

### Evaluation: Wirksame Wege zur Nutzung - Wege zur wirksamen Nutzung

Stamm, Margrit

Veröffentlichungsversion / Published Version

Sammelwerksbeitrag / collection article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:

W. Bertelsmann Verlag

#### Empfohlene Zitierung / Suggested Citation:

Stamm, M. (2008). Evaluation: Wirksame Wege zur Nutzung - Wege zur wirksamen Nutzung. In M. Ant, A. Hammer, & O. Löwenbein (Hrsg.), *Nachhaltiger Mehrwert von Evaluation* (S. 145-158). Bielefeld: W. Bertelsmann Verlag. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-52366-4>

#### Nutzungsbedingungen:

Dieser Text wird unter einer CC BY-SA Lizenz (Namensnennung-Weitergabe unter gleichen Bedingungen) zur Verfügung gestellt. Nähere Auskünfte zu den CC-Lizenzen finden Sie hier: <https://creativecommons.org/licenses/by-sa/3.0/deed.de>

#### Terms of use:

This document is made available under a CC BY-SA Licence (Attribution-ShareAlike). For more information see: <https://creativecommons.org/licenses/by-sa/3.0>

# 10. Evaluation: Wirksame Wege zur Nutzung – Wege zur wirksamen Nutzung

MARGRIT STAMM – UNIVERSITÄT FRIBOURG – SCHWEIZ

## 10.1 Einleitung: das Problem



Das Referat befasst sich mit der Frage, ob und inwiefern die Gesellschaft von Evaluation lernen kann und lernen will. Auf den ersten Blick erscheint eine solche Frage allerdings als überflüssig, da sich Evaluation ja geradezu als Wissenschaft für die Praxis versteht und deshalb die Bereitstellung nützlichen Wissens für Abnehmer- und Anwendersysteme als selbstverständliche Leistung erachtet. Dieses Selbstverständnis findet denn auch in der gesellschaftlichen Akzeptanz seinen Niederschlag: Evaluation ist zu einer Standardaufgabe nationaler, regionaler und privater Bildungssysteme geworden. Dabei überwiegt der allgemeine Konsens, Evaluation habe Anstöße zur Entwicklung, Optimierung und Veränderung pädagogischer Praxis zu übernehmen. Ein solcher Konsens basiert auf der unhinterfragten Annahme, ein forcierter Einsatz von Evaluation führe automatisch zu einer Qualitätssteigerung von Bildungsleistungen und Bildungsangeboten. Entsprechend hat sich Evaluation bis heute kaum darüber ausweisen müssen, ob sie ihre intendierten Wirkungen tatsächlich erzielt und ob das, was sie produziert, auch für die Adressaten gut ist und von ihnen genutzt wird.

## 10.2 Die Rolle der Erziehungswissenschaft in der Evaluationsforschung

Die Frage nach der Nutzung von Evaluationswissen trifft die Erziehungswissenschaft relativ unvorbereitet. Und dies, obwohl die Vertrautheit mit Evaluation dazu

geführt hat, sie als selbstverständlichen Teil der Pädagogischen Psychologie zu bezeichnen (Krapp/Weidenmann, 2001; Rost, 2001) und auf ihre unverzichtbare Rolle in der Steuerung oder Entwicklung unserer Bildungssysteme zu verweisen (Altrichter/Posch, 1997; Burkard/Eikenbusch, 2000; Helmke, 2000; Lange, 1999; Tillmann/Vollstädt, 2001). Allerdings mangelt es auch nicht an Abhandlungen, welche die praktische Evaluationsrelevanz herausstreichen und davor warnen, Evaluation zum Selbstzweck werden zu lassen (Beywl, 1988/1999; Guba/Lincoln, 1981; Hellstern/Wollmann, 1983; Kordes, 1983; Patton, 1988; 1994; Wottawa/Thierau, 1998). Merkwürdig bleibt jedoch, dass die Folgen von Evaluation nicht häufiger thematisiert werden, obwohl gerade der erziehungswissenschaftliche Bereich Evaluation mit besonderen Erwartungen an die Umsetzung verbindet. Geht man beispielsweise von der Satzung aus, dass Schulleistungsvergleichsstudien eine neue und vorrangige Aufgabe nationaler Bildungssysteme werden, Evaluation jedoch ein vorwiegend pädagogisches Geschäft sein soll, so müsste die Erziehungswissenschaft in diesem Bereich künftig gehaltvoll mitreden. Dann wird sie allerdings nicht darum herumkommen, sich mit ihrem alten, in neuem Zusammenhang gestellten und ungelösten Problem zu beschäftigen, nämlich ob und wie theoretisches Wissen in die pädagogische Praxis überführt werden kann, respektive unter welchen Bedingungen Gesellschaft und Bildungspolitik bereit sind, von Evaluation zu lernen. In der Konsequenz bedeutet dies, dass die Erziehungswissenschaft mit Vorteil die breiten Forschungserfahrungen aus den achtziger Jahren zur Verwendung sozialwissenschaftlichen Wissens in außerwissenschaftlichen Kontexten berücksichtigt (Beck/Lau, 1983; Beck/Bonß, 1989a; 1989 b; Drerup, 1979; 1982, 1989; Drerup/Terhart, 1990; Evers/Nowotny, 1989; Weber, 1994).

Die Wissensverwendungsforschung zeigt, dass von der Erkenntnis zur Verwendung wissenschaftlichen Wissens in pädagogische Praxiszusammenhänge eine große, ernüchternde Kluft besteht und dass es keinen gradlinigen Transfer von der Wissenschaft in die (pädagogische) Praxis gibt. Es empfehlen sich jedoch auch Anleihen bei der angloamerikanischen *Research on evaluation utilization* (Alkin/Coyle, 1988; Cousins/Leithwood, 1986, 1993; Daillak, 1982; Heller, 1986; Patton, 1986; 1997; Patton, 2000; Siegel/Tuckel, 1985; Weiss, 1972c; 1982). Ein differenzierter Blick in diese Forschung zeigt jedoch klar, dass die Theorie der Evaluationsnutzung bei Weitem die Empirie übersteigt und deshalb keine genügend große empirische Basis für klare Urteile über die Evaluationsnutzung vorliegt. Letztlich konkretisiert sich das Problem der Evaluationsnutzung damit in seinen unzulänglichen Erklärungsversuchen, sodass nichts weiter bleibt, als bei der konkreten Evaluationspraxis anzusetzen. Dies geschieht anhand der Schulleistungsvergleichsevaluation PISA und der Feldstudie über Schweizer Bildungsevaluationen.

## 10.3 Feldstudie und empirische Typenbildung

Die Feldstudie umfasst alle in der deutschen Schweiz verfügbaren, von kantonalen Bildungsdepartementen an externe Evaluatorinnen und Evaluatoren in Auftrag

Item	C 1* N = 15	C 2* N = 8	C 3* N = 10	C 4* N = 13
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	94	44
Widerstände	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 durch Bildungsdepartement	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	92	98
Termintreue	67	33	95	33
methodische Glaubwürdigkeit	67	50	98	78
finanzielle Ressourcen	11	17	57	57
Commitment Evaluator/in	11	44	71	56
Reputation Evaluator/in	22	100	71	98
Nützlichkeit/Anwendbarkeit	33	67	50	50
Zufriedenheit mit Nutzung	11	33	93	50
Erfahrungen mit Evaluation	67	67	17	67
Engagement Entscheidungsträger	22	50	95	86
<p>Lesehinweis: Alle Angaben in Prozent. Beispielsweise weisen 44 % der Evaluationen des ersten Typus eine hohe sprachliche Qualität auf, in Cluster 4 sind es hingegen 98%. In Cluster 2 ist nur in 33 % der Fälle Widerstand aufgetreten und in 11 % besteht Zufriedenheit mit der tatsächlichen Nutzung. Ganz anders hingegen Cluster 4: Hier besteht in 78 % der Fälle Widerstand und die Zufriedenheit mit der Nutzung beträgt 50 %.</p> <p>* C 1entspricht dem nachfolgend beschriebenen Typ der Blockade, C 2 dem Typ des Alibis, C 3 dem Typ der Innovation und C4 dem Typ der Reaktion.</p>				

Tabelle 10-1: Clusterspezifische Merkmalsverteilung

gegebenen und zwischen 1995 und 1998 abgeschlossenen Evaluationen. Auf der methodologischen Basis der vergleichenden Fallanalyse (Kelle/Kluge, 1999; Kluge, 2000) mit empirischer Typologie wurden insgesamt 56 Datensätze erhoben und mit Hilfe des Textdatenbanksystems winMAX mit integrierter Clusteranalyse verarbeitet (Kuckartz, 1999). In einem ersten Schritt erfolgte die Erarbeitung relevanter Vergleichsdimensionen (Zweck/Ziel und Rezeptions- /Verwendungstyp) mit anschließender Gruppierung der Fälle anhand dieser Merkmale. Parallel dazu wurde eine Clusteranalyse mit insgesamt 40 Items gerechnet, die aus der Vercodung der Interviewtexte in das empirische Stufenmodell eingeflossen waren (Tabelle 10-1).

Sie bildete gewissermaßen die systematische Auswertung dieser Ähnlichkeitsmatrix. Dabei wurde die Methode des Complete-Linkage-Verfahrens, ein hierarchisches Verfahren nach agglomerativer Vorgehensweise, gewählt (Backhaus, 1996, S. 348). Da mit dieser Clusteranalyse ein Vergleich mit den über das empirische Stufenmodell gewonnenen Ergebnissen stattfinden sollte, ging es darum, eine Clusterlösung zu finden, welche die größte Übereinstimmung aufweist.

Letztlich waren es deshalb theoretische Erwägungen, der angestrebte Grad an Differenziertheit und die Größe der einzelnen Cluster, welche den Entscheid für eine Lösung mit vier Clustern den Vorzug gaben. Die in Tabelle 10-1 dargestellten Clusterprofile enthalten für jedes Cluster den prozentualen Anteil, welcher das Cluster an der jeweiligen Merkmalsausprägung aufweist. Diese Clusterprofile sind Bestandteil für die in Kapitel 10.5 erfolgende Interpretation.

Die Cluster ließen sich schließlich in eine Kreuztabelle einfügen, welche die Zuordnung der Evaluationen zu den einzelnen Feldern sichtbar macht (Tabelle 10-2). Anhand der Merkmalskombinationen und Sinnzusammenhänge führte die Charakterisierung der Typen zu den Bezeichnungen ‚Blockade‘, ‚Innovation‘, ‚Alibi‘ und ‚Reaktion‘. Tabelle 10-2 macht klar, dass Evaluation eindeutig vom Vorwurf der Folgenlosigkeit freigesprochen werden kann, denn mit der ‚Reaktion‘ (n = 10) und der ‚Innovation‘ (n = 13) liegen zwei Nutzungstypen vor, die nachfolgende bildungspolitische Entscheide wesentlich beeinflusst haben.

Nutzung von Evaluationswissen zur Erhellung bildungspolitischer Entscheide ist somit zumindest in etwa der Hälfte der Fälle pädagogische Realität. Gegenteiliges gilt allerdings für die Blockade (n = 15) und das Alibi (n = 8).

Rezeptions- und Verwendungstyp	Zweck / Ziel		Entwicklung /Optimierung
	Kontrolle	Legitimation	
direkte Rezeption/ Nutzung nachweisbar	C 4 REAKTION (n = 10)		C 3 INNOVATION (n = 13)
direkte Rezeption/ Nutzung unklar	C 1 BLOCKADE (n = 15)	C 2 ALIBI (n = 8)	
direkte Rezeption/ Nutzung nicht nachweisbar			

Tabelle 10-2: Vier Typen der Beziehung zwischen Rezeption und Ziel/Zweck

## 10.4 Die vier Typen von Evaluationsnutzung

Das kombinierte Modell empirischer Typenbildung hat im Ergebnis zur Formulierung der 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.

### **Die Typik der Blockade**

Evaluationen, welche zu gleichen Teilen der Überprüfung, der Kontrolle und der Umsetzung dienen sollen (je 33 %), jedoch kaum direkte Nutzung von Evaluationswissen aufzeigen (33 %), 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. Entsprechend gering ist die Zufriedenheit mit der Nutzung der Ergebnisse (11 %). Gleiches gilt für die Einschätzung der Verwendbarkeit der Ergebnisse (33 % Positivwerte). Sowohl das Engagement der Evaluatorenschaft wie auch der Entscheidungskader sind mit 11 % resp. 22 % relativ gering.

## ***Die Typik der Innovation***

In diesem Typus finden sich Evaluationen, welche eindeutig entwicklungs- und umsetzungsorientierten Charakter haben (60 %). Sie lassen sich durch geplante, umfassende und anhaltende Umsetzungsmaßnahmen kennzeichnen, welche den Charakter von Innovationen annehmen. Idealtypische Konturen ergeben sich aus der hohen Qualität der Evaluation (methodische Glaubwürdigkeit: 78 %, sprachliche Qualität 100 %), den organisationalen Merkmalen, welche Autonomie und Selbstverantwortung fördern (Bottom-up-Struktur: 78 %) und letztlich Dissemination (93 %) und tatsächliche Nutzung als Folgen eines fruchtbaren Prozesses erscheinen lassen. Entsprechend hoch ist die Zufriedenheit mit den Evaluationsergebnissen (79 %) und deren Nutzung (93 %). Die Gesamtbilanz ist vollumfänglich positiv (100 %).

## ***Die Typik des Alibis***

Dieser Typus setzt sich aus Evaluationen zusammen, die in erster Linie legitimierenden Charakter (67 %) haben und nur zu 33 % direkte Evaluationsfolgen nachweisen können. Die Evaluationen werden deshalb als ‚Alibi‘ bezeichnet, weil sie kein pädagogisches Erkenntnisinteresse, sondern lediglich der legitimatorische Wert, produziert worden zu sein, miteinander verbindet. Und dies, obwohl sich dieser Evaluationstypus durch den höchsten Anteil dienstleistungsorientierter Evaluatoredinnen und Evaluatoren auszeichnet (67 %). Alibi-Evaluationen, in allen Fällen (100 %) in top-down organisierten Bildungsorganisationen durchgeführt, sind Antworten mit reinem Bestätigungs- und Rechtfertigungscharakter, auf von bestimmten Interessengruppen ausgehende Forderungen nach Evaluation. Die Gesamtbilanz fällt entsprechend nur zu 17 % positiv aus.

## ***Die Typik der Reaktion***

Dieser Typus beinhaltet Evaluationen, die zwar auf die Überprüfung (44 %) und Kontrolle (55 %) ausgerichtet sind, jedoch eine klare Rezeption und Nutzung der Ergebnisse vorweisen können (100 %). Am augenfälligsten ist dabei, dass dies als Reaktion geschieht, bestimmte Beschlüsse oder Maßnahmen infolgedessen direkte Antworten auf Forderungen bestimmter Anspruchsgruppen darstellen. Die Evaluation, durchgeführt von reputierten Evaluatoredinnen und Evaluatoren (100 %), hat relativ hohe bildungspolitische Bedeutung (78 %), muss aber auch mit ausgeprägtem Widerstand einzelner Akteurguppen kämpfen (78 %).

## 10.5 Pädagogische Konsequenzen als Politikberatung

Im Ganzen gesehen legt diese Nutzungstypologie nahe, dass direkte 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 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 verschlungene Wege gehen, sind beträchtlich. Solche Ergebnisse provozieren den Versuch, die Anwendungsfähigkeit von Evaluation auf eine neue Basis zu stellen. Das soll im Folgenden als Politikberatung in zwei Perspektiven geschehen, zum einen als ‚wirksame Wege zur Nutzung‘ und zum anderen als ‚Wege zur wirksamen Nutzung‘. Dabei wird davon ausgegangen, dass Evaluatorinnen und Evaluatoren wirksame Wege vorzeichnen können, welche einer Nutzung durch die Abnehmer-systeme förderlich sind. Die zweite Perspektive betrifft die Entscheidungsträger und Projektleitungen selbst, die bestimmte Strategien verfolgen und Wege beschreiten können, die eine wirksame Nutzung nach sich ziehen. 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.

### 10.5.1 Wirksame Wege zur Nutzung

#### *Optimierung der Methodologie*

Die Forderung nach methodischer Qualität legitimiert sich nicht nur durch auf die Erhöhung der technischen Glaubwürdigkeit und der generellen Akzeptanz ausgerichtete 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 Sophistikation. 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 im Sinne der Auftraggeber entsprechende ‚Richtige‘ und ‚Wahre‘ ausstrahlen, was die Evaluationsrolle als „Wahrheitsgarantin“ (Oelkers, 2001a, 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 einerseits 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 die in der Evaluation selbst verankerten Rezeptions- und Verwendungsschranken, welche 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 Praxisausrichtung („Nützlichkeitsstest“). 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 Evaluatorinnen und Evaluatoren in der Lage sind, den Entscheidungsträgern ihren selektiven, marginalen oder kursorischen Wissensgehalt bewusst zu machen (vgl. die Typik der Innovation).

## 10.5.2 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 Entscheidungsdichte 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 Umsetzungerwartungen und -forderungen der bildungspolitischen Behörde parallel zur Anzahl ihrer Ergebnisrückmeldungen.

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, sodass 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.

## **10.6 Bilanz**

Die Nutzbarmachung von Evaluationswissen, die über die Schaffung wirksamer Wege und auch über die Schaffung wirksamer Nutzung erleichtert werden könnte,

bleibt ein ungeklärtes und schwieriges Geschäft, in erster Linie dort, wo sie mit Hoffnungsszenarien der Veränderung verbunden wird. Dem verbreiteten Glauben an die Kraft der Evaluation als moralische Instanz eines bildungspolitisch schlechten Gewissens ist aufgrund der vorliegenden Erkenntnisse mit einem gewissen Maß an Skepsis zu begegnen. Die problematischen Wirksamkeitsbefunde der empirischen Studie und die Erkenntnisse über die verschlungenen Wege des Praktischwerdens von Evaluationswissen lassen Evaluation 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, die Legitimation von Evaluation auf das Postulat der Verwendbarkeit auszurichten, d.h. auf die Hauptaufgabe, in geschickter Art und Weise Verwendung zu provozieren. 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.

## 10.7 Literatur

---

**Alkin, M.C./Coyle, K.** (1988). Thoughts on evaluation utilization, misutilization and non-utilization. *Studies in Educational Evaluation*, 14, 331-340.

**Altrichter, H., Posch, P.** (1997). Mikropolitik der Schulentwicklung. Förderliche und hemmende Bedingungen für Innovationen in der Schule. Innsbruck: StudienVerlag.

**Backhaus, K.** (1996). *Multivariate Analysemethoden – eine anwendungsorientierte Einführung*. Berlin: Springer.

**Beck, U./Bonß, W.** (1989a). 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./Bonß, W.** (1989b). Weder Sozialtechnologie noch Aufklärung? Analysen zur Verwendung sozialwissenschaftlichen Wissens. Frankfurt a. M.: Suhrkamp.
- Beck, U./Lau, Ch.** (1983). 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.
- Benner, D.** (2002). Die Struktur der Allgemeinbildung im Kerncurriculum moderner Bildungssysteme. Ein Vorschlag zur bildungspolitischen Rahmung von PISA. Zeitschrift für Pädagogik, 1, 68-90.
- Beywl, W.** (1988/1999). Zur Weiterentwicklung der Evaluationsmethodologie. Grundlegung, Konzeption und Anwendung eines Modells der responsiven Evaluation. Bern: Lang (Reprint 1999).
- Burkard, C., Eikenbusch, G.** (2000). Praxishandbuch Evaluation in der Schule. Berlin: Cornelsen.
- Campbell, D.** (1969). Reforms as experiments. American Psychologist, 24, 409-429.
- Cousins, J.B./Leithwood, K.A.** (1986). Current empirical research on evaluation utilization. Review of Educational Research, 56, 331-364.
- Cousins, J.B./Leithwood, K.A.** (1993). Enhancing knowledge utilization as a strategy for school improvement. Knowledge: Creation, Diffusion, Utilization; 14, 305-33.
- Dailak, R.H.** (1982). What is evaluation utilization? Studies in Educational Evaluation, 8, 157-162.
- Drerup, H.** (1979). Wissenschaftstheorie und Wissenschaftspraxis. Probleme der Vermittlung zwischen metawissenschaftlichen Forschungsprogrammen und einer Praxis der Spezial- / Erziehungswissenschaft. Grundmann: Bonn.
- Drerup, H.** (1982). Anwendungsprobleme in der Evaluation. Bildungsforschung / Bildungspraxis, 4, 154-170.
- Drerup, H.** (1989). Probleme außerwissenschaftlicher Verwendbarkeit von Erziehungswissenschaft. Zum Einfluss von Erziehungswissenschaft im politisch-administrativen Bereich. In: E. König/P. Zedler (Hrsg.). Rezeption und Verwendung erziehungswissenschaftlichen Wissens in pädagogischen Handlungs- und Entscheidungsfeldern (S. 143-166). Weinheim: Deutscher Studienverlag.
- Drerup, H., Terhart, E.** (1990). Erkenntnis und Gestaltung. Vom Nutzen erziehungswissenschaftlicher Forschung in praktischen Verwendungskontexten. Weinheim: Deutscher Studienverlag.
- Dubs, R.** (2000). Teilautonomie der Schulen: Annahmen, Begriffe, Probleme, Perspektiven. Paderborn: Paderborner Universitätsreden.

- Evers, A./Nowotny, H.** (1989). Über den Umgang mit Unsicherheit. Anmerkungen zur Verwendung sozialwissenschaftlichen Wissens. In: U. Beck/W. Bonß (Hrsg.). *Weder Sozialtechnologie noch Aufklärung? Analysen zur Verwendung sozialwissenschaftlichen Wissens* (S. 355–383). Frankfurt a. M.: Suhrkamp.
- Guba, E.G./Lincoln, Y.S.** (1981). *Effective evaluation: Improving the usefulness of evaluation results through responsive and naturalistic approach*. San Francisco: Jossey-Bass.
- Hagemeister, V.** (1999). Was wurde bei TIMMS erhoben? Über die empirische Basis einer aufregenden Studie. *Die Deutsche Schule*, 2, 160–177.
- Hagemeister, V.** (2000a). Irrwege und Wege zur „Testkultur“. *Die Deutsche Schule*, 1, 322–328.
- Hagemeister, V.** (2000b). Die TIMSS-Leistungen bleiben zweifelhaft. *Die Deutsche Schule*, 3, 480–490.
- Heller, F.** (1986). *The use and abuse of social science*. London: Sage.
- Hellstern, G.-M./Wollmann, H.** (1983). *Evaluierungsforschung. Ansätze und Methoden – dargestellt am Beispiel des Städtebaus*. Stuttgart: Birkhäuser.
- Helmke, A.** (2000). Von der externen Leistungsevaluation zur Verbesserung des Lehrens und Lernens. In: U. P. Trier (Hrsg.). *Bildungswirksamkeit zwischen Forschung und Politik* (S. 135–164). Chur: Rüegger.
- Kelle, U./Kluge, S.** (1999). *Vom Einzelfall zum Typus*. Opladen: Leske + Budrich.
- Kluge, S.** (2000). Empirisch begründete Typenbildung in der qualitativen Sozialforschung. *Forum Qualitative Sozialforschung 1 (1)* [On-line]. Available: <http://www.qualitative-research.net/fqs> (22. 09.2006).
- Kordes, H.** (1983). Evaluation in Curriculumprozessen. In: U. Hameyer, K. Frey/H. Kraft (Hrsg.). *Handbuch der Curriculumforschung*. Weinheim/Basel: Beltz.
- Krapp, A./Weidenmann, B.** (Hrsg.). (2001). *Pädagogische Psychologie. Ein Lehrbuch*. Weinheim: Beltz, Psychologie Verlags Union.
- Kuckartz, U.** (1999). *Computergestützte Analyse qualitativer Daten. Eine Einführung in Methoden und Arbeitstechniken*. Opladen: Westdeutscher Verlag.
- Lange, H.** (1999). Qualitätssicherung in Schulen. *Die Deutsche Schule*, 1, 144–159.
- Oelkers, J.** (2001a). Bildung, Bildungsforschung und Bildungspolitik. Vortrag am Züricher Festival des Wissens, 10. Mai.
- Patton, M.Q.** (1986, 1997). *Utilization-focused evaluation*. Thousand Oaks, London, New Delhi: Sage.

- Patton, M.Q.** (1988). The evaluator's responsibility for utilization. *Evaluation Practice*, 9, 5–24.
- Patton, M.Q.** (1994). Developmental evaluation. *Evaluation Practice*, 3, 311–319.
- Patton, M.Q.** (2000). A vision of evaluation that strengthens democracy. Referat an der 4. Jahreskonferenz der Europäischen Evaluationsgesellschaft in Lausanne, 14. Oktober.
- Rost, D.H.** (2001). Handwörterbuch Pädagogische Psychologie. Weinheim: Beltz, Psychologie Verlags Union.
- Siegel, K./Tuckel, P.** (1985). The utilization of evaluation research. *Evaluation Review*, 9, 307–328.
- Terhart, E.** (2002). Wie können die Ergebnisse von vergleichenden Leistungsstudien systematisch zur Qualitätsverbesserung in Schulen genutzt werden? *Zeitschrift für Pädagogik*, 1, 91–110.
- Tillmann, K.-J./Vollstädt, W.** (2001). (Hrsg.). Politikberatung durch Bildungsforschung. Das Beispiel: Schulentwicklung Hamburg. Opladen: Leske + Budrich.
- Von der Groeben, A.** (2000). Wo liegen die Wurzeln von Schulqualität? Eine Antwort auf Hermann Lange. *Die Deutsche Schule*, 2, 339–354.
- Von der Groeben, A.** (2002). Nicht in Maßnahmen stecken bleiben. Ein Plädoyer für eine radikale Frage nach PISA. *Pädagogik*, 4, 38–42.
- Weber, U.** (1994). Regulation und Wissen: Implikationen neuerer Ergebnisse der Verwendungsforschung für eine Theorie der Regulation. Frankfurt a. M./Bern: Lang.
- Weinert, F.E.** (2002). Perspektiven der Schulleistung – mehrperspektivisch betrachtet. In: F.E. Weinert (Hrsg.). Leistungsmessung in Schulen (S. 353–365). Weinheim/Basel: Beltz.
- Weiss, C.H.** (1972a). *Evaluation research: Methods of assessing program effectiveness*. Englewood Cliffs, NJ: Prentice-Hall.
- Weiss, C.H.** (1972c). Utilization of evaluation. Toward comparative study. In: C.H. Weiss (Ed.). *Evaluation action programs: readings in social action and education* (pp. 318–326). Boston: Allyn and Bacon.
- Weiss, C.H.** (1982). Measuring the use of evaluation. *Evaluation Studies Review Annual*, 7, 129–145.
- Wottawa, H./Thierau, H.** (1998). *Lehrbuch Evaluation*. Bern: Huber.