

■ Vor kurzem ist ein Buch erschienen, das die Times Educational Supplement als »... heiligen Gral der Didaktik« gewürdigt hat, weil es in einem Band zusammenfasse, was man heute aus empirischen Studien über Bedingungen für einen erfolgreichen Unterricht wisse: »Visible Learning« von John Hattie (2009, dt. 2013): Daten zu mehr als 130 Merkmalen aus über 50.000 Basisstudien. Ist damit wirklich endgültig geklärt, was »guten Unterricht« ausmacht?

Hatties Zusammenschau wird auch in Deutschland immer häufiger zitiert, um bestimmte »Methoden« oder »Ansätze« als mehr oder weniger sinnvoll zu bewerten². Denn Hattie ist noch einen Schritt weitergegangen als die Kolleg/inn/en vor ihm: Er hat die Werte aus über 800 Metaanalysen zum Schulerfolg unter verschiedenen Bedingungen in einer Meta-Meta-Analyse auf einer noch höheren Ebene verdichtet und findet in der Tat einige beeindruckende Zusammenhänge. Hier die Plätze 1-10 der angeblich einflussreichsten Faktoren³ (s. ergänzend unten die Auswahl-Übersicht auf Deutsch):

- .7 klare Anweisungen des Lehrers (»teacher clarity«)
- .7 Schüler als Lehrer (»reciprocal teaching«)
- .6 Fortbildung von Lehrer/inne/n
- .5 Schüler als Tutor_inn_en (»peer tutoring«)
- .0 Schülerkontrolle über Lernen
- .0 offen vs. traditionell
- .0 Altersmischung
- .0 Fachwissen Lehrer
- .0 Förderung von Wahrnehmung und Motorik

Wie kompliziert das Bild tatsächlich ist, zeigen beispielhaft die Werte zum Faktor »Computereinsatz« im

Metaanalysen der Unterrichtsforschung

Hat Hatties »Visible Learning« einen Nutzen für die pädagogische Arbeit vor Ort?¹

von Hans Brügelmann



Prof. Hans Brügelmann,
Grundschulpädagogik,
Universität Siegen

Die empirische Bildungsforschung – vor allem in den angelsächsischen Ländern – hat in den vergangenen 20, 30 Jahren eine Unzahl von Befunden aus Einzelprojekten zu verzeichnen, die kaum jemand mehr überblicken kann. Angesichts dieser zunehmenden Fülle von Detailstudien hat die Bedeutung sog. Metaanalysen zugenommen. Um zu sehen, ob es eine Beziehung zwischen Faktoren wie z.B. Fernsehkonsum von Kindern und ihrer Lesefähigkeit gibt und wie stark sie ausgeprägt ist, verrechnet man statistische Effekte über verschiedene Studien hinweg. Man tut dies, obwohl die Daten unter verschiedenen Bedingungen, in verschiedenen Kontexten (Bildungssystemen, Sprachen, Kulturen...) und mit unterschiedlichen Instrumenten gewonnen worden sind. Auf diese Weise hofft man robustere Werte zu gewinnen und den Einfluss von Ausreißer-Befunden aus Einzelstudien zu minimieren. Auf den ersten Blick ein überzeugender Ansatz – und auf alle Fälle eine Hilfe, um einen groben Überblick zu gewinnen.

Die formal aufgelisteten Merkmale zeigen allerdings Spannungen auf, die den Sinn von Ranglisten, wie Hattie sie publiziert, in Frage stellen⁴. Zum Beispiel wirke positiv, wenn Lehrer/innen klare Anweisungen bei der Aufgabenstellung geben – aber auch wenn Schüler/innen anderen Schüler/innen etwas beibringen oder wenn sie sich gar selbst beurteilen dürfen. Auch die folgende Auswahl über das ganze Spektrum der Effektstärken zeigt, dass es darauf ankommt, was man jeweils unter den einzelnen Etiketten konkret versteht – sowohl in den verschiedenen Studien, die als »gleich« miteinander verrechnet werden, als auch bei der Übersetzung in andere Sprachen (s. die Originalbegriffe in Klammern) wie auch in jeder Auslegung der stark verknappten »Etiketten« durch einzelne Leser/innen:

- 1.4 Selbstbeurteilung
- 1.3 Piaget-orientierte Programme
- .9 förderorientierte Beurteilungen
- .9 Einüben von methodischen Fertigkeiten (»micro teaching«)

Rank	Influence	Studies	Effects	ES
1	Self-reported grades	209	305	1.44
2	Piagetian programs	51	65	1.28
3	Providing formative evaluation	30	78	.90
4	Micro teaching	402	439	.88
5	Acceleration	37	24	.88
6	Classroom behavioral Comprehensive interventions for learning disabled students	160	942	.80
7	Teacher clarity	na	na	.75
8	Reciprocal teaching	38	53	.74
9	Feedback	1287	2050	.73

Unterricht. Mit einer durchschnittlichen Effektstärke von .39 ist sein Einfluss als mäßig positiv einzuschätzen. Wenn die Lernschritte durch das Programm vorgegeben werden, verschwindet dieser sogar ganz (-.02). Andererseits steigt der Effekt, wenn die Schüler das Vorgehen im Programm selbst steuern konnten, auf .49, und wenn sie mit einem Partner arbeiteten, sogar auf .96. Das heißt aber, dass nicht die Mittelwerte das eigentlich interessante Datum sind, sondern eher die Extremwerte und damit die Streuung. Diese verweisen darauf, dass es auf die Art des Computereinsatzes und auf die Bedingungen seiner Nutzung ankommt, ob er didaktisch produktiv wird oder nicht. Was aber nützt dann der Lehrperson das Wissen um einen Mittelwert von 0.39, wenn sie eine Form des Computereinsatzes wählt, die einen Wert von 0.0 erzielt? Während Hattie Abweichungen vom Mittelwert sozusagen als Messfehler betrachtet, verweisen sie aus meiner Sicht auf den Einfluss bedeutsamer Drittfaktoren.

Diese Konzept- und Kontextabhängigkeit von Wirkungen einer »Methode« zeigt sich auch bei anderen Interventionen. So hat nach Hattie ein auf Problem-Aufgaben fokussierter Unterricht mit einer Effektstärke von .15 insgesamt nur einen geringen positiven Effekt – mit .57 aber eine hohe Wirksamkeit, wenn vorher Beispielaufgaben durchgearbeitet worden sind. Dabei kommt es zusätzlich auf die angezielte Kompetenzebene an: für die Sicherung von Prinzipien steigt der Effekt bis zu auf .75, für den Erwerb von Basiswissen ist die Beziehung dagegen sogar negativ, d. h. ein traditioneller Unterricht ist in dieser Hinsicht erfolgreicher.

Entsprechend relativieren Hattie selbst und auch seriöse Interpreten wie Terhart (2011), Helmke (2012) oder Meyer (2013) einzelne Aussagen bzw. ihren Geltungsbereich. Trotzdem operieren sie alle mit den Ranglisten, als spiegelten sie aussagekräftige Merkmale von Unterricht. Und die machen Furore. So wird vielerorts die Rückkehr zu traditionellen Lehrmethoden gefordert, denn »offener Unterricht« sei nicht wirksamer ($d = .01$). Aber was bei Hattie bzw. in den von ihm zusammengefassten Meta-Studien konkret als »offen« erfasst worden ist, kann nicht mehr nachvollzogen werden. Aus anderen Analysen wissen wir, dass sich in diesem Topf ganz unterschiedliche Ansätze⁵ befinden. Und so überrascht es nicht, dass – entgegen den verbreiteten Plädoyers für »mehr Frontalunterricht« (vgl. etwas Spiewak 2013) – bei Hattie als »sehr wirksame« Faktoren u. a. auftauchen:

- .90 »formative assessment« (inhaltsbezogene Rückmeldungen)
- .82 Diskussionen zwischen Schüler/inn/en im Unterricht
- .74 reziprokes Lernen (wechselseitige Unterstützung der Schüler/inn/en untereinander)
- .69 »meta-cognitive strategies« (über das eigene Lernen nachdenken)
- .54 kooperatives statt konkurrenzorientiertes Lernen
- .47 Überzeugung der Schüler/inn/en von ihrer Selbstwirksamkeit
- allesamt Merkmale eines Unterrichts, der Schüler/inn/en Raum für eigene Lernwege gibt und ihnen Verantwortung für das eigene Lernen

zugesteht/abverlangt. Das Bild ist also viel komplizierter, als herausgegriffene Einzelmerkmale vermuten lassen – gerade im Hinblick auf die Diskussion über lehrerzentrierten vs. offenen Unterricht.

Schon die SCHOLASTIK-Studie von Weinert/Helmke (1997) operierte mit Merkmalen wie »Strukturiertheit«, die – ohne Kenntnis der Originalskalen – ganz Unterschiedliches bedeuten können und entsprechend unterschiedlich gedeutet werden: inhaltliche Systematik, methodische Kleinschrittigkeit, soziale Regelungen. Zudem können »Strukturen« von der Lehrperson vorgegeben oder mit den Schüler/inn/en ausgehandelt werden: Regeln für Arbeits- und Verkehrsformen, gemeinsame Projektplanung, individuelle Lernverträge ...⁶ In der didaktischen Diskussion werden Vorteile von »Strukturiertheit« aber ohne Umstände als Argumente für Lehrerzentrierung verbucht, ohne dass – erst recht in Metastudien wie der von Hattie – klar ist, wie der Gegenpol in den einzelnen Basis-Studien aussah.

Für die Einschätzung des Erfolgs einer Methode kommt es also erstens darauf an, an welchem Kriterium man diesen misst. Dann ist zu prüfen, in welcher Form unter welchen Bedingungen sie umgesetzt wird. Die für einen Überblick über die Forschung und für bildungspolitische Entscheidungen verdienstvolle Verdichtung der Daten in Hatties Meta-Meta-Analyse bedeutet also, dass diese genau durch die Ablösung von den konkreten Kontexten für Praxis weitgehend ihre Bedeutung verlieren. Denn die Zahlen sind mehrdeutig⁷. Was bedeutet zum Beispiel eine mittlere Korrelation oder Effektstärke im konkreten Fall⁸: leicht positiver Einfluss in allen Fällen ODER starker in einigen und gar keiner in anderen ODER hoher im Zusammenwirken mit Faktor X, aber niedriger im Zusammenwirken mit Variable Y?

Das heißt aber: Selbst wenn 99 von 100 Studien eine Überlegenheit der Methode B zeigen, dürfte das kein Grund sein, einer Lehrerin, die mit Methode A erfolgreich arbeitet, deren

Nutzung zu untersagen (vgl. wieder das Beispiel der Computernutzung: einige »Ausreißer« unter den Metaanalysen weisen besonders positive Effekte auf). Es wäre auch kein Grund, weiteren Lehrer/inne/n, die Methode A ausprobieren wollen, dies zu verbieten. Allenfalls könnte man eine besondere Begründung verlangen, warum jemand von einer weithin bewährten Praxis abweichen will. Juristisch gesprochen, erzeugen statistische Befunde aus großen Stichproben dann eine »Beweislast«, d. h. eine Begründungspflicht für Abweichungen – sie können diese aber nicht abwerten oder gar ausschließen. Schon Weinert/Helmke (1997, 472) fanden unter ihren erfolgreichsten Lehrer/innen »bizarre Merkmalsprofile« in den untersuchten Dimensionen, also auf dieser abstrakten Ebene gerade nicht eine Idealform.

Kontexte verändern Beziehungen – das gilt selbst in den Naturwissenschaften, wenn wir an die Gesetze der Mechanik im Mikrobereich der Atome denken. In den Sozialwissenschaften wiederum lassen sich Befunde aus kontrollierten Experimenten nicht einfach auf Alltagssituationen übertragen, Befunde aus einem (schul)kulturellen Kontext nicht unbesehen auf einen anderen.

Insofern gerät die vertraute Grundfigur der Forschung und ihres Verhältnisses zur Praxis ins Wanken: zunächst Ableitung allgemeiner Regeln aus der Häufigkeit von Beziehungen bei einer großen Zahl von Fällen und dann deduktive Ableitung einer Handlungsanweisung für den Einzelfall aus diesen Regeln. Für Lehrer/innen bedeutet das: Jede Situation ist ein neuer Fall und bedarf einer eigenen Einschätzung. Die Durchschnittsbefunde aus (Meta-)Metaanalysen können dafür hilfreiche Hypothesen liefern – aber keine Vorschriften⁹.

(Fußnoten auf der übernächsten Seite 15. Das Literaturverzeichnis kann auf der Homepage der BLZ nachgelesen werden.)

■ Bereits zum siebten Mal jährt sich der Bremer Friedenslauf in der letzten Schulwoche vor den Sommerferien. Damit ist die Veranstaltung mitten auf dem Marktplatz bereits eine feste Institution an vielen Bremer Schulen geworden. Das Prinzip des Friedenslaufes setzt auf die Eigeninitiative der Bremer Schülerinnen und Schüler. Vor dem Lauf suchen sich die Kinder Sponsoren und verhandeln selbständig den Betrag, den sie für jede gelaufene Runde bekommen. Im letzten Jahr liefen so 1300 Schülerinnen und Schüler von 10 Bremer Schulen bei bester Laune für den Frieden.

Die erlaufenen Spenden kommen der Friedensarbeit des Zivilen Friedensdienstes weltweit und in Deutschland zu Gute. In diesem Jahr gehen die Spenden an das Bremer Projekt »Alle Kinder wollen schwimmen«, das Schwimmunterricht für Kinder aus einkommensarmen Familien ermöglicht. Außerdem werden die Friedensprojekte des Forums Ziviler Friedensdienst im Libanon gefördert, die Jugendliche bei der Aufarbeitung des



Bürgerkriegs und bei der Versöhnungsarbeit unterstützen.

Die Schulen, die am Friedenslauf teilnehmen, haben auch die Möglichkeit ein kostenfreies Bildungsprogramm des Forums Ziviler Friedensdienst (forum ZFD), zu nutzen. Hier geht es konkret um Themen wie zivile Konfliktbearbeitung, Gewaltprävention und interkultureller Dialog. Die Schülerinnen und Schüler können die konkrete Arbeit der Friedens- und Konfliktberater aus dem Libanon oder Palästina kennenlernen oder sich im Rahmen von Projekttagen mit den Themen (un)fairer Handel, Migration und Friedensarbeit in Zusammenhang mit Sport auseinandersetzen.

Auch in diesem Jahr hat Willi Lemke, Sonderberater des Generalsekretärs der Vereinten Nationen für Sport im

Dienst von Entwicklung und Frieden die Schirmherrschaft für den Lauf übernommen. Ermöglicht wird die Veranstaltung durch die Träger Forum Ziviler Friedensdienst e.V., der Landessportbundes Bremen e.V., das LicideHaus und den Bremer Jugendring.

7. Bremer Friedenslauf am 25. Juni 2013

Schüler engagieren sich

Anmeldungen für den 7. Bremer Friedenslauf sind bis zum 11. Juni 2013 möglich.

Weitere Infos zum Friedenslauf und zum Bildungsprogramm finden Sie unter www.bremen.run4peace.eu
www.gew-hb.de/Binaries/Binary12584/Aktionsheft_Bremen.pdf.
Kontakt: Alexandra Weidmann, weidmann@forumzfd.de oder bremen@run4peace.eu

Bis zum 4. Juni wählen!

Bremischer Gewerkschaftstag der GEW:

Noch bis Dienstag, den 4. Juni 2013, können Mitglieder über die Zusammensetzung des Bremischen Gewerkschaftstages entscheiden. Gewählt wird in den Stadtverbänden Bremen und Bremerhaven jeweils getrennt in den Wahlbereichen Schule, Hochschule und Forschung, Sozialarbeit und Jugend-

hilfe, Erwachsenen- und Weiterbildung und SeniorInnen. Alle Mitglieder in Bremen sollten per Post Briefwahlunterlagen, Mitglieder in Bremerhaven sollten Wahlausweise erhalten haben. Insgesamt können bis zu 86 Vertreterinnen und Vertreter für das höchste Gremium des GEW-Landesverbandes bestimmt werden. Sie entscheiden über die grundlegenden Fragen der Politik unserer Gewerkschaft und wählen u.a. die

Mitglieder des Landesvorstandes. Erfolge bei der Durchsetzung unserer gewerkschaftlichen Ziele können wir nur erreichen, wenn die Mitglieder die Politik bestimmen. Um den zu wählenden Delegierten dabei den Rücken zu stärken, ist eine hohe Wahlbeteiligung notwendig. Die Stimmabgabe muss bis spätestens 4. Juni 2013 erfolgt sein – Wahlbriefe müssen bis dahin beim Wahlausschuss eingegangen sein.

Metaanalysen der Unterrichtsforschung

- 1 Nachdruck der für »Lehren und Lernen« überarb. und erw. Fassung meines Beitrags Die Hattie-Studie: Der heilige Gral der Didaktik? In: Grundschule aktuell, H. 121 (2013), 25-26. Wir danken dem Grundschulverband und »Lehren und Lernen« für die freundliche Genehmigung zum Abdruck.
- 2 Vgl. etwa Köller (2012), Spiwak (2013); differenzierte Auseinandersetzungen mit Hatties inhaltlichen Aussagen finden sich bei: Terhart (2011), Meyer (2013).
- 3 ES = Effektstärke; ab 0,5 gilt ein Effekt gemeinhin als beachtenswert – der Anschaulichkeit halber oft gleichgesetzt mit dem Vorsprung von etwa einem Schuljahr.

- 4 Vgl. zu weiteren Kritikpunkten, z. B. der Beschränkung der Erfolgsmaße für »erfolgreichen Unterricht« auf Ausschnitte aus (nur) den Fächern Mathematik und Sprache, aber auch zu forschungsmethodischen Bedenken Lind (2011), Brügelmann (2013), Rolff (2013).
- 5 Vgl. etwa meine Übersicht (1998, insbesondere 14-19).
- 6 Vgl. zu diesen und weiteren Einschränkungen der SCHOLASTIK-Befunde meine Kritik (2005, 43-44).
- 7 Rolff (2013) vergleicht die Hattie-Studie und ihre Rezeption sogar mit einem »Rorschach-Test«.
- 8 Ganz abgesehen von dem häufig übersehenen Problem, dass ein gleichzeiti-

- ges Auftreten von zwei Faktoren nicht gleichbedeutend ist mit einer inhaltlichen Beziehung: Korrelation bedeutet nicht Kausalität. Ein hoher TV-Konsum geht häufig einher mit schwacher Leseleistung – aber ist sie die Ursache? Oder kann nicht auch eine schwache Lesekompetenz verhindern, dass Kinder zum Buch greifen (und damit ihre Lesefähigkeit verbessern) und stattdessen lieber fern sehen? Oder ist gar ein Drittfaktor (z. B. Vernachlässigung durch die Eltern) möglicherweise Ursache für beide Verhaltensweisen?
- 9 Vgl. ausführlicher zum Status statistischer Befunde und zu den Anforderungen an eine für die Praxis nützliche Forschung: Brügelmann (2011), Rolff (2013).