

# Das **A-B-C** der **wirkungsorientierten** **Evaluation**

Glossar – Deutsch / Englisch – der wirkungsorientierten Evaluation

Wirkung  
Indikator  
Outcomes  
Wirkstrang  
Meta-Evaluation

Programmbaum  
mock evaluation  
Effektivität  
Simulationssstudie

## Institut für Evaluation

Dr. Beywl & Associates GmbH  
Hohenstaufenring 63

D-50674 Köln

Tel.: +49 (0) 221/424 80 71

Fax: +49 (0) 221/424 80 72

[www.univation.org](http://www.univation.org)

ISBN 978-3-00-028910-1

Das Werk einschließlich seiner Teile ist urheberrechtlich geschützt. Jede Verwertung außerhalb der engen Grenzen des Urheberrechtsgesetzes ist ohne Zustimmung des Herausgebers unzulässig und strafbar. Dies gilt insbesondere für Vervielfältigungen, Übersetzungen, Mikroverfilmungen und die Einspeicherung und Verarbeitung in elektronischen Systemen.

**Impressum:**

Autoren: Wolfgang Beywl/Melanie Niestroj  
Herausgeber: Univation Institut für Evaluation  
Dr. Beywl & Associates GmbH  
ISBN 978-3-00-028910-1  
Layout/Satz: grafixx | marion schmidt  
☛ info@grafixx-koeln.de

Vertrieb: Univation Institut für Evaluation  
Dr. Beywl & Associates GmbH  
Hohenstaufenring 63  
D-50674 Köln  
Tel.: +49 (0) 221/424 80 71  
Fax.: +49 (0) 221/424 80 72  
info@univation.org

# Das **A-B-C** der wirkungsorientierten Evaluation

Glossar – Deutsch / Englisch – der wirkungsorientierten Evaluation

Herausgegeben von

**Univation** Institut für Evaluation GmbH  
Dr. Beywl und Associates, Köln  
2., vollständig bearbeitete und ergänzte Auflage

– Landmarke wirkungsorientierter Evaluation<sup>1</sup> –

Wolfgang Beywl/Melanie Niestroj

Nach einer kurzen Darstellung der Absichten dieses Beitrages wird der Entstehungshintergrund des Programmbaus skizziert, um dann das Modell selbst und seinen Nutzen für Evaluierende sowie Programmverantwortliche zu erläutern. Es wird aufgezeigt, wie Resultate eines Programms ausdifferenziert werden können und wie sich die Bezeichnung des Programmbaus als 'Landmarke wirkungsorientierter Evaluation' begründen lässt. Der Programmbaum ermöglicht die Unterscheidung unterschiedlich herausfordernder Evaluationsniveaus, welche Anforderungen an eine wirkungsorientierte Evaluation verdeutlichen. Diese werden vorgestellt und ergänzt um einige Methoden und Instrumente, welche die Arbeit mit dem Programmbaum unterstützen.

Zielsetzung des Beitrags

Folgende Überlegungen haben uns bewogen, den Programmbaum in diesem Glossar vorzustellen:

... Der Programmbaum veranschaulicht, wie ein zusammenhängendes Bündel definierter Begriffe strukturiert ist, macht exemplarisch einen Teil des großen Netzes sichtbar, das die Begriffe des Glossars verbindet. Dieses Begriffs-Netz ist ein Orientierungsangebot für die komplexe und anspruchsvolle Arbeit in der → *Evaluation*. Die Querverweise in diesem Beitrag ermöglichen darüber hinaus eine Vorausschau auf die benachbarten begrifflichen Teilnetze.

... Die visuelle Darstellung bezieht sich mit dem → *Programm* auf einen der Zentralbegriffe der Evaluation. Er bildet einen wichtigen Knotenpunkt im Gesamtnetz. Für viele → *Beteiligte* und → *Betroffene* ist dieser Terminus sperrig. Es ist daher eine 'didaktische' Aufgabe, den möglichen → *Nutzen* heraus zu arbeiten, wenn man mit dem Konstrukt des Programms arbeitet. Dies gilt analog für zahlreiche andere Fachbegriffe der Evaluation, die mit der jeweiligen Sprache des → *Evaluationsfeldes* vermittelt, d. h. kommuniziert, abgewogen, eventuell angepasst werden müssen.

... Der Programmbaum unterstützt → *Evaluierende* substantiell bei der Planung und praktischen Durchführung einer → *wirkungsorientierten Evaluation*.

... Er hilft darüber hinaus Programmverantwortlichen und Mitarbeitenden nicht nur, ihr eigenes Programm besser zu verstehen, seine Logik zu entschlüsseln und die darauf gerichtete Evaluation besser nachzuvollziehen. Er hilft auch,

<sup>1</sup> Zu allen 79 gekennzeichneten Begriffen finden sich Definitionen im Glossarteil. Dabei ist jeder Begriff – wie auch in den Erläuterungen im Glossar – jeweils nur bei seiner ersten Verwendung gekennzeichnet.

U	triangulation	Triangulation
	trump criterion point [deal-breaker]	Trumpf-Kriterienpunkt
	undesired results (of a programme)	unerwünschte Resultate (eines Programms)
	unit of analysis	Untersuchungseinheit
	unit of collection	Erhebungseinheit
	unit of description	Aussageeinheit
	unit of sampling	Auswahleinheit
	use (of the evaluation/ of the evaluation findings)	Nutzen (der Evaluation/ der Evaluationsergebnisse)
	users (of the evaluation process or findings)	Nutzende (von Evaluationsprozess oder Evaluationsergebnissen)
	utilisation (of evaluation)	Nutzung (der Evaluation)
V	utilisation-focused evaluation	nutzungsgesteuerte Evaluation
	utility (of an evaluation)	Nützlichkeit (einer Evaluation)
	utility analysis	Nutzwertanalyse
	validity	Gültigkeit; Validität
	value (of an evaluand)	Wert (eines Evaluationsgegenstandes)
	value clarification method	Wertklärungsmethode
	value clarification techniques	Wert-Klärungs-Techniken
	value distant evaluation models	wertedistanzierte Evaluationsmodelle
	value of a variable	Merkmalsausprägung; Variablenwert
	value positioned evaluation models	wertepositionierte Evaluationsmodelle
W	value prioritising evaluation models	wertepriorisierende Evaluationsmodelle
	value relativistic evaluation models	wertrelativistische Evaluationsmodelle
	variable	Variable
	working session (within evaluations)	Arbeitssitzungen (in der Evaluation)
	worth (of a programme)	Tauglichkeit (eines Programms)
	written report (of an evaluation)	Bericht, schriftlicher (einer Evaluation)

Programme besser zu planen und damit deren → Wert für → Zielgruppen und Gesellschaft zu gewährleisten.

## Entstehungshintergrund des Programmbaums

In der Geschichte der Evaluation gibt es mehrfach Versuche, der Komplexität von → Evaluationsgegenständen – hier Programme aus verschiedensten Gegenstandsfeldern – durch Modellierung Herr zu werden. Es geht darum, die relevanten Elemente der Wirklogik des betreffenden Programms zu identifizieren, diese in ihrem Zusammenhang systematisch zu beschreiben und damit evaluierbar zu machen. Vielfach bedient man sich dafür visueller Darstellungen, die bis zu einem Dutzend Elemente enthalten. Es gibt verschiedene Modelle, die als Heuristik genutzt werden können, wenn man sich einem Gegenstand nähert.

US-amerikanische Evaluationstheoretiker begannen in den späten 1960er Jahren damit, richtungsweisende Modellierungen zu entwickeln: Bob Stake (1972) unterscheidet in seinem an (schul-)didaktische Modelle angelehnten Countenance-Ansatz „Voraussetzungen, Prozesse und Ergebnisse“ und kombiniert diese mit fünf Evaluationsdimensionen. Daniel Stufflebeam (1972) entwickelt das → CIPP-Modell<sup>2</sup> (context-input-process-product). Seit den 1980er Jahren werden solche Modellierungen von Evaluationsgegenständen als → logische Modelle bezeichnet.<sup>3</sup>

Ein einfaches logische Modell verbindet Informationen zum → Konzept, zur → Intervention und den → Resultaten. Ein Beispiel aus der Beschäftigungsförderung ist in der Abbildung 1 skizziert.

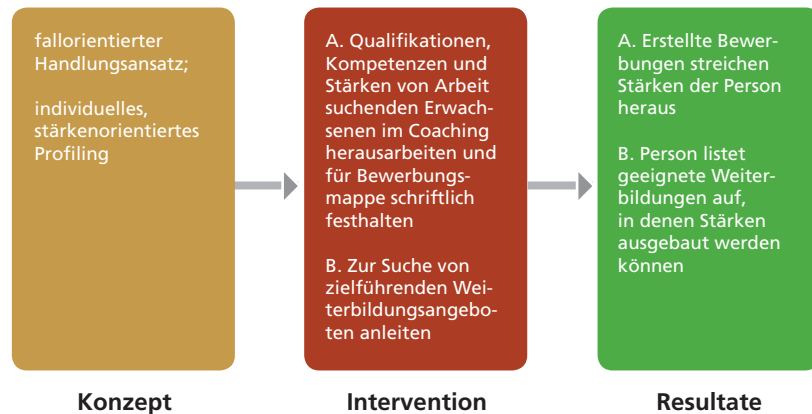


Abbildung 1: Einfaches logisches Modell am Beispiel Beschäftigungsförderung

<sup>2</sup> Eine Übersetzung der aktuellen CIPP-Checkliste findet sich auf: [www.evaluationsstudium.ch](http://www.evaluationsstudium.ch) > Weiterführende Informationen > Checklisten.

<sup>3</sup> Einen geschichtlichen Abriss der logischen Modelle in der Evaluationstheorie bieten McLaughlin/Jordan (1999).

Diese Darstellung ist stark verkürzt. So bleibt offen, was die Ausgangsbedingungen der zu beratenden Zielgruppe sind und in welcher Lebenslage sie sich befindet. Beispielsweise können vorliegende gesundheitliche oder Suchtprobleme die Erfolgswahrscheinlichkeit eines Profilings stark einschränken, ggf. eine Anpassung des Konzepts erforderlich machen.

In ausgearbeiteten Modellen, z. B. im CIPP-Modell, wird daher den → Bedingungen große Aufmerksamkeit geschenkt, unter denen das Programm agiert und die einen wichtigen Einfluss auf die Gestaltung von Konzept und Umsetzung haben. Was im CIPP-Modell unter `C` gefasst wird, ist im Programmbaum in vier Elemente ausdifferenziert:

- *Kontext I*, der ökonomische, politische, kulturelle und soziale Rahmenbedingungen des Programms umfasst
- *Struktur*, d. h. organisationale Merkmale der programmtragenden Organisationen sowie des aktiven Kooperationsnetzes
- *Incomes*, also Kenntnisse, Haltungen sowie weitere Voraussetzungen der Zielgruppen
- *Inputs*, also von Programmfinanziers und -trägern eingebrachte finanzielle und Humanressourcen

Vom Kontext zu den Inputs ist die Stabilität der Bedingungen unterschiedlich ausgeprägt: Der Kontext existiert weitgehend unabhängig vom Programm und ist durch dieses – wenn überhaupt – nur längerfristig beeinflussbar. Je nachdem ist die Struktur ebenfalls sehr stabil und eine kaum beeinflussbare Größe. Hingegen lassen sich Inputs und Incomes mittel- oder gar kurzfristig beeinflussen, sei es durch Personaleinstellungen oder -qualifizierung oder durch Gewinnung oder Auswahl von Programmteilnehmenden.

Auch die Resultate lassen sich nach zeitlichen Aspekten und unterschiedlicher Komplexität ausdifferenzieren: einerseits in eng mit der Programmumsetzung verbundene → Produkte und Leistungen (im Beispiel: Anzahl der erstellten und versandten Bewerbungsmappen), andererseits in → Outcomes, d. h. erwünschte Veränderungen und Stabilisierungen bei Zielpersonen (im Beispiel: Kenntnisse zu eigenen Stärken), sowie meist länger- und langfristige → Impacts, Veränderungen/Stabilisierungen in beteiligten Organisationen und im gesellschaftlichem Subsystem, in dem diese agieren (im Beispiel: verbesserte Zusammenarbeit zwischen Arbeitsverwaltung und Weiterbildungseinrichtungen).

Auf Basis einer Vielzahl von Vorbildern und Vorarbeiten hat Univation das umfassende logische Modell *Programmbaum* entwickelt. Dieser wird ob seiner Anschaulichkeit und `didaktischen Anleitungskraft` in zahlreichen Feldern nicht nur von Evaluierenden, sondern auch von Programmverantwortlichen und -mitarbeitenden angewendet.

## Nutzen des Programmbaums

Der Programmbaum richtet sich besonders an Evaluierende und Programmverantwortliche, also die beiden Hauptakteure in der → *Programmevaluation*.

Er unterstützt Programmverantwortliche wesentlich dabei, ...

- ihr Programm zu planen und zu steuern, d. h. den logischen Zusammenhang zwischen Voraussetzungen und → *Aktivitäten* / → *Interventionen* sowie den erwünschten Resultaten herzustellen und dies in eine systematische Form zu bringen. Sie werden unterstützt, auf alle relevanten Aspekte zu achten, bspw. die Bedingungen ihres Handelns in die Planung und Steuerung einzubeziehen. So können Änderungen in Bedingungen auf Grundlage des Programmbaums auf Konsequenzen für die gesamte Programmlogik geprüft werden.

- (über) das Programmkonzept angemessen zu kommunizieren (bspw. mit Evaluierenden, aber auch anderen Personengruppen, wie Finanziers); teils wird der ausgefüllte Programmbaum zur Öffentlichkeitsarbeit genutzt.

Der Programmbaum unterstützt Evaluierende wesentlich dabei, ...

- den Evaluationsgegenstand systematisch zu beschreiben und dabei zentrale Elemente und Verknüpfungen zu identifizieren, die ggf. im Rahmen von → *Evaluationsfragestellungen* aufzugreifen sind.

- die → *Evaluierbarkeit* des Programms abzuschätzen.

- mit Programmverantwortlichen und → *Auftraggebenden* zu kommunizieren und mit ihnen eine gemeinsame Sprache zu finden.

- → *Daten*, → *Informationen*, → *Schlussfolgerungen* etc. zuzuordnen und sie zu verbinden. Daraus kann auch eine Gliederung für → *Rückmeldungen* und → *Berichterstattung* abgeleitet werden.

## Der Programmbaum mit seinen Elementen

Die folgende Tabelle 1 gibt Erläuterungen zu den Elementen des unten abgebildeten Programmbaums. Dabei sind alle Definitionen der → *Programmelemente* gegenüber den Glossardefinitionen sinngemäß gekürzt und werden um Beispiele, wiederum aus dem Bereich der Beschäftigungsförderung, ergänzt.

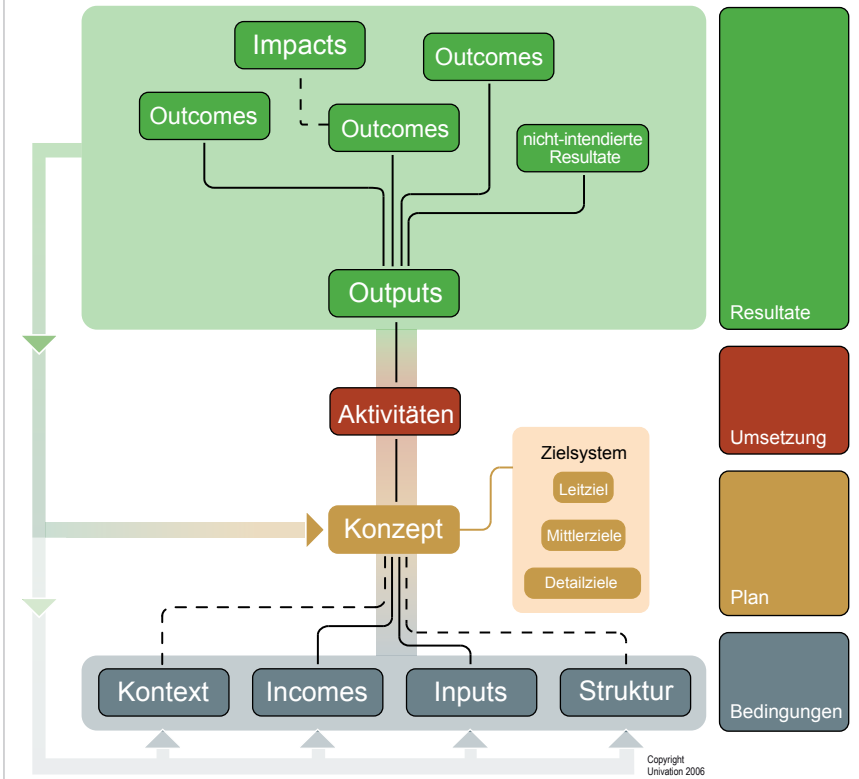


Abbildung 2: Der Programmbaum von Univation

Tabelle 1: Erläuterung der Elemente des Programmbaums

Element	Glossardefinition	Beispiele aus dem Bereich Beschäftigungsförderung
Kontext I *	Gesamtheit der umgebenden Systemumwelt eines Programms je nach dessen Reichweite auf lokaler, nationaler und/oder internationaler Ebene. Kann sich auf geographische, rechtliche, soziale, demographische, politische, kulturelle oder andere Aspekte beziehen, die sich in der Regel langfristig und weitgehend unabhängig vom Programm selbst ändern.	Gesetze wie SGB II, Anteile von Arbeit suchenden Personen an der Bevölkerung einer Region, gesellschaftliches Klima gegenüber Arbeitslosigkeit.

\*) Im Programmbaum wird nicht von 'Kontext I' gesprochen, um Nutzende (insb. Programmmitarbeitende) nicht zu verwirren. Die Unterscheidung zwischen dem → *Kontext I* des Programms und dem → *Kontext II*, der sich auf die Evaluation bezieht, soll ausschließlich aus evaluationstheoretischer Sicht, bspw. in diesem Glossar, getroffen werden.

Struktur	Bedingungen, die beim Träger (oder: darüber hinausgehend in einem Verbundsystem) eines Programms vorliegen. Darüber hinaus weitere an einem Programm beteiligte Institutionen und/oder Personen mit ihren Beziehungen und Verantwortungsbereichen.	Kooperationen, bspw. arbeitsteilige Beratung von Arbeit suchenden Personen durch ARGE und private Dienstleister, Beschäftigungsträger etc., Rechtsform, Kapitalausstattung und Leitbild der das Programm tragenden Organisation.
Incomes	Ressourcen, die Mitglieder der Zielgruppen in das Programm einbringen.	Lerneingangsbedingungen der Teilnehmenden von Qualifizierungsmaßnahmen, (Vorwissen, Motivation etc.), körperliche Verfassung.
Inputs	Finanzielle, personale oder andere Ressourcen, die in ein Programm investiert werden, teils als monetäre Kosten erfassbar.	Personal- und Sachaufwendungen, Anzahl und Qualifikationen der im Programm Tätigen, bspw. spezielle Fortbildungen in lösungsorientierter Beratung.
Konzept	Gedanklicher (Grob-)Entwurf von Verantwortlichen eines Programms dazu, mit Hilfe welcher Interventionen welche Resultate bis wann, wo/bei wem ausgelöst werden sollen und wie der Programmprozess insgesamt gesteuert und überwacht werden soll.	Projekt-/Förderantrag für eine Qualifizierungsmaßnahme oder ein Profiling-Angebot, Curricula für Weiterbildungsangebote etc.
Zielsystem	Gesamtheit expliziter, schriftlich dargelegter Ziele für ein Programm auf den drei Konkretionsstufen Leitziele, Mittlerziele und Detailziele. Detailziele konkretisieren/spezifizieren Mittlerziele, Mittlerziele tun dies für Leitziele. Die Ziele einer niedrigeren Konkretionsstufe sind logisch darauf geprüft, dass ihre Erreichung zur Verwirklichung der Ziele auf der höheren Stufe beiträgt.	---

Leitziel	Ein weit gestecktes, abstraktes Ziel, das durch ein Programm erreicht werden soll. Ein solches Leitziel bleibt über lange Zeit, oft mehrere Jahre stabil; es ist selten möglich, es in vollem Umfang zu erreichen.	Mögliches Leitziel: Die Beschäftigungsfähigkeit von Jugendlichen nach dem Hauptschulabschluss ist erhöht und die Lehrstellensuche ist erfolgreich.
Mittlerziele	Verbinden Leitziele mit den Detailzielen eines Programms und gliedern/konkretisieren die Schwerpunkte im Bereich des Leitziels. Mittlerziele sind mittelfristig gültig und können in diesem Zeitraum verwirklicht werden.	Mögliches Mittlerziel: Jugendliche nach dem Hauptschulabschluss meistern die für eine Bewerbung erforderlichen Schritte und verfügen über die erforderlichen Instrumente.
Detailziele	Konkretisierte Mittlerziele, die im Idealfall spezifisch, messbar, akzeptabel, realistisch und terminiert (S.M.A.R.T.) sind.	Mögliches Detailziel: Alle Jugendlichen der Klasse haben vor Erhalt des Abschlusszeugnisses der Hauptschule eine aktuelle, die individuellen Stärken klar benennende Bewerbungsmappe erstellt.
Aktivitäten	Handlungen der beruflich oder ehrenamtlich im Programm Tätigen: (a) Interventionen, die sich auf die Zielerreichung richten, (b) Hilfsaktivitäten, welche die Interventionen ermöglichen oder unterstützen, und (c) Zielgruppen-Aktivitäten, die eine aktive Beteiligung der Zielpersonen erfordern.	Interventionen: Beratung, Profiling...  Hilfsaktivitäten: Teamaufbau, Einarbeiten neuer Mitarbeitender...  Zielgruppen-Aktivitäten: Lernen, Bewerben...
Outputs	In unmittelbarem Zusammenhang mit den Interventionen eines Programms stehende Resultate in Form (zählbarer) Leistungen/Produkte, Teilnahme-/Benutzungsmengen sowie Zufriedenheit der Benutzenden/Zielgruppen.	Anzahl Teilnehmerstunden, verteilte Informationsbroschüren, Hits auf der Webseite der Organisation, Anzahl Beratungsgespräche/Profilingen.

Outcomes	Intendierte Resultate bei Zielgruppen, wie z. B. Veränderungen bzw. Stabilisierungen im Wissen, den Einstellungen, sozialen Werten oder dem Können, im Verhalten oder in der Lebenslage/dem Status der Zielpersonen.	Erweitertes Wissen über Bewertungsmöglichkeiten, erhöhte Motivation zur Bewerbung oder zum Besuch einer Fortbildung, verstärkte Bewerbungstätigkeit, Integration in den Arbeitsmarkt.
Impacts	Resultate eines Programms, die über die bei Zielgruppen auftretenden Outcomes hinausgehen, z. B. in Form von Merkmalen sozialer Systeme, einer Organisation, eines Sozialraums, eines regionalen Weiterbildungssystems, der Bildungspolitik eines Bundeslandes.	Verbesserte Zusammenarbeit der Arbeitsverwaltung mit Weiterbildungsanbietern, Ausbau eines bedarfsgerechten Angebots an Beratungsangeboten sowie Qualifizierungsmaßnahmen, bedarfsgerechter öffentlicher Nahverkehr für Berufspendler und -pendlerinnen.
nicht intendierte Resultate	Resultate eines Programms auf verschiedenen möglichen Ebenen, die in dessen Konzept nicht als zu erzielend vorgesehen sind, also außerhalb der im Konzept festgehaltenen Ziele liegen.	Wohn-Attraktivität eines Stadtteils, in dem intensiv Maßnahmen zur Beschäftigungsförderung betrieben werden, steigt an, dadurch Beeinflussung von Wohnungsmarkt/Mietpreisen.

## Stufenfolge von Resultaten

In der praktischen Arbeit (und auch hier im Glossar) werden sowohl → *Outputs* als auch *Outcomes* abgestuft. So werden jeweils *Outputs* und *Outcomes* I., II. und III. Ordnung unterschieden (siehe Abbildung 3). Die Überlegung hinter der so genannten Resultaten-Treppe ist, dass es oft eines Resultats auf einer niedrigeren Stufe bedarf, damit ein Resultat auf einer höheren Stufe eintreten kann. So ist bspw. die erweiterte Kenntnis von Recherchemöglichkeiten nach Ausbildungsplätzen (*Outcome* I) eine Voraussetzung dafür, dass Jugendliche eine verstärkte und gezielte Bewerbungstätigkeit entfalten können (*Outcome* II). Ein Denken und Planen in diesen Abfolgen erhöht die Konsistenz eines Programmentwurfs und zeigt die verschiedenen Ansatzpunkte für Fragestellungen und Datenerhebungen in der Programmevaluation.

Für die → *Güte* und → *Tauglichkeit* von Programm und Evaluation ist es in der Regel ein wichtiges → *Kriterium*, dass die Kette von den → *Outputs* I bis hin zu den *Impacts* geschlossen, jedes Glied sichtbar und beschrieben ist. Die Abbildung unterstützt die Programmverantwortlichen dabei, ihre → *Ziele* zu klären und dabei genau festzulegen, wie herausfordernd diese ausfallen sollen (bis auf welche Stufe der Treppe die Programmverantwortlichen gehen wollen). Dabei ist es stets abhängig vom jeweiligen Programm (seinen *Inputs*, *Incomes* etc.), ob das Erreichen des → *Output-Ziels* „Erreichung der richtigen Zielgruppen“ bereits eine große Herausforderung darstellt (wie bei vielen Benachteiligtenprogrammen) oder ob es sich (fast) von selbst einstellt (wie z. B. für die 1. Klasse der Grundschule).

### RESULTATE EINES PROGRAMMS

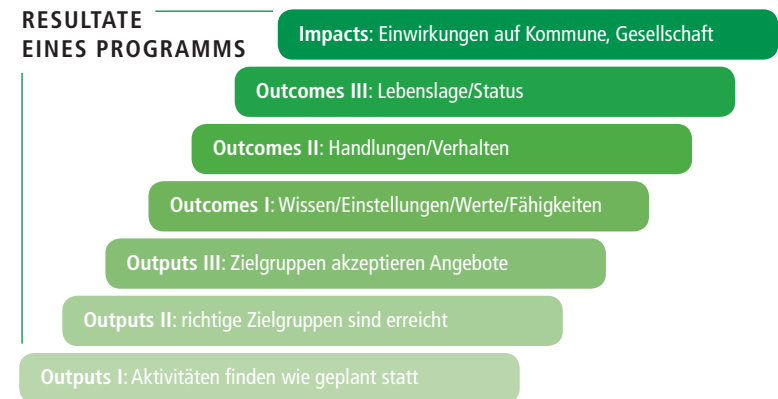


Abbildung 3: Die Resultate-Treppe von Univation



D. h. jedes Programm braucht seine eigene Gradierung an Zielen für die verschiedenen Resultatebenen (im Konzept auszuweisen).

## Der Programmbaum unterstützt die wirkungsorientierte Evaluation

Der Programmbaum drängt dazu, sowohl umfassend und vernetzt zu denken, als auch für jedes Element spezifizierte Beschreibungen zu erstellen. Er lenkt den Blick auf die Outcomes und Impacts, die sich in der Krone befinden.

Die Outcomes sind die 'Früchte' des Programms, auf die es ankommt. Dies erinnert Programmverantwortliche daran, dass die Formulierung von Output-Zielen in der Regel nicht ausreicht, um ein Programm von hoher → *Bedeutsamkeit* zu entwerfen. → *Outcome-Ziele* sind im Programmkonzept zwingend erforderlich.

Der Programmbaum kann – wie im Titel dieses Artikels – als 'Landmarke der wirkungsorientierten Evaluation' bezeichnet werden:

- da er Programmverantwortliche anhält, alle ihre Aktivitäten auf die erwünschten Resultate (insbesondere Outcomes) auszurichten
- da er es Evaluierenden erleichtert die Programmverantwortlichen beim Anstreben und Erzielen dieser Resultate zu unterstützen (→ *formative Evaluation*) oder das Programm mit Bezugnahme auf die gesetzten Outcome-Ziele (und auch andere Kriterien) bilanzierend zu bewerten (→ *summative Evaluation*).

Wie viele Outputs nötig sind, welche Arten fachlicher Aktivitäten gewählt werden und von wem (z. B. beruflich/ehrenamtlich Tätige) sie sinnvollerweise erbracht werden, hängt ab von der begründeten Fähigkeit der Interventionen, gewünschte Outcomes auszulösen. Ein nach dieser Logik entworfenes Konzept, das auch Analysen zu den Bedingungen des Programms aufnimmt, verspricht → *Wirkung*. Ob und in welchem Umfang diese tatsächlich eintritt, ist nicht garantierbar, aber → *Erfolg* kann mit diesem Planungsinstrument wahrscheinlicher gemacht werden.

Aus der Sicht der Evaluierenden leistet die Ausarbeitung des Programmbaums eine → *Wirkungsmodellierung*, die gemeinsam mit der Überprüfung der → *Zielerreichung* plausibel macht, dass die gemessenen Resultate tatsächlich Wirkungen des Programms, also auf dessen Interventionen rückführbar sind. Noch sicherer wird eine → *Wirkungsfeststellung*, wenn über das logische Modell hinaus eine → *Programmtheorie* erstellt wird, bei der die Verknüpfungen zwischen den einzelnen Elementen theoretisch, auf der Basis von Forschungsergebnissen oder bereits vorliegenden Evaluationen detailliert begründet und auch empirisch überprüft werden.

## Niveaus wirkungsorientierter Evaluation

Der Programmbaum ermöglicht es, verschiedenen anspruchsvolle Evaluationsniveaus zu unterscheiden. In einer fünfstufigen Folge nehmen von Ebene zu Ebene Aufwand und evaluationskonzeptionelle Herausforderungen ebenso zu, wie das Potenzial, zu einer wirksamen und wirtschaftlichen öffentlich verantworteten Programmpraxis beizutragen. Die höheren Niveaus setzen jeweils die Bearbeitung der vorangehenden voraus<sup>4</sup>:

Tabelle 2: Fünf Niveaus wirkungsorientierter Evaluation (Quelle: eigene Darstellung)

### 1 Bestimmung der Input-Output-Relation mittels → *Kosten-Leistungs-Analyse*

Es wird festgestellt, wie viele (geldwerte) → *Ressourcen* eingesetzt wurden, um die gezählten Outputs zu produzieren (vgl. z. B. Rieder 2004). Die ist eine prä-evaluative Aufgabe, d. h. wenn sich die Untersuchungen auf diesen Aspekt beschränken, kann noch nicht von Evaluation gesprochen werden.

### 2 Messung und → *Bewertung der (Outcome-)Zielerreichung*

Evaluation stellt fest, in welchem Ausmaß die im Zielsystem des Programms gesetzten Outcome-Ziele erreicht werden. Programme gelten dann als erfolgreich, wenn sie die gesetzten Ziele in befriedigendem Maße erreichen (vgl. z. B. W. K. Kellogg Foundation 1998).

### 3 Messung und Bewertung der Income-Outcome-Relation

Der Zuwachs an Wert oder Nutzen, den das Programm für die Zielgruppen geleistet hat, wird durch Evaluation festgestellt. Bestimmt wird über die Zielerreichung (also z. B. Ausprägung der Bewertungsmotivation) hinaus der Unterschied, den das Programm bei den Zielgruppenmitgliedern macht (der Zuwachs gegenüber der Einstiegssituation in das Profiling). Hierfür ist ein methodisch angemessenes → *Erhebungsdesign* erforderlich, da der Messvorgang allein (z. B. ein Interview vor Beginn des Profiling) zu einem erhöhten Outcome führen kann (Motivation ist erhöht, da Interview als wertschätzende Aufmerksamkeit gewertet wurde).

### 4 Feststellung und Bewertung der → *Wirksamkeit des Programms*

Evaluation erbringt Belege, in welchem Maße empirisch nachgewiesene Outcomes (also z. B. abgeschlossene Lehrverträge Bildungsbenachteiligter) auf das Programm, insbesondere seine Interventionen zurückzuführen sind (bspw. ein Bewerbungstrainings oder Übungsfirma-Simulationen). Dies kann – nach Kostenaufwand für die Evaluation gestaffelt – jeweils unter Berücksichtigung identifizierter Einflussfaktoren aus dem Kontext I des Programms geleistet werden durch:  
(a) → *Wirkungsabschätzung*: Einschätzung des Programm-Beitrags zu den gemessenen Outcomes durch Feldexperten und -expertinnen, die weder Mitglieder des Programm- noch des Evaluationsteams sind  
(b) *Wirkungsmodellierung*: Füllung des Programmbaums und Messungen bei den programmtheoretisch relevanten Elementen, insbesondere von Outcomes und Outputs  
(c) → *Wirkungsnachweis*: empirische, statistisch abgesicherte Wirkungsmessungen im Rahmen → *experimentaldesigngesteuerter Evaluation* oder → *quasi-experimentaldesigngesteuerter Evaluation* unter Nutzung von → *Experimentalgruppen* und → *Vergleichsgruppen* bzw. → *Kontrollgruppen* (vgl. bspw. Bortz/Döring 2006)

### 5 Bestimmung und Bewertung der → *Wirtschaftlichkeit* mittels → *Kosten-Wirkungs-Analyse*

Evaluation überprüft, wie kostengünstig (Kosten im weitesten Sinne) der festgestellte Programmnutzen (bei Bilanzierung von Outcome-Wirkungen, erwünschten und unerwünschten nicht vorhergesehenen Resultaten) erbracht wird (Hinweise geben Levin/McEwan 2001).

<sup>4</sup> In der Praxis ist dies ein oft verpasstes Ideal: Bei vielen Evaluationen des Niveaus 2 fehlen Angaben über Outputs, die mittels → *Monitoring* bereitgestellt werden müssen. Noch mehr aber fehlen selbst rudimentäre Angaben zu den Kosten des Evaluationsgegenstands. Die Gewinnung dieser Daten ist insbesondere für Evaluierende in → *externen Evaluationen* oft sehr schwierig. Sie sollten – zusammen mit den Geldgebenden – darauf drängen, dass die Programmverantwortlichen zumindest eine vereinfachte Form der Kosten-Leistungs-Analyse als festen Bestandteil in ihren operativen Auftrag übernehmen, wenn dies nicht bereits mit Steuerungsverfahren z. B. des New Public Management implementiert ist.



## Instrumente und Methoden rund um den Programmbaum

Im Verlauf der letzten Jahre sind beim Einsatz des Programmbaums in Evaluationsprojekten eine Reihe von → *Evaluationsmethoden* und → *Datenerhebungsinstrumenten* entstanden, welche die Logik des Programmbaums mit seinen einzelnen Elementen aufnehmen und es ermöglichen, die Elemente im Verlauf einer Evaluation systematisch zu konkretisieren und angemessen einzubeziehen. Zu nennen ist hier bspw. die → *Struktur-Landkarte*, eine grafische Darstellung der das Programm tragenden bzw. an seiner Umsetzung beteiligten Institutionen und Personen mit ihren Beziehungen und Zuständigkeiten. Zudem gibt es mit dem → *Bedingungsmonitoring* die Möglichkeit, relevante Änderungen in Bedingungen des Programms regelmäßig zu erheben und ihre Konsequenzen für die Programmumsetzung im Blick zu behalten.

Die → *Bedingungsklärung* ist erforderlich, um die → *Übertragfähigkeit* des Programms bzw. einzelner Elemente oder auch → *Wirkstränge* zu beurteilen. Schließlich wird die → *Zielklärung*, die fast schon als klassische Aufgabe der Evaluation gilt, wenn Ziele des Programms nicht ausreichend konkret bzw. expliziert sind, mit der Einbindung des → *Zielsystems* in den Programmbaum systematisch fundiert.

## Literatur zu diesem Artikel

Bortz, Jürgen/Döring, Nicola (2006): *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler*. 4. Aufl. Berlin: Springer.

McLaughlin, John A./Jordan, Gretchen B. (1999): "Logic models: a tool for telling your program's performance story". In: Wholey, Joseph S./Hatry, Harry P./Newcomer, Katheryn E. (Hg.) (1999): *Evaluation and program planning*. San Francisco: Jossey Bass. 65-72.

Levin, Henry M./McEwan, Patrick J. (2001): *Cost-effectiveness analysis: Methods and Applications*. 2nd edition. Thousand Oaks: Sage.

Rieder, Lukas (2004): *Kosten-/Leistungsrechnung für die Verwaltung*. Bern: Haupt.

Stake, Robert E. (1972): „Verschiedene Aspekte pädagogischer Evaluation“. In: Wulf, Christoph (Hg.): *Evaluation*. München: Pieper. 92-112.

Stufflebeam, Daniel L. (1972): „Evaluation als Entscheidungshilfe“. In: Wulf, Christoph (Hg.): *Evaluation*. München: Pieper. 113-145.

W. K. Kellogg Foundation (1998): *Evaluation handbook: philosophy and expectations*. Battle Creek: W. K. Kellogg Foundation.

## ÜBER DEN AUTOR/DIE AUTORIN

**Wolfgang Beywl** ist wissenschaftlicher Leiter und Gesellschafter der Univation – Institut für Evaluation GmbH in Köln sowie stellvertretender Direktor des Zentrums für universitäre Weiterbildung (ZUW) an der Universität Bern; dort leitet er das Weiterbildungsprogramm Evaluation mit dem Abschluss „Master of Advanced Studies in Evaluation“.

Studium der Soziologie, der Erziehungs- und Politikwissenschaft in Bonn mit Abschluss Magister Artium 1977; 1988 Promotion in Wirtschaftswissenschaft und ihre Didaktik an der Universität zu Köln mit einer Arbeit zur Evaluationsmethodologie; 1998 Weiterbildungsabschluss „Organization Consulting“ Cleveland Gestalt Institut.

Bis 1999 Forscher und Evaluator an den Universitäten Bonn, Köln und Wuppertal; Gründung und Leitung freier Forschungsinstitute. Durchführung von Evaluationen und Evaluationsberatungen für Stiftungen, Ministerien, Träger von sozialen Dienstleistungen, Aus- und Weiterbildungseinrichtungen, Hochschulen. Lehrt Evaluation in mehreren Weiterbildungsprogrammen der Universität Bern, sowie als Gastdozent an in- und ausländischen Hochschulen und Weiterbildungseinrichtungen. 1997 Gründungsmitglied der Gesellschaft für Evaluation; bis 2005 Vorstandsmitglied (DeGEval).

Ca. 100 Fachpublikationen auf dem Gebiet Evaluation, darunter Handbuch der Evaluationsstandards (VS-Verlag 3. Auflage 2006); Evaluation – ein systematisches Handbuch (VS-Verlag 2009), sowie zahlreiche veröffentlichte Evaluationsberichte. Gründer der deutschsprachigen elektronischen Diskussionsliste forum-evaluation.

**Melanie Niestroj** ist Diplom Erziehungswissenschaftlerin und seit 2002 wissenschaftliche Mitarbeiterin im Team der Univation – Institut für Evaluation GmbH in Köln. Während dieser Zeit leitete und begleitete sie Evaluationsprojekte in verschiedenen Feldern. Neben der Tätigkeit als Evaluatorin und Evaluationsberaterin ist sie aktiv in der Vermittlung von Evaluationskompetenzen bspw. im Rahmen des Weiterbildungsprogramms Evaluation an der Universität Bern. Sie wirkt aktiv mit an der Entwicklung und Weiterentwicklung von Evaluationsmethoden und -instrumenten wie dem Programmbaum.