

Ines Langemeyer

Evidenzbasiertes Forschen oder das Problem des Beweisens

[Ergänzendes Material zur Videoreihe „Forschen Lernen“, siehe: lehr-lernforschung.org]

Aus dem Englischen kommt die Begrifflichkeit *evidence*, aus welcher die Leitidee einer *evidence-based* Forschung abgeleitet wird. Gemeint ist, kurz gesagt, eine hypothesenprüfende Forschung. Nach dem Vorbild der Medizin werden im engeren Sinne damit Experimentalstudien mit einer großen randomisierten Stichprobe und einem Interventions-/Kontrollgruppendesign bezeichnet. Als Anglizismus ist daraus im Deutschen „Evidenz“ und „evidenzbasierte Forschung“ geworden. Diese Übersetzungen scheinen auf der Hand zu liegen, sind aber problematisch.

Der Philosoph Hans Jörg Sandkühler stellt klar, dass es zwischen dem deutschen Begriff „Evidenz“ und dem Englischen „evidence“ einen Unterschied gibt. Und beides bezeichnet den gemeinten Sachverhalt nur schlecht. Denn Ziel der hypothesenprüfenden Forschung ist ein „forschungsmethodisch erzeugtes empirisches Wissen“ (Sandkühler 2011, 33). Das Augenmerk liegt also auf dem Methodischen, wie dieses Wissen erzeugt wird. Sieglinde Jornitz gibt zu bedenken, dass im Deutschen Evidenz „eine Offenbarung ohne weitere Legitimation: sich selbst evident sein“ bedeutet (Jornitz 2008, S. 207). Hingegen könne im Englischen „evidence“ [...] jedes ‚Mittel der Bestätigung und Rechtfertigung einer Annahme‘ sein, auch alles, ‚was Grundlage einer Meinung‘ [...] ist“ (ebd.). Damit ist aber weder im Englischen noch im Deutschen klar ausgedrückt, dass man eine bestimmte hypothesenprüfende Art der Wissenserzeugung meint.

Die Bedeutung von „evidence“ ist zu schwach, um präzise auf „forschungsmethodisch erzeugtes empirisches Wissen“ zu verweisen. Noch problematischer ist aber die Verwendung des deutschen Ausdrucks „Evidenz“, d.h. der Verweis auf einen Vorgang der Selbstoffenbarung.

„Evidenz soll jenes Wissen auszeichnen, dessen Gründe stärker sind als die Gründe, die eine bloße Meinung für sich in Anspruch nehmen kann (vgl. Chisholm 1979); als stärker gelten die Gründe dann, wenn sie von ‚der Realität‘ verbürgt sind und so ausgesagt werden, dass sie eine *von der Subjektivität unabhängige Geltung* haben. Evidenz-Ansprüche sind Evidenz-Anrufungen: Man beruft sich auf das ‚Offensichtliche.‘ Evidentes soll die Garantie seiner Wahrheit außer sich haben und deshalb als fester Grund der Zurückweisung skeptischer Einwände gelten können.“ (Sandkühler 2011, S. 35)

Und damit kommt man zu dem problematischsten Punkt: „evidence soll die Voraussetzung eines Wissens sein, das, weil es empirisches Wissen ist, immer schon gerechtfertigt ist.“ (Ebd.)

Um den starken Bezug zu „forschungsmethodisch erzeugtem empirischen Wissen“ deutlich zu machen, lässt sich der Begriff des Beweises bzw. der Beweisführung wieder aufgreifen. Dies ist mit Blick auf die Verwendung dieses Begriffs in der Mathematik nicht unumstritten, wird er doch hier nicht in einem Bezug zur Empirie, sondern zur Logik verwendet.

Historisch veränderte sich jedoch die Verwendung der Begriffe „Beweis“ und „Beweisführung“ mit Newton. Newton verknüpfte das Experiment mit der exakten Messung und so mit der Mathematisierung von Phänomenen. Dieses Vorgehen ist nicht nur für die Natur-, sondern auch für die Sozialwissen-

schaften zum Vorbild geworden, obwohl es in diesen Disziplinen nicht die mechanischen Kausalitätsbeziehungen gibt, die in der newtonschen Physik gelten.

Viele wissenschaftliche Disziplinen würden zudem im Anschluss an den Falsifikationismus Karl Poppers unterstreichen, dass ein *absoluter* Beweis unmöglich ist. Dies ist wichtig. Dennoch muss der Begriff des Beweises nicht aus der natur- und sozialwissenschaftlichen Forschung ad acta gelegt werden. Beweise sind hier eher ein regulatives Ideal, das anzuerkennen notwendig ist, weil die Wissenschaft darum ringt, wahre Aussagen über die Wirklichkeit zu finden. Denn für die Erklärung von Phänomenen in der Wirklichkeit gibt es selten nur eine einzige Theorie, sondern häufige mehrere, die miteinander konkurrieren. Deshalb muss eine Beweisführung in jedem Fall ein regulatives Ideal von Forschung bleiben.

Wie geht man aber mit dem Beweisen als einem regulativen Ideal um?

Im Folgenden werden zwei Beispiele aus der Sozialforschung diskutiert, die veranschaulichen, dass die fehlende Beweislast von Studien dann deutlich wird, wenn man anfängt, ihre Daten theoretisch zu interpretieren und sie für die Untermauerung einer Theorie zu belasten.

Das erste Beispiel ist eine Untersuchung zu Leistungsunterschieden von Schülern bezüglich ihrer Lesekompetenz. Die Analyse stammt von Monika Jungbauer-Gans, die einen Ländervergleich zwischen Deutschland, Frankreich und der Schweiz vornimmt. In dieser Studie werden folgende Datenauswertungen mittels logistischer Regressionsanalyse präsentiert (Jungbauer-Gans 2004, S. 389).

Zunächst erscheint diese Analyse mit einem Ländervergleich eine nüchterne, rein statistische Betrachtung zu sein, eine, die sich rein auf konkret Beobachtbares, eindeutig Registrierbares und Zählbares bezieht, ohne Deutungen zu enthalten. Durch logistische Regressionsrechnungen kann man hier in der Tabelle sehen, wie wahrscheinlich es ist, dass einzelne (dichotome) Merkmale zusammen mit dem (intervallskalierten) Merkmal „Lesekompetenz“ auftreten. Positive Werte scheinen dabei einen positiven Zusammenhang mit Lesekompetenz aufzuzeigen, negative Werte negativ. Man könnte also das erste intuitive Verständnis bzw. die erste Evidenz so in Worte fassen: Hohe Werte indizieren einen starken, niedrige einen schwachen Einfluss.

Die Ergebnisse scheinen selbst-evident. Aber sie sind es nicht. Ich versuche einmal ein paar Aspekte aufzuzeigen: Was bedeutet es z.B., dass in Deutschland das Abitur der Mutter einen stärkeren positiven Einfluss auf die Lesekompetenz der Schüler zu haben scheint als in der Schweiz oder in Frankreich? Verbringen die Mütter hier weniger Zeit mit ihren Kindern? Oder wird Lesekompetenz in den dortigen Schulen besser gefördert? Oder ist vielleicht das Niveau des Abiturs in den Ländern gar nicht vergleichbar? All das wissen wir nicht. Zieht man aus der Statistik die Varianzaufklärung hinzu, die hier in Prozent angegeben ist, so erfährt man, dass die Regressionsanalyse, die auf die Daten angewendet wurde, bei der Schweiz etwas besser zu den Daten passt als in den beiden anderen Ländern. Die angegebenen Einflussstärken sind gewissermaßen nur Bruttoeffekte. Die verschiedenen Einflussfaktoren, die man in der Studie berücksichtigt hat, sind noch nicht gegeneinander getestet worden. Und vor allem: Man muss die Lebensverhältnisse in den Ländern kennen, damit man den Ergebnissen überhaupt irgendeinen Sinn zuweisen kann. Sinnbezüge sind sprachliche Phänomene, die häufig nur eine begriffslogische oder alltagslogische Evidenz besitzen. Man muss auf

eine konkrete Realität rekurren, um eine gewisse Plausibilität des Ergebnisses entstehen zu lassen, etwa: „Ach ja, weil Mütter mit Abi ihren Kindern im Alltag wahrscheinlich mehr vorlesen oder stärker aufs Lesen achten, deshalb sind diese Kinder wohl auch in den Leistungstests besser.“ Aber diesen Zusammenhang hat man mit dem Item „Mutter - Abi“ nicht empirisch erfasst und schon gar nicht überprüft. Es könnte theoretisch auch sein, dass das Ergebnis eine Vererbung von Intelligenz bestätigt. Dann würde man von einem weitestgehend biologischen und nicht von einem sozialen bzw. pädagogischen Zusammenhang ausgehen.

Man kann bei der Studie von Jungbauer Gans sehen, dass sie als eine Untersuchung zum Einfluss des „sozialen und kulturellen Kapitals“ auf Lesekompetenz angelegt ist. Was zunächst so erscheint, als wäre es fast schon auf den ersten Blick verständlich, wirft plötzlich eine Unmenge an Fragen auf. Warum wird Kapital bedeutsam? Das ist doch eine ökonomische Kategorie. Was ist daran sozial oder kulturell? Ich gehe im Folgenden nur auf das kulturelle Kapital ein, da es mir nur um ein Beispiel für Sprach- bzw. Theoriefragen in Forschungsprozessen geht.

Man muss bei dem gewählten theoretischen Rahmen erst einmal wissen, dass der Begriff „Kapital“ nicht wortwörtlich, sondern eher metaphorisch bzw. als Analogie verwendet wird. Mit dem Begriff kulturelles Kapital wird auf akademische Titel, die Teilhabe an kulturellen Gütern, wie Museums-, Theater-, Opernbesuche, auf die Beschäftigung mit Literatur und Wissenschaft rekuriert. Der Gedanke ist, dass Menschen so etwas wie kulturelles Kapital anhäufen, auch wenn es dafür kein Konto und keine Bank gibt.

Genau deshalb sieht man in der Tabelle einige Rubriken, die die Wissenschaftlerin mit Absicht gewählt hat. Sie bezieht sich in ihrem gedanklichen

Handeln dadurch im Weiteren auch auf die Theorie Pierre Bourdieus. Bourdieu war ein französischer Soziologe.

Der Begriff „Kapital“ hat bei Bourdieu nicht direkt die Bedeutung eines Zahlungsmittels, sondern eher die eines Investitionsgutes. Man investiert Zeit und Geld in den Erwerb von Bildungstiteln oder in die Besuche von kulturellen Veranstaltungen oder in den Besuch von Bildungsinstitutionen. Man kann nicht direkt damit „bezahlen“, wenn man einen Beruf erlangen möchte. Aber der Bildungstitel ist für den Zugang zu bestimmten Positionen bedeutsam. Nach Bourdieu lässt sich also das kulturelle Kapital durch das Erreichen besser bezahlter und einflussreicherer Positionen in der Gesellschaft in ökonomisches „konvertieren“, auch wenn es keine Wechselkurse gibt, die wie bei Devisen unabhängig von der Person gelten würden.

Soviel zu den eigentümlichen Sinnbezügen der bourdieuschen Theorie. Vielleicht hat man als Leser nun ein gewisses Unbehagen: Muss man an diese Kapital-Metaphern glauben, um das zu sehen, was die Soziologen sehen? Ja, in der Tat. Das wirkt zuweilen eher wie eine Glaubens- und nicht wie eine wissenschaftliche Gemeinschaft.

Wie zumindest zu sehen ist, hat Empirie in jedem Moment einen theoretischen Bezug und braucht ihn auch. Sonst weiß man nicht, was die Forschung eigentlich besagt oder besagen soll.

Gehen wir noch einmal zurück zu der international vergleichenden Studie zu Lesekompetenz.

Die entscheidende Frage ist, wenn ich nicht einfach nur in eine Glaubensgemeinschaft aufgenommen werden möchte: Kann ich mit so einem Begriff

wie dem kulturellen Kapital, der sprachlich betrachtet bloß als Metapher gebildet wurde, eigentlich irgendetwas in Bezug auf Lesekompetenz erklären?

Bourdieu verwendete den Begriff „kulturelles Kapital“, um in der Soziologie zu erklären, dass sich Eliten in der Gesellschaft nicht nur über ökonomisches Kapital reproduzieren, sondern auch die Bildungsinstitutionen dabei eine Rolle spielen. Anders gesagt, spielt sich nach Bourdieu bei der Wiederherstellung von sozialer Ungleichheit in modernen Gesellschaften, von Arm und Reich, oben und unten, nicht nur ein rein ökonomisches Muster ab. Man müsse nicht nur den Bereich der Wirtschaft studieren; ein wichtiger Teil von Machtmechanismen würde auch im Bereich des Bildungssystems liegen, obwohl dieses Feld selten bis gar nicht nach ökonomischen Regeln funktioniert. Wer sich hierzulande z.B. einen Dokortitel kaufen und nicht erarbeiten will, verstößt gegen Regeln und Normen.

Nun hat aber die PISA-Studie, deren Rohdaten hier verwendet wurden, ein ganz anderes Problem. Sie versucht über die Messung von Schülerleistungen im großen Maßstab zu erklären, wie leistungsfähig die jeweiligen Bildungssysteme sind. Sie liefert für den Wettbewerb zwischen Ländern Daten, die zu Aussagen führen sollen, welche Länder besser sind.

Wissenschaft muss sich deshalb immer damit beschäftigen, in welcher Weise es überhaupt sinnvoll ist, solche Untersuchungen anzulegen. Und inwiefern, sind Forschungen aus einem Bereich übertragbar auf einen anderen. Schauen Sie einmal selbst – wir sind damit zugleich mit Sprachfragen konfrontiert:

Man kann zwar sagen, dass nach Bourdieus Begriffsverständnis das „kulturelle Kapital“ unter anderem den Bücherbesitz umfasst. Er funktioniert anscheinend wie eine Oberkategorie, so wie Obst Äpfel und Birnen usw.

umfasst. Während man zwar sagen kann, dass Äpfel Obst sind, so kann man doch nicht umgekehrt sagen, dass der Bücherbesitz bereits „kulturelles Kapital“ *ist*. Man merkt, wie nur die Sprache Besonderheiten deutlich machen kann: Angemessen ist der Begriff „kulturelles Kapital“ nur verwendet, wenn auch tatsächlich Literatur gelesen wird, also Zeit in die Bildung durch literarisch oder wissenschaftlich wertvolle Bücher investiert wird. Telefon-, Koch-, Steuerberaterbücher würden eher nicht darunter zählen. Bei der Empirie dieser Studie von Jungbauer-Gans ist das übrigens gar nicht ganz klar. Man unterstellt bei dem erfassten Merkmal „Bücherbesitz“ ganz grob, dass auch bildende Bücher gelesen werden, wenn davon viele im Haushalt vorhanden sind. Ist diese Ungenauigkeit wissenschaftlich eigentlich zulässig?

Des Weiteren ist zu fragen: Ist es überhaupt sinnvoll, ein soziologisches Konstrukt wie „kulturelles Kapital“ auf das pädagogisch-psychologische Konstrukt „Lesekompetenz“ zu beziehen? Konkretisiert meint doch auch letzteres nichts anderes, als dass durch ein Testverfahren beobachtet wurde, dass jemand Texte lesen (also beim Textlesen ein angemessenes Sinnverstehen herstellen) konnte. Läuft man nicht Gefahr, dass man hier -etwas vereinfacht gesprochen- Lesen durch Lesen erklären will? Alles nur Tautologie?

Nicht ganz, weil man das Lesen einmal als Merkmal für das familiäre Milieu annimmt (d.h. hier können auch andere Familienmitglieder als das Kind selbst lesen), das andere Mal nimmt man Lesen als eine individuelle Fähigkeit (konkret auf das Kind bezogen). Trotzdem: Da scheint's ein G'schmäcke zu geben.

Was sich hiermit exemplarisch zeigen lässt: Man kommt möglichen Problemen in der Forschung nur auf die Schliche, wenn man mit der eigenen

Theorie ernst machen (Lewin), und das heißt, wenn man mit der Sprache, mit begrifflichem Denken, bewusst umgehen kann. Sprache ist gedankliches Handeln. Man braucht eine hohe Sensibilität für Sinnzusammenhänge, um Widersprüche und Ungenauigkeiten erkennen zu können.

Dies gilt im Übrigen nicht nur für die Geistes- und Sozialwissenschaften, weil sie zweifellos stärker textbasierte Fächer sind. Auch die Mathematik betrifft es: Wenn sie z.B. Zahlenräume voneinander abgrenzt, weil z.B. Rechenoperationen, die in der Menge der Rationalen Zahlen gemacht werden, in der Menge der Natürlichen Zahlen unlogisch sind, dann geht es auch hier um ein begriffliches Denken. Der jeweiligen Zahlenraum gibt die Logik vor, in der die einzelne konkrete Zahl steht und davon abhängig sind nur bestimmte Rechenoperationen zulässig. Lernende müssen den Bezug zu den unterschiedlichen Zahlbegriffen richtig herstellen können. Auch die Statistik hat ihre Begriffe wie Wahrscheinlichkeit. Deshalb muss man z.B. vorsichtig sein, die negativen Werte bei der Hausaufgabenhilfe in Bezug auf Lesekompetenz so auszulegen, sie würde den Lesefähigkeiten der Schüler schaden. Wahrscheinlichkeitsangaben sind keine Kausalerklärungen.

Ein weiteres Beispiel: Der Bildungssoziologe Rolf Becker zeigt ebenfalls mit den Daten aus einer PISA-Erhebung durch logistische Regressionsanalysen, dass Allgemeinbildende Schulen in der Schweiz Kinder mit Migrationshintergrund eigentlich nicht institutionell diskriminieren, sondern Faktoren wie Geschlecht, der ökonomische Status und das Bildungsniveau der Eltern sowie der familiäre Sprachhintergrund eine größere Rolle spielen.

Konkret soll die Datenanalyse nachweisen, dass Kinder mit Migrationshintergrund von Lehrerinnen und Lehrern bei der Benotung durchschnittlich also nicht schlechter behandelt werden. Die Untersuchung hat also aus Sicht der Autoren gegen die These der institutionellen Diskriminierung eine Art

Gegenbeweis geliefert – was allerdings schwächer als bloßer „Hinweis“ bezeichnet wird:

„Es finden sich keine empirischen Hinweise für institutionelle Diskriminierung, dass Migranten aufgrund ihres Migranten- oder Ausländerstatus als ‚Problemfälle‘ im Bildungssystem etikettiert und stigmatisiert werden.“ (Becker/Jäpel/Beck 2011, S. 15)

Aber dennoch scheint es den Autoren darum zu gehen, dass sie mit der Studie die Theorie der „institutionellen Diskriminierung“ widerlegen zu können, und das hieße, den bloßen „Hinweisen“ ein sehr viel größeres theoretisches Gewicht zu verleihen.

Schaut man die Daten dieser Studie an, so berichten sie auch, dass z.B. Kinder mit Migrationshintergrund durchschnittlich schlechtere Noten haben. Und das hat, so könnte man argumentieren, sicher ganz konkrete Bedeutungen im Einzelfall für die Kinder, aber auch für die Lehrerinnen und Lehrer.

Die Frage ist, für wie gewichtig man die „Hinweise“ (bzw. die nicht zu findenden Hinweise auf eine institutionelle Diskriminierung) aus den logistischen Regressionsmodellen von Becker et al. hält. Gereichen sie zu einer echten Falsifikation? Darf man z.B. folgern, dass weitere Studien, die sich mit einem alltäglichen Rassismus in Schulen beschäftigen, gar nicht länger notwendig sind?

Dies führt wieder zu kritischen Fragen an die Methode der Studie selbst: Ist es legitim, das Diskriminierungsproblem komplett an Noten festzumachen? Das wäre sicher nicht hinreichend, um einen Gegenbeweis gegen Bourdieus Theorie zu liefern, der ja die ökonomischen Dimensionen bei der Repro-

duktion der Eliten nicht ausklammerte, sondern durch Erkenntnisse über das Feld der Bildungsinstitutionen erweiterte (Bourdieu/Passeron 1970/2018).

Wie die beiden Beispiele zeigen, ist es wichtig, immer wieder und wieder den Blickwinkel zu wechseln, von einem konsequent eingenommenen Denkstandpunkt zurückzutreten und gegen ihn die stärksten Gegenargumente zu entwickeln. Nur dann, wenn man nach einer gründlichen Abwägung ein Argument für das stärkste hält, das sich durch Gegenargumente nicht mehr entkräften lässt, kann man davon sprechen, dass es sich um wissenschaftlich geprüftes Wissen handelt. Wenn gesagt werden kann, dass Daten dieses Gewicht haben, dann entspricht dies weder einer „Selbstevidenz“ noch einer bloßen Ansammlung von „Mitteln der Begründung“, zu denen im Englischen sogar Meinungen gehören können. Es geht dann eher um das, was im Englischen „*genuine evidence*“ genannt wird, was im Deutschen weder einem bloßen Hinweis oder Nachweis gleichkommt, sondern treffender als ein *Beweisen* zu bezeichnen ist, wobei der Beweis nicht absolut erbracht werden kann, sondern *im Sinne eines regulativen Ideals* existiert.

Denn das „mit Überzeugungen bzw. Evidenzen und ihrer Rechtfertigung verknüpfte epistemologische Problem besteht nicht darin, dass das Erkennen ‚eigentlich‘ darauf gerichtet wäre, die *Realität* als solche zu repräsentieren. [...] Überzeugungen hängen nicht davon ab, wie es sich in der Realität nun einmal verhält; sie sind Einstellungen *innerhalb* der phänomenalen Wirklichkeit zu dem, was nach menschlichem Maß für Realität gehalten wird. Überzeugungen können falsch sein, wenn sie sich aufgrund von Sinnestäuschungen, falscher Verallgemeinerung einzelner Erfahrung, semantischen Fehlern, Argumentationsfehlern und Fehlschlüssen oder ideologischen Zielen bilden.“ (Sandkühler 2011, S. 53)

Literatur:

Becker, R., Jäpel, F., & Beck, M. (2013). Diskriminierung durch Lehrpersonen oder herkunftsbedingte Nachteile von Migrant*innen im Deutschschweizer Schulsystem? *Schweizerische Zeitschrift für Soziologie / Swiss journal of sociology / Revue suisse de sociologie*, 39(3), 517-549.

Bourdieu, P., & Passeron, J. C. (2018). *La reproduction: éléments pour une théorie du système d'enseignement*. Minuit.

Jornitz, S. (2008): Was bedeutet eigentlich "evidenzbasierte Bildungsforschung"?, *Die Deutsche Schule*, 100/2, S. 206-216.

Jungbauer-Gans, M. (2004). Einfluss des sozialen und kulturellen Kapitals auf die Lesekompetenz. Ein Vergleich der PISA 2000-Daten aus Deutschland, Frankreich und der Schweiz. *Zeitschrift für Soziologie*, 33(5), 375-397.

Sandkühler, H. J. (2011). Kritik der Evidenz. In Bellmann, J. (Hrsg.). *Wissen, was wirkt* (S. 33-55). Springer: VS Verlag für Sozialwissenschaften.