

Empirische Gewissheit gibt es nicht

Lernprozesse zu vermessen, ist völlig aussichtslos. Doch es spricht auch wenig dafür, dass sich Lernergebnisse genau beziffern lassen.

Von Wolfram Meyerhöfer

Anfang Oktober werden Bildungsforscher den Kultusministern erklären, warum es weniger wichtig ist, sich mit dem Lernen zu beschäftigen, als Geld in die Vermessung der Ergebnisse des Lernens zu stecken. Zum Jahreskolloquium des DFG-Schwerpunktprogramms „Kompetenzmodelle zur Erfassung individueller Lernergebnisse und zur Bilanzierung von Bildungsprozessen“ werden die Bildungsentscheider nach Frankfurt am Main geladen. Die Bildungspolitik soll auf diese Weise dazu gebracht werden, noch mehr Geld aus der Verbesserung von Schule abzuziehen und stattdessen in das Testen von Schülern zu stecken. Kompetenzmodelle sind in der Empirischen Bildungsforschung immer Kompetenzstufenmodelle. Das liegt daran, dass das Ziel dieser Disziplin nicht ist, das Lernen zu verstehen und zu verbessern. Man müsste sich dann mit Lernen und Vergessen beschäftigen, mit Interesse und Langweile, mit Nutzen und Nutzlosigkeit, mit der Motivation und der Demotivation durch Schulnoten. Die Empirische Bildungsforschung will das Gelernte vermessen.

Der Geist der Kinder soll also auf einer Zahlenskala abgebildet werden. Man möchte sagen können: Du, Kind A, hast 503 Punkte und bist somit ein schlechterer Mathematikschüler als Kind B mit 506 Punkten. Die Menschenmesser folgen also dem Bild, dass das Menschliche auf Skalen gepresst werden muss und dass dies gesellschaftlich wertvoll ist, so dass Kultusministerien dafür Millionen ausgeben. Sie folgen aber offenbar auch einer gewissen Lust am Wahn, denn nur in solchen Momenten wird man ernsthaft glauben können, dass man Kindern Schulnoten mit der Genauigkeit einer Hundertstel Stelle nach dem Komma zuweisen kann – nichts anderes ist die Behauptung, dass 503 Punkte etwas anderes seien als 506 Punkte.

Dass Kompetenzstufenmodelle der neueste Schrei des Bildungssektors sind, merkt der Student in der Vorlesung, der Referendar im Studienseminar, der Lehrer anhand der Fortbildungskataloge. Dort, wo früher pädagogische Bekenntnisse abgefragt wurden, muss man heute in Prüfungen Kompetenzstufenraster aufsagen, die genauso leer sind. Man mag diesen „Wachwechsel“ als Legitimationstrick des pädagogischen Apparats mit dem Mittel der Scheininnovation abtun. Wichtiger aber ist es zu verstehen, warum es nun gerade Kompetenzstufenmodelle sind, in die Kräfte des Bildungssystems gesteckt werden.

Bei Kompetenzstufenmodellen handelt es sich um Versuche, die Resultate der Empirischen Bildungsforschung als relevant erscheinen zu lassen: Wenn deutsche Schüler bei einem internationalen Schultest 63 Prozent der Aufgaben korrekt ankreuzen, finnische Schüler aber 66 Prozent, dann sieht man, dass der Unterschied gering ist. Deshalb werden diese Tests mit Skalen versehen, die kleine Unterschiede groß erscheinen lassen. Deutschland hat dann vielleicht 495 Punkte (leider nur „Mittelmaß“), Finnland hingegen 533 Punkte („Spitzengruppe“).

Diese Punkte erzählen leider gar nichts. Zudem kann man die Skala beliebig strecken oder stauchen, und wenige statistische Entscheidungen – von denen man sehr viele treffen muss – lassen die Skala und auch die Länderreihenfolge kräftig durcheinanderpurzeln. Trotzdem wird bei Tests nie gesagt, wie viele Aufgaben so ein finnischer Schüler denn nun mehr korrekt angekreuzt hat als ein deutscher Schüler. Denn dann würde sichtbar, dass die Tests kaum eine Aussage über Bildungssysteme zulassen.

Der Ausweg aus dem Dilemma sind nun die Kompetenzstufenmodelle. Sie sollen „die Skala zum Sprechen bringen“, sie sollen die Illusion erwecken, dass man aus Testpunktwerten inhaltliche Aussagen ableiten kann. Dazu macht man eine Liste, auf der die Aufgaben nach Lösungshäufigkeit geordnet sind. Diese Liste unterteilt man nun in gleich große Abschnitte. Wenn die leichteste Aufgabe von 85 Prozent der Schüler gelöst wird, die schwerste aber nur von 10 Prozent der Schüler, dann kann man 5 Stufen festlegen, die jeweils 15 Prozentpunkte „breit“ sind. Nun kann man sich die Aufgaben anschauen, die jeweils auf den so entstandenen „Kompetenzstufen“ liegen. Die Idee ist, dass man beschreiben kann, welche Fähigkeiten mit den Aufgaben auf den jeweiligen Stufen gemessen werden. Hinterher tut man dann so, als ob diese Kompetenzstufen irgendetwas über den Lernprozess sagen könnten.

Dieses Vorgehen wirkt technisch raffiniert und löst bei manchem Begeisterung aus, weil der statistische Apparat suggeriert, dass hier Wissenschaft am Werke sei und dass das Pädagogische endlich einmal harte Zahlen liefert. Das Problem dieses Ansatzes ist, dass er nicht funktioniert. Schaut man auf das statistische Procedere, so leuchtet unmittelbar ein, dass man auf diese Weise keinerlei ernsthafte Aussage über Bildungsprozesse oder Bildungssysteme generieren kann. Das funktioniert aber nicht einmal für die Deutung von Testresultaten. Wenn ein Kochlehrling A einen exzellenten Braten machen kann, ein anderer Lehrling B hingegen eine wunderbare Quiche, dann kann die Empirische Bildungsforschung nicht einfach das Essen genießen. Die Lehrlinge müssen unbedingt in eine Kompetenzreihenfolge gebracht werden, das ist das zentrale Ziel der Kompetenzmodelle. Dazu lässt man viele Lehrlinge Braten und Quiche herstellen. Wenn viele Lehrlinge einen guten Braten hinbekommen und wenige Lehrlinge eine gute Quiche, gilt der Quiche-Meister B als der bessere Koch. Dieses Ergebnis ist empirisch fundiert und nachgerade objektiv. Wenn Sie also Braten-Fan sind und im Exzellenz-Lokal der Empirischen Bildungsforscher essen, dann zeigt Ihr Unwille gegenüber dem schlechten Braten dem Forscher lediglich, dass der Empirie Ihres Gaumens keinerlei Signifikanz zukommt.

Was beim Koch auf den ersten Blick einleuchtet, bedarf in der Mathematik eines zweiten Blickes. Im Kasten sehen Sie eine Aufgabe, die das Pisa-Konsortium als beispielhaft bezeichnet. Die Seite $A-B$ des Dreiecks $A-B-C$ ist 6 cm lang. Es werden die Mittelpunkte E und F der Seiten eingezeichnet. Wie lang ist $E-F$? Man kann die Aufgabe auf verschiedene Weise lösen, unter anderem durch Messen, durch Intuition, mit Hilfe von Strahlensätzen oder mit einer Formel, die (zu Recht) kaum eine Rolle im Mathematikunterricht spielt. Die Formel sagt, dass die sogenannte „Mittellinie des Dreiecks“ (EF) halb so lang ist wie die „Grundlinie des Dreiecks“ (AB).

Wenn ein Schüler die Aufgabe löst oder nicht löst, wissen wir nicht, wie er die Aufgabe gelöst hat oder wie er gescheitert ist. Wir wissen also nicht, was wir mit der Aufgabe eigentlich messen. Nur sehr triviale Aufgaben „messen scharf“, sind also auf genau eine Weise lösbar. Fast alle Aufgaben „messen unscharf“, sind also auf verschiedenen Wegen lösbar. Wir wissen also nicht genau, welche Fähigkeiten wir gemessen haben – außer der Fähigkeit, solche Aufgaben zu lösen, also sein Kreuz an die gewünschte Stelle zu setzen, die „richtige“ Zahl hinzuschreiben oder auf eine Weise offene Antworten aufzuschreiben, die in

den Kategorienkatalog der Tester passt. Inwiefern das alles „mathematische Leistungsfähigkeit“, „Lesefähigkeit“ sein soll, bleibt hingegen unklar.

Diese Unschärfe macht es völlig unmöglich, ein Kompetenzstufenmodell aufzustellen: Wenn eine Aufgabe sehr viele verschiedene Fähigkeiten gleichzeitig misst, dann kann man nicht benennen, welche Fähigkeiten ein Schüler hat, der die Aufgabe löst. Alle Kompetenzstufenmodelle fußen aber darauf, dass man genau benennen kann: Ein Schüler, der diese Aufgabe löst, der kann das und das. Wenn das nicht benennbar ist, dann kann man die Modelle nicht inhaltlich füllen. Wenn ein Kompetenzstufenmodell mit Aufgaben entworfen wird, die mehrere Lösungswege zulassen, lässt es sich nicht empirisch absichern. Wenn es aber mit Aufgaben erstellt wird, die nur einen Lösungsweg zulassen, dann ist es didaktisch kontraproduktiv, weil solche Aufgaben in die geistige Verarmung führen. Eine gute Testaufgabe – bei der man genau benennen kann, was sie misst – ist das Gegenteil einer bildenden Schulaufgabe, welche geistige Vielfalt und Debatte über diese Vielfalt herausfordert. Deshalb führt der Weg der immer stärkeren Standardisierung von Bildung – getragen von einem nicht hinterfragten Glauben an die mathematische Modellierbarkeit des menschlichen Geistes – das Schulsystem in eine Kultur der intellektuellen Armut.

Der Autor lehrt Mathematikdidaktik an der Universität Paderborn.