkeiten (z. B. mit Datenbanken) und auf wirksame sowie unaufwändige Institutionalisierungsformen diskutiert werden soll.

Klärung und Übersetzung von Forschung als Dienstleistung für die pädagogische Praxis

Plädoyer für die Einrichtung einer „Evaluationsstelle für nutzerorientierte Bildungsforschung“

von Hans Brügelmann & Hans Werner
Heymann
Universität Siegen

[Werkstattfassung, 25.3.2006]

Derzeit erleben wir in der Bildungsforschung einen Boom an empirischen Untersuchungen, den zweiten seit der ersten „empirischen Wende“ um 1970-75. Angeregt durch die internationalen Leistungsstudien TIMSS, PISA, IGLU usw. vergeben Ministerien große Evaluationsstudien. Parallel unterstützen Fördereinrichtungen wie die DFG Programme zur Entwicklung der Grundlagenforschung im Bildungsbereich. So verdienstvoll diese Aktivitäten und so nützlich ihre Ergebnisse oft auch sind - es zeichnen sich einige Risiken dieses erneuten empirischen Aufbruchs ab: Die Unübersichtlichkeit der Befunde und die mangelnde Durchschaubarkeit ihrer komplexen Gewinnung führen auf Seiten der potenziellen AbnehmerInnen zu naiven (unzulässig vereinfachenden) Deutungen oder Verständnislosigkeit und verleiten umgekehrt auf Seiten der WissenschaftlerInnen zur Inanspruchnahme von Deutungsmonopolen und Immunisierung gegen fachliche Kritik.

Probleme der zweiten „empirischen Wende“

Die plötzlich verfügbaren Forschungsmittel haben in den interessierten Einrichtungen Hektik ausgelöst. Vor allem bildungspolitisch initiierte Studien stehen unter einem erheblichen Zeitdruck, der eine gründliche Vorbereitung, aber auch eine nachhaltige Auswertung der Untersuchungen erschwert. Kaum ist ein Projekt begonnen, wird schon das nächste beantragt. Das gilt inzwischen auch für universitäre Forschungsprojekte, da sich Hochschulinstitute zunehmend über Drittmittel finanzieren oder zumindest legitimieren müssen. Es entstehen Berge von Daten, deren Bedeutung für Außenstehende und zum Teil auch für Insider kaum mehr zu überschauen ist.

Damit verbunden ist ein zweites Problem. Jede Studie entfaltet ihr eigenes Panorama. Das Set an Fragestellungen, der jeweilige theoretische Hintergrund, die eingesetzten methodischen Verfahren, die Stichproben und Kontexte variieren erheblich. Diese Unterschiede erschweren den interessierten BildungspolitikerInnen und SchulpraktikerInnen die Einschätzung, wie relevant die Befunde einer konkreten Studie für ihre aktuell anstehenden Aufgaben und Entscheidungen sind. Noch schwieriger ist es, die Ergebnisse unterschiedlicher Studien zueinander in Beziehung zu setzen und Differenzen oder Widersprüche zwischen ihnen aufzulösen. In Ansätzen leisten das besser ausgestattete Projekte in ihren Berichten selbst. Aber auch sie beziehen andere Studien nur selektiv ein - und interpretieren ihre Ergebnisse durch die eigene Brille. Wir haben bereits früher den Vorschlag gemacht, zumindest die Ergebnisse von politisch bedeutsamen Studien wie PISA oder VERA nicht (allein) von den ForscherInnen selbst auswerten und interpretieren zu lassen¹. Die Idee, schon für den Erstbericht theoretisch differierende Interpretationen auszuwählen, hat bisher keine Resonanz gefunden. Dies ist wegen des zusätzlichen Zeitaufwands - und auch im Blick auf die Karriereinteressen der ProjektmitarbeiterInnen - verständlich, aber für die Transparenz wissenschaftlicher Befunde hinderlich: Der gewünschten wissenschaftlichen Rationalität würde man durch die avisierte soziale Kontrolle der unvermeidbaren Sub-

¹ Vgl. Brügelmann (2005, 343) - konkret zum ersten Mal vorgeschlagen (und gescheitert...) für die Publikation des deutschen Berichts zur IEA-Lesestudie von 1991 (Lehmann u. a. 1995). Wir kommen am Ende unseres Beitrags auf diesen Vorschlag zurück.

jektivität im Forschungsprozess besser gerecht werden.

Nun gibt es in der Wissenschaft neben der Durchführung von Einzelprojekten auch den Anspruch und eine Tradition, die Ergebnisse bereichsbezogen zu sichten und zu systematisieren. In den angelsächsischen Ländern sind es sogenannte *Reviews*, die Überblicke schaffen sollen. Analoge Formen finden sich im deutschsprachigen Raum. Aber auch sie haben ihre Grenzen.

Rezensionen beschränken sich meist auf einzelne Publikationen und selbst Sammelbesprechungen beziehen nur Neuerscheinungen ein. Zeitlich weiter ist die Perspektive von **Projektanträgen** und **Dissertationen**, aber deren Kapitel zum *State of the Art* sind jeweils auf das konkrete Untersuchungsinteresse fokussiert. Grundsätzlicher und umfassender sind die allerdings nicht sehr häufigen **Habilitationsschriften** angelegt. Ihr Ziel ist es zudem, den *theoretischen* Rahmen der Forschung auszu-leuchten und zu systematisieren. Die Bedeutung der Befunde für die Lösung praktischer Probleme wird nur selten thematisiert.

Für die Praxis bedeutsamer, aber ebenfalls selten sind von Politik, Verwaltung oder Verbänden vergebene **Gutachten**, die den Stand der Forschung im Blick auf konkrete Problemlagen resümieren².

Das Problem aller genannten Formen von *Research synthesis* ist, dass die Auswahl, Ordnung und Deutung der Forschungsergebnisse in hohem Maße von den Personen abhängt, die die Sichtung vornehmen. Als

² In Deutschland werden sie noch am ehesten von der Bund-Länder-Kommission für Bildungsplanung vergeben, vgl. etwa die Bestandsaufnahme von Entwicklungs- und auch Forschungsprojekten zur Förderung von MigrantInnen in Gogolin u. a. (2003).

Alternative ist deshalb vor gut 30 Jahren in den USA das Instrument der **Metaanalyse** eingeführt worden³ und dort inzwischen gut etabliert. In Metaanalysen werden die statistischen Daten verschiedener Studien nach explizit definierten Kriterien miteinander verrechnet. Damit wird der Anspruch erhoben, den Einfluss der jeweils besonderen Kontextbedingungen in den Einzelstudien sowie den persönlichen Einfluss der AuswerterInnen zu reduzieren. Ausschalten lässt sich das subjektive Moment aber auch hier nicht. Es kommt bereits zum Tragen bei der Entscheidung über die anzulegenden Kriterien, wenn die Frage ansteht, welche Studien überhaupt als forschungsmethodisch adäquat in die geplante Metaanalyse einbezogen werden sollen. Denn für die „methodische Qualität“ von Studien gibt es unterschiedliche Maßstäbe. Auch für die Alternativen, ob man die berücksichtigten Studien einzeln zählt, also ihre Ergebnisse gleichgewichtig verrechnet, oder ob man die Kennwerte nach der Zahl der jeweils in der Studie untersuchten Fälle gewichtet, gibt es jeweils gute Gründe. Am stärksten kommt die persönliche Position der WissenschaftlerInnen, die die Metaanalyse durchführen, in der verbalen Zusammenfassung der Rechenergebnisse zum Ausdruck. An dieser Stelle wird notwendigerweise fokussiert, gewichtet, geglättet, gedeutet - denn Zahlen sprechen nicht für sich.

Für LeserInnen aus Politik und Praxis sind die Schritte, die von der Fülle der Einzeluntersuchungen zu einem komprimierten Stand der Forschung führen, nur schwer nachvollziehbar. Oft hat auch die inhaltliche Zusammenfassung noch eine technische Komplexität, die selbst Fachleute aus

³ Vgl. Glass (1976; 1977); Hunter u. a. (1982/2004); für den deutschen Sprachraum: Fricke/Treinius (1985).

Politik und Praxis nicht auflösen können⁴. Zudem werden in solchen Forschungsberichten Fachbegriffe wie „Signifikanz“⁵ oder „Risiko“⁶ von vielen LeserInnen nicht richtig verstanden⁷ - oder schon von den ForscherInnen selbst unzutreffend bzw. zumindest missverständlich verwendet. Insofern bedürfen Forschungsberichte einer fachkundigen und verständlichen Übersetzung.

Aber leisten nicht genau das **WissenschaftsjournalistInnen**? In der Tat gibt es Beiträge in Zeitschriften wie „Spektrum der Wissenschaft“ und für den Bildungsbereich seit einigen Jahren in der ZEIT, die sich diesem Anspruch zu stellen versuchen. Die sprachpsychologischen Publikationen des ZEIT-Mitarbeiters Dieter E. Zimmer (z. B. 1986) haben da Maßstäbe gesetzt. Doch meist entstehen solche Artikel aus aktuellem Anlass, die AutorInnen haben nur wenige Wochen oder gar nur Tage Zeit für ihre Recherchen und sie beschränken sich in der Regel auf die Darstellung einzelner Studien oder Projekte, die gerade publiziert worden sind.

Festzuhalten ist: Bei allem Bemühen um wissenschaftliche Rationalität ist empirische Forschung nicht völlig loslösbar von der Subjektivität und persönlichen Sicht der ForscherInnen. Das gilt bereits für die Durchführung und Auswertung von Einzelstudien, bei denen schon die AutorInnen in der Interpretation der Daten nicht

⁴ Deshalb werden z. B. die AutorInnen nach Veröffentlichung ihrer Berichte von Ministerien zu mehr oder weniger vertraulichen Hintergrundgesprächen eingeladen, um die „tatsächliche“ Bedeutung der Ergebnisse zu erläutern - aus ihrer persönlichen Sicht...

⁵ Vgl. Sedlmeier (1996).

⁶ Vgl. Gigerenzer (1998; 2002).

⁷ Vgl. Krämer (1994); Beck-Bornholdt/ Dubben (1997; 2001).

selten deren Aussagekraft überziehen oder Randbedingungen nicht deutlich genug ausweisen⁸. Es gilt erst recht für die Erstellung von Forschungssynthesen und Metaanalysen auf der Basis dieser Studien. Wie am Beispiel der Metaanalysen erläutert wurde, sind subjektive Deutungen unvermeidlich und machen die besondere Qualität von interpretativen Zusammenfassungen im Vergleich zu einer (nur scheinbar) rein numerischen Verrechnung von Studien aus⁹. Zu bedenken sind zudem forschungsmethodische Einschränkungen bzw. Unzulänglichkeiten oder gar Interessensbindungen, die für Laien-LeserInnen nicht immer leicht zu durchschauen sind¹⁰.

Dann aber geht es insgesamt um Zweierlei:

⁸ Vgl. die allgemeine Kritik von Rauin (2004) und beispielhaft die Kritik von Block/ Klemm (2005) an den Auswertungen von PISA-2003 und von Brügelmann (1998) zur SCHOLASTIK-Studie von Weinert/ Helmke; als einschlägige Arbeiten älteren Datums sei auf die exemplarische Kritik an der Gesamtschulstudie von Treiber/Weinert (1985) hingewiesen: Beck/ Bromme/ Heymann u. a. (1988a u. 1988b); ähnlich im Ansatz Heymann (1984).

⁹ Vgl. z. B. Sedlmeier (1996, 59) zur Deutung von Signifikanztests: „Bei jeder Interpretation von Daten ist explizit subjektives Urteil mit im Spiel. Was angestrebt werden kann, ist ein durch eine geeignete Datenanalyse gewonnener Konsens unter den Experten in einem bestimmten Gebiet, nicht aber das Ersetzen eines Rituals durch ein anderes.“ Hilfreiche Hinweise zur Unterscheidung von statistischer „Signifikanz“ und inhaltlicher „Bedeutung“ finden sich bei Lind (2005).

¹⁰ Aus diesem Grund ist gerade in den USA eine Evaluationsstelle für die Prüfung von privat finanzierten Studien gegründet worden: "Education reports released by private think tanks are now being scrutinized by expert, academic reviewers associated with the just-launched Think Tank Review Project. Within two weeks of a report's release the project provides reviews to policy makers, the media, and the public. Reviews are posted on the ASU Education Policy Studies Laboratory Web site at: <http://thinktankreview.org>".

1. das persönliche Urteil bei der *Sichtung* und *Zusammenfassung* von Befunden aus verschiedenen Quellen transparent zu machen und - so weit möglich - durch soziale Mechanismen zu kontrollieren¹¹;
2. die Bedeutung von Befunden - und die Bedeutung eventueller Differenzen zwischen ihnen - für Handlungsoptionen in Politik und Praxis *verständlich* zu übersetzen.

Perspektiven für eine fachliche fundierte Sichtung und Übersetzung von Forschungsbefunden

Wir schlagen vor, auf der Grundlage unserer bereits vorliegenden Vorarbeiten¹² eine „Evaluationsstelle für nutzerorientierte Bildungsforschung (EVANUB)“¹³ aufzubauen.

Deren Angebote könnten interessant sein für Ministerien, Verbände, JournalistInnen. Die Arbeitsstelle könnte bei aktuellen Anfragen sehr rasch auf relevante und solide Arbeiten verweisen, die bereits vorliegen. Zu mittelfristig anstehenden Entscheidungen könnte sie eigene Stellungnahmen erarbeiten und publizieren, vor allem aber Auftragsgutachten übernehmen. Diese werden je nach dem finanziellen Rahmen unterschiedlich anspruchsvoll ausfallen können.

¹¹ Vgl. dazu und zum Folgenden ausführlicher unsere Vorschläge in Brügelmann (1974).

¹² Vgl. Heidenreich/ Heymann (1973); Brügelmann (1983); Balhorn/ Brügelmann (1986ff.); Heymann (1984); Beck/ Bromme/ Heymann u. a. (1988a/b); Heymann (1996/2003); Brügelmann, H., u. a. (1998ff.); Brügelmann (2005); Arbeitsgruppe Primarstufe (2006).

¹³ Es ist noch ein inhaltlich treffenderer und „kürzelkräftigerer“ Name zu suchen...

Nimmt man die in dieser Denkschrift geäußerte Kritik ernst, müsste die Arbeitsstelle sowohl über fachlich-inhaltliche als auch über forschungsmethodische Kompetenz verfügen. Dies bedeutet in der Aufbauphase eine Konzentration auf ausgewählte Arbeitsgebiete¹⁴, die sich aber durch - ggf. punktuelle, d. h. projektbezogene - Kooperation mit KollegInnen an anderen Hochschulen leicht erweitern lässt.

Entscheidend ist, wie diese Evaluationsstelle ihre eigene Funktion definiert. Sie sollte sich nicht zum Obergutachter über die Richtigkeit inhaltlicher Deutungen aufschwingen. Vielmehr geht es darum, die *Sichtung* in einem Verfahren organisieren, das Levine (1973; 1974) analog zum juristischen Prozess als „adversary model“ konzipiert hat. Die Aufgaben der Arbeitsstelle bestehen dann darin,

- **Kriterien** zu entwickeln, mit deren Hilfe die Qualität von Studien bewertet werden kann - vor allem, a-

¹⁴ Zurzeit verfügen wir über folgende Fachkompetenzen:

- Sprachdidaktik, vor allem im Vor-, Grund- und Sonderschulbereich (Backhaus, Brinkmann, Brügelmann, Coelen, Franzkowiak)
- Mathematikdidaktik, vor allem im Sekundarbereich (Heymann, Siewert)
- Schulpädagogik und allgemeine Didaktik beider Schulstufen (Brügelmann, Heymann, Müller-Naendrup)

Außerdem decken die genannten Personen - gestützt durch die Kooperation mit anderen KollegInnen an dieser Hochschule - folgende Methodenkompetenzen ab:

- interpretative („qualitative“) Erhebungs- und Auswertungsverfahren
- standardisierte („quantitative“) Erhebungs- und statistische Auswertungsverfahren.

ber nicht ausschließlich in forschungsmethodischer Hinsicht¹⁵;

- für jeden Auftrag klare **Fragestellungen** zu formulieren bzw. die Anfragen des Auftraggebers aus ihrer Fachkenntnis heraus zu spezifizieren, um die Gutachten in einer **bearbeitbaren** Form zu **fokussieren**;
- mindestens zwei (Gruppen von) **GutachterInnen auszuwählen**, die unterschiedliche Positionen in den zu prüfenden Fragen vertreten, sozusagen als „Anwälte“ für diese Positionen;
- in Zusammenarbeit mit diesen GutachterInnen diejenigen Studien als **„Evidenz“ zu bestimmen**, die im Blick auf die Fragen ausgewertet und bewertet werden soll¹⁶;
- die getrennt entstandenen Gutachten - sozusagen als „stellvertretende Leserin“ - auf **Übereinstimmungen und Differenzen** hin zu **sichten** und mögliche Gründe für Abweichungen anzubieten.

Das Verfahren verzichtet also auf den (illusionären) Anspruch, Subjektivität aus dem Interpretationsprozess methodisch-technisch auszuschalten. Es macht diese vielmehr explizit und durch den sozialen Mechanismus der Pro- und Contra-Anwälte

¹⁵ Vgl. etwa zu den *Guidelines des 'Evidence for Policy and Practice Information and Coordinating Centre (EPPI-Centre)'* u. a. Harlen/ Deckin Crick (2002, 19-29); Harlen (2004, 22-32); Newman u.a. (2004).

¹⁶ Neben Sekundäranalysen von Einzelstudien können auch bereits vorliegende Metaanalysen und Forschungssynthesen genutzt werden.

durchsichtig und damit in ihren Auswirkungen kontrollierbar.

Die Aufgabe der EVANUB selbst wäre es, am Ende die zentralen Aussagen der Stellungnahmen - sozusagen als „stellvertretende Leserin“ - in fairer Weise zusammenzufassen (ohne Differenzen aufzulösen) und so zu übersetzen, dass sie handlungsrelevant werden können. Statt nach Kriterien wie „richtig“ vs. „falsch“ zu urteilen, würde die Zusammenfassung Potenziale und Risiken von Handlungsoptionen herausarbeiten.

Kumulativ könnte so ein Fundus an Überblicken zu zentralen bildungspolitischen und unterrichtspraktischen Fragen entstehen. Zudem könnte die Arbeitsstelle stilbildend wirken, in dem sie die hierarchische verstandene Expertenautorität von Wissenschaft ersetzt durch dialogische Formen einer mehrperspektivischen Diskussion von empirischen Daten, die deren Mehrdeutigkeit im Bewusstsein der Öffentlichkeit hält.

Auf einer zweiten Stufe könnte die Kapazität der Arbeitsstelle ausgebaut werden für die Durchführung eigener Reanalysen¹⁷ und Metaanalysen. Diese sind aber deutlich aufwändiger und zurzeit auch nicht notwendig, weil der Fundus der vorhandenen Einzelstudien und Metaanalysen bei weitem noch nicht ausgeschöpft ist.

¹⁷ Reanalysen sind eigenständige Auswertungen auf der Grundlage der Urdaten einer Fremdstudie, während Sekundäranalysen sich auf die Auswertungen der Originalstudie beziehen und diese kritisch bewerten.

Literatur

Arbeitsgruppe Primarstufe (2006, in Vorb.): Ziffernoten und alternative Beurteilungsformen. Eine Expertise der Arbeitsgruppe Primarstufe an der Universität Siegen (Hans Brügelmann u.a.) im Auftrag des Grundschulverbands e.V.: Frankfurt.

Balhorn, H., Brügelmann, H., u. a. (1986ff.): Jahrbücher Lesen und Schreiben der Deutsche Gesellschaft für Lesen und Schreiben. Faude Verlag: Konstanz/ Libelle: Lengwil.

Beck, M./ Bromme, R./ Heymann, H. W. u. a. (1988a): Gefangen im Datenlabyrinth. Kritische Sichtung eines Forschungsberichts zum schulischen Chancenausgleich. In: Zeitschrift für Pädagogische Psychologie, 2. Jg., 91-111.

Beck, M./ Bromme, R./ Heymann, H. W. u. a. (1988b): Chancenausgleich: Ideologie und Empirie. Eine Antwort auf Weinert. In: Zeitschrift für Pädagogische Psychologie, 2. Jg. 173-178.

Beck-Bornholdt, H.-P./ Dubben, H.-H. (1997): Der Hund, der Eier legt. Erkennen von Fehlinformationen durch Querdenken. Rororo 60359: Reinbek.

Beck-Bornholdt, H.-P./ Dubben, H.-H. (2001): Der Schein der Weisen. Irrtümer und Fehlurteile im täglichen Denken. Hoffmann und Campe: Hamburg.

Block, R./ Klemm, K. (2006): PISA 2003: differenzierende Bemerkungen zum neuen Ländervergleich. In: Schulverwaltung NRW, H. 2/2006, 38-40.

Brügelmann, H. (1974): Towards checks and balances in educational evaluation - On the use of social control in research design. In: MacDonald/ Walker (1974, 25-66).

Brügelmann, H. (1983): Kinder auf dem Weg zur Schrift - eine Fibel für Lehrer und Laien. Ekkehard Faude: Konstanz (7. Auflage 2000).

Brügelmann, H. (1998): Längsschnittstudie in der Diskussion. Rezension von: Franz Weinert/ Andreas Helmke „Entwicklung im Grundschulalter“. In: Grundschulzeitschrift, 12. Jg., H. 117, 55-56.

Brügelmann, H. (2005): Schule verstehen und gestalten - Perspektiven der Forschung auf Probleme von Erziehung und Unterricht. Libelle: CH-Lengwil. Ab 1.10.2005 fortlaufend [etwa alle sechs Monate] aktualisiert unter: <http://www.agprim.uni-siegen.de/schuleverstehen>

Brügelmann, H., Fölling-Albers, M., Richter, S., Speck-Hamdan, A. (1998ff.): Jahrbücher Grundschule. Friedrich Verlag: Seelze/ Grundschulverband: Frankfurt.

- Fricke, R./ Treinies, G. (1985): Einführung in die Metaanalyse. Methoden der Psychologie, Bd. 3. Hans Huber: Bern u. a.
- Glass, G.V. (1976): Primary, secondary, and meta-analysis of research. In: Educational Researcher, Vol. 11, 3-8.
- Glass, G.V. (1977): Integrating findings: The meta-analysis of research. In: Shulman (1977, 351-379).
- Gigerenzer, G. (1998). We need statistical thinking, not statistical rituals. In: Behavioral and Brain Sciences, Vol. 21, 199-200.
- Gigerenzer, G. (2002): Das Einmaleins der Skepsis. Über den richtigen Umgang mit Zahlen und Risiken. Berlin Verlag: Berlin.
- Gogolin, I., u. a. (2003): Förderung von Kindern und Jugendlichen mit Migrationshintergrund. Gutachten. Materialien zur Bildungsplanung und zur Forschungsförderung, Heft 107. Bund-Länder-Kommission für Bildungsplanung und Forschungsförderung: Bonn.
- Harlen, W. (2004b) A systematic review of the evidence of the impact on students, teachers and the curriculum of the process of using assessment by teachers for summative purposes. In Research Evidence in Education Library. EPPI-Centre, Social Science Research Unit, Institute of Education: London.
- Harlen, W./ Deakin Crick, R. (2002): A systematic review of the impact of summative assessment and tests on students' motivation for learning (EPPI-Centre Review, version 1.1*). In: Research Evidence in Education Library. Issue 1. EPPI-Centre, Social Science Research Unit, Institute of Education: London.
- Heidenreich, W.-D./ Heymann, H. W. (1976): Lehr-Lern-Forschung. Neuere unterrichtswissenschaftliche Literatur im Spiegel eines neuen Forschungsansatzes. In: Zeitschrift für Pädagogik, 22. Jg., 225-251.
- Heymann, H. W. (1984): Modellierungsprobleme bei der Erforschung des Lehrer- und Schülerverhaltens. In: Unterrichtswissenschaft, 12. Jg, 232-251.
- Heymann, H. W. (1996): Allgemeinbildung und Mathematik. Beltz: Weinheim (engl. 2003)
- Heymann, H. W. (2003): Why teach mathematics - A focus on general education. Kluwer: Dordrecht et al. (dt. 1996).
- Hunter, J., et al. (1982): Meta-Analysis: Cumulating research findings across studies. Sage: Beverly Hills/ Newbury Park, Cal. (new ed. 1990; 2004).
- Lehmann, R.H., u. a. (1995): Leseverständnis und Lesege-wohnheiten deutscher Schülerinnen und Schüler. Beltz: Weinheim/ Basel.
- Levine, M. (1973): Scientific method and the adversary model. In: Journal of Educational Evaluation, Vol. 2, No. 4, 1-3.
- Levine, M. (1974). Scientific method and the adversary model. Some preliminary thoughts. In: American Psychologist, Vol. 29, 661-677.
- Lind, G. (2005): Effektstärken: Statistische versus praktische und theoretische Bedeutsamkeit. → www.uni-konstanz.de/ag-moral/pdf/Lind-2005_Effektstaerke-Vortrag.pdf [Abruf: 16.3.06]
- MacDonald, B./ Walker, R. (eds.) (1974): Innovation, Evaluation, Research and the Problem of Control. Some interim papers. SAFARI project/ Centre for Research in Education. University of East-Anglia: Norwich.
- Newman, M., et al. (2004): Improving the usability of educational research: Guidelines for the reporting of empirical primary research studies in education. Evidence for Policy and Practice Information and Coordinating Centre (EPPI- Centre)/ Social Science Research Unit (SSRU). Institute of Education/ University of London.
- Rauin, U. (2004): Die Pädagogik im Bann empirischer Mythen - Wie aus empirischen Vermutungen scheinbare pädagogische Gewissheit wird. In: Pädagogische Korrespondenz (Zeitschrift für kritische Zeitdiagnostik in Pädagogik und Gesellschaft), H. 32/2004, 39-49.
- Sedlmeier, P. (1996): Jenseits des Signifikanztest-Rituals: Ergänzungen und Alternativen. In: Methods of Psychological Research Online, Vol. 1, No. 4, 41-63.
→ <http://www.mpr-online.de/issue1/art3/sedlmeier.pdf> [Abruf: 9.11.2005]
- Shulman, L.S. (ed.) (1977): Review of research in Education. Vol. 5. Peacock: Itasca, Ill.
- Treiber, B./ Weinert, F. (1985): Gute Schulleistungen für alle? Psychologische Studien zu einer pädagogischen Hoffnung. Aschendorff: Münster.
- Zimmer, D. E. (1986): So kommt der Mensch zur Sprache. Über Spracherwerb, Sprachentstehung und Sprache & Denken. Haffmanns Verlag: Zürich.