

Thomas Höhne

# Evaluation als Wissens- und Machtform

**Text:**

Im vorliegenden Papier wird Evaluation als ein komplexes Kontrollwissen aufgefasst, dessen zentrale Funktion in der Optimierung von Steuerungsleistungen liegt. Dies beinhaltet vor allem die Kontrolle von Kausalzusammenhängen, die aber – so die konstruktivistische These – erst durch das Instrument Evaluation hergestellt werden. Kontrolle wird dabei als ein wesentlicher Faktor von Macht in Informations- bzw. Wissensgesellschaften aufgefasst. Jede Gesellschaft entwickelt historisch ihre eigenen Kontrollmittel für ein geeignetes soziales Kausalitätsmanagement zur Systemsteuerung. ‚Evaluation‘ stellt dabei ein Steuerungsinstrument dar, das im Kontext des Taylorismus und der positivistisch orientierten Testpsychologie zuerst in den USA entwickelt wurde und sich als tayloristisches Steuerungsinstrument in den 60er Jahren gesellschaftlich verallgemeinerte. In vier Phasen wird der Evaluationsdiskurs rekonstruiert und es wird danach gefragt, welche Funktion das tayloristische Steuerungsmittel Evaluation heute in Zeiten eines neoliberalisierten Postfordismus besitzt. So lässt sich am Beispiel von Evaluation zeigen, dass die neoliberale Transformation des Staates im Kern mit der weitreichenden Durchsetzung repressiv-restriktiver Mittel des Ausschlusses und der rigiden Standardisierung zum Zweck der Selektion einhergeht. Was Evaluation betrifft, so besteht der Wandel darin, dass es sich von einem politischen Mittel, das ehemals zur Bewertung von Sozialprogrammen im politischen und pädagogischen Bereich (Aufhebung von Bildungsungleichheit) eingesetzt wurde, hin zu einem reinen Ökonomisierungs- und Standardisierungsinstrument entwickelt hat. Als zweites wichtiges Element lässt sich beobachten, dass Evaluation tief in sozialen Praktiken der Subjekte verankert ist und dort ‚Haltungen‘ kreiert, durch welche die Subjekte sich permanent selbst evaluieren und sich unter Selbstbeobachtung stellen. Ganz im Sinne der Foucaultschen Analysen zur Gouvernementalität zeigt sich hier eine Koformierung öffentlich-staatlicher Praktiken und Subjektivierungsweisen im Neoliberalismus.

**Autor:**

Dr. Thomas Höhne  
thomas.hoehne@erziehung.uni-giessen.de  
<http://www.erwachsenenbildung.uni-giessen.de/hoehne>

**Zitation:**

Dieser Text ist in der Schriftenreihe der eb.giessen veröffentlicht worden.  
Sie finden diese in der virtuellen Bibliothek der Universität Giessen.

[http://geb.uni-giessen.de/geb/portal/eb\\_giessen/](http://geb.uni-giessen.de/geb/portal/eb_giessen/)

Thomas Höhne: *Evaluation als Wissens- und Machtform*. [Abrufbar unter: [http://geb.uni-giessen.de/geb/portal/eb\\_giessen](http://geb.uni-giessen.de/geb/portal/eb_giessen) Gießener elektronische Bibliothek].

## Evaluation als Wissens- und Machtform

### Problemaufriss und Perspektive

Warum wird in einer Gesellschaft, in einer Organisation oder Gruppe Evaluation betrieben? Die Antwort scheint auf der Hand zu liegen: Um das Bessere vom Schlechteren unterscheiden zu können, um knappe Ressourcen effizienter einsetzen zu können oder um Leistung messen zu können. Demnach ist es das Ziel von Evaluation herauszufinden, welche Auswirkungen ein Handeln oder eine Maßnahme hat (Wirkungsannahme) und in welcher Weise *Wirkungen* gesteigert werden können (Effektivitätskriterium). Dies schließt wiederum die Annahme ein, dass das vorliegende Wissen noch verbesserungs- und steigerungsfähig ist. Doch mit diesen Antworten treten neue Fragen auf: Welches sind die Kriterien von Evaluation? Wie lässt sich ‚Besseres‘ und ‚Schlechteres‘ unterscheiden? Was bedeutet Verbesserung? Existieren nicht schon immer allgemeine, soziale und normative Vorstellungen über Besseres/Schlechteres oder Effektiveres/Ineffektiveres, die in das Evaluationswissen unabdingbar einfließen? *Welches Wissen* wird mithilfe von Evaluation generiert und welches wird dabei vorausgesetzt, das selbst im Evaluationsdiskurs nicht mehr weiter thematisiert oder problematisiert wird?

Erst ein historisch-genealogischer Blick auf die Voraussetzungen von Evaluationswissen verrät etwas über dessen Komplexität und darüber, in welcher Weise es als wirksames Mittel für Modernisierung und Rationalisierung fungiert. Denn Evaluation stellt ein Instrument dar, politisches, soziales und ökonomisches Handeln effektiver zu machen, d.h. es zu rationalisieren. Die folgenden Ausführungen setzen daher an dem zweiten Fragenkomplex an, mit dem eine *diskursanalytisch-epistemische Perspektive* auf Evaluation angestrebt wird. Dies bedeutet, dass die Prämissen

und die Art des generierten Evaluationswissens systematisch sowie historisch beleuchtet werden sollen.

Einen Ausgangspunkt stellt die Annahme dar, dass es sich bei Evaluation nicht nur um eine Technik oder Methode zur Rationalisierung von Entscheidungen handelt, die schlechter oder besser eingesetzt werden kann oder die ideologisch ge- oder missbraucht wird. Vielmehr ist die Frage nach der Art und den Voraussetzungen des Evaluationswissens jenseits ideologiekritischer Unterstellungen von Bedeutung. Warum? Weil sich erst auf diese Weise die mehr implizit vorhandenen Vernetzungen mit anderen Wissensformen herauspräparieren lassen und Machtfunktionen und –konstellationen sichtbar werden. So können etwa Bürokratisierung und industriegesellschaftliche Organisation der kapitalistischen Produktion als wichtige, der Evaluation vorausgehende Formen der Rationalisierung angesehen werden<sup>1</sup>. Jede Form herrschenden sozialen Wissens impliziert, wenn auch auf unterschiedliche Art und Weise, einen Machtanspruch, der nicht unabhängig von den Trägern, den Verfechtern dieses Wissens, d.h. den sozialen Akteuren existiert. Somit ist diese Sichtweise notwendig mit der Frage nach der sozialen und politischen Legitimation dieses Wissens verknüpft. Da es zu den Merkmalen (post)modernen Wissens gehört, keine klaren Linien der Differenzierungen und Positionierungen zu schaffen, an denen sich die Geister eindeutig scheiden oder respektive die Akteure sichtbar unterscheiden, ist es lohnend, die komplexen, ineinander „verflochtenen Diskursverwurzelungen“ zu entwirren und in nachvollziehbare Linien zu rekonstruieren.

Ein weiterer Aspekt der hier zu eröffnenden diskursanalytischen Perspektive stellt die These dar, dass nicht Subjekte sozusagen frei über Wissen verfügen und sich autonom eine Ordnung schaffen. Vielmehr konstituiert umgekehrt das Wissen qua Formierung einer bestimmten Diskurs- und Wissensordnung die Subjekte, die in einer imaginären Verkehrung sich selbst als Urheber, Inhaber und Herrscher dieses Wissen wähnen. Daraus ergibt sich, dass außer den *Prämissen*, die jedes historisch rele-

---

<sup>1</sup> Weiter unten soll gezeigt werden, wie das moderne Disziplinwissen des 19. Jahrhunderts, wie es Foucault herausgearbeitet hat, und der Taylorismus wesentliche Vorstufen zur modernen Evaluation darstellen, die sich als explizites Wissens erst in der zweiten Hälfte des 20. Jahrhunderts entwickelt hat.

vant gewordene und institutionalisierte Wissen sowie neben den *Machteffekten*, die dieses Wissen auch immer auszeichnet, seine *subjektkonstitutive Funktion* von zentraler Bedeutung. Diese drei Aspekte – Prämissen, Macht und Subjektconstitution – bilden wichtige Eckpunkte der folgenden Analyse, die methodisch durch die Diskursanalyse realisiert werden soll<sup>2</sup>.

## **Evaluation als komplexes soziales Wissen**

Fasst man Evaluation zunächst allgemein als „Bewertung“ oder „Einschätzung“, dann kann sie sich auf ganz unterschiedliche Gegenstände beziehen, denn alles lässt sich in irgendeiner Form evaluieren (Widmer 2000: 78). Evaluation stellt ein Beobachtungswissen dar, das systematische Kriterien für die Bewertung von Dingen, Situationen, Handlungen, d.h. anderem Wissen enthält. Dass Evaluation durchaus etwas mit Alltagserfahrung zu tun hat, darauf hat beispielsweise Herbert Grüner hingewiesen:

„Jeder Mensch betreibt in seiner Erfahrungswelt ständig Evaluation, indem er auf der Grundlage kognitiver und affektiver Kriterien Gutes von Schlechtem und Sinnvolles von Sinnlosem trennt“ (Grüner 1993: 29).

So stellt sich Evaluation ein Teil von Alltagserfahrung dar, die schon aufgrund des Umstandes, dass Kriterien für Bewertung vorliegen, systematisierte Erfahrung und in dem Sinne Beobachtung<sup>3</sup> ist. Beobachtet und bewertet man sich selbst oder andere, dann entsteht durch permanente Rückkopplung ein Kreislauf der Kontrolle, aufgrund dessen dann Veränderungen erreicht und Interventionen bestimmt werden können. Dabei nimmt man Wirkungen und zugleich die Ursachen dieser Wirkung wahr. Dies deutet auf einen spezifisches Kausalitätsverständnis hin, nämlich einen systemisch-kybernetischen Beobachtungszusammenhang, der mit jeder

---

<sup>2</sup> Für weitere vertiefende methodologische Überlegungen zur Diskursanalyse, die hier nicht geleistet werden können, vgl. Höhne 2003a.

<sup>3</sup> Der Begriff der Beobachtung, wie der hier verwendet wird, setzt eine Systematisierung von Wissen und Erfahrung voraus. Er bezieht auf direkte personale Beobachtung – etwa im Alltag – oder kann auch im Zuge der medialen Expansion apersonale, soziale Beobachtungsmedien einschließen. So stellt jede Statistik in einer Zeitung oder ein Bericht über gute und schlechte Universitäten eine soziale Beobachtung respektive ein soziales Beobachtungswissen dar, mit dem sich eine Gesellschaft beurteilt.

Evaluationslogik implementiert wird. Diese besteht in der Erzeugung von Kontrollkreisläufen in Form von Rückkopplungsschleifen, die je nach Beobachtungsrichtung selbst- oder fremdregulierend eingesetzt werden können.

Im grundlegenden Sinne geht es um den gesellschaftlichen Umgang mit bzw. die soziale Konstruktion von Komplexität. So kann der Begriff der Kausalität in einer prinzipiellen Weise auf die Art bezogen werden, in der in einer Gesellschaft Zusammenhänge überhaupt erst hergestellt und als legitimes Wissen akzeptiert werden. Kommt der Regen von den Göttern, werden die Kinder vom Storch gebracht, entsteht Persönlichkeit durch Erziehung, steigen die Preise aufgrund mangelnder Nachfrage oder mangelt es an Nachfrage aufgrund der schon zu hohen Preise? Die Vorstellungen über Kausalität haben sich vor allem mit der Entwicklung des philosophischen und naturwissenschaftlichen Wissens gewandelt (vgl. von Saldern 1991: 130-148) und mithin haben sich auch die Praktiken der Einwirkung und das Wissen darum verändert. So lässt sich ‚Erziehung‘ als spezifische Einwirkungspraktik auffassen: Wie kann man erzieherisch auf Kinder einwirken, damit sie zu politisch und ökologisch bewussten Bürgern heranwachsen? Wo liegen die Grenzen von Verständnis für das Erreichen pädagogischer Ziele? Wenn Schülern zuviel Freiheit gewährt wird, geht ihnen dann Orientierung verloren? In diesen und anderen Fragen sind Kausalitäts- und Wirkungsunterstellungen impliziert (oder expliziert), die soziale Vorstellungen über Machbarkeit und Effektivität von Erziehung repräsentieren. Betrachtet man Erziehung daher als eine Art soziales Kausalitätsmanagement, so erfordert dies ein spezielles Wissen um die Einflussmöglichkeiten, Wirkungen, Methoden und Ziele, die zu erreichen man bestrebt ist. Mit welchen Mitteln können die Wirkungen erzielt werden? Wie können sie verbessert, gesteigert und optimiert werden? Auf dieser Ebene des Beobachtungswissens geht es um systematische Wirkungsüberprüfung und stete Verbesserung dieses Wissens sowie der damit verbundenen Wirkungen. Kurzum: *Es geht in einem grundlegenden Sinne um Evaluation und um gesellschaftlich akzeptierte Vorstellungen*

von *Kausalität, Komplexität und Einwirkung*<sup>4</sup>. Dieses soziale Kausalitätsmanagement kann als eine implizite Form von Evaluation insofern aufgefasst werden, als es bei Evaluation immer auch um Wirkungskontrolle geht. In der weiter unten folgenden historischen Rekonstruktion von Evaluation als einer spezifischen Art des pädagogisch-psychologischen Kontrollwissens wird auf diesen Punkt der Evaluation noch genauer eingegangen.

Selbst- und Fremdregulierung stellen ein zentrales Mittel der Rationalisierung und Subjektformierung der modernen Gesellschaft dar, wie dies Max Weber für den Zusammenhang von Kapitalismus und Religion und Norbert Elias auf der Ebene der Psycho- sowie Soziogenese bezüglich der Regulierung des individuellen Trieb- und Affekthaushalts gezeigt haben. Diese Beobachtungen 1. Ordnung<sup>5</sup> erfahren aufgrund der wissenschaftlichen und informationell-medialen Expansion im 18. und vor allem 19. Jahrhundert eine Steigerung und erhalten so den Status eines systematischen, institutionellen Beobachtungswissens 2. Ordnung, wie es sich in Statistiken, Akten und anderen Systematisierungspraktiken äußert. Dieses systemische Wissen ist emergent, da es eine neue Strukturebene bildet, auf der Informationen nach spezifischen Regeln und Verfahrensweisen systematisiert und ausdifferenziert werden. Die moderne europäische Gesellschaft zeichnet sich seit dem 18. Jahrhundert durch ein evaluatives Beobachtungswissen aus, das sich zunehmend verfeinert und ausdifferenziert hat. In diesem Sinne kann sie auch als „Evaluationsgesellschaft“ (Prondczynsky 2001: 92, Fn 2) bezeichnet werden.

Legt man ein solch weites Verständnis von Evaluation zugrunde, so ermöglicht dies, den historischen Zusammenhang von Subjekt, Rationalität und Methode in Bezug auf Evaluation im Auge zu behalten. Dabei stellt Evaluation nicht nur irgendeine wissenschaftliche Methode zur Bewertung von Prozessen dar, sondern beinhaltet als spezifische Wissensform historisch ein bestimmtes Rationalisierungs- und Disziplinierungsniveau auch

---

<sup>4</sup> Erst Anfang der 80er Jahre wurde in grundlegender Weise das Technologiedefizit der Pädagogik zum Thema gemacht (Luhmann/Schorr 1982).

<sup>5</sup> Beobachtungen 1. Ordnung tauchen historisch als moralisch-ethische Ordnungen auf, die eine systematische Lebensführung (Ziele, Planungen, Programme usw.) ermöglichen. Dazu gehören die umgangssprachlich in ihrer Bedeutung für diesen Prozess zu unrecht abgewerteten Sekundärtugenden.

auf Seiten der Subjekte selbst<sup>6</sup>. Es geht also um eine Verklammerung von subjektiver Rationalisierung etwa in Form methodischer Lebensführung auf der Ebene von Praktiken und objektiver Rationalisierung in technisch-wissenschaftlicher, politischer und ökonomischer Hinsicht.

Hierbei ist noch notwendig der Aspekt der Systematisierung der Bewertung hervorzuheben, wodurch aus der einfachen Alltagswahrnehmung eine *strukturelle Beobachtung* wird, durch die nachhaltig ein Verhalten beobachtet, systematisiert, kategorisiert und schließlich geändert werden kann. In diesem Sinne erfährt ein derart angesammeltes (Selbst)Beobachtungswissen in der wissenschaftlichen, medialen und institutionellen Systematisierung eine Steigerung – oben als Wissen 2. Ordnung bezeichnet –, wodurch es erst zu einem wichtigen Medium der Rationalisierung wird<sup>7</sup>. Der Medienbegriff hat hierbei die doppelte Bedeutung von Mittel der Beobachtung und Medium als die Form des Wissens – so etwa, wenn das Wissen in Buchform oder elektronischer Form als Formel, Tabelle, Statistik, Text usw. auftritt.

In jedes Evaluationswissen gehen soziale und normative Vorstellungen ein und es ist grundlegend von politischen und ökonomischen Kontexten abhängig (z.B. Stockmann 2000, Grüner 1993: 37). Entsprechend spiegeln sich unterschiedliche Logiken im Evaluationswissen wider, dessen Gegenstände und Formen wiederum in Abhängigkeit des Feldes, in dem evaluiert wird, vielfältig sein können:

„Unter dem Begriff der Evaluation subsumieren sich die unterschiedlichen Theorien, Konzepte und Verfahrensweisen und auch im wissenschaftlichen Sprachgebrauch werden darunter verschiedene Gegenstände verstanden. So finden wir z.B. in den Wirt-

---

<sup>6</sup> In institutioneller Hinsicht hat dies Foucault mit dem Begriff der Disziplin beschrieben, die sich historisch als Gefängnis, Schule, psychiatrische Anstalt usw. durchsetzt (Foucault 1994).

<sup>7</sup> Es soll nicht behauptet werden soll, dass Alltagsevaluation primär gegenüber wissenschaftlicher Evaluation sei oder dass sie als subjektiver und objektiver Ableger einer einheitlichen Rationalisierungsstufe fungierten. Vielmehr korrelieren sie in einer *bestimmten Weise*, nämlich dort, wo wissenschaftliche Evaluation die *Möglichkeit der Verhaltensänderung und –kontrolle auf Seiten der Subjekte voraussetzt und dies systematisch in dem Wissen, das durch Evaluation generiert wird, mit berücksichtigt wird*. Mit anderen Worten: Wissenschaftliche Paradigmen implizieren, um historisch wirksam zu werden, eine bestimmte Form(ierung) des Subjekts und eine Durchsetzung des Wissens auf der Ebene institutioneller oder alltäglicher Praktiken. In dem Sinne kann das System des Scientific Management nach Frederic Taylor als ein umfassendes Evaluationsprojekt aufgefasst werden, auch wenn es nicht diesen Namen trägt. Ein entsprechende Beobachtungs- und Kontrollwissen findet sich in unterschiedlichen Bereichen und institutionellen Praxisformen, ohne entsprechend eingeordnet zu werden. Dabei geht es um die Ähnlichkeit der Praktiken, die zu untersuchen ist.

schaftswissenschaften die Begriffe Kosten/Nutzen bzw. Effizienzmessung, in der Psychologie die Wirkungsforschung oder in der Pädagogik u.a. die Lernerfolgskontrolle“ (Grüner 1993: 29).

So folgt etwa die „Messung der Wertertragssteigerung des Produkts A zum Zeitpunkt X gegenüber Zeitpunkt Y“ als ökonomischer Gegenstand einer anderen Evaluationslogik als der Gegenstand „Der Einfluss von Gewalttrainings auf die psychosoziale Entwicklung von männlichen Jugendlichen“. Dies klingt trivial, doch zeigt sich an diesen Beispielen, dass die Frage der Gegenstandskonstitution in der Evaluationsforschung für die Abschätzung der Adäquatheit der Methoden und der unterschiedlichen Instrumente von elementarer Bedeutung ist. Gemeinhin entzündet sich daran auch der klassische wissenschaftsinterne Streit um qualitative und quantitative Methoden (vgl. Stockmann 2000. Lee 2000). Grob können hierbei Leistungsevaluation und Wirkungsevaluation als zwei grundlegende Formen unterschieden werden. In beiden Fällen stehen am Anfang normative Vorstellungen, die sich in Gestalt eines Programms und Zielen oder Leistungs- und Effizienzvorstellungen ausdrücken. Der zentrale Unterschied besteht darin, dass Evaluation als Wirkungsforschung danach fragt, *ob* vermutete oder erhoffte Wirkungen überhaupt erzielt wurden, während Leistungsevaluation (erfolgte oder angenommene) Wirkungen schon immer voraussetzt, wenn statisch der ‚Leistungsstand‘ oder Effizienz gemessen werden. Zudem wird auch kritisch vermerkt, dass nicht Leistung, sondern Erfolg gemessen werde, von dem aus auf Leistung zurück geschlossen werde (Hofacker 2000: 98).

Der Blick bei der Funktionsbestimmung von Evaluationsformen ist nicht nur auf die verschiedenen Bezugsdisziplinen und Kontexte zu richten, da die Frage nach der Art des Wissens, das durch Evaluation als spezifische epistemische Praktik generiert wird, von grundlegendem Stellenwert für eine Theorie der Evaluation (vgl. Lee 2000). Hierzu ist eine historische Perspektive auf die Genealogie der Evaluation als moderne, wissenschaftlich systematisierte Beobachtungstechnologie notwendig, denn erst durch sie wird deutlich, in welchem Ausmaß Evaluation schon immer in umfassendere wissenschaftliche Paradigmen bzw. allgemeine

diskursive Formationen eingebunden (gewesen) ist und welche Veränderungen Evaluationswissen dabei durchlaufen hat (vgl. Lee 2000: 141 ff.).

Wenn von Wissensformen und epistemischen Praktiken die Rede ist, so wird in mehrfacher Weise eine an Michel Foucault orientierte Perspektive eröffnet. Zum einen wird das Wissen und die Art seiner Formierung *im Diskurs zum Objekt der Analyse*, zum anderen gilt es, unabdingbare Verknüpfung von Wissen mit Macht zu untersuchen, d.h. die Machteffekte, die dominante Wissensformen und Wissenspraktiken produzieren. Hier zeigt sich, dass etwa die gouvernementalitätstheoretischen Studien, die in Anschluss an Foucaults historische Untersuchungen zur Analyse des Neoliberalismus durchgeführt wurden, von einigem Interesse sind (vgl. Bröckling/Krasmann/Lemke 2000). Sie erweitern den Blick noch einmal wesentlich um die *staatlich-politische Dimension*, die in der Evaluationsforschung wenig thematisiert wird. In dieser Hinsicht zeigen sich zwei interessante Untersuchungsrichtungen. Zum einen macht die historische Analyse deutlich, in welcher Weise Evaluation als umfassende politisch-staatliche Technologie im Zusammenhang mit sozialstaatlichen Reformprogrammen in den USA entstanden ist. Zum anderen spielt Evaluation auch im Kontext der gesellschaftlichen Verallgemeinerung von Kontrolltechnologien eine Rolle, die eine informationelle Kontrolle weiter Teile der Gesellschaft erlauben (Beniger 1986). Zu diesem Set an objektivierenden Kontrolltechniken, die sich vor allem in den 90er Jahren über das Vordringen neoliberaler Praktiken gesellschaftlich verallgemeinert haben, gehören beispielsweise Total Quality Management, Audits, Qualitätszirkel, Supervision, Mediation usw. Evaluationswissen kann daher als ein *systematisches kybernetisches Kontrollwissen* begriffen werden, durch das in Form von Feedbackschleifen Punkte und Ausmaß möglicher Interventionen bestimmt werden.

Die Repräsentationsformen, in denen Evaluation erscheint, sind wesentlich für die Analyse der Art der Rationalität, die ihnen unterliegt. So eröffnet beispielsweise jede Statistik einen möglichen Raum für politisches Handeln (Miller/Rose 1994: 58), so wie sie ihrerseits eine *Form des Wissens* darstellt, das auf bestimmten Prämissen und Voraussetzungen beruht. Dies bedeutet, dass eine Statistik innerhalb einer Evaluation mehrere

Funktionen hat. Neben der Funktion, sogenannte Fakten quantitativ darzustellen, mobilisiert die Form der Repräsentation des Wissens (Kurven, Zahlen, Linien, X und Y- Geraden usw.) ein komplexes Statistikwissen auf der Rezipientenseite, das in schulischer Sozialisation bzw. schulischem Lernen langjährig angeeignet wurde (vgl. Höhne 2003). Impliziert ist dabei der Glaube an die Abbildung von Realität, der wissenschaftlich durch die Unterscheidung von harten und weichen Faktoren oder die bereits erwähnte quantitativ/qualitativ-Differenz genährt wird. Dies bedeutet, dass das Evaluationswissen selbst auf seine *Repräsentationsformen* hin untersucht werden muss, um die Wirkungen, die es auf Subjektseite erzeugt, zu analysieren. Zu diesem Komplex von Funktionen gehört auch die Normalisierung, d.h. es werden aufgrund wissenschaftlich erzeugter Daten bestimmte Gegenstände hervorgebracht, die als legitimes, relevantes und beforschbares Wissen gelten.

## **Definitionen und begriffliche Bestimmungen von Evaluation**

Einleitend war angedeutet worden, dass Evaluation sich begrifflich auf die beiden Ebenen 1) der Alltagserfahrung als mehr oder minder systematisch betriebene individuelle Selbst- und Fremdkontrolle und 2) auf die systematisch-wissenschaftliche Bewertung und Kontrolle von Prozessen beziehen kann. Auf der zweiten Ebene erfolgt durch Institutionalisierung in Form bereichsspezifischer, verfestigter und regelhafter sozialer Praktiken eine gesellschaftliche Verallgemeinerung dieses Beobachtungswissens:

„Der Vorgang, Dinge, Systeme und Prozesse zu bewerten, wird auch im öffentlichen Leben verstärkt betrieben und findet in Bereichen wie z.B. der Wirtschaft, der öffentlichen Verwaltung oder dem Gesundheits- und Bildungswesen verstärkt Eingang“ (Grüner 1993: 29).

Dergestalt manifestiert sich das Beobachtungs- und Kontrollwissen auf einer eigenen staatlich-institutionellen Ebene, auf der es normalerweise definiert wird.

So wird Evaluation etwa zur Verbesserung pädagogischer Praxis verstanden als die „systematische Sammlung von Daten, die als Erfolgs- o-

der Misserfolgsindikatoren für ein pädagogisches Programm interpretiert werden in der Absicht, die pädagogische Praxis zu verbessern“ (Rhyne 2001: 187). Schließlich wird auch die Dimension pragmatisch-programmatischer Entscheidungen bei der Definition von Evaluation hervorgehoben: „Evaluation is the systematic investigation of the merit or worth of an object (programm) for the purpose of reducing uncertainty in decision making“ (Mertens nach Stockmann 2000a: 13). Eine umfassende forschungspraktische Bestimmung von Evaluation geben Peter Rossi und Howard Freeman, wenn sie Evaluation definieren als

„systematische Anwendung sozialwissenschaftlicher Forschungsmethoden zur Beurteilung der Konzeption, Ausgestaltung, Umsetzung und des Nutzens sozialer Interventionsprogramme. Evaluationsforschung bezeichnet den gezielten Einsatz sozialwissenschaftlicher Forschungsmethoden zur Verbesserung der Planung und laufenden Überwachung sowie zur Bestimmung der Effektivität und Effizienz von (...) sozialen Interventionsmaßnahmen“ (ebd.).“

Evaluation wurde im Kontext staatlich-politischer Planung zum ersten Mal großflächig eingesetzt und ist daher eng verknüpft mit dem Steuerungsoptimismus der 60er und 70er Jahre. Sie ist Teil eines komplexen Kontrollwissens, das von Seiten des Staates als panoptische Technologie akkumuliert wurde. In unterschiedlichen Bereichen wurden „Massenfeedbacktechnologien“ (Beniger 1986: 20) wie etwa Meinungsumfragen, Konsumentenforschung, Massenkommunikationsforschung oder psychologische Eignungstests etabliert, die ein ungeheures Daten- und Informationsnetz bildeten. Die dadurch erfolgende Datenexplosion ist ein zentraler Ausgangspunkt der Informationsgesellschaft (ebd.: 21), in der technisch-mediale Entwicklungen wesentlich zu den Möglichkeiten, die Daten zu erzeugen, zu verwalten und auszuwerten, beitrugen.

### **Evaluation als panoptische Kontrolltechnologie**

In der Definition von Rossi/Howard spiegeln sich die Vorstellungen einer technokratisch-bürokratischen Planungsrationalität wider, die durch den aktiven Staat bzw. eine „aktive Politik“ (Mayntz 1997:68) sowie durch eine „aktive Gesellschaft“ (Etzioni 1968) erreicht werden sollte. Dabei löst sich

der Staat aber nicht einfach auf, sondern dringt tiefer in die Gesellschaft ein, indem staatlich-politische – und zunehmend ökonomische – Praktiken sich gesellschaftlich verallgemeinern, veralltäglichen und normalisieren<sup>8</sup>. Dies zeigt sich besonders deutlich anhand der Evaluation, die sich mittlerweile quer zu den unterschiedlichen sozialen Bereichen in allen Organisationen wieder findet. Angefangen bei staatlichen Verwaltungen, Schulen, Kindergärten, sozialen Einrichtungen bis hin zu Sportvereinen wird mittlerweile alles evaluiert. Aber auch die individuellen Lebensläufe, Erwerbs- und Bildungsbiographien bleiben von der allgemeinen Evaluationswelle nicht verschont, denn auch pädagogische Konzepte wie lebenslanges Lernen und Kompetenzerwerb stellen Selbstevaluations-techniken auf individueller Ebene dar.

Damit soll, wie Anfangs schon angedeutet, eine Haltung der steten Selbstüberprüfung der Subjekte erzeugt und als allgemeine Lebenspraxis etabliert werden, bei der die Differenz von ‚privat/öffentlich‘ aufgehoben ist. Angesichts der ‚strukturellen Entstrukturierung‘ und des Abbaus institutioneller Sicherheiten ist eine solche evaluative Selbstbeobachtung (Wird das richtige Wissen angeeignet? Wird heute eine Entscheidung getroffen, die sich morgen als falsch herausstellt? usw.) eine Reaktion auf die Individualisierung sozialer und vor allem ökonomischer Risiken, da die Verantwortung für die positiven oder negativen Risikoeffekte auf den Einzelnen verschoben wird (Bröckling/Krassmann/Lemke 2000). Hier können evaluative Praktiken als Selbsttechnologien wesentlich dazu beitragen, soziale Distinktion und Selektion zu legitimieren und sie mit zu erzeugen<sup>9</sup>, indem

---

<sup>8</sup> Es ist auf keinen Fall anzunehmen, dass der Staat im Neoliberalismus einfach verschwindet oder gar ‚abstirbt‘. Dies würde bedeuten, dass ‚der Staat‘ lediglich den Subjekten ‚aufgesetzt‘ sei und nicht auch ihre Praktiken bestimmen würde. Diese Vorstellung ist ein Effekt des neoliberalen Diskurses selbst, der ‚den Staat‘ als etwas Externes und Fremdes darstellt, der die Autonomie und Freiheit der Individuen nur einschränke und bürokratisch gängele, wo er nur könne. Demgegenüber ist von einer notwendigen (zivil)gesellschaftlichen Verankerung des modernen Staates auszugehen, der die Subjekte auch in alltäglichen Handlungen prägt. Hier könnte man vom impliziten oder „unsichtbaren“ Staat sprechen (Leisner 1994). Der oft vermutete ‚Abbau‘ oder ‚Rückzug‘ ist eher als Transformation des Staates zu deuten und Evaluation stellt hierbei eine von staatlich-bürokratischen Stellen ursprünglich eingeführte Kontrollpraktik dar, die sich gesellschaftlich verallgemeinert hat.

<sup>9</sup> Mit Evaluation aufgrund von Makrodaten wird normalerweise nicht nach sozialen Hintergründen oder ökonomischen Umständen für Leistung gefragt, sondern Leistung wird im abstrakten Sinne individualisiert, indem gerade von sozialen Bedingungen abgesehen wird (etwa was im universitären Bereich die Zahl und Dauer von Abschlüssen betrifft). Dies macht deutlich, dass Evaluationsmodelle stets auch daraufhin zu prüfen sind, was sie *nicht erheben, dethematisieren und ausblenden*, um die (Macht)Effekte einschätzen zu können, die unweigerlich damit einhergehen.

soziale und ökonomische Risiken den Individuen zugeschrieben und naturalisiert werden, da ‚das Soziale‘ der Risiken nicht mehr hervortritt.

Zu den Effekten von Evaluation als einer Technologie systematischer, institutioneller Dauer(selbst)beobachtung gehört, dass durch sie a) Selbstbeobachtungs- und Selbstüberwachungspraktiken auf Seiten der Individuen induziert werden und b) der Rechtfertigungsdruck auf die Einzelnen erhöht wird, ständig effektiv zu handeln. Hierdurch vollzieht sich eine totalitäre Normalisierung von Kontrolltechniken, da sich im gewissen Sinne die panoptische Utopie durch Dauerbeobachtung vollendet. Das von Jeremy Bentham ersonnene Panoptikum repräsentiert die (mögliche) allumfassende Beobachtung 1) aller von einem Beobachter im Zentrum, der aber selbst nicht gesehen werden kann sowie 2) eines jeden durch den anderen innerhalb des Gefängnisses, für das er diese Überwachungstechnologie geplant hatte. Es bildet aber ein „verallgemeinerungsfähiges Funktionsmodell (...), das die Beziehungen der Macht zum Alltagsleben der Menschen definiert“ (Foucault 1994: 263). So wird es auch als „Programm“ bezeichnet, das eine „individualisierende Beobachtung“, „Charakterisierung und Klassifizierung“ und eine „analytische Aufteilung des Raums“ (ebd.: 261) ermögliche:

„Das Panopticon ist eine Maschine zur Scheidung des Paares Sehen/Gesehenwerden: im Außenring wird man vollständig gesehen, ohne jemals zu sehen; im Zentralturm sieht man alles, ohne je gesehen zu werden. Diese Anlage ist deswegen so bedeutend, weil sie die Macht automatisiert und entindividualisiert“ (ebd: 259).

Foucault weist darauf hin, dass die Wirkung der Überwachung permanent sei, auch wenn ihre Durchführung sporadisch sei. Dies vollzieht auf der Seite der Subjekte in Form der Internalisierung von Machtmitteln, wodurch es zum Prinzip der eigenen Unterwerfung werde (ebd.: 260). So wird die Macht tendenziell unkörperlich und durchzieht netzförmig Institutionen und Gesellschaft. Unwillkürlich werde jeder zur Information, aber „niemals Subjekt in einer Kommunikation“ (ebd.: 257). Schließlich wird von Foucault festgehalten, dass sich das panoptische Prinzip grundsätzlich in jede Funktion integrieren lasse – sei es Erziehung, Heilung, Produktion oder Bestrafung (ebd.: 265). Es steigere die jeweilige Funktion nachhaltig und

trage zu deren gesellschaftlicher Verallgemeinerung bei. In dem Sinne repräsentiert das panoptische Modell also ein elementares Beobachtungsprinzip moderner, institutionell verfasster Gesellschaften und besitzt zugleich die Funktion eines Mediums, denn gesellschaftliche Kommunikations- und Rationalisierungsmöglichkeiten lassen sich nachhaltig durch entsprechende apersonale Beobachtungsmedien steigern (vgl. Fußnote 3). In ihnen ist ein Wissen – in Form von Schülerakten, Krankenberichten, wissenschaftlichen Texten, bürokratischen Regeln usw. sedimentiert –, durch das Subjekte konstituiert werden. Individuen werden als Subjekte in Gestalt von „Fällen“, Typologisierungen, Definitionen von Rand- oder Teilgruppen usw., d.h. qua Diskurse konstituiert. Evaluation stellt eine (von vielen) panoptische Beobachtungstechnologien dar, durch die sich historisch, institutionell spezifisch und diskursiv ein *komplexes Wissen um Wirkungen und Kausalzusammenhängen* politisch-staatlichen Handelns akkumuliert wurde. Ein Blick auf die historische Entwicklung der Evaluation zeigt, dass sich dieses Wissen auch verändert.

## **Zur historischen Entwicklung des Evaluationsdiskurses**

Innerhalb der Evaluationsforschung werden verschiedene Phasen und Modelle unterschieden (z.B. Grüner 1993: 37, Prondczynsky 2001). So differenzieren etwa Prondczynsky, Grüner, Lachenmann und Stockmann zwischen mehreren Phasen der „Evolution der Evaluationsforschung“ (Prondczynsky 2001: 103-109). Hier sollen nur einige Schlaglichter der historischen Entwicklung der Evaluationsforschung genannt, aber einige wichtige Ereignisse hervorgehoben werden, wobei die diskurshistorischen Akzente etwas anders gesetzt werden<sup>10</sup>.

Ausgegangen wird von vier Phasen der Entwicklung des Evaluationsdiskurses. Die erste Phase, welche etwa die zweite Hälfte des 19. Jahrhunderts umfasst, stellt eine Vorphase des Evaluationsdiskurses dar. Dieser beginnt explizit um die Jahrhundertwende, und ist durch den Taylorismus sowie durch psychologische Messtechnologien gekennzeichnet (2.

---

<sup>10</sup> Die von Grüner (1993: 31 ff.) oder Stockmann (2000) präsentierten Phasenmodelle zeichnen zum Teil nur die immanente Entwicklung von Evaluation nach, sind aber im Unterschied zu dem hier präsentierten 4-Phasenmodell feiner in der Darstellung. Um die größeren Diskursbewegungen und z.T. komplexen Vernetzungen des Evaluationsdiskurses darzustellen, muss daher notwendig auf ein größeres Schema zurückgegriffen werden.

Phase). Anders als in den USA, wo Evaluation in Form von Lernzielüberprüfungen im Bildungsbereich von Ralph Tyler im dem 40er Jahren wesentlich Grund gelegt wurden – er gilt als „Vater der Evaluation“ (Grüner 1993: 34) – wurde in Deutschland erst in den 60er Jahren durch die Bildungsreform, die Förderung der Bildungsökonomie als eigenständiger Forschung und im Kontext der Curriculums- und Schulforschung der Rahmen für Evaluation geschaffen. Hierbei ist nicht zu vergessen, dass internationale Ereignisse wie der „Sputnickschock“ und internationale Organisationen wie die OECD und die UNESCO dazu beigetragen haben, Bildungssysteme und deren Entwicklung zum Topthema zu machen (3. Phase). Die vierte Phase kann in den 90er Jahren mit der gesellschaftlichen Durchsetzung neoliberaler Praktiken (z.B. New Public Management) angesetzt werden, in der Evaluation als Instrument von Leistungsmessung und Effizienzsteigerung eine Schlüsselstellung einnimmt. Die unterschiedlichen Diskursphasen stellen dabei keine streng abgrenzbaren Abfolgen dar, sondern bieten eine Heuristik, um die komplexen Konfigurationen unterschiedlicher Diskurse sichtbar zu machen, die sich in bestimmten Perioden verdichten.

## **1.Phase: Die erste Verdichtung des Kontrollwissens im 19. Jahrhundert**

Im Rahmen sich durchsetzender Institutionalisierungen im Sozial- und Bildungsbereich entwickelte sich im 19. Jahrhundert ein dichter Diskurs um die „psychologische und psychophysikalische Messung“ (ebd.: 105). So stellt Kahlert fest, dass speziell um die Wende vom 19. zum 20. Jahrhundert ein zunehmendes Interesse an unterrichtspsychologischen Untersuchungen „zum Thema Zeit- Kraftökonomie des Lernens und Lehrens“ (Kahlert 1978: 117) zu verzeichnen sei. Es erschienen Studien zur Ermüdung, zum Ablauf des Unterrichts, zur Lage und Dauer der Unterrichtsstunden, zur Fächerfolge und zur Funktion von Pausen. Diese erste frühe Phase der gezielten wissenschaftlichen Wirkungsermittlung von sozialen und erzieherischen Maßnahmen war getragen von einer positivistischen Empirievorstellung, die von einer engen, linearen Kopplung von Maßnahme und Wirkung ausgeht, was die Anfänge der Evaluationsforschung

kennzeichnet: „Vor allem in den Anfangsjahren der Evaluationsforschung herrschte ein positivistisch bestimmter methodologischer Rigorismus vor“ (Stockmann 2000a: 17). Dies hatte unter anderem zur Folge, dass über die festen *zeitlichen* und *räumlichen* Ordnungsmuster der Disziplinarinstitutionen – spezifisch die raumzeitliche Parzellierung (Gstettner 1981: 49 ff.) – hinaus eine *kausalmechanistische Einwirkungspsychologie* zum Zuge kam, wodurch ein eigenes „pädagogisch-psychologisches Tableau“ als Wissensfeld etabliert wurde: „Die Disziplinargesellschaft hat eine Theorie der Konditionierbarkeit vorbereitet und durch praktische Erprobung einen konkreten Beitrag zur ‚Mikrophysik der Macht‘ geliefert“ bemerkt Gstettner (ebd.: 61-62). Im Sinne Foucaults kann von einer „positiven Ökonomie“ gesprochen werden, aufgrund deren das explizite Strafsystem durch ein konditionierendes Signalsystem ersetzt wurde (ebd.: 53). Hierbei ist die Nähe der Pädagogik zur Ökonomie festzustellen und besteht weniger, wie oft vermutet wird (Stichwort „Drillschule“), in der Affinität zum Militär, denn zentral war dabei der ökonomische, aber nur pädagogisch umsetzbare Gedanke, das Subjekt als eigene Produktivkraft zu sehen<sup>11</sup>. Genuin ökonomische Techniken der zeitlich-räumlichen und kausalen Wirkungskontrolle dienten zur adäquaten Steuerung und Kontrolle von Entwicklung. Ökonomie ist hierbei als eine enge *Kopplung von Zeit, Raum und Kausalität zum Zweck der Effizienz- und Effektivitätssteigerung* zu sehen, was *Quantifizierung* und *Messung* von Variablen- und Wirkungszusammenhängen impliziert. Sie ist hierbei nicht im Sinne eines abgetrennten ökonomischen Bereichs zu denken, sondern als eine *spezifische Form der Rationalität*, die das Denken, Handeln und die Praktiken im Allgemeinen bestimmen. Evaluation ist hierbei als ein spezifisches Kausalitätsmanagement zu verstehen, mit dem Zusammenhänge konstruiert, gleichzeitig rationalisiert und auf diese Art pragmatisch hergestellt werden. Foucault

---

<sup>11</sup> Bildungstheorien enthalten, wenn man so will, auch einen bildungsökonomischen, am Humanen als Kapital orientierten Kern – neben dem religiösen der Selbsterschaffung des Subjekts durch Bildung. Bildung kann als ein fortgeschrittener Modus der Selbstreproduktion begriffen werden und es muss nicht einmal eine rein instrumentalistische Ausrichtung eines solchen Bildungsbegriffs unterstellt werden. Selbstreproduktion meint hierbei ein (Grund)Vermögen, das steigerungsfähig ist – dies wäre dann ein möglicher historischer Ort von Bildung, in der eine positive Ökonomie der eigenen Entwicklung mit der Entfaltung der eigenen Kräfte ununterscheidbar beieinander lägen. Schon ein kurzer Blick auf verschiedene Bildungstheorien lässt die Affinität einer solchen positiven Bildungsökonomie des Subjekts und dem klassischen Topos der Selbstentfaltung erahnen, wenn man sich Semantiken wie „Vermögen“, „Selbsttätigkeit“ (Vierhaus 1972: 515) oder „Verbesserung“ (ebd.: 516) anschaut.

hat darauf verwiesen, dass es bei der Analyse darum gehen müsse zu untersuchen, „wie Rationalitätsformen sich selbst in Praktiken oder Systemen von Praktiken einschreiben und welche Rolle sie in ihnen spielen“, da es keine Praktiken ohne „bestimmte Rationalitätsordnung“ gebe (Foucault nach Bröckling/Krasmann/Lemke 2000a: 20). Eine solche Rationalitätsordnung begann sich ab der Mitte des 19. Jahrhunderts zu entwickeln.

So gab es innerhalb der Pädagogik in der zweiten Hälfte des 19. Jahrhunderts eine starke Tendenz, empirische Beobachtungen auf naturwissenschaftliche Füße zu stellen – aus Gründen der Theoriesicherheit. Diese Entwicklung bildet das diskursive Vorfeld der Evaluation, die ein bestimmtes Verständnis von Kausalität, Wirkung und ihrer Beeinflussung impliziert. Die Kontingenz von Seele und Subjekt konnte auf Dauer kein ernsthafter Bezugspunkt für eine pädagogische Technologie sein, die sich wissenschaftlich absichern wollte. Daher wurden die äußeren und beobachtbaren Faktoren für die Genese eines entsprechend wissenschaftlich-positivistisch abgesicherten Wissens bedeutungsvoller, wobei als Untersuchungsgegenstand der *Körper gegen die Seele* ausgetauscht wurde. Diese quasi materialistisch orientierte, durch Experimente und Techniken abgesicherte Erkenntnisweise versuchte durch Präzision und dichte, empirische Beobachtungen aller denkbaren, äußeren Einflussfaktoren das Innere des Subjekts, d.h. die ‚black-box‘ möglichst klein zu halten und auf diese Art eine Ökonomie der Einwirkungen auf den Körper zu etablieren. Foucault hat darauf hingewiesen, dass die „Macht der Ökonomie auf einer ihr vorgängigen ‚Ökonomie der Macht‘ beruht“ (Bröckling/Krasmann/Lemke 2000a: 26). Für den ökonomischen Bereich bedeutete dies:

„Die ökonomische Ausbeutung erfordert eine ‚politische Besetzung des Körpers‘ (Foucault). Daher muss die Arbeitskraft zunächst einmal als Arbeitskraft konstituiert werden, bevor sie ausgebeutet werden kann: Lebenszeit muss in Arbeitszeit synthetisiert, die Individuen an den Ablauf des Produktionsprozesses fixiert und dem Zyklus der Produktion unterworfen werden; Gewohnheiten müssen ausgebildet, Zeit und Raum in feste Schemata eingefügt werden“ (ebd.).

Analog dazu lassen sich die wissenschaftlichen Untersuchungen zu Entwicklung, Lernen, Motivation usw. als pädagogisch-psychologische Besetzung des Körpers auffassen. Schließlich wird der pädagogisch formbare Körper, der Körper der Erziehung über die entsprechenden Praktiken erst hervorgebracht, denn mit jeder subtilen Einwirkung korreliert ein detailliertes Einwirkungswissen, d.h. ein Technologiewissen und die Fähigkeit des Kausalmanagements. Am Beispiel der sogenannten Überbürdungsdebatte lässt sich die zunehmende Kontrolldichte aufgrund mikrologischer Kausal- und Wirkungsanalysen demonstrieren, die dieses wissenschaftliche Beobachtungswissen charakterisieren.

Der Nexus von schulischem Lernen und Schülergesundheit wurde in der zweiten Hälfte des 19. Jahrhunderts unter dem Stichwort „Überbürdung“, d.h. Überforderung der Schüler intensiv diskutiert (vgl. Oelkers 1998: 248). In dezidierter und ziselierender Weise wurden räumliche, zeitliche, motorisch-körperliche und biophysiologische Bedingungen in der Schule untersucht. Dies reichte vom „Bedarf an Atemluft“ und Schadensvermeidung wie „Kurzsichtigkeit“ bis hin zur Erforschung des Einflusses von Schulmobiliar auf die Konzentrationsfähigkeit der Schüler. Dabei ging es zum einen um Gesundheit, aber auch untrennbar damit verknüpft, um Optimierungsstrategien wie etwa die Vermeidung „unnützer Muskelermüdungen“ (ebd.: 251). Ziel dieser postmetaphysischen Pädagogik war es, den Seelenbegriff durch naturwissenschaftlich-materialistische Theorien und positivistisch-empirische Forschung zu ersetzen. Pädagogik, Medizin und Physiologie waren im Diskurs eng verflochten. Dadurch kam es auch zu einer engen thematischen Kopplung von Lernen und Arbeit, denn beide stellen in individueller und sozialer Hinsicht die zentralen (Re)Produktionsfaktoren von Entwicklung dar. So stand logisch die Suche nach der Bestimmungsgröße im Fordergrund, mit der diese Entwicklung gemessen werden konnte. Kategorial ging es dabei um „Leistung“ und mithin um die Messung des entsprechenden Leistungsvermögens:

„(...) exakte Leistungsmessung stand am Ende der Überbürdungsdebatte. Die letzte große Diskussion des lange skandalfähigen Themas in den neunziger Jahren war *arbeitsphysiologisch* bestimmt. Die sanitätspolizeilichen Argumente wurden ergänzt und sehr rasch überlagert durch Messungen der Arbeitsleistungen von Schulkindern, For-

schungen zur Lernökonomie und Beiträgen zur Erziehung des Gedächtnisprozesses (...) Das verlagerte die Problemstellung von Reformen des Milieus zur Rationalisierung des schulischen Lernens und so der Arbeitszeit“ (Oelkers 1998: 271).

Konsequent deutete beispielsweise Emil Kraepelin Ermüdung dann auch nicht mehr als gesundheitsschädigendes Phänomen, sondern als „fortschreitendes Sinken der Arbeitsleistung“, was für körperliche und geistige Arbeit zugleich gälte. Um eine physiologische Feinsteuerung der entsprechenden muskulär-sensomotorischen Prozesse zu erzielen, musste ein Instrument zur Messung und Quantifizierung von Belastung bzw. Belastungsfähigkeit gefunden werden, mit welcher der Grenznutzen physiologischer Belastungen bestimmt werden konnte. Es war der italienische Physiologe Angelo Mosso, der als „Anreger der Psychophysiologie der Ermüdung“ den Ergographen erfand, das von Helmholtz' Myographen abgeleitet war, mit dem Muskelkontraktionen gemessen werden konnten (Vatin 1998: 356-357). Mit dem Ergographen konnte darüber hinausgehend quantifiziert werden, wie hoch die Belastung sowie Aus- und Überlastung von Muskeln lag. Der Ergograph fand auch in der Schulforschung rasch Eingang, der als sicheres Instrument zur Ermittlung von Ermüdungsprozessen galt (Oelkers 1998: 273). Während sich bei Kraepelin seine Analogisierung von körperlicher und geistiger Arbeit fand, unterließ es etwa Wilhelm Wundt, seine physiologische Psychologie auf die Pädagogik auszuweiten (ebd.: 275). Er sah zum einen die Kontingenz individueller Entwicklung und zum anderen, dass einfache physiologische Phänomene Teile komplexerer Gebilde darstellten, zu denen sie im Laufe der Entwicklung synthetisiert wurden. Kurzum: Eindeutige Aussagen mit auch experimentell zu erzeugenden und zu wiederholenden Kausalitäten, wie sie physiologische Prozesse auszeichnen, ließen sich nicht einfach auf Prozesse höherer Ordnung übertragen. Für den Übermüdungstheoretiker Mosso bedeutete dies, dass beispielsweise zwischen der einfachen Kausalität von Maschinen – hier wurde zumeist die Energiemetapher zur Darstellung von Verausgabung im Bereich menschlicher Arbeit verwendet – und dem „komplexeren Mechanismus der tierischen Ökonomie“ unterschieden werden musste:

„Unser Körper lässt sich nicht mit einer Lokomotive vergleichen, die für jeden Kilometer zurückgelegter Strecke eine bestimmte Menge Kohle verbrennt. Wenn dagegen unser Körper müde ist, hat eine geringe Menge Arbeit verheerende Folgen“ (Mosso nach Vatin 1998: 359).

So beinhaltet die Beachtung der Differenz zwischen einfachen physiologischen Phänomenen (einfache, mechanisch wiederholbare Reiz-Reaktion-Verknüpfungen) und komplexeren (neuro)psychologischen Zusammenhängen eine empirisch kaum noch beobachtbare *Komplexität*, die sich aus der Synthese verschiedener Entwicklungsschritte und –niveaus, einem nicht mehr durchschaubaren Spiel von materiellen und immateriellen Prozessen und fehlender Theorien, die noch die Gewähr geboten hätten, zumindest die Erklärungslücke erklären zu können, zusammensetzte. So stellt sich etwa als Hauptproblem bezüglich der Weiterentwicklung der Myographie zur Ergographie<sup>12</sup> das dar, was eigentlich das Ziel des Instruments sein sollte, nämlich die Messung des Reizes:

„Im Unterschied zum elektrischen Reiz, der vom Experimentator gelenkt wird, verfügt man über keine Möglichkeit, den Nervenreiz direkt zu messen. Der Impuls wird allein vom Probanden gesteuert, der im Unterschied zum Froschmuskel kein passives Objekt in den Händen des Experimentators ist (ebd.: 360).

Infolgedessen wurde die direkte Verknüpfung von Physiologie mit der Psychologie der Arbeit scharf kritisiert (ebd.: 362) und auch die Analogisierung von geistiger und körperlicher Arbeit musste aufgegeben werden, da sich unüberwindbare begriffliche Schwierigkeiten einstellten – etwa die Definition von Arbeit (ebd.: 367). Die „Ökonomisierung geistiger Arbeit“ stellte auch in Meumanns experimenteller Pädagogik zu Anfang des 20.

---

<sup>12</sup> Auch im Produktionsbereich wurde der Ergograph bald in Frage gestellt. So schrieb Hendrik de Man im Rahmen einer Untersuchung zur „Arbeitsfreude“ in 20er Jahren: „Die mechanistische Denkweise der meisten Psychotechniker und ‚instrumentalistischen‘ Psychologen hindert sie leider noch immer daran, die Wichtigkeit der voluntaristischen und psychogenen Faktoren (...) vom Gesamtzustand her zu erfassen. Der Ergograph in der Fabrik verzeichnet keine Kurve, die erklärbar wäre ohne Beziehung auf all die seelischen Tatbestände der Arbeiterpsyche, deren Ursprung man sowohl in dem Familienleben, in der Vorgeschichte und in der politischen und sozialen Umwelt jedes einzelnen Arbeiters, wie in seiner eigentlichen Arbeitsleistung zu suchen haben wird“ (de Man nach Walter-Busch 1996: 148, Fn 3). Bemerkenswert ist der Rückgriff auf den Seelenbegriff, um die mechanistische Deutung von Handlungen, Dispositionen usw. zu kritisieren, denn sie stellt auch für die Pädagogik seit dem 17. Jahrhundert die zentrale Bezugsgröße dar, um die Kontingenzen von erzieherischen Einwirkungen zu begründen.

Jahrhunderts eine wichtigen Forschungsschwerpunkt dar (Oelkers 2002: 17). Gegenüber einer spekulativen geisteswissenschaftlichen Pädagogik ging es bei ihm um die „experimentelle Evaluation der Wirksamkeit von Erziehungsabsichten (ebd.: 17). Die Möglichkeit von Einwirkungen kann aber nur mithilfe einer entsprechenden *Gegenstandstheorie* erforscht werden, mit der sich *Angemessenheit* und *Optimierungsformen* der jeweiligen Einwirkungsmethoden genauer bestimmen lassen:

„Ein großer Teil der Erziehung und des Unterrichts (ist) der pädagogischen Forschung ganz entrückt (...) Die Pädagogik kann nur das prüfen, ob die von Staat und Gesellschaft bestimmten oder von dem einzelnen Pädagogen geforderten Ziele den kindlichen Individuen und den Entwicklungsstadien des Kindes, für welche sie aufgestellt werden, angemessen sind und wie sie am besten erreicht werden“ (Meumann nach Oelkers 2002: 18).

Pädagogik hat traditionell Erziehungswirkungen erforscht und evaluiert, doch tritt die Kausalitätsproblematik erst explizit und verschärft in der empirisch orientierten Pädagogik bzw. Psychologie in der Absetzung von geisteswissenschaftlicher Pädagogik auf, die die Kontingenz von Erziehungsprozessen mit der Eigenlogik der Seele des Kindes zu erklären suchte. Evaluation als die Analyse und Bewertung erzieherischer Maßnahmen impliziert daher eine mehr oder minder komplexe Theorie von Einwirkungsverhältnissen, aus denen sich weitere Bestimmungen wie die oben angeführten Kriterien der Angemessenheit oder Optimierung ableiten. Die Beschreibung von Kausalverhältnissen hängt genuin mit der Theorie zusammen, mit deren Hilfe ein Phänomen beobachtet und beschrieben wird. Neben Kausalität stellt Standardisierung ein zentrales Element von Evaluation dar, der in der Phase des Taylorismus eine wichtige Rolle spielt.

## **2. Phase: Taylorismus und Testpsychologie**

Zwei Elemente sind für die historische Rekonstruktion von Evaluation als einer umfassenden Handlungs-Wirkungs-Forschung von Interesse. Zum einen die systematische Suche nach Kausalitäten bis hinein in die Mikrobereiche von Handlung und Psyche und zum anderen die dezidierte und genaue Beobachtung, die ein komplexes Wissen um das Protokollieren,

Systematisieren und Bewerten der Beobachtungen voraussetzt. Foucault hat dies in der Entwicklung der Disziplinarinstitutionen als zentrale Technologien rekonstruiert und metaphorisch als „Disziplinarblick“ (1994: 224) bezeichnet, der sich in den USA bis in den Taylorismus bzw. das Scientific Management und den Fordismus der 20er und 30er Jahre verfolgen lässt. Während Taylor in seinen Betrieben auf kleinschrittige Überprüfung der Arbeitseffektivität setzte<sup>13</sup>, versuchte Ford durch Standardisierung<sup>14</sup> Einsparungen und Optimierung von Arbeitsprozessen zu erzielen (Kieser 1999: 86). Taylor baute nicht auf ein bestimmtes Organisationsprinzip als vorgegebenes Passepartoute, sondern es ging ihm um eine „Lösungsmethode“, „die beschreibt, wie unter Berücksichtigung unterschiedlicher Ausgangsbedingungen ein gewünschtes Ziel realisiert wird“ (ebd.: 76). Beobachtung und ein detailliertes und systematisiertes Sammeln von Informationen waren dafür notwendig. Durch eine wenn auch spät einsetzende Rezeption des Taylorismus entstand so ein regelrechtes „Effizienz-Fieber“ in den USA:

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

Diese Expertokratie spiegelte sich etwa im Bildungsbereich im sogenannten Akkreditierungsmodell wieder, das nach Beywl das zweite Modell neben den standardisierten Messverfahren darstellt, bei dem Expertenkommissionen Schulen bewerteten (Grüner 1993: 34).

Evaluation erweist sich unter anderem deshalb als ein wichtiges Element innerhalb eines Sets von Wissensformen und -praktiken, die sich seit Taylorismus bzw. scientific management durchgesetzt und verändert

---

<sup>13</sup> Taylor beobachtete 26 Jahre lang Arbeiter an Drehbänken, unternahm 30.000 Versuche und investierte 200.000 Dollar zur Optimierung dieses spezifischen Arbeitsprozesses. Dabei entwickelte er zwölf Parameter wie Kühlung, Schneidezeit, Schneidewinkel, Durchzugskraft usw. (vgl. Kieser 1999: 78).

<sup>14</sup> Zwei andere wichtige Formen der Standardisierung ist die sozialwissenschaftliche Konstruktion des Durchschnittsmenschen („homme moyen“) durch Adolphe Quetelet (vgl. Link 1997: 143) und die Einführung der deutschen Industrienorm seit 1917 (ebd.: 270 f).

haben, weil sich im Fall der USA schon früh eine parallele Entwicklung im Bildungs- und Produktionsbereich feststellen lässt:

„Mit dem Aufstieg des *scientific management* und dem *school survey movement* in den Vereinigten Staaten zwischen 1900 und 1930 erwiesen sich dann auch die Wünsche des Schulverwaltungsbeamten Sears als zeitgenössische des *Age of Efficiency and Testing*. Besonders im Erziehungswesen entsteht damals ein immenser Bedarf an ‚objective test of all kind‘ (...), den schulbezogen zunächst Prüfungsbögen des ‚new type‘ befriedigen. Neu an ihnen ist, dass sie ‚large numbers of objective items rather than a handful of essay questions‘ enthalten (...) Die Verlockungen solcher Testbatterien werden durchaus in dem Versprechen ihrer *Objektivität* und der Möglichkeit einer gesellschaftsweiten *Vergleichbarkeit* der erhobenen Daten gesehen. Der Reiz, den *standardization* und *standards* damals auszuüben beginnen, hält sich ungebrochen bis heudet in den Bemühungen um Bundesstaatliche und nationale Standards der Lehrerausbildung und der Anstellung von Lehrern, der Schulcurricula und der *teaching effectiveness* (Prondczynsky 2001: 105, Hervorh i. Orig., T.H.).

Hieran zeigt sich, in welcher Weise Wissen aus unterschiedlichen aneinander anknüpft und sich verstärkt. Neben dem Taylorismus bildet die Testpsychologie eine zweite wichtige Diskurslinie, die in Deutschland dominierte. Die Messung individueller Leistungsfähigkeit über standardisierte Verfahren wurde mithilfe von Intelligenztests zu erreichen versucht, die wesentlich auf das Programm der individuellen Psychologie von Alfred Binet zurückgingen (Grubitzsch 1991: 78 ff.). Die Binet-Simon Testreihen zwischen 1908 und 1911 stießen auf weitgehende Akzeptanz, so dass in einer UNESCO-Untersuchung von 1948 eine Verwendung des Tests bereits in 22 Ländern festgestellt werden konnte (ebd.: 79). In einer Art Normalverteilungskurve ging es um die Zuordnung der Individuen zu Gruppen, die idealtypisch nach Entwicklung gestaffelt waren. Die standardisierten Verfahren der Testpsychologie zur Ermittlung von Leistungsfähigkeit bilden direkte Vorläufer der Evaluation, was anhand der Art der Testkonstruktion deutlich wird:

„Eigenschaften als Mittel zur Differenzierung zwischen Personen und Eigenschaften als homogene, quantitative Größen. Fast ausnahmslos sind die Bemühungen von Testkonstrukteuren darauf beschränkt, solche Verhaltenstendenzen ausfindig zu machen, bezüglich derer sich die Personen unterscheiden (...) Demnach wird für die ‚Intelligenz‘ fraglos

unterstellt, dass Intelligenz über Personen quantitativ variiert, dass es Personen mit ‚hoher Intelligenz‘ und mit ‚niedriger Intelligenz‘ geben muss. Ergänzt wird diese Auffassung durch ein Vorgehen, das die Überzeugung scheinbar rechtfertigt, dass diese Unterschiede lediglich quantitativer Natur seien (...) Dass es auch Unterschiede qualitativer Art geben könnte, wird nicht ernsthaft in Betracht gezogen“ (Krüger/Tröger nach Grubitzsch 1991: 81).

Wenn Intelligenz durch Kategorien wie ‚Leistung‘ oder ‚Effektivität‘ ausgetauscht wird, so gelten die genannten Konstruktionsprinzipien auch für standardisierte Modelle von Evaluation, in denen aufgrund standardisierter Normvorgaben (Indikatoren) auf Unterschiede (Differenzen) zwischen Individuen oder einzelnen Organisationen aufgrund einer Normalverteilung auf individuelle Leistungsfähigkeit geschlossen wird<sup>15</sup> (Vergleich). Organisationen wie Schulen, Kindergärten, Universitäten oder Krankenhäuser werden hierbei ‚individualisiert‘, d.h. wie Personen individuell positioniert und bewertet. Der unmittelbare Effekt ist, dass die Skalierung solcher Messungen eo ipso eine Konkurrenz der bewerteten Personen oder Organisationen impliziert, die auf einer abstrakten Ebene in ein direktes Verhältnis gesetzt werden und deren Positionen je nach Bezugsgruppe und Durchschnitt positiver oder negativer ausfallen (z.B. die Rankings, in denen scheinbar ein objektiv Leistung evaluiert wird). In diesem Zusammenhang kritisierte Holzkamp in den 70er Jahren die Zirkularität derartiger Tests, die immer wieder die Selektion im Bildungssystem auf analysierte Intelligenzunterschiede zurückführen wollten:

„Schulsysteme mit rigoroser Auslese von ‚unten‘ nach ‚oben‘ erscheinen so als sinnvolles Mittel, der Heraussonderung der ‚von Natur aus‘ Besten, die demgemäß dazu bestimmt sind, gesellschaftliche Führungsfunktionen einzunehmen (...) Dabei kann hier auf zirkuläre Weise der Anteil, der die institutionellen Selektionsfilter passieren kann, als Anteil der durch natürliche Begabung und Intelligenz besonders Ausgezeichneten gedeutet werden“ (Holzkamp 1976: 45).

### **3. Phase: Bildungsreform, Humankapital und Curriculum**

---

<sup>15</sup> Daher ist es auch nicht verwunderlich, dass Lee J. Cronbach als einer der einflussreichsten Testpsychologen auch entscheidende Beiträge zur Evaluation von Curricula geliefert hat (vgl. Wulf 1972).

In Deutschland wurde gegenüber den USA Evaluation mit erheblicher „Verspätung“ eingeführt und zum ersten Mal Anfang der 70er Jahre im Rahmen der Haushaltskonsolidierung im politisch-administrativen Bereich vorgeschrieben (vgl. Lachenmann 1978: 61), doch erlebte sie erst im Kontext von Neoliberalismus und New Public Management der 90er Jahre einen Boom, der bis heute anhält.

Für den Bildungsbereich lassen sich zwei Felder ausmachen, in denen Evaluation in den 70er Jahren besonders zum Einsatz kam: In der Schulforschung wurde die Gesamtschule spezifisch evaluiert und in der Curriculumsforschung hat Evaluation eine wichtige Rolle gespielt (Wulf 1972). Es ist unschwer zu erkennen, dass beide Einsatzgebiete der Evaluation im direkten Zusammenhang mit der Bildungsreform stehen, in der es wesentlich um die Überprüfung von Programmen ging. Evaluation ist daher in ihrer institutionellen Form eng an den Programmbegriff geknüpft, der seinerseits historisch auf sozialpolitische Reformprogramme in den USA seit den 50er und 60er Jahren zurückgeht (Stichwort „Great Society“). In unterschiedlichen Bereichen wie (Aus)Bildung, Gesundheit, Ernährung, Infrastruktur oder später Entwicklungshilfe ging es um die Bewertung von Reformprogrammen. Dies verdeutlicht die grundlegend politisch-staatliche Dimension von Evaluation zum Zweck sozialstaatlicher Intervention. Evaluation stellt in diesem Sinne eine aktive, intervenierende Forschung dar, also eine Art „staatliche Aktionsforschung“, bei der die *Wirkungen* eines Programms bei den Zielgruppen untersucht werden (Stockmann 2000a: 21). Ursprünglich handelt es sich bei Evaluation daher um eine staatlich-politisch initiierte und mithilfe der Wissenschaft umgesetzte Wirkungsforschung in der Absicht, sozialen Wandel gezielt zu steuern und dadurch Rationalität zu erhöhen:

„Ziel eines solchen 'systematischen Denkens für soziale Aktion' (Rivlin) ist also die Erhöhung der Rationalität der öffentlichen Verwaltung in ihrer Funktion als aktiv in die Gesellschaft eingreifende, planende – d.h. vernünftige Verteilung knapper Ressourcen und eine Verbesserung der (Sozial)Politik“ (Kantowsky 1977a: 25).

Somit spiegelt Evaluation die *politisch-wissenschaftliche Doppelstruktur der Reformprogramme* wider. Die angestrebten Veränderungen sind an

bestimmte Ziele und Intentionen geknüpft, auf die Evaluation direkt bezogen werden kann:

„Evaluation erfolgte in einer fortwährenden Überprüfung der Realisierung von Intentionen bzw. normativen Kriterien und als Entwicklung von Alternativen auf Grund unbefriedigender Ergebnisse, wobei die Modifikationen sich häufig auf die Mittel der Realisierung, oft genug aber auch auf die Intentionen selbst bezogen“ (Wulf 1972: 10).

Der Begriff der Intention, der zum einen den normativen Gehalt und zum anderen die Interessen der Akteure begrifflich deutlich werden lässt, ist mittlerweile durch den Zielbegriff ersetzt worden, der scheinbar Objektivität verbirgt. Dies deutet auf eine weitere Verschiebung im Evaluationsdiskurs hin, die sich im vorherrschenden positivistischen Verständnis von Evaluation als wissenschaftlich-rationalem, objektivem und neutralem Verfahren der Leistungsmessung äußert.

Konnte seinerzeit Christoph Wulf als Herausgeber eines Bandes zu Evaluation Anfang der 70er Jahre noch allgemeine ideologiekritische Bedenken zur Evaluation äußern, so ist der gegenwärtige Diskurs durch eine auf praktische Anwendung zielende ‚Evaluatoren-gemeinde‘ gekennzeichnet. Doch die kritischen Anmerkungen von Wulf zur Evaluation können generell Geltung für eine kritisch-reflektierte Sicht auf die Funktionen von Evaluation beanspruchen, denn das Problem der einseitigen Auftragsforschung und damit der Legitimation (Wulf 1972: 11) stellt sich in jedem Evaluationsbereich. Zudem – und dieses Argument hat angesichts der aktuellen neoliberalen Veränderungen einen noch gewichtigeren Stellenwert – kann durch Evaluation grundsätzlich alles, d.h. der status quo einer jeden Institution, Situation, Organisation usw. in Frage gestellt und zur Veränderung ‚getrieben‘ werden (ebd.: 11). Die posthume Legitimation von Interventionen wie auch die evaluativ erhobenen Defizitdiagnosen legitimieren vergangenes und zukünftiges Handeln auf Seiten der Akteure.

Historisch waren es vor allem im Schulbereich die Gesamtschulen, die von den 60er bis in die 80er Jahre evaluiert wurden, doch waren die Ergebnisse derart unterschiedlich, dass

„hinsichtlich des methodologischen Ansatzes, der Vorgehensweise und der eingesetzten Datenerhebungsmethoden (...) kein abschließendes Urteil über die Effektivität der neuen Schulform gegenüber dem traditionellen gegliederten Schulsystem möglich war“ (Stockmann 2000a: 26).

Dieser Kontext der Bildungsreform verweist auf einen zweiten Diskursstrang, der für die Genealogie der Evaluation neben der wissenschaftlichen und ökonomischen Rationalisierung wesentlich ist. Die in den 60er Jahren anhebende Debatte um den Wert des Humankapitals<sup>16</sup> und die damit verbundene bildungsökonomische Forschung stellte die Reformbedürftigkeit des Bildungssystems heraus. Hierbei spielten internationale Organisationen wie die OECD und die UNESCO eine Vorreiterrolle. So wurden auf der Washingtoner OECD-Konferenz von 1961 die Ziele für den Ausbau von Bildungssystemen festgelegt und „Weichen für eine weltweite Bildungsexpansion“ gestellt, wobei die Bildungsökonomie als Steuerungswissenschaft für eine systematische Bildungsplanung etabliert wurde (vgl. von Recum 1998: 10). Gegenüber der zweiten Phase liegt die Differenz in der Internationalisierung von Vergleichsstudien im Bildungsbereich mit der Absicht einer Globalsteuerung über die nationalen Grenzen hinweg.

#### **4. Phase: Die Durchsetzung des Neoliberalismus**

Die gegenwärtige Phase des Neoliberalismus ist durch eine weitreichende Ökonomisierung des Sozialen bzw. eine „Generalisierung der ökonomischen Form“ (Bröckling/Krassmann/Lemke 2000: 16) gekennzeichnet:

„Die Ökonomie gilt nicht mehr als *ein* gesellschaftlicher Bereich mit spezifischer Rationalität, Gesetzen und Instrumenten, sie besteht vielmehr aus der Gesamtheit menschlichen Handelns, insofern dieses durch die Allokation knapper Ressourcen zu konkurrierenden Zielen gekennzeichnet ist“ (ebd.)

Im neoliberalen Kontext wird Evaluation als Instrument zur Effektivitäts- und Effizienzmessung zu einem Türöffner für Ökonomisierung, mit der

---

<sup>16</sup> Die Humankapital- und Bildungsökonomie-debatte reicht historisch bis ins 18. Jahrhundert zurück (vgl. Kahlert 1978), weshalb eher von einer „Wiederentdeckung des Humankapitalkonzeptes“ zu sprechen ist (Hüfner 1970: 30).

eine kausalmechanistische (Re)Taylorisierung nicht-ökonomischer Bereiche realisiert wird. Konzepte wie lebenslanges Lernen und Kompetenzen lassen sich hinsichtlich der Subjekte als evaluative Selbsttechniken begreifen. Den Individuen wird suggeriert, dass ihre Fähigkeiten nie zuende entwickelt und stets verbesserungswürdig seien, was ihre Leistung betrifft. Daher ist die *normativ-moralische Appellstruktur* des Evaluationsdiskurses von zentraler Bedeutung, wie sie in Evaluationsberichten, Rankings oder Statistiken enthalten ist. Durch diese spezifische Form sozialen Beobachtungswissens werden die Subjekte in Konkurrenz und unter Handlungsdruck gesetzt. Etabliert wird so ein Klima des Verdachts, da derjenige, der einer Evaluation unterzogen wird, bereits unter Verdacht geraten ist. Denn es wird unterstellt, dass entweder Ressourcen nicht optimal genutzt wurden, es nicht effektiv genug produziert oder effizient gespart wurde. Evaluation fungiert dabei als scheinbar objektives Instrument der *Datenermittlung*. Im Unterschied zur Datenerhebung zu Diagnosezwecken folgen aus der Ermittlung positive und negative Sanktionen: Die Guten sollen belohnt werden (z.B. Eliteuniversitäten), während die Schlechten weniger Geld erhalten. Veröffentlichte Rankings wirken wie ein Pranger, an den jeder als ‚Schuldiger‘ gestellt wird, der unverantwortlich gewirtschaftet und nur mit ungenügender Leistung aufwarten kann.

Evaluation wird von offizieller Seite her mit dem Rationalitätstechnischen Argument verteidigt, dass es einen vernünftigen Ressourceneinsatz und somit vernünftiges Sparen ermögliche. Sparen erscheint hierbei als (Sach)Zwangsmechanismus und als unanfechtbare Form ökonomischen Kalküls. Demgegenüber hat Torsten Bultmann darauf hingewiesen, dass es kein unpolitisches Sparen gäbe (1996: 329). Geld und Gesetze sind klassische Steuerungsmedien im Bildungssystem, so dass jede Ressourcenverknappung unmittelbar Einfluss auf die jeweiligen Organisationen und deren Operieren hat. Die nachhaltige Diskussion um das Sparen – und in dieser Diskursintensität wird Sparen zu einer „ideologischen Form“ (ebd.: 330) – blendet aus, dass die Universitäten seit dem Doppelbeschluss der Kultusministerkonferenz von 1977 zwar für alle Studierwilligen offen gehalten, aber nicht mehr weiter ausgebaut wurden. Seitdem ist der Anteil der Hochschulausgaben von 1,3% in den 70er Jahren auf 0,9% zu

Beginn der 90er Jahre zurückgegangen (ebd.: 336). So wie ‚leere Kassen‘ als Naturereignis im Diskurs dargestellt werden, findet eine Naturalisierung sozialer und politischer Verhältnisse über die Marktmetapher statt. Die ungleichen sozialen Ausgangsbedingungen bleiben dabei außen vor, so dass nur noch die Individuen mit ihren unterschiedlichen Fähigkeiten oder die Universitäten mit ihrer unterschiedlichen Qualität berücksichtigt werden. Macht strategisch einzusetzen und sich Vorteile gegenüber ‚den anderen‘ zu verschaffen, werden als legitime Mittel und Indizien der eigenen Marktstärke angesehen. Rechte hat man lediglich noch als (zahlen-der) Kunde, womit das Menschenrecht auf Bildung keine Rolle mehr spielt.

Auf der Ebene psychischer Befindlichkeit der Subjekte erzeugen diese Diskurse eine Haltung der Prävention. Angesichts der zunehmenden Risiken (ökonomische Risiken, Arbeitslosigkeit) wird so Bildungserwerb zum Risikomanagement, das eng an Faktoren wie Zeit und Geld gekoppelt wird. ‚Zeit ist Geld‘ oder, wie als ironisch Wortspiel formuliert werden könnte: ‚Zeit isst Geld‘. Daher lautet das Motto: Möglichst das beste in kürzester Zeit erreichen, denn man weiß nie, wie es weitergeht und was noch folgt. Die strukturell durch finanzielle Verknappung produzierte Unsicherheit im Bildungssystem erzeugt also eine Disposition der permanenten Selbstbeobachtung und Selbstevaluation, in der jeder Schritt, den man tätigt, wohl überlegt sein will. Im Sinne einer veralltäglichten Chaostheorie könnten kleine Entscheidungsfehler von heute schon morgen große, sprich: existenzielle Auswirkungen haben. Auf diesen fruchtbaren Boden der (selbst)evaluativen Grunddisposition der Subjekte, die permanent angerufen sind, ihre Ängste und Unsicherheiten zu rationalisieren und ihrem individuellen Lebenssicherheitsmanagement zu unterwerfen, treffen ihrerseits Rankings, Evaluationsstudien und Universitätsstatistiken. Sie gaukeln sicheres Wissen und damit Entscheidungssicherheit, wo keine Sicherheit zu erreichen ist. So erzeugen Diskurse auf der einen Seite Dispositionen der Angst und Unsicherheit, für deren rationalen Umgang sie gleichsam auch die Mittel bereit stellen.

Im Unterschied zu Evaluationen der 70er Jahre im Rahmen der Reformprogramme geht es heute nicht mehr darum, ein Programm auf seine

Wirkungen zu überprüfen (Input-Orientierung, Wirkungsevaluation), sondern die Leistungsevaluation dient der *Legitimation des Abbaus von Strukturen*, deren Ineffektivität eigentlich schon feststeht – und mithin auch die Alternativen, mit denen gleichzeitig aufgewartet wird. So wissen die Protagonisten des New Public Management bereits, dass die Einführung von Quasi-Märkten und Konkurrenzsituationen im Bildungsbereich bereichernd und fördernd auf die Entwicklung wirken wird, was aber bei jeder empirischen Überprüfung lediglich den Status einer Behauptung hat (Hofacker 2001). So gibt es kaum einen eindeutigen bildungsökonomischen Befund, der auf empirisch sicheren Füßen steht. Weder die Produktivitätshypothese des Humankapitals noch die Grenzproduktivitätshypothese werden durch die Datenlage eindeutig gestützt oder entkräftet (Timmermann 2002: 97). Und ob „höhere Bildungsausgaben eine Folge des allgemeinen Einkommenswachstums seien oder ob höhere Bildungsausgaben ein höheres Wirtschaftswachstum induzieren“, kann ohne eine überzeugende Theorie nicht entschieden werden (ebd.). Nicht nur die makroökonomischen Untersuchungen lassen keine klaren Schlussfolgerungen zu, auch die mikroökonomischen Ressourcenstudien werfen mehr Probleme und Fragen auf als sie Antworten liefern. Kritisiert wird u.a., dass die einzelnen Bildungsinstitutionen eine „black box“ blieben, dass den Studien eine lerntheoretische Basis fehle und dass der sozialisations-theoretisch zunehmend bedeutsamere soziale Kontext nur unzureichend berücksichtigt werde (ebd.: 111).

Darüber hinaus wird ‚der Staat‘ in neoliberalen Diskursen als Störgröße dargestellt, dessen politisches, legislatives und administratives Handeln automatisch zu chronischer Bevormundung führe. Diese Kritik zielt strategisch auf einen *bestimmten Staatsbegriff*, nämlich den Sozialstaat keynesianischer Prägung, und nicht etwa auf den autoritär-feudalen Staat preußischer Provinienz, der eine wesentlich höhere administrative Regelungsdichte hatte. Eine Ökonomisierung des Bildungssystems etwa in Form der bruchlosen Übertragung von Evaluation als betriebswirtschaftlichem Steuerungsmittel würde, organisationstheoretisch betrachtet, als ein Versuch anzusehen sein, die Entscheidungslogik der Organisation umzukodieren. Die Folge dieses Eingriffs ist eine nachhaltige Irritation des Fel-

des durch die Implementation einer feldfremden Logik. Diese Umkodierung impliziert aber einen *höheren und strikteren Regulierungsbedarf*.

„Die Verbürokratisierungsproblematik durch ‚Entstaatlichung‘ und Subordination unter die Mechanismen von Markt und Wettbewerb zu lösen, würde vom Regen bürokratischer Reglementierung in die Traufe einer noch stärkeren externen Regie der Hochschule durch ökonomische Steuerungsmechanismen führen“ (Hofacker 2000: 109).

Kulturell gesehen arbeiten Organisationen immer mit einer bestimmten „Trägheit ihres Wissens“, das die Beharrungskraft von Organisationskulturen ausmacht – und zwar von allen Organisationen einschließlich der ökonomischen (vgl. Höhne/Nousch/Seyfert 2004). Daher stellt sich die Frage, inwieweit Veränderungen des Feldes überhaupt von außen beeinflusst, geschweige denn gesteuert werden können, wenn die Organisation nicht nachhaltig geschädigt oder gar aufgelöst werden soll. Daher müssen die Effekte von Veränderungen, die den Zweck und die Logik des Feldes in elementarer Weise betreffen, reflektiert werden. *Denn hierbei zeigt sich, dass Evaluation in grundlegender Weise zu einer Strukturveränderung des Feldes beitragen bzw. sie legitimieren kann.* Da Institutionen bzw. Organisationen an soziale Funktionen und Übereinkünfte gebunden sind, entsteht auf der normativen Ebene durch das strukturelle Unterlaufen dieses sozialen Funktionsanspruchs ein Legitimationsproblem. Hierbei findet in der Diskussion eine semantische Verschiebung des Legitimationsbegriffs von einer am Gemeinwohl orientierten Legitimation hin zu einer steuerungstechnokratischen Orientierung statt:

„Die New Public Management Modelle lösen diesen Legitimationsrahmen organisatorisch auf, indem die demokratische Legitimation nur für die Zieldefinition als erforderlich angesehen wird, während Effizienz und Effektivität in Abgrenzung gegenüber dem Politischen als Aufgaben des Managements ausgewiesen werden, die nach dessen Regeln in eigener Zuständigkeit, ‚Managers right to manage‘, zu verwirklichen seien. In vielen Bereichen stellt sich die Legitimationsproblematik aber gerade hinsichtlich der Art und Weise der Instrumentierung der Zielvorgabe. So besteht ein ganz erheblicher Unterschied, ob das Ziel der ‚Qualitätssicherung‘ von Studienabschlüssen repressiv durch scharf selektive Zugangsregelungen und Prüfungen oder vorrangig über Strategien der Verbesserung der Lehre und der Studienbedingungen verfolgt wird“ (Hofacker 2000: 127).

Die Ökonomie wird gegenüber der Politik zum entscheidenden Akteur von Veränderung, obwohl sie zur Umwelt des Bildungssystems gehört. „Autonomie“ wird in diesem Kontext zum Kampfbegriff, mit dem die Interventionen von außen legitimiert und bestehende Strukturen als anachronistisch dargestellt werden sollen, indem sie als Abhängigkeiten-produzierend und aufrechterhaltend dargestellt werden. So wird die eigene „unfreundliche Übernahme“ – da im Bourdieuschen Sinne „feldfremde Übernahmen“ – euphemistisch zur „freundlichen Unterstützungsleistung“, mit dem Effekt, dass die eigentlich bestehende Autonomie des Feldes, die durch rechtlich-gesetzlicher Steuerung ist, abgebaut und schließlich destruiert wird. Evaluation kann hierzu ein Mittel darstellen. Zum einen wird ihm die bereits erwähnte Objektivität zugeschrieben, zum anderen „vermittelt“ es zwischen externen und internen Akteuren des Feldes, um das es geht. Wer könnte sich gegen Evaluation als objektive Leistungs-, Effektivitäts- und Wirkungsmessung aussprechen, wenn er sich nicht dem Verdacht aussetzen wollte, als strukturkonservativ eingestuft zu werden? Hier greift ein narratives Schema als Subdiskurs, in dem die Ökonomie als Befreiersubjekt (= Held) die Autonomie bringt, indem sie die in Ketten der Abhängigkeit gehaltene Bildung (= potentielle Braut) dem Staat (= Ungeheuer) entreißt.

In dem Maße, in dem Evaluation zum methodischen Instrumentarium für die Messung der Wirksamkeit von Sozialprogrammen wurde, lieferte es gleichzeitig die Legitimationsgrundlage für entsprechendes staatliches Handeln. Nach Gerd-Michael Hellstern und Hellmut Wollmann stellt Evaluationsforschung Zeuge, Rechtfertigungsinstrument und Stimulator für diese Politik dar, die in den USA mit zahlreichen gesetzlichen Initiativen und Institutionalisierungen in Form von Behörden und Programmen verbunden war (vgl. Stockmann 2000a: 21). Den zentralen *Funktionswechsel von Evaluation* und die Verlagerung hin zu Kosten- Einsparungs- und Effizienzprogrammen ist durch die ökonomische Krise markiert, die in den 70er Jahren begann, und in den 80er Jahren zur konservativ-neoliberalen Wende führte. Insofern hat das Spar- und Krisenszenario eine viel längere Vorgeschichte, während der Bedeutungszuwachs von Evaluation sich we-

sentlich durch die Einführung des New Public Management erklären lässt. Dessen Programm lautet:

„Kundenorientierung, Kostensenkung und Effizienz, Trennung der Zieldefinierungs- und strategischen von den operativen Kompetenzen, Trennung der Funktionen von Leistungsfinanzierung und Leistungserbringung, flache Hierarchien mit möglichst weitgehender operativer Selbstständigkeit der jeweiligen Ebenen (...)“ (Hofacker 2000: 89).

Hierbei soll die rechtliche Steuerung soweit wie möglich zurückgeschraubt werden und input-Steuerung durch output-Steuerung ersetzt werden. Diese Umstellung des Steuerungsmodus entspricht analog der Wechsel von der Wirkungsevaluation zur Leistungsevaluation, die zum zentralen Instrument der Effektivitäts- und Effizienzmessung im Rahmen der New Public Management-Programme wird. Die Frage ist nur, ob damit auch eine Steigerung von Qualität erreicht wird, die normalerweise in Aussicht gestellt wird.

## **Evaluation und Qualität**

Die Frage nach der Qualität bzw. dem Verständnis von Qualität der Leistung, die eine Organisation für die Gesellschaft erbringt, ist unmittelbar mit Evaluation verknüpft. Ähnlich dem Evaluationsbegriff ist der Qualitätsbegriff historischen Diskurskonjunkturen und –paradigmata unterworfen (vgl. Fend 1996) und die Tendenz, Qualität an reinen Outputfaktoren (Abschlussrate, bestandene Prüfungen usw.) zu messen, stellt eine Reduktion des Qualitätsbegriffs dar. Von zentraler Bedeutung ist daher eine Klärung von Qualität, da Evaluation primär an das Ziel der Verbesserung im Sinne von Qualitätssteigerung gebunden und aus dem Qualitätsbegriff abgeleitet wird. So hat etwa Helmut Fend darauf verwiesen, dass es sich bei Qualität um ein Phänomen auf mehreren Ebenen handelt, bei dem die Reduzierung auf eine Ebene automatisch einen reduzierten Qualitätsbegriff zur Folge hat. Daher kann in Anschluss daran nur für einen anspruchsvollen und nicht ausschließlich formal-quantitativen Qualitätsbegriff plädiert werden. Für den Schulbereich bedeutet dies:

„Der Marsch durch die verschiedenen Formen und Ebenen der Qualität im Bildungswesen legt die Interpretation nahe, dass das phänomenal erscheinende, beobachtbare, sichtbare und erlebbare alltägliche Schulgeschehen mit seinen divergierenden Wirkungen und Widersprüchen das Ergebnis des *konfigurativen Zusammenspiels von Gestaltungsfaktoren auf verschiedenen Ebenen* ist“ (Fend 1996: 93).

Quantifizierende Methoden von Evaluation reduzieren strukturell die Komplexität von Organisationen, so dass sich etwa im Kontext der Bildungsfinanzierung durch internationale Organisationen wie OECD und Weltbank grundsätzlich die Frage nach der „demokratischen Legitimation international vermittelten bildungspolitischen Steuerungshandelns“ stellt (von Recum 1998: 16). Denn „wer legitimiert die OECD, die Indikatorenforschung zum Schlüsselglied des Verhältnisses von Wissenschaft und Bildungspolitik zu machen?“ (Weiler nach von Recum 1998: 17). Hier verlagert sich die Gratwanderung zwischen „Vereinheitlichung und Differenzierung“ (Norbert Wenning) als das klassische Problem des Nationalstaats des 19. Jahrhunderts auf die transnationale Ebene und wirft die Frage nach der demokratischen Legitimation von politischen Top-Down-Prozessen auf, durch welche die historisch gewachsenen Bottom-Up-Strukturen auf lokal-regionaler und nationaler Ebene verändert oder nachhaltig erschüttert werden können. Dieselbe Problematik tritt bei der quantitativen, output-orientierten Bestimmung von Qualität als zentralem Ziel von Evaluation auf und daher verweist Fend zurecht auf die Notwendigkeit der Differenzierung nach unterschiedlichen „Beurteilungsebenen“:

„Jedes Bildungswesen einer Nation repräsentiert eine je individuelle Konstellation von Gestaltungsmerkmalen auf der nationalen und regionalen Systemebene und Gestaltungsmerkmalen auf der lokalen Schulebene und je einmaligen Persönlichkeiten von Lehrern und ihrem Repertoire an Möglichkeiten des Schule-Haltens (...) Wir müssen die Qualitätsfrage systematisch nach der Beurteilungsebene differenzieren. So kann es zum Beispiel geschehen (...), dass auf der Schulebene ein Land, zum Beispiel in den USA im Vergleich etwa zu deutschsprachigen Schulen, ungleich attraktivere Einzelschulen zu finden sind. Auf der Systemebene zeigt sich aber ein insgesamt deutlich niedrigeres Niveau. Der Hintergrund ist unschwer auszumachen: Der auf dem Marktprinzip beruhende ausgedehnte private Anteil am Schulwesen führt in den USA dazu, dass an einzelnen Orten Vorteile kumuliert und maximiert werden“ (Fend 1996: 93).

Aufgrund der jeweiligen Konstellation kann über allgemeine Indikatoren nicht auf allgemeine Ursachen und ein entsprechendes Handeln geschlossen werden (ebd.: 94) <sup>17</sup>. Vielmehr ist kritisch auf die Gefahr zu verweisen, dass durch die Indikatorisierung eine Abstraktionsebene konstruiert wird, auf der – metaphorisch gesprochen – alle Katzen grau sind.

Allgemeine Indikatoren als Grundelemente von Evaluation und internationalen Bildungsvergleichen bergen grundlegend zwei Probleme. Zum einen die schon von Fend angedeutete Ausblendung von Strukturebenen des Bildungssystems unterhalb der Makroebene, die in der konsequenten politischen Umsetzung auf eine homogenisierende Weltgesellschaft auf ein eindimensionales Bildungsverständnis hinausliefere (vgl. von Recum 1996: 17 f.). Zum anderen werden mit Makroindikatoren „funktionale Äquivalente“ zu Marktsteuerungsmechanismen eingeführt (Hofacker 2000: 95), die als formalabstrakte Größen genau die quantifizierenden Vergleiche und Umrechnungen erst ermöglichen, durch deren Konstruktion von anderen Ebenen und Variablen abgesehen wird.

## **Fazit**

Evaluation stellt sich als komplexes soziales Beobachtungswissen auf den Ebenen alltäglicher und wissenschaftlicher Praktiken dar. Die historische Rekonstruktion zeigte, dass es weit verzweigte historische Diskurslinien gibt, in denen ein implizites Evaluationswissen identifiziert werden konnte - exemplarisch an der Überbürdungsdebatte und der Entstehung psychologischer Tests aufgezeigt.

Weiterhin fungiert Evaluation als panoptische Kontrolltechnologie, durch welche Gesellschaft und Subjekte unter (gegenseitige) Dauerbeobachtung und Dauerbewertung gestellt werden. Sie stellt Teil eines Sets kybernetischer Praktiken wie Assessment, Supervision, 360 Grad Feedback auf der Ebene von Organisationen dar. Darüber hinaus wurden aufgrund der technisch-medialen Entwicklung seit den 70er Jahren komplexe, sozialwissenschaftliche und politische Beobachtungssysteme wie Massenkommunikationsforschung, Konsumentenforschung, Umfragen, Datenbanken

---

<sup>17</sup> Fend fügt hier als Beispiel die völlig unterschiedliche Zahl von Studienabbrechern in drei verschiedenen Ländern an und der Grund dafür liege im „erwähnten komplexen Wechselspiel von Gestaltungskräften auf verschiedenen Ebenen“ (ebd.).

usw. geschaffen – Stichwort Massenfeedback – die ein zentrales Signum einer „Evaluationsgesellschaft“ (Prondczynsky) bilden. In diesem gesellschaftlichen Rahmen hat sich Beobachtung als wesentlicher Steuerungsmodus postmoderner Gesellschaften etabliert und über vielfältige Praktiken normalisiert, wozu Evaluation als Instrument und Methode wesentlich beigetragen hat. Die dahinter stehende und historisch über 150 Jahre zurückreichende Utopie ist die einer völligen Kontrolle von Komplexität, Kausalität und Wirkungen von Handeln über Wissen und Information.

Durch (Selbst)Beobachtung werden kybernetische Kontrollkreisläufe in Gang gesetzt, durch die punktuell und permanent ‚Korrekturen‘ eines gestörten Gleichgewichts vorgenommen werden können. Angezielt ist im Kontext des Neoliberalismus eine Selbstadjustierung der Subjekte an ‚normale Risikolagen‘, wobei pädagogische Konzepte des lebenslangen Lernens und Kompetenzerwerbs die Anrufung zur steten Selbstevaluation pädagogisch rationalisieren. Individualisierung von Risiken und die Aufforderung, Leben und Lernen selbst zu managen, stehen komplementär zueinander.

In grundsätzlicher Weise kann Evaluation auch als spezifischer gesellschaftlicher Umgang mit Komplexität bzw. Kausalität verstanden werden. Dergestalt bildet es einen Teil eines weit verzweigten sozialen Wissens um Wirkungen, Kausalität und Zusammenhänge. In Form wissenschaftlichen Wissens zweiter Ordnung kann Evaluationswissen methodisch-systematisch gesteigert werden, wodurch es zur Rationalisierung von Entscheidungen und Handeln beiträgt. Hierfür steht die Phase der Planungs-rationalität der 60er und 70er Jahre, in der Evaluation explizit als Teil sozialer Reformen angewendet wurde. Gegenüber dieser Programmevaluation, in der Wirkungen untersucht wurden, geht es in den 90er Jahren wesentlich um die Evaluation von Leistung, Effektivität und Effizienz. Ein Effekt der indikatoren-gesteuerten Leistungsmessung und quantitativen Evaluation ist die Etablierung einer neuen, ökonomischen Rationalität, mit der eine andere Entscheidungslogik durchgesetzt werden soll. Die Metapher der ‚black box‘, ist auf Bildungsinstitutionen genauso wie auf die Bildungssubjekte gleichermaßen anwendbar, da ihre ‚inneren Strukturen‘ auf dieser Ebene der Abstraktion irrelevant sind. Evaluation stellt im Rahmen des

New Public Management ein zentrales Mittel des Gleichnamig-Machens und des abstrakten Vergleichens dar. Dergestalt wirkt sie als Selektionsmittel oder zumindest als rationales Mittel der Legitimation von Selektion und trägt zur Rationalisierung von Risikoängsten bei. Eine theoretische Alternative zu dieser Art von Evaluation könnte die Ableitung des Evaluationsbegriffs aus einem anspruchsvollen und komplexen Qualitätsbegriff sein, der auch deutlich machte, dass Evaluation von spezifischen Ziele, Bereichen und Gegenstände abhängt, bei denen sie zur Anwendung kommt. Darüber hinaus ist ein sozialhistorischer Blick auf Evaluation notwendig, um die subtilen Machteffekte, die unzweifelhaft damit einhergehen, zu analysieren.

## **Bibliographie**

Beniger, James R. (1986): *The Control Revolution*. Cambridge u.a.

Etzioni, A. (1968): *The Active Society*. New York/London.

Fend, Helmut (1996): *Schulkultur und Schulqualität*. In: *Zeitschrift für Pädagogik. Die Institutionalisierung von Lehren und Lernen* (34. Beiheft), hg. von Achim Leschinsky. Weinheim/Basel. S. 85-98.

Foucault, Michel (1994): *Überwachen und Strafen*. Frankfurt/M.

Grüner, Herbert (1993): *Evaluation und Evaluationsforschung im Bildungswesen*. In: *Pädagogische Rundschau*, 47, S. 29-52.

Höhne, Thomas (2003): *Schulbuchwissen. Umriss einer Wissens- und Medientheorie des Schulbuchs*. Frankfurt/M.

Höhne, Thomas (2003a): *Die thematische Diskursanalyse – dargestellt am Beispiel von Schulbüchern*. In: Keller, R. u.a. (Hg.): *Handbuch Sozialwissenschaftliche Diskursanalyse. Band 2: Forschungspraxis*. Opladen. S. 389-420.

Höhne, Thomas/Nousch, Daniela/Seyfert, Gerd (2004): *Universität als Organisation. Eine organisationssoziologische und diskursanalytische*

- Studie zu Hochschule und ihren Experten/Expertinnen. (abrufbar unter: [www.i-wb/publikationen](http://www.i-wb/publikationen)).
- Holzkamp, Klaus (1976): Begabung-Intelligenz. In: Wulf, Christoph (Hg.): Wörterbuch der Erziehung. München. S. 44-49.
- Hüfner, Klaus (1970): Bildungsinvestition und Wirtschaftswachstum. Stuttgart.
- Kahlert, Helmut (1978): Prolegomena zu Geschichte der Bildungsökonomie. Weinheim/Basel.
- Kantowsky, Detlev (Hg. ) (1977): Evaluierungsforschung und –praxis in der Entwicklungshilfe. Zürich.
- Kantowsky, Detlev (1977a): Die Rationalität von entwicklungspolitischen Maßnahmen der deutschen Bundesregierung und deren Evaluierung. In: Kantowsky, Detlev (Hg. ) (1977): Evaluierungsforschung und –praxis in der Entwicklungshilfe. Zürich. S. 1-24.
- Keiner, Edwin (Hg.) (2001): Evaluation (in) der Erziehungswissenschaft. Weinheim/Basel.
- Kieser, Alfred (Hg.) (1999): Organisationstheorien. Stuttgart/Berlin/Köln.
- Lachenmann, Gudrun (1977): Evaluierungsforschung – historische Hintergründe, sozialpolitische Zusammenhänge und wissenschaftliche Einordnung. In: Kantowsky, Detlev (Hg. ) (1977): Evaluierungsforschung und –praxis in der Entwicklungshilfe. Zürich. S. 25-87.
- Lee, Barbara: Theories of evaluation. In: Stockmann, Reinhard (Hg.) (2000): Evaluationsforschung. Grundlagen und ausgewählte Forschungsfelder. Opladen, S. 127-164.
- Leisner, Walter (1994): Der unsichtbare Staat. Berlin.
- Link, Jürgen (1997): Versuch über den Normalismus. Wie Normalität produziert wird. Opladen.
- Luhmann, Niklas/Schorr, Karl Eberhard (1982): Zwischen Technologie und Selbstreferenz. Fragen an die Pädagogik. Frankfurt/M.
- Mayntz, Renate (1997): Soziologie in der öffentlichen Verwaltung. Heidelberg.

- Oelkers, Jürgen (1998): Physiologie, Pädagogik und Schulreform im 19. Jahrhundert. In: Sarasin, Philip/Tanner, Jakob (Hg.) (1998): Physiologie und industrielle Gesellschaft. Frankfurt/M., S. 2456-285.
- Oelkers, Jürgen (2002): Einige historische Erfahrungen im Verhältnis von Psychologie und Pädagogik. In: Reichenbach, Roland/Oser, Fritz (Hg.): Die Psychologisierung der Pädagogik. Übel, Notwendigkeit oder Fehldiagnose. Weinheim/München. S. 12-28.
- Prondczynsky, A. von (2001): Evaluation der Lehrerausbildung in den USA: Geschichte, Methoden, Befunde. In: Keiner, E. (Hg.)..., a.a.O., S. 91-140.
- Recum, Hasso von (1998): Bildungspolitische Steuerung oder Die Kunst, das Unmögliche möglich zu machen. Frankfurt/M.
- Rhyn, Heinz (2001): Bildungsevaluation und Bildungspolitik – Pädagogische Ambitionen zwischen Forschung und Politik. In: Keiner, Edwin (Hg.) (2001): Evaluation (in) der Erziehungswissenschaft. Weinheim/Basel. S. 179-194.
- Stockmann, Reinhard (Hg.) (2000): Evaluationsforschung. Grundlagen und ausgewählte Forschungsfelder. Opladen.
- Stockmann, Reinhard (2000a): Evaluation in Deutschland. In: Ders.: Evaluationsforschung. Grundlagen und ausgewählte Forschungsfelder. Opladen. S. 11-40.
- Timmermann, Dieter (2002): Bildungsökonomie. In: Tippelt, Rudolf (Hg.): Handbuch Bildungsforschung. Opladen. S. 81-122.
- Vatin, Francois (1998): Arbeit und Ermüdung. Entstehung und Scheitern der Psychophysiologie der Arbeit. In: Sarasin, Philip/Tanner, Jakob (Hg.) (1998): Physiologie und industrielle Gesellschaft. Frankfurt/M., S. 347-368.
- Vierhaus, Rudolf (1972): Bildung. In: Brunner, Otto/Conze, Werner/Koselleck, Reinhart (Hg.): Geschichtliche Grundbegriffe. Stuttgart. S. 508-551.
- Von Saldern, Matthias (1991): Erziehungswissenschaft und Neue Systemtheorie. Berlin.
- Walter-Busch, Emil (1996): Organisationstheorien von Weber bis Weick. Amsterdam.

- Widmer, Thomas (2000): Qualität der Evaluation. In: Stockmann, Reinhard (Hg.) (2000): Evaluationsforschung. Opladen, S. 77-102.
- Wulf, Christoph (Hg.) (1972): Evaluation. Beschreibung und Bewertung von Unterricht, Curricula und Schulversuchen. München.