

SKBF Staff Paper #27

«Bitte wenden» – John Hatties Navigationsfehler im Forschungsdschungel

Chantal Oggenfuss

Unter Mitarbeit von Viviane Zbinden



SKBF | CSRE

Schweizerische Koordinationsstelle
für Bildungsforschung

Centre suisse de coordination pour
la recherche en éducation

Centro svizzero di coordinamento
della ricerca educativa

Swiss Coordination Centre for
Research in Education

«Bitte wenden» – John Hatties Navigationsfehler im Forschungsdschungel

Chantal Oggenfuss, unter Mitarbeit von Viviane Zbinden

Einleitung

Wer mit dem Auto unterwegs ist, kennt die irritierende Situation: Mitten auf der Autobahn meldet sich das Navigationssystem mit der Ansage «Bitte wenden». Natürlich weiss man, dass das nicht möglich ist, und doch haftet dieser Stimme etwas Verlockendes an. Sie klingt präzise, sie nimmt uns das Denken ab, sie suggeriert Sicherheit. Verlockend wirkt auch John Hatties neues Werk *Visible Learning: The Sequel* (2023). Auf 400 Seiten, gespeist aus über 2100 Meta-Analysen, präsentiert er Lehrkräften, Schulleitungen und Fachpersonen der Bildungspraxis eine Art Landkarte des Lernens. Das Versprechen: Orientierung inmitten des Forschungsdschungels. Die Frage ist jedoch, ob diese Orientierung tatsächlich verlässlich ist oder ob sie uns, wie das Navigationsgerät mitten auf der Autobahn, in die Irre führt.

Kritik an der ersten Ausgabe aus dem Jahr 2008

Bereits die Erstveröffentlichung von *Visible Learning* (Hattie, 2008) wurde schnell zu einem internationalen Erfolg. Die Vermarktung wurde von umfangreichen Weiterbildungsangeboten begleitet und stiess in der Bildungslandschaft auf grosses Interesse. Die von Hattie präsentierte Übersicht über mehrere hundert Einflussfaktoren auf den Lernerfolg wirkte wie ein Befreiungsschlag. Lehrkräfte sowie Bildungsforscherinnen und -forscher griffen dankbar zu. Doch die Publikation blieb nicht ohne Kritik. Unter anderem stellte Terhart (2011) die Frage, ob Hattie mit der «Mega-Analyse» tatsächlich den «Heiligen Gral» der Unterrichtsforschung entdeckt habe. Viele Befunde würden bereits Bekanntes bestätigen, andere würden durch neuere Studien relativiert oder widerlegt. In diesem Spannungsfeld stellte Terhart die Frage nach dem Nutzen von Hatties Werk. Bemerkenswert ist in diesem Zusammenhang, dass es für einen differenzierten Überblick zur Forschung längst Handbücher gibt. Diese fanden jedoch weder millionenfachen Absatz noch Eingang in die Bildungspraxis. Pant (2014) kritisierte, dass Hatties Meta-Meta-Analyse äusserst heterogene Meta-Analysen zu sehr unterschiedlichen Konstrukten und Interventionen zusammenführe, wodurch relevante Informationen verloren gehen. Werden zudem methodisch schwache oder unzureichend kontrollierte Studien integriert, könne dies die Aussagekraft der aggregierten Ergebnisse zusätzlich erheblich einschränken (garbage in, garbage out). Darüber hinaus birgt die Präsentation der Wirkungsfaktoren die Gefahr kausaler Fehlinterpretationen. Beruhen die ausgewiesenen Effekte lediglich auf bivariaten Zusammenhängen, bilden sie nur Korrelationen ab. Komplexe Moderatoren wie Kontext, Forschungsdesign oder Forschungsmethode bleiben häufig unberücksichtigt, obwohl diese bereits bei der Auswahl der Studien systematisch hätten berücksichtigt werden müssen. Oggenfuss und Wolter (2015) warnten vor einem unkritischen Forschungskonsum, zu dem die vereinfachte, visuelle Darstellung der Ergebnisse verleite, und wiesen auch auf die begrenzte Übertragbarkeit von Hatties Befunden hin. Was in einem Klassenzimmer in Sydney, Christchurch oder Nelson erfolgreich sein kann, muss nicht zwangsläufig auch in Zürich, Chur oder Neuenburg Gültigkeit haben.

Hatties Reaktion und warum Quantität ohne Qualität die Aussagekraft verliert

In *Visible Learning: The Sequel* (2023) greift Hattie zentrale Kritikpunkte an seiner ersten Ausgabe auf und diskutiert diese ausführlich. Er verteidigt seinen Ansatz, Meta-Analysen mit sehr unterschiedlichen methodischen Grundlagen einzubeziehen. Eine strikte Beschränkung auf methodisch «perfekte» Studien nach dem sogenannten *gold standard* hält er für nicht praktikabel, weil dadurch eine ungenügende Anzahl an Studien übrig bliebe. Zur Veranschaulichung verweist er auf das *What Works Clearinghouse*, das ausschliesslich streng kontrollierte Designs wie randomisierte Kontrollstudien zulässt. Dort liege die durchschnittliche Anzahl der Studien pro Review bei lediglich zwei. Auf dieser Basis liessen sich kaum belastbare Schlussfolgerungen ziehen («*It is impossible to conclude much based on two studies other than more research is needed*», Hattie, 2023, S. 31). Aus Hatties Sicht reicht methodische Strenge deshalb nicht aus, um die Qualität der Evidenz zu gewährleisten. Stattdessen sei ein breites Spektrum an Studien sowie eine vorsichtige Interpretation der Befunde erforderlich.

Der zentrale Schwachpunkt der Argumentation bleibt damit auch in der überarbeiteten Ausgabe unverändert: eine grössere Anzahl an Studien erhöht die Aussagekraft der Ergebnisse nicht automatisch – insbesondere dann nicht, wenn methodisch weniger belastbare Studien berücksichtigt werden. Besonders problematisch ist dabei der Einschluss nicht-kausaler Studien. Bei Arbeiten der sogenannten *Clearinghouses* genügen von anfänglich rund vierzig Studien oft nur sehr wenige den strengen und transparenten Auswahlkriterien. Hattie hingegen stützt seine Effekte zusätzlich auf die übrigen 95%, die von den *Clearinghouses* als methodisch unzureichend eingestuft werden, und präsentiert das Ergebnis aufgrund der Quantität als die robustere Evidenz. Das bedeutet, dass in manchen Fällen der Durchschnitt aller Studien einen positiven oder gar keinen Effekt zeigt, während der durchschnittliche Effekt der methodisch belastbaren und kausalen Studien negativ ausfällt und umgekehrt. Die Folge ist, dass Befunde zu völlig entgegengesetzten Aussagen führen können.

Hattie schliesst zwar inzwischen einzelne Meta-Analysen aus, etwa dann, wenn Effektstärken ungewöhnlich hoch liegen ($d > 2.0$) oder methodische Mindestanforderungen nicht erfüllt werden. Zudem führt er den sogenannten *Robustness Index* ein, der die Stabilität der Gesamtergebnisse beschreiben soll. Auf einer Skala von 1 bis 5 werden

dabei fünf Indikatoren berücksichtigt, unter anderem die Zahl der Meta-Analysen und der Primärstudien, die Grösse der Stichproben sowie die *fail-safe number* (Anzahl zusätzlicher Null-Studien bis zur Aufhebung des Gesamtergebnisses).

Hattie verweist auch darauf, dass sich in seinen Analysen keine systematischen Zusammenhänge zwischen Studiendesign, Stichprobengrösse oder Anzahl der Studien und der Höhe der Effektstärken finden. Die vorgenommenen Anpassungen ändern jedoch nichts daran, dass Fehlschlüsse nur vermieden werden können, wenn nicht-kausale Studien konsequent ausgeschlossen oder zumindest eindeutig getrennt ausgewiesen werden. Vor diesem Hintergrund wirkt es unglaublich, wenn Hattie den Vorwurf zurückweist, seine Darstellung der Befunde suggeriere kausale Wirkungen. Er betont, dass die von ihm genutzten Meta-Analysen grösstenteils keine kausalen Aussagen zulassen. Wenn seine Ergebnisse dennoch in der öffentlichen Debatte als Belege für kausale Zusammenhänge präsentiert werden oder ihm unterstellt werde, er selbst präsentiere solche Effekte, sei dies ein Missverständnis.

Die Tücken der Vereinfachung

Auf die vorsichtigen methodischen Relativierungen folgen auf rund 400 Seiten wie in der ersten Ausgabe erneut weitreichende Vereinfachungen. In der überarbeiteten Ausgabe präsentiert Hattie die Effekte neu als Thermometer statt als Barometer und wählt damit wieder eine Darstellung, die komplexe Zusammenhänge scheinbar klar und auf einen Blick zugänglich macht. Für Fachpersonen der Bildungspraxis wirkt diese Übersicht wie eine verlässliche Orientierungshilfe. Die zugrunde liegende Forschung wirft jedoch häufig weitere Fragen auf oder liefert heterogene Ergebnisse, die Hattie nur teilweise erläutert. Mit den in der Einleitung vorgenommenen Relativierungen setzt er bei gleichzeitiger, scheinbar eingängiger Vereinfachung fundierte Methodenkenntnisse voraus – etwa darüber, wie Studien angelegt sind, welche Verzerrungen entstehen können und warum Ergebnisse nicht vorschnell als kausal interpretiert werden dürfen. Was das für die Verwendung von Visible Learning als Orientierungshilfe bedeutet, wird deutlich, wenn man drei exemplarische Einflussfaktoren genauer betrachtet:

Der **relative Alterseffekt (RAE)** beschreibt das Phänomen, dass innerhalb eines Jahrgangs jüngere Kinder im Durchschnitt schlechtere Schulleistungen erzielen als ihre älteren Mitschülerinnen und Mitschüler. Hattie stützt sich dabei auf eine einzelne, nicht veröffentlichte Studie, wodurch die Nachvollziehbarkeit der Evidenzbasis nicht gegeben ist (Maynard, 2013). Aktuelle, methodisch solide Einzelstudien und eine neue Meta-Analyse belegen den RAE hingegen robust und zeigen deutliche Folgen für Leistung, Bildungslaufbahnen und Selbstkonzept. Hatties ausgewiesener Effekt stimmt inhaltlich zwar mit dem aktuellen Forschungsstand überein, beruht jedoch auf einer nicht transparent dokumentierten Studienauswahl. Um die Aussage kritisch zu prüfen und die Tragfähigkeit des Effekts verlässlich zu beurteilen, ist eine eigenständige Recherche nach weiteren Forschungsarbeiten notwendig.

Mobbing und Leistung: Die Frage, ob Kinder schlechtere Schulleistungen zeigen, weil sie Mobbing erfahren oder ob sie Opfer von Mobbing werden, weil sie in der Schule schlechter abschneiden, ist nicht unbedeutend – nicht zuletzt, wenn es darum geht, Risikogruppen zu identifizieren und Präventionsprogramme gezielt anzusetzen. Zwar warnt Hattie in der Einleitung allgemein vor der Fehlinterpretation kausaler Zusammenhänge, doch fehlt an dieser Stelle ein expliziter Hinweis. Eine neuere Meta-Analyse berücksichtigt hingegen Studien, die methodisch anspruchsvolle Verfahren wie *Propensity Score Matching* oder *Regression Discontinuity Designs* einsetzen, um den kausalen Zusammenhang zwischen Mobbing und Leistung differenzierter zu prüfen (Schoeler et al. 2018). Diese Arbeit, ebenso wie weitere Einzelstudien (Eriksen et al., 2014; Gorman et al., 2021), die explizit die Frage der Kausalität aufnehmen, könnten das Bild deutlich schärfen, werden in Hatties Analyse jedoch nicht berücksichtigt (Hattie zitiert zu Mobbing und Schulleistungen folgende Studien: Gardella et al., 2013; Moore et al., 2017).

Sommerferien-Effekt: Die Forschung zum sogenannten Sommerferien-Effekt untersucht, wie sich lange unterrichtsfreie Phasen auf die Lernentwicklung von Schülerinnen und Schülern auswirken. Im Zentrum steht die Frage, ob und in welchem Ausmass Lernrückstände entstehen. Hattie stützt sich in seiner Darstellung auf eine ältere Studie aus den 1990er-Jahren, die im Durchschnitt Lernverluste, also einen negativen Effekt, nachweist (Cooper et al., 1996). Neuere Analysen zeichnen jedoch ein deutlich differenzierteres Bild: Während für einige Schülerinnen und Schüler die Sommerferien erhebliche Leistungseinbussen mit sich bringen, machen andere sogar Lernfortschritte

(Atteberry & McEachin 2021; Fitzpatrick & Burn, 2019). Hattie führt diese aktuelleren Arbeiten in seiner Literaturübersicht zwar auf, kann sie jedoch in seiner Meta-Analyse von Meta-Analysen nicht berücksichtigen. Durch die Nennung dieser Studien können die Lesenden zumindest erkennen, warum der ausgewiesene Gesamteffekt mit Vorsicht zu interpretieren ist und die tatsächlichen Auswirkungen möglicherweise deutlich heterogener ausfallen, als die vereinfachte Darstellung vermuten lässt. Gleichzeitig wird ersichtlich, wie stark die ausgewiesenen Effekte von der Auswahl der berücksichtigten Studien abhängen.

Schlussfolgerung

Diese drei Beispiele, man könnte weitere nennen, verdeutlichen das eigentliche Problem von Visible Learning. Die Hilfe zur Orientierung, die Hattie geben möchte, kann leicht in die Irre führen. Sie vermittelt eine Genauigkeit, die so nicht existiert und läuft damit Gefahr, falsche Gewissheiten zu schaffen. Die komplexe Forschungslage wird auf eine Zahl reduziert, während zugleich davor gewarnt wird, diese Zahl vereinfachend weiterzuverwenden.

Darin liegen gleich zwei Widersprüche, die auch in der neuen Ausgabe ungelöst bleiben. Zum einen warnt Hattie ausdrücklich vor vorschnellen Ursache-Wirkungs-Schlüssen, präsentiert die Einflussfaktoren aber in einer Form, die genau diesen Eindruck vermittelt. Zum anderen fordert er auf, die präsentierten Effekte zu kontextualisieren und kritisch zu prüfen, was voraussetzen würde, dass Lesende über ein hohes Mass an methodischem Verständnis verfügen. Zugleich scheint der Erfolg seines Werks gerade darin zu liegen, dass es sich an ein Publikum richtet, für das eine solche Prüfung angesichts des Umfangs und der Komplexität der Themen kaum realistisch ist. Wer hingegen über das nötige methodische Wissen verfügt, erkennt rasch, dass die präsentierte Orientierung auf keiner verlässlichen Grundlage beruht.

So wirkt auch «*Visible Learning: The Sequel*» am Ende wie ein Navigationssystem, das mitten auf der Autobahn zur Kehrtwende auffordert: Die Stimme klingt klar und bestimmt, doch sie ist nicht verlässlich – und wir fahren weiter, um uns an der nächsten Ausfahrt neu zu orientieren.

Literatur

- Atteberry, A. & McEachin, A. (2021). School's out: The role of summer in understanding achievement disparities. *American Education Research Journal*, 58(2), 239–282.
<https://doi.org/10.3102/0002831220937285>
- Cooper, H., Nye, B., Charlton, K., Lindsay, J. & Greathouse, S. (1996). The Effects of Summer Vacation on Achievement Test Scores: A Narrative and Meta Analytic Review. *Review of Educational Research*, 66(3), 227–268.
- Eriksen, T. L. M., Nielsen, H. S., & Simonsen, M. (2014). Bullying in elementary school. *The Journal of Human Resources*, 49(4), 839–871.
- Fitzpatrick, D. & Burn, J. (2019). Single-track year-round education for improving academic achievement in U.S. K-12 schools: Results of a meta-analysis. *Campbell Systematic Reviews*, 15.
<https://doi.org/10.1002/cl2.1053>
- Gardella, J. H., Fisher, B. W., & Teurbe-Tolon, A. R. (2017). A systematic review and meta-analysis of cyber-victimization and educational outcomes for adolescents. *Review of Educational Research*, 87(2), 283–308.
<https://doi.org/10.3102/0034654316689136>
- Gorman, E., Harmon, C., Mendolia, S., Staneva, A., & Walker, I. (2021). Adolescent school bullying victimization and later life outcomes. *Oxford Bulletin of Economics and Statistics*, 83(4), 995–1030.
<https://doi.org/10.1111/obes.12432>
- Hattie, J. (2008). Visible learning. A synthesis of over 800 meta-analyses relating to achievement. Routledge: London, New York.
- Hattie, J. (2023). Visible Learning: The sequel. A synthesis of over 2100 meta-analyses relating to achievement. Routledge: London, New York.
- Maynard, B. R. (2013). Is there an academic benefit to participating in extracurricular activities? A systemic review and meta-analysis (unveröffentlichte Dissertation).
- Moore, S. E., Norman, R. E., Suetani, S., Thomas, H. J., Sly, P. D., & Scott, J. G. (2017). Consequences of bullying victimization in childhood and adolescence: A systematic review and meta-analysis. *World Journal of Psychiatry*, 22(1), 60–76.
<https://doi.org/10.5498/wjp.v7.i1.60>
- Oggenfuss, C. & Wolter, S. C. (2015). Die scheinbare Leichtigkeit des Forschungskonsums. *Bildung Schweiz* 6/2015, 40–41.
- Pant, H. A. (2011). Aufbereitung von Evidenz für bildungspolitische und pädagogische Entscheidungen: Metaanalysen in der Bildungsforschung. *Zeitschrift für Erziehungswissenschaft* 17, 79–99.
- Schoeler, T., Duncan, L., Cecil, C. M., Ploubidis, G. B., & Pingault, J.-B. (2018). Quasi-experimental evidence on short- and long-term consequences of bullying victimization: A meta-analysis. *American Psychological Bulletin*, 144(12), 1229–1246.
<https://doi.org/10.1037/bul0000171>
- Terhart, E. (2014). Has John Hattie really found the holy grail of research on teaching? An extended review on Visible Learning. *Journal of Curriculum Studies*, 43(3), 425–438.