

Prof. Dr. Roland Reichenbach ist Professor für Allgemeine Erziehungswissenschaft an der Universität Zürich. Seine Interessen betreffen die Bildungsphilosophie, Pädagogische Ethik, Politische Bildung und Verhandlungs- und Einigungsprozesse.



#### ■ Roland Reichenbach

### „Was war es, was wir wissen wollten?“ 15 Jahre vergleichende Leistungsstanderhebungen im Bereich der Bildung – eine Kritik

Hans Blumenberg schreibt in *Die Lesbarkeit der Welt* (1981): „Hält man sich an Kants Katalog der letzten grossen Fragen, so bleibt es nicht vordringlich bei dem: *Was können wir wissen?* Enttäuschung gerade an dem, was sich als gekanntes Wissen herausstellt, erfordert zu fragen: *Was war es, was wir wissen wollten?* Das gibt auch der anderen Hauptfrage des Kanons: *Was dürfen wir hoffen?* eine Abwandlung, die zu fragen unausweichlich macht: *Was war es, was wir erhoffen durften?* (Blumenberg 1981, 9).

Diese Fragen seien hier auf die zurzeit grassierenden Leistungsstanderhebungen (v. a. PISA, TIMSS, IGLU, VERA) angewendet, welche bezeichnenderweise als Paradebeispiel für empirische Bildungsforschung gelten, wiewohl sie – als reine Evaluationsstudien – freilich nur einen spezifischen Fall von Bildungsforschung darstellen, der ohne grossen theoretischen Anspruch bleibt. Im Unterschied zu Evaluationsstudien hat sozialwissenschaftlich gehaltvolle Forschung *theoriegeleitet* zu sein. Theorien sind Grundlage und Gegenstand des einschlägigen wissenschaftlichen Austausches. Damit sie dies sein können, müssen sie sprachlich ausformuliert sein und eine Form aufweisen, die intersubjektiv nachvollziehbar ist. Die Bemühung um Transparenz ist nur dann einigermaßen vollständig oder zufriedenstellend, wenn die ontologischen, epistemo-

logischen und methodologischen (Vor-)Annahmen, Zugänge und Fragestellungen (sowie damit verbundene Probleme) verständlich thematisiert werden<sup>1</sup>.

## 1. Was wollten wir wissen? Leistungsstanderhebungen und Bildungssteuerung

Wir wollten wissen, wie die Leistungen der Bildungssysteme im internationalen und intranationalen Vergleich zu beurteilen sind, wie sie zu erklären sind und wie sie zu verändern, d. h. zu verbessern sind. Das erste betrifft gewünschtes (1) *Standortwissen* (zumindest relatives Standortwissen), das zweite (2) *Erklärungswissen* und das dritte (3) Veränderungs- bzw. *Steuerungswissen*.

Und haben wir dieses Wissen mit den Möglichkeiten der in Frage stehenden Leistungsstanderhebungen erhalten bzw. ernsthaft erhoffen dürfen? Wir durften sicher erhoffen, dass wenn repräsentative Gruppen von Schülerinnen und Schülern in den unterschiedlichen Nationen und Bundesländern hinsichtlich gleicher Kompetenzaufgaben getestet werden, ein Wissen zu Unterschieden und Ähnlichkeiten produziert wird, das als eine mögliche Orientierungshilfe im Sinne des Standortwissens ernst zu nehmen ist und welches einen wichtigen, jedenfalls interessanten Beitrag dazu leisten soll, was heute gerne Systemmonitoring genannt werden möchte.

Dass aber „politisch gewünschte und plausible Strukturveränderungen“ *nicht* aus beschreibenden Daten und Korrelationen des Systemmonitorings abgeleitet werden können, wie Prenzel, Baumert und Klieme (2008, 3) hervorgehoben haben, wissen freilich nicht nur einschlägige empirische Bildungsforscher. Dennoch müsse, so die gleichen Autoren, wer „steuern möchte, (...) zunächst einmal wissen, wo er oder sie sich befindet“ (Prenzel/Baumert/Klieme 2008, 1). So produziere PISA „Steuerungsinformationen“, welche die Kompetenzen von Fünfzehnjährigen in verschiedenen Staaten

---

<sup>1</sup> Ontologische Perspektiven explizieren den Gegenstandsbereich der Theorie, meist einen Aspekt der (sozialen) Wirklichkeit. Die zentralen Fragen lauten: Wovon handelt diese Theorie (wovon aber nicht?) und wie lautet die Grundfrage, mit welcher sie eine Antwort zu geben verspricht oder vorgibt? Die Frage, was die Frage ist, wird häufig unterschätzt und leicht vernachlässigt. So begegnen uns viele (Forschungs-)Arbeiten, die zwar ein Thema oder mehrere Themen haben, aber es bleibt unklar, was eigentlich die Frage ist. Epistemologische Perspektiven behandeln das Problem, welche Sorte von Wissen oder Erkenntnissen mit der vorgeschlagenen Theorie gefunden werden kann und welche Aspekte des Gegenstandsbereichs und damit verbundene Perspektiven, Wissensformen oder -typen damit ausser Acht gelassen werden (müssen). Welche Antworten können erwartet werden, wenn so gefragt wird (und nicht anders)? Auf welches Wissen muss damit verzichtet werden? Methodologische Fragen betreffen im weitesten Sinne die (möglichen und nicht möglichen) Mittel der Erkenntnisgewinnung und der damit verbundenen Probleme. Es geht dabei keineswegs allein um die Angabe der forschungsmethodischen Entscheidungen und datenanalytischen Verfahren.

betreffen. Diese „Steuerungsinformationen“ würden sich aber nicht in der relativen Standortbestimmung erschöpfen, denn zugleich erfahre man ja auch, „welches Kompetenzniveau derzeit weltweit erreicht werden kann“ (2). Daher weiss Bremen heute nicht nur: „Wir stehen auf Platz 16 von 16 in Deutschland, und dies schon länger und offenbar immer wieder“, sondern auch „theoretisch könnten wir mindestens so gut wie Bayern sein, theoretisch, d. h. wenn nur sehr vieles anders wäre“. Und Brasilien weiss: „Uns trennen viele, viele Punkte von Finnland, aber man kann ja wie die über 540 Punkte erreichen! Doch wir haben nicht einmal 400!“ Mit diesen relationalen Standortinformationen (die natürlich keine „Steuerungsinformationen“ sind, ein wahrlich unpassendes Wort, das Prenzel, Baumert und Klieme damals benutzt haben, im Grunde eine Überredungsvokabel) wissen wir nun, dass Bayern irgendwie wie Schweden oder auch Österreich ist, Bremen aber ziemlich so wie Mexiko und Brasilien! Es bleibt der Trost für das in dieser Hinsicht exotische Bremen, dass Bayern wenigstens nicht wie Finnland oder Kanada oder Korea ist. Nun kann sich Bremen an Bayern orientieren, Brasilien (Rang 38 von 38) an Finnland (Rang 1)<sup>2</sup>.

Die PISA-Ergebnisse haben mit Steuerungswissen weder im engeren noch im weiteren Sinn etwas zu tun. Sie liefern Informationen zu einer möglichen Standortbestimmung. Aus Standortwissen kann weder Erklärungswissen noch aber Steuerungswissen generiert werden. Dass man meint, diese durchwegs bekannte Tatsache dennoch erwähnen zu müssen, spricht für sich.

## 2. Was war es, was wir erhoffen durften?

Klieme, Hartig und Kenk beschreiben das Vierpunkteprogramm der einschlägigen Bildungsforschung, um die es hier geht, so: Es geht erstens um die Entwicklung und empirische Prüfung theoretischer Kompetenzmodelle. Darauf aufbauend zweitens um die Entwicklung von psychometrischen Messmodellen. Drittens um die Entwicklung von Messverfahren zur empirischen Erfassung von Kompetenzen, die sich daraus ableiten. Und schließlich um die Frage „wie die Nutzung von Diagnostik und

---

2 Ist es das, was „wir“ wissen wollten? Und wer ist dieses „Wir“? Persönlich zähle ich mich nicht dazu. Doch solches Wissen interessiert freilich dennoch sehr, die Olympiadenhaftigkeit der Ranglisten hin oder her. So schätze ich beispielsweise das Schweizerische Jahrbuch für Statistik oder die Sichtung der regelmässigen WHO Youth Surveys. Dort z. B. habe ich gelernt, dass die Jugendlichen in der Schweiz von allen untersuchten Nationen jene sind, die am meisten kiffen! Ich weiss nicht, ob dies auch als Steuerungswissen gehandelt wird. Die Schweizer Jugendlichen sind aber auch Spitze beim Zähneputzen. Ob ein Zusammenhang zwischen Kiffen und Zahnpflege besteht, und wenn ja, ob, wenn weniger Zähne geputzt würden, auch der Drogenkonsum etwas reduziert werden könnte, ist den Ergebnissen leider nicht zu entnehmen ...

Assessment zu fundierten und präzisen Entscheidungen in der pädagogischen und bildungspolitischen Praxis beiträgt“ (Klieme/Hartig/Kenk 2010, 10).

Vermögen nun also die Ergebnisse von TIMSS, PISA, IGLU und/oder VERA eine Basis für *fundierte* und *präzise* Entscheidungen in der *pädagogischen* und *bildungspolitischen* Praxis zu liefern? Kaum. Das hängt noch am wenigstens mit dem Sein-Sollen-Fehler zusammen, der mit solchen Schlüssen notwendigerweise verbunden wäre („bloß“ ein logischer Fehler), sondern vor allem mit der Komplexität des Feldes, welches untersucht bzw. befragt wird. So schreiben Prenzel, Baumert und Klieme (2008), dass die Faktoren, die für den größeren oder eben geringen Erfolg eines Bildungssystems führen können – Erfolg immer und allein gemessen an den PISA-Ergebnissen –, sehr vielfältig sind. Neben dem Schulsystem würden auch „kulturelle Normen und Werte, soziale und ökonomische Faktoren, der politische Umgang mit Migration, aber auch Unterrichtstraditionen oder die Qualität der Lehrerbildung eine Rolle“ spielen (2008, 2). Damit musste wohl gerechnet werden. War die Hoffnung, Leistungsstandeshebungen würden Steuerungswissen generieren, also berechtigt? Inwiefern denn überhaupt? Ist es wirklich erstaunlich, dass zwischen Schulsystemmerkmalen und erreichten Kompetenzniveaus keine klare Beziehung besteht? Hat jemals eine schulisch einschlägig informierte Person einen solchen Zusammenhang erwartet? Vielleicht ja, vielleicht nein. Es spielt keine Rolle und spricht keineswegs gegen PISA, TIMSS, IGLU oder VERA. Die Produktion von Nichtwissen ist nicht nur ein Topos des ältesten abendländischen Bildungsideals, sondern gehört – man denke an das Falsifikationsprinzip oder den Fallibilismus – zu reflexiver Wissenschaft und Forschung. (Man kann durchaus Zusammenhänge untersuchen, von denen man fast sicher weiss, dass sie nicht wirklich gefunden werden können ...)

Die mit PISA gefundenen Beziehungen können, wie man weiss, nicht als Ursache-Wirkungs-Zusammenhänge betrachtet werden, weil das nichtexperimentelle Querschnittsdesign der betreffenden Studien dies nicht zulässt: Zu viele bedeutsame unabhängige Variablen, die in keiner Weise auch nur irgendwie systematisch variiert werden könnten. Wir haben also nicht nur das Problem der korrelativen Zusammenhänge, welche keine Aussagen über Kausalität zulassen, sondern vor allem – m. E. wesentlich gewichtiger – das Problem der Komplexität. Die kaum zu bewältigende Komplexität, die durch die nur teilweise sichtbare Vernetzung bedeutsamer Variablen entsteht, wäre wohl zu meistern, stünden uns nicht nur eine, sondern viele Welten zu Verfügung und dazu die Mittel, dieselben nach Forschungsfragen experimentell zu modellieren, dann wäre ein akzeptables Verhältnis von interner und externer bzw. ökologischer Validität dieser Experimente vielleicht erzielbar. Aber dazu wird es, weil wir bloss eine und

dann noch eine komplexe Welt zur Verfügung haben, nicht kommen, selbst wenn die Forschungsgelder, die Zeit und die Menschen dafür vorhanden wären.

### 3. Kapitulation vor hoher Komplexität

Die Kapitulation vor hoher Komplexität ist eine vernünftige Verhaltensweise. Jedenfalls scheint es mir vernünftiger zu sein, vor so hoher Komplexität zu kapitulieren als weiterhin zu suggerieren, sie könnte gezähmt oder kontrolliert oder solle einfach souverän ignoriert werden. Wünschenswert wäre also eine gewisse Kapitulationskompetenz: Um zu kapitulieren, muss man sich eingestehen können, dass man gescheitert ist und wahrscheinlich weiterhin scheitern wird. Solche Kapitulationen sind in der Wissenschaft völlig normal, wenn auch jeweils frustrierend. Erinnern wir uns etwa an die Aptitude-Treatment-Interaction-Forschung. Welch grossartiges Programm, und wie wenig ist davon übrig geblieben? Freilich nicht nichts, sondern die Einsicht in die Begrenztheit der Möglichkeiten, schulisches Lernen und seine Bedingungen möglichst individuell zu erfassen und dann in der Unterrichtspraxis zu optimieren.

In dieser Hinsicht gefallen mir die Amerikaner. Sie leben in einer grossartigen Nation und machen daher auch immer wieder grossartige Fehler – im Unterschied zu kleinen Nationen, welche die grossen Nationen Fehler machen lassen und dann vielleicht doch nichts davon lernen. Die Amerikaner haben auch die Grösse, sich ihre Fehler kurz darauf rückhaltlos einzugestehen. Die American Evaluation Association (AEA 2006) fasste die Erfahrungen mit der Wirkung bzw. Wirkungslosigkeit regelmässiger Leistungsstanderhebungen vor einigen Jahren wie folgt zusammen: „Obwohl mehr als zwei Dekaden in Gebrauch, hat staatlich verordnetes Testen mit einschneidenden Folgen (für Schüler, Lehrer und/oder Schulen) die Qualität der Schulen nicht verbessert noch hat es die Disparitäten im Bildungserfolg zwischen den Geschlechtern, Rassen oder Klassen beseitigt, noch hat es das Land in moralischen, sozialen oder ökonomischen Begriffen vorangebracht“ (zit. u. übers. nach R. Münch 2009, 38)<sup>3</sup>.

Wenn kein valides Erklärungswissen und auch kein entsprechendes Steuerungswissen aus PISA et al. entwickelt werden kann, so heisst das freilich nicht, dass internationale Vergleiche überflüssig sind. „Wenn man aber“, wie Prenzel, Baumert und Klieme (2008, 3) schreiben, „herausfinden will, welche Folgen eine spezielle politische oder pädagogische Massnahme hat, muss man die Entwicklung der betroffenen

---

<sup>3</sup> Kaum vernimmt man im deutschsprachigen Raum von offizieller Seite vergleichbare „Eingeständnisse“. Hier scheint niemand zu scheitern und niemand verantwortlich zu sein, die Bildungsforscher sowieso nicht.

Schulen und Schüler systematisch und über einen ausreichend langen Zeitraum mit der Entwicklung anderer Schulen und nicht betroffener Schüler vergleichen“. Das ist freilich ein ganz anderes Untersuchungsdesign, welches auch tatsächlich eher zu versprechen vermag, Erklärungswissen und allenfalls Steuerungswissen zu generieren.

Die Autoren schreiben weiter: „Wenn Grenzen und Nutzen der empirischen Bildungsforschung realistisch gesehen werden, dann ist sie dort angekommen, wo andere empirische Disziplinen mit längerer Forschungstradition längst sind: im Bereich der normalen Wissenschaft“ (3). Ja genau, im Bereich der normalen Wissenschaft. Dann wird mit Bildungsforschung Bildungspraxis so wenig verändert (und zu verändern versucht), wie mit Politikwissenschaft Politik gestaltet oder mit Wirtschaftswissenschaft die Wirtschaft geregelt wird; eine politikwissenschaftlich basierte Politik wird es nie geben. Das ist eben normal. Und wenn die Idee aufgegeben wird, dass Bildungspolitik auf Bildungsforschung basieren könnte, wird das eine wahre und gute Ent-Täuschung sein.

Noch aber ist die Bildungsforschung nicht ganz „normal“. Daher stimme ich Prenzel, Baumert und Klieme zu, auch wenn diese Autoren „Normalität“ hier wohl etwas anders verstanden haben möchten. Die Bildungsforschung, um die es hier geht, ist noch stark eine „embedded science“, versteht sie sich doch offenbar als Instrument der Bildungspolitik und Bildungsadministration. Dieses bloss vorgetäuschte oder tatsächliche szientifische Selbstmissverständnis hat sich in den letzten zwei Jahrzehnten als funktional erwiesen (es gibt und gab Synergien zwischen Forschung und Politik, basierend auf wechselseitigem Nutzen), aber wissenschaftstheoretisch und wissenschaftspolitisch ist es längerfristig wahrscheinlich nicht überzeugend. Ob und wann die Bildungsforschung je normal sein wird, ist aber nicht das, was mich hier interessiert.

#### 4. Was folgt aus Evaluationen?

Margrit Stamm hat in ihrer Habilitationsschrift *Evaluation und ihre Folgen für die Bildung* (Stamm 2003) herausgearbeitet, wie problematisch die implizite oder auch explizite Gleichsetzung von Relevanz und Wirkung im Bereich der Bildungsevaluation sein kann. So wurde nun lange angenommen, „der von der Evaluation produzierte Wissenskorpus sei nicht nur relevant, sondern auch wirksam und in der Lage, die Probleme zu lösen“, die der Grund dafür sind, dass „Evaluationsforschung um Rat gefragt worden ist“ (Stamm 2003, 357). Zum Fazit ihrer Studie schreibt Stamm: „Die problematischen Wirksamkeitsbefunde meiner empirischen Studie und die Erkenntnisse über die verschlungenen Wege des Praktischwerdens von Evaluationswissen, aber auch die über die Evaluationsforschung aufgearbeitete Literatur, lassen

Evaluationen in einem problematischeren Licht erscheinen, als dies in Anbetracht der aktuellen bildungspolitischen Grosswetterlage wünschbar wäre. Dazu kommt, dass es immer mehr evaluationserfahrene Laien gibt, die Evaluationswissen auch in Frage stellen können, sodass unsicher bleibt, ob die Hoffnung, mehr Evaluation erzeuge verstärkt Veränderung, nicht doch als berufspolitischer Zweckoptimismus einer ganzen Profession bezeichnet werden muss“ (ebd.).

Wer Evaluationen allein durch ihren Nutzen rechtfertigen will, muss letztlich unterstellen, so Stamm, Nutzung sei immer gut, Nicht-Nutzung aber „falsch, schlecht oder missbräuchlich“ (359). Auch wenn die Instrumentalisierungsthese nicht vertreten wird – denn sie steht ja auf sehr wackeligen Füßen –, ergibt sich nach Stamm das Problem, „dass Evaluationswissen nur in seltenen Fällen einem Rationalitätskontinuum entsprechend direkt und vollständig in Umsetzungsprozesse überführt wird. Viel häufiger gehen sie mit aufwändigen Verarbeitungsprozessen einher, die zu selektiven Konklusionen führen und deshalb im Blickwinkel der Evaluationstheorie immer Verfälschung und Zensur sein müssen“ (358 f.). Wenn gefragt werde, „unter welchen Bedingungen Evaluation als Königsweg zu einer neuen Bildungsqualität bezeichnet werden kann und die Antwort kulturelle und politische Kontextverhältnisse mitberücksichtigen soll, dann entflechten solche Überlegungen die Evaluation von ihrer fiktiven Konzentration auf deduktiven und instrumentellen Gebrauch und erweitern den *Impact* auf andere beeinflussende Grössen“ (359).

Kurz, es gibt keinen direkten Weg von relativem Standortwissen zu Steuerungswissen. Zu fragen ist, wann ersteres überhaupt Hilfe darstellt, um zu letzterem zu kommen. Man kann auch fragen, was gegeben und gezeigt werden müsste, um die Behauptung aufrechtzuerhalten, dass Evaluationen der Qualitätssicherung dienen würden. In einem konkret praxeologischen Verständnis ist dies unproblematisch. Jeder Handwerker, ja jeder irgendwie tätige Mensch evaluiert seine Arbeitsprozesse in der einen oder anderen Weise; in der Regel begleitet sukzessives, stetiges Evaluieren den Prozess, denn nur so kann die Qualität des Arbeitsproduktes gesichert werden. Hier geht es um Sorgfalt und Berufsethos. Weil dies so evident ist und ganz den täglichen Erfahrungen des Lebens und Arbeitens entspricht, folgern wir wohl auch etwas zu schnell, dass Evaluationen im komplexen, institutionellen und systemischen Feld, wenn sie etwa Leistung und Wirkung von Bildungssystemen betreffen, nach dem gleichen engen Zusammenhang relevant und wirksam sein müssen. Ganz abgesehen davon, dass hier die Überprüfung immer erst erfolgt, wenn in den Prozess der Entstehung des Evaluationsgegenstandes gar nicht mehr eingegriffen werden kann, haben wir es hier mit der Gemengelage der Mehrebenen zu tun: Was auf Ebene eines nationalen

Bildungssystems oder Bundeslandes zutreffen mag, hat vielleicht im Kopf des Lerners überhaupt keine Entsprechung.

2007 hat Helmut Heid gezeigt, welche Kausalketten und Wirkungszusammenhänge angenommen und vorausgesetzt werden können müssen, damit die so plausibel erscheinende Behauptung, wonach Leistungsstanderhebungen im Bildungsbereich der Qualitätssicherung dienen würden, aufrechterhalten werden kann. *Erstens* müsse gezeigt werden, dass die „Verwirklichung des in Standards Kodifizierten (...) Resultat (‚Wirkung‘) abgrenzbarer bzw. messbarer Lernaktivitäten, allgemeiner: erfolgreichen Lernens“ ist. *Zweitens*, dass erfolgreiches Lernen das „Resultat (‚Wirkung‘) identifizierbarer Lehraktivitäten, allgemeiner: erfolgreichen Lehrens“ darstellt. *Drittens*, dass erfolgreiche Lehraktivitäten Ausdruck (Indikator) hoher Lehrkompetenz darstellen. *Viertens*, dass hohe Lehrkompetenz das Resultat („Wirkung“) erfolgreicher Lehrerbildung bzw. Lehrfortbildung darstellt, und schliesslich, *fünftens*, dass erfolgreiche Lehrerbildung das Resultat („Wirkung“) guter Bildungsforschung ist (Heid 2007, 37). Wenn diese Zusammenhänge nicht nachgewiesen werden könnten, so Heid, „dann besagt die Erfüllung eines Standards nur eines, nämlich dass der Standard erfüllt ist – und nichts darüber hinaus!“ (Heid a. a. O., 37).

Der scheinbar klare Zusammenhang von Evaluation und Qualitätssicherung, so evident er auf der Mikroebene des Unterrichts ist, entpuppt sich auf der Bildungssystemebene in methodischer, forschungspraktischer, infrastruktureller, zeitökonomischer und nicht zuletzt finanzieller Hinsicht als so voraussetzungsreich, dass die grossen Hoffnungen, die dahinein gesetzt werden (wenn es denn um Systemsteuerung gehen soll), m. E. nahezu grundlos sind.

## 5. Alte Skepsis an Evidenzbehauptungen

Diese Skepsis ist keineswegs neu. Vor gut hundert Jahren schrieb Alfred North Whitehead: „Ich behaupte, dass kein System externer Tests, das primär darauf abzielt, einzelne Studierende zu prüfen, zu irgendetwas führt, ausser zu erzieherischer Verschwendung“. Und: „Es sind primär die Schulen und nicht die Studierenden, die überprüft werden sollen“ (Whitehead, 2012, 54 [Original 1912]). Dafür sei Voraussetzung, dass jede Schule „auf der Grundlage ihres eigenen Curriculums ihre eigenen Abschlusszeugnisse ausstellen“ sollte (ebd.). Denn: „Jede Schule muss den Anspruch haben, mit Bezug auf ihre speziellen Verhältnisse betrachtet zu werden. Die Klassifizierung von Schulen für bestimmte Zwecke ist notwendig. Aber kein absolut strenges Curriculum, das nicht vom eigenen Personal modifiziert worden ist, sollte

zulässig sein“ (ebd.). Das mag aus heutiger Sicht reichlich radikal erscheinen. Dies wäre aber kein gutes Argument gegen den Wahrheitsanspruch, der von Whitehead vor so langer Zeit schon erhoben worden ist.

Nun ist sicher zu beachten, dass die nationalen Forschungsprogramme, die etwa in Deutschland an die PISA-Erhebungen angelagert worden sind, vermehrt zu Einsichten geführt haben, welche über die blossen Olympiade-Ergebnisse hinausgehen. Nach Prenzel, Baumert und Klieme (2008) gehören dazu die Einsichten aus Überprüfungen der „Lehrplanvalidität der Tests, innovative Testerweiterungen, Ländervergleiche, Erweiterung der Migrantenstichprobe, ergänzende Untersuchungen zu Bildungsgerechtigkeit, Unterrichts- und Schulqualität, Leistungsentwicklung und Leistungsstagnation oder Untersuchungen zur professionellen Kompetenz von Lehrkräften“ (3). Diese zusätzlichen, PISA ergänzenden Themen hätten „den Gedanken einer stärker evidenzbasierten Bildungspolitik überhaupt erst entstehen lassen“ (4). Insofern ist es sicher richtig zu sagen, dass in der Folge der internationalen Leistungsvergleiche eine beachtliche Reihe von Untersuchungen initiiert worden ist, welche nicht nur das oben genannte Standortwissen betreffen, sondern auf wesentlich besserer Grundlage auf Erklärungswissen abzielen können. Nur werden die genannten Grenzen zum Zusammenhang von Evaluation und Qualitätssicherung sowie das Problem der Steuerung im Bildungswesen damit nicht aufgehoben, und die unbescheidene Formel der „evidenzbasierten Bildungspolitik“ verliert ihren Chimärencharakter damit ebenso wenig.

Es ist einigermaßen problematisch, dass der Begriff der Evidenz in einem so komplexen Feld wie der Erfassung von Leistungen und Wirkungen von Bildungssystemen bzw. überhaupt im Bereich der Bildung neuerdings verwendet wird. Evidenzbasierte Bildungspolitik erscheint mir etwa so wahrscheinlich und begrüßenswert zu sein wie eine wissenschaftlich fundierte Lebensführung. Mit einer Person, die so ihr Leben führt, möchte ich jedenfalls nicht verheiratet sein. Wenn im Bereich der Sozialwissenschaften von „empirischer Evidenz“ gesprochen wird, dann meint man ja in der Regel, dass solide Befunde empirischer Untersuchungen für oder aber gegen eine Theorie sprechen. In der Philosophie meint Evidenz vornehmlich die unmittelbare Erkenntnis, das Selbstverständliche und Offensichtliche, das keiner Erklärung bedarf und vielleicht auch keiner Erklärung fähig ist. Wohl ist davon auszugehen, dass PISA-Leistungspunkte, sagen wir 525 oder 497, nicht für sich allein sprechen können. Philosophische Evidenz kann also nicht gemeint sein. Bleiben wir bei den Sozialwissenschaften: hier könnten 525 und 497 die Bedeutung haben, Teil einer empirischen Evidenz zu sein, wenn denn klar wäre, welche Theorie damit bestätigt oder widerlegt werden könnte. Doch diese Theorie gibt es ja nicht. Theorieentwicklung ist ja auch nicht Ziel von Evaluationsbe-

mühungen, allenfalls zufälliges Nebenprodukt. Es ist aber anzunehmen, dass jene, die eine evidenzbasierte Bildungspolitik fordern oder wünschen, oder schlimmer noch: glauben, dass sie sie selbst praktizieren, davon ausgehen, Bildungsforschung könne genau solch erwünschte Evidenz produzieren. Von solch anachronistisch anmutender Wissenschaftsgläubigkeit oder Wissenschaftsfrömmigkeit sollte sich der empirische Bildungsforscher nicht geschmeichelt fühlen, besser sollte er davor zurückschrecken.

Jede sozialwissenschaftliche Forschung versucht zu beschreiben, was ist. Wie es sich verhält. Sie versucht, wenn es geht, den Gegenstand zu verstehen und damit verbundene Phänomene zu erklären. Das geht nicht ohne Theorie, d. h. reflektierte Abstandspraxis. Würde im Bereich der Bildung jene (Selbst-)Evidenzen „gefunden“, welche in der Bildungspolitik offenbar gewünscht werden, dann wäre Bildungsforschung im Grunde nahezu überflüssig. Der Bildungsforscher würde sozusagen höchstens auf die Rolle reduziert, die Wirklichkeit der Bildung für sich sprechen zu lassen. Nicht, dass diese Idee ganz absurd wäre. Vielmehr gehört sie zur Geschichte der Wissenschaft und der *Lesbarkeit der Welt*. Sie stellt einen Oberflächenrealismus dar, der von Wissenschaftlern nicht vertreten werden dürfte.

Daher zurück zu Blumenberg. Er schreibt: „Es ist die Grundidee des Empirismus, die Natur würde von sich aus ihre Geschichte erzählen, wenn man sie nur liesse, sich nicht ihr aufdrängte und vordrängte, sie nicht durch Vorwissen und Vorurteile an ihrer Selbstkundgabe hinderte./Eine der sprachlichen Voraussetzungen für diese Grundvorstellung ist, dass die Naturwissenschaft eben nicht ‚Naturwissenschaft‘ hiess, sondern ‚Naturgeschichte‘, *historia naturalis*. Es ist eine Umdeutung des Sinnes, dass der Mensch diese Geschichte erzählt, wenn der frühe Empirismus mit dem Gedanken spielt, man müsse die Natur selbst sie zwanglos preisgeben lassen“ (Blumenberg 1981, 86).

Wer aber heute noch glaubt, die soziale Wirklichkeit liesse sie von sich aus so unmittelbar zur Sprache bringen, wenn man sie nur liesse, sodass Bildungspolitik quasi aufgrund von Evidenzerlebnissen praktiziert werden könne, der darf freilich von ontologischen, epistemologischen und methodologischen Fragen keine Ahnung haben, um seine oberflächenrealistische Irritationslosigkeit aufrechtzuerhalten.

Woher kommt nun dieser Wunsch, es jetzt immer mehr ganz genau wissen zu wollen? Sodass man aufgrund dieses genauen Wissens auch genau weiss, was zu tun ist? Das kann sich offenbar nur ein Wesen wünschen, das nicht weiss, was zu tun ist, wo es steht, woher es kommt und wohin es geht. Ein Wesen, das sich nicht verorten kann. Eines, das meint, vielleicht auch vermeintlich, es wüsste nicht, wo es steht und was es will – offensichtlich ein Wesen mit einem Identitätsproblem. In einem seiner letzten Interviews sprach der Soziolinguist Basil Bernstein im Zusammenhang der

Transformationen des Bildungssystems und der Kompetenzorientierung von „pädagogischer Panik“: „Ich denke, was wir gerade erleben, ist eine pädagogische Panik, die die moralische Panik maskiert, eine tiefe Panik in unserer Gesellschaft, die nicht weiß, was ist und wohin es geht. Und das ist eine Periode der pädagogischen Panik. Und es ist das erste Mal, dass pädagogische Panik die moralische Panik maskiert bzw. verschleiert“ (Bernstein, zit. n. u. übers. v. Serfl 2004, 26; vgl. dazu auch Bernstein 1996).

Warum reicht es heute nicht mehr, von gerechtfertigten Meinungen, guten Ideen und einer Sensibilität für Probleme auszugehen, um Bildungspolitik zu betreiben? Wäre es nicht plausibler und wünschenswerter, statt einer evidenzbasierten Bildungspolitik einfach eine problemorientierte Bildungspolitik zu haben? Liegt nicht darin überhaupt die wesentliche Aufgabe von Politik, Probleme zu behandeln?

Was durften wir erhoffen? „Zu fragen, ob unsere Welt die Welt sei, die wir wollen, gehört zur politischen Rhetorik“, schreibt Blumenberg (1981, 10). „Sie muss mit ihrem Recht den Eindruck erwecken, es gebe Welten zum Ausschauen oder habe sie gegeben; dabei könne man sich in der Wahl vergriffen haben, was Korrekturen erfordere und legitimiere. Doch was so in den grossen Gestus der Weltverwaltung eingeht, lässt sich tatsächlich als Fragestellung an jede geschichtliche Epoche herantragen: *Welches war die Welt, die man haben zu können glaubte?*“ (Blumenberg a. a. O., 10).

## 6. Schulen als organisierte Anarchien

Natürlich ist es funktional für manche Reformen, nicht nur im Bildungssystem, wenn diese Frage nicht gestellt wird und – falls sie schon einmal gestellt worden ist – zu vergessen; vergessen, „dass das, was man vorhat, schon einmal (oder mehrmals) versucht worden und gescheitert ist. Die wichtigste Ressource der Reformer scheint daher eine Leistung des Systemgedächtnisses zu sein, nämlich das Vergessen“ (Luhmann 2002, 167). In einer guten Welt, so könne man argumentieren, wäre evidenzbasierte Bildungspolitik möglich und Bildungssteuerung verlief rational, immer im Dienste der Verbesserung des Systems und der Sicherung der Qualität. In einer solchen Welt leben wir allerdings zweifelsohne nicht, sondern vielmehr in einer nur sehr begrenzt steuerbaren, aber dann doch irgendwie nicht allzu schlechten und schlimmen Welt (wenigstens in unseren Breitengraden). Wenn man auf einem Dampfer ist, der zwar kaum steuerbar ist, aber glücklicherweise meist in die richtige Richtung fährt, dann sollte einen, glaube ich, der Wunsch nach Steuerbarkeit nicht zu einem Aktivismus verleiten, mit der der Kahn von seinem Weg immer mehr abzuweichen droht. Man muss, glaube ich, kein Buddhist sein, um zu erkennen, dass zur Steuerung und Führung

auch eine meist unterschätzte Strategie gehört: die Inaktivität. Einzuwenden, diese Sicht sei eine Rechtfertigung des Status quo, ja, so würde man die ganzen Ungerechtigkeiten im Bildungssystem einfach in Kauf nehmen wollen, ist ein recht dummer, moralisierender Vorwurf. Denn aus der Krisenhaftigkeit der Welt folgt keineswegs ihre Steuerbarkeit. Wer aber Probleme ernst nimmt, will sie zu lösen versuchen, das ist klar. Aber die Probleme im Bildungsbereich werden nicht durch noch vermehrte und intensivere, systematischere Leistungsstanderhebungen gelöst oder auch besser verstanden werden. Wer Bremen voranbringen will, der soll das tun. Was auch immer Bremen helfen wird, weitere Bildungsforschung wird es nicht sein: Praxis lässt sich nur in der Praxis verbessern.

Die Organisationstheoretiker Cohen, March und Olson haben Schulen schon vor drei Jahrzehnten als „organisierte Anarchien“ verstanden (1990) – sowie übrigens auch Universitäten und Verbrecherorganisationen. Organisierte Anarchien sind nach den Autoren durch drei Merkmale geprägt: (1) durch problematische Präferenzen, (2) unklare Technologien und (3) fluktuierende Partizipation (Cohen u. a. 1990, 330). Mit Steuerungsmodellen würden Kontrollmechanismen und Koordinationsmechanismen eingeführt, die „sowohl von der Existenz wohldefinierter Ziele und einer wohldefinierter Technologie als auch von einer persönlichen Verwicklung der Beteiligten und die Angelegenheiten der Organisation ausgehen. Wo hingegen die Ziele und die Technologie verschwommen und die Teilnahme fluktuierend ist, brechen viele der Axiome und Grundprozeduren des Managements zusammen“ (Cohen u. a. 1990, 331). Die Einsichten bzw. Behauptungen des Modells organisierter Anarchien sind mindestens dreifach: (1) Mit Handlungen werden die Präferenzen der Entscheidungsträger erst aufgedeckt. (2) Probleme werden behandelt, ohne dass rationale präferenzgesteuerte Entscheidungen getroffen werden können. (3) Vorhandene Lösungsinstrumente – z. B. Leistungsstanderhebungen – definieren vielmehr das Problem und die Präferenz, als dass sie als Mittel von vorher beanstandeten Problemen fungieren.

Dies hat auch „Folgen“ für die Bildungsforschung. Zum Beispiel kann auch dort geforscht werden, wo es kaum etwas zu erkennen gibt (außer etwas mehr oder flüchtig-sportliche Ranglisten). Weiter werden Evaluationen auch dann durchgeführt, wenn ihre Notwendigkeit gar nicht plausibel gemacht werden kann. Und Bildungsforschung steht unter (dem vermeintlichen) Druck, für politisches Handeln regelmäßig „Evidenzen“ vorzufinden oder nötigenfalls zu erzeugen (ohne diese Erkenntnispolitik öffentlich zu hinterfragen, z. B. dahingehend, was denn überhaupt als „Evidenz“ zählen kann und was nicht). Immer wieder müssen Äpfel mit Birnen und Karotten mit Kartoffeln verglichen werden, ohne dass vertieft über eine Theorie des Obstes oder des Gemüses nachgedacht

wird. Kurz: Die Erzeugung von Problemen, Präferenzen und Problempräferenzen durch Lösungsmittel führt dazu, dass kaum noch gefragt wird, welcher Gegenstand eigentlich und warum erforscht werden soll.

Evident ist im Bereich der Bildung sicher die normative Überdetermination des Schulsystems: Zu viel Sollen, zu wenig Können. Zu viele Probleme, zu wenig Mittel. Was ist in dieser Lage tun? Niklas Luhmann meinte: „Man kann mehr auf Förderung der Individuen abstellen oder mehr auf Herstellung von Chancengleichheit und Ausgleich von Benachteiligungen durch Schichtung und Geschlecht. Man kann mehr auf Förderung der Schulklasse Wert legen oder mehr auf Förderung der Zurückbleibenden. Wie geht das System mit solchen Inkonsistenzen um? Auf der Ebene der Professionen mag man sie leugnen und in den Geheimnisbereich des individuellen Könnens und der Erfahrung hineinziehen. Die Lösung des Problems liegt in der Leugnung des Problems ...“ (Luhmann 2002, 165). Die Leugnung des Problems der normativen Überdeterminiertheit des Schulsystems und die Leugnung der damit verbundenen Kontingenz der Problembehandlung sowie Bildungsforschung sind so evident wie der Umstand, dass mit der Veränderung der Steuerungsphilosophie oder -phantasie die zentralen Probleme des Bildungssystems nicht gelöst werden können. Es gehört zur Komplexität der gesellschaftlichen Subsysteme, dass man sich letztlich meist irgendwie „durchwursteln“ muss – was Lindblom bekanntlich als *science of muddling-through* bezeichnet hat (Lindblom 1959) –, ob einem das passt oder nicht. In der Antike kannte man noch den Unterschied zwischen Praxis und Technik und wusste daher, dass Praxis nicht primär in der Anwendung von Regelwissen bestehen kann. Die Schule ist nicht nur eine Bildungs- und Ausbildungsanstalt (Personalisations- und Qualifikationsinstanz), sondern auch Eingliederungsanstalt (Sozialisationsinstanz), Legitimationsanstalt (Instanz politischer Integration), Ausleseanstalt (Selektionsinstanz/ soziale Reproduktionsinstanz), eine wirtschaftliche Anstalt (ökonomische Institution), eine Verwaltungseinrichtung (administrative Institution), eine verselbständigte und selbständige Anstalt (autozentrische und autonome Institution), eine Zwangsanstalt und eine nützliche Aufbewahrungsanstalt, ohne welche modernes Leben nicht möglich wäre. PISA et alter haben mit den meisten dieser gesellschaftlich zentralen Funktionen des Bildungssystems gar nichts zu tun, sie sind nicht Gegenstand der internationalen oder nationalen Evaluationen. Wie könnte man ernsthaft annehmen, dass ein so komplexes gesellschaftliches Subsystem auf der Basis von Punkte-Resultaten der Schülerinnen und Schüler in weniger als einer Handvoll Fächern gesteuert werden könnte oder auch gesteuert werden sollte! Oder auch bloss, dass damit ein angemessenes Systemmonitoring möglich sei? Da reicht doch ein Parameter (die Kompetenzpunkte

...) kaum aus, außer man wolle partout in die Falle des Campbell'schen Gesetzes laufen (Campbell 1975)<sup>4</sup>. Monitoring heisst Überwachung – und die sollte dort bleiben, wo sie auch herkommt und nützlich ist: in die Ozeanologie, die Hydrologie, die Seismologie, in die Intensivstation. Wasserstand, Erdbewegungen und Herzfrequenzen lassen sich wenigstens „wirklich“ messen (und diese Messung muss nicht im Lichte von Traditionen, kulturellen Normen und moralischen Werten gedeutet werden). Kurz, auf der Basis von selektivem Wissen zum Kompetenzstand lässt sich Schulqualität nicht sichern (vgl. z. B. Fischer/Kauertz/Neumann 2008, 32).

## Zum Schluss

Und was bleibt denn am Schluss eigentlich konkret übrig von diesem Weinert'schen Kompetenzverständnis, das zunächst viele pädagogische und noch viel mehr bildungspolitische Köpfe nicken liess? Kompetenz, so hiess es damals, sei das Zusammenspiel von kognitiven und praktischen Fähigkeiten (skills), von Wissen (knowledge and tacit knowledge), Motivation, Werte-Orientierung, Einstellungen, Emotionen und weiteren Komponenten, die für das Ausführen einer Handlung benötigt würden (vgl. Weinert 1999, 8). Es wäre mir neu, dass PISA dergleichen erfasst oder auch erfassen könnte. Allein weiß ich etwa, dass Daten zur Motivation und zum Interesse miterhoben werden und wurden. Und dass die Südkoreaner, obwohl sie akademisch Topleistungen wie die Finnen zeigen, bei PISA die niedrigsten Interesse- und Motivationswerte aufweisen. Wahrscheinlich ist das die empirische Evidenz dafür, dass man nicht motiviert sein muss, um Höchstleistungen zu erbringen. Ja eben, die Kultur, die Werte, die Unterrichtstraditionen, irgendwie spielt das alles in diesem Bereich mit: „Everything is connected, somehow“ ...

Zusammenfassend sei dreierlei bemerkt und wiederholt:

- *Erstens*: Mit PISA, TIMSS, IGLU und VERA (u. a.) wird kein Steuerungswissen generiert, diese Erhebungen können nicht die Basis von Bildungspolitik darstellen. Statt Steuerungswissen hat Bildungspolitik, aber auch Bildungspraxis auf *Orientierungswissen* zu setzen: dasselbe hat in demokratischen, liberalen Gesellschaften viele Quellen. Auf welche man aber hauptsächlich setzt, wird weiterhin umstritten bleiben (müssen).

<sup>4</sup> Campbells These besagt im Wesentlichen, dass je stärker ein einzelner quantitativer Faktor soziale Entscheidungen begründen soll, desto stärker verzerrenden Einflüssen ausgesetzt ist und selbst dazu führt, die sozialen Prozesse zu verfälschen, die untersucht und verbessert werden sollen: „The more any quantitative social indicator is used for social decision making, the more subject it will be to corruption pressures and the more apt it will be to distort and corrupt the social processes it is intended to monitor“ (Campbell 1976, 49).

- *Zweitens*: Es wäre wünschenswert und nun an der Zeit, dass sich die sogenannte evidenzbasierte Bildungspolitik deutlich bescheidener geben und einfach problemorientiert weiterarbeiten würde.
- *Drittens*: Die empirische Bildungsforschung ist zwar noch nicht ganz normal, aber sie wird es vielleicht wieder werden.

## Literatur

- Bernstein, Basil 1996: *Pedagogy, Symbolic Control and Identity*. London.
- Blumenberg, Hans 1981: *Die Lesbarkeit der Welt*. Frankfurt/M.
- Campbell, Donald T. 1975: *Assessing the Impact of Planned Social Change*. In: Lyons, Gene Martin (Hrsg.): *Social Research and Public Policy*. Hanover, S. 49 ff.
- Cohen, Michael D. u. a. 1990: Ein Papierkorb-Modell für organisiertes Wahrverhalten. In: March, J. G. (Hrsg.): *Entscheidung und Organisation. Kritische und konstruktive Entwicklungen und Perspektiven*. Braunschweig, S. 329-372 (amerikanisches Original 1972).
- Fischer, Hans E. u. a. 2008: *Standards and Bildung*. In: Mikelskis-Seifert, Silke u. a. (Hrsg.): *Four Decades of Research in Science Education – from Curriculum Development to Quality Improvement*. Münster, S. 29-41.
- Heid, Helmut 2007: Was vermag die Standardisierung wünschenswerter Lernoutputs zur Qualitätsverbesserung des Bildungswesens beizutragen? In: Benner, Dietrich (Hrsg.): *Bildungsstandards. Chancen und Grenzen, Beispiele und Perspektiven*. Paderborn, S. 29-48.
- Klieme, Eckhard u. a. 2010: Kompetenzmodellierung. Eine aktuelle Zwischenbilanz des DGF-Schwerpunktprogrammes. In: *Zeitschrift für Pädagogik*, Jg. 56., Beiheft 56, S. 9-11.
- Lindblom, Charles E. 1959: *The Science of Muddling-Through*. *Public Administration Review*, Jg. 19, S. 79-88.
- Luhmann, Niklas 2002: *Das Erziehungssystem der Gesellschaft*. Frankfurt/M.
- Münch, Richard 2006: *Die akademische Elite*. Frankfurt/M.
- Prenzel, Manfred u. a. 2008: *Steuerungswissen, Erkenntnisse und Wahlkampfmunition: Was liefert die empirische Bildungsforschung? – Eine Antwort auf Klaus Klemm*. Frankfurt/M.
- Sertl, Michael 2004: *A Totally Pedagogized Society. Basil Bernstein zum Thema*. In: *Schulheft* 116/2004, S. 17-29.
- Stamm, Margrit 2003: *Evaluation und ihre Folgen für die Bildung. Eine unterschätzte pädagogische Herausforderung*. Münster.
- Weinert, Franz E. 1999: *Konzepte der Kompetenz*. Paris.
- Whitehead, Alfred N. 2012: *Die Ziele von Erziehung und Bildung und andere Essays*. Frankfurt/M. (amerik. Original 1967).