

Evaluation und ihre Folgen: Eine unterschätzte pädagogische Herausforderung

KURZFASSUNG: Dieser Beitrag präsentiert Ergebnisse einer umfangreichen empirischen Arbeit, die sich mit Rezeption, Transfer und Nutzung von Evaluationsbefunden in der pädagogischen Praxis befasst. Ausgehend von der Feststellung, dass die Nicht-Präsenz der Erziehungswissenschaft in der Evaluationsforschung dazu geführt hat, dass aus pädagogischer Perspektive bislang kaum Antworten auf die Nutzungsproblematik möglich sind, wird eine Typologie von Evaluationsnutzung als mögliche Antwort vorgestellt. Da die Befunde insgesamt jedoch zwiespältig sind, wird in der Konsequenz versucht, die Anwendungsfähigkeit von Evaluation auf eine neue Basis zu stellen. Dies geschieht im Sinne einer Politikberatung über zwei verschiedene, an Evaluatorinnen und Evaluatoren einerseits und an Entscheidungskader andererseits gerichtete Perspektiven. Sie münden schliesslich in die Reflexion, dass die Zukunft von Evaluation mit einem selbstbescheidenden Einsatz einhergehen sollte.

1. Einleitung

Bestimmte Ideen haben die Fähigkeit, mit einer besonderen Nachhaltigkeit grosse Teile der geistigen Welt zu besetzen, indem sie die Lösung fundamentaler Probleme oder Fragestellungen versprechen. Ein Beispiel dafür bildet die Idee der Evaluation in der aktuellen Bildungslandschaft, die kaum auf empirischer Basis untersucht oder in historischer Perspektive beleuchtet, jedoch umso stärker propagiert wird. Ohne Zweifel ist sie zu einer „*grande idée*“ (GEERTZ 1999) geworden, vergleichbar mit dem Begriff der ‚Innovation‘ der siebziger Jahre (BROCKMEYER 1978, FEND 1980). Mit ungeheurer Wirkung infiziert sie alle Bildungsbereiche und bestimmt weitgehend und fast frei von jeglicher Skepsis den Bildungsdiskurs. Obwohl ein solcher inflationärer Einsatz schon fast beklagenswert erscheint, muss man ihn als Ausdruck einer als gültig vorausgesetzten Prämisse akzeptieren, der Beizug von Evaluationswissen bilde gewissermassen eine *via regia* für die Optimierung von Bildungsleistungen. Auf ein solches Geschäft kurzgeschlossener Kausalität hat sich nun auch die Pädagogik erfolgreich eingelassen, ohne sich dabei zu überlegen, dass sie damit in gewissem Sinn ein ihr inhärentes altes Problem unter dem neuen Blickwinkel der Evaluationsforschung betrachtet, das Problem nämlich, wie theoretisches Wissen in die pädagogische Praxis überführt werden kann. Solche Probleme stellen sich der Erziehungswissenschaft, seit sie als Pädagogik versucht, sich als eigene Wissenschaft zu begründen, so wie dies HERBART oder SCHLEIERMACHER getan haben. Eigentlich dürfte man vermuten, dass sie inzwischen eine Lösung gefunden und mit der entsprechenden Erfahrungsbasis ihren Platz auch in der Evaluationstheorie gefunden hat. Das ist jedoch nicht, auch nicht ansatzweise, geschehen: Die Nicht-Präsenz der Erziehungswissenschaft in der Evaluationsforschung widerspiegelt sich nur allzu deutlich in der entsprechenden Fachliteratur. Selbstkritische Bestandesaufnahmen, theoretische Grundlagen oder die Klärung wissenschaftstheoretisch strittiger Fragen sind für den pädagogischen Bereich rar, methodische Aufsätze und Relevanzdiskussionen, analog dem angelsächsischen Raum, fehlen fast gänzlich. Auch die empirische Bildungsforschung, welche wesentlich zur Entfaltung einer systematisierten und methodisierten, des-

kriptiven und erklärenden Betrachtungsweise taktischer Entwicklungen im Bildungswesen beigetragen hat, ist es nicht gelungen, einer pädagogisch orientierten Evaluationsforschung zum Durchbruch zu verhelfen. Bis heute scheint sie nicht existent zu sein, so dass Fragen, beispielsweise was mit dem zur Verfügung gestellten Evaluationswissen geschieht, wie sich sein Stellenwert für die Weiterentwicklung der Praxis präsentiert, ob dieses Wissen richtig oder falsch oder überhaupt rezipiert und adäquat weiterverwendet wird, aus pädagogischer Perspektive nahezu unbeantwortet geblieben sind. Geht man nun allerdings von der Satzung aus, dass Evaluation ein pädagogisches Geschäft sein soll, das sich mit Praxisrelevanz auseinanderzusetzen hat, so ist Evaluationsforschung für eine erfolgreiche Legitimation nicht nur auf Selbstwertargumente, sondern auch auf Verwendbarkeits- und Anwendungsnachweise angewiesen.

Solche Reflexionen waren es, die mich angesichts der Hochkonjunktur, die dem ‚Prinzip Evaluation‘ zur Zeit beschieden ist, dazu führten, auf der Basis einer breit angelegten Feldstudie eine empirisch fundierte Diskussion über Verwendungsprobleme und Verwendungserfolge zu lancieren, die als kritische Diagnose wie auch Prognose Antworten zur Rolle der Evaluation als pädagogische Vermittlungswissenschaft liefern kann. Dabei handelt es sich freilich nicht um eine gänzlich neue Fragestellung. Sie ist bereits am Rande der Forschungsarbeiten zur Curriculumevaluation der siebziger Jahre (FREY 1975, HAFT & HAMEYER 1975, HAMEYER et al. 1983, NEUL 1977, WULF 1975) und der wissenschaftlichen Begleitforschung (FEND 1980, GSTETTNER & SEIDL 1974, LICHTENSTEIN 1983, MITTER & WEISHAUPT 1977, 1979) verschiedentlich aufgetaucht. Darüber hinaus bietet die Wissensverwendungsforschung Einsicht in die Transformation von wissenschaftlichem Wissen in praktische Erkenntnis (BECK & LAU 1982, BECK & BONSS 1989, DEWE 1988, DEWE & FERCHHOFF 1989, DRERUP 1982). Ferner liefert die amerikanische *research on evaluation utilization* zahlreiche Beispiele von Bestandesaufnahmen zur Verwendungsproblematik von Evaluationswissen (ALKIN et al. 1979, ALKIN 1980, DICKEY 1981, GREENE 1987, WEISS & BUCUVALAS 1980).

Leitintention meiner Arbeit war dabei, die Problemformel zu klären, Evaluation könne sich nur legitimieren, wenn sie ihren Anspruch nach praktischer Nutzbarmachung einlöse und die pädagogische Praxis das zur Verfügung gestellte Problemwissen auch umsetze. In diesem Aufsatz möchte ich einen spezifischen Teil dieser Problematik herausgreifen und danach fragen, wie die vorhandenen Informationen genutzt werden, welche Charakteristika sich zeigen und ob sich unterschiedliche „operative Muster“ (DEWE & FERCHHOFF 1989, S. 23) als Nutzungstypen ergeben. Dabei gehe ich so vor, dass ich zuerst die Feldstudie inklusive der empirischen Typenbildung vorstelle (2) und anschliessend eine Charakterisierung der vier Typen – ‚Blockade‘, ‚Alibi‘, ‚Reaktion‘ und ‚Innovation‘ – vornehme (3). Sie führt mich zu einer differenzierten, insgesamt jedoch zwiespältigen Beurteilung, die in drei pädagogische Konsequenzen im Sinne einer Politikberatung (4) mündet. Dass Evaluation bilanzierend schliesslich als fragliches Geschäft bezeichnet werden muss, eröffnet ihr gleichzeitig die Chance, zukünftig als knapper eingesetztes Gut mehr Wirksamkeit als Bildungsprozess zu entfalten (5).

2. Die Feldstudie: Erhebungsmethoden und empirische Typenbildung

Das Untersuchungskonzept der Feldstudie basierte auf einem theoretischen Modell mit insgesamt neun Schlüsselfaktoren, sechs unabhängigen Variablen (Evaluationsprozess, Entscheidungsträger, EvaluatorIn, Organisationsmerkmale, Kontext und Interessen, Nützlichkeit und Anwendbarkeit der Befunde) und drei abhängigen Variablen (Dissemination und Diffusion, Verwendung, Auswirkungen). Nebst der Überprüfung von sieben Hypothesen, die hier nicht weiter erörtert werden, diente es dazu, die Varianz von Rezeption, Transfer und Nutzung von Evaluationswissen zu klären. Um die Fragestellung auf kontrollierte Weise angehen zu können, wurde die Methodologie der vergleichenden Fallanalyse (KELLE & KLUGE 1999, KLUGE 2000) mit empirischer Typologie als komparativer Ansatz gewählt, der aufgrund seiner vergleichsdimensionenorientierten Basis in der Lage ist, möglichst alle mit der Nutzung in Zusammenhang stehenden Erscheinungen in ‚dichter Weise‘ zu beschreiben (GEERTZ 1999). Auf der Basis solcher Überlegungen und unter Vorgabe eines Bedingungsrastrers wurde schliesslich eine Vollerhebung aller in der deutschen Schweiz verfügbaren, von externen Expertinnen und Experten verantworteten und zwischen 1995 und 1998 abgeschlossenen Evaluationen als Querschnittsanalyse durchgeführt. Mit der Vollerhebung konnte die Varianz des Untersuchungsfeldes sichergestellt und die Vermeidung von Verzerrungen garantiert werden. Da alle 18 in Frage kommenden Evaluationsprojekte einbezogen werden konnten, sind Repräsentativität und Verallgemeinerbarkeit der Ergebnisse für diesen geografischen Raum gewährleistet.

Um die Rezeptions-, Transfer- und Transformationsprozesse in der Umsetzungsphase empirisch eruieren zu können, gelangten insgesamt drei unterschiedliche Methoden zum Einsatz, die Dokumentenanalyse, das leitfadengestützte fokussierte Interview sowie quantifizierte Befragungsinstrumente in Form von Fragebögen. Die Interviews wurden ab Tonband transkribiert und produzierten insgesamt 1064 Seiten Auswertungstext. Dazu war eine Codierung des Textmaterials notwendig, deren Gültigkeit mit dem Verfahren der Inter-Coder-Reliabilität überprüft wurde. Demzufolge lagen zu jedem Fall qualitative Aussagen und quantifizierte Daten vor, die anschliessend mit einem kombinierten Modell empirischer Typenbildung (KUCKARTZ (1999a, b; KELLE & KLUGE 1999, KLUGE 1999, 2000) verarbeitet werden konnten. Die Datenauswertung erfolgte dabei mit dem Textdatenbanksystem winMAX, das die Verwaltung und Auswertung des umfangreichen Datenmaterials ermöglicht und erlaubt, die individuellen Daten für den Einzelfall, aber auch für den Fallvergleich, darzustellen. Mit winMAX besteht zudem die Möglichkeit, Textinformation in Form von Variablenwerten auf einfache Weise zu vercoden und mit diesen Variablen quantitative Auswertungen, beispielsweise Faktoren-, Korrespondenz- oder Clusteranalysen, durchzuführen. In einem ersten Schritt konzentrierte ich mich gemäss dem Stufenmodell nach KLUGE (1999, S. 260) auf die Erarbeitung relevanter Vergleichsdimensionen, die auch für die Forschungsfrage aussagekräftig waren, auf die ‚direkte Rezeption und Verwendung‘ und die ‚Zielsetzung der Evaluation‘. Die Fälle wurden anschliessend anhand dieser Merkmale gruppiert, die ermittelten Gruppen hinsichtlich der empirischen Regelmässigkeiten untersucht und kreuztabellelliert, was die Zuordnung der 18 Evaluationen zu den einzelnen Feldern sichtbar machte. Parallel dazu wurde eine Clusteranalyse gerechnet, gewissermassen als systematische Auswertung dieser Ähnlichkeitsmatrix. Die vergleichende Analyse des empirischen Datensatzes erfolgte mit verschiedenen Algorithmen (Complete-,

Single- und Average-Linkage-, Median-, Zentroid- und Ward-Verfahren. Dabei wurde untersucht, welche der insgesamt fünf Clusterlösungen den Ergebnissen der Vergleichsdimensionen am ehesten entsprach. Mit dem Complete-Linkage-Verfahren, dem Ähnlichkeitsmass der quadrierten euklidischen Distanz, ergab sich mit vier Clustern eine nahezu identische Lösung zum klassischen Verfahren. In Tabelle 1 werden die Clusterprofile dargestellt. Sie enthält für jedes Cluster den prozentualen Anteil, welcher das Cluster an der jeweiligen Merkmalsausprägung aufweist. Diese Matrix war Bestandteil für die nachfolgende Interpretation.

Item	Cluster 1* n Fälle = 5	Cluster 2* n Fälle = 3	Cluster 3* n Fälle = 6	Cluster 4* n Fälle = 4
Überprüfung	33	33	29	44
Legitimation	33	67	21	55
Umsetzung	33	0	60	0
bildungspolitische Bedeutung	67	50	56	78
politische Brisanz	67	33	50	67
Zufriedenheit mit Ergebnissen	44	33	79	33
positive Gesamtbilanz	44	17	100	44
Widerstand	33	0	14	78
Bottom-up Struktur	33	0	78	44
Top-down Struktur	67	100	21	56
direkte Verwendung	33	33	100	100
Veränderungsbereitschaft	44	33	86	78
Unterstützung ED	33	33	86	67
Berufsorientierung ‚Umsetzung‘	22	67	29	22
Berufsorientierung ‚Wissenschaft‘	78	33	71	78
Zensuransätze	44	78	71	78
Umfang Dissemination	33	70	93	78
sprachliche Qualität	44	67	100	100
Termintreue	67	33	100	33
methodische Glaubwürdigkeit	67	50	100	78
finanzielle Ressourcen	11	17	57	57
Commitment Evaluator/in	11	44	71	56
Reputation Evaluator/in	22	100	71	100
Nützlichkeit/Anwendbarkeit	33	67	50	50
Zufriedenheit mit Nutzung	11	33	93	50
Erfahrungen mit Evaluation	67	67	17	67
Engagement Entscheidungsträger	22	50	100	86
<i>Lesehinweis:</i> • Alle Angaben in Prozent. • 44 % der Evaluationen des ersten Typus weisen eine hohe sprachliche Qualität auf. In Cluster 1 beispielsweise ist nur in 33 % der Fälle Widerstand aufgetreten und in 11 % bestand Zufriedenheit mit der tatsächlichen Nutzung. Ganz anders hingegen Cluster 4: Hier bestand in 78 % der Fälle Widerstand und die Zufriedenheit mit der Nutzung umfasste 50 %. • *Cluster 1 entspricht der ‚Blockade‘, Cluster 2 dem ‚Alibi‘, Cluster 3 der ‚Innovation‘ und Cluster 4 der ‚Reaktion‘.				

Tabelle 1: Clusterspezifische Merkmalsverteilung

Insgesamt lagen nun die Ergebnisse der Clusteranalyse und des empirischen Stufenmodells vor. Der Vergleich der beiden Lösungen verdeutlichte eine grosse Übereinstimmung. Die Cluster 2 und 4 stimmten vollständig überein, Cluster 1 und 3 unterschieden sich in Bezug auf einen Fall voneinander. Nach Rückgriff auf die Originaldaten und die Einzelfallanalyse wurde schliesslich der *Lösung des Stufenmodells* den Vorzug gegeben (Tabelle 2) und anhand der Merkmalskombinationen und Sinnzusammenhänge eine Charakterisierung der Typen vorgenommen, die schliesslich zu den Bezeichnungen ‚Blockade‘, ‚Innovation‘, ‚Alibi‘ und ‚Reaktion‘ führte. Die Kunst der Typenbezeichnung lag dabei darin, eine Formulierung für das Charakteristische in den Einzelfallanalysen zu finden, die zugleich ihr gemeinsam Geteiltes und ihr zu anderen Fällen Differentes darstellen, ohne dass es dabei zu Verkürzungen oder Verzerrungen der Komplexität des Untersuchungsmaterials kommen sollte.

Rezeptions- und Verwendungstyp	Zweck/Ziel		
	Kontrolle	Legitimation	Entwicklung/ Optimierung
direkte Rezeption/ Nutzung nachweisbar	REAKTION (n = 4)		INNOVATION (n = 5)
direkte Rezeption/ Nutzung unklar	BLOCKADE (n = 6)	ALIBI (n = 3)	
direkte Rezeption/ Nutzung nicht vorhanden			

Tabelle 2: Vier Typen der Beziehung zwischen Rezeption und Ziel/Zweck

3. Die vier Typen von Evaluationsnutzung

Das kombinierte Modell empirischer Typenbildung hat im Ergebnis zur Formulierung dieser vier nachfolgend charakterisierten Typen geführt, auf die hin die Einzelfälle konvergieren. Sie bilden damit die Varianz, mit der Evaluationsnutzung insgesamt verstanden werden kann. Damit stehen vorerst vier Schemata zur Verfügung, die sich methodologisch gesehen als Elemente einer ‚Grammatik‘ der Rezeptions- und Nutzungsfrage von Evaluation begreifen lassen.

Der Blockade-Typus

Evaluationen, welche primär der Überprüfung dienen und kaum direkte Nutzung von Evaluationswissen aufzeigen, bilden den Typus der Blockade und zwar deswegen, weil bestimmte Gründe unterschiedlicher Natur eine unmittelbare Umsetzung der Befunde verhindern. Damit trifft die Bezeichnung ‚Blockade‘ ins Schwarze, weil sie die metaphorische Bedeutung einer Sperre annimmt, welche an der Verhinderung weiterer Aktivitäten beteiligt ist.

Die Blockade als Nutzungstypus lässt sich vorerst über die beiden Vergleichsdimensionen Überprüfung als Hauptzielsetzung sowie unklare oder nicht vorhandene direkte Rezeption definieren. Da die Entscheidungskader der Blockade-Typik mehrheitlich über einen erziehungswissenschaftlich-pädagogischen Hintergrund verfügen, sind sie zwar in der Lage, das evaluativ gewonnene Wissen

autonom zu bewerten. Trotzdem muss ihre Problemerkfassung beschränkt bleiben, da Interpretationen nur soweit möglich sind, als dies ihre hierarchischen Organisationsstrukturen und die ausgeprägten Delegationsstrukturen zulassen (vgl. dazu auch DUBS 2000). Sie bilden ungeeignete Bedingungen für Lernprozesse und Diskurspolitik, aber geeignete Blockaden, die den Sprung zur Weiterverwendung der Ergebnisse verhindern oder verdunkeln können. Besonders deutlich offenbart sich im Typus der Blockade die Notwendigkeit, die strukturellen Dimensionen und latenten Schattierungen der Makroebene berücksichtigen zu müssen. Dies lässt die Evaluation allerdings zum Teilstück eines politischen Spiels werden. Eine Blockade-Evaluation kann deshalb nur kontextuelle Relevanz erzielen, wenn sie die Bedingungen des Makrokontexts berücksichtigt, was die lokale Verwendung jedoch problematisch werden lässt.

Der Blockade-Typus unterscheidet sich von den drei anderen Typen vor allem in der Dimension der Nutzen- und Anwendbarkeitseinschätzung. Die Erfahrung mit Evaluation und ihr fachliches Know-how erlaubt den Verantwortlichen, aus der Vielzahl von Rezeptionsmöglichkeiten für den Umgang mit diesem ‚anderen‘, herausfordernden Wissen eine geeignete Variante auszuwählen, nach dem Motto ‚Wie stelle ich die Suppe beiseite, damit wir sie nicht zu heiss löffeln müssen?‘ Als an Wissenschaft gewöhnte Professionelle sind sie in der Lage, in verwissenschaftlichten Formen zu argumentieren, Befunde zu relativieren und sie Evaluationserkenntnissen aus anderen Untersuchungen gegenüberzustellen. Deshalb werden die Befunde zuerst einmal zur Kenntnisnahme an andere Gremien weitergereicht und über viele Umwege dann schliesslich doch in Absichtserklärungen mit hinausschiebendem Charakter überführt, um die aktuellen bildungspolitischen Zielsetzungen nicht zu gefährden. Da solche Entscheidungskader die Einschätzung von Nützlichkeit und Anwendbarkeit von Evaluationsinformation immer stark auf ihre politische Verwendungstauglichkeit hin befragen, sind Evaluationen dieses Typs vordergründig gar nicht in der Lage, den für die Auftraggeber relevanten Informationsbedarf zu befriedigen. Damit wird die Blockade zum Paradebeispiel der Frage, welchen Voraussetzungen denn Evaluationen genügen müssen und genügen können, um einen höheren Einfluss zu erlangen.

Der Typus der Innovation

In diesem Typus finden sich Evaluationen, welche eindeutig entwicklungs- und umsetzungsorientierten Charakter haben. Sie lassen sich kennzeichnen durch geplante, umfassende und anhaltende Umsetzungsmassnahmen, welche den Charakter von Innovationen annehmen. Idealtypische Konturen ergeben sich aus der hohen Qualität der Evaluation, den organisationalen Merkmalen, welche Autonomie und Selbstverantwortung fördern und letztlich Dissemination und tatsächliche Nutzung als Folgen eines fruchtbaren Prozesses erscheinen lassen.

Die Hauptkennzeichen der Innovation liegen in der Umsetzungsorientierung und der direkten Rezeption und Nutzung. Sie unterscheidet sich von allen anderen Typiken durch die höchste Zufriedenheit mit der tatsächlichen Umsetzung bei gleichzeitig hohem persönlichem Einsatz der Entscheidungsverantwortlichen und der Evaluationsforschenden. Die eher bottom-up strukturierten Bildungsadministrationen zeigen relativ grosse Veränderungsbereitschaft, die sich auch in nicht-hierarchischen Beziehungsstrukturen zwischen Auftraggebern und Auftragnehmern äussern. Im Ganzen gesehen zeichnet sich der Innovationstypus durch eine

hohe Qualität der Evaluation (sprachliche Verständlichkeit, Termintreue, Methodenvwahl) und eine hohe Verwendungstauglichkeit der Befunde aus.

Als Gegenpol des Blockade- und des Alibi-Typus beinhaltet die Innovation viele Aspekte einer idealtypischen Konstruktion von Evaluationsnutzung. Im Gegensatz zur Blockade sind Entscheidungsverantwortliche der Innovation allerdings auch geneigt und in der Lage, den Evaluationsgegenstand selbst zum Problem zu machen und sich persönlich in Lernprozesse einzulassen, Veränderung anzustreben und nicht primär als bedrohlich zu empfinden. Wirklichkeitserfassung wird damit in der Innovations-Typik zur Problemerkfassung. Den Entscheidungsträgern, die häufig in Personalunion auch Projektleitung und Auftraggeberrolle innehaben, kommt dabei eine Schlüsselrolle zu. Mit ihrem hohen Engagement schaffen sie für alle Beteiligten ein günstiges, offenes Arbeitsklima und sichern die Überführung der Ergebnisse in gestaltende Massnahmen. Die günstigen politischen Konstellationen, etwa die Unterstützung durch die bildungspolitische Behörde – auch im Falle personeller Wechsel – tragen dazu bei, dass die Evaluation zur persönlichen Angelegenheit der Verantwortlichen wird und Erkenntnis- und Handlungsinteressen konvergieren können. Zur idealtypischen Konturierung tragen jedoch auch starke Evaluatorenpersönlichkeiten bei, die nicht nur über hohe fachliche Qualität und Feldkenntnis verfügen, sondern auch im Disseminationsprozess eine bedeutende Rolle spielen und bei der Dateninterpretation unterstützend zur Seite stehen.

Allerdings ist die Innovations-Typik auch gefährdet. Ihr Hauptrisiko besteht in der prozessorientierten Arbeitsweise, die grosse zeitliche und finanzielle Ressourcen bindet und den evaluativen Charakter zu Gunsten einer allgemeinen Schulentwicklungsperspektive zu verlieren droht. Die Gefährdung betrifft auch Position und Rolle der evaluierenden Personen, die aufgrund des eher formativ anmutenden Evaluationscharakters und der Familiarität mit den Entscheidungsverantwortlichen ihre externe neutrale Position zu Gunsten einer Insiderposition aufgeben und so ungewollt in die Rolle einer verbündeten Person schlüpfen. Innovationstypen, die als externe Evaluationen deklariert werden, machen jedoch nur so lange Sinn, als Evaluatorinnen und Evaluatoren tatsächlich den Expertenstatus beibehalten und sich gegenüber den Auftraggebern und Akteuren auch eindeutig als solche darstellen und solange, als die pädagogische Basis tatsächlich ein Laiengremium bleibt und nicht zu einer selbsternannten Expertengruppe wird. Dazu kommt, dass die Typik der Innovation mit ihren prozessorientierten, bottom-up gesteuerten Evaluationen den bildungspolitischen Behörden weit mehr Sorge bereitet als top-down-orientierte Ansätze, die besser kontrollierbar und deshalb weniger beängstigend sind. Demgegenüber sind die rationalen Taktiken der Reaktion oder die politisch gefärbten Taktiken des Alibis und der Blockade kontrollierbarer und deshalb akzeptierter.

Die Typik des Alibis

Dieser Typus setzt sich aus Evaluationen zusammen, die in erster Linie legitimierenden Charakter haben und keine direkten Evaluationsfolgen nachweisen können. Dabei handelt es sich um sogenannte Bestätigungsevaluationen, deren erste Motivation in der Beglaubigung oder Bestätigung vermuteter Ergebnisse liegt und hier deshalb als Alibi-Evaluationen bezeichnet werden.

Legitimierende Zielsetzung, keine oder unklare Rezeption und Verwendung – auf diese Kurzformel gebracht bildet die Typik des Alibis den Gegenpart der

Innovation. Dort Optimierung, Entwicklung und Nachhaltigkeit, hier Wirkungslosigkeit, Frustration, Bestätigung des schon vorher Gewussten. Im Ganzen gesehen unterscheidet sich das Muster des Alibis von den drei anderen Nutzungstypen durch den Legitimationsdruck, der auf der Evaluation lastet, durch die bescheidene Finanzierung sowie durch bereits vorhandene Erfahrung mit Evaluation. Trotz Handlungsdruck werden Alibi-Evaluationen eher mit Widerwillen durchgeführt, bringen jedoch wegen der bestätigenden Ergebnisse den Verantwortlichen massive Entlastung, nicht zuletzt auch, weil sie Material für eine ‚Absicherung für alle Fälle‘ liefern und einen Schlusspunkt unter die bisherige Projektarbeit setzen. Alibi-Evaluationen bringen somit auch Entwarnung. Zwar mit einer ähnlichen Zielsetzung wie Evaluationen des Musters Reaktion bedacht, unterscheiden sie sich von diesen jedoch durch die bereits vor Evaluationsbeginn feststehende und von den Ergebnissen unabhängige Entscheidungsfindung. Alibi-Typen kennzeichnen weder Erkenntnis- noch Handlungs- oder Verwendungsinteresse. Sie haben lediglich Bestätigungs- oder Abschlusscharakter, gefragt ist deshalb Zweckrationalität, jedoch keine Sinnrationalität, und es gilt die Devise: Evaluation durchführen und sich nicht beirren lassen. Deshalb steht auch die Frage nicht zur Diskussion, ob Evaluationen in der Lage sind, den für weitere Entscheidungen relevanten Informationsbedarf zu befriedigen. Evaluationswissen, auch wenn adressatenorientiert aufbereitet, wird kaum weiterverarbeitet, sondern eher ignoriert oder bestenfalls argumentativ zur Rechtfertigung festgelegter Entscheide benutzt. Fundierte methodologische oder technische Evaluatorenarbeit hat ebenso geringe Relevanz wie die Verwendungstauglichkeit der Befunde, obwohl sie in Evaluationen dieses Typs durchwegs positiv eingeschätzt wird. Damit wird nur allzu deutlich, dass Rezeption und Gebrauch nicht lediglich davon abhängen, ob Nutzen- und Anwendungseinschätzung positiv ausfallen, sondern weit stärker mit anderen kontextuellen Mechanismen verbunden sind. Deshalb gestaltet sich die Arbeit von Evaluatorinnen und Evaluatoren in einer Alibi-Typik als schwierige Aufgabe. Sie sind nicht nur stark auf sich gestellt, sondern können sich auch kaum entfalten und werden darüber hinaus mit strikten Vorgaben und Regeln eingedeckt, die ihre Rolle im Hinblick auf Dissemination und Datennutzung bestimmen. Verständlich also, wenn sich Evaluatorinnen und Evaluatoren ärgern und sich als Opfer irreführenden Verhaltens und falscher Vorstellungen empfinden, wenn sie feststellen müssen, dass ihre Arbeit nicht ernst genommen wird, die Befunde lediglich zur Rechtfertigung der erwünschten Praxis dienen und den Einsatz politischer Macht verfolgen, jedoch nicht als Instanzen kritischer Handlungskontrolle und rationaler Handlungssteuerung betrachtet werden. Sie müssen erkennen, dass ihre Evaluationen Informationen liefern, die nicht gebraucht, nicht zur Kenntnis genommen werden und ohne vordergründige Folgen bleiben. Verdeckt bleibt allerdings die wichtige, strategisch-symbolische Nutzung, die in der nachträglichen Legitimierung dessen besteht, was vorgängig aus anderen Gründen entschieden worden ist, ein Vorgang, für den DREURPS Bezeichnung „legitimatorische Legendebildung“ (1982, S. 165) zutreffend ist.

Die Typik der Reaktion

Dieser Typus beinhaltet Evaluationen, die zwar auf die Überprüfung und Kontrolle ausgerichtet sind, jedoch eine klare Rezeption der Ergebnisse vorweisen können. Am augenfälligsten ist dabei, dass dies als Reaktion geschieht, bestimmte Be-

schlüsse oder Massnahmen infolgedessen direkte Antworten auf Forderungen bestimmter Anspruchsgruppen darstellen.

Die Definition dieser Typik gründet in der Schnittstelle zwischen direkter Rezeption und der Vergleichsdimension Überprüfung/Kontrolle. Obwohl die Entscheidungsverantwortlichen die Schlüsselrolle im Umsetzungsprozess übernehmen, sind es die bildungspolitischen Behörden, welche die Entscheide treffen. Das mag ein Grund dafür sein, dass sich teilweise verkürzte Verwendungsinteressen asymmetrischer oder gar komplementärer Art durchsetzen können, die – ohne Rücksicht auf die Botschaft der Evaluation – technologische, praktische, aber auch ideologische Ansprüche verfolgen und dem Zweck oder Sinn entgegengesetzte Nutzungsanstrengungen nach sich ziehen, ohne Rücksichtnahme auf das persönliche Wohlbefinden einzelner Individuen oder Gruppen. In dieser Hinsicht unterscheidet sich die Reaktion von der Innovation, die eher auf diskursiv ausgehandelte Nutzungswege und Lerneffekte setzt. Dort, wo die Reaktions-Typik von vorwiegend technologischen Verwendungsinteressen geprägt ist, gelten solche Erkenntnisse als verwendungsrelevant, die objektiv sind und strengen wissenschaftlichen Kriterien genügen. Von Evaluatorinnen und Evaluatoren wird deshalb erwartet, dass sie die letzten Sicherheiten pädagogischer Argumentation liefern, um die Evaluation mit dem Siegel wissenschaftlicher Redlichkeit auszeichnen zu können.

Im Mittelpunkt der Reaktions-Typik stehen deshalb Versachlichungshoffnungen und ein hohes Handlungsinteresse der Entscheidungsverantwortlichen. Folglich besteht grosses Interesse an Evaluationsergebnissen, die spezifisch methodisch abgesichert und bestätigt sind und deshalb als gesicherte Grundlagen gegenüber Mutmassungen, spekulativem Vorwissen und dem Widerstand der Feldakteure eingesetzt und der bildungspolitischen Bedeutung der Evaluation gerecht werden können. Deshalb wird im Reaktions-Typus die Umsetzung der Ergebnisse, das Praktisch-Werden der Befunde, prioritär behandelt, der Erkenntniszuwachs scheint aber als willkommenes Nebenprodukt auf. Ähnlich dem Blockade-Typus, jedoch in deutlicher Abgrenzung zur Innovation, basieren Umsetzungshandlungen auf der Einschätzung der Verwendbarkeit in Anlehnung an politisch gefärbte Konstellationen. Positive Ergebnisse bringen dort Bestätigung und Entlastung, wo sie wie erwartet ausfallen, negative Erfolgsbilanzen führen jedoch zu Kritik, auf die mit schnellen Entscheiden reagiert wird. Dass auch Evaluationen mit der Zielsetzung Überprüfung/Kontrolle direkte Rezeption und Verwendung erzeugen, widerspricht dem in der Evaluationsliteratur postulierten üblichen Bild. Allerdings verkörpert die Reaktion nicht das Beispiel einer reibungslosen Transferevaluation, die Wissen gradlinig in Umsetzung befördert, sondern ein von erheblichen Unruhen begleitetes Unternehmen, das Folgeaktivitäten als ‚Notmassnahmen‘ geradezu provoziert.

4. Pädagogische Konsequenzen als Politikberatung

Im Ganzen gesehen legt diese vier Nutzungstypen nahe, dass direkte Rezeption und Nutzung deutlich häufiger festgestellt werden kann als in der einschlägigen amerikanischen Literatur behauptet, dass sie jedoch nicht nur als gradliniger Transfer von den Evaluatoren zu den Anwendersystemen vonstatten geht. Direkte Nutzung von Evaluationsinformation ist somit kein „unhaltbarer Mythos“ (DRERUP

1982, S. 104), keine Wunschvorstellung der Forschung, sondern pädagogische Realität, die in den beiden Typiken Innovation und Reaktion ihren Ausdruck findet und damit Annahmen über die ausbleibende direkte Verwendung von Evaluationswissen demontiert. Zugleich gibt die Typologie zu erkennen, dass direkte Nutzung auch negative Wirkung entfalten kann oder dass das Ausbleiben direkter Evaluationsfolgen trotzdem zu Wirksamkeit führen kann. Die Anteile, welche nicht im intendierten Schema ablaufen, sondern verschlungener Wege gehen, sind beträchtlich. Jedenfalls geschieht Nutzung nicht ausschliesslich über eine lineare Wissensakkumulation oder -assimilation im Sinne eines Transfers von Empfehlungen oder Folgerungen, sondern grossenteils als Transformation, als vielschichtige Umformung und Umgestaltung von Evaluationswissen in alltägliche „Praxisdeutungen“ (DEWE 1988), bis sie mit dem Sinnhorizont der Abnehmer und Anwender kompatibel sind. Erst wenn sich die Anwendungsfähigkeit der vorgezeigten Optimierung- und Veränderungsalternativen erwiesen hat, d. h., wenn sie mit den Perspektiven und Interessenkontexten der Auftraggeber und Akteure, mit ihrer kognitiven Kompetenz und ihren Vorstellungen von Wahrheit konvergieren, kann Verwendung erfolgen. Aus Evaluatorensicht sind solche Transformationsleistungen allerdings ein Problem, vielleicht sogar eine Kränkung, weil sie ihr sorgfältig generiertes Wissen grossenteils verändert, selektiv oder vereinfachend gebraucht oder nicht gebraucht sehen. Sie müssen beobachten, wie Entscheidungsträger und Feldakteure den Sinnzusammenhang der Befunde auf ihre Weise konstruieren, ihn mit ihrem Verstehenshorizont verbinden und somit immer Reduktionen, Vereinfachungen oder Umwandlungen vornehmen.

Solche Ergebnisse provozieren den Versuch, die Anwendungsfähigkeit von Evaluation auf eine neuen Basis zu stellen. Das soll im Folgenden als Politikberatung in zwei Perspektiven geschehen, als ‚wirksame Wege zur Nutzung‘ und als ‚Wege zur wirksamen Nutzung‘. Dabei wird davon ausgegangen, dass Evaluatoreninnen und Evaluatoren *wirksame Wege* vorzeichnen können, welche einer Nutzung durch die Abnehmersysteme förderlich sind. Sie betreffen die Optimierung der Methodologie und die Reflexion der Anwendungspraxis. Die zweite Perspektive betrifft die Entscheidungsträger und Projektleitungen selbst, die bestimmte Strategien verfolgen und Wege beschreiten können, die eine *wirksame Nutzung* nach sich ziehen. Dazu gehörten in erster Linie eine stringente Planung des Evaluationsvorhabens inklusive der Begleitung der Evaluation sowie einer Evaluierbarkeitsprüfung, welche den Nachweis liefern soll, dass Evaluation tatsächlich das geeignete Instrument für die intendierte Überprüfung, Kontrolle oder Entwicklung der Massnahme darstellt. Beide Perspektiven werden jedoch von einer unsichtbaren dritten Perspektive begleitet, die sich auf nicht-kontrollierbare Faktoren ausrichtet und deshalb die Angelegenheit anfälliger und unsicherer werden lässt.

4.1 Wirksame Wege zur Nutzung

Optimierung der Methodologie

Die Forderung nach methodischer Qualität legitimiert sich nicht nur durch solche, auf die Erhöhung der technischen Glaubwürdigkeit und der generellen Akzeptanz ausgerichteten Faktoren, sondern auch durch die Ausrichtung der Evaluation auf

ihre Funktion als Lieferantin gesicherten Wissens. Ihr wesentlicher Beitrag für die Nutzungsrelevanz liegt allerdings nicht in der methodischen Sophistikaion, vergleichbar mit den Standards der wissenschaftlichen Begleitforschung der siebziger Jahre, die bekanntlich trotz grosser Bemühungen um methodische und technische Perfektion von den Anwendersystemen aufgrund ihrer Komplexität nur schwer verstanden wurde und deshalb Quelle des Nutzungsdefizits geworden war. Gefragt ist eine Ausrichtung auf die Hauptaufgabe der abnehmergerechten Ergebnisaufbereitung, die sich auf neues, gültiges, organisiertes und übergebbares Wissen konzentriert und die soziale Wirklichkeit des Evaluandums einsichtig zu erklären imstande ist. Es soll verstehendes Wissen generiert, in einen praxisgerechten Sprachstil übersetzt und populärwissenschaftlich aufbereitet werden. Methodische Qualität muss das dem Sinn der Auftraggeber entsprechende ‚Richtige‘ und ‚Wahre‘ ausstrahlen, was die Evaluationsrolle als „Wahrheitsgarantin“ (OELKERS 2001, S. 12) freilich in einem spezifischen, problembeladenen Licht erscheinen lässt. Da nämlich der Transfer der Evaluationsinformation im Unterschied zur Forschung nicht in Richtung exklusiver Fachöffentlichkeit, sondern interessierter Laienschaft stattfindet, welche die Ergebnisse in ihre Vorstellungen transformiert und sie mit ihren eigenen Konzepten kombiniert, kommt Evaluation nicht umhin, sich in das Prokrustesbett der „verkürzten Forschung“ (CAMPBELL 1969) zu zwängen. Gleichwohl muss sie sich aber erinnern, dass Validität und Reliabilität eine gute Methodologie ausmachen und ausgewogene Untersuchungsdesigns und -analysen intelligent und sensitiv in die entsprechenden Kontexte eingebettet werden sollen. Saubere, für die Anwenderschaft nachvollziehbare methodische Kriterien gewinnen insbesondere dann an Relevanz, wenn Ergebnisse anders als erwartet ausfallen.

Reflexion der Anwendungspraxis

Evaluation setzt den Erwerb von Verstehen über ihre mehrfache Abhängigkeit von spezifischen, kontextuellen und politischen Bedingungen voraus. In erster Linie betrifft dies das Verständnis der Konturierung pädagogischer Akteure als mit reichem Erfahrungsschatz versehene Feldexperten, welche die Evaluation als Profilierungsfeld betrachten und mit ihrem persönlichen Erfahrungsfeld eine starke Konkurrenz bilden. Evaluatorinnen und Evaluatoren stehen damit vor einer besonderen Herausforderung: Sie sollen die Befunde zwar mit angemessenen Alternativen verknüpfen, die an Erfahrungen und kognitive Verstehenshorizonte anknüpfen, andererseits trotzdem weiterführende, neue Ergebnisse und Erkenntnisse liefern. Verstehen beinhaltet jedoch auch die Berücksichtigung des Wissens über in der Evaluation selbst verankerte Rezeptions- und Verwendungsschranken, die nur zu überwinden sind, wenn die Reflexion der Anwendungspraxis in Relation zu den institutionellen Rahmenbedingungen erfolgt. Ob Evaluation tatsächlich Lieferantin gesicherten Wissens sein kann, hängt nicht zuletzt vom Einfühlungs- und Urteilsvermögen der Evaluatorinnen und Evaluatoren als dem unverzichtbaren Verbindungsglied zwischen der Generierung von Evaluationswissen und dem Verwendungsinteresse der Abnehmer ab. Darüber hinaus müssen sie sich vergegenwärtigen, dass Ergebnisse, Alternativdeutungen und Handlungsvorschläge von den Abnehmern immer zweifach gewürdigt werden, erstens in Bezug auf die Sinnrationalität als der wissenschaftlichen Glaubwürdigkeit (*‚Wahrheitstest‘*) und zweitens in Bezug auf die Zweckrationalität als der Praktikabilität und Praxis-

ausrichtung (*Nützlichkeitstest*). Anwendung und Aneignung reagieren somit stark auf aktuelle und persönliche Bedürfnisse und weniger auf die historisch gewachsenen Hintergründe des Evaluationsgegenstandes. Nicht wissenschaftliche, sondern gesellschaftliche Ursachen bestimmen und prägen die Verwendungsinteressen der Entscheidungsträger. Damit ist auch zu erklären, warum das Störende einfach abgestossen (vgl. die Typik der Reaktion) oder blockiert wird (vgl. die Typik der Blockade). In positivem Sinne können theoretische Befunde jedoch auch in ein Konkurrenzverhältnis zum Vorverständnis und den Insiderkenntnissen des pädagogischen Feldes treten, dann nämlich, wenn Evaluаторinnen und Evaluatoren in der Lage sind, den Entscheidungsträgern ihren selektiven, marginalen oder kursorischen Wissensgehalt bewusst zu machen (vgl. die Typik der Innovation).

4.2 Wege zur wirksamen Nutzung

Planungskonzeption

Wirksame Nutzung ist Aufgabe der Entscheidungsträger und Auftraggeber. Ihr erster Ort bildet eine wirksame Planungskonzeption, die sich zuerst einmal auf die Kommunikation der Zielsetzungen der Evaluation konzentriert. Die Explizierung von Zielsetzung und Umgang mit den Daten ist vor allem in Projekten mit hoher bildungspolitischer Bedeutung wesentlich. Von Bildungsadministrationen werden sie jedoch oft absichtlich nicht exakt definiert, um sie allenfalls später den Erfordernissen entsprechend anpassen zu können. Vollständig ist eine Planung dann, wenn auch mögliche unbeabsichtigte und unerwartete Wirkungen diskutiert und Verbreitung und Verwendung vom ersten Moment der Konzeptionsphase an mitbedacht werden. Dazu gehört auch die Antizipation möglicher negativer Erfolgsbilanzen. Im Weiteren wäre in einer Planungskonzeption die Vermeidung von Evaluationsredundanz zu berücksichtigen. Eine solche ist gegeben, wenn verschiedene Reformprojekte oder Massnahmen parallel evaluiert werden und Steuerungsinstanzen und Abnehmersysteme gleichzeitig mit vielen Daten aus verschiedenen Überwachungs- und Systemsteuerungsprozessen konfrontiert werden, die letztlich eher verwirrende denn steuernde Wirkung entfalten können. Muss ein Organisationssystem mit so hoher Entscheidungsichte operieren, kann keine der Aufgaben seriös ausgeführt werden. Das hat zur Folge, dass der Verwendungskontext im Rahmen von Entscheidungsvorbereitung und politischer Fortschreibung nicht bedacht wird und Umsetzungsleistungen zufällig bleiben müssen. Wenn nicht genügend Zeit zur Verbreitung der Ergebnisse und zum produktiven Umgang mit den Erkenntnissen gegeben ist, steigen die Vorbehalte des pädagogischen Feldes gegenüber Umsetzungserwartungen und -forderungen der bildungspolitischen Behörde parallel zur Anzahl ihrer Ergebnissrückmeldungen.

Zu einem Planungskonzept gehört auch ein offener Diskurs zwischen Evaluatorenschaft und Auftraggeber über die Vor- und Nachteile der Methodenwahl. Dabei ist zu bedenken, dass dem Handlungsparadigma verpflichtete, auf qualitativen Methoden beruhende und die *stakeholders* einbeziehende Evaluationen immer eine diachrone Hauptaufgabe zukommt: Seine gezielt lernorientierte Ausrichtung bedingt, dass Evaluаторinnen und Evaluatoren nicht nur über ein gutes Mass an Responsivität, sondern dass auch Entscheidungsträger und Feldakteure selbst

über genügend zeitliche Ressourcen, mentale Kräfte und Diskursfähigkeit und -bereitschaft verfügen müssen, um sich in Lernprozesse einlassen zu können. Zudem ist die Kongruenz zwischen der Aufwändigkeit von Verfahren und der zeitlichen Verfügbarkeit zu berücksichtigen, denn diskursiv angelegte Verfahren sind ressourcenintensiv und produzieren keine schnell bereitstellbaren Ergebnisse.

Der letzte Punkt einer Planungskonzeption betrifft ein angemessenes, objektives Auswahlverfahren. Der Status quo zeigt, dass Vertrautheit mit der Person des Evaluators zwar häufigstes Auswahlkriterium ist, jedoch wenig geeignet scheint, um Unabhängigkeit, Fachkompetenz und Objektivität garantieren zu können. Ein submissionsorientiertes Auswahlverfahren fördert nicht nur die Professionalisierung des Berufsstandes, sondern macht auch ‚entgegengesetzte‘ Transferleistungen kontrollierbar, nämlich Entscheidungsträger, die mit Evaluationsforschenden in einem freundschaftlichen Vertrauensverhältnis stehen und sie in ihren Ergebnisdarstellungen, Interpretationen, Konklusionen und Handlungsempfehlungen beeinflussen.

Evaluierbarkeitsprüfung

Evaluation gilt heute a priori als erfolgreich, weil sich der Begriff im gesellschaftlichen Diskurs einen wichtigen Platz gesichert hat und in das allgemeine Selbstverständnis eingedrungen ist. Mit der Zunahme der Bedeutung verbunden ist jedoch die Problematik ihres Einsatzes. Überblickt man das Heer aktueller Evaluationen allein in der deutschen Schweiz, so lässt sich vermuten, dass Zielsetzungen, Fragestellungen oder auch politisches Klima nicht in allen Fällen die Durchführung einer Evaluation zu legitimieren in der Lage sind, so dass sich die Frage nach dem Zusammenhang zwischen Einsatz und Eignung aufdrängt. Wünschenswert scheint deshalb, anhand einer Evaluierbarkeitsprüfung die geeignete Methode zu wählen. Mit Evaluierbarkeitsprüfung ist die Frage gemeint, ob die bei der Problembenennung formulierten Fragen mit dem Instrument der Evaluation unter Einsatz anerkannter Forschungsmethoden beantwortet werden können. Ist eine Massnahme oder ein Projekt anhand eines solchen Verfahrens nicht evaluierbar (beispielsweise, weil noch keine brauchbaren Daten vorliegen, zu komplexe Systemprobleme vorhanden sind, eine ausgeprägte Entwicklungsorientierung oder sehr einfache Fragestellungen vorliegen), kann dies zur Entscheidung führen, die Evaluierung auf einen späteren Zeitpunkt zu verschieben, eine neue, realistischere Vorgehensweise festzulegen oder ein anderes Überprüfungs-, Kontroll- oder Entwicklungsinstrument zu wählen wie beispielsweise Expertengutachten, Hearings, Assessmentverfahren oder Shadowings.

4.3 Erschwerte Zugänge – erschwerte Nutzung: nicht-kontrollierbare Faktoren

Will Evaluation den brüchig gewordenen Anspruch, als Lieferantin gesicherten Wissens zu gelten, in der öffentlichen Diskussion verteidigen, so kommt sie nicht darum herum, neben der Konzentration auf die kontrollierbaren Faktoren gleichzeitig die nicht-kontrollierbaren Faktoren zu berücksichtigen und sie als mögliche Störvariablen anzuerkennen. Symptomatisch hierfür sind zwei Bereiche, die politische Kontextverflochtenheit und negative Ergebnisbilanzen. Bedingt durch den

Umstand, dass Evaluation politischer Akt und öffentliches Gut zugleich ist, an dem sowohl Bildungspolitik und Bildungsadministration, pädagogische Praxis und scientific community, aber auch zahlreiche Zaungäste teilhaben, die sich alle als selbsternannte Experten von Erziehungs- und Bildungsfragen verstehen, sind Evaluationen im Bildungsbereich „Musterbeispiele für Politikverflechtungen“ (KLÖTI 1995, S. 80). Eingebettet in einen Kontext voller heimlicher Nebenwirkungen produziert Evaluation immer Daten, die verschiedenen Zwecken dienen. Die Praxis hat somit aufgrund ihrer unterschiedlichen Bedingungen und Zwänge andere Vorstellungen als die Theorie, die ihrerseits unter der Präponderanz der scientific community steht. Trotzdem ist die Differenz zwischen der theoretischen Produktion von Evaluationswissen und ihrer Verwendung nicht schliessbar. Da Evaluation auch keine therapeutischen Funktionen hat, sind Konflikte oder Entscheidungen im Umgang mit den Ergebnissen demzufolge Angelegenheit der Steuerungsinstanzen, während Evaluationsforschung lediglich Daten für den bildungspolitischen Diskurs liefern kann. Evaluatoreninnen und Evaluatoren sind zwar aufgerufen, Defizite aufzuspüren und manifest zu machen, ohne jedoch die Verantwortung für ihre Bedeutung und für nachfolgende Prämissenkorrekturen übernehmen zu müssen. Für Entscheidungsträger hingegen, eingespannt in einen Bildungsapparat, der über die Sozialkontrolle funktioniert, stellen inopportune Befunde oder generell negative Erfolgsbilanzen, die nicht den Erwartungen entsprechen – eine spezifische Herausforderung dar, die Hilflosigkeit erzeugen kann und die zur Anwendung ‚neutralisierender‘ Strategien im Sinne von Transformationen, Vertuschungen oder Beschönigungen führt. Daten, welche Annahmen und Erwartungen nicht bestätigen, können mit solchen Verwandlungshandlungen trotzdem zur Systemsteuerung eingesetzt werden.

5. Bilanz

Die Nutzbarmachung von Evaluationswissen, die über die Schaffung wirksamer Wege und auch über die Schaffung wirksamer Nutzung erleichtert werden könnte, bleibt *cum grano salis* ein ungeklärtes, schwieriges und nicht gerade verlässliches Geschäft, in erster Linie dort, wo sie mit Hoffnungsszenarien der Veränderung verbunden wird und in der Verbesserung pädagogischer Selbstreflexions- und Entfaltungsfähigkeit der Akteure ein wesentliches Ziel sieht. Dem verbreiteten Glauben an die Kraft der Evaluation als moralische Instanz eines bildungspolitisch schlechten Gewissens, welche die Qualität der Bildungssysteme lenken könne, ist aufgrund der vorliegenden Erkenntnisse mit einem gewissen Mass an Skepsis zu begegnen, ohne jedoch ihre kulturelle Bedeutung als ‚grande idée‘ vorschnell aufs Spiel zu setzen und sich in die Reihe derer einzuordnen, die Zweifel anbringen, wenn Relevanzpostulate zu ihrer Verteidigung formuliert werden. Die problematischen Wirksamkeitsbefunde der empirischen Studie und die Erkenntnisse über die verschlungenen Wege des Praktischwerdens von Evaluationswissen lassen Evaluation allerdings in einem problematischeren Licht erscheinen als dies in Anbetracht der aktuellen bildungspolitischen Grosswetterlage wünschbar wäre.

Sollten Evaluationen angesichts der skizzierten Befunde nicht eher unterbleiben, wenn Entscheidungen grossmehrheitlich auf der selektiven Wahrnehmung von Evaluationsbefunden basieren und diese häufig nicht mehr als eine Stimulation des Wissens der Abnehmer bewirken können, jedoch damit rechnen müssen,

nicht gebraucht, uminterpretiert, relativiert und trivialisiert zu werden? Wäre es eine vordringliche Aufgabe der Evaluationsforschung, vor Evaluation zu warnen? Eine derartige Konsequenz wäre nun allerdings doch eine zu krasse Folgerung aus den bisherigen Ausführungen. Die empirischen Befunde dieses Aufsatzes legen vielmehr nahe, sich unterhalb solcher Verlegenheiten zu konsolidieren und sich unter Verzicht auf positive Meinungspflege und kritiklose Akzeptanz im Prokrustesbett zurechtzufinden. Viel eher gilt es, die Legitimation von Evaluation auf das Postulat der Verwendbarkeit auszurichten, d. h. auf die Hauptaufgabe, in geschickter Art und Weise Verwendung zu provozieren, aber mit der Einsicht, dass sie nie oder nur in seltenen Fällen den Vorstellungen der Evaluationsforschung entsprechen kann, also inadäquat bleiben muss. Das Fatale daran ist, dass diese Erkenntnis immer sowohl Belastung wie auch Entlastung sein kann und einer erheblichen Ideologieanfälligkeit unterliegt. Wenn Nutzung zwar provoziert werden soll und auch tatsächlich nachgewiesen werden kann, jedoch nur ausnahmsweise den Standards der Evaluationsforschung entspricht, dann hat Evaluation Verantwortung für diese provozierenden Aktivitäten zu tragen. Gleichzeitig wird sie jedoch auch frei für eine Reformierung von Massstäben der Nutzen-erwartung. Um dieses Dilemma lösen zu können, sind keine destruirenden Rundumschläge oder enthusiastischen Überzeugungsarbeiten gefragt, die sich etwa in Extrembekenntnissen artikulieren, eine evaluatorenbesetzte Gesellschaft solle die Welt demokratisieren und aus ihr einen besseren und weiseren, weniger von Konflikten besetzten Platz machen. Eher gefragt scheint ein Plädoyer für mehr Distanz zu Evaluation, für einen reflektierteren Einsatz, der Evaluation zwar zu einem knapperen Gut werden lässt und zwangsläufig zu einer Redimensionierung des Evaluationskults führt, ihr aber vielleicht die Chance eröffnet, gezieltere Wirksamkeit zu entfalten.

Literatur

- Alkin, M.C., Daillak, R. & White, B. (1979). *Using evaluations*. Beverly Hills, CA: Sage.
- Alkin, M.C. (1980). Naturalistic study of evaluation utilization. *New Directions for Program Evaluation (Utilization of Evaluative Information)*, 5, 19–28.
- Beck, U. & Bonss, W. (1989). Verwissenschaftlichung ohne Aufklärung? In U. Beck & W. Bonss (Hrsg.). *Weder Sozialtechnologie noch Aufklärung? Analysen zur Verwendung sozialwissenschaftlichen Wissens* (S. 7–45). Frankfurt a. M.: Suhrkamp.
- Beck, U. & Lau, Ch. (1982). Die „Verwendungstauglichkeit“ sozialwissenschaftlicher Theorien: Das Beispiel der Bildungs- und Arbeitsmarktforschung. In U. Beck (Hrsg.). *Soziologie und Praxis. Erfahrungen, Konflikte, Perspektiven. Sonderband 1 der Zeitschrift Soziale Welt* (S. 369–402). Göttingen.
- Brockmeyer, R. (1978). Wissenschaftliche Begleitung – Zusammenarbeit von Verwaltung und Wissenschaft. *Unterrichtswissenschaft*, 2, 113–125.
- Campbell, D. (1969). Reforms as experiments. *American Psychologist*, 24 (4), 409–429.
- Dewe, B. (1988). Wissensverwendung in Bildungsprozessen. Zur Transformation wissenschaftlicher Information in Praxisdeutungen. *Studien zum Umgang mit Wissen, Bd. 6*, Baden-Baden: Nomos.
- Dewe, B. & Ferchhoff, W. (1989). Über kraftlose Theorien und ihre „starke“ Verwendung. *Emile. Zeitschrift für Erziehungskultur*, 3, 13–28.
- Dickey, B. (1981). Utilization of evaluation of small-scale educational projects. *Educational Evaluation and Policy Analysis*, 2 (6), 65–77.

- Drerup, H. (1982). Anwendungsprobleme in der Evaluation. *Bildungsforschung/Bildungspraxis*, 4 (2), 154–170.
- Dubs, R. (2000). *Teilautonomie der Schulen: Annahmen, Begriffe, Probleme, Perspektiven*. Paderborn: Paderborner Universitätsreden.
- Fend, H. (1980). Fördert Evaluation Innovation? Vom Verhältnis von Innovation und Evaluation. In J.-P. Meylan (Hrsg.). *Innovation und Evaluation von Schulversuchen. Bericht über die erziehungswissenschaftliche Arbeitstagung* (S. 163–177). Montreux, 25.–29. Sept. 1978.
- Frey, K. (1975). *Handbuch der Curriculumforschung*, 3 Bände. München: Juventa.
- Geertz, C. (1999). *Dichte Beschreibung. Beiträge zum Verstehen kultureller Systeme*. Frankfurt a. M.: Suhrkamp.
- Greene, J.C. (1987). Stakeholder partizipation in evaluation design: is it worth the effort? *Evaluation and Program Planning*, 10, 379–394.
- Gstettner, P. & Seidl, P. (1977). Sozialpsychologische Aspekte wissenschaftlicher Begleitforschung. In W. Mitter & H. Weishaupt (Hrsg.). *Ansätze zur Analyse der wissenschaftlichen Begleitung bildungspolitischer Innovationen* (S. 84–119). Weinheim/Basel: Beltz.
- Haft, H. & Hameyer, U. (1975). *Curriculumplanung – Theorie und Praxis*. München: Kösel.
- Hameyer, U. et al. (1983). *Handbuch der Curriculum-Forschung*. Weinheim/Basel: Beltz.
- Kelle, U. & Kluge, S. (1999). *Vom Einzelfall zum Typus*. Opladen: Leske + Budrich.
- Klöti, U. (1995). Politische Evaluation im Bildungswesen. In U.P. Trier (Hrsg.). *Wirksamkeitsanalyse von Bildungssystemen. Symposium Bern 30.–31. März 1995* (S. 80–92). Aarau: Schweizerische Koordinationsstelle für Bildungsforschung.
- Kluge, S. (1999). *Empirisch begründete Typenbildung. Zur Konstruktion von Typen und Typologien in der qualitativen Sozialforschung*. Opladen: Leske + Budrich.
- Kluge, S. (2000). Empirisch begründete Typenbildung in der qualitativen Sozialforschung. *Forum Qualitative Sozialforschung* 1 (1) [On-line]. Available: <http://qualitative-research.net/fqs>.
- Kuckartz, U. (1999a). *Computergestützte Analyse qualitativer Daten. Eine Einführung in Methoden und Arbeitstechniken*. Opladen: Westdeutscher Verlag.
- Kuckartz, U. (1999b). *winMAX Professionelle Version. Handbuch zum Textanalysesystem winMAX für Windows 95/98/NT*. Westdeutscher Verlag: Opladen.
- Lichtenstein, J. (1983). Lehrplanentwicklung und Curriculumforschung. In J.-P. Meylan (Hrsg.). *Evaluation und Innovation im Bereich der Primarschule* (S. 39–58). Bern: Haupt.
- Mitter, W. & Weishaupt, H. (1977). *Ansätze zur Analyse der wissenschaftlichen Begleitung bildungspolitischer Innovationen*. Weinheim/Basel: Beltz.
- Mitter, W. & Weishaupt, H. (1979). *Strategien und Organisationsformen der Begleitforschung. 4 Fallstudien*. Weinheim/Basel: Beltz.
- Neul, W. (1977). *Strategien der Evaluation von Curricula und Curriculumentwicklung*. Stuttgart: Klett.
- Oelkers, J. (2001). *Ein kleiner Nachtrag zum Vortrag von Detlef Müller-Böling*. Vortrag auf der Tagung „Welche Qualität in den Sozialwissenschaften? Ansätze, Erfahrungen und Perspektiven“, Bern: Uptown Gurten, 17. Mai.
- Weiss, C.H. & Bucuvalas, M.J. (1980). *Social science research and decision-making*. New York: Columbia University Press.
- Wulf, C. (1975). Planung und Durchführung der Evaluation von Curricula und Unterricht. In K. Frey (Hrsg.). *Curriculum-Handbuch, Bd. II* (S. 567–579). München: Juventa.