

Evaluation – eine Begriffsdefinition

von Reinhard Stockmann

Evaluation steht nicht nur für spezifisches Handeln, das die Bewertung von empirisch gewonnenen Informationen zum Ziel hat, auf deren Basis rationale Entscheidungen getroffen werden können, sondern auch für das Ergebnis dieses Prozesses. Wissenschaftlich durchgeführte Evaluationen zeichnen sich dadurch aus, dass sie (a) auf einen klar definierten Gegenstand bezogen sind; (b) dass für die Informationsgewinnung objektivierende empirische Datenerhebungsmethoden eingesetzt werden und dass (c) die Bewertung anhand präzise festgelegter und offengelegter Kriterien, (d) mit Hilfe systematisch vergleichender Verfahren vorgenommen wird. Die Evaluation wird (e) in der Regel von dafür besonders befähigten Personen (Evaluatoren) durchgeführt, um (f) auf den Evaluationsgegenstand bezogene Entscheidungen zu treffen.

Damit das Nutzungspotenzial von Evaluationen möglichst optimal ausgeschöpft wird, hat sich jede professionell durchgeführte Evaluation mit folgenden fünf Fragen auseinanderzusetzen: (1) *Was* (welcher Gegenstand), wird (2) *wozu* (zu welchem Zweck), (3) *anhand welcher Kriterien*, (4) *von wem*, (5) *wie* (mit welchen Methoden) evaluiert?

(1) Im Prinzip gibt es bei der Wahl des Evaluationsgegenstands kaum Einschränkungen. Objekte der Bewertung können Gesetze, Produkte, Dienstleistungen, Organisationen, Personen, Prozesse sowie soziale Tatbestände jedweder Art oder gar Evaluationen selbst sein. Häufige Evaluationsgegenstände sind allerdings Reformmaßnahmen, Projekte, Programme oder Policies. Wenn dies der Fall ist, dann interessiert in einer Evaluation zumeist, ob die damit verbundenen Ziele erreichbar sind (performativ / formativ) oder erreicht wurden (summativ). Dabei werden die Soll-Vorgaben mit dem zum Evaluationszeitpunkt gemessenen Ist-Zustand verglichen. Je mehr der Ist-Wert den Soll-Vorgaben entspricht (oder diese sogar übertrifft), umso besser wird das Ergebnis bewertet. Mit *zielorientierten Evaluationen* sind jedoch eine Reihe von Problemen verbunden, die ihre Durchführung enorm erschweren können. Dies ist insbesondere dann der Fall, wenn Programmziele nicht klar formuliert

wurden oder sogar ganz fehlen, wenn neben den offiziell deklarierten Zielen andere, konkurrierende (informelle) Zielvorgaben existieren, wenn nicht alle Akteure (z.B. einer programmdurchführenden Organisation) die gleichen Ziele verfolgen und wenn sich die Ziele über die Zeit hinweg stark verändert haben.

Unabhängig von diesen Problemen besteht bei zielorientierten Evaluationen zudem die Gefahr, dass nicht-intendierte Effekte systematisch ausgeblendet werden (Tunnelblick). Doch gerade diese können sich als interessant und wichtig erweisen und deshalb für die Bewertung des Erfolgs, der Wirksamkeit oder Nachhaltigkeit eines Programms von entscheidender Bedeutung sein.

Als Ausweg aus dieser Problematik bietet sich als *Alternative* eine *wirkungsorientierte Evaluation* an. Diese geht nicht primär von den Zielen eines Programms aus, sondern versucht, hypothesengeleitet potentiellen Wirkungen auf die Spur zu kommen. Hierfür eignet sich z.B. der vom CEval entwickelte Evaluationsansatz, mit dem sich die Suche nach intendierten und nicht-intendierten Wirkungen steuern und strukturieren lässt.

Obwohl der Wirkungsbegriff in der Evaluation eine Schlüsselgröße darstellt, ist nicht immer klar, was er bedeutet. Häufig wird er mit dem Begriff ‚Output‘ verwechselt. *Outputs* sind die Produkte oder Leistungen, die von einer Organisation erbracht werden, wie z.B. die Anzahl verteilter Essenspakete, die Anzahl behandelter Kranker, die Anzahl der durchgeführten Beratungen etc. *Wirkungen* stellen die Veränderungen dar, die Folgen dieser Leistungen sind, also z.B. gesättigte Menschen, verbesserter Gesundheitszustand der Behandelten oder Personen, die nach der Beratung wissen, was zu tun ist.

Werden die Wirkungen eines Programms bilanziert, dann dürfen die *unbeabsichtigten Folgen* nicht vernachlässigt werden, denn die Qualität eines Programms kann nicht isoliert betrachtet werden, sondern nur in seiner gesamten Komplexität. Hierzu gehören auch nicht erwartete oder unerwünschte Wirkungen. Wirkungen können danach unterschieden werden, ob sie *intendiert (geplant)* sind und mit den Zielen eines Programms der Leistungsangebots in Einklang stehen oder ob es sich um *nicht-intendierte (ungeplante)* Wirkungen handelt. Dabei werden intendierte Wirkungen in der Regel im Hinblick auf die Zielerreichung positiv zu bewerten sein, während nicht-intendierte Wirkungen sowohl positiv –

also die Zielerreichung unterstützend – als auch negativ, d.h. der Zielerreichung zuwiderlaufend, ausfallen können. Intendierte negative Wirkungen sind durchaus möglich, z.B. wenn bestimmte Nachteile, die mit einem Programm verbunden sind, bewusst in Kauf genommen werden. Ob eine Wirkung als intendiert oder nicht-intendiert, positiv oder negativ bewertet wird, hängt selbstverständlich von den Zielen eines Programms ab, und nicht zuletzt von der Perspektive des Betrachters.

Da das Ziel von *Wirkungsevaluationen* darin besteht, mit größtmöglicher Zuverlässigkeit festzustellen, ob eine Intervention die intendierten Wirkungen auslöst, sind die Einflüsse anderer Faktoren, die ebenfalls für die gemessenen Veränderungen verantwortlich sein könnten, auszuschließen. D.h. in dem Geflecht von beobachteten Wirkungen sind differenzierte Ursachenzuschreibungen vorzunehmen. Diese Aufgabe stellt eine der größten Herausforderungen einer Evaluation dar. Dies liegt vor allem daran, dass die soziale Welt einen hohen Komplexitätsgrad aufweist, d.h. die meisten sozialen Phänomene auf vielen Ursachen basieren. Interventionen haben zudem in der Regel nur einen geringen Eingriffsspielraum und ein niedriges Veränderungspotenzial. Oft sind die Programm- oder Leistungswirkungen nur schwach ausgeprägt und es besteht selbst bei professionellem Einsatz von Auswertungsverfahren die Gefahr, dass sie im allgemeinen ‚Rauschen‘ gar nicht erkannt werden.

- (2) Evaluationen können folgenden vier unterschiedlichen, aber miteinander verknüpften Funktionen dienen: (a) Der Gewinnung von Erkenntnissen, um die Prozessabläufe oder die Wirkungszusammenhänge in einem Programm zu verstehen. (b) Um Kontrolle auszuüben, indem festgestellt wird, ob die in der Planung festgelegten Ziele erreicht wurden. (c) Um Lernpotenziale zu eröffnen, die für die Weiterentwicklung von Programmen genutzt werden sollen. (d) Um Programme zu legitimieren, in dem öffentlich belegt wird, wie nützlich, wirksam oder nachhaltig sie waren.

Da Evaluationen mittlerweile als Ausdruck moderner, „evidence based policy“ gelten, werden Evaluationen mitunter auch missbraucht, indem sie dazu verwendet werden politisch bereits getroffene Entscheidungen nachträglich mit Hilfe von Evaluationsergebnissen zu legitimieren. Diese „taktische“ Funktion von Evaluation lässt sich jedoch nicht mit ihrem eigentlichen Zweck begründen, sondern stellt eher ihre pathologische Seite dar.

Die Festlegung auf eine prioritäre Funktion steuert die Herangehensweise und bestimmt das Design und die Durchführung von Evaluationen. Diese können nicht nur verschiedene Funktionen erfüllen, sondern im Rahmen der einzelnen Phasen der Programmentwicklung auch unterschiedliche Analyseperspektiven und Erkenntnisinteressen verfolgen. Evaluationen können dazu genutzt werden, (1) die Planung eines Programms oder einer Maßnahme zu verbessern (ex-ante Evaluation), (2) die Durchführungsprozesse zu beobachten (on-going Evaluation) oder (3) die Wirksamkeit und Nachhaltigkeit von Interventionen ex-post zu bestimmen (ex-post Evaluation).

Evaluationen können demnach mehr *formativ*, d.h. aktiv-gestaltend, prozessorientiert, konstruktiv und kommunikationsfördernd angelegt sein, oder mehr *summativ*, d.h. zusammenfassend, bilanzierend und ergebnisorientiert.

Die *Begleitforschung* kann als eine besondere Form der Evaluation gelten, die sich ex-ante mit den Voraussetzungen bzw. der Planung oder (on-going) mit der Implementation von Programmen beschäftigt. Anders als die meisten Evaluationen wird sie nicht nur punktuell, zu einem bestimmten Zeitpunkt im Programmverlauf eingesetzt, sondern kontinuierlich – programmbegleitend. Dadurch können Längsschnittdaten gewonnen werden, die nicht nur kontinuierlich über Veränderungsprozesse informieren, sondern auch Ursache-Wirkungszuschreibungen erleichtern.

(3) Im Unterschied zu *Normenreihen*, wie sie von ISO aufgestellt werden oder den im Rahmen von Qualitätsmanagementmodellen wie EFQM festgelegten Parametern, kann Evaluation nicht auf einen fixierten Kanon von Bewertungskriterien zurückgreifen. Dies wäre angesichts der sehr verschiedenen Aufgabenstellungen und Untersuchungsgegenständen von Evaluation auch nicht sinnvoll. Sehr häufig orientieren sich die Bewertungskriterien allerdings am Nutzen eines Gegenstands, Sachverhalts oder Entwicklungsprozesses für bestimmte Personen oder Gruppen.

In der Entwicklungszusammenarbeit gibt es z.B. einen breiten Konsens darüber, dass Kriterien wie Relevanz, Effektivität, Effizienz, Wirkungen und Nachhaltigkeit in einer Evaluation nicht fehlen sollten. Die Festlegung, welche Kriterien für die Informationsbewertung in einer Evaluation verwendet werden sollen, kann auf verschiedene Weise erfolgen. Existieren Vorgaben, wie die vom DAC, dann werden sie häufig *direktiv* vom

Auftraggeber festgelegt. Manchmal wird die Bestimmung der Evaluationskriterien aber auch dem Evaluator überlassen, da man ihn für den Experten hält, der am besten wissen sollte, nach welchen Kriterien ein bestimmtes Programm zu beurteilen ist. Diese Kriterienauswahl könnte man als *wissens- oder erfahrungsbasiert* bezeichnen. Eher selten ist die Festlegung der Bewertungskriterien durch die Zielgruppe, also die Personen, die aus einem Programm Nutzen ziehen sollen. Bei einem solchen – emanzipativen – Vorgehen werden die Belange der (möglicherweise benachteiligten) Zielgruppe in den Vordergrund gestellt. Die subjektiv wahrgenommene Sicht der Betroffenen bei der Auswahl der Kriterien soll sicherstellen, dass deren Bedürfnisse und Erfordernisse prioritär in die Bewertung von Evaluationsergebnissen eingehen. *Partizipativ* könnte ein Verfahren genannt werden, bei dem Auftraggeber, Evaluatoren, Vertreter der Zielgruppen und andere Stakeholder gemeinsam die Bewertungskriterien für die Evaluation festlegen, um möglichst viele Perspektiven zu berücksichtigen.

- (4) Evaluationen können prinzipiell von internen oder externen Experten durchgeführt werden. Als *intern* werden Evaluationen bezeichnet, wenn sie von der gleichen Organisation vorgenommen werden, die auch das Programm oder das Projekt durchführt. Wird diese interne Evaluation von Mitarbeitern der Abteilung (dem Referat) durchgeführt, die gleichzeitig mit der operativen Durchführung des Programms betraut sind, dann wird von ‚*Selbstevaluation*‘ gesprochen. Nehmen Mitarbeiter einer anderen Abteilung des Hauses (z.B. einer Evaluations- oder Qualitätssicherungsabteilung) die Evaluation vor, dann handelt es sich zwar um eine interne Evaluation, aber nicht um eine Selbstevaluation.

„*In-house*“-Evaluationen haben den Vorteil, dass sie rasch und mit geringem Aufwand durchgeführt werden können, dass die Evaluatoren in der Regel über eine hohe Sachkenntnis verfügen und dass die Ergebnisse sich unmittelbar umsetzen lassen. Schwächen der internen Evaluation werden vor allem darin gesehen, dass die Evaluierenden zumeist nicht über eine ausreichende Methodenkompetenz verfügen, dass es ihnen an Unabhängigkeit und Distanz mangelt und dass sie möglicherweise so sehr mit ihrem Programm verhaftet sind, dass sie aussichtsreichere Alternativen nicht erkennen.

Externe Evaluationen werden von Personen durchgeführt, die nicht dem Fördermittelgeber oder der Durchführungsorganisation angehören. In der Regel weisen externe Evaluatoren deshalb eine größere Unabhängigkeit, eine profunde Methodenkompetenz und professionelles Evaluationswissen auf und kennen das Fachgebiet, in dem das Programm bzw. das Projekt angesiedelt ist. Zudem können externe Evaluationen reformerischen Kräften innerhalb einer Organisation zusätzliche Legitimität und Einflussstärke verleihen, die sie benötigen, um Veränderungsprozesse in Gang zu setzen.

- (5) Grundlegend für die Frage ‚wie evaluiert wird‘ ist die Wahl des Forschungsparadigmas. Grob kann zwischen *zwei Hauptrichtungen* unterschieden werden. Die einen betrachten Evaluation als ein *empirisch-wissenschaftliches Verfahren*, das der kritisch-rationalen Forschungslogik folgt und prinzipiell alle bekannten empirischen Forschungsmethoden für einsetzbar hält. Evaluation ist somit als angewandte Sozialforschung zu verstehen, die besondere Forschungsbedingungen zu berücksichtigen und ein spezifisches Erkenntnis- und Verwertungsinteresse hat, bei dem der Nutzen der Evaluationsergebnisse für die ‚Praxis‘ im Vordergrund steht. Die zweite Hauptrichtung verbindet mit Evaluation einen anderen Anspruch und geht von anderen Voraussetzungen aus. Das Vorhandensein einer real existierenden Welt, die prinzipiell erkannt und „objektiv“ mit Hilfe empirisch-wissenschaftlicher Verfahren erfasst werden kann, auch wenn diese Instrumente unvollständig und teilweise fehlerhaft sein können, wird bestritten. Stattdessen wird angenommen, dass „*Realität*“ aus verschiedenen Perspektiven *sozial konstruiert* ist, die in Konflikten zueinander stehen können. Deshalb fordern die Anhänger dieses Ansatzes ein ‚qualitatives‘ Denken, um die verschiedenen Sichtweisen und Interpretationen der ‚Realität‘ erfassen zu können.

In der Evaluationspraxis wird häufig ein sogenannter Methodenmix angewendet, der aus qualitativen und quantitativen Verfahren besteht. Dadurch sollen die Schwächen einzelner Ansätze durch die Stärken anderer ergänzt werden.

Literatur:

Fitzpatrick, J. L.; Sanders, J.R.; Worthen, B.R. (2004): Program Evaluation. Alternative Approaches and Practical Guidelines. 3. Auflage, Boston u.a.: Pearson.

Rossi, P. H.; Lipsey, M. W.; Freeman, H. E. (2004): Evaluation. A systematic Approach. Thousand Oaks u.a.: Sage.

Shaw, I. F.; Greene, J. C.; Mark, M.M. (2006) (Hg.): The SAGE Handbook of Evaluation. Thousand Oaks u.a.: Sage.

Stockmann, R. (2006): Evaluation und Qualitätsentwicklung. Münster: Waxmann.

Stockmann, R. (2007) (Hg.): Handbuch zur Evaluation. Münster: Waxmann.

Stockmann, R.; Meyer W. (2010): Evaluation. Eine Einführung. Opladen, Farmington Hills: Verlag Barbara Budrich.

Bibliographische Angaben

Stockmann, Reinhard:

Evaluation – eine Begriffsdefinition

Saarbrücken: Centrum für Evaluation, 2011.