

WISSENSCHAFTLICHE DISKUSSIONSPAPIERE

Heft 86

Barbara Hupfer

Wirkungsorientierte Programmevaluation

Eine Synopse von Ansätzen
und Verfahren einschlägiger Institutionen
in Deutschland

Schriftenreihe
des Bundesinstituts
für Berufsbildung
Bonn

**Bundesinstitut
für Berufsbildung**

BiBB ▶

- ▶ Forschen
- ▶ Beraten
- ▶ Zukunft gestalten

Die WISSENSCHAFTLICHEN DISKUSSIONSPAPIERE des Bundesinstituts für Berufsbildung (BIBB) werden durch den Präsidenten herausgegeben. Sie erscheinen als Namensbeiträge ihrer Verfasser und geben deren Meinung und nicht unbedingt die des Herausgebers wieder. Sie sind urheberrechtlich geschützt. Ihre Veröffentlichung dient der Diskussion mit der Fachöffentlichkeit.



Der Inhalt dieses Werkes steht unter einer Creative Commons Lizenz (Lizenztyp: Namensnennung – Keine kommerzielle Nutzung – Keine Bearbeitung – 3.0 Deutschland).

Das Werk wird durch das Urheberrecht und/oder einschlägige Gesetze geschützt. Jede Nutzung, die durch diese Lizenz oder Urheberrecht nicht ausdrücklich gestattet ist, ist untersagt.

Weitere Informationen finden Sie im Internet auf unserer Creative Commons-Infoseite <http://www.bibb.de/cc-lizenz>

Vertriebsadresse:

Bundesinstitut für Berufsbildung
A 1.2 Kommunikation - VÖ
53142 Bonn

Bestell-Nr.: 14.086

Copyright 2007 by Bundesinstitut für Berufsbildung, Bonn
Herausgeber:
Bundesinstitut für Berufsbildung, Bonn
Internet: www.bibb.de
E-Mail: zentrale@bibb.de
Umschlaggestaltung: Hoch Drei Berlin
Herstellung: Bundesinstitut für Berufsbildung, Bonn
Druck: Bundesinstitut für Berufsbildung, Bonn
Printed in Germany

ISBN 978-3-88555-811-8

Diese Netzpublikation wurde bei Der Deutschen Bibliothek angemeldet und archiviert.
URN: [urn:nbn:de:0035-0212-5](http://nbn-resolving.org/urn:nbn:de:0035-0212-5)

Inhaltsverzeichnis

Vorwort	5
Einleitung	6
1. Das Evaluationssystem der Gesellschaft für Technische Zusammenarbeit (GTZ)	7
1.1 Ziele und Qualitätsmerkmale von Evaluationen in der GTZ	7
1.2 Selbstevaluierung und Fremdevaluierung	8
1.3 Verortung von Selbst- und Fremdevaluierung im Kontrollsystem der Entwicklungszusammenarbeit	11
1.4 Reformen des GTZ-Evaluierungssystems	14
1.5 Literatur	16
2. Der Evaluationsansatz des Centrums für Evaluation (CEval)	17
2.1 Aufgaben und Programmatik	17
2.2 Evaluationsansatz	18
2.2.1 Interne oder externe Evaluation?	21
2.2.2 Summative und formative Evaluation	22
2.2.3 Orientierung an den „Standards für Evaluation“ und wissenschaftstheoretische Position	23
2.2.4 Der Analyseleitfaden	24
2.3 Beispiel: Die Evaluation des Kommunikations- und Informationssystem Berufliche Bildung (KIBB)	24
2.4 Literatur	28
3. Die Gesellschaft für Evaluation (DeGEval) – Dialogforum zur Professionalisierung von Evaluationen	29
3.1 Aufgaben	29
3.2 Die „Standards für Evaluation“	30
3.3 Arbeitskreise	32
3.3.1 Beispiel: AK „Evaluation von Entwicklungspolitik“	32
3.3.2 Beispiel: AK „Berufliche und betriebliche Bildung“	33
4. Die Evaluation des Mehrebenenprogramms EQUAL (BMWA/BMAS)	34
4.1 Partizipativer Ansatz	34
4.2 Evaluation der Entwicklungspartnerschaften (EP)	35
4.2.1 Vielfalt von Evaluationsansätzen und -verfahren	35
4.2.2 EP-Evaluation vermittelt zwischen Teilprojekten und Programmebene	36
4.2.3 Interne und externe Evaluationsansätze	36
4.2.4 Vorgaben der Programmebene für die EP-Evaluation in der zweiten Förderrunde: Evaluationen müssen ausgeschrieben werden	38

4.3	EQUAL- Evaluation auf Programmebene	38
4.3.1	Anforderungen an das Evaluationsdesign.....	39
4.3.2	Überprüfung der Grundannahme: Netzwerkbildung führt zu Innovationsgewinnen?	39
4.3.3	Mix aus verschiedenen Ansätzen und Methoden.....	40
4.3.4	Kombination formativer/offener Evaluation und summativer Evaluation.....	41
4.3.5	EQUAL- Evaluationsdesign	41
4.4	Literatur	45
5.	Ansätze und Verfahren der Programmevaluation des Deutschen Jugendinstituts (DJI)	47
5.1	Allgemeine Gestaltungsgrundlagen	47
5.1.1	Externe und unabhängige wissenschaftliche Begleitung und Evaluation von Programmen.....	47
5.1.2	Interne und externe Evaluation, Selbst- und Fremdevaluation.....	48
5.1.3	Partizipativer Ansatz: Unterschiede in der Einbindung der Beteiligten und Betroffenen	50
5.1.4	Multidisziplinäres und plurales methodisches Selbstverständnis	50
5.1.5	Summative, formative und wissensgenerierende Evaluation	52
5.1.6	Verwendungstheoretische Probleme der Nutzenorientierung	53
5.1.7	Wirkungsevaluation: Nur bei ausgereiften Programmen sinnvoll?	54
5.1.8	Meta-Evaluation und Evaluations-Synthese.....	55
5.2	Beispiel Cluster-Analyse	55
5.2.1	Anwendungsbereiche: Komplexe und multizentrische Programme	55
5.2.2	Konzeptionelle Lösungen für konfligierende Evaluationsanforderungen.....	56
5.2.3	„Logic Models“ für die Analyse interner Wirkungszusammenhänge	57
5.3	„Intime“ Kenntnisse und Vertrauen fördern qualifizierte Evaluation	58
5.4	Literatur	59
6.	Programmevaluation des BMBF am Beispiel des Förderschwerpunkts „Sozial-ökologische Forschung“	61
6.1	Auftrag, Konzeptentwicklung, Beteiligte	61
6.1.1	Auftrag.....	61
6.1.2	Diskursiver Ansatz und formative Funktion	62
6.1.3	Eckpunkte der Programmevaluation	63
6.1.4	Leitkriterien: Grundlage für die Evaluationsleitfäden	64
6.1.5	Methode des „Expert Review“: Unabhängige Gruppe von Experten evaluieren mithilfe der Leitfäden	65
6.1.6	Interne Evaluation und Selbstevaluation	66
6.1.7	Exkurs: Evaluationsforschung und -praxis im Wissenschaftsbereich – Guidelines für die Programmevaluation	66
6.2	Literatur	68
7.	Zusammenfassender Vergleich	69

Vorwort

Dieses Diskussionspapier befasst sich mit Ansätzen und Verfahren wirkungsorientierter Programmevaluation ausgewählter, einschlägiger Institutionen in Deutschland. In den Blick genommen wurden:

Die Gesellschaft für Technische Zusammenarbeit (GTZ),

Das Centrum für Evaluation (CEval),

Die Deutsche Gesellschaft für Evaluation (DeGEval),

Das Bundesministerium für Wirtschaft und Arbeit (BMWA/BMAS) am Beispiel der Programmevaluation EQUAL,

Das Deutsche Jugendinstitut (DJI),

Das Bundesministerium für Bildung und Forschung (BMBF) am Beispiel der Evaluation des Förderschwerpunktes „Sozial-ökologische Forschung“.

Die Studie enthält eine Synopse der Gemeinsamkeiten und Unterschiede der untersuchten Evaluationsansätze und -verfahren im Hinblick auf die Verknüpfung von interner und externer Expertise beziehungsweise interner und externer Evaluation sowie Literaturübersichten zu den untersuchten Beispielen.

Als zunächst interne Recherche konzipiert, diente die Studie ursprünglich der Vorbereitung des BIBB-Forschungsprojekts: „Reflexive Meta-Evaluation von Modellprogrammen zum betrieblichen Lernen“ im Arbeitsbereich 3.3. unter Leitung von Dr. Dorothea Schemme (BIBB) in Kooperation mit Prof. Dr. Susanne Weber (Hochschule Fulda). Sie wurde nun für die Veröffentlichung in Form eines Diskussionspapiers aufbereitet.

Die Sichtung der Evaluationskonzepte erfolgte via Internet, Telefoninterviews, E-Mail-Anfragen und Literaturanalysen. Die Ergebnisse wurden mit den Institutionen rückgekoppelt. Dennoch handelt es sich um „unvollständige“ Rekonstruktionen und selektive Zusammenstellungen expliziter Informationen, die den Selbstdarstellungsbedürfnissen der Institutionen sicher nicht umfänglich gerecht werden. Auch „Verzerrungen“ aufgrund der fehlenden Innensicht – Insiderwissen, Erfahrungswissen oder „tacit knowledge“ – sind unvermeidbar; andererseits gilt hier, was für Evaluationen immer wieder eingefordert wird: der möglichst „unabhängige“ Blick von außen. Auf kritische Fragen und Kommentierungen wurde dennoch (weitgehend) verzichtet, um nicht den Eindruck einer externen Evaluation von Evaluationsansätzen zu erwecken.

Dr. Dorothea Schemme

BIBB Arbeitsbereich Modellversuche

Barbara Hupfer

Diplom-Soziologin / Wissenschaftsjournalistin

Einleitung

Knappe Mittel und Sparmaßnahmen in nahezu allen Bereichen öffentlichen Handelns machen es erforderlich, in möglichst wirkungsvolle und erfolgversprechende Fördermaßnahmen und Programme zu investieren. Evaluation spielt dabei nicht mehr nur auf Projekt – und Programmebene eine wichtige Rolle, sondern wird zu einem wichtigen Instrument der Qualitätssicherung und Organisationsentwicklung.

Insgesamt ist eine Schwerpunktverlagerung weg von der Planungsfokussierung hin zur Wirkungs- und Nutzenorientierung, zum Empowerment und zur Erkenntnisgenerierung durch Evaluation zu beobachten. Der Fokus richtet sich auf die Überprüfung der Nachhaltigkeit von Wirkungen, gleichzeitig sollen durch Evaluationen Lernchancen auf individueller und institutioneller Ebene eröffnet werden.

Obwohl die Bedeutung von Evaluation steigt, ist ein Mangel an institutionalisierter Evaluation in der öffentlichen Verwaltung zu konstatieren¹ (vgl. STOCKMANN 2004: 30,31). Das BIBB-Forschungsprojekt „Reflexive Meta-Evaluation von drei Modellprogrammen zum betrieblichen Lernen“ ist daher auch institutionelle Pionierarbeit: Es kann dazu beitragen,

- ## den institutionellen Diskurs zum Thema Evaluation anzuregen,
- ## die theoretische und methodische Weiterentwicklung von Evaluation durch die systematische Verknüpfung von Grundlagenforschung und anwendungsorientierter Forschung voranzutreiben ²
- ## und den fachübergreifenden und Theorie-Praxis-Dialog zu befruchten.

Die Evaluationsansätze und -verfahren wurden entsprechend dem Rechercheablauf dargestellt. Wiederholungen und Redundanzen ließen sich aufgrund der vorgenommenen Untergliederung nach Institutionen nicht vermeiden. Der Vorteil dieser Darstellung ist, dass der institutionelle Kontext bewahrt wurde.

Abschließend wird versucht, Trends, Gemeinsamkeiten und Unterschiede der Ansätze zu identifizieren. Ein nächster Schritt könnte die Erarbeitung einer Synopse-Matrix sein, um die Ansätze und Verfahren im Hinblick auf relevante Analysedimensionen und -kriterien vergleichen zu können.

¹ Vgl. Reinhard STOCKMANN (2004): Evaluation in Deutschland. In: Ders. (Hrsg.): Evaluationsforschung. Opladen, S. 30, 31

² STOCKMANN fordert, dass die Kluft zwischen sozialwissenschaftlicher Grundlagenforschung und anwendungsorientierter Evaluationsforschung überwunden werden müsse (STOCKMANN, Reinhard (2004): Evaluation in Deutschland. In: ders. (Hrsg.): Evaluationsforschung. Opladen, S. 37) und fordert eine anwendungsorientierten Professionalisierung der Evaluationsforschung in einer disziplinär und Grundlagenforschung bezogenen universitären Forschungslandschaft (vgl. HELLSTERN und WOLLMANN 1984, zit. nach STOCKMANN 2004: 34).

1. Das Evaluationssystem der Gesellschaft für Technische Zusammenarbeit (GTZ)

Die Entwicklungszusammenarbeit (EZ) ist ein Politikfeld, in dem die Evaluation von Vorhaben (Projekten und Programmen) schon zu einem vergleichsweise frühen Zeitpunkt institutionalisiert worden ist. Das Bundesministerium für wirtschaftliche Zusammenarbeit und Entwicklung (BMZ) hat bereits in den 70er-Jahren ein „Inspektionsreferat“ eingerichtet, das später in „zentrale Erfolgskontrolle“ umbenannt wurde. Auch die Durchführungsorganisationen des BMZ, die GTZ und die Kreditanstalt für Wiederaufbau (KfW) verfügen über weit entwickelte Evaluationssysteme.³

1.1. Ziele und Qualitätsmerkmale von Evaluationen in der GTZ

Evaluierungen⁴ sind Bestandteil des Qualitätsmanagementsystems der GTZ. Das heißt, sie liefern Erkenntnisse, um die eigene Arbeit kontinuierlich verbessern zu können. Überprüft werden Zielerreichung, Signifikanz, Relevanz, Effizienz und Wirkungen von entwicklungspolitischen Projekten und Programmen sowie deren Nachhaltigkeit. Das Monitoring von Wirkungen ist seit Einführung des entwicklungspolitischen Auftragsrahmens (AURA), der zwischen BMZ und GTZ vereinbart wurde, verpflichtend (www.gtz.de/Wirkungen). Überprüft werden sowohl die während der Projektlaufzeit auftretenden positiven und negativen, intendierten und nicht intendierten Wirkungen, als auch die Nachhaltigkeit von Projekt- oder Programmwirkungen nach Abschluss der Vorhaben (Schluss- und Ex-post-Evaluierungen). Als Wirkungen gelten dabei nur solche Veränderungen, die eindeutig auf die entwicklungspolitischen Interventionen zurückzuführen sind (www.gtz.de/Wirkungen).

Evaluierungen haben den Zweck der Rechenschaftslegung gegenüber dem BMZ und der Öffentlichkeit – hierzu gehört zum Beispiel die Frage der effizienten und sinnvollen Verwendung von Steuergeldern –, sie liefern Informationen für die Steuerung der Vorhaben und unterstützen das individuelle und institutionelle Lernen (vgl.: Internes Papier zum GTZ-Evaluierungssystem vom 3. Juni 2005 und www.gtz.de/evaluierungssystem).

³ Reinhard STOCKMANN, Leiter des Centrums für Evaluation (CEval) in Saarbrücken, hat in seinem Lehrbuch „Evaluationsforschung“ auf einige Defizite der früheren Evaluationspraxis von BMZ, GTZ und KfW hingewiesen. Im Fokus seiner Kritik standen Theorie- und Methodendefizite, fehlende gemeinsame Standards und Bewertungsmaßstäbe oder die Vernachlässigung von Ex-Post-Evaluationen und Wirkungsanalysen (STOCKMANN 2004: 375). Die aktuelle Evaluationspraxis und weiterentwickelte Evaluationssysteme in der EZ legen den Schluss nahe, dass seine Defizitdiagnose heute nicht mehr aktuell und gültig ist (A. d. V.).

⁴ In der GTZ wird der Begriff Evaluierung anstelle von Evaluation verwendet.

Die GTZ orientiert sich bei ihren Evaluierungen an den Grundsätzen des Development Assistance Committee (DAC) der OECD. Evaluierung wird hier definiert als eine „möglichst systematische und objektive Bewertung der Planung, Durchführung und Ergebnisse von laufenden oder bereits abgeschlossenen Vorhaben“ (Internes Papier der GTZ zu den DAC-Grundsätzen, 3.12.2005).

Qualitätsmerkmale einer Evaluierung sind dem DAC zufolge:

- ## „**Unabhängigkeit und Unparteilichkeit** bei der Konzipierung und Durchführung der Evaluierungen (TOR, Gutachterausswahl, Berichterstattung). Evaluierungsaufgaben sollten deshalb grundsätzlich vom operativen Management getrennt wahrgenommen werden. Die institutionelle Trennung sollte aber nicht so weit gehen, dass die Verknüpfung zwischen Evaluierungsergebnissen und operativen Entscheidungen verlorengeht.
- ## **Glaubwürdigkeit und Transparenz** werden erzielt, wenn sowohl Erfolge als auch Misserfolge beleuchtet und die Tatsachen, Schlussfolgerungen und Empfehlungen präzise auseinander gehalten werden.
- ## **Nützlichkeit:** Die Evaluierungsergebnisse sollten zu entscheidungs- und handlungsrelevanten Aussagen und Lerneffekten führen, damit die Qualität der Entwicklungszusammenarbeit verbessert und die öffentliche Rechenschaftslegung gesichert werden kann.
- ## Die **Verbreitung** der Evaluierungsergebnisse ist ein integraler Bestandteil der Evaluierung. Das DAC empfiehlt darüber hinaus die Einrichtung von Feedback-Mechanismen, die sicherstellen, dass die Evaluierungsergebnisse in die künftige Arbeit einfließen.
- ## Die **Beteiligung der Partner**, mit denen ein Entwicklungsvorhaben durchgeführt wird, sollte bei jeder Evaluierung weitestgehend praktiziert werden. Deren Sichtweise und Expertise gilt als integraler Bestandteil der Evaluierungen.“

(Internes Papier der GTZ zu den Evaluierungsgrundsätzen nach DAC, 3.12. 2005)

1.2 Selbstevaluierung und Fremdevaluierung

Das Evaluierungssystem der GTZ besteht aus den zwei Säulen Selbst- und Fremdevaluierung. „In den Evaluierungen wird gefragt: Tun wir das Richtige? (Relevanz); erreichen wir die Projekt- und Programmziele? (Effektivität); tragen wir zu übergeordneten Zielen und Wirkungen bei? („Impact“); handeln wir wirtschaftlich? (Effizienz); sind die Wirkungen von Dauer? (Nachhaltigkeit).“ (www.gtz.de/evaluierung; Stand 31.05.06). Einen hohen Stellenwert hat die Überprüfung der Nachhaltigkeit von Projekten und Programmen, das heißt, ob die angestrebten Wirkungen auch von Dauer sind.

Als **Selbstevaluierung** wird die systematische Überprüfung eines Vorhabens durch die operativen, projektführenden Arbeitseinheiten bezeichnet, bei der den DAC-Grundsätzen folgend auch externe, unabhängige Gutachter einbezogen werden. Dieses Verfahren wird Selbst-

evaluierung genannt, weil die Einschätzung der Beteiligten und Betroffenen im Fokus steht. Selbstevaluierungen sollen gegenüber Fremdevaluierungen den Vorteil haben, dass sie individuelles und institutionelles Lernen in besonderem Maße fördern, weil die Lernbereitschaft der Beteiligten sehr hoch ist. Dabei wird auf Untersuchungen verwiesen, die ergeben haben, dass sich die Ergebnisse von Selbstevaluierungen als sehr zuverlässig erweisen (vgl. Internes Paper zum GTZ-Evaluierungssystem, 3. Juni 2005 und www.gtz.de/Evaluierung).

Die zentralen Instrumente der Selbstevaluierung sind die Projektfortschrittskontrolle (PFK), die Querschnittsanalyse (QSA) und der Schlussbericht. E-VAL, ein neues elektronisch gestütztes Evaluierungsinstrument, wird nicht nur bei Selbst-, sondern zunehmend auch bei Fremdevaluierungen eingesetzt (www.gtz.de/evaluierungssystem Stand: 31.05.06). Eine ausführliche Beschreibung der einzelnen Instrumente ist auf der Internetseite der GTZ nachzulesen. (Ich halte mich bei ihrer Darstellung bewusst an die Selbstbeschreibung der Institution, um Verzerrungen eines Beobachters, der nicht Mitglied der GTZ Welt ist, zu vermeiden. A. d.V.).

„Die **Projektfortschrittskontrolle (PFK)** ist eine kritische, extern begleitete Reflexion über Stand und Wirkungen eines Vorhabens. Sie unterstützt die Projektsteuerung und das projektinterne und unternehmensweite Lernen. Gegen Ende einer jeden Phase eines Vorhabens wird verpflichtend eine PFK durchgeführt. Die Verantwortung liegt bei der projektführenden Einheit. Die PFK analysiert die Wirkungen des Vorhabens und beurteilt die Zielerreichung sowie die Relevanz der angestrebten Ziele für übergeordnete entwicklungspolitische Wirkungen. Sie überprüft die Konzeption und Strategie des Vorhabens, um Lern- und Veränderungsprozesse bei allen Beteiligten anzuregen. Wirkungsorientiertes Monitoring und e-val-Befragungen liefern die Grundlage für die interne Reflexion und den Blick nach vorn. Unabhängige Gutachter geben zusätzlich die Außensicht. So kann die Fortführung eines Vorhabens gesteuert werden und können die Erkenntnisse in das Wissensmanagement der GTZ einfließen.

Seit 1994 führt die GTZ jährlich eine **Querschnittsanalyse (QSA)** der laufenden und abgeschlossenen Vorhaben der Technischen Zusammenarbeit (TZ) durch. Mit diesen Analysen werden die Qualität der Vorhaben und ihre entwicklungspolitische Wirkung systematisch beobachtet. Sie bilden die Grundlage für die im zweijährigen Abstand erstellte Broschüre „Erfolgreiche Zusammenarbeit – nachhaltige Wirkungen“, mit der die GTZ öffentlich Rechenschaft über ihre Projektarbeit ablegt. Grundlage der QSA sind Erhebungsbögen, die u. a. Einschätzungen über die Wirkungen und Wirtschaftlichkeit der Vorhaben dokumentieren.⁵ Für laufende Projekte werden die Erhebungsbögen am Ende jeder Projektfortschrittskontrolle oder BMZ-Evaluierung vor Ort durch die Gutachter ausgefüllt und mit Projektteam und Partnern abgestimmt. Für abgeschlossene Vorhaben erfolgt dies im Zusammenhang mit der Erstellung des Schlussberichts.

Die Verantwortlichen des deutschen Beirats (Auftragsverantwortliche der GTZ) erstellen nach Beendigung jedes Vorhabens einen Schlussbericht an den Auftraggeber. Im Mittelpunkt des Schlussberichts steht die Auskunft über die erzielten Wirkungen und die Zielerreichung sowie die Dokumentation der wichtigsten Ereignisse und Entscheidungen im Verlauf eines Vorhabens. „Lessons learned“ und mögliche Übertragungen von Instrumentarien auf

⁵ Siehe die aktuelle Neunte Querschnittsanalyse der GTZ, in der 261 abgeschlossene TZ-Vorhaben der Jahre 2003/2004 und eine zusätzliche Zufallsstichprobe von 136 Vorhaben derselben Jahre ausgewertet wurden (www.gtz.de/QSA Stand 31.05.06).

andere Projekte oder Programme sind dabei wichtige Aspekte. Darüber hinaus ist der Schlussbericht ein Hilfsmittel für den Auftraggeber zur Wahrnehmung seiner entwicklungspolitischen und strategischen Steuerungsaufgabe für ähnliche Vorhaben.

E-VAL ist ein computergestütztes Verfahren für Evaluierung, das die GTZ in Zusammenarbeit mit einer Unternehmensberatung entwickelt und bis Januar 2006 in 470 Vorhaben eingeführt hat. E-VAL löst sich zunächst vom Anspruch der Objektivität und erkundet die subjektive Sicht der Personen, die gut über ein Vorhaben Bescheid wissen (Vertreter der Zielgruppen, Partnerfachkräfte, Mittler, GTZ Mitarbeiter/innen und andere Beteiligte). Die subjektiven Sichtweisen der befragten Personen werden von der Software aggregiert, und es ergeben sich sehr verlässliche quantitative Zusammenfassungen der Sichtweisen der verschiedenen Gruppen über Erfolg, Zielerreichung, Qualität der Beiträge, Stärken und Schwächen. Diese Ergebnisse werden in Kennzahlen ausgedrückt, die sich mit anderen Vorhaben in einem Sektor oder Land vergleichen lassen. E-VAL wird im Verbund mit Projektfortschrittskontrollen, Schlussberichten sowie auch Fremdevaluierungen eingesetzt“ (www.gtz.de/evaluierungssystem Stand 31.05.06).

Fremdevaluierung dient im Wesentlichen der Rechenschaftslegung und wird definiert als „unabhängige und unparteiische Überprüfung der Vorhaben, bei der die Verantwortung und die Steuerung entweder außerhalb der GTZ liegt oder innerhalb der GTZ bei der unabhängigen Evaluierungseinheit“ (Internes Paper zum GTZ Evaluierungssystem vom 3. Juni 2005).

Das Unterscheidungskriterium zur Selbstevaluierung ist damit der Ort der Steuerungsverantwortung, die bei Fremdevaluierungen entweder außerhalb der GTZ oder innerhalb der GTZ bei der unabhängigen Evaluierungsstelle liegt. Die Einbeziehung externer und unabhängiger Evaluator/-innen bzw. Gutachter/-innen ist demgegenüber kein hinreichendes Kriterium für die Unterscheidung von Selbst- und Fremdevaluierung, da Externe sowohl Selbst- als auch Fremdevaluierungen begleiten (A. d. V.).

Fremdevaluierungen erfolgen durch das BMZ, den Bundesrechnungshof, Wirtschaftsprüfungsgesellschaften und durch unabhängige, von der GTZ Stabsstelle Evaluierung beauftragte Institute (z.B. CEval).

Die BMZ Evaluierungen beziehen sich auf strategische und projektübergreifende Fragestellungen mit dem Ziel, das deutsche System der Entwicklungspolitik konzeptionell, organisatorisch und institutionell weiter zu entwickeln. Bei seinen strategischen Evaluierungen wird das BMZ durch die Stabsstelle Evaluierung der GTZ unterstützt (www.gtz.de/evaluierungssystem).

Instrumente der strategischen Evaluierungen des BMZ sind Themen- und Instrumentenevaluierungen sowie Evaluierungen von Länderprogrammen.

Themenevaluierungen beziehen sich auf bestimmte thematische Schwerpunkte der EZ, wie Kinder- und Jugendarmut oder die Entwicklungszusammenarbeit in Krisengebieten, die projekt- und länderübergreifend in den Blick genommen werden.

Instrumentenevaluierungen untersuchen bestimmte entwicklungspolitische Instrumente, wie beispielsweise die Kooperation zwischen technischer und finanzieller Zusammenarbeit.

Seit 2006 treten *Evaluierungen von Länderprogrammen* an die Stelle der bis dahin üblichen Sektorevaluierungen. „Sie betrachten das gesamte Portfolio der deutschen staatlichen Entwicklungszusammenarbeit (GTZ, CIM, KfW, DED, InWEnt) eines Kooperationslandes im Rahmen der festgelegten Schwerpunkte (z. B. Ressourcenschutz, Gesundheit, Energie)“ (www.gtz.de/Evaluierungssystem).

Innenrevision, entwicklungspolitische Bewertungen und Plausibilitätskontrollen sowie die Prüfung der Jahres- und Schlussrechnungen erfolgen durch eine von der GTZ beauftragte Wirtschaftsprüfungsgesellschaft. Diese wird seit 2001 auch mit der prozessbegleitenden Überprüfung einer repräsentativen Zufallsstichprobe von 100 laufenden Vorhaben anhand von Projektunterlagen beauftragt. Hinzu kommen regelmäßige Prüfungen und Sonderprüfungen des Bundesrechnungshofs.

Seit 2005 wird das Evaluierungssystem der GTZ durch externe Schluss- und Ex-post-Evaluierungen ergänzt. Neben diesen externen Fremdevaluierungen werden auch laufende Vorhaben fremdevaluert.

1.3 Verortung von Selbst- und Fremdevaluierung im internen und externen Kontrollsystems der Entwicklungszusammenarbeit

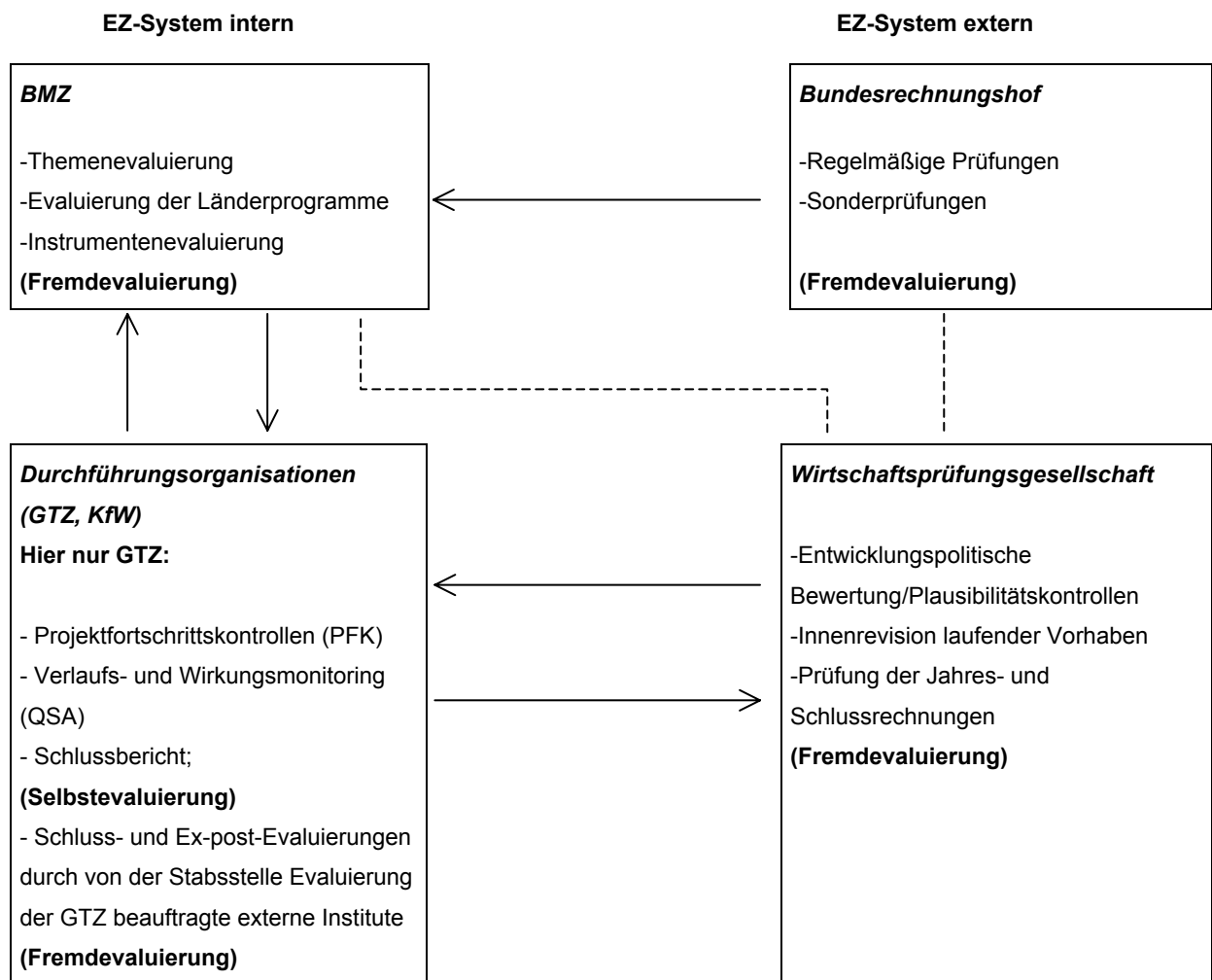
Die verschiedenen Instrumente und Aktivitäten der Selbst- und Fremdevaluierung werden im Folgenden in das Systemmodell der Erfolgskontrolle der deutschen Entwicklungszusammenarbeit integriert. (Diese Einordnung beruht auf Schlussfolgerungen der Verfasserin). Dabei ergeben sich Kombinationen von interner Selbstevaluierung, externer Fremdevaluierung einerseits und von interner und externer Fremdevaluierung sowie interner Selbst- und Fremdevaluierung andererseits.⁶

⁶ Diese Kombinatorik macht zumindest deutlich, dass es eine falsche Vereinfachung wäre, Selbstevaluierung mit interner Evaluierung oder Fremdevaluierung mit externer Evaluierung gleichzusetzen. Außerdem unterscheiden nicht alle hier vorgestellten Ansätze zwischen interner und externer Evaluation einerseits und Selbst- und Fremdevaluierung andererseits, was die Vergleichbarkeit der existierenden Evaluationssysteme erschwert (A. d. V.).

So sind die strategischen Evaluierungen des BMZ aus Sicht des Gesamtsystems der EZ dem internen Kontrollsystem zuzuordnen (interne Fremdevaluierung). Auch bei den durch unabhängige Gutachter und externe Institute durchgeführten Evaluierungen handelt es sich um interne Fremdevaluierungen. Die Fremdevaluierungen von Bundesrechnungshof und Wirtschaftsprüfungsgesellschaft sind demgegenüber dem externen Kontrollsystem zugeordnet (externe Fremdevaluierung), weil hier sowohl der Ort der Evaluierung als auch der Steuerung nicht in der GTZ angesiedelt sind.

Nur die Selbstevaluierungen mit den Instrumenten Projektfortschrittskontrolle (PFK), Querschnittsanalyse (QSA) und Schlussbericht sind gleichzeitig auch interne Evaluierungen (interne Selbstevaluierung).

Übersicht: Selbst- und Fremdevaluierung der GTZ im System der deutschen Entwicklungszusammenarbeit



(Quelle: BMZ, 10/1997 nach STOCKMANN 2004: 396 und eigene Ergänzungen 05/2006)

Aus Sicht der GTZ unterscheidet sich das, was als intern und extern definiert wird, von der Perspektive des BMZ oder auch der gesamten Entwicklungszusammenarbeit (A.d.V.). Die Evaluierungsaktivitäten des BMZ werden aus GTZ – Sicht nicht dem internen, sondern dem externen Kontrollsystem zugeordnet (externe Fremdevaluierung).

Diese Befunde machen zweierlei deutlich:

Erstens: Die Verwendung der vier Konzepte, Selbst- und Fremdevaluierung, interne und externe Evaluierung führt zur Vermehrung der Kombinationsmöglichkeiten und zu Mischformen und verweist auf die Perspektivität und Relativität dessen, was als interne oder externe Evaluierung definiert wird.

Zweitens: Die Begriffe interne und externe Evaluierung stehen aufgrund der Perspektivität dessen, was als intern oder extern definiert wird – und der Tatsache, dass externe Gutachter sowohl bei Selbst- als auch Fremdevaluierungen beteiligt sind - nicht notwendig für ein unterschiedliches Ausmaß an Objektivität und Unabhängigkeit von Evaluierungen (A. d. V.).

1.4 Reformen des GTZ-Evaluierungssystems

Die kontinuierliche Optimierung des Evaluierungssystems der GTZ ist selbsterklärtes Ziel. In den letzten Jahren wurden zusammenfassend folgende Reformen eingeleitet und realisiert:

Im Sinne des Unabhängigkeitspostulats nach den OECD/DAC Kriterien wurde die Bedeutung unabhängiger und unparteiischer Gutachter bei Selbst- und Fremdevaluierungen gestärkt. Externe Institute, wie zum Beispiel das „Centrum für Evaluation“ (CEval) unter Leitung von Professor Reinhard STOCKMANN, übernehmen im Auftrag der GTZ die Evaluierung laufender Vorhaben, Schlussevaluierungen und Ex-post-Evaluierungen.⁷

Die stärkere Gewichtung unabhängiger Evaluierungen erfolgt auch durch die Aufwertung der hausinternen Evaluationsstelle, der Stabsstelle Evaluation, die direkt der Geschäftsstelle unterstellt wurde und vom operativen Bereich getrennt ist. Die Stabsstelle „arbeitet mit einem übergeordneten, auf das Gesamtportfolio und Unternehmensstrategien ausgerichteten Fokus. Sie entwickelt Evaluierungsverfahren und Instrumente, unterstützt die operativen Bereiche und Vorhaben bei der Selbstevaluierung und steuert die Fremdevaluierungen“ (Internes Papier zum GTZ-Evaluierungssystem vom 3. Juni 2005).

Zu den einzelnen Aufgaben der Stabsstelle zählen das Controlling strategischer Projekte, die Durchführung der Querschnittanalysen (QSA), die Beratung von TZ-Vorhaben und GTZ-Organisationseinheiten, die Optimierung des GTZ-Evaluierungssystems, die Verbreitung von

⁷ vgl. Kapitel 2: Evaluationsansatz des Centrums für Evaluation – CEval

Evaluationsergebnissen und die Zusammenarbeit mit den Evaluierungseinheiten anderer Geber und Durchführungsorganisationen (GTZ-internes Papier vom Juni 1999 zit. nach STOCKMANN 2004: 398).

Ein weiterer Schwerpunkt der Reformen betrifft die 2005 eingeführten Schluss- und Ex-Post-Evaluierungen, um die Nachhaltigkeit von Projekt- und Programmwirkungen, deren Relevanz, Signifikanz und Effizienz beurteilen zu können. Bei den Schlussevaluierungen, die einige Monate vor Beendigung eines Vorhabens beginnen, werden Chancen für die Nachhaltigkeit von Wirkungen eingeschätzt und letzte Steuerungen ermöglicht. Im Fokus der Ex-post-Evaluierungen, die zwei bis fünf Jahre nach Abschluss eines Vorhabens stattfinden sollen, steht die Beurteilung der tatsächlich eingetretenen Nachhaltigkeit von Wirkungen.

Verfahrenstechnische Verbesserungen wie das neu eingeführte computergestützte E-Val-System sollen dazu beitragen, die Validität von Evaluierungen zu erhöhen; auch die methodische und konzeptionelle Weiterentwicklung von GTZ-Evaluierungen (- in Anlehnung an die CEval-Konzeptionen -), die Orientierung an den Standards der empirischen Sozialforschung und den Standards für Evaluation der Gesellschaft für Evaluation (DeGEval) sollen die Qualität der Evaluierungen sichern.

Weiteres Ziel der Reformen ist die methodische Weiterentwicklung der Evaluierung von Vorhaben der programmorientierten Gemeinschaftsfinanzierung mit bi- und multilateralen Gebern und Partnerländern (vgl. Internes Papier zum GTZ-Evaluierungssystem vom 3. Juni 2005).

1.5 Literatur

BMZ (Hg.) (2000): Leitfaden zur Durchführung von BMZ-Evaluierungen. Bonn: BMZ

BODEMER, Klaus (1997): Erfolgskontrolle der deutschen Entwicklungshilfe – improvisiert oder systematisch? Meisenheim

BORRMANN, Axel u. a. (2001): Reform der Erfolgskontrolle in der deutschen Entwicklungszusammenarbeit. Eine Zwischenbilanz. Baden-Baden: Nomos

BORRMANN, Axel (1999): Erfolgskontrolle in der deutschen Entwicklungszusammenarbeit. Baden-Baden: Nomos

BRÜNE, Stefan (Hrsg.)(1998): Erfolgskontrolle in der entwicklungspolitischen Zusammenarbeit. Hamburg: Deutsches Übersee-Institut

DONNER, Franziska (1998): Das Qualitätsmanagement der GTZ – zum Grundverständnis. In: Entwicklung und Zusammenarbeit, Jg. 39, Nr. 516, S. 1313-133

GTZ (Hrsg.) (1998): Nachhaltige Wirkungen durch Qualitätsmanagement – eine Herausforderung für die Technische Zusammenarbeit. Eschborn: GTZ

GTZ (1999): Wegweiser für die PFK. Eschborn: GTZ

GTZ (2005): Das GTZ Evaluierungssystem. Internes Paper der GTZ vom 3. Juni 2005. Eschborn: GTZ

KEVENHÖRSTER, Paul; STOCKMANN, Reinhard (2001): Wissenschaftlicher Rigorismus oder praxisorientierter Pragmatismus? Zum Verhältnis zwischen Entwicklungspolitik und Wissenschaft. In: E u.Z, Nr. 4, 42. Jg.

MICHAELOWA, Katharina; BORRMANN, Axel (2003): Wer evaluiert was, wie und warum? Eine politökonomische Analyse am Beispiel der deutschen Entwicklungszusammenarbeit. Vortragsmanuskript

REUBER, Marianne (2005): Inwieweit erfüllt die GTZ die DAC-Evaluationskriterien? Internes Paper der GTZ vom 15.12.2005. Eschborn: GTZ

STOCKMANN, Reinhard (2004): Evaluation staatlicher Entwicklungspolitik. In: ders. (Hrsg.) (2004): Evaluationsforschung. Opladen, S. 375-410

2. Der Evaluationsansatz des Centrums für Evaluation (CEval)

2.1 Aufgaben und Programmatik

Das *Centrum für Evaluation (CEval)* ist ein der Universität des Saarlandes, Fachrichtung Soziologie, angegliedertes wissenschaftliches Institut. Die Gründung des Instituts (2002) wurde im Zusammenhang mit den Reformprozessen der Evaluationssysteme der Entwicklungszusammenarbeit (BMZ, GTZ, KfW) von Professor Stockmann initiiert, der das Centrum leitet. Zunehmend agiert CEval wie ein Beratungsunternehmen auf dem freien Markt und übernimmt Aufträge für die Beratung und externe Durchführung von Evaluationen. So wurde zum Beispiel ein Beratervertrag mit dem BMZ geschlossen, der die Unterstützung in methodischen und evaluativen Frage- und Problemstellungen vorsieht. Im Mittelpunkt eines Rahmenvertrags mit der GTZ stehen regelmäßige Evaluationen von GTZ-Vorhaben, insbesondere von Schluss- und Ex-post-Evaluationen sowie die Weiterentwicklung der Evaluationskonzeption und -Methodik; diese orientiert sich an den im CEval entwickelten Konzeptionen (www.ceval.de/Jahresbericht).

Der Aufgabenbereich von CEval umfasst:

- ≠ Die Erarbeitung und Weiterentwicklung der theoretischen und methodischen Grundlagen sowie der fachlich-wissenschaftlichen Standards der Evaluation von Programmen und Maßnahmen,
- ≠ die Qualifizierung von (internen und externen) EvaluatorInnen und Gutachtern,
- ≠ die Beratung öffentlicher und privater Einrichtungen bei der Durchführung von Evaluationen,
- ≠ die (externe) Durchführung von Evaluationen,
- ≠ die Förderung des fachlichen Informationsaustauschs (vgl. www.ceval.de).

Erkenntnisgenerierung und Überprüfung der Nachhaltigkeit von Wirkungen stehen im Vordergrund

Evaluationsgegenstände des CEval sind Modellversuche, Projekte, (Förder-)programme und Organisationen. Im Vordergrund stehen die Erkenntnisgenerierung und die Überprüfung der Nachhaltigkeit von Projekt- und Programmwirkungen. Ein Blick auf die Liste der

Auftrags-Evaluationen vom 9. 2. 06 (www.ceval.de/Projekte) verweist auf auftraggebende Institutionen⁸ und Evaluationsgegenstände in den Schwerpunktbereichen Umweltforschung und Umweltkommunikation, Bildung, Arbeitsmarkt und Entwicklungszusammenarbeit. Konstatiert wird darüber hinaus eine verstärkte Nachfrage nach Evaluationen im Kulturbereich (www.ceval.de/Jahresbericht 2005).

2.2 Evaluationsansatz

Der Evaluationsansatz von CEval ist interdisziplinär und Politikfeld übergreifend ausgerichtet; dadurch kann und soll aus den Erfahrungen in anderen Bereichen gelernt werden (Lernfunktion von Evaluation). Grundsätzlich gilt, anwendungsbezogene Fragestellungen mit wissenschaftlichen Erkenntnissen zu verknüpfen.

Als besondere Merkmale des Evaluationsansatzes des CEval werden genannt:

- Methodenvielfalt,
- Partizipation,
- Theorieorientierung,
- Qualifizierung.

Methodenvielfalt liefert „aussagekräftigeres Bild“

CEval Evaluationen können aus der Vielfalt soziologischer Methodenkompetenz der MitarbeiterInnen profitieren. Multimethodendesigns, bestehend aus quantitativen und qualitativen Methoden der empirischen Sozialforschung, sollen die Validität und Reliabilität der Evaluationsergebnisse sichern. „Will man Feststellungen über relevante Programmbedingungen und -wirkungen durch ein Gefüge von sich wechselseitig stützenden Evidenzen absichern, so liefert ein multipler methodischer Zugang im allgemeinen ein reichhaltigeres und aussagekräftigeres Bild als ein monomethodischer Ansatz“ (BRANDSTÄDTER 1999 zit. nach STOCKMANN 2004: 15).

⁸ Auftraggeber sind z.B.: GTZ, EED (Evangelischer Entwicklungsdienst), InWent (Internationale Weiterbildung und Entwicklung), Alexander von Humboldt Stiftung, Competence-Center Virtuelle Saar Universität, DBU (Deutsche Bundesstiftung Umwelt), Marie-Curie-Programm der EU, DAAD, Bundesinstitut für Berufsbildung (mit bisher zwei Evaluationsaufträgen: Evaluation des KIBB „Kommunikations- und Informationssystems Berufliche Bildung“ 08/03 – 12/05 und die Beratung von „IBQM-Initiativstelle Berufliche Qualifizierung von Migrantinnen und Migranten bei der Evaluation der Beruflichen Qualifizierungsnetzwerke“ 04/04-11/06); Mehr zu Auftraggebern und Evaluationsprojekten ist im Jahresbericht 2005 des CEval nachzulesen (zu bestellen unter: ve.schneider@ceval.de oder jetzt neu auf der Website www.ceval.de/Jahresbericht).

Angestrebt wird zudem die systematische Analyse von Wirkungen – gemessen an den Programmzielen – sowie möglichst weitgehende Kausalitätsprüfungen.

Die Auswahl passgenauer Evaluationsmethoden und die Entwicklung von Instrumenten der Datenerhebung orientieren sich am Evaluationsgegenstand, den Zielen und Aufgaben einer Evaluation, dem gewählten Untersuchungsparadigma⁹, dem Erkenntnisinteresse, dem theoretischen Rahmen und den EvaluatorenInnen (vgl. STOCKMANN 2004: 15).

Zur Analyse prozessbezogener Daten (Programmsteuerung, -ablauf etc.) werden überwiegend qualitative Erhebungsmethoden als geeignet befunden, für die Überprüfung der Zielerreichung, Wirkungsevaluation und Kausalbetrachtungen werden quantitative Erhebungs- und Auswertungsverfahren eingesetzt.

Erhebungsmethoden sind u. a.:

- ## Sekundäranalysen vorhandener Materialien,
- ## Leitfadeninterviews,
- ## Standardisierte Befragungen,
- ## Fallstudien (STOCKMANN 2004: 15).

Partizipation der Beteiligten als Prinzip

Evaluationsansätze, auf die Bezug genommen wird, sind beispielsweise der „Nutzungsgesteuerte“ (BEYWL 2000), der „Wirkungsorientierte“ (BEYWL et al. 2004) oder die „Empowerment-Evaluation“ (FETTERMAN et al. 1996). Bei all diesen Ansätzen gilt das Prinzip, die Beteiligten in den Evaluationsprozess einzubeziehen.¹⁰

„Dem Prinzip Partizipation der Beteiligten“ liegt eine interaktionistische Forschungsmethodologie zugrunde. Partizipation beinhaltet die gemeinsame Entwicklung von Qualitätszielen und -kriterien. Sie soll gewährleisten, dass die Vielfalt der Nutzungsinteressen und Sichtweisen einbezogen und Vertrauen in das Evaluationsvorhaben geschaffen wird. Andererseits birgt dieses Vorgehen Konfliktpotenziale, da Qualitätsbegriffe und -ziele unterschiedlich sein können.

⁹ Evaluationsparadigmen oder -Funktionen, die Einfluss auf die Methodenwahl, geeignete Instrumente und den Evaluationsansatz haben, sind: Erkenntnisfunktion, Kontrollfunktion, Legitimitätsfunktion und Lernfunktion (vgl. STOCKMANN 2004. Was ist eine gute Evaluation? Ceval Workpaper 9: 3).

¹⁰ Der Ansatz der Cluster-Evaluation (s. DJI) werde, so Krapp, ihrer Kenntnis nach nicht eingesetzt (Telefongespräch am 11. Februar 2006)

Wichtiges Ergebnis ist, dass die Beteiligten in allen Phasen der Evaluation einbezogen werden; in der Design- und Verwertungsphase als aktive Mitgestalter, in der Durchführungsphase als Ressourcepersonen für die Informationsgewinnung und in der Abschlussphase als Impulsgeber für die Verwertungsziele:

In der Designphase entwickeln die EvaluatorInnen gemeinsam mit den Beteiligten das Evaluationskonzept. Im Mittelpunkt steht die Konkretisierung eines Analyseleitfadens/Analyserasters und es werden verbindliche Absprachen bezüglich des Vorgehens getroffen. Partizipation und prozessbegleitende, formative Evaluation korrespondieren mit einem entwicklungsoffenen Evaluationskonzept, das kontinuierliche Anpassungen der eingesetzten Evaluationsinstrumente ermöglicht (vgl. STOCKMANN 2004: CEval Workpaper 9: 13).

In der Durchführungsphase der Datenerhebung fungieren die Beteiligten als Ressourcepersonen für die Informationsgewinnung. Durch regelmäßiges Feedback der Auswertungsergebnisse und gemeinsame Workshops wird deren Einbindung gesichert.

Am Ende eines Evaluationsvorhabens findet in einem Abschlussworkshop die gemeinsame Diskussion und Bewertung der Ergebnisse statt und es werden Strategien zur Umsetzung/Implementierung der Ergebnisse entwickelt (Verwertungsphase).

Theorieorientierung

Das theoretische Konzept des CEval versteht sich als hypothesengenerierend und wirkungsorientiert; es nimmt beispielsweise Bezug auf die Lebenszyklusforschung, Organisationstheorien, Innovations- und Diffusionstheorien (vgl. Evaluation des Kommunikations- und Informationssystem berufliche Bildung (KIBB) des Bundesinstituts für Berufliche Bildung).

Qualifizierung: Aufbau von Evaluationskompetenz

Evaluationen können als Instrumente der Qualitätsentwicklung und -sicherung eingesetzt werden. Denn sie tragen dazu bei, Ablaufprozesse zu optimieren, Zielerreichung und Wirksamkeit zu überprüfen, Relevanz und Signifikanz von Zielen und Wirkungen zu hinterfragen sowie Ursache-Wirkungszusammenhänge aufzudecken (www.ceval.de/Jahresbericht). Um eine kontinuierliche Qualitätssicherung zu ermöglichen, wird angestrebt, Evaluationen zunehmend mit Qualifizierungsmaßnahmen für MitarbeiterInnen der evaluierten Organisationen zu verbinden (Aufbau interner Evaluationskompetenz). Zielgruppe sind vor allem Verantwortliche für Qualitätsentwicklung oder Monitoring & Evaluation.

2.2.1 Interne oder externe Evaluation?

Auch wenn CEval selbst als externer Evaluator auftritt, wird nicht generell für den Einsatz externer Evaluation votiert (A.d.V.). Bevor die Entscheidung für eine interne oder externe Evaluationen getroffen wird, sollten die jeweiligen Institutionen jedoch Vor- und Nachteile im Hinblick auf die Evaluationsziele abwägen. Die Unterscheidung zwischen interner und externer Evaluation bezieht sich auf den Ort der Evaluation, also, ob sie von derselben Organisation vorgenommen wird, die das Projekt oder Programm durchführt (intern) oder von einer externen Institution.¹¹

Derzeit ist ein Trend zur externen Evaluation spürbar, da ihr im allgemeinen größere Unabhängigkeit und Objektivität attestiert wird (A.d.V.).

Als Vorteile externer Evaluation werden genannt:

- ## größere Unabhängigkeit,
- ## profundere Methodenkompetenz externer EvaluatorInnen,
- ## zusätzliche Legitimierung und Einflussstärke bezogen auf organisationale Reformkräfte.

Nachteile externer Evaluation können sein:

- ## Mangelndes Gegenstandswissen und Prozesswissen: Dies kann dazu führen, dass externe EvaluatorInnen abgelehnt werden. Insbesondere bei komplexen und/oder transdisziplinären Forschungsprojekten(-objekten) kann mangelndes Gegenstandswissen externer EvaluatorInnen zur Evaluationshürde werden (A.d.V.).
- ## Fehlende Vertrauensbasis: Externe Evaluationen können Angstgefühle, von außen kontrolliert zu werden, hervorrufen. Diese können wiederum Abwehrreaktionen (Reaktanz) und Boykothaltungen nach sich ziehen, was sich negativ auf die Qualität und Validität der erhobenen Daten auswirken kann.
- ## Es können zusätzliche Kosten entstehen.
- ## Die spätere Implementierung oder Umsetzung der Beschlüsse/Ergebnisse könnte problematisch werden (fehlende Begleitung).

¹¹ Zu unterscheiden wäre darüber hinaus, ob die EvaluatorInnen identisch sind mit den Personen, die mit der operativen Durchführung des Projektes/Programms betraut waren (z. B. Selbstevaluierung bei der GTZ) oder um *Peers* aus einer anderen Organisationseinheit (interne Fremdevaluierung).

Als Vorteile interner Evaluation gelten:

- ≠# Sie ist im allgemeinen weniger aufwändig und kostengünstiger.
- ≠# Interne EvaluatorenInnen verfügen über eine größere Sachkenntnis bezogen auf das Evaluationsobjekt.
- ≠# Insbesondere in formativen und prozessbegleitenden Evaluationen können interne EvaluatorenInnen in der Regel notwendige Korrekturen im Projektverlauf schneller umsetzen.

Schwächen interner Evaluation können sein:

- ≠# Fehlende Methodenkompetenz interner Mitarbeiter;
- ≠# Mangel an Distanz und Unabhängigkeit/Projektblindheit, so „dass sie aussichtsreiche Alternativen nicht kennen“ (STOCKMANN 2004: Workpaper 9: 8).

Fällt die Entscheidung für eine interne Evaluation, so sei darauf zu achten, dass die internen projekt- oder programmverantwortlichen MitarbeiterInnen keine qualitativen Auswertungen oder Interpretationen der Evaluationsergebnisse vornehmen, so Stefanie Krapp vom CEval (Telefongespräch am 11.01.06). Subjektive Verzerrungen könnten jedoch durch quantitative, „objektive“ Methoden der Datenerhebung abgeschwächt werden sowie durch Bewertungen im Team mit externen EvaluatorenInnen.

Beispiel für ein internes Steuerungsinstrument ist das Projektmonitoring, das von den MitarbeiterInnen selbst durchgeführt wird.

2.2.2 Summative und formative Evaluation

Bei der Evaluationsdurchführung wird zwischen summativer (bilanzierender) und/oder formativer (gestaltender) Evaluation unterschieden:

- ≠# Formative Evaluationen: Sie liefern steuerungsrelevante Informationen für das Programmmanagement und erfolgen prozessbegleitend; Inkonsistenzen des Projektplanes oder der Programmplanung können aufgedeckt und Korrekturmöglichkeiten des Vorgehens aufgezeigt werden. Bei der formativen Evaluation stehen die Reflexion auf notwendige Veränderungen, Lern- und Erkenntnisziele im Vordergrund. Beispiel: formative Evaluation des Kommunikations- und Informationssystems Berufliche Bildung (KIBB) des Bundesinstituts für Berufsbildung.
- ≠# Summative Evaluationen: Vor allem im Rahmen von Abschluss- und Ex-post-Evaluationen wird summativ evaluiert, da hier kaum Korrekturmöglichkeiten eines Projekts

oder Programms gegeben sind. Die summative Evaluation bilanziert den Grad der Zielerreichung und untersucht die Projektwirkungen anhand vorliegender Dokumente und Ergebnisse der prozessbegleitenden Evaluation.¹² Aufgrund der Durchführung von Schluss- und Ex-post Evaluationen für die GTZ erhält dieser traditionelle Schwerpunkt des CEval neue Bedeutung. Ein Beispiel für formative und summative Evaluation ist die Entwicklungspartnerschaft MENTOR im Rahmen der Gemeinschaftsinitiative EQUAL.

- ## Aufbau von wirkungsorientierten Monitoring- und Evaluationssystemen: CEval berät und unterstützt Organisationen beim Aufbau interner Monitoring- und Evaluationssysteme, die von den Mitarbeitern und Mitarbeiterinnen selbst bedient werden.

2.2.3 Orientierung an den „Standards für Evaluation“ und wissenschaftstheoretische Position

Das CEval bezieht sich laut Jahresbericht 2005 bei seinen Evaluationen auf die „Standards für Evaluation“ der Deutschen Gesellschaft für Evaluation (DeGEval); diese sollen die Qualität von Evaluationen sichern und Transparenz schaffen über Evaluationen als professionelle Praxis (STOCKMANN 2004: CEval Workpaper 9: 12).¹³ „Bewertungskriterien richten sich nicht nach vorgegebenen Normen (ISO) oder festgelegten Parametern, sondern nach Kriterien, die den situativen Bedingungen des Evaluationsgegenstandes angepasst werden“ (JAHRESBERICHT 2005: 5).

Das CEval orientiert sich tendenziell an einem kritisch-rationalen Wissenschaftsverständnis, wobei auch systemische Zusammenhänge in den Blick genommen werden (S. KRAPP in einem Telefongespräch am 11.01.06). Diesbezüglich ist auch auf Professor STOCKMANN zu

¹² Dennoch können Abschluss- oder Ex-post-Evaluationen sowie Meta-Evaluationen formativ angelegt sein und müssen sich nicht auf bilanzierende (summative) Aktivitäten beschränken. So besteht die formative Funktion des BIBB Projekts „Reflexive Meta-Evaluation von drei Modellversuchsprogrammen zum betrieblichen Lernen“ (2006-2008) darin, reflexiv neues Wissen für eine zukunftsfähige Gestaltung der betrieblichen Bildungsarbeit und einer innovativen Programmgestaltung zu generieren und nicht nur Ertrag und Wirkungen zu bilanzieren (vgl. BIBB Forschungsprojekt 8: Reflexive Metaevaluation von Modellprogrammen zum betrieblichen Lernen; Leitung der Entwicklungspartnerschaft (intern/extern): Dr. Dorothea SCHEMME, BIBB und Prof. Dr. Susanne WEBER, Hochschule Fulda).

¹³ In diesem Zusammenhang ist auf einen Vortrag von Prof. Stockmann zum Thema „Evaluation als Instrument der Qualitätssicherung“ am 10. November 2005 in Bonn hinzuweisen (Veranstaltung mit der Deutschen Bundesstiftung Umwelt in Kooperation mit dem BIBB). Evaluation wurde hier nicht mehr ausschließlich auf Evaluationsgegenstände bezogen, sondern als Instrument von Qualitätssicherung und Organisationsentwicklung thematisiert.

verweisen, der konstatiert, dass der „kalte Krieg der wissenschaftstheoretischen Positionen“ mittlerweile dem Bemühen weiche, Gemeinsamkeiten zu erkunden (vgl. STOCKMANN 2004: CEval Workpaper 9).

2.2.4 Der Analyseleitfaden

Als zentrales Steuerungsinstrument einer Evaluation gilt der Analyseleitfaden, der sich aus dem Theoriegerüst (s. Punkt 2.3) generiert. Er hat den Zweck, den Informationsbedarf zu strukturieren und ist Grundlage der Instrumentenkonstruktion sowie der Datenauswertung, -analyse und -interpretation. Der Analyseleitfaden orientiert sich an den spezifischen Bedürfnissen der Beteiligten „und verbindet die theoretisch fundierten, allgemeinen Erfolgsparameter von Programmen und Projekten mit den konkreten Zielen, Inhalten und Wirkungsindikatoren“ (KRAPP 2005).

2.3 Beispiel: Die Evaluation des Kommunikations- und Informationssystems Berufliche Bildung (KIBB)

Das KIBB-Portal ist Teil des „Kommunikations- und Informationssystem Berufliche Bildung“ des Bundesinstituts für Berufsbildung, das mit finanzieller Unterstützung des BMBF aufgebaut wurde.¹⁴ KIBB verfolgt drei Hauptziele:

- ## Die zeitnahe Beantwortung von Anfragen aus Verwaltung und Politik der Berufsbildung,
- ## den Aufbau eines Portals der deutschen Berufsbildungsforschung,
- ## die Förderung des Wissensaustauschs in der Arbeitsgemeinschaft Berufsbildungsforschungsnetz (AG BFN).

Der theoretische Rahmen der projektbegleitenden Evaluation von KIBB wurde in einem Beitrag von Krapp (2005) dargestellt (vgl. KIBB Newsletter 1/2005; www.kibb.de/berichte/newsletter). Er steht exemplarisch für unterschiedliche Evaluationsgegenstände und beinhaltet die bereits genannten theoretischen Modelle: Lebenszyklus-Modell, organisationstheoretisch hergeleitetes Wirkungsmodell, Diffusions- und Nachhaltigkeitsmodell.

¹⁴ Die KIBB-Wissenslandkarte wird kontinuierlich um Wissensbestände aus der Arbeitsgemeinschaft Berufsbildungsforschungsnetz (AG BFN) ergänzt.

Lebenszyklus-Modell betont den Prozesscharakter von Programmen

Das Lebenszyklus-Modell ist ein heuristisches Modell für die (insbesondere) formative und prozessbegleitende Evaluation. Die Lebensverlaufsperspektive betont den Prozesscharakter von Projekten und Programmen, die einem typischen Phasenmuster folgen. Nach der gelungenen Implementationsphase ist vor allem die Nachhaltigkeit der Innovationen ein Indikator für den Erfolg. „Das phasenorientierte Entwicklungskonzept von KIBB sowie eine auf die erfolgreiche Implementation und Nachhaltigkeit bezogene Ergebnisevaluation findet im Lebenszyklus-Modell somit einen geeigneten theoretischen Rahmen“ (KRAPP 2005). Die einzelnen Projekt- oder Programmphasen sind aus Sicht der Lebensverlaufsperspektive kausal miteinander verbunden. Demnach wird die Möglichkeit nachhaltiger Projektimplementierung bereits durch die Projektauswahl und -konzeption beeinflusst. Aber auch die während der Projektlaufzeit geschaffenen Strukturen sind Voraussetzungen für die Möglichkeit langfristiger und nachhaltiger Projektwirkungen.

Wirkungsmodell: Organisationstheoretischer Kontingenzansatz zur Evaluation der Wirkungen

Der Evaluationsansatz rekonstruiert KIBB als technisch-organisationale Innovation. In dem verwendeten Wirkungsmodell wird die Innovation gleichgesetzt mit dem Projektauftrag/Ziel von KIBB, nämlich der Entwicklung eines Kommunikations- und Informationssystems für Expertinnen und Experten im Bereich der beruflichen Bildung.¹⁵ Die Wirkungsevaluation bezieht sich demnach auf den Projektauftrag/die Ziele und ist bis hierher ein Ist-Soll-Vergleich. Wie die Diskussion um Methoden und Ansätze der Evaluation in der EZ (GTZ) bereits gezeigt hat, ist die planungsfokussierte Evaluation/Ist-Soll-Vergleich um eine Wirkungsanalyse zu ergänzen, in der auch nicht intendierte Effekte zu berücksichtigen sind.

Die (angestrebten) Ziele und Wirkungen sind aus organisationaler beziehungsweise institutioneller Perspektive Innovationen, deren langfristiger Bestand Voraussetzung für die Nachhaltigkeit der Projektwirkungen ist. Für die Analyse der eingesetzten Instrumente von KIBB (Wissenslandkarten etc.) und ihrer spezifischen Wirkungen im Hinblick auf organisationale

¹⁵ Weitere (implizite) Ziele sind eine effektivere Nutzung und Bereitstellung von Expertenwissen und (aktuellen) Informationen sowie der Förderung des Austauschs zwischen unterschiedlichen Akteuren der Berufsbildung. Zu überprüfen ist auch das Ziel von KIBB, die Stellung des BIBB als Kompetenzzentrum für Entwicklung und Forschung in der Berufsbildung zu fördern und die Zusammenarbeit mit dem BMBF und intern zu verbessern.

Veränderungen wurde ein organisationstheoretischer Kontingenzansatz gewählt.¹⁶ Inhaltlich geht es bei der Analyse um die Fähigkeit der Organisation, Innovationen zu verbreiten.

Das Innovations-/Diffusionsmodell: Diffusion als zentraler Indikator für die Nachhaltigkeit und den Erfolg einer Innovation

Das Innovations-/Diffusionsmodell adressiert vor allem produktbezogene Faktoren, die zur Verbreitung bestimmter Angebote und deren Wirkungen beitragen. Der diffusions-theoretische Ansatz berücksichtigt, dass KIBB verschiedene Aufgaben und implizite Ziele miteinander verbindet: Einerseits die Verbesserung der Kooperationsbeziehungen mit anderen Institutionen, andererseits die Verbreitung und Vernetzung von Wissen in der Berufsbildung. Diffusion wird somit zu einem zentralen Indikator für Nachhaltigkeit und Erfolg.

Neben produktspezifischen Kriterien, die für einzelne Instrumente bzw. Angebote - Wissenslandkarte, Expertenmonitor, AG BFN-Community - im Projektverlauf zu operationalisieren sind, gibt es eine Reihe von Variablen, die für die meisten der einzuführenden Innovationen gelten. Insgesamt können vier Variablengruppen unterschieden werden, die den Diffusionsprozess beeinflussen:

- 1) die Konzeption,
- 2) die spezifischen Eigenschaften der Innovationen selbst,
- 3) die Eigenschaften der Organisation, welche die Innovation einführt und
- 4) externe Variablen (gesetzliche Rahmenbedingungen, aktuelle Agenden der öffentlichen Diskussion, etc.).

Das Nachhaltigkeitsmodell

Die Nachhaltigkeit einer Innovation wird mithilfe eines Ansatzes untersucht, der sich nicht auf den Wirkungshorizont der Projektdauer beziehungsweise der Implementationsphase der Produkte beschränkt. Damit werden Zukunftsfähigkeit, langfristige Tragfähigkeit und Folgen von KIBB berücksichtigt. Neben der dauerhaften Implementierung neuer organisatorischer

¹⁶ Dieser wurde hier nicht näher erläutert. Die Kontingenztheorie betont die Abhängigkeit („Contingency“) zwischen Organisation und Umwelt, wobei Ziel der Interventionen von Beratung und Management die Optimierung der Umwelthanpassung ist. Das Konzept der „Abhängigkeit“ wird u.a. von Systemtheoretikern kritisch reflektiert, weil es das Beharrungsvermögen komplexer Organisationen gegenüber einer sich verändernden Umwelt unterschätze. Der Kontingenzansatz ist ein Vorläufer der später entwickelten „Interorganisationstheorie“ und der „Netzwerktheorie“ von Organisationen (vgl. WILLKE, Helmut 1999: Systemtheorie II. Interventionstheorie. Stuttgart, S. 170,171).

und inhaltlicher Strukturen sind die dauerhafte Nutzung der erarbeiteten Angebote und Konzepte und die effektive Erreichung der spezifischen Zielgruppen und der für diese Zielgruppen geschaffene Mehrwert zu erfassen. Neben dem Fortbestand der eigentlichen Projektergebnisse ist ein weiterer Indikator für Nachhaltigkeit, dass KIBB und die beteiligte(n) Organisation(en) auch nach der Pilotphase innovativ auf sich ändernde Rahmenbedingungen und Problemlagen reagieren und Projektergebnisse entsprechend innovativ anpassen können.

Aus diesem Theoriegerüst - bestehend aus Lebenszyklus-Modell, organisationstheoretischem Kontingenzansatz, Innovations-/Diskussionsmodell und Nachhaltigkeitsmodell - wurde der Evaluationsanalyse-Leitfaden gemeinsam mit den Projektbeteiligten generiert¹⁷. Dieser strukturierte den Informationsbedarf entsprechend der Anforderungen von KIBB. Er war Grundlage für das methodische Vorgehen und die Instrumentenauswahl sowie die Datenauswertung, -analyse und -interpretation. Dabei wurden die allgemeinen Erfolgsparameter von Programmen und Projekten mit den konkreten Zielen, Inhalten und Wirkungsindikatoren von KIBB verbunden (vgl. KRAPP 2005).

¹⁷ Weitere Informationen zum Analyseleitfaden und zur KIBB-Evaluation können bei Franz SCHAPFEL-KAISER (BIBB) angefragt werden (schapfel-kaiser@bibb.de).

2.4 Literatur

BRANDSTÄTTER, Jochen (1990): Entwicklung im Lebenslauf. In: MAYER, Karl Ulrich (Hrsg.): Lebensverläufe und sozialer Wandel, Sonderheft der Kölner Zeitschrift für Soziologie und Sozialpsychologie). Opladen: Westdeutscher Verlag

CHELIMSKY, Eleanor (1995): New dimensions in evaluation. In: World Bank Operations Evaluations Department (OED): Evaluation and development proceedings of the 1994 World Bank Conference. Washington D.C., S. 3-11

KRAPP, Stephanie (2005): Theoretischer Rahmen der KIBB-Evaluation durch das Centrum für Evaluation. In: KIBB Newsletter 1/2005 (www.kibb.de/Projekte/Berichte/Newsletter)

KROMREY, Helmut (2001): Evaluation – ein vielschichtiges Konzept. Begriff und Methodik von Evaluierung und Evaluationsforschung. Empfehlungen für die Praxis. In: Sozialwissenschaft und Berufspraxis, Nr. 24

MERTENS, Donna M. (2000): Institutionalizing Evaluation in the United States of America. In: STOCKMANN, Reinhard (Hrsg.): Evaluationsforschung. Opladen: Leske und Budrich, S. 41-57

POLLITT, Christopher (2000): Qualitätsmanagement und Evaluation in Europa: Strategien der Innovation oder der Legitimation. In: Müller-Kohlberg u. a. (Hrsg.): Qualität von Humandienstleistungen. Opladen: Leske und Budrich

ROSSI, Peter und Howard E. FREEMAN (1999): Evaluation. A Systematic Approach. 6. Auflage. Thousand Oaks u.a.: Sage

SCHAPFEL-KAISER, F.; WEIS, U.; ZÖLLER, A.: Wissensmanagement in der Berufsbildungsforschung. Das Kommunikations- und Informationssystem berufliche Bildung (KIBB) unterstützt die Arbeitsgemeinschaft Berufsbildungsforschungsnetz (AG BFN). In: BWP 5 / 2005

STOCKMANN, Reinhard (2004): Was ist eine gute Evaluation? Centrum für Evaluation. Saarbrücken. Arbeitspapiere Nr. 9

STOCKMANN, Reinhard (Hrsg.) (2004): Evaluationsforschung. 2. Auflage. Opladen. Leske und Budrich

STOCKMANN, Reinhard (2006): Evaluation und Qualitätsentwicklung. Münster. Waxmann Verlag

VEDUNG, Evert (2000): Evaluation Research and Fundamental Research. In: STOCKMANN, Reinhard (Hrsg.): Evaluationsforschung. Opladen, S. 103-127 (2. Auflage 2004, S. 111-135)

WIDMER, Thomas (2004): Qualität der Evaluation – Wenn Wissenschaft zur praktischen Kunst wird. In: STOCKMANN, Reinhard (Hg.): Evaluationsforschung. 2. Auflage. Opladen, S. 83-111

3. Die Gesellschaft für Evaluation (DeGEval) – Dialogforum zur Professionalisierung von Evaluationen

3.1 Aufgaben

Die Gesellschaft für Evaluation e.V.¹⁸ wurde im Jahr 1997 gegründet und ist ein Zusammenschluss von Personen und Institutionen, die im Bereich der Evaluationen tätig sind (vgl. § 2 der Satzung). Der Verein versteht sich als Antwort auf die wachsende Bedeutung von Evaluationen in einer Vielzahl von Bereichen öffentlichen Handelns; Beispiele sind die Entwicklungspolitik, Schulen und Hochschulen, berufliche und betriebliche Bildung, Strukturpolitik, Medien, Umwelt, Gesundheit, Forschung und Technologie, Verwaltung und Soziale Dienstleistungen (www.degeval.de).

Mit der Unterschiedlichkeit der Evaluationsgegenstände und Evaluationsziele ist aus Sicht der DeGEval eine Differenzierung von Evaluationsansätzen und -methoden zu beobachten: Darüber hinaus geht es nicht mehr nur um die Bewertung von Projekten oder Programmen und deren Wirkungen, sondern um die Verbindung mit Ansätzen des Qualitätsmanagements und der Organisationsentwicklung.

Aus diesen Entwicklungen ergibt sich ein wachsender Diskussionsbedarf über Theorien, Modelle und Methoden bei Personen und Institutionen, die Evaluation erforschen und lehren, Evaluationen durchführen und Evaluationen in Auftrag geben (www.degeval.de).

Die DeGEval verfolgt

- ## die Professionalisierung von Evaluation,
- ## die Zusammenführung unterschiedlicher Perspektiven der Evaluation sowie
- ## Information und Austausch über Evaluation.

Vorstand und themenspezifische Arbeitskreise realisieren diese Ziele durch Empfehlungen und Leitlinien, Veranstaltungen und Veröffentlichungen. Die DeGEval führt selbst keine Evaluationen durch und vertritt keine einheitliche Position zu Ansätzen und Verfahren. Vielmehr versteht sie sich als Dialogforum, in dem in Arbeitskreisen oder auf Jahrestagungen Ansätze, Methoden und Ergebnisse von Evaluationen vorgestellt und diskutiert werden. Darüber hinaus bietet sie Weiterbildungen und Fachveranstaltungen zu den „Standards“ an.

¹⁸ Kürzliche Namensänderung der Deutschen Gesellschaft für Evaluation in Gesellschaft für Evaluation; das Akronym DeGEval wurde jedoch beibehalten.

3.2 Die „Standards für Evaluation“

Die in den USA entwickelten „Standards for Evaluation“ („Joint Committee of Standards for Evaluation“) waren Vorbild für die 2001 von der Mitgliederversammlung der Gesellschaft für Evaluation (DeGEval) verabschiedeten "Standards für Evaluation".¹⁹ Evaluationen sollen demnach vier Eigenschaften aufweisen: **Nützlichkeit, Durchführbarkeit, Fairness** und **Genauigkeit**.

Verbreitung und Anwendung der Standards sollen:

- ≠# die zunehmende Professionalisierung von Evaluation und einen konstruktiven Dialog über deren Nutzen und Qualität erreichen (Qualitätssicherung),
- ≠# die Kommunikation von Evaluatoren/Evaluatorinnen, Auftraggebern/Auftraggeberinnen, Adressaten/Adressatinnen und Beteiligten unterstützen,
- ≠# Orientierungshilfen für Meta-Evaluationen bieten,
- ≠# Transparenz von Evaluationsaktivitäten als professionelle Praxis schaffen.

Die „Standards“ beanspruchen Gültigkeit für die Evaluationspraxis in vielen Bereichen öffentlichen Handelns, für unterschiedliche Evaluationsgegenstände und -zwecke.²⁰ Will man mit den Standards arbeiten, sind jedoch Anpassungen an den Evaluationsgegenstand, inhaltliche Konkretisierungen und Operationalisierungen notwendig. Außerdem ist vorab zu prüfen, welche Standards herangezogen werden können (Passung) sowie deren Intention und Reichweite. Es wird empfohlen, vorab zu prüfen,

- ≠# ob die jeweiligen Standards tatsächlich anwendbar sind; dabei sind Begründungen für oder gegen die Anwendbarkeit der Standards anzugeben,
- ≠# ob Konkurrenzen oder Widersprüche zwischen den Anforderungen von Standards in der konkreten Evaluation bestehen (z. B. Nützlichkeit in forschungsorientierten und wissensgenerierenden Evaluationen, vgl. DJI).

In einem bis 2004 durchgeführten Review haben Institutionen und Personen (Review-Group), die selbst Evaluationen in Auftrag geben oder durchführen, die Standards diskutiert und kritische Stellungnahmen abgegeben²¹. Beurteilt wurde die Passung der Standards für eine

¹⁹ Die Standards können unter www.degeval.de heruntergeladen oder als Broschüre mit Erläuterungen online bestellt werden.

²⁰ Gesellschaft für Evaluation (Hrsg.) 2004: Standards für Evaluation. Zu beziehen über www.degeval.de.

²¹ In dieser Review Group war auch das Deutsche Jugendinstitut (DJI) vertreten, dessen Evaluationsansätze in Kapitel 5 dargestellt werden.

Vielzahl von Evaluationsfeldern, wobei vor allem Klarheit und Verständlichkeit der Formulierungen geprüft wurden.²² Im Hinblick auf das geplante BIBB Projekt ist besonders auf den Standard G 9 (Genauigkeitsstandard) **Meta-Evaluation** hinzuweisen:

„Um Meta-Evaluationen zu ermöglichen, sollen Evaluationen in geeigneter Form dokumentiert und archiviert werden. (...). Solche Meta-Evaluationen dienen dem Zweck, die Qualität von Evaluationsprozessen und -ergebnissen zu beurteilen und zu verbessern. Im Rahmen einer Meta-Evaluation **können** (Hervorhebung d. V.) die hier vorliegenden Standards eingesetzt werden (WIDMER 1996). Eine systematische Meta-Evaluation kann im Rahmen des Evaluationsteams (intern) oder durch Außenstehende (extern) erfolgen. (...).

Darüber hinaus ist die wissenschaftliche Publikation über Evaluationsprojekte und ihre Verfahren, Probleme und Ergebnisse grundsätzlich wünschenswert. Sie fördert den Fortschritt der Evaluation und die Verbreitung von Wissen im jeweiligen Gegenstandsbereich, ist ein Instrument der Qualitätsentwicklung und fördert darüber hinaus die Verbreitung und Akzeptanz von Evaluationen.“ (www.degeval.de und Deutsche Gesellschaft für Evaluation e.V. (Hrsg.) 2004: Standards für Evaluation).²³

Gutachten und Meta-Evaluationen, die sich auf die Standards beziehen, sollten der DeGEval zufolge immer im Kontext der professionellen Reflexion und des kritischen Dialogs stehen. „Die DeGEval fordert Gutachter, die mit den Standards arbeiten auf, anzugeben, welche Standards sie heranziehen und welche nicht, insbesondere, ob sie Beteiligte und Betroffene identifizieren und ob sie eine unparteiische Position beziehen“ (www.degeval.de/Leitlinien für die Erstellung von Gutachten oder Meta-Evaluationen). Darüber hinaus werden Meta-EvaluatorInnen, die auf der Grundlage der Standards arbeiten, gebeten, ihre Ergebnisse der DeGEval für Archivierungs- und Forschungszwecke zur Verfügung zu stellen.

²² Vgl. www.degeval.de/standards Ansprechpartner: Prof. Dr. Hans-Dieter DANIEL, Vorstandsvorsitzender DeGEval und Dr. Wolfgang BEYWL, Koordinator DeGEval-Standards.

²³ Für Meta-Evaluationen und externe Gutachten können auf die „Standards“ bezogene „Leitlinien für die Erstellung von Gutachten oder Meta-Evaluationen“ (www.degeval.de) angefordert werden. Darüber hinaus ist auf die „Leitlinien zur Aus- und Weiterbildung“ und zur „Selbstevaluation“ hinzuweisen.

3.3 Arbeitskreise

Mittlerweile haben sich eine Vielzahl von Arbeitskreisen in 14 Bereichen unter dem Dach der DeGEval konstituiert, die mit den „Standards“ arbeiten. Im Folgenden sind Ziele und Aktivitäten zweier Arbeitskreise, die in engem Bezug zu dieser Synopse stehen, exemplarisch dargestellt:

3.3.1 Beispiel: AK „Evaluation von Entwicklungspolitik“

Der Arbeitskreis „Evaluation von Entwicklungspolitik“ (www.degeval.de) hat sich auf der ersten Jahrestagung der Deutschen Gesellschaft für Evaluation (DeGEval) im September 1998 auf Initiative von Prof. Dr. Reinhard Stockmann konstituiert.

Die zentralen Aufgaben des Arbeitskreises sind:

- ≠# die Förderung des interinstitutionellen Lernens durch dauerhaften Austausch über die Erfahrungen in unterschiedlichen Handlungsfeldern,
- ≠# die Mitwirkung an der Weiterentwicklung von Theorie, Praxis und Methoden der Evaluation durch offene Diskussion über in Planung oder in Anwendung befindliche Konzeptionen und Austausch der jeweiligen Erfahrungen und Ergebnisse,
- ≠# die Mitwirkung an der Erarbeitung eines allgemein akzeptierten Begriffsverständnisses (Definitionen etc.) sowie der Entwicklung von Standards, Gütekriterien und Qualitätsindikatoren als Basis einer gemeinsamen Evaluationskultur.

Als Handlungsziele wurden vereinbart:

- ≠# Entwicklung von Standards und Qualitätskriterien als professionelle Orientierungshilfe für entwicklungspolitische externe Gutachter/-innen (Grundlage/Input für die Standards der Evaluation),
- ≠# Bereitstellung einer Internet-Infobörse mit Materialien zur Evaluation in der EZ (z. B. Evaluierungsleitfäden, Berichtsraster, Evaluierungsberichte etc.),
- ≠# Beiträge zur Fortbildung und Qualifizierung von Gutachtern, zunächst durch eine recherchierbare Zusammenfassung vorhandener Angebote (langfristig möglicherweise sogar durch eigenständige Organisation von Fortbildungsveranstaltungen),
- ≠# Entwicklung praxisbezogener methodischer Handreichungen für Gutachter und Gutachterinnen auf Grundlage der im gegenseitigen Austausch gewonnenen Erkenntnisse. (www.degeval.de)

3.3.2 Beispiel: AK: Berufliche und betriebliche Bildung²⁴

Im AK „Berufliche und betriebliche Bildung“ war unter anderen die AG: „Evaluation von Mehrebenenprogrammen und Netzwerken“ angesiedelt. Die Ergebnisse der AG wurden in drei Veröffentlichungen zur Evaluation der Entwicklungspartnerschaften (EP) im Programm EQUAL präsentiert.²⁵

Eine Erörterung zentraler Ergebnisse findet sich im Kapitel „Ansätze und Verfahren der Programmevaluation des BMWA am Beispiel von EQUAL“. Gleichzeitig reflektieren diese Publikationen den in der DeGEval (und im BMWA/BMAS) geführten Diskussionsprozess zum Thema Evaluation von Netzwerken und Mehrebenenprogrammen.

²⁴ Ansprechpersonen: Christoph Diensberg, Universität Rostock und Dr. Dorothea Schemme, Bundesinstitut für Berufsbildung

²⁵ Voß, Rainer (Hrsg.) (2004): Evaluation von Netzwerken und Kooperationsmanagement. Institut für regionale Innovationsforschung (IRI). Institutsschriften Nr. 25;
KAEWNETARA, Eva und USKE Hans (Hrsg.) (2004): Netzwerkevaluation im Prozess. Schriftenreihe des Rhein-Ruhr-Instituts für Sozialforschung und Politikberatung (RISP) an der Universität Duisburg-Essen. RISP-Texte 1/2004;
POTTER, Philip und Herbert KLEMISCH (Hrsg.) (2003): Evaluationsansätze in Kooperationsvorhaben. Klaus NOVY Institut- KNI Papers 1/03

4. Die Evaluation des Mehrebenenprogramms EQUAL (BMWA/BMAS)

4.1 Partizipativer Ansatz

Mit der EQUAL-Programmevaluation wurden – so das Ergebnis der Recherche²⁶ – drei externe Institutionen beauftragt. Sie bildeten ein Konsortium, das die Programmevaluation auf nationaler Ebene konzipiert und durchgeführt hat. Mit der Gesamtkoordination wurden die ICON-Institute in Köln beauftragt (www.icon-institute.de), Kooperationspartner waren die COMPASS GmbH, Bremen (www.compass-bremen.de) und die PIW GbH in Teltow (www.piw.de). Die wissenschaftliche Gesamtleitung lag in Händen von Dr. Herbert SALLEN, Universität Köln.

Das Team der EQUAL-Programmevaluation

(1. Förderrunde 2003-2005; vgl. www.evaluation-equal.de)

Dr. Herbert Sallen, Uni Köln Wissenschaftliche Gesamtleitung herbert.sallen@uni-koeln.de Köln 0211 - 7201604	Rudi Frey , PIW, Teltow piw-teltow@t-online.de Teltow 03328 - 303011
Axel Weise, COMPASS, Bremen axel.weise@compass-bremen.de Bremen 0421 - 16068 - 31 Dr. Karsten Schuldt, PIW, Rostock	Thomas Gebur, COMPASS, Bremen thomas.gebur@compass-bremen.de Bremen 0421 - 16068 - 30
Henriette Meseke, COMPASS, Bremen henriette.meseke@compass-bremen.de Bremen 0421 - 16068 - 33	Benno Savioli, COMPASS & Feedback, Bremen savioli@feedback-bremen.de Bremen 0421 - 16068 - 0
Nils Pagels, COMPASS, Bremen nils.pagels@compass-bremen.de Bremen 0421 - 16068 - 40	Judith Braun, COMPASS, Bremen judith.braun@compass-bremen.de Bremen 0421 - 16068 - 30
Klaus Kohlmeyer, ProBeruf, Berlin kohlmeyer@proberuf.de Berlin 030 - 78709590	Barbara Wirth-Bauer, ICON-INSTITUTE Gesamtkoordination, köln 0221-93743204 barbara.wirthbauer@icon-institute.de

²⁶ Internetrecherche, Telefongespräche (z. B. mit dem wissenschaftlichen Leiter Dr. Herbert SALLEN, Uni Köln), Literaturanalyse; EQUAL ist nun beim Bundesministerium für Arbeit und Soziales (BMAS) angesiedelt (Nationale Koordinierungsstelle EQUAL, Referat VI a 5; Leiter ist Dr. Michael Heister).

Die systematische Zusammenarbeit von Programmevaluation und EP-Evaluationen war Bestandteil der Leistungsbeschreibung und des Evaluationskonzeptes. Ergebnis der Kooperation ist beispielsweise der „Common Minimum“, ein standardisierter Minimalkatalog von Fragestellungen und Indikatoren, der seitens der EP-Evaluationen bedient wurde.

Der Evaluationsauftrag beinhaltete die Kooperation mit den EvaluatorsInnen der Entwicklungspartnerschaften (EP), dem BMWA/BMAS als Programmverwaltungsbehörde, der Technischen Hilfe, der Steuerungsgruppe des Begleitausschusses, der Europäischen Kommission, den ProgrammevaluatorsInnen der anderen Mitgliedstaaten und der EU-weiten Meta-Evaluation. Diese Zusammenarbeit sollte dazu beitragen, dass Lernprozesse der verschiedenen an der Programmumsetzung beteiligten Akteure und Ebenen nicht verloren gehen und deren Ergebnisse für die Programmevaluation nutzbar sind (www.evaluation-equal.de).

4.2 Evaluation der Entwicklungspartnerschaften (EP)

4.2.1 Vielfalt von Evaluationsansätzen und -verfahren

Alle Maßnahmen innerhalb der Europäischen Gemeinschaftsinitiative (GI) EQUAL wurden in sogenannten Entwicklungspartnerschaften (EP) bzw. Netzwerken zusammengefasst, die wiederum aus unterschiedlichen Teilprojekten bestehen.

Die Evaluation der EP wurde in der ersten Förderrunde von den Programmverantwortlichen nicht einheitlich geregelt. Dementsprechend vielfältig sind die Evaluationsansätze, sowohl im Hinblick auf externe oder interne Evaluationen als auch im Hinblick auf die Designs und verwendeten Evaluationsinstrumente.²⁷ Sie umfassen qualitative und quantitative Untersuchungs- und Analyseinstrumente wie Dokumentenanalyse, Leitfadengespräche, Fallstudien, standardisierte Befragungen der Zielgruppen und ExpertInnen. Die Vielfalt und Heterogenität der Ansätze entspricht dem experimentellen Charakter von EQUAL und wird als programm-funktional bewertet.

Das Evaluationsverfahren der Cluster-Evaluation, das für Mehrebenen-Kontexte prädestiniert ist, wurde in der ersten Förderrunde auf EP-Ebene nicht eingesetzt (und soll auch in der

²⁷ vgl. die Ergebnisse zur Mid-Term-Evaluation im Equal-Programm der AG „Evaluation von Mehrebenen-Programmen und Netzwerken“ im Januar 2004, hrsg. von KAEWNETARA und USKE. RISP-Texte 1/2004)

zweiten Förderrunde nicht verwendet werden), weil dieser Ansatz zu aufwändig und zu kostspielig sei (vgl. dazu POTTER 2004: 41-50).²⁸

4.2.2 EP-Evaluation vermittelt zwischen Teilprojekten und Programmebene

Aufgabe der EP-Evaluation war es, die Teilprojekte zu unterstützen und die Anforderungen des Programms zu vermitteln. Dieses geschah in halbjährigen Workshops mit thematischen Inputs, an denen Trägervertreter/-innen und Projektleiter/-innen teilnahmen. Ein wichtiges Thema war die Selbstevaluation der Teilprojekte und die Einhaltung bestehender Standards zur Selbstevaluation. (Selbstevaluation erfolgte auf der Ebene der Teilprojekte, interne und externe Fremdevaluation auf der Ebene der EP). MÜLLER-KOHLBERG und BEYWEL (2002) haben im Rahmen der DeGEval einen Katalog von Standards zur Selbstevaluation formuliert (www.degeval.de), der auf Projektebene verwendet wird (vgl. SCHMITZ/MAAS 2004: 27).

4.2.3 Interne und externe Evaluationsansätze

Auf der **Ebene der Entwicklungspartnerschaften (EP)** existieren sowohl interne als auch externe Evaluationsansätze, wobei die Programmkonstruktion eine externe Evaluation als Teilprojekt vorsah (vgl. SCHMITZ und MAAS 2004: 24). In der EP „Gender Mainstreaming in der Informationsgesellschaft“ kam es beispielweise zum Konstrukt einer **internen Fremdevaluation**: Es wurde ein separates Evaluationsteam gebildet, das sich als Teilprojekt-Partner und als Teil des Kern- oder Leitungsteams verstand. „Da es selbst nur in der Rolle der Evaluation tätig ist, handelt es sich um Fremdevaluation, im Unterschied zur Selbstevaluation auf der Ebene der Teilprojektevaluation“ (SCHMITZ und MAAS 2004: 24).

Die fehlende Trennschärfe von interner und externer Evaluation scheint für die EP-Ebene typisch und wurde von den Durchführungsverantwortlichen insgesamt eher positiv gesehen. EQUAL-EvaluatorInnen seien, so Katrin BALTES, Gerlinde STRUHKAMP und Klaus KOHLMAYER, zum Teil „wie die Jungfrau zum Kinde zur Evaluation“ gekommen (BALTES u. a. zit. nach KAEWNETARA und USKE 2004: 3). Sie waren Mit-EntwicklerInnen, zentrale IdeengeberInnen für die EP, also Teile des Prozesses, den sie evaluieren (vgl. KAEWNETARA/USKE 2004: 32). Trotz der möglichen Probleme (z.B. fehlende Objektivität), die eine Doppelrolle von Evaluation und Managementfunktion/Mitentwicklungsfunktion, Beratungs- und Unterstützungsfunktion mit sich

²⁸ Cluster Evaluationen sind dort angezeigt, wo es Ähnlichkeiten zwischen den einzelnen Programmelementen gibt, so dass Ergebnisse und Zusammenhänge der verschiedenen Teile/Elemente aggregiert werden können. Dies sei bei EQUAL nur eingeschränkt der Fall, so Dr. Herbert SALLEN, wissenschaftlicher Gesamtleiter des Programms von der Universität Köln. Der (weitgehende) Verzicht der Cluster-Evaluation bei EQUAL habe den „sachlichen Hintergrund“, dass die EP nach Thema und Zielsetzung im wesentlichen Unikate seien.

bringen könnte, überwogen die Stimmen für dieses Verfahren. Die folgenden Argumente für die interne Evaluation beziehen sich überwiegend auf die prozessbegleitende und formative Evaluation der Projekte und EP, können teilweise aber auch auf Ex-post- und Meta-Evaluationen sowie Programmevaluationen übertragen werden.

Argumente für die interne Evaluation:

- ≠# Die Einbeziehung von internen ExpertInnen in das Evaluationsteam bringt einen Mehrwert an Wissen und Erfahrung für die Reprogrammierung; externe EvaluatorInnen oder Institutionen verfügen häufig nicht über das notwendige, EQUAL relevante Know-how, wenn sie bisher noch nicht in irgendeiner beratenden oder operativen Funktion in EQUAL eingebunden waren. Die Gefahr, dass einzelne Personen einen steuernden Einfluss nehmen, wird in komplexen Evaluationsverfahren mit einer größeren Anzahl von BewerterInnen als relativ gering erachtet (vgl. SHEIK/DORR/HECKL (2004): Halbzeitbewertung EQUAL Österreich: 8).
- ≠# Die vielfältigen und wechselnden Aufgaben als MitentwicklerInnen, IdeengeberInnen ProzessbegleiterInnen und Verbindungsglieder zur Programmebene bewirken eine spezifische Qualität von Beobachtungsfähigkeit: Sie ist geprägt durch Erfahrung, Wissen und einen Blick für den Gesamtprozess. Diese Wissens- und Prozessperspektive kann ebenso für Meta-Evaluationen und Programmevaluationen (ex-post) gewinnbringend sein (vgl. KAEWENTARA/USKE (2004: 33). „Objektivität“ ist dabei nicht ausgeschlossen: Sie wird durch eine reflexive Innenperspektive erzeugt.²⁹

²⁹ Bei transdisziplinären (Forschungs-)projekten, wie z.B. *Evalunet*, das vom BMBF gefördert wurde, wird die externe Evaluation und das einfache *Peer Review-Verfahren* aus vergleichbaren Gründen kritisch beurteilt. Für die Evaluation interdisziplinärer und/oder transdisziplinärer Vorhaben seien diese Ansätze unterkomplex; notwendig sei demgegenüber ein Expert-Review, das die Perspektivenvielfalt der Disziplinen einfangen kann. Auch die Multiperspektivität der an den Modellversuchen des BIBB beteiligten Akteure erfordert es, dass im Rahmen einer Meta-Evaluation die Vielfalt der Interessen und Perspektiven, sowie ggf. verschiedener wissenschaftlicher Disziplinen, die sowohl im wissenschaftlichen als auch lebensweltlichen oder unternehmens-bezogenen Kontext angesiedelt sind (Theorie und Praxis), mit einfließt. Ein *Peer-Review* – also eine Programmevaluation von Fachkollegen, wie es etwa bei der DFG lange Zeit üblich war – ist nicht geeignet. In Anlehnung an die Erfahrungen der Evaluation von sozial-ökologischen Forschungsprojekten wäre auch für das BIBB Vorhaben (Projekt 8: Reflexive Meta-Evaluation von Modellversuchen zum betrieblichen Lernen. Laufzeit: 2006-2008) ein *Expert-Review* empfehlenswert: d. h., die Gruppe der Evaluierenden sollte aus ExpertInnen zusammengesetzt sein, die den Evaluationsgegenstand aus einzeldisziplinärer und/oder fachlich integrierter Perspektive beurteilen können und WissenschaftlerInnen, die Erfahrungen oder Expertise in der Durchführung, Konzeption und Bewertung von komplexen Akteur-Settings besitzen (vgl. dazu auch Bergmann, M. u. a., Qualitätskriterien transdisziplinärer Forschung, 2005: 11).

Interne EvaluatorenInnen sind Vertrauensgeber: Eine EP als soziales Netzwerk wird wesentlich durch Vertrauen zusammengehalten. Die evaluierenden Personen/Teams haben die Funktion des Monitorings und können schneller auf Fehlentwicklungen reagieren (PASCHER 2004: 39). Vertrauen spielt auch bei Meta- oder Programmevaluationen eine wichtige Rolle, das bei einer externen Evaluation fehlt; dieser Vertrauensmangel kann zur Reaktanz bis hin zur Boykothaltung der Evaluierten/Evaluanden führen (vgl. SCHMITZ/MAAS (2004: 27). In diesem Zusammenhang wird außerdem auf die Gefahr der „Überevaluation“ (Doppelgleisigkeiten und Mehrfachbelastungen der Evaluanden) durch die verschiedenen Ebenen der Projekt-, der EP- und der Programmevaluation hingewiesen und eine stärkere Integration und Abstimmung gefordert. Notwendig seien koordinierte Kooperationsstrukturen zwischen internen bzw. ModulevaluatorInnen (EvaluatorenInnen der EP) und den ProgrammevaluatorInnen, so wie sie zwischen nationalen und transnationalen EvaluatorenInnen auf europäischer Ebene bereits existieren (vgl. SHEIK/DORR/HECKL 2004: 20).

4.2.4 Vorgaben der Programmebene für die EP- Evaluation in der zweiten Förder- runde: Evaluationen müssen ausgeschrieben werden

Die bisherige Vielfalt der Evaluationsansätze wird sich in der zweiten Förderrunde (2005-2007) „auf Einheitsmaß reduzieren“ (KAEWNETARA und USKE 2004: 3). Denn auf Programmebene wurde beschlossen, dass Evaluationen der EP (auch die begleitende und formative Evaluation) künftig auszuschreiben sind. Damit soll ausgeschlossen werden, dass MitentwicklerInnen, so wie bisher, auch als EvaluatorenInnen tätig sein können. „Aus den vielen formativen Ansätzen werden formatierte, die zwar noch den Auftrag der Prozessbegleitung bekommen werden, aber nicht mehr die dazu nötigen Mittel. Damit erübrigen sich auch (künftig) andere aufwändige Evaluationsvorhaben wie etwa der Ansatz der Cluster-Evaluation für Mehrebenen-programme“ (KAEWNETARA und USKE 2004: 3).

4.3 EQUAL- Evaluation auf Programmebene

Im Folgenden wird die EQUAL-Programmevaluation vom Juni 2003 (Untersuchungsdesign Stufe 2) dargestellt, die sich auf die Leistungsbeschreibung des damaligen BMA und dem darauf basierenden Angebot bezieht. Dopplungen zu den Evaluationsdesigns auf EP-Ebene beruhen darauf, dass die Programmevaluierung auf die EP-Evaluationen Bezug nimmt. Die wesentlichen Unterschiede zu den EP-Evaluationen bestehen im Evaluationsgegenstand (Programm), den daraus abgeleiteten Untersuchungsfragen und darin, dass es sich um eine externe Evaluation handelt.

Der Evaluationsauftrag umfasste die Analyse, Bewertung und Dokumentation der bis zum Jahr 2005 dauernden ersten Förderrunde. Wegen des experimentellen und mehrdimensionalen Charakters der Gemeinschaftsinitiative (GI) EQUAL wurde eine Kombination unterschiedlicher methodischer Untersuchungsansätze sowie empirischer Erhebungsinstrumente eingesetzt – genauso wie bei der Evaluation der einzelnen EP.

4.3.1 Anforderungen an das Evaluationsdesign

Das Evaluationsdesign sollte den Besonderheiten und der spezifischen Programmlogik – im Unterschied etwa zu anderen Strukturfondsinterventionen – gerecht werden. Besonderheiten sind:

-
- ≠# Die experimentelle Ausrichtung,
- ≠# die Förderung von partnerschaftlich-netzwerkartigen Kooperationszusammenhängen (EP-Entwicklungspartnerschaften),
- ≠# die Verbindung von Innovationsentwicklung, transnationaler Zusammenarbeit und Mainstreaming,
- ≠# die Generierung transferierbarer Problemlösungen und die Nachhaltigkeit von Innovationen, die Vorrang haben vor der individuellen Statusverbesserung einzelner Akteure und unmittelbarer arbeitsmarktpolitischer Ergebnisse,
- ≠# die prozesshafte Programmsteuerung und der Verzicht auf kleinteilige Ergebnisvorgaben,
- ≠# das Mehrebenensystem der Programmumsetzung (regional/sektoral, national/transnational).

(vgl. Jahresbericht 2002 Mid-term-Evaluation In: www.evaluation-equal.de)

4.3.2 Überprüfung der Grundannahme: Netzwerkbildung führt zu Innovationsgewinnen?

Der programmatische Grundgedanke von EQUAL war bzw. ist, dass Netzwerkbildung in Form von Entwicklungspartnerschaften mit strategischen und operativen Partnern sowie transnationalen Kooperationen zu Produktivitäts- bzw. Innovationsgewinnen führen. Diese (implizite) Programmtheorie wurde auf der Ebene der Evaluation der EP - aufgrund der Erfahrungen bei der Implementierung - kritisch reflektiert.

Vor allem BALTES und STRUHKAMP (2004) haben den Automatismus von „Netzwerk- Förderung = Innovations-Förderung“, der dem EQUAL-Programm zugrunde liegt, infrage gestellt. Anhand einer Matrix – die als heuristisches Instrument für die EQUAL-Programm Evaluation vorgeschlagen wird (BALTES und STRUHKAMP 2004: 7) - werden die beiden Konzepte „Netzwerke“ und „Innovation“ ins Verhältnis gesetzt. Ergebnis ist, dass der programmatische Ansatz von EQUAL nur dann erfüllt ist, wenn sowohl die Netzwerk-Arbeit gut funktioniert als auch durch das Netzwerk Innovationen entwickelt und verbreitet werden. In der Realität könne letztlich aber auch weder das eine noch das andere funktionieren und die Kooperation fruchtlos sein. „Insofern ist eine verbindliche Programmvorgabe wie „Netzwerk-Bildung“ mit Vorsicht zu genießen, zumindest dann, wenn sie als Vorstufe einer anderen Leistung, hier der Innovations-Generierung, verstanden wird“ (BALTES und STRUHKAMP 2004: 9). Eine solche direkte (kausale) Verbindung anzunehmen, verkenne das vergleichsweise fragile Konzept des Netzwerks, das unberechenbarer sei, als zum Beispiel eine Organisation (ebd.). Aus dieser Sicht sei die Frage nach Innovationen und Innovationsgewinnen für die Meta-Evaluation und/oder Programmevaluation von EQUAL zentral.

Das Evaluationskonzept müsse deshalb alle erfolgskritischen programm-, umsetzungs- und kontextbezogenen Faktoren berücksichtigen und analysieren.

Als zentrale Programmziele werden von den EvaluatorenInnen „Innovation und Mainstreaming“ genannt und die daraus ableitbaren Erfolgs- oder Misserfolgskriterien.

4.3.3 Mix aus verschiedenen Ansätzen und Untersuchungsmethoden

Die Gewährleistung einer gegenstandsadäquaten Programmevaluation erforderte eine kontinuierliche Auseinandersetzung und fortschreitende Entwicklungsarbeit in den Kontexten der Theoriebildung, der Methodendiskussion und der Evaluationsinstrumente.

Evaluationsaufgabe war gleichzeitig die Entwicklung eines gegenstandsangemessenen Evaluationsansatzes, d. h. einer Kombination aus verschiedenen Evaluations- und Untersuchungsmethoden. Der nachfolgende Mix resultiert aus den Evaluationsaufgaben und *Evaluationsparadigmen*.³⁰

³⁰ In Anlehnung an E. CHELIMSKY unterscheidet KROMREY (2001) folgende „**Evaluationsparadigmen**“: Das Forschungsparadigma zur Verbreiterung der Wissensbasis, das Kontrollparadigma zu Kontrollzwecken sowie das Entwicklungsparadigma zu (Programm- und/oder Projekt-) Entwicklungszwecken. Vgl. hierzu die Ausführungen zur wissensgenerierenden Evaluation im Kapitel: Ansätze und Verfahren der Programmevaluation des DJI.

Funktionen von Evaluation: Erkenntnisfunktion, Kontrollfunktion, Dialogfunktion, Legitimationsfunktion (nach STOCKMANN) sind angemessen zu berücksichtigen, wenn sie den EQUAL-Programmevaluationsauftrag berühren.

4.3.4 Kombination formativer/offener Evaluation und summativer Evaluation

Dem Evaluationskonzept liegt das Prinzip der „formativen“, d. h. der programmgestaltenden Evaluation zugrunde. Es eröffnet die Möglichkeit, aus den Erfahrungen der ersten Förderphase zu lernen und Schlussfolgerungen für die darauf folgenden Förderphasen zu ziehen. Dieses Vorgehen wird auch als „offene“ Evaluation bezeichnet: Die Problem- und Fragestellungen sowie das methodische Vorgehen und die Bewertungskriterien der Evaluation stehen nicht vom vornherein fest, sondern werden im Prozess der Evaluation entwickelt. Neben der „formativen“ bzw. „offenen“ Evaluation wird die summative Evaluation zum Beispiel für die Identifizierung arbeitsmarktpolitischer Ergebnisse und Effekte und der Analyse von Monitoringdaten eingesetzt.

Das angelsächsische Konzept der „Cluster-Evaluation“ ließe sich teilweise auf bestimmte Untersuchungsschwerpunkte der GI EQUAL anwenden, da dieser Ansatz speziell für regional verstreute Projekte mit relativ offenen Strukturen entwickelt wurde. Aus Ressourcengründen konnte dieses Konzept nicht genutzt werden, böte aber einen möglichen Zugang für die Analyse und Bewertung der Mainstreamingprozesse (www.evaluation-equal.de Stand: 06/06).

4.3.5 EQUAL- Evaluationsdesign

Das Evaluationsdesign beinhaltet die drei Untersuchungsdimensionen:

- A: Programmkonzeption und Rahmenbedingungen;
- B: Programmdurchführung und Ergebnisse und
- C: Programmwirkungen und Zielerreichung.

Diese sind untergliedert nach Untersuchungsschwerpunkten und passgenauen Untersuchungsmethoden bzw. Evaluationsinstrumenten.³¹ Operationalisierungen und Methodenauswahl werden damit begründet, dass es kein „*Musterdesign für Evaluationen geben kann, sondern je nach Konstellation von Gegenstand und Fragestellungen ‚maßgeschneiderte‘ Vorgehensweisen zu entwickeln und zu begründen sind*“ (www.evaluation-equal.de). Dabei sei auf das gängige Methodeninstrumentarium der quantitativen und qualitativen Sozialforschung zurückzugreifen³¹.

³¹ Verwiesen sei exemplarisch auf: Key issues for the mid-term evaluation of EQUAL CIP in the member states; Guidelines for Systems of Monitoring and Evaluation for the Human Resource Initiative EQUAL in the Period 2000 - 2006; Methodische Arbeitspapiere: Arbeitspapier Nr. 8: Die Halbzeitbewertung der Strukturfondsinterventionen; Methodische Arbeitspapiere, Arbeitspapier 3: Indikatoren für die Begleitung und Bewertung: eine Indikative Methode.

A: Programmkonzeption und der Rahmenbedingungen

Untersuchungsschwerpunkte	Untersuchungsmethoden
<i>Innere Kohärenz des Programms und Zielarchitektur</i>	formative Evaluation, Expertengespräche, Dokumentenanalysen
<i>Sozioökonomischer Kontext und äußere Relevanz</i>	Summative Evaluation, Dokumentenanalysen, sekundärstatistische Analysen
<i>Operationeller Kontext: Steuerbarkeit und Steuerungsfähigkeit</i>	formative Evaluation, Gruppeninterviews, ExpertInnenengespräche, Dokumentenanalysen

B: Programmdurchführung und Ergebnisse

Untersuchungsschwerpunkte	Untersuchungsmethoden
Arbeitsmarktpolitische Effekte	summative Evaluation der Monitoringdaten, Informationen der EP-Evaluationen (quantitativ)

Auf eine umfassende Darstellung soll hier verzichtet werden (siehe www.evaluation-equal.de Stand: 06/06).

C: Programmwirkungen und Zielerreichung (Zielevaluation/Ist-Soll-Vergleich und Wirkungsevaluation)

Untersuchungsschwerpunkte	Untersuchungsmethoden
<i>Innovationen</i>	
-Innovationsansätze	Dokumentation und Bewertung, formative Evaluation, Gruppeninterviews,

	ExpertInnengespräche, Fallstudien, Informationen der EP-Evaluationen (qualitativ)
-Strategische Instrumente	
---Kooperation in EP/Netzwerkansatz	formative Evaluation, Gruppeninterviews, ExpertInnengespräche, Fallstudien, Informationen der EP-Evaluationen (qualitativ)
---Transnationalität	summative Evaluation: Dokumentenanalysen, Monitoringdaten, Fallstudien
---Empowerment	formative Evaluation, Gruppeninterviews, ExpertInnengespräche, Fallstudien
Querschnittsziele	
-Bekämpfung des Rassismus	formative Evaluation, Gruppeninterviews, ExpertInnengespräche, Fallstudien
-Nutzung der Potenziale der IuK- Technologien	formative Evaluation, Gruppeninterviews, ExpertInnengespräche, Fallstudien

Gender Mainstreaming	formative Evaluation: Cluster-Evaluation, Gruppeninterviews, Ex- pertInnengespräche, Fallstudien, Informatio- nen der EP-Evaluationen (qualitativ)
Mainstreaming der Innovationen	formative Evaluation, Cluster-Evaluation, Gruppeninterviews, ExpertInnengespräche, Fallstudien, Informatio- nen der EP-Evaluationen (qualitativ)

(Aus: www.evaluation-equal.de Stand: 06/06)

4.4 Literatur

BALTES, Katrin und Gerlinde STRUHKAMP (2004): Kennzeichen von Netzwerken und Innovationen und Implikationen für Evaluationen am Beispiel der EU-Gemeinschaftsinitiative EQUAL. In: VOß, Rainer, Institut für regionale Innovationsforschung (IRI) (2004): Evaluation von Netzwerken und Kooperationsmanagement. Beiträge zum Workshop der Arbeitsgruppe „Evaluation von Mehrebenenprogrammen und Netzwerken des Arbeitskreises „Berufliche und betriebliche Bildung“ der Deutschen Gesellschaft für Evaluation am 27. Juni 2003. Institutschriften Nr. 25: 2-13

DIENSBERG, Christoph (2004): Probleme des Wissensmanagements als evaluative Herausforderung in Netzwerken. In: VOß, Reiner (Hrsg.)(2004): Evaluation von Netzwerken und Kooperationsmanagement. Institutsschriften Nr. 25: 21-46

DIENSBERG, Christoph und Cornelia RASKOPP (2003): Vorwort. In: POTTER, Philip. und Herbert KLEMISCH (Hrsg.) (2003): Evaluationsansätze in Kooperationsvorhaben. Beispiele aus der Praxis von Entwicklungspartnerschaften im EQUAL Programm. KNi Papers 01/03: 1

HÜTTE, Ursula und Susanne MÄDER (2003): Innova-Entwicklungspartnerschaft für Selbsthilfegenossenschaften. In: POTTER, Philip und Herbert KLEMISCH (Hrsg.) (2003): Evaluationsansätze in Kooperationsvorhaben. KNi Papers 01/03: 33-39

KOHLMEYER, Klaus (2003): Innovation durch Akteurskooperation: Lehren aus der Gemeinschaftsinitiative Beschäftigung. In: POTTER, Philip und Herbert KLEMISCH (Hrsg.) (2003): Evaluationsansätze in Kooperationsvorhaben. KNi Papers 01/03: 9-18

KAEWNETARA, Eva und USKE Hans (2003): Kann Evaluation einen Beitrag zur Förderung von Innovationen in Netzwerken leisten? Das Beispiel zweier Entwicklungspartnerschaften. In: POTTER, Philip und Herbert KLEMISCH (Hrsg.) (2003): Evaluationsansätze in Kooperationsvorhaben. Beispiele aus der Praxis von Entwicklungspartnerschaften im EQUAL Programm. KNi Papers 01/03: 18-27

KAEWNETARA, Eva und USKE Hans (Hrsg.)(2004): Netzwerkevaluation im Prozess. Aktuelle Ansätze in komplexen sozialen Programmen. Schriftenreihe des Rhein-Ruhr-Instituts für Sozialforschung und Politikberatung an der Universität Duisburg-Essen. RISP-Texte 1/2004

KAEWNETARA, Eva und USKE Hans (2004): Halbzeitevaluation von Netzwerken als Co-management. In: Dies. (ed.): Netzwerkevaluation im Prozess. Aktuelle Ansätze in komplexen sozialen Programmen. Schriftenreihe des Rhein-Ruhr-Instituts für Sozialforschung und Politikberatung an der Universität Duisburg-Essen. RISP-Texte 1/2004

POTTER, Philip und Herbert KLEMISCH (2003): Handlungsebenen in der Evaluation der Entwicklungspartnerschaft „kompakt“. In: dies. (ed.): Evaluationsansätze in Kooperationsvorhaben. Beispiele aus der Praxis von Entwicklungspartnerschaften im EQUAL Programm. KNI Papers 01/03: 48-58

POTTER, Philip und Herbert KLEMISCH (2003): Einleitung: Evaluation in Kooperationsvorhaben. In: dies. (ed.). Evaluationsansätze in Kooperationsvorhaben. KNI Papers 01/03: 2-9

SHEIK, Sonja; DORR, Andrea und Eva HECKL (2004): Methoden und Ergebnisse der Halbzeitbewertung EQUAL Österreich. In: KAEWNETARA, Eva und USKE Hans (Hrsg.)(2004): Netzwerkevaluation im Prozess. RISP-Texte 1/2004

POTTER, Philip (2004): Cluster evaluation in multi-level contexts. In: KAEWNETARA, Eva und USKE Hans (Hrsg.)(2004): Netzwerkevaluation im Prozess. RISP-Texte 1/2004

VOß, Rainer (Hrsg.), Institut für Regionale Innovationsforschung (IRI): Evaluation von Netzwerken und Kooperationsmanagement - Beiträge zum Workshop der Arbeitsgruppe „Evaluation von Mehrebenenprogrammen und Netzwerken“ des Arbeitskreises „Berufliche und Betriebliche Bildung“ der Deutschen Gesellschaft für Evaluation am 27. Juni 2003. Institutsschriften Nr. 25 (2004). Zu beziehen über: TFH Wildau, Bahnhofstr., 15745 Wildau, Tel. + Fax: 03375-508214

5. Ansätze und Verfahren der Programmevaluation des Deutschen Jugendinstituts (DJI)

5.1 Allgemeine Gestaltungsgrundlagen

5.1.1 Externe und unabhängige wissenschaftliche Begleitung und Evaluation von Programmen

Das Deutsche Jugendinstitut in München (DJI) ist das bundesweit größte außeruniversitäre sozialwissenschaftliche Forschungsinstitut im Bereich Kinder, Jugendliche, Frauen und Familien. Neben der Forschung nimmt das DJI Informations-, Service- und Beratungsaufgaben wahr und stellt sozialpolitisch relevante Daten und Orientierungshilfen für Politik und Praxis zur Verfügung (www.dji.de Stand: 06/06).

Ein wichtiges Aufgabengebiet des Instituts ist die wissenschaftliche Begleitung und Evaluation von (Modell) Programmen des Bundes (Bundesministeriums für Familien, Senioren, Frauen und Jugend, BMFSFJ) aber auch von Länder- und EU-Programmen. Als externer Evaluator und wissenschaftlicher Begleiter legt das DJI besonderen Wert auf seine Unabhängigkeit. „Wir sind keine Anhänger der Vermischung von Programmsteuerung und Evaluationsverantwortung“, erläutert Karin HAUBRICH von der Konzeptstelle für Evaluation³² in einem Telefoninterview. In vielen Fällen wird das DJI mit der externen Evaluation von Programmen beauftragt, die von anderen entwickelt, durchgeführt und zum Teil auch über Programmagenturen koordiniert werden. In einigen Fällen wird am DJI auch eine wissenschaftliche Konzeptentwicklung durchgeführt. Um Rollenkonflikte³³ zu vermeiden, die die Unparteilichkeit der Evaluation gefährden könnten, werden Lösungen gesucht, die auf eine klare Rollenverteilung abzielen, indem zum Beispiel an der Programmentwicklung nicht beteiligte MitarbeiterInnen für Evaluationen eingesetzt werden. Unparteilichkeit soll auch durch die institutionelle Trennung von Programmentwicklung und -durchführung gewährleistet werden. „Das DJI regt in manchen Fällen Programme an und entwickelt sie, aber andere Träger führen sie durch. Die Rollen sollen in jedem Fall transparent gemacht werden. So wird ein möglicher Bias reduziert, aber auch erkennbar gemacht“ (HAUBRICH).

³² Genaue Bezeichnung der Konzeptstelle für Evaluation: „Strategien und Konzepte externer Evaluation in der Kinder- und Jugendhilfe“ der Abteilung J (Jugend und Jugendhilfe)

³³ Rollenkonflikte entstehen u. a. dann, wenn MitarbeiterInnen Programme evaluieren, die sie (mit) entwickelt haben oder EvaluatorenInnen beratende Aufgaben übernehmen sollen.

Neben der externen Evaluation von Programmen engagiert sich das Institut in der Debatte über Verfahren, Strategien und Konzepte der Evaluation.³⁴ Ziel ist die konzeptionelle, methodologische und theoretische Weiterentwicklung der Evaluation im Feld der Kinder- und Jugendhilfe in Deutschland, wobei auf die internationale Diskussion, insbesondere auf US-amerikanische Ansätze, rekurriert wird.³⁵

5.1.2 Interne und externe Evaluation, Selbst- und Fremdevaluation

Für die Fragestellung der Synopse ist besonders auf die Ausführungen zum Thema Selbst- und Fremdevaluation, interne und externe Evaluation hinzuweisen (vgl. HAUBRICH/LÜDERS 2004:327).

Als **interne Evaluation** wird im DJI eine Evaluation bezeichnet, wenn sie von internen EvaluatorInnen gesteuert und durchgeführt wird, die die Arbeit ihrer Kollegen und Kolleginnen bewerten.³⁶ Wenn die internen EvaluatorInnen nicht an der Programmentwicklung und -steuerung beteiligt sind, weisen sie eine gewisse Unabhängigkeit zum Evaluationsgegenstand auf.³⁷ In der Praxis gibt es bei internen Evaluationen häufig Mischformen in der Verteilung der Programmverantwortung und Evaluationsverantwortung, wodurch fließende Übergänge zur Selbstevaluation entstehen können.

Im Gegensatz dazu werden **externe Evaluationen** von Personen oder Institutionen vorgenommen, die nicht Mitglieder der Organisation sind, in der der Untersuchungsgegenstand angesiedelt ist. Hier besteht also per Definition eine relative Unabhängigkeit von Programmverantwortung und Evaluationsverantwortung (vgl. HAUBRICH/HOLTHUSEN/STRUHKAMP 2005:2).

³⁴ Beispiel hierfür ist das vom BMFSFJ geförderte DJI-Projekt „Strategien und Konzepte externer Evaluation in der Kinder und Jugendhilfe“, in dem die aktuelle Forschungslage gesichtet wird und Vorschläge für die konzeptionelle Weiterentwicklung der Forschung und Praxis erarbeitet werden (vgl. HAUBRICH/LÜDERS 2004: 331).

³⁵ Das DJI ist Mitglied der DeGEval-Gesellschaft für Evaluation und Karin HAUBRICH ist persönliches Mitglied der American Evaluation Association (AEA).

³⁶ An anderer Stelle dieser Synopse (z.B. Kapitel zu Ansätzen des BMBF) wird dies als Peer Review definiert (A. d. V.).

³⁷ Vgl. dazu die Definition interner Evaluation des ISOE (vgl. Ansätze und Verfahren der Programmevaluation des BMBF): Als intern wird hier ein Evaluationsvorhaben definiert, wenn es auf eigene Initiative der jeweiligen Institution erfolgt; dabei ist die Zusammenarbeit von internen Mitarbeitern und externen Evaluatoren/Gutachtern die Regel; BERGMANN (ISOE) verweist in diesem Zusammenhang auf die Grenzen von Peer Reviews in transdisziplinären Projekten.

Die **Selbstevaluation** ist in verschiedenen Praxisfeldern wie der Kinder- und Jugendhilfe (vgl. die Broschürenreihe „Qualitätssicherung in der Kinder- und Jugendhilfe“ des BMFSFJ zit. nach HAUBRICH/LÜDERS 2004: 327) und der beruflichen Weiterbildung (vgl. KREKEL/RASKOPP 2003: 9 zit. nach HAUBRICH/LÜDERS 2004: 327) anzutreffen.³⁸ Besonderes Kennzeichen der Selbstevaluation – so wie sie im Feld der Kinder- und Jugendhilfe verstanden wird – ist, dass die für die Programmplanung und -umsetzung Verantwortlichen auch Evaluationsverantwortung tragen.

Sowohl interne als auch externe Evaluation werden im DJI als **Fremdevaluation** definiert. Mit der Definition der Begriffe interne und externe sowie Selbst- und Fremdevaluation knüpfen Haubrich und Lüders an das vorherrschende Begriffsverständnis in der Kinder- und Jugendhilfe ein, wie es von Maja HEINER, Wolfgang BEYWL und Hiltrud VON SPIEGEL geprägt worden ist. Gleichwohl wird darauf hingewiesen, dass es andere Definitionen dieser vier Begriffe gibt, die auf der Unterscheidung zwischen dem Ort der Steuerung einer Evaluation (Selbst-, Fremdevaluation) und dem Ort der Umsetzung (interne, externe Evaluation) beruhen.³⁹ Eine solche Differenzierung zwischen Steuerung und Durchführung von Evaluationen wird als hilfreich eingeschätzt, um die faktischen Verteilungen steuerungsrelevanter Kompetenzen auf interne und externe Akteure genauer in den Blick zu nehmen. In der Realität sind die Kompetenzen der Evaluationssteuerung jedoch häufig komplex angelegt. Nicht selten gibt es extern und intern besetzte Steuerungsrunden für Evaluationen. Daher sei - unabhängig von den gewählten Begrifflichkeiten - eine höhere Transparenz der Steuerungsarchitektur notwendig (vgl. HAUBRICH/LÜDERS 2004: 327).

Welcher Typ von Evaluation sinnvoll ist, müsse im Einzelfall entschieden werden. Dabei sei zu berücksichtigen, dass der jeweilige Evaluations-Typ „Konsequenzen hinsichtlich der Art der Informationen, die gewonnen werden können, der Reichweite und Gültigkeit der Ergebnisse sowie der benötigten Voraussetzungen und Ressourcen (hat)“ (HAUBRICH/HOLTHUSEN/STRUHKAMP 2005: 2).

³⁸ Auch in der GTZ werden die Begriffe Selbstevaluierung und Fremdevaluierung, interne und externe Evaluierung verwendet (vgl. Kapitel: Ansätze und Verfahren der Programmevaluation in der GTZ).

³⁹ Vgl. Ansatz der GTZ.

5.1.3 Partizipativer Ansatz: Unterschiede in der Einbindung der Beteiligten und Betroffenen

Wie bei allen anderen hier vorgestellten Institutionen (CEval, BMWA/BMAS, GTZ, BMBF), werden auch im DJI *zum Teil* partizipative Evaluationsansätze gewählt. „Für die Kinder- und Jugendhilfe ist vor allem die Frage der Einbindung bzw. der Partizipation von Beteiligten und Betroffenen in Evaluationen bedeutend. Denn Partizipation ist als Gestaltungsprinzip sozialer Interaktion in der Kinder- und Jugendhilfe selbst grundlegend. Dies entspricht nicht nur dem Selbstverständnis der professionell Tätigen in diesem Feld, sondern ist auch im Hinblick auf den Einbezug von Kindern und Jugendlichen gesetzlich im Kinder- und Jugendhilfe-Gesetz (KJHG) festgeschrieben“ (HAUBRICH/HOLTHUSEN/STRUHKAMP 2005: 4).

Unter dem Konzept „Partizipativer Ansatz“ können sich allerdings unterschiedliche Formen und Verfahren der Einbindung von Beteiligten und Betroffenen versammeln. Für das Evaluationsdesign maßgeblich sind deshalb Fragen nach der Steuerung der Evaluation, nach den Beteiligtegruppen und zu welchem Zweck und in welcher Intensität sie eingebunden werden sollen. „Im Zusammenspiel dieser drei Dimensionen ergeben sich diverse Formen der Partizipation, die Verlauf und Ergebnisse einer Evaluation prägen“ (COUSINS/WHITMORE 1998 zit. nach HAUBRICH/HOLTHUSEN/STRUHKAMP 2005: 4).

5.1.4 Multidisziplinäres und plurales methodisches Selbstverständnis

Die Breite des Forschungsinstituts, weitgefächerte Aufgabenstellungen und Evaluationsgegenstände korrespondieren mit einem multidisziplinären und pluralen methodischen Selbstverständnis. Über 140 wissenschaftliche MitarbeiterInnen decken ein breites Forschungsspektrum in fünf Abteilungen⁴⁰ und einem Forschungsschwerpunkt ab.⁴¹

Zum Einsatz kommen je nach Untersuchungsgegenstand und Fragestellung sowohl quantitative als auch qualitative Methoden der empirischen Sozialforschung, interaktive Sozialforschung und Handlungsforschung (früher häufiger, heute nur noch selten), Cluster-Evaluation, Empowerment Evaluation, Responsive Evaluation, Kosten-Nutzen-Analysen, Outcome-Evaluationen oder Programmtheorie-Evaluationen. Die Durchsicht der DJI-Veröf-

⁴⁰ Die fachlichen Abteilungen des DJI sind: Kinder und Kinderbetreuung, Jugend und Jugendhilfe, Familie und Familienpolitik, Geschlechterforschung und Frauenpolitik, Zentrum für Dauerbeobachtung und Methoden; außerdem gibt es einen Forschungsschwerpunkt Übergänge in Arbeit.

⁴¹ Damit unterscheidet sich das DJI zum Beispiel vom Centrum für Evaluation (CEval), das einen vergleichsweise „einheitlichen Ansatz“ vertritt (HAUBRICH, Telefoninterview).

fentlichungen zu Evaluationen bestätigt die Pluralität von Ansätzen: Es finden sich Empowermentansätze (advocative Ansätze), systemische und handlungstheoretische Ansätze, Verknüpfungen externer Evaluation mit Ansätzen zur Selbstevaluation, die Cluster-Evaluation und programmtheoriebasierte Evaluationen.⁴²

Die Wahl und Kombination von Ansätzen und Methoden sollte sich dabei am Evaluationsgegenstand, den Zielen und den Aufgaben einer Evaluation sowie der Entwicklungsphase oder „Reife“ eines Programms orientieren⁴³ (vgl. HAUBRICH/LÜDERS 2004: 329 bis 331).

⁴² Mit dem Ansatz der „Cluster Evaluation“ wurde an die internationale Diskussion angeknüpft. Er greife aber nur bei multizentrischen Programmen bis zu maximal 30 Projekten; der Ansatz müsse weiterentwickelt und für Deutschland anschlussfähig gemacht werden (Haubrich, Telefoninterview).

⁴³ Bezugnehmend auf OWEN/ROGERS 1999 werden von HAUBRICH und LÜDERS fünf Aufgabentypen von Evaluation unterschieden:

Die **proaktive Evaluation** erhebt den Bedarf für ein geplantes Programm, good practice Beispiele, den aktuellen Forschungsstand, um die Konzeption des Programms abzusichern oder zu verbessern.

Die **klärende Evaluation** hat den Zweck das Konzept zu verfeinern und zu konkretisieren (in der Anfangsphase der Programmimplementation). Fragen sind, ob das Programmkonzept schlüssig ist oder die Rekonstruktion der „impliziten Theorie, wie bestimmte Programmelemente zu bestimmten Ergebnissen führen sollen und deren Überprüfung und Weiterentwicklung im Hinblick auf Kontextbedingungen“ oder die Überprüfung der Evaluierbarkeit.

Die **interaktive Evaluation** hat den Zweck das Programm während der Laufzeit zu verbessern (= formative Evaluation). Konkret genannte Evaluationsansätze sind Responsive (antwortende) Evaluation, Qualitätsüberprüfung, Empowerment Evaluation, Cluster Evaluation und interaktive Sozialforschung beziehungsweise Handlungsforschung (vgl. HAUBRICH/LÜDERS 2004: 330). Diese Ansätze liefern Erkenntnisse über das „Wie“ programmbezogener Umsetzungsformen, ob sie funktionieren und wie sie weiterentwickelt werden können damit das Programm effektiver gestaltet wird (vgl. HAUBRICH/LÜDERS 2004: 329). **Monitoring** dient der Feinsteuerung des Programms und die Wirkungsevaluation der bilanzierenden Beurteilung der erreichten Ziele und (nicht) intendierter Wirkungen (HAUBRICH/LÜDERS 2004: 330)

Impact Evaluationen werden am Ende oder nach Ablauf eines Programms durchgeführt, wenn dessen Umsetzung bereits ausreichend erprobt worden ist und klare Kenntnisse über die Konzeption und Implementation des Programms vorliegen. Sie dienen in der Regel der bilanzierenden Beurteilung, ob ein Programm seine angestrebten Ergebnisse erreicht hat oder nicht. Typische Fragen sind dabei, ob das Programm so wie geplant implementiert worden ist, ob die angestrebten Ziele erreicht wurden, welche nicht-intendierten Wirkungen aufgetreten sind, ob und in welchem Ausmaß die Intervention zu den festgestellten Ergebnissen geführt hat, welche Auswirkungen unterschiedliche Formen der Implementation auf die Ergebnis-erreichung haben oder ob das Programm kosteneffizient gearbeitet hat.

5.1.5 Summative, formative und wissensgenerierende Evaluation

Die Unterscheidung von summativer (bilanzierender) und formativer (gestaltender) Evaluation geht zurück auf M. SCRIVEN (SCRIVEN 1991 a, 1996 zit. nach HAUBRICH/LÜDERS 2004:), wobei die beiden Begriffe unterschiedlich ausgelegt, definiert oder konnotiert werden. So betonen Wissenschaftler wie M. Q. PATTON oder T. WIDMER die Nutzenorientierung und den Bezug des Begriffspaars auf unterschiedliche Verwendungsformen („uses“) von Evaluation.⁴⁴ HAUBRICH und LÜDERS sehen die Begriffe primär mit Zielen und Absichten der Evaluation verbunden, wobei die Entscheidung „summativ oder formativ“ Folgen für „das Evaluationsdesign, die Fragestellungen, relevante Indikatoren, Erhebungs- und Auswertungsverfahren, die Form der Ergebnisvermittlung und die Art der Berichterstattung“ hat (HAUBRICH/LÜDERS 2004: 328).

Sowohl bei der formativen als auch der summativen Evaluation steht der unmittelbare instrumentelle Nutzen der Ergebnisse im Vordergrund. Bei der formativen Evaluation, die primär bei der Programmentwicklung und -durchführung eingesetzt wird, besteht dieser Nutzen in der Verbesserung des laufenden Programms, bei der summativen Evaluation, die in der Regel nach Abschluss eines Programms durchgeführt wird, besteht er darin, Entscheidungen über die Fortführung oder Modifikation eines Programms zu treffen.

Eine dritte Form von Evaluation, die sich nicht am instrumentellen Nutzen oder der unmittelbaren Verwendbarkeit orientiert, ist die sogenannte **wissensgenerierende Evaluation** nach M. Q. PATTON. Sie zielt darauf ab, Bewertungen und gewonnene Erfahrungen eines Programms als Grundlage für die Gestaltung späterer Programme und Maßnahmen verwenden zu können; insofern ist sie von konzeptionellem Nutzen (vgl. PATTON 1997 nach HAUBRICH/LÜDERS 2004: 328). Da der (spätere) Verwendungs- bzw. Verwertungskontext der Evaluationsergebnisse nicht vorhersehbar ist, bildet die wissensgenerierende Evaluation eine eigenständige Kategorie (ebd.)⁴⁵.

⁴⁴ M.Q. PATTON (1997) definiert formativ und summativ als unterschiedliche „uses“ von Evaluation; auch T. WIDMER (2000, S.80) betont die Nutzungsdimension der Begriffe (zit. nach HAUBRICH/LÜDERS (2004): Evaluation – mehr als ein Modewort? S.328)

⁴⁵ Zu vertiefen wäre m. E. hier dennoch die Frage nach der Trennschärfe von wissensgenerierender und summativer/formativer Evaluation, denn auch summative und formative Evaluation generieren Wissen, das von konzeptionellem Nutzen sein könnte. Werden summative und formative Evaluation als **Verfahren** definiert, denen bestimmte Methoden und Instrumente zugeordnet werden, so stellt sich auch die Frage nach davon unterscheidbaren, trennscharfen Verfahren/Methoden/Instrumenten einer wissensgenerierenden Evaluation. Werden die Begriffe allerdings wie hier, als **Zwecke** definiert, können sie mehr oder weniger trennscharf sein (A.d.V.).

Daraus könnte man schlussfolgern, dass der wissensgenerierenden Evaluation, deren Zweck nicht die unmittelbare instrumentelle Nutzung des generierten Wissens ist (vgl. HAUBRICH/HOLTHUSEN/STRUHHKAMP 2005: 2) auch kein instrumentell verengter, auf den unmittelbaren Nutzen bezogener Wissens- oder Erkenntnisbegriff zugrunde liegen kann (A.d.V.).

Das DJI strebt vor allem wissensgenerierende Evaluationen an, da bei Bundesmodellprogrammen die Frage der Übertragbarkeit zentral ist. (Formative und summative Elemente können hier durchaus enthalten sein). Das generierte Wissen soll der Weiterentwicklung der Fachlichkeit in der Kinder- und Jugendhilfe, wie auch der Familienpolitik und Familienhilfe dienen, allgemein und nicht nur situativ gültig und damit übertragbar sein.⁴⁶ Wissensgenerierung zielt damit vor allem auf die Weiterentwicklung des fachlichen und konzeptionellen Know-how für die Fachpraxis und Politik ab (vgl. HAUBRICH 2004: 229). Das heißt konkret: Es geht darum, zu gültigen Aussagen darüber zu gelangen, „warum unter gegebenen Rahmenbedingungen bestimmte Arbeitsformen erfolgreiche Ansätze zur Lösung eines gesellschaftlichen Problems darstellen oder auch nicht“ (HAUBRICH/HOLTHUSEN/STRUHKAMP (2005). Dieser Anspruch korrespondiert mit der überregionalen Anlage und dem bundesweiten Interesse aller wissenschaftlichen Begleitungen und Programmevaluationen für das BMFSFJ.⁴⁷ In diesem Zusammenhang sei auf die Problematik des Transferanspruchs von Modellprogrammen hinzuweisen (HAUBRICH), die hier jedoch nicht vertieft werden soll.

5.1.6 Verwendungstheoretische Probleme der Nutzenorientierung:

Kritische Auseinandersetzung mit den Nützlichkeits-Standards der DeGEval

Für Evaluationen maßgeblich seien die Standards des wissenschaftlichen Arbeitens⁴⁸ und die „Standards für Evaluation“ der DeGEval. „Wir vertreten die Standards, es gibt aber auch eine kritische Auseinandersetzung dazu“ (HAUBRICH, Telefoninterview). Insbesondere das Nutzenpostulat, das in den Nützlichkeits-Standards der DeGEval unterstrichen wird, werde diskutiert. Primär gehe es um die Gewinnung wissenschaftlich fundierter Erkenntnisse.⁴⁹ Den Nutzungsinteressen der Beteiligten müsse zwar Rechnung getragen werden, indem die „erwartbaren Verwendungskontexte und die möglichen Nebenfolgen für andere Beteiligte und Betroffene berücksichtigt werden“ (HAUBRICH/LÜDERS 2004:324). Es könne aber leicht zur Überforderung eines Evaluationsvorhabens führen, alle impliziten, expliziten und möglicherweise konfligierenden Nutzungsinteressen zu berücksichtigen (vgl. HAUBRICH/LÜDERS/STRUHKAMP 2005: 8).

⁴⁶ Auch hier zeigt sich, dass die wissensgenerierende Evaluation nicht zweckfrei ist, sondern die Erzeugung nützlicher und verwertbarer Erkenntnisse im Blick hat. Der Nützlichkeitsstandard wird also durchaus bedient (siehe dazu weiter unten).

⁴⁷ Das DJI wird zu einem hohen Anteil vom Bund (BMFSFJ) finanziert.

⁴⁸ Die wissenschaftliche Evaluationsforschung ist an möglichst verallgemeinerbaren Aussagen interessiert und die Verfahren orientieren sich an den Standards wissenschaftlich-systematischer Erkenntnisgewinnung; leitend sind die Regeln und Gütekriterien sozialwissenschaftlicher Forschung (vgl. HAUBRICH/LÜDERS 2004: 326, Fußnote 18)

⁴⁹ Die Erkenntnisfunktion wird auch im CEVAL (Prof. STOCKMANN) als Ziel von Evaluation genannt. Die Kombination von Erkenntnisfunktion und Nützlichkeitsanspruch hat einen instrumentalisierten Begriff von Erkenntnis zur Folge (A.d.V.).

Außerdem erfordere die Beurteilung der Nützlichkeit der Ergebnisse einen voraussehbaren Verwendungskontext, was „nach aller Erfahrung sehr selten und nur in begrenztem Maße der Fall ist“ (HAUBRICH/LÜDERS 2004: 324). Denn sozialwissenschaftlich generiertes Wissen werde von der Praxis autonom genutzt.⁵⁰

Aufgrund der verwendungstheoretischen Probleme, die ein auf Nutzenorientierung ausgerichtetes Evaluationsverständnis aufwirft, plädieren HAUBRICH und LÜDERS (vgl. HAUBRICH/LÜDERS 2004: 324) für die deutliche Unterscheidung zwischen den konkreten Verwertungsinteressen des Auftraggebers oder anderer Beteiligter (Nützlichkeit) und dem allgemeinen Anspruch an Evaluationen, nutzbare und brauchbare Ergebnisse (Nutzbarkeit) zu erzielen.

5.1.7 Wirkungsevaluation: nur bei ausgereiften Programmen sinnvoll?

Die Evaluation von Wirkungen ist in allen Politikfeldern (also nicht nur in der Kinder- und Jugendhilfe) ein zunehmend wichtiges Ziel. Im Sinne einer Überprüfung von Wirkungsannahmen sei diese, so HAUBRICH u. a., jedoch nur bei ausgereiften, bereits bewährten, nicht aber bei „unreifen“ Modellprogrammen sinnvoll „Wenn man Modellprogramme als „exemplarische Feldexperimente“ versteht, in deren Rahmen neue Strategien, Ansätze, Konzepte erprobt und entwickelt werden können und stellvertretend für den jeweiligen Bereich Erfahrungen, insbesondere mit innovativen Ansätzen gesammelt werden sollen, um daraus Anregungen zur Weiterentwicklung fachlicher Praxis zu gewinnen, kann man nicht zugleich weitreichende Erwartungen hinsichtlich ihrer Wirksamkeit hegen“ (HAUBRICH/LÜDERS/STRUHKAMP 2005: 6).

Da die Wirkungsfrage jedoch auch bei der Evaluation von (innovativen oder „unreifen“) Modellprogrammen gestellt wird, bei denen die gängigen sozialwissenschaftlichen Verfahren der Wirkungsforschung (Kausalanalysen) aber nicht greifen, könne die Rekonstruktion des internen Zusammenspiels der Programmelemente – zum Beispiel mithilfe der Cluster-Evaluation oder Ansätzen der Programmtheorie-Evaluation - aufschlussreich sein (HAUBRICH/LÜDERS/STRUHKAMP 2005:7). Diese „zum Teil kleinteilige Rekonstruktion der Programmelemente und ihres konzeptionellen Zusammenwirkens“ (ebd.) sei Voraussetzung und Basis für regional und zeitlich begrenzte Wirkungsanalysen von Programmen (ebd.).

Ohne das „innere“ Verständnis, wie Aktivitäten über bereitgestellte Angebote zu bestimmten Ergebnissen führen sollen, bliebe das grundsätzliche Problem jeder Wirkungsanalyse bestehen: Dass beobachtete Effekte zwar identifiziert, nicht aber als kausal auf das jeweilige Programm zurückgeführt werden können (ebd.: 6).

⁵⁰ Die autonome Nutzung der Evaluationsergebnisse ist nur ein Argument gegen die Evaluierbarkeit und Nützlichkeit, nicht aber gegen die Möglichkeit, nützliche Erkenntnisse zu erzielen.

5.1.8 Meta-Evaluation und Evaluations-Synthese

Der Begriff Meta-Evaluation wird in den Leitlinien für Gutachten und Meta-Evaluationen der DeGEval als Evaluation von Evaluation(en) definiert und sei damit „besetzt“ (HAUBRICH).⁵¹ Ziel der Meta-Evaluation ist dieser Definition zufolge die Überprüfung der Qualität von Evaluationen. Im Fokus der Meta-Evaluation stehen die benutzten Verfahren, Kriterien und Ergebnisse einer oder mehrerer Evaluation, die anhand von vorher festgelegten Standards oder Gütekriterien (wie zum Beispiel den „Standards für Evaluation“) überprüft und bewertet werden sollen.⁵²

Von der Meta-Evaluation wird die Evaluations-Synthese unterschieden, die sich auf bereits existierende Evaluationsergebnisse aus vorliegenden Studien bezieht. So könnte zum Beispiel am Ende eines Programms eine Evaluationssynthese von Ergebnissen der Evaluationen einzelner Projekte vorgenommen werden.

„Von Forschungs- und Evaluationssynthesen wird dann gesprochen, wenn der Untersuchungsgegenstand nicht ein einzelnes Programm ist, sondern die Erkenntnisse aus mehreren Studien in einem thematischen Feld zusammengefasst werden sollen. Diese können sich auf die Auswertung von Ergebnissen allein beziehen, oder auch auf Evaluationskonzepte, eingesetzte Methoden u. a. Als ein spezifisches Verfahren der (statistischen) Zusammenfassung von Forschungs- und Evaluationsergebnissen kann hierzu unter bestimmten Voraussetzungen eine Meta-Analyse durchgeführt werden“ (HAUBRICH/HOLTHUSEN/STRUHKAMP 2005: 4).

5.2 Beispiel „Cluster-Evaluation“ als ein im DJI weiterentwickelter Ansatz

5.2.1 Anwendungsbereiche: Komplexe und multizentrische Programme

In ihrem Beitrag „Cluster-Evaluation – Wirkungen analysieren und Innovationen fördern“ (HAUBRICH 2004a) erläutert Haubrich Anwendungsbereiche und Besonderheiten der Cluster Evaluation.⁵³ Der aus den USA stammende Ansatz wurde erstmals 1998 vom DJI in der

⁵¹ In den DeGEval Standards finden sich die entsprechenden Definitionen und Erläuterungen zur Meta-Evaluation unter G9 (Genauigkeitsstandards).

⁵² Vgl. dazu WIDMER, Thomas (1996): Meta-Evaluation. Bern u. a.

⁵³ Die Cluster-Evaluation ist nur einer von vielen im DJI angewendeten Evaluationsansätzen, wenngleich der prononcierteste.

wissenschaftlichen Begleitung und Evaluation von Bundesmodellprogrammen in der Kinder- und Jugendhilfe angewendet.⁵⁴

„Cluster“ meint eine Gruppe von heterogenen und weitgehend autonomen Projekten mit gleichen Zielen, Strategien und/oder gleichen Zielgruppen oder eine Gruppe von Projekten, die „gemeinsam auf einen *systemischen* Veränderungsprozess in Institutionen und Fachpolitik hinwirkt“ (W.K. KELLOGG FOUNDATION 1995: 5 zit. nach HAUBRICH 2004a: 156).

Anwendungsbereiche der Cluster-Evaluation sind:

- ## die Evaluation komplexer, multizentrischer und Mehrebenen-Programme mit regional autonomen und inhaltlich sowie strukturell heterogenen Projekten, die auf *systemische* Entwicklungsprozesse abzielen.⁵⁵
- ## Die Evaluation explorativer Initiativen (formative Evaluation);
- ## Die Evaluation von Modellprogrammen auf Landes- oder Bundesebene.

5.2.2 Konzeptionelle Lösungen für konfligierende Evaluationsanforderungen

In den Evaluationsanforderungen an Modellprogramme treten häufig Zielkonflikte auf. Beispiele sind die geforderte Kombination formativer, summativer und wissensgenerierender Evaluationsleistungen sowie die Wirkungsanalyse bei noch nicht ausgereiften, entwicklungs-offen angelegten Programmen. Wie bereits weiter oben erläutert, zielt die Cluster-Evaluation vor allem darauf ab, Wirkungen zu analysieren und die Entwicklung von Innovationen zu fördern. Damit beansprucht der Ansatz, konzeptionelle Lösungen bereit zu stellen, diese teilweise konfligierenden Anforderungen zu bewältigen (ebd.:157) und die unterschiedlichen Interessen der Programmbeteiligten sowie der Fachöffentlichkeit zu berücksichtigen (ebd.: 159). Das Evaluationsdesign soll dem wissenschaftlichen Gütekriterium der Wahrheitsfindung und dem Gütekriterium der Nützlichkeit, die sich teilweise auch widersprechen können, Rechnung tragen (ebd.). Die Schwierigkeiten, die sich daraus ergeben, wurden weiter oben bereits erwähnt.

⁵⁴ in dem Modellprogramm "Mobile Jugendsozialarbeit für junge Menschen ausländischer Herkunft" (www.dji.de/materialien_zur_Qualitätssicherung_in_der_Kinder-_und_Jugendhilfe)

⁵⁵ Ein prominentes Beispiel für ein solches komplexes Mehrebenen Programm ist EQUAL (BMWA); hier wurde der Ansatz aufgrund seiner Aufwändigkeit jedoch nicht eingesetzt. Dies ergaben meine Recherchen für die EP- Evaluation in Deutschland und die deutsche Programmevaluation.

Während viele multizentrische Programmevaluationen Unterschiede in der Implementierung eines Programms als Störfaktor wahrnehmen und deshalb versuchen, Abweichungen zu kontrollieren, will die Cluster-Evaluation von diesen Unterschieden profitieren (ebd.: 161). „Eine Vereinheitlichung der Programmimplementation (wäre) im Sinne der Programmzielsetzungen geradezu kontraproduktiv. Gleichwohl sind Cluster-Evaluationen wirkungs- und ergebnisorientiert“ (vgl. HAUBRICH 2004a: 161).

Untersucht werden die „relativen Wirkungen der lokalen, den jeweiligen Rahmenbedingungen angepasste Implementationen. [...] Durch vergleichende Betrachtung einer Gruppe von Projekten werden gemeinsame Themen und Zusammenhänge auf der Programmebene identifiziert, Erfolge und Misserfolge analysiert“ (HAUBRICH 2004a: 161); so können die Projekte aus den Ergebnissen der Programmevaluation lernen und ihre Arbeit verbessern (formative Funktion) (ebd.). Die wissensgenerierende Leistung besteht darin, dass Fachpraxis und -politik die gesammelten Erfahrungen für die Weiterentwicklung des Arbeitsfeldes nutzen und über zukünftige Programme oder gesetzliche Initiativen entscheiden können. (Letzteres wäre auch eine summative Leistung). Formative, wissensgenerierende und summative Evaluation erfordern jedoch Prioritätensetzungen und Kompromisse im Evaluationsdesign aufgrund einhergehender möglicher Konflikte (dies.: 168).

Inkonsistenzen und unterschiedliche Interpretationen sollen möglichst frühzeitig während des Evaluationsprozesses aufgedeckt und den Programmbeteiligten rückgespielt werden. „Partizipation, Austausch und Kooperationen finden statt auf Netzwerktagungen, telefonisch, schriftlich oder in Internet-Foren sowie in separaten Foren für „heikle Themen“ (dies.: 165). Die endgültigen Schlussfolgerungen aus der Evaluation werden den Programmbeteiligten rückgekoppelt und kommunikativ validiert, insbesondere hinsichtlich der Genauigkeit und Vollständigkeit (ebd.).

5.2.3. „Logic Models“ für die Analyse interner Wirkungszusammenhänge

Die Beschreibung der „internen Logik“ eines Programms, beziehungsweise „das Verständnis darüber, wie Veränderungen erreicht werden sollen“ und wie sich das interne Zusammenspiel der Programmelemente gestaltet, sind Voraussetzungen, Aussagen über Programmwirkungen machen zu können (vgl. HAUBRICH/HOLTHUSEN/STRUHKAMP 2005: 3). Logische Modelle („Logic Models“) sind ein Instrument um dieses Zusammenspiel zu beschreiben und grafisch abzubilden. Dabei geht es auch darum, die häufig impliziten Programmannahmen zu klären. Logische Modelle bringen die Faktoren „Input“ (personelle, finanzielle Ressourcen) – „Output“ (Produkte, Dienstleistungen) – „Outcome“ (Ergebnisse) – „Impact“ (langfristige Wirkungen) in

Form von Wenn-dann-Aussagen in einen logischen Zusammenhang (dies.: 3,4).⁵⁶ Damit ist eine einfache Form des logischen Modells beschrieben. Diese Modelle können jedoch – wie die Entwicklungen in den USA, Kanada und Australien zeigen – je nach Gegenstand und Forschungsfrage vielfältige, auch komplexe Formen annehmen.

5.3 „Intime“ Kenntnisse und Vertrauen fördern qualifizierte Evaluation

Grundsätzlich gilt: (Externe) EvaluatorInnen, die in der formativen und prozessbegleitenden Evaluation involviert waren, zum Beispiel in Form der wissenschaftlichen Begleitung, können auch in summative Evaluationen einbezogen werden. Diese Rollen-Kombination kann folgende Vorteile haben:

- ≠# Erstens fördert die genaue, innere („intime“) Kenntnis des Evaluationsgegenstandes eine qualifizierte summative Beurteilung, insbesondere bei komplexen Programmen, die durch „Nicht Eingeweihte“ nur schwer möglich wäre.
- ≠# Zweitens fördert das aufgebaute Vertrauen während der wissenschaftlichen Begleitung die Wahrscheinlichkeit, glaubwürdige Informationen zu erhalten.

Diese Einschätzung kann auch auf interne Evaluationen übertragen werden (A.d.V.). Wichtig sei jedoch eine klare Rollendefinition, „die zwischen Rückkopplung von Ergebnissen und Beratung der Projekte oder gar Programmsteuerung unterscheidet und eine kritische Reflexion der Grenzen zwischen Evaluation und beratender Einflussnahme in einem Evaluationsteam [...], [um] die notwendige Distanz der Evaluatorinnen und Evaluatoren zu erhalten“ (dies.: 168).

⁵⁶ Ob einfache mechanistische Kausalmodelle geeignet sind, systemische oder dialektische Wirkungszusammenhänge zu erfassen, kann an dieser Stelle nicht diskutiert werden (A.d.V.).

5.4 Literatur

ARBEITSSTELLE KINDER- UND JUGENDKRIMINALITÄTSPRÄVENTION (Hrsg.): Evaluierter Kriminalitätsprävention in der Kinder- und Jugendhilfe. Erfahrungen und Ergebnisse aus fünf Modellprojekten. München: DJI 2003, 215 S.

BEYWL, W., SCHEPP-WINTER, E.: Zielgeführte Evaluation von Programmen – ein Leitfaden, QS – Materialien zur Qualitätssicherung in der Kinder- und Jugendhilfe, Nr. 29, hrsg. vom Bundesministerium für Familie, Senioren, Frauen und Jugend, Bonn 2000

GERZER-SASS, Annemarie, ERLER, Wolfgang, SASS, Jürgen und Monika JAECKEL: Familienselbsthilfe und ihr Potential für eine Reformpolitik von „unten“. Deutsches Jugendinstitut, München 1999

HAUBRICH, Karin, LÜDERS, Christian und Gerlinde STRUHKAMP: Möglichkeiten und Grenzen für Evaluationen von Modellprogrammen angesichts der Frage nach ihrer Wirksamkeit und Nachhaltigkeit. In: entimon-Newsletter, Jahrg.: 15, 2005, Heft 4

HAUBRICH, Karin, LÜDERS, Christian und Gerlinde STRUHKAMP: „Wirksamkeit, Nützlichkeit, Nachhaltigkeit“ - was Evaluationen von Modellprogrammen realistisch leisten können.“ In: Deutsche Kinder- und Jugendstiftung (Hrsg.): „Evaluation - Chance oder Rotes Tuch?“. „Verschiedene Fremdevaluationsansätze im Kontext von Praxiserfahrung“, Opladen 2005 a

HAUBRICH, Karin, HOLTHUSEN, Bernd und Gerlinde Struhkamp: Evaluation – einige Sortierungen zu einem schillernden Begriff. In: DJI Bulletin 72 Plus, München 2005

HAUBRICH, Karin: Die Konstruktion des Untersuchungsgegenstandes in der Evaluation innovativer multizentrischer Programme. In: Karl-Siegbert REHBERG (Hrsg.): Soziale Ungleichheit - Kulturelle Unterschiede. Verhandlungen des 32. Kongresses der Deutschen Gesellschaft für Soziologie in München 2004. Frankfurt 2005

HAUBRICH, Karin: Cluster-Evaluation: Wirkungen analysieren und Innovation fördern. In: MEISTER, Dorothee M.; TERGAN, Sigmar-Olaf; ZENTEL, Peter (Hrsg.): Evaluation von E-Learning. Zielrichtungen, methodologische Aspekte, Zukunftsperspektiven. Münster 2004a, S. 155-170

HAUBRICH, Karin: Evaluation practice around the world. Germany. Transferring Cluster Evaluation to Model Programs in Germany. In: MATHISON, Sandra (Hrsg.): Encyclopedia of Evaluation. London, 2004b, S. 67-68

HAUBRICH, Karin und Christian LÜDERS: Evaluation - mehr als ein Modewort? In: Recht der Jugend und des Bildungswesens Jahrg.: 52, 2004, Heft 3, S. 316-337

HAUBRICH, Karin: Die Evaluation von Modellprogrammen im Spannungsfeld zwischen Wirkungs- und Nutzungsorientierung. In: LOIDL-KEIL, R.; VIECHTBAUER, K. (Hrsg.): Evaluation in der außerschulischen Kinder- und Jugendarbeit in Österreich. Berichte aus der Praxis. Linz, 2003, S. 205-222

HAUBRICH, Karin und Christian LÜDERS: Qualitative Evaluationsforschung. In: SCHWEPPE, Cornelia (Hrsg.): Qualitative Forschung in der Sozialpädagogik. Opladen 2003

HAUBRICH, Karin und Christian LÜDERS: Evaluation - hohe Erwartungen und ungeklärte Fragen. In: DISKURS Jahrg.: 11, 2001, Heft 3, S. 69-73

HAUBRICH, Karin: Cluster-Evaluation: lokale Kontexte berücksichtigen, Innovation und Beteiligung fördern. In: BMFSFJ (Hrsg.): QS 35 Materialien zur Qualitätssicherung in der Kinder- und Jugendhilfe, S. 65-68, 2001 (zum download: www.bmfsfj.de)

HAUBRICH, Karin und Kerstin FRANK. Vom Aufsuchen zur beruflichen Integration. Evaluationsstudie zum Bundesmodellprogramm „Mobile Jugendsozialarbeit für junge Menschen ausländischer Herkunft“. DJI-Arbeitspapier Nr. 1-156. München [WWW-Dokument] Verfügbar unter: http://www.dji.de/l_auslaender

HOOPS, Sabrina und Hanna PERMIEN: Die Programmevaluation des Bundesmodellprogramms „Ambulante Intensive Begleitung“ (AIB) - Chancen und Herausforderungen einer qualitativen Follow-up-Studie. In: Recht der Jugend und des Bildungswesens, 2004, Heft 3, S. 389-400

VOSSLER, Andreas und Andrea M. OBERMAIER: Netze knüpfen - Integration fördern. Evaluationsstudie zum Bundesmodellprogramm „Interkulturelles Netzwerk der Jugendsozialarbeit im Sozialraum“. Abschlussbericht der wissenschaftlichen Begleitung, hrsg. vom Interkulturellen Netzwerk der Jugendsozialarbeit. München 2003

WÜSTENDÖRFER, Werner: Erprobung des "Single Subject Research" im Modellprogramm "Kompetenzagenturen". DJI-Arbeitspapier 19/2005, Reihe: Wissenschaftliche Texte 4/2005 (Download: www.dji.de)

6. Programmevaluation des BMBF am Beispiel des Förderschwerpunkts „Sozial-ökologische Forschung“

Das Rahmenkonzept für die Evaluation des Förderschwerpunkts „Sozial-ökologische-Forschung“ bildet die zentrale Grundlage für exemplarisch aufzuzeigende Ansätze und Verfahren wirkungsorientierter Programmevaluation im BMBF. Darüber hinaus beziehe ich mich auf den Ergebnisbericht des im Förderschwerpunkt integrierten Projektes EVALUNET⁵⁷ und telefonische Expertengespräche mit AutorInnen, um die für die Programmevaluation relevanten Aspekte zu vertiefen.⁵⁸

6.1 Auftrag, Konzeptentwicklung, Beteiligte

6.1.1 Auftrag

Das BMBF hat im März 2003 eine externe Evaluationseinrichtung, die EvaConsult Berlin (info@ EvaConsult.de), beauftragt, ein Rahmenkonzept für die Evaluation seines Förderschwerpunktes „Sozial-ökologische Forschung“ mit den Programmelementen „Verbundforschung“, „Infrastrukturförderung“, „Nachwuchsförderung“ zu erarbeiten. Auftrag war:

- ≠ Die Festlegung von Eckpunkten einer Programmevaluation des Förderschwerpunkts
- ≠ Die Vereinbarung von Evaluationszielen und Leitkriterien für den Förderschwerpunkt insgesamt und die einzelnen Programmelemente
- ≠ Das Konzept des Evaluationsverfahrens, der Organisation und des Zeitplans
- ≠ Entwürfe von Musterleitfäden für die Evaluation (Simon und Stucke 2003a: 2)

⁵⁷ Vgl. BERGMANN, Matthias u. a. (2005): Qualitätskriterien transdisziplinärer Forschung. Ein Leitfaden für die formative Evaluation von Forschungsprojekten. ISOE-Studentexte, Nr. 13, Institut für Sozial-ökologische Forschung. Frankfurt a. M.

⁵⁸ Eine kritische Auseinandersetzung mit der früheren Evaluationspraxis des BMBF hat Stefan KUHLMANN (1998) in seiner Habilschrift „Politikmoderation“ geleistet. Leider ist das Buch vergriffen. Nach Aussage einer der Autoren und Mitglied der Evalunet-Projektgruppe, Engelbert SCHRAMM (ISOE, Frankfurt a. Main), sei der Ergebnisbericht von EVALUNET die Grundlage für die Entwicklung eines Meta-Evaluationsverfahrens auf Programmebene (BMBF). Matthias BERGMANN (Projektleiter von Evalunet) und Dagmar SIMON (WZB, EvaConsult, Berlin) sind gerade dabei, Schlussfolgerungen aus der Evalunet-Studie für die Programmevaluation zu ziehen.

6.1.2 Diskursiver Ansatz und formative Funktion : Einbeziehung der Beteiligten fördert Akzeptanz der Evaluationsmaßnahmen und ermöglicht Lernerfolge

Die Entwicklung und Abstimmung des Evaluationsdesigns (Rahmenkonzept und Evaluationsleitfäden) erfolgte in mehreren Rückkopplungsschleifen mit den am Verfahren beteiligten relevanten Akteuren, wobei auch externe Gutachter/Experten mit einbezogen wurden. Dieser *diskursive und partizipative Ansatz* verfolgt den Zweck, die Akzeptanz aller Beteiligten zu sichern und Transparenz über Verfahrensweisen und Entscheidungen zu schaffen. Die Diskursivität von Evaluationsvorhaben war auch der zentrale Verfahrensaspekt des Projektes EVALUNET⁵⁹ und gilt sowohl für die Designphase, als auch für die Auswertung und die Bewertung der Ergebnisse (Durchführungsphase und Verwertungsphase).

Das von EvaConsult⁶⁰ entwickelte Rahmenkonzept wurde am 12. Mai 2003 dem Strategiebeirat „Sozial-ökologische Forschung“ des BMBF zur Stellungnahme vorgelegt und anschließend auf einem vom Projektträger (GSF-Forschungszentrum für Umwelt und Gesundheit, München) veranstalteten Workshop (1,5 Tage) präsentiert. Teilnehmer waren die Projektnehmer und Projektnehmerinnen aus den geförderten Verbundprojekten und aus der Infrastruktur- und Nachwuchsförderung. Die Ergebnisse der Diskussion wurden bei der abschließenden Abstimmung des Konzepts zwischen Auftraggeber (BMBF) und Auftragnehmer (EvaConsult) berücksichtigt (vgl. SIMON und STUCKE 2003a: 2,3).

In Phase zwei der Entwicklung des Evaluationsdesigns wurden auf der Grundlage von Experteninterviews mit internen und externen Akteuren Evaluationskriterien (Leitkriterien) für den Förderschwerpunkt entwickelt und das Verfahren präzisiert. Die Operationalisierung der Leitkriterien mündete schließlich in sogenannte Musterleitfäden für die Programmelemente Verbundforschung, Infrastrukturforschung und Nachwuchsförderung. Diese Leitfäden stellen das Frage- und Indikatorenraster für die Programmbewertung durch eine *unabhängige Evaluierungsgruppe* dar. Sie wurden im August 2003 von EvaConsult vorgelegt, um durch Vertreter aus den Förderprogrammen und der Evaluierungsgruppe diskutiert zu werden (vgl. SIMON und STUCKE 2003b: 2).

⁵⁹ „Bei der in *Evalunet* angewendeten diskursiven Evaluation wird von Evaluierenden und Evaluierten gemeinsam ein analytischer Diskurs über die Hintergründe von Erfolgen und Misserfolgen des betrachteten Forschungsvorhabens geführt. Nur im direkten Gespräch war es möglich, ein tiefergehendes Verständnis für die komplizierten Abläufe, für Integrations-schritte, Planrevisionen u. v. m. zu bekommen, um so zu einer angemessenen Bewertung des Projektes in der Lage zu sein“ (BERGMANN u. a. 2005: 12).

⁶⁰ AutorInnen : Dr. Dagmar SIMON, EvaConsult und Mitarbeiterin des Wissenschaftszentrums Berlin und Dr. Andreas STUCKE, Wissenschaftsrat, Köln.

Da sich der Förderschwerpunkt als „lernender Schwerpunkt“ begreift und Änderungen der Förderstrategie beabsichtigt sind, sind gegebenenfalls auch entsprechende Modifikationen in den Evaluationsverfahren und -kriterien vorzunehmen. In diesem Zusammenhang wird der formativen Funktion von Evaluation (KUHLMANN: 2003) ein zentraler Stellenwert eingeräumt. Denn bei formativen Evaluationen wird das Lernziel sowohl für noch andauernde als auch für neu zu entwerfender Vorhaben betont. „Aus- und Bewertung geschehen also im Hinblick auf Schlussfolgerungen für die eigene Forschungsarbeit“ (BERGMANN u. a. 2005:12).⁶¹

6.1.3 Eckpunkte der Programmevaluation

Die Programmevaluation versteht sich als eine Kombination aus:

- Ex-post-Bewertung der bisherigen Förderpraxis,
- begleitende und formative Bewertung der laufenden Förderaktivitäten und
- Ex-ante-Beurteilung der zukünftigen Wirksamkeit des Förderschwerpunkts (SIMON und STUCKE 2003a:5).

Dabei wird auf die internationalen Standards der Evaluationsforschung (OECD) und der Gesellschaft für Evaluation (DeGEval) Bezug genommen.

Eckpunkte/Analysedimensionen der Programmevaluation waren:

- ## *die Grundannahmen des Förderprogramms* (z.B.: Beurteilung des Bedarfs für sozial-ökologische Forschung),
- ## *die Ziele* (Beurteilung der gesellschaftspraktischen und wissenschaftlichen Zielprojektionen: z. B. sind neue, transdisziplinäre Formen der Erkenntnisproduktion wünschbar und realisierbar?),
- ## *die Wirksamkeit der Förderinstrumente* bezogen auf die Ziele (z. B.: Beurteilung der Instrumente auf ihre nachhaltige Wirksamkeit)
- ## *die Organisation, Steuerung und Qualitätssicherung der Programmelemente* (Beurteilung der Kooperation der beteiligten Akteure: BMBF, Projektträger und Strategiebeirat/

⁶¹ Demzufolge sind formative Evaluationen nicht nur prozessbegleitend, sondern auch bei Abschluss (Ex-Post) oder Ex-ante (Vorbereitung neuer Vorhaben) einzusetzen. Dieser Blickwinkel unterscheidet sich von der gängigen Evaluationspraxis, in der formative Evaluationen prozessbegleitend und summative (zumeist quantitativ bilanzierend) bei abgeschlossenen Vorhaben angewendet werden.

Steuerung; Wirksamkeit der Begleitprojekte; Beurteilung der Verfahren und Kriterien für die Antragsbegutachtung; Beurteilung der internen Kommunikationsstrukturen),

≠ *Prospektion* (Entwicklung der Umweltforschung/Nachhaltigkeitsforschung in den nächsten 10-15 Jahren; Einordnung des BMBF-Förderprogramms in die mittel- und langfristige Entwicklung des Forschungsfeldes) (vgl. SIMON und STUCKE 2003a: 6,7).

Im Mittelpunkt stand die **Frage der Wirksamkeit**, die auf der Grundlage qualitativer Daten der Leistungen der ProjektnehmerInnen ermittelt wurde. Die zentralen Adressaten der Programmevaluation waren jedoch der Zuwendungsgeber/Auftraggeber (BMBF), der Projektträger (GSF) und der Strategiebeirat als Beratungsorgan des Schwerpunktes „Sozial-ökologische Forschung“.

Auf der Grundlage dieser Eckpunkte/Analysedimensionen wurden Leitkriterien entwickelt, die für jeden Programmbereich operationalisiert und in Evaluationsleitfäden übersetzt worden sind. Zudem wurden jedem Eckpunkt geeignete Methoden der Datengewinnung zugeordnet (quantitative: Statistik und Bibliometrie; Indikatoren: Antrags- und Förderstrukturen, Publikationsanalysen, Kooperationsaktivitäten; qualitativ: „Expert Review“ auf der Grundlage der Kriterienkataloge und Leitfäden)⁶².

6.1.4 Leitkriterien: Grundlage für die Evaluationsleitfäden

Die Programmelemente Verbundforschung, Infrastruktur- und Nachwuchsförderung des Förderschwerpunkts wurden unter wissenschaftlichen, wissenschaftspolitischen und förderpolitischen Zielen bewertet (SIMON und STUCKE 2003a: 7). Leitkriterien der Analyse sind:

≠ *Struktur des Förderprogramms*

≠ *Prozesssteuerung*

≠ *Ergebnisse*

Diese Leitkriterien wurden mehrstufig operationalisiert (SIMON und STUCKE 2003a: 11) und schließlich in Evaluationsleitfäden (- für jedes Programmelement wurde ein separater Leitfaden entwickelt -) übersetzt (SIMON und STUCKE 2003b:5-19). Bei der Konzeption der Leitfäden wurde das Wissen und die Expertise aller Beteiligten sowie externer Experten einbezogen: Akteure aus dem Förderschwerpunkt, Vertreter des Zuwendungsgebers BMBF,

⁶² Die detaillierte Zuordnung von Eckpunkten und entsprechenden Methoden ist im Rahmenkonzept noch nicht geleistet.

des Projektträgers (GSF-Forschungszentrum für Umwelt und Gesundheit, München), einzelne ProjektnehmerInnen sowie potenzielle Nutzer.⁶³

6.1.5 Methode des „Expert Review“: Unabhängige Gruppe von Experten evaluieren mithilfe der Leitfäden

Bezugnehmend auf die internationale Diskussion über Evaluationsstandards müsse sich Evaluation von Wissenschaft und Forschung am Maßstab internationaler Exzellenz der wissenschaftlichen Leistungen orientieren. „Dieses Kriterium muss auch für Formen wissenschaftlicher Forschung angelegt werden, die sich durch ihren inter- oder transdisziplinären Zuschnitt dem bekannten Peer-Review-Verfahren⁶⁴ der Beurteilung teilweise entziehen“ (vgl. OECD zit. nach SIMON und STUCKE 2003a: 5). Was hier für die Programmevaluation formuliert wird, gilt gleichermaßen auch für die Projektebene.⁶⁵ Die heterogene Struktur der Beteiligten, die aus unterschiedlichen wissenschaftlichen Disziplinen, aus Wissenschaft und Praxis kommen, erfordert eine analoge, umfassende Evaluationskompetenz, die durch ein einfaches Peer-Review (Evaluation durch FachkollegInnen) nicht gewährleistet werden kann. Um der Heterogenität, Komplexität und Perspektivenvielfalt inter- oder transdisziplinärer Programme oder Projekte gerecht zu werden, ist das Gutachterteam aus ExpertInnen zusammenzusetzen, die den Untersuchungsgegenstand aus einzelfachlicher und fachlich integrierter Perspektive beurteilen können, und WissenschaftlerInnen, die ihre Expertise hinsichtlich des Durchführens, Konzipierens und Bewertens von heterogenen oder transdisziplinären Forschungsvorhaben einbringen (BERGMANN u. a. 2005: 11). Da dieses Expertengremium unabhängig sein soll, gleichzeitig aber umfassende Kenntnisse des Untersuchungsgegenstandes besitzen muss, ist eine ausschließlich externe Evaluation durch eine unabhängige, „monolithische“ Evaluationseinrichtung, genauso wie ein Peer-Review nicht

⁶³ GesprächspartnerInnen waren: Dr. Matthias BERGMANN (Wissenschaftskolleg zu Berlin); Ingrid BALZER/Dr. Monika WÄCHTER (GSF/Projektträger), Prof. Reinhard HÜTTL (TU Cottbus); Dr. Thomas JAHN (ISOE, Frankfurt a. M.); Thomas KORBUN/Esther HOFFMANN (Institut für ökologische Wirtschaftsforschung, Berlin), Prof. Dr. Lenelis KRUSE-GRAUMANN (Fern Universität Hagen), Prof. Dr. Werner POLLMANN (ehem. Umweltbevollmächtigter der Daimler Chrysler AG); Dr. Martina SCHÄFER (TU Berlin); Dr. Angelika WILLMS-HERGET/Bernd FISCHER (BMBF) (vgl. SIMON und STUCKE 2003b: 4).

⁶⁴ Das Peer-Review-Verfahren wird/wurde zum Beispiel in der DFG praktiziert (vgl. KUHLMANN 2003) und der GTZ.

⁶⁵ Das im Förderschwerpunkt „Sozial-ökologische Forschung“ integrierte Projekt *Evalunet* hat sich intensiv mit der Frage gelingender transdisziplinärer Forschungsprojekte und dem Problem ihrer Bewertung auseinandergesetzt. Ergebnis von EVALUNET ist ein Leitfaden mit Kriterien zur formativen Evaluation, der sowohl prozessbegleitend als auch für die abschließende Bewertung des Projekterfolgs konzipiert wurde. Das diskursive Prinzip gilt auch hier in allen Phasen der Evaluation (Design- Durchführungs- und Verwertungsphase/Interpretation der Ergebnisse). Vgl. BERGMANN, Matthias u. a. (2005): Qualitätskriterien transdisziplinärer Forschung. Ein Leitfaden für die formative Evaluation von Forschungsprojekten. Institut für sozial-ökologische Forschung (ISOE) GmbH. Frankfurt a. M.

angemessen. Vielmehr soll, entsprechend dem diskursiven und partizipativen Ansatz sowie dem Konzept des „Expert-Review“, Evaluation in der Kombination von Beteiligten, Programmverantwortlichen und externen ExpertInnen durchgeführt werden.

6.1.6 Interne Evaluation und Selbstevaluation

Als interne Evaluation wird eine Evaluation definiert, wenn sie aus eigenem Antrieb der Institution/en erfolgt, die programmverantwortlich sind (Zuwendungsgeber, Programmentwickler, Durchführungsorganisation). „Intern“ bedeutet nicht, dass die Evaluierenden ausschließlich aus diesen Institutionen kommen, sondern es können auch externe Experten hinzugezogen werden. Demgegenüber werde Selbstevaluation ausschließlich von „internen“, also Projektbeteiligten und/oder Programmverantwortlichen durchgeführt (vgl. BERGMANN u. a. 2005: 22). Demzufolge wäre die Programmevaluation des Förderschwerpunktes eine interne Evaluation, obgleich sie von einem unabhängigen Expertengremium durchgeführt wird und das Evaluationskonzept von einer externen Einrichtung – diskursiv - entwickelt wurde.⁶⁶

6.1.7 Exkurs: Evaluationsforschung und -praxis im Wissenschaftsbereich - Guidelines für die Programmevaluation

Evaluation spielt im Hochschulbereich und in außeruniversitären Wissenschafts- und Forschungseinrichtungen seit vielen Jahren eine wichtige Rolle. Bereits seit 1980 sind Evaluationen zur Wirkungsanalyse forschungs- und technologiepolitischer Programme fest etabliert, die im Gegensatz zu anderen europäischen Ländern überwiegend von externen Forschungsinstituten durchgeführt werden (vgl. KUHLMANN/HOLLAND 1995 zit. nach SIMON und STUCKE 2003: 3).

Heute stehen sogenannte Systemevaluationen im Mittelpunkt, bei denen einzelne Segmente des Forschungssystems im Hinblick auf ihre Integration in die deutsche Forschungslandschaft untersucht werden. Dabei rücken Fragen der Forschungsorganisation und -struktur sowie des Managements zunehmend in den Blick und es wird auf die Funktion der Qualitätsentwicklung von Evaluation („objectice of improvement“) fokussiert (vgl. RÖBBECKE/SIMON 2001 zit. nach SIMON und STUCKE 2003a: 3,4).

Das Rahmenkonzept von EvaConsult nimmt auch Bezug auf die Position des Wissenschaftsrates, der die Verbesserung der Leistungsfähigkeit und der Qualität von Forschung und Lehre als wichtiges Ziel von Evaluationen hervorhebt (Wissenschaftsrat 2002 zit. nach SIMON und STUCKE 2003a). Er stellt fest⁶⁷, dass trans- und interdisziplinäre Settings gegenüber einzel-

⁶⁶ Ob diese Definition auch vom Auftraggeber BMBF geteilt wird, wäre zu überprüfen.

⁶⁷ in seiner Stellungnahme zu Querschnittsbegutachtungen in der Forschung (2002)

disziplinären Forschungskontexten bevorzugt gefördert würden, weil neues, innovatives Wissen immer seltener in einzeldisziplinären Kontexten erzeugt und von dort aus in gesellschaftliche Anwendungsbereiche transferiert würde. Vorteil integrierter und transdisziplinärer Entwicklungsarbeit sei, dass mögliche Anwender in frühen Phasen des Forschungsprozesses beteiligt sind (WISSENSCHAFTSRAT 2000 zit. nach SIMON und STUCKE 2003a: 4). Diese Kooperationsform gilt insbesondere für die sozial-ökologische Forschung.

In den Studien zu interdisziplinären und transdisziplinären Projekten und deren Evaluation wird insbesondere auf die Arbeiten von BERGMANN und JAHN (1998 und 2003) Bezug genommen, die in das Projekt EVALUNET eingeflossen sind und auch für die Konzeption und Durchführung der Programmevaluation des Förderschwerpunkts wegweisend waren (Verfahren der „diskursiven Evaluation“, s. o.).

6.2 Literatur

BERGMANN, Matthias (2003): Indikatoren für eine *diskursive* Evaluation transdisziplinärer Forschung. In: „Technikfolgenabschätzung“, Nr. 1, 12. Jahrgang - März 2003, S. 65-75

(Matthias Bergmann arbeitet am Wissenschaftskolleg zu Berlin und für das Institut für sozial-ökologische Forschung, Frankfurt a. M. Der Autor leitete das im Förderschwerpunkt „Sozial-ökologische Forschung“ integrierte Projekt EVALUNET)

BERGMANN, Matthias u. a. (2005): Qualitätskriterien transdisziplinärer Forschung. Ein Leitfaden für die formative Evaluation von Forschungsprojekten. ISOE-Studientexte, Nr. 13. Institut für sozial-ökologische Forschung (ISOE) GmbH, Frankfurt a. M.

KUHLMANN, Stefan (2003): Leistungsmessung oder Lernmedium? Evaluation in der Forschungs- und Innovationspolitik. In: „Technikfolgenabschätzung“, Nr. 1, 12. Jahrgang - März 2003, S. 11-19

(Prof. Stefan Kuhlmann arbeitet am Fraunhofer Institut für Systemtechnik und Innovationsforschung (ISI))

RÖBBECKE, M. und Dagmar SIMON (2001): Reflexive Evaluation. Ziele, Verfahren und Instrumente der Bewertung von Forschungsinstituten. Berlin

SIMON, Dagmar und Andreas STUCKE (2003a): Rahmenkonzept für die Evaluation des Förderschwerpunkts „Sozial-ökologische Forschung“ des Bundesministeriums für Bildung und Forschung. Paper für den Strategiebeirat vom 12. Mai 2003, hrsg. von EvaConsult Berlin

(Dr. Dagmar Simon ist Mitarbeiterin des WZB für Sozialforschung und Leiterin von EvaConsult; Dr. Andreas Stucke arbeitet beim Wissenschaftsrat in Köln)

SIMON, Dagmar und Andreas STUCKE (2003b): Muster-Leitfäden für die Evaluation des Förderschwerpunkts „Sozial- ökologische Forschung“ des Bundesministeriums für Bildung und Forschung. Paper vom August 2003, hrsg. von EvaConsult. Berlin/Köln

7. Zusammenfassender Vergleich

Abschließend wird der Versuch eines Vergleichs (ohne expliziten Vergleichsmaßstab) unternommen, d.h. Gemeinsamkeiten/Trends und Unterschiede der Ansätze kursorisch und thesenartig zusammengefasst.

Vergleichsproblematik:

Ein Vergleich der vorgestellten Ansätze und Verfahren im Hinblick auf die Frage „intern und/oder extern“ ist schwierig, da es hier unterschiedliche Konstrukte und Kombinationsmöglichkeiten gibt. Die Begriffe interne und externe Evaluation bezeichnen den Ort der Evaluation, also, ob sie von der Institution durchgeführt wird, die auch die Programmverantwortung trägt (intern) oder von einer externen Institution. Im Rahmen der Evaluation des Förderschwerpunkts „Sozialökologische Forschung“ des BMBF (vgl. Projekt *Evalunet*) wird demgegenüber eine Evaluation bereits als intern definiert, wenn sie aus *eigenem Antrieb* der projekt- oder programmverantwortlichen Institutionen erfolgt, wobei diese nicht zwingend auch die Steuerungsverantwortung übernehmen muss. Externe Experten und Gutachter werden mit einbezogen oder können sogar die Federführung übernehmen. Die zusätzliche Unterscheidung von Fremd- und Selbstevaluierung erschwert die Vergleichbarkeit durch einen weiteren Differenzierungsgrad. So sind fließende Übergänge von interner Evaluation und Selbstevaluierung möglich oder auch Kombinationen wie die interne Fremdevaluierung (z. B. bei der GTZ).

Die Definition von interner und externer Evaluation ist abhängig vom Beobachterstandpunkt; (Außen- oder Innensicht, vgl. GTZ-Kapitel).

Ein zweites Vergleichsproblem resultiert aus den unterschiedlichen Positionen der analysierten Institutionen - als Auftraggeber (GTZ, BMBF, BMAS) oder Auftragnehmer (DJI, CEval) und berührt letztlich auch die Frage „Ownership“ bzw. die Zuordnung der Ansätze: Sind die extern entwickelten Ansätze und Verfahren (z. B. vom CEval, von EvaConsult oder dem Institut für Sozial ökologische Forschung/ISOE) gleichzeitig Ansätze der auftraggebenden Institutionen?⁶⁸

⁶⁸ Die Ansätze und Verfahren des DJI (Auftragnehmer) wurden nicht dem BMBFSFJ (Auftraggeber) zugeordnet. Im Fall des BMWA und des BMBF wurden die extern entwickelten Ansätze den Auftraggebern zugeordnet.

Gemeinsamkeiten und Trends

- ≠# In keinem der untersuchten Designs wird die Kombination von interner und externer Evaluation als Programmatik oder innovatives Konstrukt in den Mittelpunkt gestellt. Die Gegenüberstellung „intern-extern“ scheint immer noch ein zentraler Unterscheidungsparameter der Evaluationstheorie und -praxis zu sein. Gleichwohl existieren in der Praxis zum Teil intern und extern besetzte Steuerungsrunden und es wird die Notwendigkeit betont, interne und externe Expertise miteinander zu verknüpfen. Die Kombination von interner und externer Evaluation und entsprechende Arrangements (z. B. Entwicklungspartnerschaften) wäre i. S. von TORRES/PRESKILL/PIONTEK eine Lösung, um die Vorteile beider Ansätze zu maximieren und die Nachteile zu begrenzen.
- ≠# Evaluation gilt zunehmend als Instrument der Organisations- und Qualitätsentwicklung (vgl. STOCKMANN 2006: Evaluation und Qualitätsentwicklung. Münster).
- ≠# Betont wird die Wirkungs- und Nachhaltigkeitsevaluation im Unterschied zu planungs-fokussierten Zielkontrollen (Ist-Soll-Vergleiche). Die Kausalitätsproblematik bei Wirkungsevaluationen (lineare oder zirkuläre Ursache-Wirkungszusammenhänge) wurde zwar teilweise angesprochen, bedarf m. E. aber noch einer vertiefenden Theorie-diskussion.
- ≠# Partizipative Ansätze sind Bestandteil aller beschriebenen Evaluationsdesigns; allerdings versammeln sich unter dem Begriff Partizipation unterschiedliche Formen, Grade und Zwecke der Einbeziehung der Beteiligten/Betroffenen (vgl. DJI-Kapitel).
- ≠# Eine wichtige Rolle spielen offene und formative (gestaltende) Evaluationen, die prozessbegleitend durchgeführt werden. Sie liefern relevante Informationen für die Projekt- oder Programmsteuerung und können auf die Verbesserung der Evaluations-objekte abzielen („objective of improvement“). Dabei geht es um Lernprozesse und Erkenntnisgewinne, die auch für die Weiterentwicklung von Evaluationsmethoden und -ansätzen nützlich sein können. Formative Evaluation ist nicht gleich zu setzen mit Prozessbegleitung. Dennoch werden die Begriffe formative und prozessbegleitende Evaluation oft synonym verwendet. Die summative Evaluation (überwiegend quantitative Datenerhebung wie z. B. Bibliometrie) wird vor allem bei Schluss- und Ex-post-Evaluationen eingesetzt. Dabei ist ergänzend auf KUHLMANN (2003) hinzuweisen, der von formativer Funktion von Evaluationen spricht; diese kann auch für Abschlussevaluationen und Ex-ante-Evaluationen gelten.

- ≠# Es zeichnet sich ein Trend zu entwicklungsoffenen, „lernenden“ Programmen (bzw. Förderschwerpunkten) und damit entsprechend entwicklungsoffenen Evaluationsverfahren ab (s. o.).
- ≠# In allen Institutionen werden nach Möglichkeit Multimethodenansätze - ein Mix aus quantitativen und qualitativen Methoden der Sozialforschung – angewendet. Der aus den USA stammende Ansatz der „Cluster-Evaluation“ wird bislang in diesem Sample nur im DJI eingesetzt und weiterentwickelt (z. B. für Programmevaluationen im Auftrag des BMFSFJ).
- ≠# Der Bezug auf die nationalen (DeGEval) Standards und/oder auch die internationalen Standards (DAC/OECD) ist Bestandteil aller untersuchten Designs; dabei wird einstimmig die Notwendigkeit einer unabhängigen Evaluation bekräftigt.

Unterschiede in den Verfahren und Ansätzen beziehen sich auf:

- ≠# Intensität, Umfang, Zweck und Art und Weise der Einbindung der Beteiligten, Adressaten sowie externer Experten/Wissenschaftler in den verschiedenen Phasen der Evaluation (Design-, Durchführungs- und Bewertungs-/Verwertungsphase): Bei komplexen, trans- und interdisziplinären Untersuchungsgegenständen zeichnet sich ein Trend zum „Expert-Review“ ab, im Unterschied zum traditionellen „Peer-Review“. (Beispiel: Die Evaluation des Förderschwerpunkts „Sozial-ökologische Forschung“ des BMBF und das integrierte Projekt EVALUNET).
- ≠# Die Bedeutung der Nutzenorientierung (vgl. Nützlichkeitsstandards der DeGEval): Sie wird in den einzelnen Institutionen unterschiedlich „hoch aufgehängt“. (Das DJI misst der Nutzenorientierung zum Beispiel eine geringere Bedeutung bei als das CEval).
- ≠# Unterschiede im Stellenwert von quantitativen und qualitativen Verfahren respektive die Gewichtung von formativer und summativer Evaluation: Obwohl ein Mix aus qualitativen und quantitativen Verfahren die meisten Designs kennzeichnet, ist davon auszugehen, dass Institutionen mit einem eher kritisch-rationalen Wissenschaftsverständnis quantifizierend-objektivierende Methoden bevorzugen (z. B. CEval), während kritisch-hermeneutisch oder konstruktivistisch-systemisch orientierte Einrichtungen/Experten stärker qualitative und diskursive Verfahren einsetzen. Die Orientierung an den DeGEval Standards ist bei Ansätzen, die nicht dem quantitativen Paradigma folgen, eingeschränkt. GUBA und LINCOLN haben deshalb für sogenannte „alternative“ Ansätze Parallelkriterien zu einigen Standards (G 9) eingeführt (vgl. WIDMER 1996).

Die Art der Programmevaluation und ihre Reichweite: Bei der Programmevaluation ist zwischen Evaluationssynthese und Meta-Evaluation zu unterscheiden. Die Unterschiede sind in den Kapiteln zur DeGEval und zum DJI beschrieben. Meta-Evaluation wird laut DeGEval und DJI (mit Bezug auf WIDMER: 1996) als Verfahren definiert, das die Güte von Evaluationen anhand von vorher festgelegten Standards (z. B. den DeGEval-Standards) überprüft. Beispiel für eine Evaluationssynthese ist die Programmevaluation von EQUAL, denn hier wurden die auf der Ebene der Entwicklungspartnerschaften und Netzwerke durchgeführten Evaluationen aus Programmsicht zusammenfassend reflektiert.

Programmevaluationen unterscheiden sich auch nach ihrer Reichweite. Sie können sich auf einzelne Aspekte beziehen 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 (vgl. SIMON und STUCKE 2003: 5).

Abstract

Sparmaßnahmen in nahezu allen Bereichen öffentlichen Handelns machen es erforderlich, in möglichst wirkungsvolle und Erfolg versprechende Projekte, Fördermaßnahmen und Programme zu investieren. Vor diesem Hintergrund - und bezugnehmend auf die Qualitätsdebatte - gewinnt das Thema wirkungsorientierte Evaluation an Bedeutung.

Das Diskussionspapier präsentiert Evaluationsansätze einschlägiger Institutionen in Deutschland, ohne den Anspruch auf Vollständigkeit oder Repräsentativität erheben zu können. Die inhaltliche Auseinandersetzung mit unterschiedlichen Evaluationsansätzen bietet eine Grundlage, den Diskurs zum Thema Evaluation in und zwischen den Institutionen der Beruflichen Bildung anzuregen. Im Mittelpunkt der Recherche stand die Frage nach dem Stellenwert von interner und externer Evaluation sowie nach Gestaltungsansätzen, in denen interne und externe Expertise miteinander verknüpft werden.

Austerity measures in nearly all areas of the public sector are making it necessary to invest in projects, support measures and programmes that are the most effective and hold the greatest promise for success. In this context - and with an eye to the quality debate – placing the focus of evaluation activities on the respective measure's effects and side effects is becoming ever more important.

This discussion paper presents approaches to evaluation used by relevant institutions in Germany, and makes no claim to being exhaustive or representative. Its examination of the substance of different approaches to the evaluation process offers a foundation for stimulating discussion in and between institutions in the vocational education and training field on the subject of evaluation. The investigation undertaken for this paper revolved around the question of the role and importance of internal and external evaluations and around organisational approaches that combine internal and external expertise.