

(Die deutsche Fassung dieses Texts steht unten nach der englischen.)
2014-10-25

Errors in John Hattie's "Visible Learning"

Rolf Schulmeister and Jörn Loviscach

rolf@schulmeister.com, joern.loviscach@fh-bielefeld.de

On close scrutiny, "Visible Learning" – the oft-cited meta-analysis of meta-analyses – displays considerable issues in terms of the selection of studies and the methods employed. This is what we found when examining sources and computations for selected parts of the book. A thorough analysis of Hattie's work is required.

What works in education and what does not? – The answers to this question are riddled with ideology stemming from the different schools of pedagogy. It is a constant desire of educational researchers to settle this issue once and for all through quantitative measurements of the learning that actually takes place under a huge range of didactic interventions and conditions. For many this hope seemed to have become a reality thanks to "Visible Learning" by John Hattie, because he collected hundreds of meta-analyses, grouped them into more than hundred "influences" and ranked these from "desired effects" down to "reverse effects".

We wanted to establish how reliable Hattie's results are. To this end, we picked a number of the influences defined by Hattie (according to the second German edition of 2014), looked at the original meta-analyses (if they were obtainable with justifiable effort) and at the numbers. Our results (Schulmeister & Loviscach, 2014) call the ranking order given by Hattie into question and cast serious doubts on both his method and his results:

1. **Questionable Studies.** Many of the meta-analyses used by Hattie are dubious in terms of methodology. Hattie obviously did not look into the individual empirical studies that form the bases of the meta-analyses, but used the latter in good faith. In addition, Hattie relies on many "unpublished dissertations" – which often can be considered exercises in applying statistics rather than thorough research. Hattie would have needed to examine the value of each meta-analysis before incorporating it into his meta-analysis of meta-analyses.
2. **Questionable Groupings.** Many of the meta-analyses do not really match the same effect group (i.e., the influence) in which Hattie refers to them. For instance, in the group "feedback", studies investigating the effect of student feedback on teachers are mixed with studies that examine the effect of teacher feedback on students.
3. **Lack of Diligence.** It seems that Hattie did not read all meta-analyses that he refers to. For instance, in the influence group "concentration, persistence, and engagement" he includes a meta-analysis that is concerned with the concentration of industrial power, not with mental focus. In addition, reading the original meta-analyses in order to verify the numbers given by Hattie we found gaps and discrepancies.

4. **Apples and Oranges.** Even where he has grouped meta-analyses correctly by their *independent* variables such as instructional interventions, Hattie has in many cases mixed apples and oranges concerning the *dependent* variables. In some groupings, however, both the independent and the dependent variables do not match easily. For instance, in the group “feedback”, a meta-analysis using music to reinforce behavior is grouped with other studies using instructional interventions that are intended to elicit effects on cognitive processes.
5. **Misleading Statistics.** The effect size given per influence is the mean value of a very broad distribution. For instance, in “Inductive Teaching” Hattie combines two meta-analyses with effect sizes of $d = 0.06$ and $d = 0.59$ to a mean effect size of $d = 0.33$ with a standard error of 0.035. This is like saying “this six-sided dice does not produce numbers from 1 to 6; rather, it produces the number 3.5 in the mean, and we are pretty sure about the first decimal place of this mean value.” However, the actual results often are much smaller or much larger than the mean. The details of the intervention matter crucially. (Hattie’s method to compute the standard error of the averaged effect size as the mean of the individual standard errors – if these are known at all – is statistical nonsense anyway.)
6. **Untenable Ranking List.** Hattie sorts the influences by the average effect sizes. The ranking thus obtained has gained much attention in the public. However, if one corrects the errors mentioned above, list positions take big leaps up or down. Even more concerning is the absurd precision this ranking conveys. It only shows the averages of effect sizes but not their considerable variation within every group formed by Hattie and even more so within every individual meta-analysis.

Conclusion

To think that didactics can be presented as a clear ranking order of effect sizes. It is a dangerous illusion. To an extreme degree, the effect of a specific intervention depends on the circumstances. Focusing on the mean effect sizes and ignoring their considerable variations and condensing the data to a seeming exact ranking order, Hattie pulls the wool over his audience’s eyes. But a more devastating error is lurking in the background: School is reduced to whatever can be “measured” in an exam-like test. As much that has been “learned” in school is quickly forgotten, this approach seems myopic to us, possibly even dangerous, as it ignores the long-term effects of school.

References

- Hattie, J. (2014). *Lernen sichtbar machen*. 2nd revised ed. Hohengehren: Schneider.
Translation of: Hattie, J. (2009). *Visible Learning*. New York: Routledge.
- Schulmeister, R., & Loviscach, J. (2014). Kritische Anmerkungen zur Studie “Lernen sichtbar machen” (Visible Learning) von John Hattie. SEMINAR 2/2014, pp. 121-130.

Fehler in John Hatties „sichtbarem Lernen“

Rolf Schulmeister und Jörn Loviscach

rolf@schulmeister.com, joern.loviscach@fh-bielefeld.de

Die viel zitierte Meta-Meta-Analyse „Visible Learning“ weist beträchtliche inhaltliche und methodische Schwächen auf. Das zeigen unsere exemplarischen Proben der Quellen und Rechnungen. Das Werk bedarf einer grundlegenden Überprüfung.

Was wirkt im Unterricht und was nicht? -- Die Beantwortung dieser Frage ist ein ideologisches Schlachtfeld verschiedener Schulen der Pädagogik. Wie schön wäre es, sie durch Messungen von Unterrichts(miss)erfolg ein für alle Mal zu beantworten. Diese Hoffnung hat sich für viele mit dem Werk „Visible Learning“ von John Hattie (2014) erfüllt: Er hat in dieser Meta-Meta-Analyse Hunderte von Meta-Analysen zu Themenkomplexen zusammengefasst und diese zu einer Rangfolge von positiv bis kontraproduktiv sortiert.

Daran interessiert, wie verlässlich Hatties Ergebnisse sind, haben wir Stichproben der von Hattie verarbeiteten Originalliteratur beschafft – soweit mit vertretbarem Aufwand möglich – und studiert. Unsere Ergebnisse (Schulmeister & Loviscach, 2014) stellen die von Hattie angegebene Rangfolge in Frage und lassen gravierende Bedenken zum Vorgehen aufkommen:

1. **Zweifelhafte Studien.** Viele der von Hattie herangezogenen Meta-Analysen sind methodisch anzweifelbar. Hattie hat offensichtlich die in diese Meta-Analysen eingegangenen empirischen Einzelstudien nicht auf ihre inhaltliche oder methodische Qualität überprüft, sondern die Stichproben (die Meta-Analysen ja darstellen) guten Glaubens übernommen. Außerdem hat Hattie viele „unpublished dissertations“ einbezogen – oftmals wenig belastbare Fingerübungen in Statistik. Vor der Einbeziehung in seine Meta-Meta-Analyse hätte Hattie die Aussagekraft der benutzten Meta-Analysen beleuchten müssen.
2. **Fragwürdige Zuordnungen.** Viele der herangezogenen Meta-Analysen passen nicht in den Themenkomplex, dem Hattie sie zugeordnet hat. So finden sich Studien, die den Effekt der Beurteilung von Lehrern durch Schüler erfassen, unter Studien, die die Rückmeldung auf Schüler messen sollen, weil beide nominell unter Feedback verzeichnet werden.
3. **Mangelnde Sorgfalt.** Augenscheinlich hat Hattie nicht alle Meta-Analysen, die er zitiert, wirklich gelesen, denn beispielsweise findet sich im Effektkomplex „Konzentration, Ausdauer und Engagement“ eine Studie, deren Thema die Konzentration industrieller Macht ist, nicht aber die Konzentration beim Lernen. Außerdem konnten wir einige von Hattie angegebene Zahlenwerte nicht so in den verwendeten Meta-Analysen wiederfinden.
4. **Äpfel und Birnen.** Selbst dort, wo er Meta-Analysen nach ihren *unabhängigen* Variablen einigermaßen treffend ausgewählt hat, um den bekannten Vorwurf des Vergleichs von Äpfeln mit Birnen zu vermeiden, hat Hattie in vielen Fällen übersehen,

dass die *abhängigen* Variablen nicht kompatibel waren. Manchmal passen aber weder die unabhängigen noch die abhängigen Variablen so recht zusammen: Im selben Effektkomplex „Feedback“ wird zum Beispiel eine Meta-Analyse, die Musik mit dem Ziel der Verhaltensverstärkung (reinforcement) einsetzt, mit anderen Studien verrechnet, die Unterrichtsinterventionen als kognitives Feedback nutzen.

5. **Irreführende Statistik.** Die pro Phänomen angegebenen Effektstärken sind die Mittelwerte von sehr breiten Verteilungen. Zum Beispiel beim Effektkomplex „Induktives Vorgehen“ zieht Hattie zwei Meta-Studien mit Effektstärken von $d = 0,06$ und $d = 0,59$ zur mittleren Effektstärke von $d = 0,33$ und einem Standardfehler von $0,035$ zusammen. Das ist so, als wenn man von einem Würfel nicht sagen würde, er liefert Zahlenwerte von 1 bis 6, sondern sagen würde: „Der Würfel liefert den Wert 3,5 und wir sind uns bei diesem Mittelwert fast auf eine Nachkommastelle sicher.“ Das reale Ergebnis ist aber oft viel kleiner oder viel größer als der Mittelwert. Will sagen: Auf die Details der didaktischen Intervention kommt es an. (Hatties Verfahren, den Gesamt-Standardfehler durch Mittelung der einzelnen Standardfehler – soweit überhaupt bekannt – zu berechnen, ist sowieso statistischer Nonsens.)
6. **Unsinnige Rangliste.** Hattie ordnet seine Themenkomplexe nach den ermittelten Effektstärken, um ein Ranking zu bilden. Diese Rangliste hat in der Öffentlichkeit die größte Aufmerksamkeit erfahren. Korrigiert man Themenkomplexe, in denen falsche Zuordnungen oder Berechnungsfehler vorkommen, springen Themenkomplexe aber in der Rangliste hin und her. Vor allem jedoch gaukelt diese Rangliste eine absurde Präzision vor, weil sie nur den jeweiligen Mittelwert abbildet, nicht die – teilweise heftigen – Schwankungen pro Themenkomplex und pro verwendeter Meta-Studie.

Fazit

Dass sich Didaktik als eine simple Rangfolge von Effektstärken abbilden ließe, ist eine gefährliche Illusion. Wie eine bestimmte Intervention wirkt, hängt extrem von den Umständen ab. Mit den kleingerechneten Streubreiten und der scheinbar exakten Rangfolge streut Hattie seinem Publikum Sand in die Augen. Im Hintergrund lauert ein noch schwerwiegenderer Denkfehler: Schule wird hier auf das reduziert, was in einer abschließenden Leistungsprüfung „messbar“ ist. Schon, weil der meiste Schulstoff schnell wieder vergessen ist, scheint uns dieser Ansatz mindestens kurzsichtig, wenn nicht sogar gefährlich, weil er die langfristigen Effekte der Schule ausblendet.

Literatur

Hattie, J. (2014). Lernen sichtbar machen. 2. korr. Aufl. Hohengehren: Schneider.
Schulmeister, R., & Loviscach, J. (2014). Kritische Anmerkungen zur Studie „Lernen sichtbar machen“ (Visible Learning) von John Hattie. SEMINAR 2/2014, S. 121-130.