

Welche Bildungsforschung sollte mehr Beachtung in der Politik finden?

Plädoyer für eine alternative Bildungsforschung: Wirksamkeitsstudien

Im Bereich der Lehr-Lern-Forschung ist die herrschende empirische Bildungsforschung höchst defizitär: Sie erhebt weder die Effektstärken noch die Effizienz der untersuchten Lehrmethoden, sie teilt weder etwas mit über das Ausmaß des Realisierungsaufwandes (Effizienz) einer empfohlenen Lern-Methode noch über deren Effektivität (Lernerfolge), mithin auch nichts über die (mögliche) Effizienz und Effektivität bildungspolitischer Maßnahmen. Der Beitrag plädiert für andere Untersuchungskonzepte, mit denen Effektivität und Effizienz für definierte Schülergruppen ermittelt werden können. Das erfordert nicht nur andere Erhebungsinstrumente, sondern qualifizierte Lehrpersonen. Der enorme finanzielle Aufwand für die vorherrschende empirische Bildungsforschung sollte eingestellt und in Effizienz- und Effekt-orientierte Wirksamkeitsstudien umgeleitet werden.

► Stichwörter: **Bildungspolitik, Politikberatung, Empirische Bildungsforschung, Lehr-Lern-Forschung, Wirksamkeitsforschung**

Nina Kolleck hat in der ZEIT vom 27.9.2018 beklagt, dass die Bildungspolitik zu wenig die Ergebnisse der Forschung berücksichtige. Dabei gab es m.E. in den letzten Jahre **eher zu viel Aufregung um solche Studien und zu viele bildungspolitische Schnellschüsse.**

Am Beispiel der Studien zum Rechtschreibunterricht zeigt Hans Brügelmann (oben in diesem Heft) auf, wie widersprüchlich die Ergebnisse der Bildungsforschung oft sind und wie leicht man falsche Schlüsse für die Bildungspolitik ziehen kann, wenn man sich auf sie stützt. Ulrich Herrmann (s. oben in diesem Heft) zeigt (wie Brügelmann), dass die Vorstellung möglicher eindeutiger „Anwendung“ von Wissenschaft auf Politik und Praxis zum einen irrig ist und zum andern das Forschungsparadigma der Praxisforschung mit der Erzeugung von praxisrelevantem Veränderungswissen an die Stelle der empirischen Bildungsforschung (mit aggregierten Massendaten) treten muss.

Worin liegt das Problem?

In den Beiträgen von Kolleck und Brügelmann werden drei Gründe angesprochen: **geringe Repräsentativität, geringer Umfang von Studien, selektive Wahrnehmung durch die Medien.** Die selektive Wahrnehmung von Befunden (man nimmt nur Studien wahr, die in das eigene Weltbild passen) ist ein großes Problem, aber keines, das die Wissenschaft ändern kann. Sie kann aber bezüglich der beiden ersten Gründe etwas tun und sollte das auch schleunigst.

Oft gilt der **Umfang einer Studie** als Gütemerkmal: Je mehr Schüler in die Studien einbezogen wurden,

umso gültiger seien ihre Ergebnisse. Dies ist aber völlig falsch. Vielmehr liegt genau in dem großen Umfang von Studien wie PISA oft ein Grund für falsche Schlussfolgerungen. Mit wachsendem Umfang wächst rein mathematisch die Wahrscheinlichkeit, dass minimale Unterschiede statistisch „signifikant“ werden, was überhaupt nichts mit „bedeutsam“ (der deutschen Bedeutung von „signifikant“) zu tun hat. **Die „signifikanten“ Unterschiede sind wegen ihrer Winzigkeit oft nicht stabil (d. h. in neuen Studien werden sie selten bestätigt) und pädagogisch meist unbedeutend.** Der Verwechslung von statistischer „Signifikanz“ mit „pädagogischer Bedeutsamkeit“ ist einer der Hauptgründe für die große Konfusion, die Ergebnisse von Studien oft anrichten.

Mit „**Repräsentativität**“ ist meist gemeint, dass die teilnehmenden Schüler in einer Studie die Vielfalt der Schülerschaft insgesamt abbilden. Aber dies ist nur dann ein Problem, wenn die Studien einen Zustand beschreiben sollen, wie etwa Wählerstudien. Für analytische Studien ist es viel wichtiger, dass sie angelegt werden, dass sie zeigen, wie stark die Effekte und die Effektivität bestimmter Lehrmethoden sind, und dass sie es möglich machen, den Effekt eindeutig auf die jeweils getestete Lehrmethode zurückzuführen.

So wie wir vom Wetterbericht wissen wollen, wie viel Grad wärmer oder kälter es heute im Vergleich zu gestern ist, und nicht, ob die Unterschiede „statistisch signifikant“ sind, so **wollen wir auch von der Bildungsforschung wissen, wie groß der absolute Lernzuwachs bestimmter Lehrmethoden im Vergleich zu anderen ist.** Zudem wollen wir wissen, wie groß der damit verbun-

dene zeitliche und finanzielle Aufwand im Vergleich zum Erfolg ist, also (a) wie viel Unterrichts- bzw. Lernzeit und (b) welche Ausbildung der Lehrperson nötig sind, um den gefundenen Lerngewinn zu erreichen.

Leider werden Effektstärke und Effizienz in fast keiner Studie berichtet, oft selbst dann nicht, wenn die entsprechenden Daten vorliegen. Statt sie dem Leser mitzuteilen, werden sie oft bis zur Unkenntlichkeit umgeformt (oft in unsägliche Rangplatzangaben). Oft werden Daten zur Effektstärke und Effizienz gar nicht erhoben. Dabei sind diese Informationen für bildungspolitische Entscheidungen von größter Wichtigkeit. Es macht einen großen Unterschied, ob wir über die Einführung einer Lehrmethode diskutieren, die nur geringe oder große Auswirkungen auf die Fähigkeitsentwicklung von Schülern hat, und ob die Einführung und Durchführung mit riesigen oder geringen Kosten verbunden ist. Darüber erfahren wird aus der Bildungsforschung meist gar nichts.

Dass dies durchaus möglich, habe ich auf dem Gebiet der moral- und demokratiepädagogischen Forschung gezeigt (Lind 2015). Die Berechnung von Maßen für die relative und absolute **Effektstärke** ist einfach, sogar einfacher als die Berechnung von statistischen Signifikanzen. Die Abschätzung von **Effizienz** ist ein schwereres Problem, das speziellen Sachverstand erfordert. Wenn Methode A 10% effektiver als Methode B ist, dafür aber die dreifache Zeit zum Lernen benötigt, wird wohl niemand fordern, diese Methode einzuführen. Leider finden sich in den Pressemitteilungen keinerlei Hinweise auf den Zeitaufwand, den eine bestimmte Lehrmethode auf Seiten der Lehrkraft und der Schüler erfordert. Manchmal ist die Lektüre der umfangreichen Berichte ergebiger; zudem muss man dort oft in Fußnoten und Anlagen wühlen, um an diese wichtigen Angaben zu kommen.

Unverzichtbar ist es auch, dass die Studien so angelegt sind, dass **alternative Faktoren**, **unerwünschte Nebenwirkungen** und **personenspezifische Effekte** erkannt und ggfs. ausgeschlossen werden können. Hier versagt die Bildungsforschung noch immer fast auf der ganzen Linie. Es gibt kaum Studien, die hierauf irgendwelche Gedanken verschwenden. Wie viel mehr wüssten wir, wenn PISA statt eine Altersgruppe einen Schuljahrgang ausgewählt hätte, und wie viel mehr wüssten wir, wenn PISA nicht nur eine, sondern zwei oder drei Schuljahrgänge ausgewählt hätte? Dann könnten wir nämlich abschätzen, wie groß die Effizienz unseres Bildungssystems und einiger seiner Gliederungen ist.

Die Frage nach der **Effizienz von Lehrmethoden und bildungspolitischen Maßnahmen** ist nicht unanständig, wie viele zu meinen glauben. Sie dient nicht nur dem wirksameren Einsatz unserer Steuergelder, was durchaus legi-

tim ist, sondern auch den Schülerinnen und Schülern, die nicht länger die Schule besuchen wollen, als eigentlich nötig ist. Und sie dient den Lehrkräften, die nicht neue Methoden lernen möchten, die vielleicht ein klein wenig effektiver sind als ihre herkömmlichen, aber viel mehr Einsatz brauchen. Wenn engagierte Lehrpersonen davon berichten, dass sie über Jahre hinweg sehr viel Freizeit in ihre Lehrmethode investieren, dann sollte das ein Warnsignal sein. Vielleicht könnten sie mit effizienteren Lehrmethoden dieselben Lerngewinne ohne Selbstaussbeutung (und ohne Nachteile für ihre Familie) erreichen.

Ich kann auch hierfür meine Erfahrungen auf dem Gebiet der Moral- und Demokratiepädagogik als Beispiel anführen (Lind 2012). *Lawrence Kohlberg* und seine Kollegen hatten mit der Dilemmadiskussion eine sehr effektive Maßnahme zur Förderung von Moralkompetenz entwickelt. Meine Meta-Analyse ergab, dass ihre durchschnittliche relative Effektstärke ($r=0.40$) deutlich über den mittleren Effektstärken von anderen erfolgreichen Lehr- und Therapiemethoden ($r=.30$) liegt. Ich fand, dass die Dilemmamethode noch verbessert werden konnte. Ich habe daher die Konstanzer Methode der Dilemma-Diskussion (KMDD) entwickelt (Lind 2015, 2018). Sie ist noch effektiver und auch effizienter, weil sie klarer strukturiert und leichter lehrbar ist. Wir verstehen heute noch nicht alles über dieses Konzept, aber viel mehr als bei Beginn. Ein gutes Verständnis des Gegenstands war absolut notwendig, um ein Messverfahren zu entwickeln, das sich trotz seiner Kürze in vielen Forschungsprojekten im In- und Ausland als hoch valide erwiesen hat. Weil das Instrument kurz und dazu in vielen Altersgruppen (ab ca. zehn Jahren) eingesetzt werden kann, können komplexe Untersuchungspläne realisiert werden, mit denen die Wirksamkeit verschiedener Methoden unter unterschiedlichen Bedingungen studiert werden kann. Nur daher können wir heute mit einiger Sicherheit sagen, dass die KMDD in den Händen von Lehrkräften, die in der KMDD ausgebildet sind, viel wirksamer und auch viel effizienter ist als viele andere Methoden. Wir können auch angeben, wie groß der Ausbildungsaufwand für die Lehrpersonen ist und wie viel Zeitaufwand für den Einsatz der KMDD notwendig ist. Er ist mit ein bis zwei Doppelstunden im Jahr so gering, dass man ihn als vernachlässigbar bezeichnen kann. Für die Ausbildung der Lehrer in der KMDD werden aber ca. 120 Lernstunden benötigt.

Meine Empfehlung aus diesen Erfahrungen

Wenn die Bildungsforschung Fragen nach der Gültigkeit ihrer Messmethoden und nach Effektivität und Effizienz von Lehrmethoden und bildungspolitischen Maßnahmen beantworten kann, dann verdienen auch ihre Ergebnisse in der Bildungspolitik große Beachtung. Ja, Bildungspolitik dürfte dann gar nicht gemacht werden, wenn sie diesen Ergebnissen nicht Rechnung trägt.

Dafür muss Bildungsforschung aber besser werden und müssen vor allem die großen Geldsummen für Studien mit sehr großen Teilnehmerzahlen und immer neue Wiederholungen mit derselben schlechten Untersuchungsanlage umgeleitet werden, um mehr Forschung

- über den Lerngegenstand (z. B. Rechtschreibfähigkeit),
- in die Ausbildung der Lehrkräfte,
- in die Entwicklung besserer Messinstrumente (die Entwicklung unseres Moralische-Kompetenz-Tests kostete fast zwei Jahre intensiver Lese-, Denk- und Erprobungsarbeit) und
- in bessere Untersuchungspläne zu ermöglichen.

Durch die Pläne muss sichergestellt sein, dass

- die Lehrkräfte in der untersuchten Lehrmethode ausreichend gut ausgebildet sind
- der Ausbildungsaufwand dokumentiert wird
- alternative Lehrmethoden mit untersucht werden
- die Länge der Intervention dokumentiert wird

- wichtige Kategorien von Lernenden (langsam/schnell, mit und ohne Hochsprachhintergrund in der Familie, geringer/hoher Lernstand bei Beginn der Untersuchung, niedrige/hohe Unterstützung durch Elternhaus usw.) in der Studie mit ausreichender Fallzahl repräsentiert sind,

um einige wichtige Faktoren zu nennen, die in Wirkungsstudien eingeschlossen werden sollten.

Literatur

Lind, G.: Ist Moral lehrbar? Berlin 2012.

Lind, G.: Moral ist lehrbar. Berlin 2015.

Lind, G.: Diskussions-Theater – eine Methode der demokratischen Bildung. In: Lehren & Lernen 44 (2018), H. 10, S. 24-33.

Prof. i. R. Dr. Georg Lind
Universität Konstanz
georg.lind@uni-konstanz.de