

# Bewertungsproblem als Herausforderung bei Programmevaluationen – methodische Lösungsbeiträge

Martina Brandt, Ulla Große

## Zusammenfassung

Programmevaluationen stellen hohe Anforderungen an die Analyse, Strukturierung, Bewertung und Validitätsprüfung von Projekt- und Kontextdaten sowie -informationen. Trotz einer großen Anzahl praktischer Evaluationsprojekte und steigender Anforderungen an ihre Durchführung gibt es in Deutschland bislang kaum empirische Untersuchungen zur Anwendbarkeit verschiedener Evaluationsansätze. Während für die Daten- und Informationsbeschaffung auf ein vielfältiges und erprobtes Methodenspektrum zurückgegriffen werden kann, liegt bislang nur für wenige Evaluationszwecke eine praktikable Bewertungsmethodik vor. An praktischen Beispielen aus Programmevaluationen werden für unterschiedliche Bewertungsprobleme methodische Lösungsansätze vorgestellt. Zugleich wird deutlich, inwiefern auch für Bewertungen in Evaluationen von Programmen vergleichbarer Komplexität aber unterschiedlichem Evaluationsauftrag eine Operationalisierung im methodischen Vorgehen möglich ist und wo die Ansatzpunkte für weitere Forschungsarbeiten liegen.

## Abstract

Programme evaluations place high demands on analysing, structuring, rating and validity-checking of project and context data and information. In spite of many practical evaluation projects and increasing requirements regarding their realization, in Germany there are only a few empirical studies about the application of different programme evaluation approaches. Whereas for collecting data many comprehensive and tried and tested methods are available, there is a practicable method for assessment only for a few tasks of evaluation. Using practical examples of programme evaluations, methodologies are demonstrated for different assessment problems. At the same time it is shown, to what degree an operationalization of procedure model is possible in the case of evaluations with the same complexity but a different purpose. Moreover, areas which would provide a basis for further research are indicated.

## 1 Spezifik von Programmevaluationen

Eine Evaluation ist die systematische Untersuchung des Nutzens oder Wertes eines Gegenstandes. Evaluationsgegenstände können z. B. Programme, Projekte, Produkte, Maßnahmen, Leistungen, Organisationen, Politik, Technologien oder Forschung sein. Die erzielten Ergebnisse, Schlussfolgerungen oder Empfehlungen müssen nachvollziehbar auf empirisch gewonnenen qualitativen und/oder quantitativen Daten beruhen [DeGEval 2008: S. 15].

Programme als Evaluationsgegenstand zeichnen sich u. a. dadurch aus, dass

- sie oft mehrjährige Laufzeiten haben,
- mehrere Projektträger unterschiedlichen Institutionentyps agieren,

- das (Haupt-)Ziel zumeist von mittlerer oder hoher Komplexität ist,
- mehrere Ziele bzw. Haupt- und Unterziele verfolgt werden,
- häufig qualitative Ziele formuliert sind,
- Projektträger innovative Instrumente und Vorgehensweisen anwenden,
- unterschiedliche Problemlösungswege erprobt werden,
- nicht selten Umsteuerungen z. B. in Projektkonzepten, erfolgen,
- Daten und Informationen sich entsprechend dem Projektfortschritt verändern und
- Programmwirkungen sich in verschiedenen Phasen von Wirkungsketten und in unterschiedlichen Zeithorizonten manifestieren können. [1]

Programme sind somit in Veränderung begriffene Evaluationsgegenstände. Da sie in aller Regel auf die Lösung aktueller Probleme zielen, die auf veränderte bzw. sich dynamisch verändernde Umfeldbedingungen zurückgehen, sind auch die Programmkontexte veränderlich. Die zu evaluierenden aufgelegten Programme werden nicht nur komplexer, auch die Erfolgsmessung wird zunehmend anspruchsvoller und die Wirkungsnachweise sind schwerer zu erbringen. Damit verbunden sind häufig Defizite in der Evaluationsrealität, auf deren Überwindung die Evaluierten zu Recht drängen. [z. B. Gerlach et al. 2003: 497]

Auch in Deutschland werden Programmevaluationen zunehmend als ein Instrument zur Steuerung politischer Entscheidungsprozesse genutzt, ihre Zahl hat stark zugenommen. Sie beziehen sich auf einzelne Aspekte oder auf ein umfassendes Design, bestehend aus:

- Ex-post-Bewertungen bisheriger Förderpraxis,
- formative und prozessbegleitende Bewertungen laufender Programme und Förderaktivitäten und
- Ex-ante-Beurteilungen zukünftiger Programme und Förderaktivitäten unter dem Gesichtspunkt der Wirksamkeit. [Simon et al. 2003, zitiert in Hupfer 2007: 72]

Mit der gewachsenen Rolle von Evaluationen sind auch die Anforderungen an Bewertungsprozesse in Evaluationen gestiegen. Nach jahrelangen Diskussionen herrscht heute weitgehend Konsens darüber, dass eine Evaluation strengen wissenschaftlichen Anforderungen genügen und quantitative und qualitative Methoden verwenden sollte (Multimethodenansätze), um für den Auftraggeber und andere Interessengruppen von maximalem Nutzen zu sein. [Stockmann 2004: 9]

Die seit 2001 vorliegenden Evaluationsstandards der Gesellschaft für Evaluation e. V. definieren z. B. Qualitätsanforderungen hinsichtlich Nützlichkeit, Durchführbarkeit, Fairness und Genauigkeit und fordern u. a.

- Grundlagen der Werturteile klar zu beschreiben,
- nur valide Informationen zu nutzen und
- Informationen systematisch auf Fehler zu prüfen.

[DeGEval 2008]

Der Evaluationszweck, also das vom Auftraggeber vorgegebene Ziel der Evaluation, kann sehr verschieden sein und sich auf einen oder mehrere Bereiche wie Erreichung des Programmziels, Kosten-Nutzen-Bewertung, Wirkungsanalyse, Qualitätssicherung, Projektsteuerung usw. beziehen. Oftmals ist in der Praxis der Evaluationszweck (gewollt oder ungewollt) zwischen »Erfolg des Programms/der Projekte«, »Zielerreichung«

und »erreichten Wirkungen« breit angelegt, aber wenig präzise formuliert. Dies stellt die Programmevaluation dann vor die Aufgabe, zunächst eine entsprechende Präzisierung des Evaluationszwecks vorzunehmen, um auf dieser Grundlage konkrete Bewertungsziele abzuleiten.

Die Notwendigkeit, hochkomplexe Gegenstände in der Evaluation adäquat abzubilden und in der Regel mehreren Evaluationszwecken gerecht zu werden, stellt hohe Anforderungen an die Analyse, Strukturierung, Bewertung und Validitätsprüfung von Projekt- und Kontextdaten sowie -informationen. Gefordert ist eine entsprechend dem Programm und den verfügbaren Ressourcen bestmöglich wissenschaftsbasierte sowie praktikable Analyse- und Bewertungsmethodik, die auch vom Auftraggeber und den Programmakteuren mit getragen wird.

Die Autorinnen haben sich in den letzten Jahren mit der Bearbeitung unterschiedlichster Evaluationsstudien für verschiedene Auftraggeber immer wieder diesem Anforderungsprofil gestellt und sich in die wissenschaftliche Diskussion zu den bestehenden methodischen Defiziten eingebracht. [Große et al. 2004]

## 2 Operationalisierungsgrad vorliegender Evaluationsmodelle und Zielsetzung des Beitrages

Angesichts einer kaum noch überschaubaren Komplexität und großen Anzahl praktischer Evaluierungsvorhaben gibt es in der Literatur zwar zahlreiche Vorschläge zu deren Strukturierung, jedoch bisher kaum empirische Untersuchungen zur Anwendbarkeit der verschiedenen Evaluationsansätze. Dies führt dazu, dass viele Evaluationsvorhaben in der Praxis ohne Bezugnahme auf bestehende Modelle auskommen (müssen) und es entsteht der Eindruck, dass das Rad der Evaluation immer wieder neu erfunden wird. [Stockmann 2007: 40]

Die Gründe hierfür sind vielfältig. So wird eine unzureichende Weiterentwicklung von Theorie und Methodik der Evaluation beklagt. »Zwar werden in den einzelnen Fächern, Sektoren und Ressorts ausgezeichnete Evaluationsstudien durchgeführt, doch (...) diese Erkenntnisse (kumulieren) nur unzureichend in einer Weiterentwicklung von Theorie und Methodik der Evaluation«. [Stockmann 2000: 34] Diese Einschätzung aus dem Jahr 2000 gilt, trotz erreichter Fortschritte, noch heute.

Als Gründe für die im internationalen Vergleich verzögerte Entwicklung der Evaluationsforschung in Deutschland werden u. a. Defizite in der Forschung sowie die institutionelle Zersplitterung und die Dominierung des Evaluationsmarktes durch Consulting- und Marktforschungsunternehmen ausgemacht. [Stockmann 2000: 31f] Die nachhaltige Institutionalisierung der Evaluation in der deutschen Forschungslandschaft als Voraussetzung für eine systematische Grundlagenforschung, der Theorie- und Methodenentwicklung sowie fachwissenschaftlicher Standards hat noch keinen zufriedenstellenden Stand erreicht.

Die Bündelung und Strukturierung des in zahlreichen Studien, Berichten und Vorträgen vorliegenden Wissens hat zu einer Identifizierung verschiedener Evaluationskonzepte und -modelle geführt. [Stockmann 2004, 2006, 2007, zit. in Filsinger 2008: 62 ff] Entsprechend der unter Punkt 1 beschriebenen Spannweite von Evaluationszwecken unterscheiden z. B. Beywl u. a. mehr als zehn Ansätze von Evaluationen, die in der Literatur reflektiert werden. [Beywl et al. 2004]

Für Programmevaluationen sind vor allem drei Modelle relevant:

- *Programmzielgesteuerte Evaluation* als der klassische Evaluationsansatz, der anhand der erreichten Resultate prüfen soll, ob die Ziele eines Programms erreicht wurden. Aus dem Grad der Zielerreichung wird auf die Verwendbarkeit des Programms geschlossen. [Beywl et al. 2004: 225]
- *Wirkungsorientierte Evaluation* [Beywl et al. 2004: 227 ff] [2] bezieht sich nicht nur auf die direkten Effekte einer Intervention auf eine Zielgruppe, sondern vor allem auch auf mittel- und langfristige Wirkungen im Hinblick auf die Programmziele und ist noch anspruchsvoller und voraussetzungsvoller. [Filsinger 2008: 63]
- *Cluster-Evaluation* sieht die Variationen in der Programmimplementation unter heterogenen Bedingungen (relativ autonome Projekte an verschiedenen Standorten mit eigener Strategien) nicht als hinderlich für die Evaluation an, ist aber nur in Fällen einsetzbar, wo keine Aussagen zu einzelnen Projekten gefordert sind, da diese nur in anonymisierter Form für die Gesamtbewertung genutzt werden. [Sanders 1997 zit. in Haubrich 2001]

Auch gibt es zahlreiche grundsätzliche Empfehlungen zur Planung und Durchführung von Evaluationsprojekten (Planung, Management, Organisation, Auswertung, Berichtlegung). [Wottawa 1998; König 2004] Der

Schwerpunkt der Modellbeschreibung liegt aber auf der Analysemethodik, die Bewertungsmethodik ist zumeist knapp oder nicht beleuchtet. Selbst Glossare zum Thema Evaluation enthalten selten Begriffe im Zusammenhang mit dem Bewertungsproblem. [z. B. univati-on 2004] Eine Ausnahme bildet die von Stockmann am Beispiel der Evaluation der Qualitätsentwicklung von Organisationen entwickelte, praktikabel nachnutzbare Evaluationskonzeption, die neben einem Muster-Evaluationsleitfaden auch ein Bearbeitungs- und Bewertungsverfahren umfasst. [Stockmann 2006]

Kernaufgaben von Bewertungsprozessen sind die Erarbeitung einer geeigneten Bewertungsmethodik und deren Anwendung auf die beschafften Daten und Informationen. Die Bewertungsmethodik umfasst die Formulierung von Bewertungskriterien, die Entwicklung geeigneter Indikatoren und die Erarbeitung einer Bewertungsmatrix, auf deren Grundlage die Ausprägungen der Indikatoren (zu beschaffende Daten und Informationen) auf geeignete Weise zu einer Gesamteinschätzung verdichtet werden. Hierfür gibt es keine allgemeingültigen Regeln. »Die Entwicklung von Indikatoren ... stellt bei jeder Evaluation einen kreativen Akt dar, da die Besonderheiten eines Programms und seines Politikfeldes sowie die Aufgabenstellung der Evaluation berücksichtigt werden müssen.« [Stockmann 2006: 204] Gleiches gilt für die Festlegung von Bewertungsskalen und eventuellen Gewichtungsfaktoren. Hinzu kommt, dass zumeist nicht auf statistische Auswertungs- und Bewertungsverfahren für quantitative Daten aus sozialwissenschaftlichen Experimenten zurückgegriffen werden kann, da das Bewerten mittels quantitativer Daten bei Programmevaluationen nur begrenzt möglich ist. Vielmehr fällt zumeist eine Überfülle qualitativer Daten und Informationen an. Auch sind die Programmziele häufig (nur) qualitativer Natur. In Lehrbüchern zur empirischen Sozialforschung sind solche methodologischen Sonderprobleme angewandter Forschung jedoch weitgehend ausgespart. [Kromrey o.J.]

Die Herausforderungen in Bewertungsprozessen liegen also weniger in der Methode der Informationsbeschaffung – hier kann auf das erprobte Instrumentarium der empirischen Sozialforschung zurückgegriffen werden –, als vielmehr im Konzept für die Informationsbewertung. Gleichwohl ist die Informationsbeschaffung immer ein zu realisierender Spagat zwischen Notwendigem und Machbarem und erfordert eine hohe Methodenkompetenz.

Mit dem vorliegenden Beitrag wird erstens das Ziel verfolgt, an praktischen Beispielen aus Evaluationen von Programmen mit qualitativer Zielsetzung für unterschiedliche Bewertungsprobleme methodische Lösungsansätze vorzustellen. Zweitens soll aufgezeigt werden, inwiefern für Programmevaluationen vergleichbarer Komplexität aber unterschiedlichem Evaluationsauftrag eine Operationalisierung im methodischen Vorgehen möglich ist und wo die Ansatzpunkte für weitere Forschungsarbeiten liegen.

Die dargestellten Erfahrungen basieren auf einer bereits abgeschlossenen und einer noch laufenden Evaluation, bei der jedoch die Bewertungsmethodik bereits für ein Teilprojekt (Versuch) erfolgreich expliziert wurde.

### 3 Operationalisierung der Bewertung an zwei praktischen Beispielen

#### 3.1 Evaluation eines Förderprogramms zur Berufsorientierung

Hauptziel der als Gemeinschaftsaktion des Ministeriums für Arbeit, Soziales, Gesundheit und Frauen (MASGF) mit dem Ministerium für Bildung, Jugend und Sport (MBS) und der Regionaldirektion Berlin-Brandenburg der Bundesagentur für Arbeit im Land Brandenburg durchgeführten 18. INNOPUNKT-Kampagne war die Systematisierung und der Ausbau des Berufsfindungs- und -wahlprozesses von Schülerinnen und Schülern. Dazu wurden sieben Teilziele definiert, wie z. B. »mehr und passgenauere Praxislernzeiten für SchülerInnen«, »Berücksichtigung regionaler Bedarfe bei der Entwicklung von Branchenkompetenzfeldern in Konzepten und Maßnahmen« oder »gestärkte Fachkompetenz und Kooperationsbereitschaft durch neue Fortbildungskonzepte«.

Sieben Projektträger stellten sich im Zeitraum von Februar 2007 bis Ende 2008 dieser anspruchsvollen Zielsetzung, entwickelten neue Konzepte und setzten diese um. Zur Erreichung des Hauptziels konnten die Träger eines oder mehrere Teilziele wählen.

Der Auftrag für die formative Evaluation als Element der Qualitätssicherung sah zwei Aufgaben vor: Neben der wissenschaftlich beratenden Begleitung der Modellprojekte mit dem Ziel, Lernprozesse und entsprechendes Handeln bei den Trägern auszulösen, oblag der Evaluation die Bewertung und Dokumentation der erreichten Ergebnisse und Wirkungen, die Ableitung von Handlungsempfehlungen hinsichtlich der flächendeckenden Einführung des Praxislernens in

Brandenburg sowie der erfolgreichen Verbindung von Schule und Wirtschaft.

Bewertungsseitig war die Evaluation damit gefordert, die Gesamtzielerreichung eines mit mehreren Teilzielen unteretzten und von mehreren Trägern auf unterschiedliche Weise realisierten Programms einzuschätzen, wobei sowohl das Hauptziel als auch die Teilziele komplexer und ganz überwiegend qualitativer Natur waren.

Hinzu kamen weitere Herausforderungen:

- Der für Programmevaluationen typische hochkomplexe Untersuchungsgegenstand hatte einen hohen Analyse- und Bewertungsaufwand zur Folge. Die Projekte wiesen eine große Bandbreite im methodischen Vorgehen, im Umfang erprobter Instrumente zur Arbeitswelt- und Berufsorientierung wie auch in der Projektorganisation auf. Das Leistungsspektrum reichte von der Konzeptentwicklung über die organisatorische Unterstützung, die Realisierung einzelner Bausteine des Praxislernens, die Qualifizierung relevanter Akteure und das Einbringen neuer Produkte bis zur individuellen Beratung von Schülerinnen und Schülern. Für nahezu alle möglichen Ansatzpunkte, die die Berufsorientierung und den Übergang in die Berufsausbildung behindern, wurden in der Kampagne Problemlösungen behandelt und dabei fast die gesamte in Deutschland genutzte Palette an Instrumenten, Methoden und Akteurskooperationen unter Brandenburger Bedingungen angewendet.
- Um die Hauptzielerreichung der Kampagne umfassend bewerten zu können, musste der in seiner Ausgestaltung weitgehend unregelmäßige komplexe Begriff der »Systematisierung von Berufsorientierung« präzisiert werden.
- Der Bewertungsprozess wurde dadurch erschwert, dass nicht alle Projekte alle Teilziele gleichermaßen im Fokus hatten, da sie eine Auswahl aus den vorgegebenen sieben Teilzielen treffen konnten. Zudem waren die Beiträge selbst von unterschiedlichem Gewicht und einige Teilziele bündelten mehrere Zielfacetten, die einen unterschiedlichen Zielerreichungsgrad haben konnten.

Zur Erfüllung des Evaluationsauftrags wurde eine Bewertungsmethodik entwickelt, die diesen Herausforderungen wie folgt gerecht wurde:

#### Begriffsklärung und Operationalisierung des Hauptziels

Systematisierung von Berufsorientierung kann sowohl ergebnisbezogen (im Sinne einer systematischen Befä-

higung von Schülerinnen und Schülern zur Berufsorientierung) als auch prozessbezogen (im Sinne eines systematischen Vorgehens bei der Einführung von Berufsorientierung an Schulen) betrachtet werden. Da die Kampagne vordergründig auf die ergebnisorientierte Betrachtung fokussierte, entwickelte die Evaluation hierfür einen entsprechenden Bewertungsrahmen aus Dimensionen, Merkmalsausprägungen und Indikatoren (vgl. Tabelle 1).

**Aufhellung des Bezugs der Teilziele zum Hauptziel**

Entsprechend des entwickelten Verständnisses von unterschiedlichen Dimensionen der Systematisierung von Berufsorientierung (als Hauptziel) wurde eine Zuordnung der sieben Teilziele zu diesen Dimensionen vorgenommen (vgl. Abb. 1). Dabei zeigte sich, dass sich zu drei Dimensionen jeweils drei Teilziele zuordnen ließen, zu einer jedoch nur eines.

**Entwicklung einer Bewertungsmatrix zur Teilzielbewertung**

Da die oben beschriebene unterschiedliche Orientierung der Projekte auf einzelne Teilziele der Kampagne eine vollständig konsistente Bewertung ausschloss, wurden die Bewertungskriterien »regionale Reichweite«, »Anzahl und Qualität der Beiträge« sowie »Ziel-

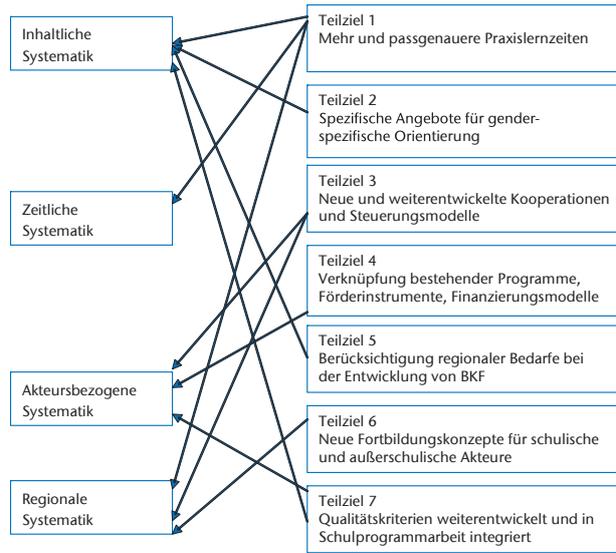


Abb. 1: Teilziele der Kampagne und Dimensionen von Systematisierung

genauigkeit« gewählt und deren Ausprägungen für eine Bewertung wie in Tabelle 2 dargestellt festgelegt.

**Strukturierte Ergebnissammlung und -dokumentation**

Zur Gewinnung von Daten und Informationen über die erzielten Ergebnisse der Projekte nutzte die Evaluation einen Mix aus qualitativen und quantitativen Methoden, wie z. B. leitfadensbasierte Vor-Ort-Gespräche und

Dimensionen	Merkmale	Indikatoren für systematische Merkmalsausprägung
Inhaltliche Systematik	Ziele und Teilziele	Vorgehen und Instrumente auf Zielsystem ausgerichtet
	Bezug zu Allgemeinbildung/Schultypen	Verzahnung externer BO-Angebote mit Lehrplänen
	Bezug zu Berufsfrüh- und Studienorientierung	gesicherte Anschlüsse
	Bezug zu dynamischer Berufs- und Ausbildungswelt	den Veränderungen angepasste Inhalte
	Vermittelbarkeit der Inhalte	Nutzung geeigneter pädagogischer Konzepte
	Qualität	eingeführte und beachtete Qualitätsstandards an allen Praxislernorten
Zeitliche Systematik	Prozesscharakter	altersgerechte Angebote aufeinander aufbauende Angebote in Jahrgangsstufen
	individueller Fortschritt von Schülerinnen und Schülern	zunehmende Individualisierung der Angebote
	Akteursbezogene Systematik	Vielfalt pflichtiger und freiwilliger Akteure
verlässliche Beteiligung		
Steuerung		
unterschiedliche Interessenlagen		politische Steuerung (bezüglich Schulbezug)
unterschiedliche Ressourcen	Abstimmung, Qualifizierung/Weiterbildung, ggf. Ressourcen zuführen	
unterschiedliche Lernorttypen	Verzahnung schulischer und außerschulischer Lernorte	
Regionale Systematik	unterschiedliche Schul-, Wirtschafts-, Branchen-, Infra- und Akteursstruktur	regionale Modell der Akteursvernetzung und Steuerung unter Beachtung von Qualitätsstandards
	unterschiedliche Zahl und Spektrum an Ausbildungs- und Studienmöglichkeiten	regionale Modelle außerschulische Anteile
	Quantität	dauerhaftes und flächendeckendes Angebot für alle SchülerInnen

Tab. 1: Dimensionen, Merkmalsausprägungen und Indikatoren systematischer Berufsorientierung

Beitrag zur Zielerreichung ist ...			kein Beitrag zur Zielerreichung
groß	mittel	gering	Zielerreichung
<ul style="list-style-type: none"> <li>in mindestens vier Projektregionen wurden Beiträge zur Zielerfüllung erbracht</li> <li>es konnten viele Belege für deutliche Veränderungen aufgeführt werden</li> <li>die Belege beziehen sich auf alle Zielaspekte</li> </ul>	<ul style="list-style-type: none"> <li>in zwei oder drei Projektregionen wurden Beiträge zur Zielerfüllung erbracht</li> <li>es gibt einige Belege für deutliche Veränderungen oder viele Belege für geringe Veränderungen</li> <li>die Belege beziehen sich auf alle Zielaspekte</li> </ul>	<ul style="list-style-type: none"> <li>nur in einer Projektregion wurden Beiträge zur Zielerfüllung erbracht</li> <li>es gibt nur vereinzelte Belege für geringe Veränderungen</li> <li>die Belege beziehen sich nicht auf alle Zielaspekte</li> </ul>	<ul style="list-style-type: none"> <li>in keiner Projektregion Belege für Veränderungen</li> </ul>

Tab. 2: Bewertungsmatrix für Beiträge zur Zielerreichung

Telefoninterviews mit allen Projektleitungen, aktive Teilnahme an Erfahrungsaustauschen der Träger, Erfassungsbögen zur Selbsteinschätzung der Träger, schriftliche Befragung aller teilnehmenden Modellschulen, und wertete die halbjährlichen Sachberichte der Projektträger aus. Sämtliche Projektergebnisse wurden teilzielbezogen gruppiert und dokumentiert.

#### Anwendung der Bewertungsmatrix auf alle Teilziele

Die Anwendung der in Tabelle 2 dargestellten Bewertungsmatrix auf die Summe der Ergebnisse aller sieben Teilprojekte führte zu der in Tabelle 3 dargestellten Bewertung des Beitrages der Kampagne zur Erreichung der gestellten Teilziele.

#### Synthetisierende Bewertung der Hauptzielerreichung

Die in Abbildung 1 dargestellte Zuordnung von Teilzielen zu den unterschiedlichen Systematisierungsdimensionen von Berufsorientierung ermöglichte nunmehr eine Bewertung der Beiträge der Kampagne differenziert

Teilziel	Beitrag zur Teilzielerreichung		
	groß	mittel	gering
mehr und passgenauere Praxislernzeiten	x		
Berücksichtigung genderspezifischer Orientierungen		x	
neue/weiterentwickelte Kooperationen/Steuerungsmodelle	x		
Verknüpfung Förderinstrumente, Finanzierungsmodelle		x	
Berücksichtigung regionaler Bedarfe	x		
neue Fortbildungskonzepte		x	
Integration Qualitätskriterien in Schulprogrammarbeit			x

Tab. 3: Übersicht zur Bewertung der Teilzielerreichung

nach der inhaltlichen, zeitlichen, akteursbezogenen und regionalen Dimension. So konnte z. B. zur regionalen Systematisierung insofern ein Beitrag bescheinigt werden, als alle Projekte bestehende Akteursnetzungen weiter profiliert und verschiedene Modelle der Steuerung erprobt und etabliert haben, die weitere Umsetzung der Qualitätskriterien für Praxislernen von vielen Projektschulen in Angriff genommen und die Notwendigkeit von Qualitätsstandards für die Berufsorientierung auch an außerschulischen Lernorten in der Netzwerkarbeit thematisiert wurde.

### 3.2 Evaluation von Maßnahmen zur Verwaltungsmodernisierung im Rahmen eines Standarderprobungsgesetzes

Gemäß des ersten Brandenburgischen Bürokratieabbaugesetzes erproben seit 2007 zahlreiche Kommunen auf Antrag für einen begrenzten Zeitraum, von landesrechtlichen Vorgaben abzuweichen und alternative oder selbstständige Lösungen zu erproben. Bürger und Unternehmen sollen von einem besseren Service beispielsweise durch eine kürzere Verfahrensdauer profitieren. Kann der Erfolg der Erprobungen in den Versuchskommunen nachgewiesen werden, ist deren landesweite Umsetzung vorgesehen.

Der Evaluationsauftrag ist auf insgesamt 28 Versuche mit zehn unterschiedlichen Versuchsgegenständen aus sieben Rechtsbereichen (Schulgesetz, Bauordnung, Wassergesetz, Straßenverkehrsrechtszuständigkeitsverordnung, Landeshaushaltsordnung, Naturschutzgesetz, Gesundheitsdienstgesetz) bezogen. Sie werden von 21 Versuchskommunen durchgeführt, darunter drei Landkreise, eine kreisfreie Stadt, 14 Gemeinden und drei Ämter, und haben verschiedene Laufzeiten. So können z. B. sechs Gemeinden über die Anträge von Eltern auf einen Schulbezirkswechsel ihrer Kinder ohne die Beteiligung des staatlichen Schulamtes entscheiden.

In einem weiteren Versuch wird acht Städten, einer Gemeinde und einem Amt die Zuständigkeit für bestimmte Vorschriften der StVO übertragen. Ein Landkreis erprobt z. B. den Verzicht der Regelung der Mindesthöhe von Aufenthaltsräumen und der Mindestgröße der Belichtungsöffnungen.

Die Evaluation hat den Auftrag, auf der Grundlage einer fundierten Analyse zu bewerten, ob die Versuche die Ziele des Standarderprobungsgesetzes erreichen und zu begründen, ob eine landesweite Umsetzung durch Rechtsänderungen zu empfehlen ist.

Bewertungsseitig ist die Evaluation damit gefordert, die Zielerreichung einer Vielzahl von Versuchen mit unterschiedlichem Erprobungsgegenstand einzuschätzen, die jeweils von einem Akteur oder mehreren Akteuren durchgeführt werden, und darüber hinaus eine begründete Übertragbarkeitsentscheidung für jeden Versuch zu treffen. Dies bedeutet in mehrfacher Hinsicht eine Herausforderung:

- Die Evaluation muss sowohl Kriterien zur Erfolgsbewertung der Versuche als auch Kriterien zur Bewertung der Übertragbarkeit (Empfehlung für landesweite Umsetzung) entwickeln, die gleichermaßen für völlig unterschiedliche Rechtsbereiche und Erprobungsgegenstände gelten.
- Die Vielzahl der Versuche und die Verschiedenartigkeit und Komplexität der Versuchsgegenstände erfordern einen hohen Aufwand für die Kontextanalyse, Daten- und Informationsbeschaffung.
- Es ist zu entscheiden, ob und wie über die Ziele des Standarderprobungsgesetzes hinausgehende Wirkungen der Versuche in die Bewertung einzubeziehen sind.
- Die Unterschiedlichkeit der Versuchsgegenstände und der einzubeziehenden Akteure (Polizeidienststellen, Bürgervereine, Schulämter, Landräte, Schulleitungen, Landesrechnungshof, Straßenbaubehörden, Unternehmen, Umweltamt, ...) stellt hohe Anforderungen an die Erhebung der benötigten Messgrößen und qualitativen Merkmale.

Zur Erfüllung des Evaluationsauftrags wurde eine mehrstufige Bewertungsmethodik mit Bewertungskriterien entwickelt, die für alle Versuch gleichermaßen gültig sind:

#### **Kriterien zur Erfolgsbewertung der Versuche**

Da die Erfolgsbewertung der Versuche vor dem Hintergrund einer landesweiten Übertragung getroffen werden muss, reicht es nicht aus, sie allein an der Errei-

chung der Ziele des Standarderprobungsgesetzes festzumachen. Vielmehr macht sich ein breiteres Verständnis von Versuchserfolg erforderlich.

Ein Versuch wird dann als erfolgreich bewertet, wenn nachfolgende drei Bedingungen gegeben sind:

1. Erfüllung mindestens eines Ziels des 1. BbgStEG [3]
  - Wirtschaftliche Entwicklung gefördert (unternehmerisches Handeln bzw. Existenzgründungen erleichtert)
  - Verwaltungsverfahren beschleunigt
  - Kosten für Unternehmen, Bürger und Verwaltung gesenkt
2. Qualitätserhaltung/-verbesserung
  - Ergebnisgüte des Verwaltungshandelns
  - Zufriedenheit von Normadressaten und weiteren Betroffenen
3. Verhältnismäßigkeit potenzieller Folgewirkungen
 

Hierbei werden nicht nur Folgewirkungen betrachtet, die unmittelbar aus der veränderten Gesetzeshandhabung resultieren, sondern auch der erforderliche Beratungs- und Schulungsaufwand in den Kommunen und bei weiteren involvierten Akteuren sowie ggf. der veränderte Aufwand für die zuständigen Ressorts.

Sofern die Versuchskommunen über die Ziele des BbgStEG hinausgehende Wirkungen erzielen, werden diese in die Bewertung von Qualität und Folgewirkungen einbezogen.

#### **Kriterien zur Bewertung der Übertragbarkeit (landesweite Umsetzung)**

Eine Standarderprobung wird dann zur landesweiten Umsetzung empfohlen, wenn nachfolgende fünf Bedingungen gegeben sind:

- Der Erfolg der/des Versuche/s konnte nachgewiesen werden (siehe oben).
- Die Verhältnismäßigkeit aller zu schaffenden Voraussetzungen für den Erfolg der Standarderprobung ist gegeben (Schulung, Controlling, ...).
- Sofern eine Erprobung nicht unabhängig vom Raumtyp (engerer Verflechtungsraum, äußerer Entwicklungsraum) und vom Typ der Versuchskommune (Landkreis/kreisfreie Stadt, Amt, Gemeind, jeweils unter Berücksichtigung der Einwohnerzahl) ist, muss der Erfolg für alle Raum- bzw. Kommumentypen nachgewiesen sein.
- In den Fällen, in denen die Erprobung nur auf dem Versuch einer einzigen Kommune basiert, konnte deren Repräsentativität belegt werden.

- Die Standarderprobung muss in ausreichender Breite und Tiefe erfolgt sein (Fallzahlen, Erprobungsinhalte).

Für nicht erfolgreiche Versuche werden die Gründe hinterfragt, da andere Versuchsbedingungen möglicherweise ein anderes Ergebnis ermöglicht hätten. In jedem Falle lassen sich die Bedingungen für eine Übertragbarkeit differenziert herausarbeiten. Es wird auch berücksichtigt, ob eine erprobte gesetzliche Regelung bereits in anderen Bundesländern Standard ist bzw. auf andere Gegenstände erfolgreich angewendet werden kann.

### Gegenstandsspezifische Indikatoren zur Erfolgsbewertung

Die Anwendbarkeit der Erfolgskriterien für Standarderprobungen erfordert eine Untersetzung mit spezifischen Indikatoren zur Erfolgsbewertung für jeden der zehn Antragsgegenstände. Diese werden entwickelt unter Berücksichtigung der spezifischen Rahmenbedingungen in den Rechtsbereichen und den von den Antragstellern formulierten Erprobungszielen, die mit den Zielen des Standarderprobungsgesetzes gespiegelt wurden. Das auf diese Weise für einen Versuch zum Schulgesetz (Schulträger als stimmberechtigtes Mitglied der Schulkonferenz) entwickelte Indikatorenset ist beispielhaft in Tabelle 4 aufgeführt.

### Versuchsspezifischer Methodenmix zur Ermittlung der Indikatorenausprägung

Die Ermittlung der Ausprägung der Indikatoren für die Erfolgsfaktoren erfolgt mit Hilfe von Messgrößen und mittels qualitativer Merkmale. Die entsprechenden Informationen werden mit einem Methodenmix aus

Expertengesprächen, Berichts- und Bewertungsbögen (Versuchskommunen, Normadressaten, Dritte), Auswertung von Berichten der Versuchskommune/n an das Ressort und weiterem Schriftmaterial empirisch ermittelt. Dieser Methodenmix wurde aufgrund der unterschiedlichen Erprobungsgegenstände jeweils versuchskonkret und auf der Grundlage einer umfangreichen Kontextanalyse ausspezifiziert. Dabei musste ein Kompromiss zwischen Wünschenswertem und tatsächlich Machbarem gefunden werden.

Die Praktikabilität der Bewertungsmethodik konnte am Beispiel des o. g. Versuchs zum Schulgesetz erfolgreich unter Beweis gestellt werden. Wesentliche Erfolgsbedingungen für die Evaluation waren die frühzeitige Abstimmung der Indikatorenauswahl mit den Versuchsakteuren und die Schaffung durchgängiger Transparenz im Bewertungsverfahren. Weitere Erkenntnisse können aus der bevorstehenden Anwendung der Evaluationsmethodik auf die übrigen Versuche aus verschiedenen Rechtsbereichen erwartet werden.

## 4 Fazit

Die Komplexität von Programmevaluationen stellt insbesondere bezüglich der vorzunehmenden Bewertungen hohe Herausforderungen an das Evaluationsteam. Für zwei unterschiedliche Evaluationsaufträge konnte ein praktikables methodisches Vorgehen aufgezeigt werden. In beiden Fällen war der Evaluationsgegenstand dadurch gekennzeichnet, dass eine größere Anzahl von Projekten/Versuchen innovative Vorgehensweisen erprobte. Der Evaluationszweck wies jedoch Unterschiede auf. Während im ersten Fall der Fokus

Erfolgskriterien	Indikatoren
Verwaltungsverfahren beschleunigt	Wegfall von Postverkehr und Telefonaten
	beschleunigte Abstimmungsprozesse
	weniger Rückfragen zum Ergebnis des Verwaltungshandelns
Kosten für Unternehmen, Bürger und Verwaltung gesenkt	Einsparung von Verwaltungskosten
Qualitätserhaltung/-verbesserung	keine Erschwernisse in Organisation und Durchführung der Schulkonferenz
	keine Erschwernisse bei Aufgabenerfüllung der Schulkonferenz
	mindestens gleiche finanzielle und personelle Sicherstellung des Schulbetriebs durch den Schulträger
	Rangerhalt pädagogischer Themen
Verhältnismäßigkeit potenzieller Folgewirkungen	Qualität der Wahrnehmung anderer kommunaler Aufgaben des Schulträgers
	Mehrbelastung von Schulgremien und Schulleitungen
	Einflussnahme auf Bestellung von SchulleiterInnen durch Schulträger

Tab. 4: Erfolgskriterien und Indikatoren am Beispiel des Versuchs »Schulkonferenz«

auf der Zielerreichung des Gesamtprogramms unter Berücksichtigung von Teilzielen lag, ging es im zweiten Fall darum, sowohl den Erfolg der einzelnen Versuche als auch die Sinnfälligkeit ihrer landesweiten Übertragung zu bewerten. Damit war in beiden Evaluationen ein zweistufiges Bewertungsproblem zu lösen und das grundsätzliche Vorgehen dementsprechend gleich (vgl. Abbildung 2).

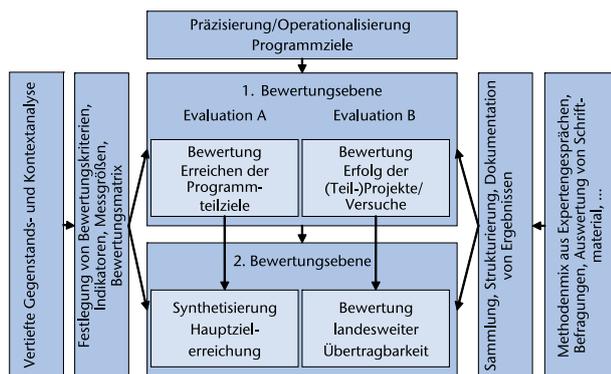


Abb. 2: Zweistufiges Vorgehensmodell für Programmevaluationen mit unterschiedlichen Bewertungszielen

Die in Kapitel 3 vorgestellte programmspezifische Umsetzung erforderte sodann konkrete Bewertungsmethodiken für vier verschiedene Bewertungsaufgaben auf zwei Bewertungsebenen (Teilzielerreichung, Versuchserfolgswertung, Gesamtzielerreichung, Übertragbarkeit). Sie schloss die Erarbeitung geeigneter Bewertungskriterien, Indikatoren und ggf. Messgrößen sowie einer Bewertungsmatrix ein und macht den hohen Stellenwert einer intensiven Gegenstands- und Kontextanalyse deutlich.

Kann überwiegend nur auf qualitative Bewertungsindikatoren zurückgegriffen werden, kommt den empirischen Erhebungen im Umfeld der Projekte eine besondere Bedeutung zu, um möglichst viele Sichtweisen betroffener Akteure in die letztlich durch die Evaluation zu treffende Bewertung der Ausprägung von Erfolgsindikatoren einfließen zu lassen und so zu möglichst objektiven Einschätzungen zu gelangen.

Ohne die Bereitstellung der erforderlichen Daten und Informationen durch die agierenden Projektakteure ist ein erfolgreiches Evaluationsvorhaben nicht vorstellbar. Deren Bereitschaft zur Mitarbeit hängt wesentlich davon ab, wie es der Evaluation gelingt, den Evaluationsprozess transparent zu gestalten.

Eine weitergehende Operationalisierung von Bewertungsprozessen in Programmevaluationen ist in zwei Richtungen vorstellbar: Erstens könnte die Ermittlung

geeigneter Bewertungsindikatoren auf der Grundlage einer weiteren theoretischen Durchdringung ihres Zusammenhangs zur Art des Untersuchungsgegenstands unterstützt werden. Zweitens wäre eine methodische Untersetzung der Erarbeitung von Bewertungsmatrizen wünschenswert, wofür der Zusammenhang zwischen Evaluationszweck und Bewertungsproblem tiefer untersucht werden müsste.

## Anmerkungen

- [1] Zu unterscheiden sind:
  - output (Leistungserbringung)
  - outcome (Wirkungen bei den verschiedenen Adressaten)
  - impact (gesellschaftliche Auswirkungen)
- [2] auch als experimentaldesign- bzw. quasi-experimentaldesign-gesteuertes Evaluationsmodell bezeichnet
- [3] vgl. Erstes Brandenburgisches Bürokratieabbaugesetz – 1. BbgBAG vom 28. Juni 2006, Artikel 1 (Gesetz zur Erprobung der Abweichung von landesrechtlichen Standards in Kommunen sowie von landesrechtlichen Zuständigkeitsverlagerungen – BbgStEG), § 1 (Ziel)

## Literatur

- Beywl, Wolfgang; Speer, Sandra; Kehr, Jochen (2004): Wirkungsorientierte Evaluation im Rahmen der Armut- und Reichtumsberichterstattung. Perspektivstudie im Auftrag und hrsg. von BMGS, Bonn.
- DeGEval – Gesellschaft für Evaluation e. V. (2008): Standards für Evaluation.
- Filsinger, Dieter (2008): Bedingungen erfolgreicher Integration – Integrationsmonitoring und Evaluation. WISO Diskurs, hrsg. von Friedrich Ebert Stiftung, Bonn.
- Gerlach, Ferdinand M.; Beyer, Martin; Szecsenyi, Raspe, Heiner (2003): Evaluation von Disease-Management-Programmen – Aktuelle Defizite, Anforderungen, Methoden. Zeitschrift für ärztliche Fortbildung und Qualität im Gesundheitswesen (ZaeFQ) (2003) 97, S. 495-501.
- Große, Ulla; Brandt, Martina (2004): Relevanz der Akteursebene in Netzwerken für die Evaluation – Problemaufriss aus einem Projektvergleich. In: Voß, R. (Hrsg.): Evaluation von Netzwerken und Kooperationsmanagement. Institut für Regionale Innovationsforschung, Institutsschriften Nr. 25, 13-19.
- Haubrich, Karin (2001): Cluster-Evaluation – ein Modell für einen dornigen Evaluationsgegenstand? DeGEval-Jahrestagung, Block 16 »Modelle der Evaluation«. Speyer, 04.10.2001.
- Hupfer, Barbara (2007): Wirkungsorientierte Programmevaluation. Schriftenreihe des BiBB, Heft 86.
- König, Joachim (2004): Selbstevaluation. Der Weg zu mehr Nachhaltigkeit, hrsg. von LASA Brandenburg GmbH, Potsdam.
- Kromrey, Helmut O. J.: Evaluation im Spannungsfeld von Wissenschaftlichkeit und Verwertbarkeit. Unter: www.hkromrey.de, eingesehen am 31.08.2009.
- Sanders, James R. (1997): Cluster Evaluation. In: Chelimsky, Eleanor; Shadish, William (Eds.): Evaluation for the 21st Century. A Handbook. Thousand Oaks, London, New Dehli, 396-404.

- Simon, Dagmar; Stucke, Andreas (2003a): Rahmenkonzept für die Evaluation des Förderschwerpunkts »Sozial-ökologische Forschung« des Bundesministeriums für Bildung und Forschung, hrsg. von EvaConsult, Berlin.
- Simon, Dagmar; Stucke, Andreas (2003b): Muster-Leitfäden für die Evaluation des Förderschwerpunkts »Sozial- ökologische Forschung« des Bundesministeriums für Bildung und Forschung, hrsg. von EvaConsult, Berlin/Köln.
- Stockmann, Reinhard (2000): Evaluationsforschung. Grundlagen und ausgewählte Forschungsfelder. Leske + Budrich, Opladen.
- Stockmann, Reinhard (2004): Evaluationsforschung – Ansatz und Methoden. EvaNet-Positionen 5/2004, Auszug aus Stockmann (2004): Evaluationsforschung. Grundlagen und ausgewählte Forschungsfelder. 2. Auflage, Leske + Budrich, Opladen.
- Stockmann, Reinhard (2006): Evaluation und Qualitätsentwicklung. Waxmann Verlag, Münster.
- Stockmann, Reinhard (2007): Handbuch zur Evaluation. Eine praktische Handlungsanleitung. Waxmann Verlag, Münster/New York/Berlin/München.
- univation GmbH (2004): Das A-B-C der wirkungsorientierten Evaluation, Köln.
- Wottawa, Heinrich; Thierau, Heike (1998): Lehrbuch Evaluation. Verlag Hans Huber, Bern/ Göttingen/Toronto/Seattle.

## Autorinnen

### **Dr. oec. Martina Brandt**

TH Wildau [FH]  
 Fachbereich Betriebswirtschaft/Wirtschaftsinformatik  
 Forschungsgruppe »Innovations- und Regionalforschung«  
 Tel. +49 3375 508-512  
 martina.brandt@tfh-wildau.de

### **Dr. rer. nat. Ulla Große**

TH Wildau [FH]  
 Fachbereich Betriebswirtschaft/Wirtschaftsinformatik  
 Forschungsgruppe »Innovations- und Regionalforschung«  
 Tel. +49 3375 508-513  
 ulla.grosse@tfh-wildau.de