

# Meta-Analysen als Wegweiser? Zur Rezeption der Studie von Hattie in der Politik.

Georg Lind

Einige Bildungspolitiker vertreten gegenwärtig die These, dass Art und Qualität des Unterrichts wichtiger für den Lernerfolg von Schülern seien als die Struktur unseres Schulsystems. Sie sehen sich in dieser These durch die empirische Bildungsforschung bestätigt, vor allem durch die Synthese von über 800 Meta-Analysen über die Bedingungen von (Schul-)Leistung von Professor John Hattie (2009), die sich aktuell großer Beliebtheit erfreut, nicht nur in der Zeit (Spiewak, 2013).

Jeder Versuch, unser Schulsystem durch Einführung von Gemeinschaftsschulen inklusiver und gerechter zu machen, würde, so ihre These, zu einer deutlichen Verschlechterung der Schulleistungen führen. Diese These steht in Zusammenhang mit der gegenwärtigen öffentlichen Diskussion über unser hochselektives und sozial stark gegliedertes Schulsystem, das nach Meinung vieler Experten ökonomische Ungleichheit produziert und nach Einschätzung der Vereinten Nationen gegen die Menschenrechte der Kinder verstößt, weil es ihnen Bildungschancen vorenthält.

Der Gegensatz – effektiver Unterricht versus Gemeinschaftsschule – wird überhaupt nur vor dem Hintergrund des aktuellen bildungspolitischen Streits verständlich. Er ist dabei in doppelter Hinsicht wenig überzeugend: Damit Unterricht effektiv ist, müssen schließlich immer auch die Rahmenbedingungen, muss also das “System” stimmen. Lehrer müssen gut ausgebildet und angemessen bezahlt werden, die Schulen müssen erreichbar und baulich so gestaltet sein, dass Lehrer wirklich lehren und Schüler wirklich lernen können. Und eben auch die Schulstruktur muss fördernd und inklusiv statt selegierend und ausschließend sein, damit Kinder gern und viel lernen. So ganz ernst scheinen die politischen Kommentatoren ihr (vermeintliches) Anti-Struktur-Argument aber auch selbst nicht zu nehmen. Sonst hätten sie nämlich nicht das 13. Schuljahr abgeschafft und das Bachelor-Studium eingeführt, sondern die Lehrerbildung reformiert und die Hochschuldidaktik verbessert. Wirklich glaubwürdig sind also weder die Verfechter noch Motive der These einer „Effizienzsteigerung“ allein mittels „besseren Unterrichts“.

Der Ansatz vom Primat des Unterrichts über die Struktur überzeugt aber auch deshalb nicht, weil die Forschung, auf die verwiesen wird, sie gar nicht stützt – und das, obwohl der Kronzeuge Hattie diese These in seinem Buch *Visible Learning* selbst propagiert. Hattie fasst in seiner “Synthese” die statistischen Befunde von

über 800 so genannten Meta-Analysen zusammen, die ihrerseits wiederum die Befunde von Einzelstudien enthalten, die zusammen genommen auf Testdaten von einigen Millionen Schülern und Schülerinnen beruhen. Das wirkt ungemein beeindruckend, und soll es wohl auch. Die Vertreter der These Pädagogik vor Struktur verweisen denn auch immer auf diese Zahlen, wenn sie Hatties Befunde als Belege oder Beweise oder gar Wahrheiten anführen. Was aber ist falsch an dieser Studie? Warum taugt sie nicht als Beleg für das Behauptete? Warum kann sie tatsächlich nicht, wie der Autor und seine Bewunderer meinen, als Anleitung für die Steigerung des Lernerfolgs unserer Schüler dienen?

Die wichtigste Kritik richtet sich, so meine ich, a) gegen die Methode der Meta-Analyse selbst, b) gegen die Anlage und Analyse der empirischen Einzelstudien, auf die sich diese stützt, und c) gegen die Verabsolutierung von „Effekten“, ohne Abschätzung von Kosten und Ertrag. Gene V. Glass, der Erfinder der so genannten “Meta-Analyse” von empirischen Studien, hatte in seiner ersten Veröffentlichung Anspruch und Grenzen dieser neuen Methode bereits präzise umrissen: “Meta-Analyse“, so schreibt er, “bedeutet eine rigorose Alternative zu der kausal-erzählenden Diskussion von Forschungsstudien“, die es sogar ermöglichen würde, aus empirischen Einzelstudien mit einem “halben Dutzend Planungs- und Analysefehlern noch valide Informationen” herauszufiltern. „Aber“, so Glass weiter, „unser Problem ist es, in diesen Informationen Wissen zu finden. Wir benötigen Methoden, um diese Studien ordentlich zusammenzufassen, so dass aus dem Wust von Einzelbefunden Wissen extrahiert werden kann” (Glass, 1976, S. 3-4). Eine gute Meta-Analyse sei daher“ das intellektuelle Gegenstück zu originärer empirischer Forschung”.

Schon Glass, der zahlreiche Meta-Analysen zur Unterrichts- und Schulqualität durchgeführt hat, hat diesen hohen Anspruch krass ignoriert. Der “intellektuelle” Anspruch einer Meta-Analyse besteht heute nur noch darin, sehr viele Dokumente von Hilfskräften nach Studien zu einem bestimmten Thema absuchen zu lassen, ein paar Ergebnisdaten (Korrelationen, Mittelwerte, Standardabweichungen und Stichprobenumfang) nebst Informationen über die erfassten Variablen in eine Tabelle zu übertragen und dem Computer zu befehlen, mit diesen Informationen vorprogrammierte Analysen durchzuführen. Die Tiefe der Meta-Analyse wird meist auf so genannte

Haupteffekte sowie auf Abhängigkeiten erster Ordnung begrenzt. Wenn zwei Bedingungen zusammenwirken (was in der Realität eher die Regel als die Ausnahme ist) oder wenn hinter der analysierten Ursache weitere Ursache wirkt, fällt das einfach unter den Tisch. Diese Einengung der Analysen bleibt nicht ohne gravierende Auswirkungen auf die pädagogischen und bildungspolitischen Schlussfolgerungen. So wird nur berichtet, dass Hausaufgaben kaum einen Effekt für die Lernleistung haben, aber nicht, dass es bestimmten Formen von Hausaufgaben (z.B. Projekte) gibt, die das Lernen sehr deutlich fördern, andere dagegen negative Wirkung haben können. Es wird analysiert, welches Lehrerverhalten mit guten Testleistungen korreliert, aber nicht, wie dieses Verhalten in der Lehrerbildung trainiert werden muss. Hattie weist nur pauschal auf Studien hin, die immerhin einen deutlichen Unterschied zwischen gut und schlecht ausgebildeten Lehrern zeigen.

Hattie lässt aber nicht immer nur die Daten sprechen, wie es die orthodoxe Meta-Analyse fordert. Wenn die Daten seiner Vorstellung von gutem Unterricht widersprechen, verfällt selbst er ins Interpretieren. So zieht er Studien in Zweifel, die zeigen, dass Schüler mit Lernstörungen am besten lernen, wenn im Unterricht sensorische Übungen ("kinesthetic learning") angeboten werden. Dass ihm der Befund nicht behagt, zeigt sich schon darin, dass er ihn nicht wie sonst als  $d$ -Wert, sondern als Korrelation ( $r = 0.64$ ) darstellt (Hattie, 2009, S. 179), was den Effekt kleiner erscheinen lässt. Der Leser muss sich das  $d$  selbst ausrechnen. Es beträgt  $d = 1,70$ . Die Effektstärke ist also sehr hoch. Anders als sonst akzeptiert Hattie diesen Befund aber nicht, sondern nimmt andere Befunde dieser Studien (dass zum Beispiel zwischen Aufrechtsitzen und Testergebnissen positive Korrelationen bestehen) zum Anlass, sie lächerlich zu machen: "Die Behauptung eines Bedürfnisses nach Aufrechtsitzen widerstreben meiner Kraft, darin einen Sinn zu sehen." (S. 197). Der wirkliche Grund der guten Lernleistungen sei keine der als „effektiv“ bezeichneten Variablen, sondern nur das auf Lernfreude ausgerichtete Unterrichtsklima. Aber weder ist das ein Gegensatz, noch verträgt sich Hatties unbegründete Kritik dieser Studien mit der naiv-affirmativen Rezeption von Daten, die er sonst in seinem Buch pflegt. Er scheint nur dann intellektuelle Anstrengung (Zweifel an den Daten; Änderung der Darstellung) für notwendig zu halten, wenn die Daten der eigenen Theorie widersprechen.

Ebenso problematisch wie die Analyse von Daten mittels oberflächlicher, mono-kausaler Erklärungsmodelle, ist die Einengung der Kriterien für "Leistung" auf standardisierte Testwerte in Mathematik und der Mutter-

sprache, und manchmal noch in den Naturwissenschaften. Andere Kriterien (wie soziale Kompetenz und soziale Ungleichheit werden gelegentlich erwähnt, aber nicht systematisch einbezogen). Diese Verkürzung wird von Hattie weder thematisiert noch begründet, obwohl die Auswahl der Kriterien für die Messung von "Effektgrößen" von zentraler Bedeutung ist. In seiner Synthese fehlen viele Kriterien, die für Pädagogik und Bildungspolitik überaus wichtig sind: Wie ist die Auswirkung bestimmter Maßnahmen auf Geographie- und Geschichtswissen? Was bedeuten sie für die Entwicklung moralisch-demokratischer Fähigkeiten? Was fördert Zusammenarbeit und die Fähigkeit Konflikte zu lösen? Vor allem aber: Wie nachhaltig ist der Lernerfolg? Sind die Lerneffekte auch Wochen, Monate und Jahre nach den Testterminen noch nachweisbar? Alles Fehlanzeige. Wie stark diese Kurzsichtigkeit die Schlussfolgerungen verfälscht, zeigt sich an vielen Beispielen, so auch beim Effekt der Klassengröße. Nach Hatties Auszählung liegt die Effektstärke dafür bei einem mageren  $d = 0.21$  (S. 86), was, so Hattie, die Kosten für Strukturmaßnahmen (hier Verkleinerung von Klassen) nicht rechtfertigt. Auch hier sind Hatties Synthese kurzsichtig und seine Schlussfolgerungen problematisch.

Ein intellektueller Beitrag hätte auch sein können, die hoch komplexe Beziehung zwischen statistischen Korrelationen und Kausalität darzulegen. Korrelationen legen manchmal eine plausible Erklärungskette nahe. Wie Hattie berichtet, zeigen Studien, dass die Ermöglichung von Rückmeldung der Schüler an den Lehrer deutlich mit deren Testleistungen korreliert ( $d = 0.72$ ; das entspricht einem  $r = 0.34$ ). Solche Befunde sind aber kein Beleg dafür, dass Schulstrukturen unwichtig sind. Die Ermöglichung von Schülerfeedback, so zeigen von ihm zitierte Studien, ist an die Qualität der Lehrerbildung gebunden. Also, sollte die Lehrerbildung verbessert werden (Strukturmaßnahme), damit Schüler bessere Testleistungen erzielen. Zum anderen sind statistische Daten immer interpretationsbedürftig, wie Hattie selbst in dem Fall oben aufgezeigt hat. Da der Faktor Schülerrückmeldung vermutlich mit vielen anderen Bedingungen für Lernerfolg korreliert (ein Problem, das als „Kolinearität“ bezeichnet wird), können wir nicht wissen, ob dieser oder ob andere Faktoren für den Zuwachs von Testleistungen kausal verantwortlich sind. Dasselbe Problem gilt auch für negative Korrelationen, wie im Fall von Noten und extrinsischen Belohnungen für gute Leistung (Hattie, S. 175 ff.). Ihre negative Wirkung kann mit anderen Bedingungen wie mit mangelnder Lehrkompetenz der Lehrperson korreliert sein. Schlecht ausgebildete Lehrer wissen sich oft nicht anders zu helfen, als die Disziplin in der Klasse mit Noten und anderen Belohnungen bzw. Bestrafungen aufrecht zu erhalten.

Noch mehr Fragen werfen niedrige Korrelationen auf. Häufig sind Korrelationen deshalb niedrig, weil die Bedingungen, deren Wirkung untersucht werden soll, in einer Studie wenig oder nicht substantiell variiert wurden oder kaum vorkommen. In vielen Studien, die die Auswirkung von offenem Unterricht erforschen sollten, scheint man die Mühe zu scheuen, Klassen mit wirklich offenem Unterricht zu finden oder ein Interventionsexperiment zu machen, in dem offener Unterricht realisiert wird. Ersatzweise analysiert man daher oft Klassen, in denen es Freiarbeit oder Stationen-Lernen gibt. Aber das ist nicht dasselbe. Wo kein wirklich offener Unterricht gemacht wird, können sich auch keine statistischen Effekte zeigen. Daten sind fast immer mehrdeutig und müssen daher immer interpretiert werden, und zwar auf der Grundlage gesicherter Forschungsergebnisse, und nicht von subjektiven Spekulationen, wie Hattie das tut. Seine ad hoc-Interpretationen taugen nicht, um Lehren und Lernen zu verbessern, wie im Klappentext behauptet wird. Hattie ist sich dieser Probleme bewusst. In der Einleitung warnt er den Leser sogar vor einer einfachen Gleichsetzung Korrelation = Kausalität (Effektivität). In den übrigen Kapiteln setzt er sich selbst aber über diese Warnung methodisch und intellektuell hinweg. Er steigert diese naive Gleichsetzung sogar noch, indem er Korrelationsstatistiken graphisch in "Effektivitäts-Barometer" übersetzt, aus denen Lehrer und Bildungspolitiker direkt Handlungsanleitungen ablesen können sollen.

Schließlich können sich Pädagogik und Bildungspolitik nicht nur auf Effektstärken stützen, sondern müssen immer auch Aufwand und Ertrag einer Maßnahme abschätzen. Hattie spricht dies an, wenn es im opportun erscheint, wie im Beispiel der Klassengröße, bleibt dabei aber auf Stammtischniveau. Auch hier zeigt sich wieder, dass das Sammeln von Statistiken nicht ausreicht, sondern eine gründliche Aufarbeitung der Fragestellung notwendig ist, wie sie in vielen Fällen auch schon vorliegt. Hattie ignoriert sie nur einfach. So waren die Auswirkung kleiner Klassen Gegenstand einer beispiellosen, aufwändigen Interventionsstudie in Tennessee (STAR-Projekt) und mehrerer Follow-up-Studien zu den langfristigen Effekten kleiner Klassen. Demnach hatte die Reduktion der Klassengröße in der Grundschule auf 13 bis 17 Schüler langfristig nachweisbare Effekte: Schüler aus den kleineren Klassen schaffen das Abitur deutlich häufiger als Schüler in normal großen Klassen (76 versus 64 Prozent), bekamen mehr Auszeichnungen (45 versus 29 Prozent), brachen weniger die Schule ab (15 versus 24 Prozent), und gingen eher aufs College (Finn et al., 2005). Addiert man die Gewinne aus diesen Auswirkungen, dann können kleinere Grundschulklassen auch ökonomisch Sinn machen. Für

den Zugewinn an Lebensqualität für die dadurch geförderten Schülerinnen und Schüler sind sie ohnehin von unschätzbarem Wert.

Der berühmte Psychologe Paul Meehl hat einmal einen guten Wissenschaftler definiert – und zwar als jemanden, der seinen Kopf statt einer Formel gebraucht. Die Meta-Analyse sollte, so schreibt einst ihr Erfinder, ein "intellektuelles Gegenstück von originärer empirischer Forschung" sein. Bei Hattie verkommt sie jedoch zur mechanischen Umsetzung von computererzeugter Statistik in Politik unter Ausschaltung des (Nach-) Denkens. Schlimmer noch: Meta-Analysen suggerieren dem Leser, dass sie die Lektüre von originärer Forschung überflüssig machen.

Meehls Definition, so scheint es, trifft indes auch auf "gute Politiker" zu. Ein guter Bildungspolitiker kommt leidernicht umhin, sich mit widerstreitenden Informationen auseinander zu setzen, um herauszufinden, was wirkliches Wissen ist und was nicht. Tatsächlich wissen wir heute aufgrund der Forschung sehr viel über die Bedingungen und Effekte guten Unterrichts und guter Schule (Bohl & Kiper, 2009). Aber dieses Wissen ist manchmal unter einem Wust von falschen Informationen und naiven Interpretationen der Forschung verborgen. Man kann aus Testdaten und auch aus Meta-Analysen, die darauf aufbauen, keine Politik "ablesen". Man kann das richtige Wissen nur finden, wenn man sich auch gegensätzliche Meinungen anhört, auf gründlichen Analysen besteht und demokratische Maximen nicht außer Acht lässt.

Sonst trifft man garantiert die falschen Entscheidungen, ob es sich nun um Strukturreform oder Unterrichtsreform handelt. Die Reduktion der Schuljahre von neun auf acht, war, wie man jetzt sieht, eine falsche Strukturreform. Die Einführung der Gesamtschule hätte eine gute werden können, wenn man sie konsequent durchgeführt hätte. Die Gesamtschule als eine unter vielen Schulformen einzuführen, war eine offenkundige Narretei. Eine "Gesamt"-Schule sollte auch allen Schülern offenstehen, nicht nur einer Minderheit. Es wäre gut, wenn die Politik mit der Gemeinschaftsschule diesen Fehler nicht wiederholt.

Die Evaluation der Unterrichtsqualität mit standardisierten Tests (PISA & Co.) war eine falsche Unterrichtsreform. Noch wehren sich allerorts Lehrer und Schüler tapfer gegen die Reduktion des Bildungsauftrags der Schule auf das Ankreuzen von Wiedererkennungsaufgaben. Aber schon wird ein Absinken der Unterrichtsqualität sichtbar, jedenfalls für die, die sich keine PISA- oder Hattie-Brille auf setzen lassen. Wie der Druck, den Unterricht auf Testleistungen zu reduzieren, auf die Qualität des Bildungssystems insgesamt beeinträchtigen kann, zeigt das Beispiel der USA (Lind, 2009).

## Quellen

\* Der obige Text beruht – leicht geändert – auf dieser Publikation: Lind, G. (2011). Pädagogik oder Struktur? Auch die Hattie-Studie weiß keinen Rat. Hessische Lehrerzeitung 11/211. 26-27.

Bohl, T. & Kiper., H., Hg. (2009). Lernen aus Evaluationsergebnissen – Verbesserungen planen und implementieren. Bad Heilbrunn: Julius Klinkhardt.

Finn, J. D., Gerber, S. B., & Boyd-Zaharissas, J. (2005). Small classes in the early grades, academic achievement, and graduating from high school. *Journal of Educational Psychology*, 97, 2, 214–223.

Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher*, 5, 3-8.

Hattie, J., (2009). *Visible learning. A synthesis of over 800 meta-analyses relating to achievement*. London: Routledge.

Lind, G. (2004). Jenseits von PISA — Für eine neue Evaluationskultur. In: Institut für Schulentwicklung PH Schwäbisch Gmünd, Hg., *Standards, Evaluation und neue*

*Methoden. Reaktionen auf die PISA-Studie*, S. 1 - 7. Baltmannsweiler: Schneider Verlag Hohengehren

Lind, G. (2009). Amerika als Vorbild? Erwünschte und unerwünschte Folgen aus Evaluationen. In: T. Bohl, T. & H. Kiper., Hg., *Lernen aus Evaluationsergebnissen – Verbesserungen planen und implementieren*, S. 78-97. Bad Heilbrunn: Julius Klinkhardt.

Meehl, P. E. (1958). When to use your head instead of the formula? In: H. Feigl, M. Scriven & G. Maxwell, Hg., *Minnesota studies in the philosophy of science*, S. 498-506. Minneapolis, MI: University of Minnesota Press.

Spiewak, M. (2013). Hattie-Studie. Ich bin superwichtig! Kleine Klassen bringen nichts, offener Unterricht auch nicht. Entscheidend ist: Der Lehrer, die Lehrerin. *Die Zeit*, 3.1.2013.

Weiterführende Literatur:

<http://www.uni-konstanz.de/ag-moral/evaluation/it-references.htm>

Zitation:

Lind, G. (2013). *Metaanalysen als Wegweiser? Zur Rezeption der Studie von Hattie in der Politik*. Download: [http://www.uni-konstanz.de/ag-moral/pdf/Lind-2013\\_meta-analysen-als-wegweiser.pdf](http://www.uni-konstanz.de/ag-moral/pdf/Lind-2013_meta-analysen-als-wegweiser.pdf)

Kontakt:

Georg.Lind@uni-konstanz.de