

Dieser Text steht unter eine Creative Commons Lizenz¹



Thomas Höhne

Evaluation

Evaluation oder Evaluieren kann in einem ganz grundsätzlichen Sinn als eine Praxis des Unterscheidens aufgefasst werden, wonach Menschen permanent die Welt nach Differenzen wie gut/schlecht, sympathisch/unsympathisch, vorteilhaft/nachteilhaft, Gefallen/Missfallen usw. differenzieren. Unterscheidungen gehen zumeist mit Bewertungen einher: Ob jemand dem politischen links/rechts-Schema zugeordnet wird, ein Land als ‚entwickelt‘, ‚Schwellenland‘ oder ‚Dritte-Welt-Land‘ klassifiziert wird oder eine ‚leistungsorientierte Schule‘ von einer ‚Brennpunktschule‘ unterschieden wird – stets werden auch und gerade in Fragen der ‚Qualität‘ nach bestimmten normativen Kriterien Ein- und Zuordnungen vorgenommen. Dies gilt auch für die Ebene formaler und offizieller Bewertungen wie Noten, Zeugnisse oder Ranglisten, die von entsprechend autorisierten Personen oder Institutionen verliehen oder auch verhängt werden. Der moderne Evaluationsbegriff, wie er als Bewertungspraxis der Leistung von Organisationen gegenwärtig in vielen Bereichen Verwendung findet, geht auf eine historische Periode zu Anfang des 20. Jahrhunderts zurück, in welcher der Staat sein Interesse an Evaluation entdeckte. Im Kontext politischer und gesellschaftlicher Reformen wurde Evaluation in den Vereinigten Staaten großflächig zuerst in den 1940er und dann in den 1960er Jahren angewandt, um die Erfolge politischer Reformprogramme zu überprüfen. Dies war schließlich der historische Startschuss für eine weltweite Expansion von Evaluation, ihren Verfahren und Instrumenten, die auch schließlich über transnationale Organisationen in verschiedene Länder eingeführt worden sind. Von Beginn an war Evaluation eng verknüpft mit einem rationalistischen Steuerungsverständnis, mit der Wirkungen von Mittelallokationen (Inputs) und Reformprozessen sowie die Optimierung von Leistung, Effizienz und Effektivität (Outputs) überprüft werden sollten.

Zur (Vor-)Geschichte des Evaluationsdiskurses

Historisch lassen sich grob drei Phasen der Entwicklung des Evaluationsdiskurses unterscheiden:

- a) Die erste Phase von der zweiten Hälfte des 19. Jahrhunderts bis Anfang des 20. Jahrhunderts kann als Vorgeschichte der Evaluation aufgefasst werden. In dieser Zeit kamen die

¹ Das Werk steht unter einer Creative Commons Namensnennung-Share Alike 4.0 International-Lizenz.

ersten Mess- und Testtechnologien in der Psychologie, teilweise auch in der Pädagogik auf. Eine wichtige Rolle für die Etablierung von Evaluation als politisches Steuerungsmittel spielten Taylorismus und das Scientific Management sowie der Fordismus im Bereich der industriellen Produktion, durch die sich gesellschaftlich ein positivistisches Verständnis rationaler Steuerung gleichermaßen von Prozessen in der Produktion wie auch in der Politik verbreitete. Während es Frederic Winslow Taylor um eine Effizienzsteigerung und damit eine Rationalisierung im Produktionsbereich ging (z.B. die Trennung von Herstellung und Management, kleinschrittige Überwachung einzelner Produktionsschritte und Optimierung der Arbeitsleistung), etablierte Henry Ford die weltweit erste Massenproduktion von Autos am Fließband (das sog. T-Modell). Taylor setzte vor allem auf die Anwendung empirisch-positivistischer Methoden, bei denen es durch Standardisierungen zu Effizienzsteigerungen und Leistungsverbesserung in Arbeitsprozessen ging, während bei Ford Steigerung der Massenproduktion mit Blick auf den Konsum von Massengütern im Vordergrund stand (Kieser 1999: 86). Die intensive Rezeption des Taylorismus und des Scientific Management hatte in den USA in den 1920er und 1930er Jahren ein regelrechtes „Effizienz-Fieber“ zur Folge:

„Universitäten wurden Effektivitätsanalysen genauso unterworfen wie Schulen und die öffentliche Verwaltung. Scientific Management ging eine Allianz mit dem Fortschrittsdenken ein. Getragen wurde die Taylorismus-Bewegung vor allem von den Collegeabsolventen der Mittelschicht, die im Scientific Management eine Technik sahen, die ihnen als Experten Einfluß nicht nur auf das Geschehen in den Unternehmen, sondern auf die gesellschaftliche Entwicklung schlechthin verschaffen würden“. (Kieser 1999: 84)

- b) Die zweite Phase des Evaluationsdiskurses verlief in einer doppelten Bewegung. Zum einen wurden in den 1940er Jahren in den USA Evaluationsverfahren im Curriculumsbereich von Ralph W. Tyler entwickelt und in Form großflächiger Lernzielüberprüfungen im amerikanischen Schulsystem verankert. Aber erst im Verlauf der 60er Jahre gewann Evaluation im Rahmen internationaler Vergleichsstudien eine zentrale transnationale Bedeutung, die von internationalen Organisationen wie OECD, Weltbank und UNESCO forciert wurden. In den USA wie schließlich auch in der BRD wurde Evaluation als politisches Steuerungsinstrument im Rahmen sozialstaatlicher Reformpolitik eingesetzt:

„Evaluationsverfahren knüpfen an so verschiedene Praktiken an wie die kameralistische Rechenschaftslegung, Bonitätsprüfungen im Finanzgewerbe, das betriebliche Controlling oder Konsumentenbefragungen im Rahmen der Marktforschung. Ihr Aufstieg als politisches Steuerungsinstrument fällt in die Zeit der Kennedy- und Johnson Administration in den USA. Dort hatte die Welle sozialpolitischer Reformen auch den Bedarf an Begleitforschung explosionsartig anwachsen lassen. Ein Heer von Sozialwissenschaftlern und -wissenschaftlerinnen hatte sich aufgemacht, um die Effekte jener Programme zu überprüfen, mit denen die Behörden die Armut und die Kriminalität zu bekämpfen, das Bildungs- und Gesundheitssystem zu modernisieren und die Städte zu sanieren versuchten. Die Evaluationsspezialisten dienten sich dabei an als ‚Wegweiser zu rationalerer Sozialplanung‘“. (Bröckling 2004: 77)

In Deutschland wurde Evaluation in den 60er Jahren im politischen und wissenschaftlichen Bereich und dort vor allem im Rahmen der Bildungsplanung als Instrument der Wirkungsanalyse genutzt. Dabei ging es um die wissenschaftliche Einschätzung der Wirkungen politischer und curricularer Programme, was etwa die Evaluation des Gesamtschulprogramms in den 1970er dokumentiert (Gruschka 1976).

- c) In der dritten Phase seit Mitte der 1990er Jahre ff. wird auf Evaluation als zentrales Element organisationaler Restrukturierung vor allem im Rahmen der Verwaltungsreform (z.B. New Public Management, Neue Steuerungsmodelle, Outputorientierung,) zurückgegriffen. Hierbei steht sie paradigmatisch für neue Steuerungsformen und ein outputorientiertes Kontrollregime, indem sie als zentrales Instrument von Leistungsmessung und Effizienzsteigerung eine Schlüsselstellung einnimmt. Internationale Vergleichsstudien wie TIMSS und PISA im Schulbereich wie auch die Etablierung eines europäischen Hochschulraums durch den Bologna-Prozess (z.B. Modularisierung) stellen hierbei wichtige Katalysatoren eines bildungspolitischen ‚Willens zum Vergleich‘ dar, der damit etabliert wurde. Denn sie stehen für Transnationalisierung von Bildungspolitik, die als Rahmen für eine „evaluationsbasierten Steuerung“ (Heinrich 2007) dient, welche als Rückgewinnung staatlicher Steuerung zu werten ist:

„Weniger im Bewusstsein der deutschen Öffentlichkeit sind die internationalen Auswirkungen von Beschlüssen der Europäischen Erziehungsministerkonferenz (Lissabon, Bologna) und der Europäischen Erziehungsdirektion in Brüssel im Rahmen von Sokrates-Programmen, der Ständigen Internationalen Konferenz der Erziehungsin-spektorate (SICI) und Kooperationsvereinbarungen im Kontext der bilateralen Kulturabkommen insbesondere zwischen den Niederlanden, in Niedersachsen und Nordrhein-Westfalen. In der Summe haben diese Einflüsse dazu geführt, dass in Deutschland die Entwicklung von Evaluationsinstrumenten in hohem Maße beschleunigt wurde und ein grundsätzlicher Wandel in der Bildungsverwaltung eingesetzt hat“ (Schnell 2007: 25)

Die Widersprüche externer Evaluation

Interne und externe Evaluation bilden mittlerweile ein Schlüsselement politisch verordneter Rechenschaftsablegung der Einzelschule gegenüber dem Staat (Rürup/Heinrich 2007: 175 ff.). Steuerungstheoretisch stellen beide Evaluationsformen zwei ganz unterschiedliche Operationen dar. Während eine interne Evaluation eine Rückmelde- und (Selbst-)Reflexionsfunktion für die Organisationsmitglieder haben soll, stellt die externe Evaluation zunächst einmal ein externes Kontrollmittel zu Leistungsmessung von Schulen dar (Rürup/Heinrich 2007: 165). Bei Evaluationen im Rahmen von Schulinspektionen werden nicht nur beispielsweise die Managementfähigkeiten von Leitungspersonen, der bauliche Zustand der Schule oder vorhandene Ressourcen bewertet, sondern auch Formen von Lernen und Unterricht, Unterrichtstile, Einsatz von Medien usw.. Mit einer solchen umfassenden ‚Qualitätskontrolle‘ schulischer Leistungen wird also auch die Professionalität von Lehrer_innen immer wieder von neuem auf den (externen) Prüfstand gestellt. Externe Evaluationen sind nicht nur im Rahmen der Schulautonomiegesetzgebung vorgeschrieben, sondern in Form der Qualitätsrahmen auch inhaltlich vordefiniert. Damit wird neben

dem Professionswissen von Lehrer_innen, deren Ausbildung bis dato Garant für die ‚Qualität‘ von Bildung war, mit dem Evaluationswissen eine zweite Wissensform etabliert, mit der die ‚Bildungsqualität‘ nun abgeprüft wird. Dieser regelmäßigen externen Intervention wird zumeist unmittelbar der Status eines objektiven Wissens zugeschrieben, wodurch eine Hierarchie des ‚subjektiven‘ Professionswissens von Lehrer_innen gegenüber dem nun vermeintlich objektiven externen Wissen konstituiert wird. Diese Akzentverschiebung steht prototypisch für neue Steuerungsformen, mit denen organisationsinternes Wissen oftmals gegenüber scheinbar objektiven Bewertungen abgewertet wird. In dieser Hinsicht wird dann auch eine grundsätzliche Kritik formuliert, nach der die evaluationsbasierte Steuerung von Schulen auch als „massiver Beeinflussungsversuch durch Berichtsformen und Berichtspflichten“ verstanden werden könne, „der sich des Mediums ‚Wissen‘ bedient“ (Kussau/Brüsemeister 2007: 38). Es stellt sich die Frage, welche Widersprüche mit einer großflächigen Einführung externer Evaluationen verbunden sind, was ein Blick auf Entwicklung der Evaluationsdiskussion in Hessen deutlich macht:

„Die Initiatoren der Schulprogrammarbeit in Hessen erkannten, dass jegliche Form von Kontrolle kontraproduktiv wirken kann, weil sie tendenziell das Versprechen der zugestandenen Selbstständigkeit dementiert. Das war sicherlich der zentrale Grund dafür, die interne Evaluation als verbindlich, die externe aber als freiwillige einzurichten. Sich freiwillig zur externen Evaluation zu melden, setzt schon eine Souveränität der zu Evaluierenden voraus, die deutlich werden lässt, dass es hier zu einem produktiven Austausch kommen kann. Ist ein solcher Grad an Selbstständigkeit und Selbstbewußtsein nicht voraus zu setzen, steht zu befürchten, dass externe Beurteilung so einen starken Einfluss gewinnt, dass es letztlich doch wieder zu einer Lenkung von außen kommt, die Bedenken und Ratschläge so nachhaltig wirken, dass der Impuls zur Autonomie der Einzelschule, der durch die Schulprogrammarbeit eigentlich gesetzt werden sollte, ins Leere läuft (...) Die Entwicklung zur Eigenverantwortlichkeit kann externe Evaluation nur fördern, wenn die Fremdkontrolle von den Akteure bereits als Instrument der erweiterten Selbstkontrolle akzeptiert ist“ (Heinrich 2007: 180)

Demnach kommt im Rahmen einer Schulautonomie im Grunde nur eine interne Evaluation in Frage, welche die externe Evaluation unterläuft und die zudem wesentlich von der Akzeptanz der Beteiligten abhängt. Ist aber die Akzeptanz bzw. Bereitschaft für Evaluation vorhanden, dann gäbe es auch keinen Grund mehr für eine externe Evaluation. Dies sind die paradoxen Effekte einer ‚fremdgesteuerten Autonomie‘, die nicht von der Einzelschule und ihren Spezifika her gedacht wird, sondern von einem standardisierten Qualitätsmodell, das ‚top down‘ verordnet wird. Ein Blick auf die Niederlande, die als Vorbild für die Qualitätsrahmen vieler Bundesländer diente, zeigt darüber hinaus einen weiteren Effekt staatlich-bürokratischer Autonomiesteuerung, nämlich das Auseinanderdriften guter und schlechter Schulen gerade durch Evaluation. Das niederländische Evaluationssystem geht vom sogenannten „Proportionalitätsprinzip“ aus: „Je qualitativ besser die Selbstevaluation ist, desto bescheidener wird die externe Evaluation durchgeführt“ (Weerts 2007: 45), d.h. gute Schulen werden zusätzlich für ihre Stärken belohnt, während weniger gute Schulen erhöhter Kontrolle unterworfen und damit in gewissem Sinne ‚bestraft‘ werden. Dieses seit Robert Merton als Matthäus-Prinzip bekannte strukturelle

Auseinanderdriften von starken und schwachen Einheiten – denen, die viel haben, wird mehr gegeben, während denen, die weniger haben, mehr genommen wird – kann steuerungstheoretisch auch als negative Rückkopplung verstanden werden. Dieses bleibt nicht aus, wenn externe Evaluationen mit Sanktionen verbunden werden, die – in welcher Form auch immer – notwendig sind, da eine folgenlose Evaluation paradox und damit keine Kontrolle im eigentlichen Sinn wäre. Hier wird die unaufhebbare Widersprüchlichkeit deutlich, die eine externe Kontrollform charakterisiert, welche mit der Legitimation der Autonomiestärkung der Bildungsinstitution daherkommt, ohne diese Autonomie bei der ‚Bildungsqualitätskontrolle‘ selbst in Rechnung zu stellen – etwa durch die Selbst-Kontrolle einer nicht standardisierbaren internen Evaluation.

Ökonomisierende Effekte evaluationsbasierter Steuerung

Die evaluationsorientierte ‚Qualitätsoffensive‘ im deutschen Bildungssystem führt zu einer verstärkten Standardisierung von ‚Bildungsqualität‘, die über externe Vorgaben (z.B. die länderspezifischen Qualitätstableaus und Qualitätsinstitute) durchgesetzt wird. Eine solche Qualitätskontrolle konsolidiert im Wesentlichen die Autonomie starker Schulen, während sie gegenüber schwachen Schulen zahlreiche Sanktionsmöglichkeiten eröffnet und damit die Heteronomie verstärkt. Die Kombination von Evaluation als Kontrollinstrument mit der Politik der ‚Schulautonomie‘ trägt damit zu einer verschärften Differenzierung der Schullandschaft in starke und schwache Schulen bei. Darüber hinausgehend ist zu beobachten, dass die stillschweigende Sanktionsdrohung, die notwendigerweise mit jeder konsequenten externen Evaluation verbunden ist, oftmals in der Semantik positiver Verstärkung (Belohnung) daherkommt – etwa wenn die Durchsetzung externer Vorgaben bzw. die Kalibrierung interner und externer ‚Qualitätskriterien‘ als ‚wichtige Rückmeldefunktion‘ und automatisch als Schritt zu einer ‚Qualitätssteigerung‘ deklariert werden (z.B. durch Zielvereinbarung). Dies ist von zentraler Bedeutung für eine bildungspolitische Steuerung, bei der Evaluation als Instrument der Schulentwicklung in den Mittelpunkt der Bemühungen gerückt wird. Mit dem Ziel der kontinuierlichen Organisations-/Qualitätsentwicklung im Schulsystem werden ‚Schulen unter Standard‘ nicht nur als eigene Gruppe definiert und dadurch sichtbar gemacht, sondern bildungspolitisch unter Dauerbeobachtung gestellt. Mit der Einführung von Evaluation im bundesdeutschen Schulsystem hat sich ein Paradigmenwechsel in der Steuerung vollzogen, der mehrere ökonomisierende Effekte beinhaltet. Zunächst einmal wird mit externer Evaluation als systematischem Steuerungsinstrument ein Wettbewerbselement eingeführt, das Schulen direkter als vorher in Konkurrenz um die ‚bessere Qualität‘ setzt. Dies kann zu Verhaltensänderungen der Art führen, dass Schulen bei erfolgreicher Evaluation ihren Evaluationsbericht strategisch veröffentlichen um sich einen kompetitiven Vorteil gegenüber anderen Schulen zu verschaffen (guter Ruf, ‚sichtbare Qualität‘). Jede positive Rückmeldung erhöht also potentiell das symbolische Kapital der Einzelschule, das im Ringen um die Elterngunst bei der ‚richtigen Schulwahl‘ entscheidend werden kann. Umgekehrt kann – muss aber nicht – ein nicht veröffentlichter Evaluationsbericht auf Elternseite als negatives Marktsignal gedeutet werden, das einen ‚Qualitätsmangel‘ befürchten lassen könnte. Wenn man schließlich davon ausgeht, dass die Effektivität von Bildungsorganisationen von professions-eigenen Kriterien abhängt (Erfahrungswissen der Lehrer_innen, Ausbildungsstandards, Weiterbildung usw.), dann stellt die evaluationsbasierte externe Definition von Bildungsqualität eine Neubestimmung von Effektivität und damit ein verändertes Ziel-Mittel-Verhältnis dar. Sie drückt

sich in der grundlegenden Verschiebung von einer professionsnahen inputorientierten Effektivität zu einer outputorientierten Effektivität aus, der von außen Qualitätskriterien vorgegeben werden. Die ‚Entmachtung‘ der Profession bzw. Deprofessionalisierung wird daher von kritischen Beobachter_innen als wichtiges Indiz für Ökonomisierung auf der Ebene der Organisation erachtet (Radtke 2009). Diese Kritik kann noch mit Blick auf die Konstruktion von Quasi-Märkten ausgebaut werden. Denn wenn die Operationen von Schulen verstärkt im Kontext von Evaluation an den vorgegebenen Außenkriterien orientiert sind (z.B. Fassadenverhalten bei Schulinspektion, Erhöhung des eigenen Kapitals durch veröffentlichte Evaluationsergebnisse und Kundenorientierung), dann führt dies langfristig, so ist anzunehmen, zu einer grundlegenden Umschreibung von Professionswissen, -praktiken und -normen. Das Problem besteht darin, dass kein Raum gegeben ist, um die möglichen Veränderungen von Professionshandeln zu thematisieren und reflektieren, da durch die vorgegebenen Qualitätskriterien immer schon alles gleichsam politisch (vor-)entschieden ist. Auf den möglichen Schereneffekt einer evaluationsbasierten Steuerung im Sinne des sogn. Matthäus-Effekt wurde bereits hingewiesen. Die mit dem Vergleichsinstrument induzierte verstärkte gegenseitige Beobachtung von Schulen, die dem gleichen standardisierten Bewertungssystem ausgesetzt sind, stellt einen logischen Folgeeffekt der Einführung von Evaluation dar. Schließlich liegen auch Hinweise aus empirischen Untersuchungen zur expliziten Marktbildung im Schulbereich vor, wo Evaluation als Instrument im Rahmen von Modellprojekten regional eingeführt wurde (Höhne/Schreck 2009). So hat sich in der Folge des SEIS-Projekts (Selbstevaluation in Schulen) der Bertelsmann Stiftung mit mehreren Bundesländern ein Markt für die Fortbildung, Multiplikatoren und Dienstleister einschließlich eines Franchise-Systems für das gleichnamige Evaluationsinstrument gebildet (ebd.: 200 ff.). Auf der einen Seite sind im Bereich der Evaluation damit die direkten ökonomisierenden Effekte öffentlich-privater Kooperationen als schulische Dienstleistung nachweisbar, auf der anderen Seite bestehen diese aber auch in der Durchsetzung eines rationalistischen Modells von Steuerung im Zeichen der Qualitätsverbesserung. Vor allem dazu hat die flächendeckende Einführung von Evaluation nach PISA in besonderer Weise beigetragen.

Literatur

- Bröckling, Ulrich (2004): Evaluation. In: ders./ Krasmann, Susann/ Lemke, Thomas (Hrsg.): Glossar der Gegenwart. Frankfurt am Main: Suhrkamp, S. 76-81.
- Gruschka, Andreas (Hrsg.) (1976): Ein Schulversuch wird überprüft. Das Evaluationsdesign für Kollegstufe NW als Konzept handlungsorientierter Begleitforschung. Kronberg: Fischer.
- Heinrich, Martin (2007): Governance in der Schulentwicklung. Von der Autonomie zur evaluationsbasierten Steuerung. Wiesbaden: VS Verlag.
- Höhne, Thomas/ Schreck, Bruno (2009): Private Akteure im Bildungsbereich. Eine Fallstudie zum schulpolitischen Einfluss der Bertelsmann Stiftung am Beispiel von SEIS. Weinheim/München: Juventa.
- Kieser, Andreas/ Ebers, Mark (Hrsg.) (1999): Organisationstheorien. Stuttgart: Kohlhammer.
- Kussau, Jürgen/ Brüsemeister, Thomas (2007): Educational Governance: Zur Analyse der Handlungskoordination im Mehrebenensystem der Schule. In: Altrichter, Herbert/ Brüsemeister, Thomas/ Wissinger, Jochen (Hrsg.): Educational Governance. Handlungskoordination und Steuerung im Bildungssystem. Wiesbaden: VS Verlag, S. 15-54.

- Radtke, Frank-Olaf (2009): Ökonomisierung. In: Andresen, Sabine (et al.) (Hrsg.): Handwörterbuch Erziehungswissenschaft. Weinheim/Basel: Beltz, S. 621-636.
- Rürup, Matthias/ Heinrich, Martin (2007): Schulen unter Zugzwang – Die Schulautonomiegesetzgebung der deutschen Länder als Rahmen der Schulentwicklung. In: Altrichter, Herbert/Brüsemeister, Thomas/Wissinger, Jochen (Hrsg.): Educational Governance. Handlungskoordination und Steuerung im Bildungssystem. Wiesbaden: VS-Verlag. S. 157-184.
- Schnell, Herbert (2007): Schulprogramm, Evaluation und sich ändernde Bildungslandschaften. In: Schöning, Wolfgang (Hrsg.): Spuren der Schulevaluation. Zur Bedeutung und Wirksamkeit von Evaluationskonzepten im Schulalltag. Bad Heilbrunn: Klinkhardt, S. 21-28.
- Weerts, Freer (2007): Das niederländische Inspektionsmodell. In: Schöning, Wolfgang: Spuren der Schulevaluation. Bad Heilbrunn: Klinkhardt, S. 42-53.