

Die Illusion der Statistiker

Seit einem Jahrzehnt vermessen die Pisa-Studien den Lernerfolg von Schülerinnen und Schülern. Die statistische Methode beansprucht Exaktheit, entzieht sich aber dem wissenschaftlichen Diskurs. Entstanden ist eine selbstreferenzielle Testindustrie mit Hunderten Millionen Franken Umsatz. Nationale Bildungstraditionen werden zerstört, weil die Statistik alles über den Kamm der Messbarkeit scheren muss, *schreibt Thomas Jahnke*

Als vor zehn Jahren die ersten Ergebnisse der Pisa-Studie (Programme for International Student Assessment) publik wurden, war man über das unerwartete Mittelmass der Leistungen der Schülerinnen und Schüler in der Schweiz, Österreich und Deutschland so bestürzt, dass selbst die einfachsten Fragen ungestellt blieben: Was wurde und wird hier in wessen Auftrag von wem zu welchem Zwecke getestet?

Auftraggeber ist die OECD, die 34 Länder mit dem vorrangigen Ziel des «nachhaltigen Wirtschaftswachstums» vereinigt. Auftragnehmer sind fünf sogenannte transnationale Bildungsdienstleister – bei vier davon handelt es sich um private Unternehmen, die Pisa entwickelt und an 67 Staaten (Pisa 2012) verkauft haben. Dass diese Firmen wesentlich an ihrem eigenen Profit interessiert sind, kann man ihnen kaum vorwerfen – auch nicht die Wahrung von Betriebsgeheimnissen, obwohl das sicher Auswirkungen auf die von ihnen angebotenen Dienstleistungen und Produkte hat. Das Auftragsvolumen für einen internationalen Pisa-Durchgang liegt im dreistelligen Millionenbereich. Die Markt-

erweiterung der Testindustrie auf die Schweiz, Österreich und Deutschland ist mit der Periodisierung der Pisa-Durchgänge im 3-Jahre-Rhythmus auch dauerhaft – zumindest bis 2015 – gelungen.

Die privatwirtschaftliche Durchführung von Pisa entzieht dieses Programm weitgehend der wissenschaftlichen Diskussion und gibt seinen Betreibern eine Gestaltungs- und Deutungshoheit, die sich einem demokratischen und auch einem nationalen Diskurs entzieht.

Getestet werden bei Pisa die «Lesekompetenz», die mathematische und die naturwissenschaftliche «Kompetenz» der Fünfzehnjährigen. Der musisch-ästhetische Bereich oder Fremdsprachenkenntnisse bleiben unberücksichtigt. In Mathematik zum Beispiel wurde und wird «mathematische Grundbildung» getestet, wobei offen zugegeben wird, dass man sich dabei um Lehrpläne nicht schert: «Der Begriff Grundbildung wurde gewählt, um zu betonen, dass mathematische Kenntnisse und Fähigkeiten, wie sie im traditionellen Curriculum der Schulmathematik definiert werden, im Rahmen von Pisa nicht im Vordergrund stehen», erläutern Pisa-Verantwortliche. Wie, könnte man fragen, kann man die Leistungen von Schüle-

rinnen und Schülern an Zielen messen, die sie gar nicht verfolgen?

«Pisa bringt eine andere Lernkultur zur globalen Vorherrschaft, als es der Tradition von klassischer Bildung und Vermittlung von Fachwissen entspricht, ohne dass man (. . .) die generelle Überlegenheit der neuen Lernkultur behaupten könnte», sagt der Bamberger Soziologe Richard Münch. Der Erziehungswissenschaftler Volker Ladenthin sieht hierin eine grundsätzliche Funktionalisierung von Bildung: «Es geht erklärermassen der OECD (als Wirtschaftsorganisation) um die Durchsetzung eines zwar als angelsächsisch bezeichneten, in Wirklichkeit aber längst auch bei uns bekannten, schlicht funktionalen Bildungsbegriffs; es geht um Universalisierung (also ökonomische Globalisierung), das heisst die Auflösung föderaler und letztlich auch nationaler Bildungssysteme; es geht um die Transformation von Selbstbestimmung und Kulturidentität in funktionale Basiskompetenzen; es geht um Normierung und Vereinheitlichung statt um Individualisierung und Differenzierung. Als Grund und Legitimation für diesen Wandel werden ausdrücklich und ausschliesslich die von der ökonomischen Situation her verursachten und definierten Qualitätsanforderungen be-

nannt.»

Als die ersten Pisa-Ergebnisse publik wurden, war man auch bildungspolitisch erleichtert. Das Unbehagen über den Stand der Schulbildung wurde dem hermeneutischen Klagen entrisen; das Menetekel ist digitalisiert. Es hat Dezimale bekommen, die den Ergebnissen des Programms den Glanz und die Sicherheit naturwissenschaftlicher Erkenntnis verleihen und zugleich eine Lösung der – nicht mehr nur gefühlten, sondern auch gemessenen – Schulbildungsmisere verheissen: Werte, die zu tief sind, muss man erhöhen. So wird Pisa selbstreferenziell und unentbehrlich, gleich wie der nächste Durchgang ausfallen mag, obwohl man weiss, dass ein Schulsystem sich in drei Jahren nicht so ändern kann, dass es wesentlich andere Schülerleistungen hervorbringt. Ob höhere Werte bei Pisa tatsächlich für eine gehobene Schulbildung sprechen, ist zudem fraglich. Wer bei Mehrfachwahlaufgaben zügig und plausibel rät, wird mit Punkten belohnt; wer über eine Aufgabe nachdenkt, ist im Nachteil. Schülerinnen und Schüler in der Schweiz, Österreich und Deutschland waren mit solchen Aufgaben bisher wenig vertraut; inzwischen ist ihre Testfähigkeit vermutlich gestiegen: Vom Wiegen wird die Sau nicht fetter, aber sie lernt, sich schwer zu machen.

Zumindest sollte man bei der Interpretation der Pisa-Ergebnisse nicht mehr von guten und schlechten Schülerinnen und Schülern oder Schulen oder Regionen oder gar Nationen sprechen, sondern von Pisa-guten und Pisa-schlechten.

Der Genauigkeitsanspruch von Pisa wird von neutralen Expertinnen und Experten angezweifelt. Der im Umgang mit grossen, schlecht konditionierten Datenmengen vertraute Physiker Joachim Wuttke stellt fest: «Allein schon unterschiedliche Schülermotivation kann, amerikanischen Studien zufolge, mit 50 oder mehr Punkten durchschlagen.» 50 Punkte entsprechen einem Lernfortschritt von fast zwei Schuljahren. Wuttke: «In Seoul wird vor der Testung die Nationalhymne gesungen; in Hamburg geben die ersten Schüler nach fünf Minuten ab. Dazu kommen Uneinheitlichkeiten bei der Stichprobenziehung, bei den Teilnahmequoten und beim Ausschluss behinderter Schüler. Allein solche Faktoren können mehr als 10 Punkte ausmachen. Und damit ist noch gar nicht die Kernfrage angeschnitten: Kann man Schülerleistungen überhaupt auf einer eindimensionalen Skala messen? In einer kultur- und sprachübergreifenden fairen Art und Weise? Man kann nicht.» Wuttke kommt zu dem Schluss, dass Pisa ein teurer Zufallszahlengenerator ist. Sicher kann man das Dümpeln der Werte etwa der Schweiz oder einzelner Kantone in den Pisa-Durchgängen mit journalistischem oder bildungspolitischem Scharfsinn wortreich kommen-

tieren, aber man sollte darüber nicht vergessen, dass die Werte jeweils von unterschiedlichen Schülern, mit unterschiedlichen Aufgaben und in Testungen mit unterschiedlichen Schwerpunkten generiert wurden und dass es sich – wenn man einmal die Spannweite üblicher Schulnoten zugrunde legt – meist um eine Interpretation im Hunderstelbereich handelt.

Zum Zweck der Testungen liest man im Vorwort des Buches Pisa 06: «Die Auftraggeber, das sind die OECD-Staaten und weitere Staaten, erwarten von dem internationalen Vergleich empirisch fundiertes Steuerungswissen. Die teilnehmenden Staaten möchten von Pisa zum Beispiel erfahren, wo sie im internationalen Vergleich stehen, welche Bildungsergebnisse andere Staaten erreichen und unter welchen Bedingungen das geschieht.» Der Erziehungswissenschaftler Klaus-Jürgen Tillmann, der Leistungsvergleichstudien und ihre politischen Folgen in vier deutschen Bundesländern detailliert untersucht hat, konstatiert ernüchternd: «Im Ergebnis zeigt sich, dass Pisa nur selten neue Programme initiiert – aber um so häufiger herangezogen wird, um die ohnehin verfolgte politische Linie zu legitimieren. Die Hoffnung, dass Pisa «Steuerungswissen» produziert, lässt sich kaum einhalten.» Auch der in Deutschland für den letzten Pisa-Durchgang 2009 verantwortliche Leiter distanziert sich inzwischen öffentlich von dem Begriff Steuerungswissen und bezeichnet ihn als Illusion. Dennoch arbeiten die Pisa-Unternehmen mit den Bildungspolitikern zu gegenseitigem Nutzen Hand in Hand. Letztere können nun – früher häufig als untätig und reformunwillig gescholten – im Takt von Wahlperioden neue Testungen und Vergleichsuntersuchungen einführen und deren Resultate in ihrem Sinn interpretieren und sich so als energisch und aktiv profilieren.

Ein Blick über den Atlantik sollte sie Besseres lehren. Die American Evaluation Association (AEA) warnt nach mehr als zwei Dekaden Test-erfahrung 2002 eindringlich vor einer Testresultat-gesteuerten Bildungspolitik. «Obwohl solche Testungen seit mehr als zwei Jahrzehnten durchgeführt werden, haben sie weder die Qualität der Schulen verbessert noch Ungleichheiten in den Leistungen verringert, noch haben sie das Land in moralischer, sozialer oder ökonomischer Sicht vorangebracht.»

Gewinner von Pisa sind nicht die asiatischen oder finnischen Schülerinnen und Schüler, sondern die empirische Bildungsforschung, was deren Programme, Stellen und Finanzierung anlangt: «An Fördergeldern herrscht kein Mangel, neue Forschungsstätten werden gegründet. Allein der For-

schungsgegenstand selbst dämpft noch die Wissenschaftler-Euphorie: «Den Schulen bringt das leider nichts», sagt Pädagoge Tillmann. Gegen diese Testergebnisse, scheinen Bildungspolitiker zu glauben, helfe vor allem weiteres Testen. Zwar hat sich etwa die deutsche Kultusministerkonferenz als Reaktion auf den Pisa-Schock sieben Verbesserungsstrategien vorgenommen – darunter Sprachkurse für Migrantenkinder, mehr Ganztagschulen und gezielte Leseförderung –, doch konsequent umgesetzt haben sie bisher nur eine einzige: Tests. «Entwicklungen gibt es zwar in allen sieben Bereichen», sagt Tillmann, «aber flächendeckend in allen Ländern sind nur die zentralen Prüfungen in den Schulen angekommen», heisst es bereits in dem Artikel «Langer Anlauf ohne Sprung» von Julia Koch im Nachrichtenmagazin «Der Spiegel» (24/2007). Geändert hat sich daran bis heute nichts. Gelder, die bisher eine – sicher vom Umfang her unzureichende – Lehrerfortbildung ermöglichten, fliessen inzwischen in Institute für Qualitätssicherung und -management.

Das Eigenartige der quantitativen empirischen Bildungsforschung ist, dass sie immer etwas hervorbringt; sie kann nicht leer ausgehen. Immer erhält sie Zahlen und damit Skalen, Intervalle, Ranglisten und, was der statistische Apparat folglich zur Verfügung stellt, Mittelwerte, Standardabweichungen, Korrelationen und so fort. Durch das Messen erhält man Daten – eigentlich sogar nur Zahlen, deren Realitätsgehalt (oder genauer Realitätsbeschreibungsvermögen) anscheinend ausser Frage steht und dem ganzen Prozess eine Art quasinaturnwissenschaftlicher Dignität verleiht. Die Frage, ob man da tatsächlich etwas misst, was man dann untersuchen kann, stellt sich gar nicht, weil der Gegenstand durch den Messprozess selbst hervorgebracht und konstituiert wird. Solche Forschung produziert ihre Begriffe (heute sagt man «Konzepte») und Ergebnisse parthenogenetisch, notfalls auch ohne jede Anleihen bei der bemessenen Realität oder Bezügen zu ihr. Eine Widerlegung der Ergebnisse von gross angelegten Vergleichsuntersuchungen, dessen sind sich die Pisaner und die involvierten Firmen sicher, ist faktisch, wenn nicht sogar prinzipiell ausgeschlossen. Die Sache ist versiegelt. Und wo Kritik einmal nicht folgenlos abprallt, freut sich die Forschungsindustrie über Folgeaufträge, die dann in der beschriebenen Art bearbeitet werden.

Die Globalisierung der empirischen Bildungsforschung formt und entleert ihre Ergebnisse in charakteristischer Weise. Weltliteratur, um ein anderes Sujet vergleichshalber heranzuziehen, spielt (häufig) in der Provinz. Nicht ein dünnes Destillat dessen, was alle oder möglichst viele vergleichbar ken-

nen oder ihnen gemein ist, macht ihren Rang aus, sondern – geradezu gegenteilig – das Eigene und Besondere, in dessen Sonderheit Welt beschrieben und erkannt wird, was dann vielen etwas sagt und folglich zu Recht als Weltliteratur bezeichnet wird.

Wenn man Bildungssysteme – schon der Begriff deutet es an – ihrer nationalen oder lokalen Traditionen und Gehalte aus Gründen der Vergleichbarkeit oder aus anderen ökonomischen oder psychometrischen Motiven entkernt, bleiben Zahlenwerke, Parameter und «Standards», die schon ausgeschlossen haben, was sie zu messen versprochen oder vorgaben. Man schert alles über den Kamm der Messbarkeit, dessen statistische Zinken nicht mehr die Welt erforschen, ihr ihre Wahrheit ablauschen, sondern eine eigene «Welt» erschaffen. Die digitale Rasterfahndung gibt dann ihre eigenen Konstrukte selbstreferenziell und zirkelhaft als Erkenntnisse aus.

Während man mit Bildungsforschung traditionell vornehmlich die Disziplin der Pädagogik assoziierte, haben inzwischen Psychologen und Psychometriker dieses Feld besetzt, was nicht nur sprachliche Auswirkungen auf das Nachdenken über die Schulbildung hat. Ökonometrie und Psychometrie liefern den Begriffsrahmen und die gedanklichen Folien. «Das meiste, was zur Durchsetzung der Optimierung des Systems durch Reformen erdacht worden ist, kommt nicht aus dem genuinen Ideenhaushalt der Pädagogik. Es baut nicht auf der abrufbaren Erfahrung mit der pädagogischen Denk- und Handlungsform auf, sondern kann als ein Import aus der angewandten Betriebswirtschaftslehre und einer «pädagogischen» Psychologie angesehen werden. Mit beidem werden fremde Universalrezepte auf die spezifischen Voraussetzungen der Schule übertragen. Die Sprache der Reform ist weitgehend die des Business: Es geht um Qualitätsentwicklung und -kontrolle, um Organisations-, Produkt- und Personalentwicklung, um Schulmarketing und Intrapreneurship», konstatiert der Frankfurter Erziehungswissenschaftler Andreas Gruschka. Verbunden mit diesem tendenziellen Zuständigkeitswechsel für Bildungsforschung von der Pädagogik zur Psychologie ist das Verblässen der Inhalte der Schulbildung, von denen in den Forschungsberichten kaum noch die Rede ist – als ginge es darum, statistische Begriffe zu erforschen und Zahlen auszubreiten und nicht Schülerinnen und Schüler und deren Bildung zu erfassen. In der Regel bleiben die Testfragen heute undiskutiert unter Verschluss; um so gewichtiger kommt die Statistik daher – übrigens mit zum

Teil unangemessenen und auch unter Experten umstrittenen Verfahren.

In der Schweiz, Österreich und Deutschland werden durch die Periodisierung von Pisa und die Vielzahl anderer Testungen die Testorientierung des Unterrichts und die Testfähigkeit und damit die erreichten Punktzahlen zwar mässig zunehmen, aber über eine bessere Schulbildung sagt das nichts aus.

Auch die Lehrerinnen und Lehrer werden in ihrer schweren Arbeit durch die Testungen nicht unterstützt, sondern unter Druck gesetzt und behindert. Die genuin Schweizer Tradition der verantwortlichen Lehrperson, die Lernprozesse individuell gestaltet, wird so konterkariert. Die Mängel der Schulsysteme in den genannten Ländern sollen nicht bestritten werden, aber durch von Tests geleitete Bildungspolitik werden sie eher verstärkt und durch neue ergänzt statt bearbeitet und möglicherweise beseitigt.