

Georg Lind

Zwischenruf: Kleine Klassen von großem Gehabe erschlagen

Eine ganz Reihe von vorbildlichen Interventionsstudien in den USA hatte in den 1980er-Jahren gezeigt, dass im Bereich der Vorschul- und Grundschulbildung (bis Klasse 3) eine **Reduzierung der Klassengröße zu einer deutlichen Lernsteigerung bei Kindern** beiträgt, u.zw. vor allem bei benachteiligten Kindern, und dass diese „Investition“ sich noch in den höheren Klassen und sogar noch im College auswirkt. Man kann sich leicht ausrechnen, dass die Mehrkosten, die kleine Klassen in der frühen Schulzeit machen, sich später voll amortisieren, abgesehen von dem persönlichen Gewinn, den benachteiligt Kinder durch diese Förderung haben, und dem Gewinn für unsere Demokratie, den eine effektivere Bildung für benachteiligte Kinder hat.

Dieser gut gesicherte Befund wurde von dem in vielen Kultusministerien noch immer gefeierten Bildungsforschungs-„Guru“ *John Hattie* („Visible Learning“) mit Hilfe einer einzigen Zahl erschlagen: „ $d = 0,13$ “. Diese Zahl („Effektstärke“) habe er aus „736 Meta-Analysen, ... die auf 50.000 Studien mit ca. 250 Millionen Lernenden zurückgreifen“ destilliert. Sie liege weit unter dem d von 0,40, das er sich als Kriterium ausgedacht hat.

Hans Brügelmann (2014) sowie *Rolf Schulmeister* und *Jörn Loviscach* (2014, 2015) haben **die Hattie-Studie einer detaillierten Kritik** unterzogen. Sie weisen auf zahlreiche Probleme bei der Berechnung und Zusammenfassung von Effektstärken hin und auf große Qualitätsmängel der Originalstudien, auf die *Hattie* nicht eingeht. *Volker Hagemeyer* (2018) hat mit einem neuen Artikel jetzt nachgelegt: „[Ich] habe die Literatur, die *Hattie* (2009) im Abschnitt „class size“ auf den Seiten 85-87 anführt, durchgearbeitet (bis auf einen Titel, der nicht erhältlich war). Dabei hat sich gezeigt, dass *Hatties* Abschnitt ‚class size‘ in ungewöhnlich großer Zahl Fehler enthält: Falsche Jahreszahlen, falsche Schülerzahlen, falsche Vorzeichen bei Effektstärken usw. (siehe die Abschnitt 2.3 und 2.4 in meinem Text von 2018). Dass die vielen Fehler, die man allein auf drei Seiten bei *Hattie* findet, bisher übersehen wurden, bestätigt, wie nachhaltig sich „die Suggestion von Objektivität durch den Einsatz standardisierter Leistungstests“ auswirkt.“

Ich habe schon früher vor der *Hattie*-Studie gewarnt (Lind 2011). Unter anderem habe ich darauf verwiesen, dass das Effektstärke-Maß „ d “ die Effekte auf die vorgefunden Streubreiten der Daten relativiert, also nichts

über die absoluten Effekte aussagt, die kleine Klassen haben.

In Stichproben, die eine geringe Streubreite aufweisen, ist die Effektstärke schon aus mathematischen Gründen höher als in Stichproben mit einer großen Streubreite. Ich habe auch darauf hingewiesen, dass riesige Fallzahlen kein Ausgleich für eine schlechte Forschungsanlage sein können. Falsche Daten bleiben falsch, auch wenn sie vielfach erhoben wurden. *Hattie* hat Studien mit sehr unterschiedlicher Qualität und sehr unterschiedlichem Umfang in einen Topf geworfen und deren Effektstärken dann auch noch arithmetisch gemittelt.

Was soll eine Bildungsadministration oder ein Lehrperson damit anfangen? Zudem beschränken sich alle Meta-Analysen von *Hattie* auf einfache Testleistungen in wenigen Fächern wie Rechnen und Lesen. Selbst da erfassen sie keine höherwertigen Lernleistungen wie Zahlenverständnis und Schreibfähigkeit, und schon gar nicht berücksichtigen diese Tests die vielen anderen Lernleistungen, die Schulen fördern sollen wie in den Fächern Geschichte, Geographie, Musik, Kunst usw., und schon gar nicht moralisch-demokratische Fähigkeiten, die für unsere Demokratie entscheidende Bedeutung haben. **Wie kann man so eine dürftige Zahl als Richtschnur für Bildungspolitik und Unterricht ausgeben?**

Otmar Patzelt (persönliche Mitteilung) hat mich auf ein weiteres Problem aufmerksam gemacht. Bei *Hattie* bleibe „die **Anwendungshäufigkeit** der Unterrichtsverfahren **unberücksichtigt**. Das ist m. E. ein Problem, weil dadurch der Eindruck entsteht, man müsse nur **Verfahren mit hohen Effektstärken möglichst häufig anwenden, um gute Lernerfolge zu erzielen, was ein Trugschluss ist.**“ *Patzelt* hat recht. *Hattie* hat die Anwendungshäufigkeit von Verfahren nirgends berücksichtigt. Ich fasse diesen Aspekt unter dem Stichwort „Effizienz“ zusammen, das neben der Anwendungshäufigkeit und noch andere Dinge berücksichtigt wie die Kosten an Zeit und Geld sowie die Ausbildung der Lehrkräfte, die eine Anwendung erfordert.

Mir ist bei der vielfachen Anwendung der Konstanzer Methode der Dilemma-Diskussion (KMDD) aufgefal-

len, dass sich auch hier das in der Wirtschaftswissenschaft bekannte Phänomen des Grenznutzens zeigt: Der Effekt nimmt zwar mit jeder Wiederholung zu, aber nicht linear. Vielmehr wird der Zuwachs an Moralkompetenz mit jeder weiteren Anwendung immer weniger. Es scheint, dass mit einer einmaligen KMDD-Veranstaltung so viel Effekt erreicht werden kann wie mit allen weiteren Anwendungen zusammen. Daher empfehle ich inzwischen, die KMDD nur ein- bis zweimal im Jahr anzuwenden, und die dadurch frei werdende Lehrkapazität dafür zu nutzen, mehr Schüler in den Genuss der KMDD kommen zu lassen.

Wie wichtig eine angemessene Ausbildung der Lehrkraft in der jeweiligen Methode ist, haben Oser und Patry (1986) schon vor einiger Zeit unter dem Stichwort „Transformationsvalidität“ aufgezeigt. Damit meinen sie das Problem, dass die Unterrichtsmethoden oft nicht so gut umgesetzt werden, wie es erforderlich wäre, damit die erhoffte Wirkung eintritt. **Methoden sind immer nur so gut wie die Lehrperson**, die sie einsetzt. Die Dilemma-Methode von Lawrence Kohlberg und Moshe Blatt ist vermutlich vor allem an der mangelnden Ausbildung der Lehrpersonen gescheitert. Kohlberg hat sie für tot erklärt (Althof 2015), obwohl die Effektstärken der Methode hoch waren (Lind 2002). Er begründete dies damit, dass die Lehrpersonen sie nicht verwenden wollten. Offenbar fühlten sie sich von der Methode mangels Ausbildung überfordert.

Eine Kollegin schrieb mir einmal, dass sie sehr enttäuscht von der Dilemmamethode sei. Sie hätte diese bei ihren Schülern sechs Wochen lang jede Woche angewendet, bei ihren Teilnehmern aber einen Niedergang der Moralkompetenz feststellen müssen. Sie war wohl dem Trugschluss erlegen, dass man einfach nur die Frequenz erhöhen muss, um einen Effekt zu erreichen. Dabei hat sie übersehen, es an ihrer mangelnden Ausbildung lag. Vermutlich hat sie die Schüler mit den Dilemmadiskussionen am Ende nur noch genervt.

Wir haben daraus gelernt. Wir bieten für die KMDD eine intensive Ausbildung mit Zertifizierung als KMDD-Lehrer an. Sie umfasst einen einwöchigen Workshop und zwei Monate *training-on-the-job*. Wer sie durchlaufen hat, erzielt mit der Methode meist deutliche Effekte (Lind 2002) und gibt die Methode auch nicht mehr auf, weil sie sich, richtig angewendet, auch in einer gesteigerten Lernmotivation der Schüler bemerkbar macht.

Weder zur Effizienz einer Methode noch zur Transformationsvalidität der untersuchten Studien finden sich bei Hattie Hinweise. Er berichtet nicht, welche Kosten die jeweiligen Verfahren verursachen, noch welche Anwendungsfrequenz optimal ist. Er sagt auch nicht, welche Ausbildung erforderlich ist, um den erwarteten Effekt einer Methode zu erzielen: z. B. ob die

Lehrkräfte, die kleine Klassen haben, dafür ausgebildet sein müssen, und ob die Lehrkräfte, die Offenen Unterricht praktizierten, wissen müssen, wie man Schüler/innen zum selbstständigen Lernen anleitet (Leitz/Müller 2016; vgl. auch Leitz 2016).

Die riesige Zahl der zitierten Untersuchungen in der Hattie-Studie lenkt von diesen Schwächen ab. Mit großem Gehabe erschlägt sie nicht nur die kleinen Klassen, deren hohe Wirksamkeit von seriösen Interventionsstudien zweifelsfrei belegt sind. Sie legt auch sonst viele falsche Spuren, weil entscheidende Informationen fehlen. Es wäre zu hoffen, dass der Hattie-Spuk endlich aus den Kultusministerien und Lehrerfortbildungen vertrieben wird.

Literatur

- Althof, W.: Just community course and transformations: A conceptual archeology of Kohlberg's approach to moral and democratic education. In: Zizek, B./Garz, D./Nowak, E. (Eds.): Kohlberg revisited. Amsterdam 2015, S. 51-90.
- Brügelmann, H.: Gilt nach Hattie: Je häufiger, desto besser? Zur Bedeutung von „Evidenzbasierung“ für pädagogisches Handeln vor Ort. In: E. Terhart, E. (Hrsg.): Die Hattie-Studie in der Diskussion, Probleme sichtbar machen. (Reihe Bildung kontrovers, 2) Stuttgart 2014, S. 38-50.
- Hagemeister, V.: Eine Analyse der bei Hattie zitierten Klassenfrequenzstudien. Forschungsergebnisse zum Thema Klassenfrequenz. In: bildungsforschung, Jg. 2018, H. 1, S. 1-26. https://open-journals4.uni-tuebingen.de/ojs/index.php/bildungsforschung/article/download/205/pdf_1
- Leitz, I.: Unterrichtsstörungen in der Hattie-Studie – eine Nachbetrachtung. In: Lehren & Lernen 42 (2016), H. 12, S. 34f.
- Leitz, I./Müller, J.: Ist Offener Unterricht tatsächlich ineffektiv? Ein kritischer Blick auf die Hattie-Studie. In: Lehren & Lernen 42 (2016), H. 7, S. 36-39.
- Lind, G.: Ist Moral lehrbar? Ergebnisse der modernen moralpsychologischen Forschung. Berlin 2002.
- Lind, G.: Pädagogik oder Struktur? Auch die Hattie-Studie weiß keinen Rat. In: Hessische Lehrer-Zeitung 11/2011, S. 26f.
- Lind, G.: Welche Bildungsforschung sollte mehr Beachtung in der Politik finden? In: Lehren & Lernen 45 (2019), H. 1, S. 35-37.
- Oser, F./Patry, J.-L.: Interventionsstudien für sozial-kognitive Kompetenz: Beispiele und theoretische Überlegungen zur Transformations-Validität. In: Unterrichtswissenschaft 14 (1986), S. 254-268.
- Schulmeister, R./Loviscach, J.: Kritische Anmerkungen zur Studie „Lernen sichtbar machen“ (Visible Learning) von John Hattie. In: Seminar 2/2014, S. 121-130; auch in: Lehren & Lernen 41 (2015), H. 4, S. 33-39.