

II. Entwicklungspolitische Wirkungserfassung: Stand der internationalen Debatte

1 Evaluationsforschung oder die Erfassung von Wirkungen

1.1 Was ist Wirkung?

Der Begriff „Wirkung“ kann u.a. physikalisch¹, rechtlich oder philosophisch definiert werden. Im Rahmen dieser Arbeit ist das Verständnis von Wirkungen im sozialwissenschaftlichen Kontext relevant: Menschliches Handeln und Maßnahmen verschiedener Akteure lösen Veränderungen auf verschiedenen Ebenen aus. In diesem Zusammenhang kommt der Kausalität eine entscheidende Bedeutung zu: Wirkung ist demnach eine Erscheinung, die auf eine andere Erscheinung, die Ursache, folgt und durch diese hervorgerufen wird.

Jede Intervention löst mindestens eine Wirkung aus. Im politischen Kontext verursachen Entscheidungen meist weitreichendere Wirkungen auf verschiedenen Ebenen und für einen größeren Kreis von Betroffenen. Die Implementierung politischer Entscheidungen erfolgt u.a. in staatlichen Programmen und Projekten, die bestimmte Wirkungen erzielen sollen, um die politischen Zielvorgaben zu erreichen.

Definition

Für den entwicklungspolitischen Kontext beschreiben Krimmel et al. (1990: 10) Wirkungen als die Effekte eines Programms / Projektes auf technische, ökonomische, soziokulturelle, institutionelle oder ökologische Veränderungen. Diese Definition kann noch um die politische Dimension erweitert werden; insgesamt betreffen die möglichen Wirkungen einer Maßnahme alle den Menschen und die Natur betreffenden Ebenen. Wirkungen sind nicht zwangsläufig singuläre Ereignisse, sondern in der Regel kontinuierliche Prozesse, die auch in einem größeren Zeitabstand zur Ursache auftreten können.

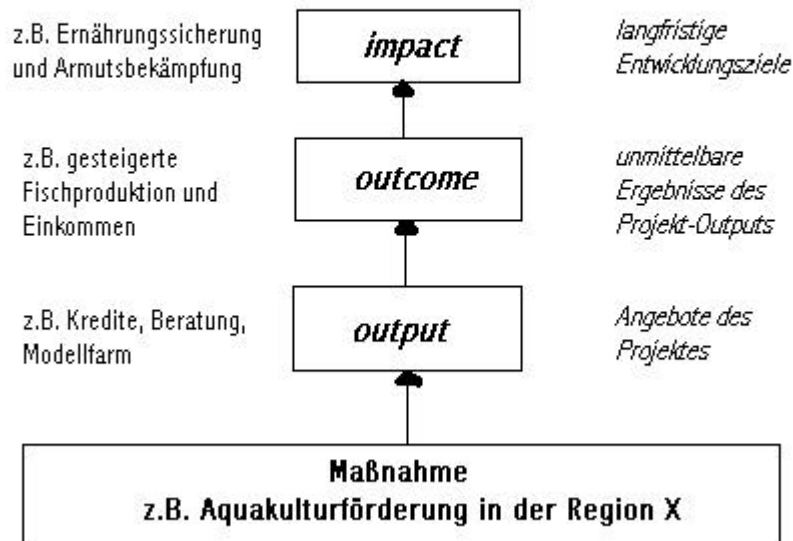
Wirkungen auf unterschiedlichen Ebenen

Mit Hilfe von konkreten Maßnahmen, Programmen oder Projekten sollen im Kontext politischer Entscheidungen Zielvorgaben umgesetzt werden. Dabei stellen die direkten Ergebnisse dieser Maßnahmen nicht unbedingt die unmittelbare Zielerreichung dar. Die Folgen einer Maßnahme lassen sich auf drei unterschiedlichen Ebenen differenzieren (siehe Schaubild 1 und vgl. Kaul 1977: 163f., Mosse/Sontheimer 1996: 8, Kumar 1995: 85 oder Mohr 1995: 283f.):

- *output*, als das unmittelbare Ergebnis eines Programms, die produzierten Güter;
- Effekte oder *outcome*, als die direkten Konsequenzen des *output*;
- Wirkungen oder *impact*, als die Konsequenzen der Maßnahme für die Veränderung der Lebensverhältnisse.

¹ In der Physik ist Wirkung eine Größe mit der Dimension eines Drehimpulses, z.B. Energie x Zeit.

Schaubild 1: Wirkungskette am Beispiel eines Aquakulturprojektes



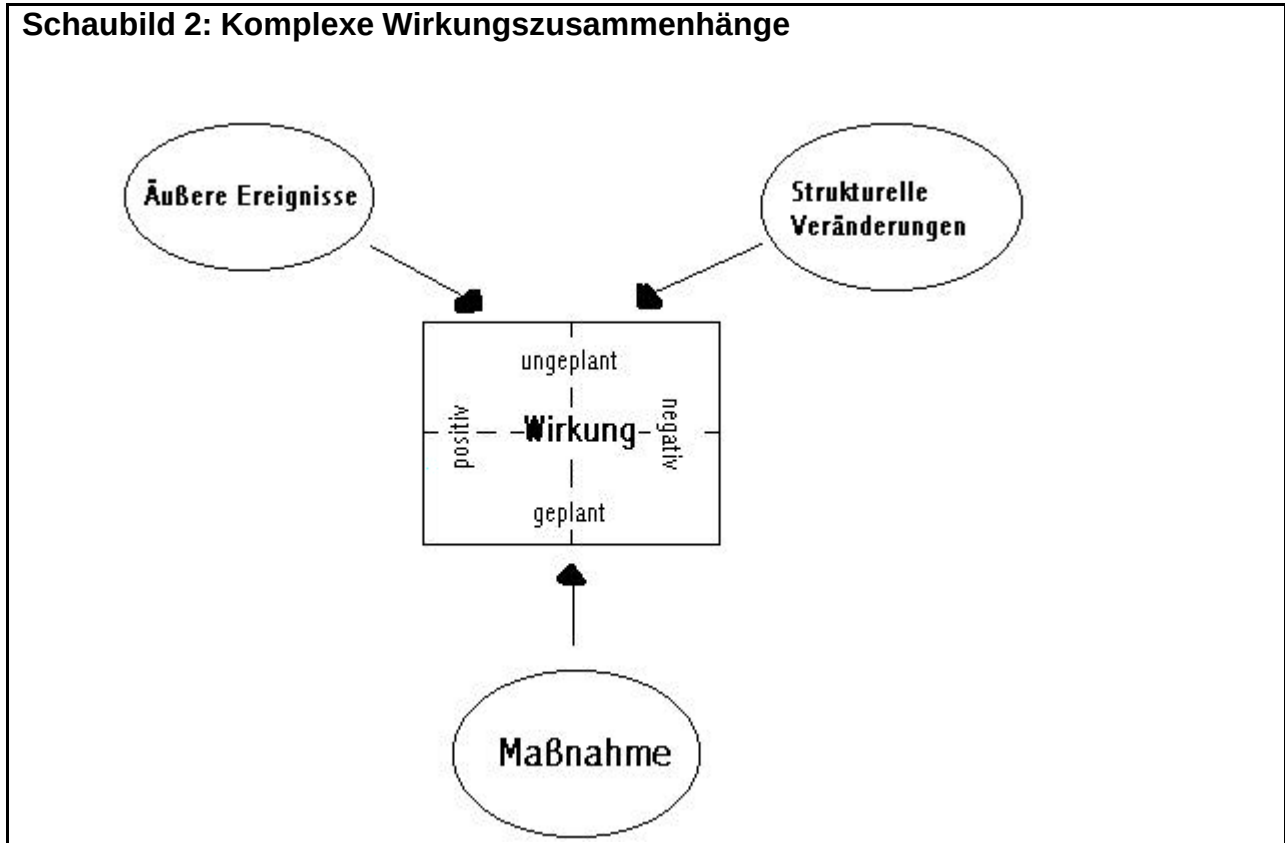
Mit dieser Unterscheidung ist neben der Intensität der Veränderung auch ein zeitliches Element in die Differenzierung eingeflossen. Wirkungen können durchaus erst mit einer längeren zeitlichen Verzögerung, zum Teil Jahre nach der Durchführung eines Programms, eintreffen. Um diese Ebene zu erfassen, sollten Untersuchungen zur Wirkungserfassung daher in größerem zeitlichen Abstand durchgeführt werden (vgl. u.a. Dolzer et al. 1998, Roche 1999, Adams 2001).

Unterschiedliche Formen von Wirkungen

Bislang war hier implizit nur von den geplanten Wirkungen einer Handlung die Rede, von Effekten, die sich einer Zieldimension eines Vorhabens zuordnen lassen. Die intendierten Wirkungen müssen nicht zwangsläufig positiv sein, obwohl dies in der Mehrzahl der Fälle so sein wird. Mitunter werden bei der Entscheidung für eine Maßnahme bereits negative Wirkungen eingeplant, die nicht vollständig vermieden werden können, aber im Zuge der Güterabwägung akzeptiert werden (vgl. Bolay 1988: 152).

Die Komplexität ökonomischer, sozialer, gesellschaftlicher und naturwissenschaftlicher Prozesse bedingt jedoch in der Regel, dass Maßnahmen auch nicht-geplante und nicht-intendierte Wirkungen auslösen (siehe Schaubild 2). Unabhängig davon, inwiefern die geplanten Wirkungen tatsächlich eintreffen und in welcher Stärke, können die nicht-intendierten Wirkungen wesentlichen Einfluss auf die Zielerreichung haben – sowohl negativ als auch positiv. Daher ist es als eindeutiges Manko anzusehen, dass die intendierten, positiven Wirkungen im Mittelpunkt der meisten Analysen stehen.

Schaubild 2: Komplexe Wirkungszusammenhänge



Neben der Unterscheidung in geplante und ungeplante Wirkungen, ist auch eine Differenzierung in irreversible und reversible *impacts* sinnvoll. Zum einen können Entscheidungen unterschiedliche Werte zugrunde liegen, auch Zielvorgaben sind politisch motiviert, so dass selbst intendierte, irreversible Wirkungen problematisch sein können. Zum anderen ist die Bedeutung nicht-intendierter, negativer Wirkungen ungleich größer, wenn es sich um nicht reversible Effekte handelt.²

Bei der Planung von Programmen/Projekten sind die Zielvorgaben zum Teil undeutlich und vielschichtig. Die geplanten Wirkungen können sich ergänzen, können aber auch mitunter im Konflikt zueinander stehen. Noch stärker ist das Konfliktpotential der Maßnahmen und ihrer Wirkungen auf der Ebene der Zielgruppen. Denn von den geplanten und ungeplanten, den positiven und negativen Wirkungen werden unterschiedliche Interessen berührt und Bevölkerungsgruppen in unterschiedlichem Ausmaß erreicht.

Ursachen und Wirkungszusammenhänge

Zu Beginn dieses Kapitels wurde bereits darauf hingewiesen, dass *qua definitionem* eine Wirkung eine abhängige Variable ist, d.h. nur existiert, wenn es eine Ursache gibt, die der Wirkung vorausgeht und diese hervorruft. Diesen Kausalzusammenhang herzustellen, eindeutig Wirkungen Ursachen zuzuschreiben, ist ein komplexes Unterfangen.

² Diese Diskussion ist vor allem im Kontext von Umweltproblemen aktuell geworden, ausführlicher hierzu z.B. Biswas/Geping (1987a) und Goodland/Edmundson (1994), die in diesem Zusammenhang die Notwendigkeit der Durchführung von Umweltverträglichkeitsprüfungen begründen, um irreversible Umweltschäden zu verhindern.

Auf philosophischer Ebene haben Bacon und später Mill maßgeblich Methoden zur Untersuchung kausaler Zusammenhänge erarbeitet.³ Dazu gehören u.a. die Methoden der Analogie, der Ähnlichkeit, des Unterschiedes sowie die vereinigte Methode von Ähnlichkeit und Unterschied mit deren Hilfe ein Kausalzusammenhang nachgewiesen werden soll.⁴

Im 20. Jahrhundert setzte sich die empirische Sozialforschung wesentlich mit der Frage auseinander, wie soziale Zusammenhänge valide und signifikant erklärt werden können (vgl. z.B. Flick et al. 1991, Kromrey 1998). Eine Definition von Validität, die häufig zitiert wird, stammt von Hammersley (1987: 79):

"An account is valid or true if it represents accurately those features of the phenomena, that it is intended to describe, explain or theorise."

Validität⁵ bezieht sich also darauf, ob gemessen wird, was gemessen werden soll. Damit stehen sowohl die Datenerhebung als auch die aus den Daten gezogenen Schlüsse auf dem Prüfstand. Wird die Validität einer Analyse in Frage gestellt, bedeutet dies eine fundamentale Kritik.

Die Zuordnung von konkreten Wirkungen zu Maßnahmen oder Programmen ist auch in der politischen und wissenschaftlichen Praxis das zentrale Problem. Während sich diese Zuschreibung auf der Ebene des *output* und teilweise auch des *outcome* noch schlüssig vollziehen lässt, ist das für den *impact* sehr viel schwieriger leistbar. Die möglichen Effekte externer Faktoren müssen idealerweise isoliert werden, um die exakte Zuschreibung zu ermöglichen.

Auf dem Hintergrund der Kausaltheorie und gleichzeitiger Bezugnahme auf die Komplexität gesellschaftlicher Realität, werden „mit und ohne“- bzw. Vorher-Nachher-Vergleiche und der Abgleich mit anderen Handlungsoptionen zur Klärung der Ursache-Wirkungskette herangezogen (vgl. Thin et al. 1997). In der sozialwissenschaftlichen Forschung zu Ursache- und Wirkungsketten gibt es aber noch mehr Lücken als eindeutig geklärte Verbindungsstücke.

Für die entwicklungspolitische Analyse versuchen Carvalho/White dies zu konkretisieren, ohne die grundlegende Problematik aufheben zu können:

³ Zu nennen sind hier Francis Bacon (*Novum organum* 1620) und John Stuart Mill (*A System of Logic* 1843).

⁴ Näheres über die Bedeutung der philosophischen Grundlagen für die empirische Sozialforschung findet sich in dem Sammelband von Sandkühler (1995).

⁵ Das Konzept der Validität lässt sich in drei Dimensionen aufteilen:

- Inhaltliche Validität – sind tatsächliches Objekt der Messung und intendiertes Objekt identisch?
- Kriterium Validität – ist eine verwendete Messgröße überhaupt prinzipiell geeignet, den Untersuchungsgegenstand zu erfassen?
- Konstrukt-Validität – verhält sich die Messgröße zu anderen Messgrößen wie es die Theorie voraussagt?

Ausführlicher hierzu bei Kromrey (1998) und Schnell et al. (1992).

„The development impact of a project is defined here as the incremental change in the standard of living of people affected by the project as a result of the project intervention“ (Carvalho/White 1996: 1).

Diese Definition beinhaltet für die Autoren drei Konsequenzen:

- den Vergleich mit und ohne eine Maßnahme;
- der Lebensstandard, nicht das Einkommen muss erfasst werden;
- die Veränderungen im Lebensstandard aller betroffenen Leute und nicht nur der Zielgruppe des Projektes müssen erfasst werden.

Die valide Zuordnung von Ursachen und Wirkungen ist mit diesen Vorschlägen jedoch nicht gelöst, sondern hierin liegt nur eine Annäherung.

Wirkungen als Kriterium zur Beurteilung

Trotz dieser Schwierigkeit, Ursachen und Wirkungen valide aufzuzeigen, d.h. Maßnahmen und Programmen ihre Folgen eindeutig zuzuschreiben, bleibt deren Analyse von zentralem Interesse. Menschliches und politisches Handeln wird neben den zugrunde liegenden Werten maßgeblich nach den Folgen beurteilt. Die Zuschreibung von Wirkungen ermöglicht Legitimation oder stellt sie in Frage und erlaubt die Behebung von Fehleinschätzungen.

Die Entwicklungspolitik muss die Ressourcen, die sie verwendet, durch den Nachweis von positiven Wirkungen legitimieren. Dieser Druck nimmt gerade in Zeiten knapper Haushaltsmittel zu. Insgesamt hat in den letzten zehn Jahren die politische und politikwissenschaftliche Diskussion über Entwicklungsmodelle und Entwicklungsprozesse an Schärfe verloren. In den 70er Jahren wurden die theoretischen Debatten über Modernisierungs- oder Dependenztheorien u.a. durch Studien zu entwicklungspolitischen Wirkungen geführt. Denn unterschiedliche Konzepte lassen sich nicht nur theoretisch hinterfragen, sondern müssen auch anhand der konkreten Wirkungen der jeweiligen Programme überprüft werden.

1.2 Wirkungen – ein Thema für die Politikwissenschaft?

Fragestellungen der Politikwissenschaft?

Die Disziplin der Politikwissenschaft widmet sich verschiedenen Fragestellungen, die zum Teil schon bei den griechischen Philosophen gestellt oder erst in jüngerer Vergangenheit entwickelt wurden. Gefragt wird nach den Zielen (z.B. Platon) und den Erscheinungsformen (z.B. Aristoteles) des Politischen. Neben dem normativen und institutionellen Fokus entwickelten sich Fragestellungen nach der Instrumentalisierung (z. B. Machiavelli) oder der Rechtfertigung (z.B. Hobbes) von Macht. Die moderne Politikwissenschaft rückt zudem die Fragen nach Konfliktursachen, Kooperationsmechanismen, nach Ursachen und Folgen des Politischen in den Mittelpunkt.⁶

⁶ Vgl. z.B. Autoren wie Dahrendorf (*Gesellschaft und Demokratie in Deutschland*, 1977, *Bürger und res publica - die Zukunft der Verantwortung*, 1986), Inglehart (*The silent revolution*, 1977, *Fundamental values across nations*, 1995).

Auch wenn die Ergebnisse politischen Handelns immer eine relevante Größe darstellen, sind sie doch erst in jüngerer Zeit stärker in den Blickpunkt politikwissenschaftlicher Forschung gerückt. Zentraler Gegenstand waren und sind jedoch weiterhin Theorien und Analysen zu Zielen und Institutionen der Politik.

Politik besteht aus *polity* – *politics* – *policy*

Was ist Politik? Der deutsche Begriff der Politik umfasst die Ebenen der Ordnung, Prozesse und Inhalte, was analog zu den angelsächsischen Begriffen in *polity*, *politics* und *policy* differenziert werden kann.

Polity repräsentiert die klassische Sicht auf Politik, auf die Ordnung und Verfahrensregeln des politischen Systems, die traditionell als Institutionenkunde betrieben wurde.⁷ Demgegenüber stellen sich *politics* den Fragen nach dem Entstehen politischer Entscheidungen, der Durchsetzung von Interessen und den Konfliktlösungsmechanismen.⁸ Die Wirkungen politischen Handelns, d.h. auch inwieweit politische Instrumente die angestrebten Ziele erreichen, fällt in die Kategorie der *policy*, ist hiervon jedoch nur ein Teilbereich. *Policy* stellt die Frage nach den Inhalten, nach Gegenstand und Aufgabe der Politik: Wie werden Ziele umgesetzt, welche Instrumente werden mit welchem Ergebnis genutzt? Diesem Teilbereich der Politik widmet sich der in den USA entstandener Ansatz der *policy analysis*, die als Politikfeldanalyse in den 70er Jahren auch in Deutschland an Bedeutung gewann.⁹

***Policy Analysis* – Politikfeldforschung**

Die kurze, klassische Definition von Dye (1977: 1) fasst den Inhalt der *policy analysis* klar zusammen: „what governments do, why they do it, and what difference it makes“. Wenn stärker die Anwendungsorientierung der Forschung in den Blick genommen wird, geht es der *policy analysis* darum „to choose the best policy alternatives among a set of alternatives with the aid of evidence and reason“ (Geva-May/Pal 1999: 262).

Aus Sicht der *policy analysis* ist Politik ein Prozess der Problemlösung (vgl. z.B. Dye 1977, Jann 1981, Schubert 1991 und Windhoff-Héritier 1993). Dieser Prozess ist dynamisch und kann in verschiedene Phasen aufgeteilt werden (siehe Schaubild 3, die klassische Darstellung von Jann)¹⁰. Idealtypisch besteht dieser *policy cycle* aus der Politikformulierung, der Implementierung und der Evaluierung.

⁷ Vgl. z.B. von Beyme (*Das politische System der Bundesrepublik Deutschland* 1998)

⁸ Stellvertretend hierfür z.B. Robert A. Dahl (*Politics, economics and welfare*, 1963), Guy Kirsch (*Ökonomische Theorie der Politik*, 1974).

⁹ Den Grundstein der *policy analysis* legten 1951 Lasswell/Lerner mit *The policy sciences*, die dazu aufforderten, dass sich die Wissenschaft stärker mit gesellschaftlichen Problemen befassen und praxisorientierte Ergebnisse liefern solle. In Deutschland fand dieser Ansatz erst während der reformorientierten Regierungszeit von Willy Brandt größere Beachtung. Die deutsche Diskussion, gerade auch in Auseinandersetzung mit der Kritik an der *policy analysis*, findet sich in einem Sonderband der Politischen Vierteljahresschrift (vgl. Windhoff-Héritier 1993).

¹⁰ Weitere Arbeiten zum *policy cycle* bei Nakamura/Smallwood (1980), Chelimsky (1985), Rist (1990).

Schaubild 3: *Policy cycle*

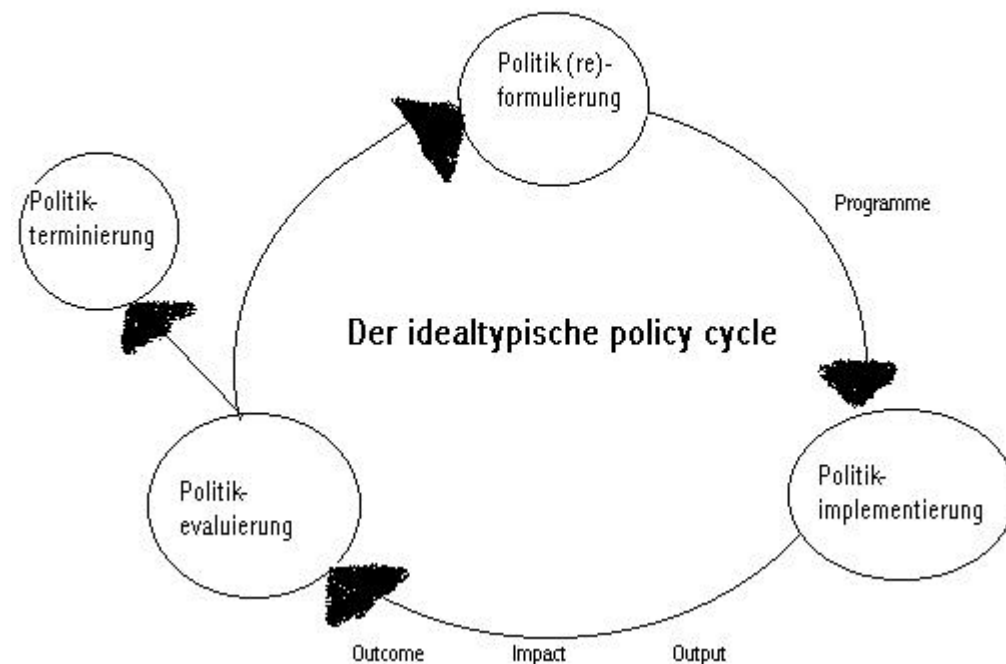


Schaubild nach Jann (1981: 34)

Der *policy analysis* in den 60er Jahren wurde zu Recht vorgeworfen, ein zu statisches Politikverständnis zu haben, das fälschlicherweise von rationalen Entscheidungsprozessen ausgeht. Die einzelnen Phasen des Politikzyklus können sich in der Realität überlagern oder verschieben. Diese Kritik wurde in den 80er Jahren aufgegriffen und die *policy analysis* versuchte, von der starren und schematischen Vorstellung des *policy cycle* abzurücken (vgl. Windhoff-Héritier 1993a).

Die Kritik an dem instrumentellen Verständnis von Politik, das demokratische Entscheidungen als Wert an sich betrachtet und grundsätzlich davon ausgeht, dass so getroffene staatliche Entscheidungen nur der Wohlfahrt verpflichtet sind, bezog sich auch auf die Evaluationsforschung (vgl. Hofferbert 1986: 511). Die Phase der Politik-evaluierung befasst sich als Teil des *policy cycle* mit der Analyse der Zielerreichung des politischen Handelns: „impact assessment is a particular form of policy analysis“ (Sadler 1994: 3).

Die Wirkungen, die von staatlichen Maßnahmen ausgehen, sollen erfasst und ausgewertet werden, wobei die Ergebnisse anschließend idealerweise wieder in den *policy cycle* eingespeist werden und entweder zur Beendigung, Bestätigung oder Abwandlung politischer Strategien führen. Wirkungen werden dabei unterschieden in:¹¹

¹¹ Dabei kommt es zum Teil zu begrifflichen Unterschieden der Nutzung von *impact* und *outcomes*. Schubert (1991: 76) benutzt bspw. *outcomes* für langfristige Wirkungen und *impact* für die

- Aktivitäten der durchführenden Akteure,
- Veränderungen bei der Zielgruppe durch diese Aktivitäten und
- umfassende Auswirkungen der Aktivitäten auf alle Betroffenen

Abgrenzung *policy analysis* - Evaluationsforschung

Die Evaluationsforschung ist ein Teil des *policy cycle* und der *policy analysis* in der Politikwissenschaft. In der theoretischen Diskussion und praktischen Anwendung verschwimmen jedoch manchmal die Grenzen zwischen beiden.

Neben der Übereinstimmung, die darin besteht, dass

- beide Wissen/Informationen im und für den politischen Prozess liefern,
- sich am Auftraggeber und den politischen Entscheidungsprozessen orientieren und
- beide dem Klienten helfen, sich systematischer mit einer Problemsituation auseinander zu setzen,

gibt es wichtige Unterschiede:

- Die Politikfeldanalyse ist zukunftsorientiert, während Evaluierungen auf laufende oder bereits abgeschlossene Maßnahmen schauen.
- Während die *policy analysis* nach politischen Alternativen sucht, die teilweise anderen Kriterien als der reinen Rationalität und Wohlfahrtsorientierung unterliegen, stehen Evaluierungen für *feedback* und *accountability*. Evaluierungen sind daher stärker der Forschung, Prüfung und Objektivität verpflichtet.
- Die Politikfeldanalyse trifft Aussagen darüber, welche Alternative zur weitestgehenden Zielerreichung führt und die Evaluierung bewertet bestimmte Maßnahmen danach, inwieweit sie das Ziel erfüllt haben.

Geva-May/Pal (1999) fordern eine klare Trennung zwischen beiden Ansätzen. Die Diskussion um die mangelnde Relevanz und politische Berücksichtigung von Evaluierungen (vgl. z.B. Chelimsky 1995) dürfe nicht dazu führen, dass die Evaluationsforschung „politischer“ werde. Dies würde nicht nur die Akzeptanz der Evaluierungsergebnisse verringern, sondern ebenfalls die *policy analysis* schwächen, weil sie auf den Ergebnissen nach wissenschaftlicher Objektivität strebenden Evaluierungen aufbaut.

Evaluationsforschung ist beides, ein Teil des politischen Prozesses und gleichzeitig ein wissenschaftliches Instrument zur Informationsgewinnung. Die Erfassung von Wirkungen ist ein wichtiger Analyseschritt zur Bewertung und Verbesserung staatlichen Handelns.

1.3 Methoden der Evaluationsforschung

Gegenstand von Evaluierungen bzw. der Evaluationsforschung sind in der überwiegenden Mehrheit konkrete Programme oder Projekte. In einem umfassenden Sinne wird ein

Veränderungen bei der Zielgruppe. Für den weiteren Verlauf dieser Arbeit sollen die Begriffe aber wie zu Beginn des Kapitel II skizziert verwendet werden, d.h. *impact* beinhaltet die langfristigen Wirkungen.

Soll-Ist-Vergleich der Ziele, des Ressourceneinsatzes, der durchgeführten Maßnahmen, der Ergebnisse und der intendierten wie nicht-intendierten Wirkungen gezogen.

Definition

Die Definitionen von Evaluierungen sind vielfältig. Sehr breit formuliert es Bolay:

„Unter Evaluierung versteht man im weitesten Sinne jede Auswertung und Bewertung von organisiertem oder öffentlichem Handeln entweder als Teilphase oder als Teilfunktion der Politikverwirklichung“ (Bolay 1988: 84).

Während einige Autoren den Begriff auf jede Programmphase anwenden (ex-ante, fortlaufend, ex-post, vgl. Forss 1985, Rossi et al. 1988, Stockmann 1996), wird mehrheitlich der Abschluss des Programms vorausgesetzt (exemplarisch Carvalho/White 1994: ix)

Da Evaluierungen im Kern die Wirkungen von Maßnahmen erfassen sollen, auch wenn sie in der Tiefe der Analyse stark abweichen, bietet die Evaluationsforschung den breitesten theoretischen Ansatz über unterschiedlichen Wege und Instrumente die Zusammenhänge zwischen Ursache und Wirkung zu analysieren.

Geschichte der Evaluierungsforschung

Auch wenn bereits die erste Verwendung von Versuchs- und Vergleichsgruppen im Alten Testament erwähnt wurde¹², ist die systematische Evaluationsforschung und -praxis noch ein relativ junges Phänomen. Mit dem Beginn der Evaluierung staatlicher Programme in den USA in den 30er Jahren, begann die politische und wissenschaftliche Auseinandersetzung und Weiterentwicklung der zugrunde liegenden Methoden.

Der *New Deal* in den USA in den 30er Jahren und die damit verbundenen groß angelegten staatlichen Reformprogramme begründeten die Notwendigkeit der Überprüfung, ob sich dieser Mitteleinsatz lohnt und damit politisch gerechtfertigt werden kann. Staatliche Investitionsentscheidungen für Sozial- oder Infrastrukturprogramme müssen sich legitimieren; dieser Druck wächst, wenn sich die Konkurrenzsituation durch knapper werdende staatliche Mittel verschärft. Neben dem Legitimationsdruck galt in den 30er Jahren, aber auch in den 60er Jahren das Interesse der Überprüfung innovativer sozialpolitischer Instrumente. In der Bundesrepublik ist der Beginn umfangreicherer Evaluierungstätigkeiten und wissenschaftlicher Begleitung Ende der 60er Jahren anzusetzen. Insgesamt werden die 60er und frühen 70er Jahre im Zuge der sozialpolitischen Reformbemühungen als das *Golden Age* der Evaluationsforschung bezeichnet (vgl. Rossi/Wright 1987, Bamberger 1991). Dies galt in besonderem Maße für die USA und vorrangig für die Überprüfung von Programmen im eigenen Land (vgl. Cracknell 1988: 75ff.).¹³

¹² Zu Beginn des Buch Daniel im Alten Testament wird ein Versuch geschildert, die Auswirkungen des Verzichts auf Wein nachzuweisen, indem eine Gruppe zehn Tage lang pflanzliche Nahrung und Wasser zu sich nimmt und eine Kontrollgruppe als Getränk nur Wein konsumiert.

¹³ Einen guten Überblick über die Entwicklung der Evaluationsforschung in anderen Ländern bis Ende der 80er Jahre findet sich in Derlien (1990) und aktuell für die USA, Europa und Deutschland sind Beiträge in dem Sammelband von Stockmann (2000a) zu finden.

Kernauseinandersetzungen

Evaluierungen haben unterschiedliche Ebenen und Maßnahmen zum Untersuchungsgegenstand und wurden und werden mit unterschiedlichen Methoden durchgeführt. In der politischen Diskussion spielen Evaluierungen nur eine geringe Rolle, da die idealtypische Beziehung zwischen politischen Akteuren und wissenschaftlicher Evaluationsforschung, wie sie die *policy analysis* voraussetzt, nicht existiert.¹⁴ Im wissenschaftlichen Raum ist gerade die Frage der angemessenen Methode bzw. der Weiterentwicklung bestehender Ansätze, sehr umstritten:

„There is much of doctrinal debate among evaluators: the profession is still young and in rapid evolution“ (Picciotto 1999: 8).

Der Kern der Auseinandersetzung bezieht sich auf die valide Erfassung der Realität, der zuverlässigen Erfassung von Wirkungen. Drei Hauptkontroversen lassen sich unterscheiden:

- experimentelle (quasi-experimentelle) und nicht-experimentelle,
- quantitative und qualitative sowie
- ökonomisch und sozial-orientierte Ansätze.

Experimentell - nicht experimentell

Unbestritten ist in der Evaluationsforschung, dass experimentelle Untersuchungen die zuverlässigsten Aussagen über Ursache- und Wirkungszusammenhänge ermöglichen (vgl. u.a. Rossi et al. 1988, Ahmed/Bamberger 1989, Kary 1992, Mohr 1995, Neubert 1999). Experimentelle Untersuchungsdesigns zeichnen sich dadurch aus, dass die Untersuchungseinheiten durch Zufallsauswahl bestimmt werden, eine Versuchsgruppe ausgewählt und bestimmten Einwirkungen ausgesetzt wird, während eine Kontrollgruppe, ohne diese Einwirkungen bleibt. Die Rahmenbedingungen werden kontrolliert.

Die zentrale Frage jeder Evaluierung, wie die Nettowirkungen erfasst werden können oder anders ausgedrückt, was passiert mit, was ohne die Maßnahme (Mit-Ohne-Vergleich), wird mit experimentellen Verfahren mit höchster Validität beantwortet.¹⁵ In der Evaluationsforschung gibt es jedoch stark divergierende Einschätzungen darüber, inwiefern es Ziel sein sollte, sich möglichst experimentellen Ansätzen anzunähern. Diese Annäherung wird beispielsweise mit quasi-experimentellen Designs versucht, die sich durch die nicht randomisierte Auswahl der Gruppen von experimentellen Verfahren unterscheiden.¹⁶ Nicht-experimentelle Designs können hingegen nur Vergleiche im Zeitablauf festhalten (Vorher-Nachher-Vergleich) (vgl. Kary 1992: 87).

¹⁴ Die damalige Präsidentin der *American Evaluation Association* beschreibt in diesem Zusammenhang „a greatly increased interest in the user's milieu“, um die Berücksichtigung der Ergebnisse der Evaluationsforschung in der Politik zu verbessern (Chelimsky 1995: 7).

¹⁵ Demgegenüber existiert jedoch auch die Auffassung, dass die Validität aufgrund der künstlichen Situation eines Experiments auch kritisch zu betrachten ist (vgl. z.B. Kary 1992: 79f.).

¹⁶ Versuchs- und Kontrollgruppen werden nach dieser Methode anhand von Merkmalsvergleichen eingeteilt. Durch dieses Vorgehen sind sie intern nicht so valide, die Durchführbarkeit wird aber erhöht (Begründer quasi-experimenteller Verfahren sind Campbell/Stamley 1969). Ausführlichere Informationen zu experimentellen und quasi-experimentellen Untersuchungsdesigns finden sich u.a. bei Rossi et al. (1988: 139ff.).

In der Boomzeit der Evaluationsforschung der 60er und 70er Jahre „the evaluation field was dominated by the randomized, controlled experimental paradigm“ (Rossi/Wright 1987: 48). Mit den enttäuschenden Ergebnissen dieser Studien setzte aber Ernüchterung und eine Abkehr vom methodischen Rigorismus ein (vgl. Valadez/Bamberger 1994: 28). Mehrheitlich sehen Evaluationsforscher heute in experimentellen Ansätzen keine praktikable Methode zur Analyse sozialer Zusammenhänge. Die Hauptargumente neben dem hohen personellen und zeitlichen Aufwand sind angesichts der Komplexität gesellschaftlicher Wirklichkeit die Unmöglichkeit der Kontrolle der äußeren Rahmenbedingungen, die Probleme bei der Auswahl von tatsächlich vergleichbaren Kontroll- und Versuchsgruppen und die moralisch problematische Entscheidung, der Kontrollgruppe Maßnahmen vorzuenthalten (vgl. z.B. Mayntz 1985: 74f., Rossi/Wright 1987, Brandstädter 1990: 218, Bamberger 1989: 234f.).

Quantitativ – qualitativ

Die Diskussion um qualitative und quantitative Methoden wird nicht zuletzt in Deutschland sehr dogmatisch geführt (ausführlicher dazu House 1980, Miles/Huberman 1987, Kelle/Erzberger 1999).¹⁷ Den scheinbar unversöhnlichen theoretischen Positionen steht aber eine Forschungspraxis gegenüber, die immer öfter versucht, beide Methoden zu verbinden.

Im Idealfall quantitativer Untersuchungen liegt der Datenerfassung ein experimentelles Design des Untersuchungsgegenstandes zugrunde. Insofern spiegelt sich in der Auseinandersetzung um qualitative und quantitative Verfahren die oben geschilderte Kontroverse. Der eigentliche Kern der Debatte liegt jedoch in der unterschiedlichen Erhebung und Auswertung der Daten zu bestimmten Wirkungszusammenhängen bzw. den entwickelten Kriterien (vgl. Gavai 1996: 115, Kelle/Erzberger 1999). Während quantitative Verfahren große Datenmengen voraussetzen und diese dann anhand statistischer Verfahren auswerten, liegt der Akzent bei qualitativen Verfahren stärker bei der intensiven Interpretation einer kleineren Anzahl von Untersuchungseinheiten (vgl. Kary 1992: 76, Ezemenari et al. 1999: 18).¹⁸ Qualitative Ansätze wollen soziale Phänomene erklären und die Wahrnehmung und Werte der unterschiedlichen Teilnehmer sozialer Prozesse zusammentragen.

Quantitative Verfahren erheben ausgehend von einer Hypothese Daten und ermitteln statistische Korrelationen. Durch valide statistische Verfahren werden präzise Aussagen möglich, komplexe Zusammenhänge sind jedoch mit quantitativen Indikatoren kaum darstellbar. Da Hypothesen und entsprechende Indikatoren vorher festgelegt werden müssen, können nachträglich keine neuen Erkenntnisse aufgenommen werden.

Die Vertreter quantitativer Ansätze betonen, dass qualitative Methoden aufgrund von Validitätsproblemen lediglich eine Ergänzung sein können, um die Fragen, die bei der quantitativen Untersuchung aufgetreten sind, näher zu analysieren (vgl. Krimmel et al. 1990: 19). Insgesamt werden qualitative Ansätze aufgrund der unzureichenden Ergebnisse experimentell-quantitativer Evaluierungen seit den 70er Jahren stärker diskutiert. Gerade auch der Misserfolg staatlicher Programme weckte das Bedürfnis, nicht nur zu

¹⁷ Lee (2000: 141) bezeichnet dies als den „cold war in evaluation“.

¹⁸ Als Begründer qualitativer Ansätze vor allem im nordamerikanischen Kontext gelten Weiss (1976), Cook/Reichardt (1979) und Patton (1987).

erfahren, ob ein Ziel erreicht wurde, sondern auch warum das Ergebnis so aussah. Die Erklärungskraft qualitativer Ansätze gewann so an Bedeutung, so dass heute in der Evaluationspraxis mehrheitlich qualitative und quantitative Verfahren gemeinsam benutzt werden und beide mit ihren Vor- und Nachteilen betrachtet werden (vgl. Campbell 1975, House 1980, Rossi/Wright 1987, Miles/Huberman 1987).¹⁹

Zwei Konzepte wurden entwickelt, um die Methodenintegration theoretisch zu untermauern:

- Phasenmodell (vorrangig Barton/Lazarsfeld 1984),
- Triangulation (vor allem Denzin 1977).

Das Phasenmodell plädiert für eine Ergänzung durch die Verwendung beider methodischen Ansätze, indem zur Hypothesengenerierung qualitative Verfahren benutzt werden, während die Datenerhebung quantitativ erfolgt. Gerade in sozio-kulturellen Kontexten, die sich vom Erfahrungshintergrund des Forschers gravierend unterscheiden, ist eine qualitative Analyse sinnvoll. Quantitativen Daten kommt bei dieser Aufgabenteilung die Aufgabe der Beschreibung zu, wohingegen qualitative Daten die Erklärung ermöglichen sollen.

Bei der Triangulation dient die Verwendung beider Verfahren zur Kontrolle/Absicherung der Ergebnisse. Zu einem Untersuchungsgegenstand werden mittels quantitativer und qualitativer Verfahren Daten erhoben und miteinander verglichen. Dieser Ansatz der Methodenintegration wird vorrangig von Vertretern qualitativer Methoden bevorzugt, hier „kommt die Idee zum Ausdruck, dass qualitative und quantitative Verfahren zwar verschieden, aber in bestimmter Hinsicht methodologisch gleichrangig sind“ (Kelle/Erzberger 1999: 514).

Ökonomische versus soziale Wirkungen

Die vorgenannten Kontroversen bezogen sich im Wesentlichen auf die Qualität des Instrumentariums. Neben methodischer Kritik an unterschiedlichen Verfahren bezieht sich die Kontroverse zwischen ökonomisch- und sozialorientierten Ansätzen vorrangig auf die Ebenen, auf denen Wirkungen beobachtet werden und die damit das Erkenntnisinteresse reflektieren.

Maßnahmen oder staatliche Programme haben in der Regel Zielvorgaben, die sich nicht auf das gesamte Spektrum gesellschaftlicher Wirklichkeit beziehen, sondern Einzelziele herausgreifen. Einen Schwerpunkt bilden hier sicherlich ökonomische Zielsetzungen und Programme. Methoden wie die Kosten-Nutzen-Analyse sollen dazu dienen, die Effizienz dieser Programme zu erfassen. Dabei liegt – methodenbedingt – ein Schwerpunkt bei den monetär fassbaren Kosten und Nutzen einer Maßnahme (vgl. z. B. Abelson 1979, Harberger 1996, Kirkpatrick/Weiss 1996b).

Kritiker betonen dagegen die Eindimensionalität dieser Ansätze und fordern Methoden, die soziale Wirkungen bei allen Maßnahmen einbeziehen. Entwicklung bestehe nicht nur aus dem *richtigen* makro-ökonomischen Ansatz, sondern vielfältigen sozialen

¹⁹ Explizit für den entwicklungspolitischen Kontext beschäftigen sich damit Chung (1996), Carvalho/White (1997) und Bamberger (2000).

Prozessen und ebenfalls Lernprozessen (vgl. z.B. Renard/Berlage 1992, Carlsson et al. 1994, Krugman 1994, Barrow 1997). Soziale, ökologische und auch gesellschaftspolitische Ebenen sollen nach diesem Verständnis nicht als Ergänzungen der ökonomischen Wirkungen analysiert werden, sondern als eigenständige, relevante Wirkungsebenen.

Evaluierungen in Entwicklungsländern

Die Evaluationsforschung und Methodendiskussion erstrecken sich auf alle Politikbereiche gleichermaßen. Für entwicklungspolitische Projekte werden vergleichbare Methoden eingesetzt wie beispielsweise für Sozialprojekte in den USA. Die Methoden wurden für die Analyse von Maßnahmen in Industrieländern entwickelt und wurden ab den 50er Jahren für die Entwicklungspolitik übernommen.²⁰

Die Ausgangslage von Evaluierungen in Entwicklungsländern weist jedoch einige Besonderheiten aufgrund der Unterschiede zwischen Programmen in Industrie- und Entwicklungsländern, der spezifischen Situation in Entwicklungsländern und den unterschiedlichen Akteuren auf:

- Maßnahmen und Evaluierungen werden in der Regel von einem anderen Land im Territorium eines Entwicklungslandes durchgeführt;
- die Systeme staatlicher Datenerfassung in Entwicklungsländern sind für quantitative Untersuchungen meist unzureichend und wenig verlässlich;
- eine verstärkte Anforderung an entwicklungspolitische Projekte ist (aufgrund des interkulturellen Ansatzes) Flexibilität und Lernfähigkeit, d.h. Projekte sind Wandel und Veränderungen unterworfen bzw. darauf angelegt (vgl. Kumar 1995: 148);²¹
- die Komplexität des Entwicklungsprozesses in Entwicklungsländern erfordert es, nicht nur Ergebnisse zu messen, wie dies beispielsweise in Programmen in den USA möglich ist, sondern auf weitergehende Wirkungen einzugehen (vgl. Kumar 1995: 150).

Die spezifische Situation von Evaluierungen in entwicklungspolitischen Zusammenhängen gibt zudem mögliche Rückschlüsse auf die Auswahl der Methoden zur Wirkungserfassung:

- die grundsätzliche Voraussetzung experimenteller Designs, die Bildung von Versuchs- und Kontrollgruppe und die Durchführung eines Programms lediglich in der Versuchsgruppe, ist im Rahmen von Entwicklungsländern moralisch noch problematischer;
- die notwendigen Voraussetzungen für quantitative Analysen liegen meist nicht vor (z.B. aufgrund der schlechten Datenlage, interkulturellen Problemen);
- Abhängigkeiten und Machtgefälle sind in entwicklungspolitischen Programmen und Evaluierungen noch größer als in Maßnahmen in Industrieländern.

²⁰ In der Evaluationsforschung gibt es insgesamt eine Dominanz der USA. Europa hat die Entwicklung nur teilweise nachgeholt, sowohl theoretisch als auch in der praktischen Anwendung (vgl. Beiträge bei Stockmann 2000a).

²¹ Hirschman hat schon 1967 in seinem höchst lesenswerten Buch *Development projects observed* Entwicklung als Lernprozess analysiert.

Methodenentwicklung und Diversifizierung

Die Gegenüberstellung von Kosten und Nutzen eines Programms oder Projektes bilden seit Beginn der Evaluationsforschung den Kern der Analysen. Die Kosten-Nutzen-Analyse ist das Instrument, das dies systematisch erfassen soll und dieser Ansatz stellt seit den 30er Jahren die am weitesten verbreitete Evaluationsmethode dar.

In den vergangenen Jahrzehnten war dieser Ansatz wiederholt starker Kritik ausgesetzt. Die methodische Debatte innerhalb der Evaluationsforschung führte einerseits zu Ergänzungen und Weiterentwicklungen der Kosten-Nutzen-Analyse und andererseits zu dem Entstehen neuer Verfahren, deren Akzent auf ökologischen, sozialen und partizipativen Elementen liegt.

Die Kosten-Nutzen-Analyse, *environmental impact assessment*, soziale Wirkungsanalyse und partizipative Verfahren bilden den Kanon der Methoden, die in entwicklungspolitischen Zusammenhängen diskutiert und angewendet werden. Neu in den Mittelpunkt und für die Fragestellung der Arbeit interessant, ist die Diskussion um die *impact analysis*. Die genannten Methoden weisen an einigen Punkten Überschneidungen auf, stellen jedoch im Kern die zentralen unterschiedlichen Herangehensweisen an die Wirkungserfassung dar. Sie sollen daher im folgenden vorgestellt und diskutiert werden im Hinblick auf folgende Kernfragen²²:

- Entwicklungsbegriff;
- Vollständigkeit / Breite des Ansatzes;
- *input-* oder *output-*Orientierung;
- Aussageebene (Mikro/Makro);
- Aussagekraft über Ursache-Wirkungszusammenhänge;
- demokratische Basis;
- Verständlichkeit / Anwendbarkeit.

Die Analyse der Evaluationsmethoden hinsichtlich der genannten Kernfragen soll die Grundlage sowohl im Hinblick auf den methodischen Ansatz der Arbeit als auch für die Auswertung vorliegender Studien bilden.

1.3.1 Kosten-Nutzen-Analyse

Stellenwert

Die Kosten-Nutzen-Analyse ist seit über 60 Jahren das dominante Instrument der Evaluationsforschung und die weltweit am häufigsten verwendete Methode, um die Effizienz von staatlichen oder privaten Programmen und Projekten zu prüfen.

Die Anfänge der Kosten-Nutzen-Analyse werden in das 19. Jahrhundert datiert und Jules Dupuit zugeschrieben (vgl. Johannsson 1993: 1). Erst in den 30er Jahren wurde

²² Die Kernfragen sollen Anhaltspunkte zur Bewertung der Methoden geben. Dahinter steht die Frage, was zeichnet eine „gute“ Methode aus? Bojö et al. (1992: 10) nennen dafür fünf Kriterien: Verhältnismäßigkeit der Kosten, umfassende und verständliche Ergebnisse, demokratische Basis und Relevanz für Entscheidungsfindung. Die Autorin hat diese Kriterien teilweise aufgegriffen und eigene hinzugefügt.

jedoch die Kosten-Nutzen-Analyse in den USA zur Grundlage politischer Entscheidungsfindung.²³ Seitdem ist die Zahl der wissenschaftlichen Arbeiten und empirischen Beispiele der Kosten-Nutzen-Analyse rasant gestiegen. In den USA ist die Kosten-Nutzen-Analyse ein zentrales Instrument der politischen Ressourcenplanung.

Definition

Grundlage der Kosten-Nutzen-Analyse ist eine ökonomische Betrachtung von Entwicklungsprozessen. Die Positionen und Definitionen differieren, inwieweit die Kosten-Nutzen-Analyse neben dem ökonomischen Fokus auch „a social assessment of the cost and benefits of public investment decisions“ (Brent 1998: 3, ähnlich Bojö et al. 1992: 57f.) erlaubt (vgl. konträre Positionen z.B. Marsden et al. 1994, Picciotto 1995, Barrow 1997).

Die Kosten-Nutzen-Analyse beruht auf der Annahme, dass alle relevanten Kosten und Nutzen einer Maßnahme erfasst und in monetären Einheiten ausgedrückt werden können. Liegt der Nutzen über den Kosten, ist der Effekt positiv, in anderen Fällen ist ein Programm negativ zu bewerten. Die *economic rate of return* drückt dieses Verhältnis aus.

Der Kosten-Nutzen-Analyse stellt sich ebenso wie anderen Evaluationsmethoden das Problem, Nutzen entsprechenden Maßnahmen valide zuzuordnen zu können. Eine Kosten-Nutzen-Analyse hat als Bemessungsgrundlage die Effekte bei Durchführung einer Maßnahme und ohne diese. Für einen Vergleich müssen aber Angaben über die Ausgangssituation vorliegen. Dies ist oftmals nicht der Fall und eine vergleichbare Kontrollgruppe ist notwendig.²⁴

Nicht in Frage stellt die Kosten-Nutzen-Analyse den Ausgangspunkt der Maßnahme, die Zieldefinition, sondern geht von einem Konsens über die zu erreichenden Ziele aus.

Kasten 1: Abgrenzung der KNA gegenüber der *cost effectiveness analysis* und *financial analysis*

Die Kosten-Nutzen-Analyse ist der dominierende Ansatz ökonomischer Evaluationsverfahren. Sie bezieht sich als Teil einer *economic analysis* auf die *social profitability*, den gesamtwirtschaftlichen bzw. gesamtgesellschaftlichen Nutzen einer Programms. Mit einer *financial analysis* wird in Abgrenzung dazu ein Vorhaben lediglich anhand betriebswirtschaftlicher Kriterien, der *private profitability*, der Kosten-Nutzen-Relation für die durchführende Institution oder das Unternehmen, bewertet.

Die *cost effectiveness analysis* geht im Gegensatz zur Kosten-Nutzen-Analyse von einem gesetzten, politisch vorgegebenen Nutzen aus, wobei die Kosten alternativer Realisierungsmöglichkeiten verglichen werden, um das Programm mit den geringsten

²³ Die Kosten-Nutzen-Analyse wird in den USA mehrheitlich *benefit-cost-analysis* genannt. In dieser Arbeit werden aber im Sinne der Einheitlichkeit die Begriffe Kosten-Nutzen-Analyse oder *cost-benefit analysis* gebraucht.

²⁴ Idealer Ausgangspunkt wäre auch in diesem Fall ein experimentelles Design, das es erlaubt, eine Kontrollgruppe ebenfalls zu untersuchen.

Kosten zur Erzielung des angestrebten Nutzen auszuwählen.²⁵ Kary nennt dies eine Untersuchung der wirkungsbezogenen Wirtschaftlichkeit, wohingegen die Kosten-Nutzen-Analyse sämtliche Wirkungen einer Maßnahme in Geldeinheiten bewertet (vgl. Kary 1992: 211).

Anwendungsbereiche

Das Hauptanwendungsgebiet für Kosten-Nutzen-Analysen ist die Projektprüfungsphase (*project appraisal*). Vor Durchführung einer Maßnahme soll die Kosten-Nutzen-Analyse entscheidende Hinweise geben, ob ein geplantes Programm die erwarteten Resultate erzielen wird bzw. in welchem Verhältnis die Erfüllung der Ziele zu den Kosten steht. Ex-ante muss zur Kosten-Nutzen-Analyse auf Annahmen oder vergleichbare andere Erfahrungen zurückgegriffen werden, um die Daten zu ermitteln. Nach Abschluss eines Programms werden Kosten-Nutzen-Analysen jedoch ebenfalls durchgeführt, um eine abschließende Bewertung zu ermöglichen.

Vorgehen

Grundannahme der Kosten-Nutzen-Analyse ist, dass Kosten und Nutzen einer Maßnahme am besten dadurch erfasst werden, dass die Bereitschaft der Menschen, für Güter zu zahlen (*willingness to pay*), abgebildet wird. Marktpreise werden als weitestgehende Annäherung für dieses Konzept verstanden. Für Güter (z.B. Umwelt oder Gesundheit), die auf dem Markt keine Preise haben, werden Hilfskonstrukte gebildet. Zum Beispiel der Verdienstaufschlag, um Gesundheit monetär darzustellen.²⁶

Das Vorgehen einer Kosten-Nutzen-Analyse ist bestimmt

- a) von der Auswahl der Kosten und Nutzen, die Berücksichtigung finden,
- b) die Bewertung dieser Kosten und Nutzen und
- c) die Bewertung im Zeitablauf (Zinsen).

Daran wird deutlich, dass auch bei dieser Methode bewusste Entscheidungen des Evaluierers getroffen werden müssen, die jenseits aller monetären Berechnungen das Ergebnis beeinflussen.

Vorteile

Der Vorteil der Kosten-Nutzen-Analyse liegt eindeutig in der quantifizierbaren Ermittlung und Gegenüberstellung von Kennziffern für Kosten und Nutzen einzelner Maßnahmen. Die resultierende *economic rate of return* symbolisiert ein eindeutiges Ergebnis und erleichtert politische Entscheidungen. Die tatsächliche Quantifizierung von Kosten und Nutzen ist die Hauptschwierigkeit, aber auch der Hauptgewinn dieser Methode.

²⁵ In der Entwicklungszusammenarbeit wird diese Methode häufig genutzt, um den kostengünstigsten Weg zur Erreichung festgelegter Ziele wie z.B. Erhöhung der Alphabetisierungsrate, Impfquote etc. zu ermitteln.

²⁶ Die vorhandene Kritik an diesem Vorgehen verdeutlichen Bojö et al. anhand des Beispiels Gesundheit: „The basic idea is to calculate tangible results of impaired health by summing the net present value of losses of earnings for the individual ... and medical costs for treatment. ... you are basically worth what you earn, plus what costs to treat you“ (Bojö et al. 1992: 75).

Die Attraktivität dieses Instruments steigt bei geringer werdenden staatlichen Mitteln, weil es scheinbar objektive Kriterien ermittelt, nach denen die Ressourcenverteilung vorgenommen oder abschließend bewertet werden kann.

Nachteile

Ökonomische Elemente und die Betonung der Fähigkeiten des Marktes dominieren diese Evaluationsmethode. Diese Charakteristika sind gleichzeitig Hauptkritikpunkte an der Kosten-Nutzen-Analyse. Zu den konkreten Kritikpunkten gehört (vgl. einige kritische Vertreter wie Amsberg 1994, Barrow 1997):

- die KNA zielt auf kurzfristige Entwicklungen/Wirkungen;
- Kosten und Nutzen, die bei Gütern entstehen, die nicht auf dem Markt gehandelt werden, werden nicht oder nur unzureichend erfasst;
- die Entstehung von Preisen auf dem Markt wird nicht hinterfragt;
- bestehende Varianzen und Unsicherheiten können nur unzureichend einbezogen werden;
- Einkommensunterschiede, die die Preisbildung beeinflussen, werden nicht berücksichtigt;
- grundsätzliche Vernachlässigung des Verteilungsaspektes;
- ein Markt funktioniert nicht von alleine bei natürlichen Ressourcen, weil es meist keinen Eigentümer gibt und man auch nicht einfach einen Markt herstellen kann, weil natürliche Ressourcen oft die Eigenschaften öffentlicher Güter haben.

Wandel der Kosten-Nutzen-Analyse

Die kritische Auseinandersetzung mit der Kosten-Nutzen-Analyse begann in den 60er Jahren. Der Vorwurf lautete, soziale und ökologische Kosten und Nutzen würden nicht berücksichtigt und die scheinbar objektive Aufschlüsselung würde die Realität nur unzureichend wiedergeben. Hirschmann, der 1967 eine grundlegende Studie zu Projektevaluierungen schrieb, sprach der Kosten-Nutzen-Analyse nur begrenzten Nutzen zu: „I do expect these concepts to have some uses in project evaluation, but almost entirely as additional elements of judgement“ (Hirschman 1967: 7).²⁷ Guy Kirsch spitzt die Kritik an der Kosten-Nutzen-Analyse und der mangelhaften Berücksichtigung sozialer Sachverhalte zu:

„In dem Ausmaß, wie die Cost-Benefit-Analyse nötig ist, ist sie nicht möglich;
in dem Ausmaß, wie sie möglich ist, ist sie nicht nötig“ (Kirsch 1975: 75).

In den 70er Jahren begann die Renaissance der Methode mit dem Versuch, neue Bereiche in die Analyse mit einzubeziehen. Verschiedene Vertreter der Kosten-Nutzen-Analyse haben in der Folge versucht, die Kritik zu berücksichtigen und das Instrumentarium daraufhin zu verändern. Vor allem die Kostenseite wurde neu definiert; das beinhaltet die eingesetzten Ressourcen und die indirekten Kosten, wie z.B. Luftverschmutzung.²⁸

²⁷ Ähnlich beschreibt dies 26 Jahre später Johannsson mit Blick auf die Vernachlässigung ökologischer Faktoren: „Cost-benefit analysis can capture and express in a single dimension (monetary units) many, but never all, of the effects of environmental projects“. Daher gelte „such analysis can seldom constitute the one and only factor in decision-making“ (Johannsson 1993: 3).

²⁸ Z. B. hat Abelson (1979) als einer der ersten versucht, Umweltkosten zu berücksichtigen.

Die Valorisierung von Umweltgütern stand im Mittelpunkt der Diskussion um eine Erneuerung und Erweiterung der Kosten-Nutzen-Analyse.²⁹ Mit Hilfe von Schattenpreisen³⁰ wird versucht, mit Marktpreisen nicht messbare Kategorien in die Analyse zu integrieren. Wenn ökologische Aspekte in eine Kosten-Nutzen-Analyse einbezogen werden sollen, ist ein *environmental impact assessment* (EIA)³¹ eine notwendige Voraussetzung, um die relevanten Umweltdimensionen vor der Gesamtbewertung festzustellen (vgl. Bojö et al. 1992: 9).

Diese Diskussion hat zu einer Erweiterung der Methode geführt, die aber weiterhin noch in der Kritik steht (vgl. Rossi et al. 1988, Berlage/Stokke 1992: 17ff., Ascher 1992, Marsden et al. 1994, House 2000). Dem ökonomischen Ansatz der Kosten-Nutzen-Analyse liegt zugrunde, dass alles seinen Wert hat, so wie sich *consumer preferences* in Geld ausdrücken. Das können auch soziale oder ökologische Güter sein. Funktioniert dieser Ansatz noch, wenn man z.B. aus ökologischer Sicht sagt, natürliche Ressourcen haben einen Wert an sich, unabhängig von den jeweiligen, momentanen Präferenzen? Und darüber hinaus: Kosten-Nutzen-Analysen beruhen auf Preisen und damit Bewertungen aus der Gegenwart, die Bedeutung natürlicher Ressourcen für zukünftige Generationen kann damit nicht erfasst werden.

Anwendung in Entwicklungsländern

Die Kosten-Nutzen-Analyse findet in der Projektprüfung und Evaluierung in Entwicklungsländern vielfältige Anwendung. Um sie in Entwicklungsländern anwenden zu können, sind jedoch einige Veränderungen vorzunehmen, die auf die Besonderheit der Situation in Entwicklungsländern eingehen.

Die Kosten-Nutzen-Analyse wurde anfangs noch unreflektiert als Instrument auch zur Bewertung der Entwicklungszusammenarbeit verwendet und spiegelte das zugrunde liegende ökonomische Entwicklungsmodell wider, dass finanzielle Unterstützung und Investitionen als Wachstumsschub ausreichen würden. Oder pointiert wie Picciotto es ausdrückt:

„Armed with cost/benefit analysis and operational research instruments, evaluation pioneers ascribed virtually no limits to the exercise of rationality in the realm of human affairs“ (Picciotto 1995: 14).

Die Differenzierung für Entwicklungsländer setzte in den 70er Jahren ein, als von führenden entwicklungspolitischen Institutionen drei verschiedene Modelle vorgelegt wurden, die versuchten, die Kosten-Nutzen-Analyse für Entwicklungsländer besser handhabbar zu machen. Die *United Nations Industrial Development Organization* (UNIDO) (Dasgupta et al. 1972), OECD (Little/Mirrlees 1974) und Weltbank (Squire/van der Tak 1975) legten drei verschiedene Studien vor, die alle auf der Kosten-Nutzen-

²⁹ Auf die umfangreiche Literatur zur Valorisierung soll hier nicht näher eingegangen werden, weil der Schwerpunkt dieser Analysen nicht auf der Erfassung der Wirkungen, sondern der monetären Bewertung derselben liegt. Vertreter dieser Richtung sind bspw. Pearce et al. (1994), Dixon et al. (1995), Atkinson et al. (1997) oder aus Weltbank-Sicht Serageldin (1995a).

³⁰ Schattenpreise sind künstliche Preise, welche die volkswirtschaftlichen Kosten von Gütern ausdrücken sollen, die keine oder mangelhafte Preise auf dem Markt haben. Zur pragmatischen Ermittlung von Schattenpreisen gibt es verschiedene Verfahren. Ausführlicher hierzu bei Brent (1998).

³¹ Auf das Instrument der EIA wird ausführlich in Abschnitt II.1.3.2 eingegangen.

Analyse basierten. Insbesondere widmeten sich diese Modelle in unterschiedlichem Maße

- den Problemen der Preisbildung in Entwicklungsländern bei Marktungenauigkeiten (*market imperfections*)³² und
- der Annahme, dass jeder selbst am besten über sein eigenes Wohlbefinden entscheiden kann und dies durch die Bereitschaft, dafür zu zahlen, ausdrückt und dadurch zur Preisbildung beiträgt. Dies funktioniert nicht in Situationen starker Einkommensunterschiede bzw. mangelnder Zahlungsfähigkeit in Armutssituationen. In Entwicklungsländern müssen daher Verteilungsaspekte integriert werden.

Die Anpassung der Kosten-Nutzen-Analyse für Entwicklungsländer lösten alle drei Ansätze der internationalen Organisationen im Wesentlichen durch die Bildung von Schattenpreisen, die Kosten und Nutzen unabhängig von tatsächlichen Marktpreisen monetär ausdrücken sollten. Diese Anpassung erscheint vielen Kritikern aber noch als unzureichend und hebt auch die grundsätzliche Kritik an den der Kosten-Nutzen-Analyse zugrunde liegenden Annahmen nicht auf.

So sieht Weiss weiterhin ein Hauptproblem aller drei Ansätze in der „Negierung der Mehrdimensionalität jedes konkreten Bewertungsproblems durch die Einengung auf ein eindimensionales Effizienzkriterium“ (Weiss 1976: 55). Die Modelle von Little & Mirlees und Squire & van der Tak werden aber auch kritisiert, weil sie *externalities* auslassen würden und Verteilungsaspekte weiterhin nicht genügend berücksichtigen (vgl. Tisdell 1991: 177, ähnlich Renard/Berlage 1992: 47f.).

Gerade auch NRO-Vertreter kritisieren die Dominanz der Kosten-Nutzen-Analyse:

„Evaluations that examine rates of return on investments proceed from a rather narrow set of principles rooted in western business practice. Such analyses assume that traditional methods can be extended to cover what are less tangible goals.“ (Marsden et al. 1994: 19)³³.

Die Weltbank verteidigt die Kosten-Nutzen-Analyse hingegen weiterhin als das angemessene Instrument zur Wirkungserfassung in entwicklungspolitischen Zusammenhängen (vgl. Harberger 1991, 1996a und 1996b, Kirkpatrick/Weiss 1996).

³² Die Kosten-Nutzen-Analyse kann in Industrie- und Entwicklungsländern nicht ohne methodische Anpassungen sinnvoll genutzt werden. Der Hauptunterschied ist, dass sich die Kosten-Nutzen-Analyse in Entwicklungsländern nicht im gleichen Maße auf Marktpreise stützt. Selbst wenn es für Güter Marktpreise gibt, gilt die Annahme, dass die Preisbildung in Entwicklungsländern zu fehlerhaft ist, um diese verlässlich nutzen zu können und Preise anders als im grundsätzlichen Konzept der KNA keinen verlässlichen Aufschluss über soziale Werte geben (vgl. Brent 1998: 3f., Harberger 1991, Renard/Berlage 1992).

³³ Marsden et al. (1994) vertreten demgegenüber das Konzept der stärkeren Berücksichtigung sozialer Faktoren und einer stärkeren Partizipation nicht nur in der Projektplanung und –durchführung, sondern auch bei den Evaluierungen. Näheres dazu in Kapitel II. 1.3.3 bis 1.3.5.

Übersicht 1: Bewertung der KNA im Hinblick auf die Kernfragen	
Entwicklungsbegriff	Die Kosten-Nutzen-Analyse entstand auf dem Hintergrund eines ökonomisch geprägten Entwicklungsbegriffs und ist diesem trotz Veränderungen verhaftet geblieben.
Vollständigkeit / Breite des Ansatzes	Fokus der KNA liegt eindeutig auf der ökonomischen, monetär quantifizierbaren Ebene. Trotz Erweiterungen der Methode zur Erfassung nicht auf dem Markt gehandelter Güter, noch deutliche Schwächen in der Erfassung sozialer und ökologischer Faktoren sowohl auf der Kosten- als auch auf der Nutzenseite.
Input- oder output-Orientierung	Sowohl <i>input</i> (=Kosten) als auch <i>output</i> (=Nutzen) stehen naturgemäß im Mittelpunkt der Kosten-Nutzen-Analyse. <i>Output</i> wird stark als <i>results</i> und weniger als <i>impact</i> interpretiert und gehandhabt.
Aussageebene (Mikro/Makro)	Die Kosten-Nutzen-Analyse will <i>social profitability</i> messen, in der Analyse greift sie aber meist kürzer und beschränkt sich auf die enge Mikroebene bzw. trifft Makro-Aussagen nur für ökonomische Kriterien.
Aussagekraft über Ursache-Wirkungszusammenhänge	Auch wenn die Kosten-Nutzen-Analyse mathematische Verfahren benutzt und Ergebnisse in Ziffern ausdrückt, fußt sie dennoch auf normativen Entscheidungen, über das, was als Kosten oder Nutzen angesehen und definiert wird: Grenzen hat die Kosten-Nutzen-Analyse zum Beispiel bei der Berücksichtigung von Einkommensverteilung, beim Intergenerationenverhältnis, bei der Bewertung von menschlichem Leben sowie bei kulturellen, historischen und ästhetischen Ressourcen. Aussagen, die die ermittelten Wirkungszusammenhänge erklären könnten, fehlen meist.
Demokratische Basis	Bei der KNA gibt es keine Partizipation der Beteiligten bzw. Rückkoppelung, was von diesen als Kosten oder Nutzen betrachtet werden. Die monetäre Festlegung von Kosten und Nutzen von Vorhaben in Entwicklungsländern beinhaltet darüber hinaus noch stärker das Problem der

	unterschiedlichen kulturellen und sozialen Bewertung von Kosten und Nutzen. Die Diskrepanz zwischen der monetären Sichtweise der Evaluierer und der Einschätzung der Zielgruppen kann sehr groß sein.
Verständlichkeit / Anwendbarkeit	Scheinbar hohe Verständlichkeit wegen der vorliegenden Kennziffern. Die Methode, die zu den Kennziffern führte, ist für Nicht-Ökonomen jedoch teilweise schwer nachvollziehbar, zumindest wenn komplizierte Preisberechnungen zugrunde liegen. Wenn die Datenlage gut ist, ist die KNA ein kostengünstiges Instrument.

1.3.2 *Environmental impact assessment (EIA)* ³⁴

Stellenwert

Ökologische Dimensionen als relevante Wirkungsebenen rückten Ende der 60er Jahre stärker in das Bewusstsein. Zu Beginn der Diskussion standen die Auswirkungen industrieller Aktivitäten in den Industrieländern im Mittelpunkt. In den 70er Jahren wurde diese Diskussion auch auf die Entwicklungsländer ausgeweitet. In der Wirkungs- und Evaluationsforschung der letzten Jahrzehnte hat die ökologische Dimension einen immer breiteren Raum eingenommen. Dies führte einerseits dazu, dass die vorherrschende Methode der Kosten-Nutzen-Analyse teilweise um Instrumente erweitert wurde, die ökologische Kosten und Nutzen einbeziehen sollen (siehe oben). Andererseits entstanden aber auch eigenständige Ansätze zur Wirkungserfassung, die ihr Schwergewicht auf die ökologische Dimension legten.

Wie für die Effizienzbewertung gilt für die Bewertung der Umweltwirkungen, dass das Schwergewicht auf der ex-ante Analyse liegt, um im Vorfeld negative Umweltwirkungen zu erkennen und zu vermeiden.³⁵ Institutionalisiert in einer verpflichtenden Umweltverträglichkeitsprüfung, sollen die Umweltwirkungen vor Beginn eines Vorhabens erfasst bzw. vorhergesagt werden.³⁶ Darüber hinaus kommt der Umweltdimension auch beim Monitoring und bei ex-post Evaluierungen eine zunehmend größere Bedeutung zu, dies hat aber nicht zur Herausbildung eines speziellen methodischen Ansatzes geführt.

³⁴ Die Begriffe *environmental impact assessment* und *environmental assessment* werden synonym gebraucht (vgl. Pearce et al. 1994: 10, Dixon et al. 1995: 9). Im Sinne der Einheitlichkeit wird in dieser Arbeit nur EIA verwendet.

³⁵ Barrow (1997:1) charakterisiert das EIA daher auch als Gegenbewegung zur „clean up later“-Haltung.

³⁶ EIA haben eine lange Tradition in den USA und Kanada. 1970 wurden EIA verpflichtend in den USA eingeführt (vgl. zur Entwicklung ausführlicher Wathern 1992, Barrow 1997: 166ff.). Die *United States Agency for International Development* (USAID) schreibt EIA seit 1976 für relevante Projekttypen vor (vgl. Kennedy 1992). In der Weltbank begannen die Überlegungen seit 1970; seit 1991 gibt es vier Kategorien von Projekten, für die ein EIA vorgeschrieben ist (vgl. Barrow 1997: 204).

Definition

Durch ein EIA sollen Umweltwirkungen erfasst, abgeschätzt und deren Bedeutung bewertet werden (vgl. Strömquist/Tatham 1992: 4, Wathern 1992: 6). Für den entwicklungspolitischen Zusammenhang definiert Bisset die Aufgabe von EIA als "identifying and assessing the environmental consequences of development projects, plans, programmes and policies in an attempt to ensure that the 'best' alternative for development is selected (Bisset 1987: 3).³⁷

Ein EIA bezieht sich ausdrücklich auf die Umweltdimension. Andere Faktoren (z.B. ökonomische, soziale etc.) werden lediglich im Hinblick auf ihre ökologischen Wirkungen analysiert.³⁸

Anwendungsbereiche

Die EIA wird entweder als selbständiges Instrument bei der Projektprüfung eingesetzt oder ist Teil bzw. Voraussetzung einer Kosten-Nutzen-Analyse, die ökologische Wirkungen ebenfalls einbeziehen möchte.

EIA werden meist auf der Mikroebene eingesetzt, die Anwendung für Regionen oder Sektoren ist selten, wenngleich sie wegen der Relevanz der Makroebene, der Bedeutung von politischen Rahmenbedingungen und der Wirkungsmultiplikation gefordert wird (vgl. Dixon et al. 1995: 20ff., Sadler 1994 und Goodland/Edmundson 1994).

Vorgehen

Die Identifizierung, Abschätzung und Interpretation der erwarteten ökologischen Wirkungen ist der Kern einer EIA. In der Realität stellt sich dabei oftmals das Problem dar, unter der Vielzahl möglicher Wirkungen, die nicht vollständig erfasst werden können, eine Begrenzung für die Wirkungsauswahl vorzunehmen.³⁹ Die *International Union for the Conservation of Nature and Natural Resources* (IUCN) hat 1980 in der *World Conservation Strategy* drei Kriterien festgelegt, nach denen Umweltwirkungen erfasst werden sollen:

- Zeitdauer und räumliches Ausmaß der Wirkung,
- Dringlichkeit,
- Grad der irreversiblen Schäden.

Zusätzlich führen Dixon et al. (1995: 10) noch das Kriterium der kumulativen bzw. synergetischen Effekte ein. Dabei wird deutlich, dass die EIA im Kern ausschließlich auf negative Umweltwirkungen zielt und positive Beiträge von Projekten oder Programmen weitgehend unberücksichtigt lässt.

³⁷ Ausführlicher zu den Eigenschaften von EIA in Barrow (1997).

³⁸ Mehrheitlich wird der Ansatz der EIA durch einen engen Umweltbegriff geprägt. Barrow kritisiert das und verwischt gleichzeitig mit einem breiteren Umweltbegriff die Grenzen zwischen EIA und *social impact assessment* (SIA) (Barrow 1997: 2).

³⁹ Die langjährigen Erfahrungen haben dabei deutlich eine Fehlerquelle des EIA aufgezeigt, wenn nicht rechtzeitig eine Auswahl der Wirkungsfaktoren stattfindet. Seitdem wird die Bedeutung der Reduzierung, des *scoping*, betont (vgl. Bisset 1987).

Zur konkreten Umsetzung einer EIA wurden seit den 70er Jahren verschiedene Instrumente entwickelt bzw. weiterentwickelt.⁴⁰ Zu den verbreitetsten gehören:

- Checkliste, eine der ersten verwendeten Methoden, relativ einfaches Verfahren; es führt die Umweltbereiche auf, die von einem Projekt wahrscheinlich betroffen sein werden;⁴¹
- *overlay mapping*, visualisiert verschiedene Auswirkungen/Veränderungen eines Projektes auf die Umwelt durch räumliche Karten;
- *network approach*, versucht in Diagrammform die Zusammenhänge der Wirkungen zu erfassen und darzustellen;
- *systems diagrams*, sollen auch komplexe Umweltsysteme als solche darstellen, ähnlich den Netzwerken;
- *simulation modelling*, stellen Wirkungen in Computermodellen dar.

Die Methoden unterscheiden sich hinsichtlich ihrer Kosten, Praktikabilität, Darstellungsebene und Aussageschwerpunkt und stellen jeweils nur Teilaspekte einer Umweltanalyse dar.

Vorteile

Der eindeutige Vorteil und die wesentliche Neuerung durch ein EIA ist, dass Umweltwirkungen im Mittelpunkt stehen und bewertet werden. Die ex-ante Anwendung von EIA zielt dabei auf die Fehlervermeidung bei der Umsetzung von Politiken und/oder privatem Handeln.

Nachteile

Im Vorfeld einer Projektdurchführung können EIA lediglich eine Annäherung an die tatsächlichen, späteren Wirkungen in einem konkreten Raum zu einer konkreten Zeit unter gegebenen Rahmenbedingungen darstellen. Ein EIA kann daher ein Monitoring und eine ex-post Evaluierung nicht ersetzen:

„While most EIAs carried out in the past, and being carried out at present, deal with the potential environmental implications of proposed development projects, there is an urgent need to monitor environmental changes once a project is operational“ (Biswas/Geping 1987b: 191).

Die möglichen Wirkungen zu identifizieren, scheint nur auf den ersten Blick eine einfache Aufgabe, oftmals liegen zu wenige Informationen über komplexe Wirkungszusammenhänge von Projekten vor. Die Datenlage ist für Umweltanalysen immer noch schlechter als für soziale Indikatoren (vgl. Serageldin 1995: 28f.). Zwar berücksichtigen EIA die Erfahrungen und Daten aus vorangegangenen, vergleichbaren Programmen. Wirkungen sind aber immer auch von der konkreten räumlichen Lage eines Projektes und den gesamten Rahmenbedingungen abhängig.

⁴⁰ Einen umfangreichen Überblick über die verschiedenen Instrumente bieten z.B. Strömquist/Tatham (1992: 4f.), Bisset (1987) und Barrow (1997: 132ff.).

⁴¹ Es gibt unterschiedliche Formen der Checkliste, die wie zum Beispiel die *multi-attribute utility* Theorie durchaus komplexere Zusammenhänge aufzeigen können (vgl. Bisset 1987: 16ff.).

Anwendung in Entwicklungsländer

Das Konzept des EIA basiert im Wesentlichen auf den Erfahrungen und Erfordernissen von Industrieländern. 1991 kritisierten die UN dieses Verfahren und forderten eine stärkere Anpassung und Beteiligung der Erfordernisse von Entwicklungsländern (UN 1991).⁴²

Ein wesentlicher Unterschied, der auch Auswirkung auf die Anwendbarkeit der Methoden hat, besteht in den unterschiedlichen Sichtweisen auf die Zusammenhänge von Umwelt und Entwicklung in Industrie- bzw. Entwicklungsländern. Priorität hat in Entwicklungsländern allgemein weiterhin das Ziel der wirtschaftlichen und sozialen Entwicklung. Diesen Entwicklungsprozess nachhaltig zu gestalten und nicht erst (wie weitestgehend in den Industrieländern) nachfolgend Umweltschutzmaßnahmen zu ergreifen, bleibt eine zentrale Herausforderung zu der die systematische Anwendung von EIA einen Beitrag leisten könnte:

„In contrast to the experience of industrialized countries where environmental concern was often viewed in the past as a cleaning-up process of the various sectors of the polluted environment, environmental concerns in developing countries are marked by a determination to achieve sustainable development in an environmentally sound manner“ (Biswas/Geping 1987b: 191 ähnlich auch DAC 2000: 10).

Ein weiterer Unterschied ist, dass die Datenlage in Entwicklungsländern meist noch schlechter ist als in Industrieländern und Probleme für die herkömmlichen Verfahren der EIA aufwirft.

Die nationale und internationale Diskussion über die Notwendigkeit einer umweltschonenden Entwicklung hat dazu geführt, dass EIA von verschiedenen Gebern bzw. in verschieden Entwicklungsländern für bestimmte Projekte zwingend vorgeschrieben wurden, ohne besondere methodische Anpassungsleistungen (vgl. Kennedy 1992). Auch in vielen Entwicklungsländern sind EIA im Rahmen der nationalen Gesetzgebung für viele Bereiche vorgesehen.⁴³ Die vorhandenen Umweltrichtlinien werden aber bislang nicht konsequent umgesetzt (vgl. DAC 2000).

Anwendungsbereiche sind in der Regel Infrastruktur- und Industrieprojekte, bei denen größere Umweltwirkungen, aber auch eine höhere öffentliche Aufmerksamkeit erwartet werden. Relevant wären EIA aber zum Beispiel auch für landwirtschaftliche Programme sowie für sektorale und nationale Maßnahmen.

⁴² Eine ähnliche Kritik formulieren die Autoren in Biswas/Geping (1987a) und Strömquist/Tatham (1992).

⁴³ In Indien gibt es seit Anfang der 80er Jahre ein EIA auf ad hoc Basis für große Projekte (siehe Kapitel III. 4.3.4.2). Eine Gesamteinordnung zur Situation in Entwicklungsländern findet sich bei Barrow (1997).

Übersicht 2: Bewertung des EIA im Hinblick auf die Kernfragen	
Entwicklungsbegriff	EIA erheben nicht den Anspruch, Entwicklung insgesamt zu erfassen, sondern wollen eine zu oft vernachlässigte Dimension (Umwelt), die weitreichende Auswirkungen haben kann, analysieren.
Vollständigkeit / Breite des Ansatzes	Im Zentrum des EIA stehen Umweltwirkungen, andere Faktoren werden ausschließlich in ihren Auswirkungen auf Umweltgüter in die Analyse mit einbezogen.
<i>Input-</i> oder <i>output</i> -Orientierung	Schwergewicht liegt auf den vermuteten Auswirkungen (=output) einer Maßnahme.
Aussageebene (Mikro/Makro)	Beschränkung auf die Mikroebene.
Aussagekraft über Ursache-Wirkungszusammenhänge	EIA versuchen eine Vorhersage von Ursachen und Wirkungen aufgrund von Daten und Vorerfahrungen. Es sind daher eher Plausibilitätsvermutungen.
Demokratische Basis	EIA beruhen mehrheitlich auf Methoden, die keine Partizipation beinhalten. Teilweise wird die betroffene Bevölkerung in die Planungsdiskussionen jedoch einbezogen.
Verständlichkeit / Anwendbarkeit	Im Vorfeld einer Entscheidung bieten EIA relevante Informationen. Komplexe Umweltwirkungszusammenhänge bedingen dabei teilweise schwer nachvollziehbare EIA. Gleichzeitig beruhen verschiedene Methoden auf einer hohen grafischen Veranschaulichung der Wirkungszusammenhänge.

1.3.3 Impact analysis

Einordnung

Bei der Analyse von Maßnahmen, Projekten oder Programmen werden die Ergebnisse betrachtet. Wie zu Beginn des Kapitel II aufgezeigt wurde, können diese Ergebnisse in *output*, *outcome* und *impact* unterschieden werden. *Impacts*, langfristige Wirkungen, sind für die Überprüfung der Zielerreichung von grundsätzlicher Bedeutung, stehen jedoch aufgrund ihrer Komplexität und hohen methodischen und zeitlichen Anforderungen meist nicht im Zentrum der Betrachtung. Dennoch hat sich das grundsätzliche Verständnis über die Notwendigkeit der Erfassung von Wirkungen gewandelt, selbst wenn die Praxis sich meist auf die Erfassung von *output* und *outcomes* beschränkt (vgl. exemplarisch die Weltbank-Diskussion in Baker 1999).

Definition

Die *impact analysis* stellt die langfristigen Wirkungen in ihren Mittelpunkt: Veränderungen sollen festgestellt, ein Kausalzusammenhang nachgewiesen und das Ausmaß der Veränderung gemessen werden. Oder:

„determining the extent to which one set of directed human activities (X) affected the state of some objects or phenomena (Y_1 , ..., Y_k) and – at least sometimes – determining also *why* the effects were as small or large as they turned out to be“ (Mohr 1995: 1).

Damit schließt die Definition der *impact analysis* alle Wirkungen eines Vorhabens mit ein und beschränkt sich nicht auf die Zielvorgaben (vgl. Bolay 1988: 152, Berlage/Stokke 1992: 13f., Roche 1999: 21). Die *impact analysis* erhebt innerhalb der Evaluationsmethoden das anspruchsvollste Ziel, das allerdings in der Praxis schwer zu erreichen ist (vgl. Barrow 1997: 22, Dolzer et al. 1998: 45f., Conway/Maxwell 1999a: 13f.).

Kasten 2: Abgrenzung *impact assessment* – *impact analysis*

Impact analysis und *assessment* werden teilweise synonym verwandt (z.B. Davis 1990). Mehrheitlich wird jedoch unter *impact assessment* eine Abschätzung der Wirkungen bestimmter Maßnahmen vor Durchführung (also ex-ante) verstanden (vgl. z.B. Thin et al. 1997)⁴⁴, wohingegen eine *impact analysis* ex-post Kausalbeziehungen aufzeigen und erklären will (vgl. z.B. Mohr 1995, Barrow 1997: 4).

Da ex-post Wirkungsanalysen sehr aufwendig und zeitraubend sind und nicht mehr in den Ablauf eines Projektes eingreifen können, geht die Tendenz hin zu einer stärkeren Nutzung von *impact assessment* als Managementinstrument vor bzw. während eines Programms (vgl. für multilaterale Organisationen die Untersuchung von Donecker/Green 1998). Folgerichtig wird ein *impact assessment* auch als Instrument zur Stärkung der *accountability*, sowie der Planung und Durchführung begriffen. Für Barrow ist ein *impact assessment* gerade durch einen Einsatz vor einer Maßnahme ein wichtiges Instrument, um nachhaltige Entwicklung zu erreichen (1997: 8f.).

Methode

In den 90er Jahren wurde verstärkt über *impact analysis* diskutiert und die Notwendigkeit grundlegenderer Analysen angemahnt. Diese Diskussion hat aber bislang nicht zu weiterführenden methodischen Überlegungen geführt, sondern beschränkt sich mehrheitlich darauf, das Ziel der Wirkungsanalysen eingehend zu beschreiben.

Da es bei der Zuschreibung von *impact* noch stärker als bei *output* und *outcome* um den schwierigen Nachweis von Kausalbeziehungen geht, sind die methodischen Fragestellungen von besonderer Relevanz bzw. Komplexität. Während einige Vertreter (z.B.

⁴⁴ Ausführlicher zu den unterschiedlichen Verfahren des *impact assessment* Barrow (1997). Dies beinhaltet sowohl EIA, SIA und technische und kulturelle Bewertungen.

Ahmed/Bamberger 1989: 13ff., Mohr 1995) die traditionelle Auffassung vertreten, ausschließlich experimentelle und quasi-experimentelle Ansätze könnten zur *impact analysis* verwendet werden, um Ursache und Wirkungszusammenhänge valide nachzuweisen, betonen andere stärker die Möglichkeiten qualitativer Ansätze. Nach diesem Standpunkt steht nicht der Nachweis der Kausalbeziehung, sondern deren Erklärung im Vordergrund, so dass auch andere Methoden in Betracht kommen (vgl. z.B. Berlage/Stokke 1992: 14ff., Valadez/Bamberger 1994, Conway/Maxwell 1999a: 21).⁴⁵ Diese Kritiker verweisen auch auf die unzureichenden Ergebnisse großangelegter quasi-experimenteller Studien in den 70er Jahren (vgl. Valadez/Bamberger 1994: 28, Binnendijk 1990: 175f.).

Die Forderung von Bolay, dass nur Analysen vorgenommen werden können über Wirkungen, die empirisch erfasst werden können (1988: 152), löst dieses Problem nicht, da die Auffassungen über die Messbarkeit verschiedener Indikatoren bzw. die Validität unterschiedlicher Methoden weiterhin stark divergieren.

In jedem Fall kann eine Wirkungsanalyse nicht alle möglichen und tatsächlichen Wirkungen eines Programms erfassen. Es muss eine Auswahl in Bezug auf die jeweiligen Forschungsfragen nach relevanten Kriterien vorgenommen werden.

Anwendung in Entwicklungsländern

Für die Situation der Entwicklungsländer ergeben sich aus der Diskussion einer *impact analysis* verschiedene Ansatzpunkte bzw. Problembereiche. Wird als notwendige Methode für diesen Ansatz das Experiment bzw. quasi-experimentelle Verfahren angesehen, muss auf die zusätzliche Problematik der Durchführung von Experimenten in diesem Kontext hingewiesen werden (siehe oben).

Des weiteren erfordert die hohe Komplexität der Wirkungsanalyse umfangreiche Kenntnisse über die jeweilige Projektsituation, die gerade von den meist externen Gutachtern, die nicht aus dem jeweiligen Kontext stammen, einen erheblichen Aufwand bedeuten.

Einen wesentlichen Teil der *impact analysis* bildet die Überprüfung der Ziele, die mit der jeweiligen Maßnahme verfolgt wurden. Da diese in der entwicklungspolitischen Projektrealität oft unzureichend formuliert und nicht operationalisiert werden, fehlt eine wesentliche Voraussetzung zur Feststellung der Zielerreichung. Donecker/Green (1998) fordern daher ein *logical framework* für alle Projekte.

Insgesamt stellen die große Komplexität entwicklungspolitischer Prozesse und kulturelle Divergenzen hohe Anforderungen an eine *impact analysis*, bedeuten aber gleichzeitig ebenfalls ihre Stärke.⁴⁶

⁴⁵ Valadez/Bamberger (1994: 24) nennen für *impact evaluations* daher auch verschiedene, mögliche Methoden neben quasi-experimentellen und Panel-Studien auch qualitative Verfahren und EIA.

⁴⁶ Bis in die 80er Jahre hinein wurden die Wirkungen von Maßnahmen der Entwicklungszusammenarbeit nur unzureichend erfasst. Eine Studie des DAC analysierte 1982 200 Evaluierungen im Hinblick auf den Beitrag der Projekte zur Armutsbekämpfung. Ein Beitrag zur Armutsbekämpfung konnte nicht festgestellt werden, weil die Evaluierungen überwiegend keine Aussagen dazu enthielten (DAC 1982).

Übersicht 3: Bewertung der <i>impact analysis</i> im Hinblick auf die Kernfragen	
Entwicklungsbegriff	Eine <i>impact analysis</i> beruht nicht per se auf einem engen oder breiten Entwicklungsbegriff. Es wird eine Auswahl der Wirkungsebenen getroffen und die wird vom Vorverständnis der Gutachter bestimmt. <i>Impact assessment</i> zum Beispiel kann sowohl zur Vorbereitung von Kosten-Nutzen-Analysen eingesetzt werden, als auch eine breite Wirkungsabschätzung vorlegen.
Vollständigkeit / Breite des Ansatzes	Umfassende Breite des Ansatzes, da verschiedene Ebenen mit einem längeren Zeithorizont erfasst werden sollen.
<i>Input</i> - oder <i>output</i> -Orientierung	Eindeutig stärker orientiert am <i>output</i> , speziell den langfristigen Wirkungen einer Maßnahme.
Aussageebene (Mikro/Makro)	Zielt explizit auf beide Ebenen.
Aussagekraft über Ursache-Wirkungszusammenhänge	Eine Verbindung von qualitativen und quantitativen Ansätzen ist die Regel. Die Erklärung von Wirkungszusammenhängen steht im Mittelpunkt der <i>impact analysis</i> .
Demokratische Basis	Mehrheitlich keine Verwendung partizipativer Verfahren.
Verständlichkeit / Anwendbarkeit	Aufwendiges, zeitintensives Verfahren und methodisch komplex. Eine <i>impact analysis</i> wird vergleichsweise selten durchgeführt und mit großem Abstand zu Ende der Maßnahme und beeinflusst daher die Entscheidungsfindung nur geringfügig.

1.3.4 Soziale Wirkungsanalyse

Einordnung

Die Erfassung sozialer Wirkungen ist ein zentraler Bestandteil der Evaluationsforschung. Es liegen aber unterschiedliche Auffassungen darüber vor, was darunter zu verstehen ist und wie soziale Wirkungen erfasst werden können.

Grob lassen sich zwei unterschiedliche Ansätze unterscheiden, die auch direkte Konsequenzen für das methodologische Vorgehen haben (vgl. z.B. Marsden/Oakley 1990a und 1990b: 5, Valadez/Bamberger 1994: 8ff., Mosse 1998):

- die soziale Dimension wird durch den Gegenstand definiert und so z.B. als die Bereiche Gesundheit, Bildung betrachtet, die quantitativ erfasst werden können; d.h. die einzelnen Wirkungsbereiche werden getrennt betrachtet bzw. ausgehend vom

Primat der Ökonomie werden positive soziale Wirkungen mit entsprechenden ökonomischen Indikatoren verbunden;

- soziale Entwicklung wird nicht auf sozialpolitische Bereiche beschränkt, sondern auf alle gesellschaftliche Verfahren und Prozesse ausgeweitet; die soziale Dimension hat Prozesscharakter; verwendete Methoden müssen daher flexibel sein.

Für die Fragestellung dieser Arbeit ist der zweite Ansatz interessant, der der Analyse sozialer Wirkungen im Gegensatz z.B. zur KNA ein besonderes Profil zuweist und im folgenden unter sozialer Wirkungsanalyse verstanden werden soll. Seit Mitte der 70er Jahren hat sich eine stärkere Diskussion um die Erfassung sozialer Wirkungen in der Evaluationsforschung entwickelt, die in der Praxis jedoch noch immer zurückhaltend angewendet wird.⁴⁷

Definition

Im Zentrum sozialer Wirkungsanalyse steht nicht eine bestimmte Methode, sondern eine Definition sozialer Entwicklung, die von traditionelleren Definitionen abweicht. Dabei geht es weniger um die materielle Bereitstellung von Gütern, sondern die Veränderung von menschlichen Fähigkeiten und Interaktionen:

„At the heart of social development efforts is the notion of increasing peoples' abilities to more effectively manage their own resources“ (Marsden et al. 1994: 11).

Nach diesem Verständnis zählen zu sozialen Wirkungen Veränderungen der Interaktionsmuster, formale und informelle Beziehungen, die sich aus diesen Interaktionen ergeben⁴⁸, soziale Organisationsformen, soziale Strukturen und Institutionen und Wahrnehmungen (vgl. Leistritz/Murdock 1981: 155ff.).

Das Bezeichnende von Ansätzen der sozialen Wirkungsanalyse ist daher die Auseinandersetzung mit sozialen Veränderungsprozessen und dem Versuch, Wirkungen auf dieser Ebene zu erfassen. Zur Beantwortung dieser Frage müssen Machtverhältnisse, Werte, Perspektiven und Wahrnehmungen hinterfragt werden, um die erwünschten sozialen Wirkungen wie Partizipation, *empowerment*, *self-reliance* oder Unabhängigkeit zu analysieren (vgl. Rahman 1990: 53ff.).

Methode

Zur Beantwortung dieser Fragestellungen steht kein einzelnes charakteristisches Instrumentarium zur Verfügung. Aus dem Verständnis von sozialer Entwicklung als Prozess und dem besonderen Anliegen, diesen Prozess zu erfassen und zu bewerten, folgt vielmehr das Postulat einer hohen methodischen Flexibilität bzw. Offenheit, die von den Kritikern auch als Beliebigkeit interpretiert wird (vgl. z.B. Kritik von Harberger 1996).

Auch die Vertreter dieses Ansatzes erkennen methodische Defizite an:

⁴⁷ „Although the significance of social impacts is often acknowledged, the utility of attempting to assess them is not“ (Leistritz/Murdock 1981: 156) – diese Feststellung gilt weitgehend auch für die 90er Jahre.

⁴⁸ In diesem Kontext kommt der Analyse des Geschlechterverhältnisses eine wesentliche Rolle zu.

„Während für die ökonomische und technische Projektbewertung ein standardisiertes Instrumentarium zur Verfügung steht, sind Methoden zur Erfassung der sozialen Dimension, die für die Wirkungsanalyse von zentraler Bedeutung sind, noch nicht ausgereift“ (Neubert 1999: I).

Im Kern führt diese Diskussion wieder auf den zu Beginn beschriebenen Disput über die Validität qualitativer versus quantitativer Verfahren zurück. Denn zur Erfassung des prozesshaften Charakters sozialer Entwicklung sprechen sich deren Vertreter für eine stärkere Nutzung qualitativer Verfahren aus (z. B. Oakley 1990: 32ff., Valadez/Bamberger 1994: 29, Neubert 1999). Als brauchbare Methoden diskutiert werden beispielsweise Sekundäranalyse, teilnehmende Beobachtung, *beneficiary assessment*, *social impact assessment* und *rapid assessment procedures* (vgl. z. B. Leistritz/Murdock 1981: 161f., Valadez/Bamberger 1994: 10ff.).

Ein grundlegendes Element der Evaluationsforschung, der Vorher-Nachher-Vergleich, muss zur Erfassung sozialer Wirkungen neu gefasst werden. Zwar bleibt der Vergleich eine wichtige Analysekategorie, der Prozess, der zu Veränderungen führt, erhält aber eine höhere, teilweise zentrale Bedeutung. Auch an diesem Punkt setzt die Kritik der Validität sozialer Wirkungsanalysen an.

Subjektivität wird jedoch im Zusammenhang sozialer Wirkungsanalysen nicht vermieden, sondern wird teilweise als notwendiger Bestandteil zur Erfassung der Realität verstanden. Einer Realität, die es nicht in einem Entwurf zu „entdecken“ gibt, sondern die aus den Wahrnehmungen der einzelnen Akteure besteht.⁴⁹ House hat daher bereits früh eine explizit abweichende Definition von Validität angeboten; demnach bedeute Validität „worthiness of being recognized“ (House 1980: 249). Der methodischen Kritik an qualitativen Verfahren zur Erfassung sozialer Wirkungen wird daher auch mit dem Vorwurf begegnet, die postulierte Objektivität gäbe es nicht und die konventionellen, quantitativen Methoden hätten einen *bias*, der menschliche Faktoren vernachlässige und die Situation armer Bevölkerungsteile ausblende (vgl. Chambers 1991: 516f.).

Kasten 3: Abgrenzung *social impact assessment* (SIA) von sozialen Wirkungsanalysen

Während soziale Wirkungsanalysen methodisch noch nicht klar umrissen und ausdifferenziert sind, hat sich seit den 80er Jahren ein Instrument herausgebildet, das *social impact assessment*, das ähnlich dem EIA zur Projektprüfung eingesetzt wird, um „peoples impacts“ (Wolf 1983: 15), „the human implications of proposed projects“ (Ingersoll 1990b: 19) zu erfassen.

Ein SIA soll analog zu einem EIA ex-ante Wirkungen analysieren und ist Teil der Projektprüfung (vgl. Dietz 1987, Taylor/Bryan/Goodrich 1990, Ingersoll 1990b: 19, Barrow 2000). Entwickelt wurde dieser Ansatz maßgeblich von Finsterbusch und Wolf

⁴⁹ Diese Position (des Konstruktivismus) steht bewusst im Gegensatz zum Wissenschaftsentwurf des Kritischen Rationalismus (mit den Vertretern Popper 1979, 1982 und Albert 1968).

(1977 und 1983⁵⁰). Es werden unterschiedliche Instrumente benutzt, um die *social impacts* festzustellen (siehe z.B. Wolf 1983: 16ff., Carley 1983). Die Weltbank setzt das Instrument des *social assessment* erst seit 1994 verstärkt ein. Indien gehört dabei zu den Hauptanwendungsländern (vgl. ausführlicher van Wicklin III 2000: 13ff.).⁵¹

Insgesamt hat sich die Anwendung von SIA im Vergleich zur Nutzung des EIA jedoch nicht durchgesetzt.

Anwendung in Entwicklungsländern

Im entwicklungspolitischen Kontext spielt die Diskussion um soziale Entwicklung eine zentrale Rolle. In den 70er Jahren fanden Ansätze, die neben quantitativen und ökonomischen Faktoren auch die Bedeutung der sozialen Dimensionen hervorhoben, stärkere Beachtung. Zielte die Diskussion anfänglich darauf ab, neben ökonomischen Kategorien wie Wirtschaftswachstum oder Sparquote auch die Berücksichtigung der Grundbedürfnisse wie Bildung oder Gesundheit durchzusetzen, trat später die Bedeutung prozesshafter Faktoren menschlicher Entwicklung in den Mittelpunkt.⁵²

Gerade im entwicklungspolitischen Kontext wird auch auf dem Hintergrund der kulturellen Unterschiede und der ungleichen Machtverhältnisse zwischen Geberorganisationen und Projektbegünstigten, von Vertretern sozialer Wirkungsanalysen die Erfassung sozialer Wirkungen mit partizipativen Elementen verknüpft (vgl. so Marsden/Oakley 1990, Marsden et al. 1994, Valadez/Bamberger 1994, Neubert 1999). So nennt Oakley (1990: 32) vier Prinzipien für die Evaluierung sozialer Entwicklung:

1. qualitative und quantitative Bewertung⁵³;
2. dynamischer Prozess;
3. enge Verbindung zum Monitoring;
4. Partizipation.

Neubert stellt in ihrem Beitrag zur sozialen Wirkungserfassung mit SWAP eine neue Methode vor, die im Wesentlichen soziale Wirkungen mittels partizipativer Elemente erfassen will (vgl. Neubert 1999). Dieses System zur Wirkungsanalyse armutsorientierter Projekte (SWAP) verbindet verschiedene, bekannte Instrumente zu einem neuen Ansatz. Im Rahmen der Projektplanung und -prüfung eingesetzte Verfahren wie Zielorientierte Projektplanung (ZOPP), *logical framework* oder *participatory* bzw. *rapid*

⁵⁰ 1983 umfasste die Zusammenarbeit zusätzlich Llewellyn (vgl. Finsterbusch et al. 1983).

⁵¹ Eine Zusammenstellung verwendeter Techniken findet sich für den Kontext der Weltbank bei Rietbergen-McCracken/Narayan (1998).

⁵² In der entwicklungspolitischen Praxis setzte sich zu diesem Zeitpunkt immer stärker die Erkenntnis durch, dass die Berücksichtigung sozialer Faktoren für den Erfolg von Projekten wichtig ist - auch wenn die Projekte einen ökonomischen oder technologischen Schwerpunkt haben. Eine Untersuchung von Kottak (1985) ergab, dass bei Weltbank-Projekten die *economic rate of return* zweimal so hoch war, wenn die Projekte sozio-kulturell integriert waren.

⁵³ Ein gutes Beispiel für den Versuch, soziale Wirkungen durch eine Verbindung von quantitativen und qualitativen Methoden zu erfassen, bietet Indhapanya et al. (1999). Mit der Methode der *spatial analysis* versucht sie, durch Indikatorenbildung auf kleinstem Raum die Auswirkungen eines Entwicklungsprogramms in Thailand zu erfassen. Dabei werden mittels partizipativer Verfahren vor Ort quantitative Indikatoren ermittelt, die es dann ermöglichen, die Auswirkungen auf kleine räumliche Einheiten wiederzugeben.

rural assessment bilden den Erfahrungshintergrund dieser Methode, die gleichzeitig stark auf Handhabbarkeit und geringen Aufwand ausgerichtet ist. Hauptdatenquelle in SWAP sind die Erinnerungsleistungen der Zielgruppen.⁵⁴ Soziale Wirkungen sollen mit Hilfe derjenigen erfasst werden, denen Projekte und Programme zugute kommen sollen (vgl. ähnliche Position Dolzer et al. 1998). Subjektivität wird dabei ausdrücklich als notwendig erachtet.

Der kulturelle Kontext wird bei sozialen Wirkungsanalysen und partizipativen Methoden stärker anerkannt und berücksichtigt, bleibt aber auch in diesen Fällen ein relevantes methodisches Problem. Ohne Berücksichtigung der kulturellen Hürden bei der Wirkungserfassung bestreiten Marsden/Oakley allerdings die Aussagekraft herkömmlicher Bewertungskriterien für Entwicklungsprojekte:

„The very nature of many social development projects and programmes makes it difficult, if not impossible, to make intelligent statements about efficiency, effectiveness, viability, and impact unless the specific cultural and socio-political environmental conditions of the particular place are given central attention“ (Marsden/Oakley 1990b: 8).

Übersicht 4: Bewertung der Sozialen Wirkungsanalyse im Hinblick auf die Kernfragen	
Entwicklungsbegriff	Diese Methode knüpft an einen Entwicklungsbegriff an, der die Verbesserung der Lebensverhältnisse in den Mittelpunkt stellt und dabei nicht allein auf die Veränderung ökonomischer Faktoren vertraut. Entwicklung wird als ein Prozess verstanden, der gleichberechtigte Teilhabe aller ermöglichen soll.
Vollständigkeit / Breite des Ansatzes	Begrenzt auf eine, wenn auch zentrale Dimension - die sozialen Wirkungen.
<i>Input-</i> oder <i>output</i> -Orientierung	Eindeutig <i>output</i> -orientiert: erfasst werden soll, was bei den Zielgruppen ankommt.
Aussageebene (Mikro/Makro)	Betrifft eher die Mikroebene, gerade in Verbindung mit zielgruppenorientierten, partizipativen Instrumenten.
Aussagekraft über Ursache-Wirkungszusammenhänge	Wirkungszusammenhänge werden aus der Sicht der Betroffenen dargestellt und analysiert. Der Ansatz ist damit subjektiv und soll es auch sein.
Demokratische Basis	Hohe demokratische Legitimation, wenn partizipative Methoden einbezogen werden.

⁵⁴ In der Durchführungsphase entwickeln die Zielgruppen Lebenslauflinien, Trendanalysen, Aktivitätenlisten und eine Einflussmatrix. Ergänzt wird diese Form der Datenerhebung durch die direkte Beobachtung der Evaluierer.

Verständlichkeit / Anwendbarkeit	Nachvollziehbare Ergebnisse mit vertretbarem Zeit- und Kostenaufwand. Bislang vergleichsweise geringe Anzahl durchgeführter Analysen.
----------------------------------	---

1.3.5 Partizipative Verfahren

Stellenwert

Die Mehrzahl der Evaluierungen wird entweder von Mitarbeitern der Geberinstitutionen oder deren Auftragnehmern durchgeführt. Die Zielgruppen der jeweiligen Maßnahmen sind meist nicht daran beteiligt. An diesem Punkt setzen neue, partizipative Verfahren an, die explizit auch aus der Kritik an den bislang gängigen Evaluationsmethoden entstanden sind und die Teilhabe der Betroffenen in Evaluierungen ermöglichen wollen:

„the inability of methods such as social cost-benefit analysis and social impact assessment to accurately and adequately reflect the dynamics of such changes has resulted in increasing calls for the involvement of those more aware of, and better able to explain, qualitative developments: i.e. the so-called ‚project beneficiaries‘ themselves“ (Marsden et al. 1994: 9).

Partizipative Verfahren haben in den 90er Jahren vor allem auch in entwicklungspolitischen Zusammenhängen an Bedeutung gewonnen. Sie beruhen auf einem vergleichbaren Entwicklungsverständnis wie soziale Wirkungsanalysen: Entwicklung wird als ein Prozess begriffen, der wesentlich durch die Beteiligten geprägt wird.

Die Diskussion um partizipative Verfahren begann in den 80er Jahren und wurde mit Verzögerung in der Praxis aufgenommen. Bryk (1983) war einer der ersten Wissenschaftler, der eine stärkere Partizipation bei Evaluierungen forderte. Die Diskussion wurde durch Guba und Lincoln (1987) vorangetrieben. Sie postulieren eine *Fourth Generation of Evaluation*⁵⁵, die vor allem die Rolle des Evaluierer verändert, indem er als Mediator und Katalysator in den Prozess eingreift. Den subjektiven Wahrnehmungen der Beteiligten kommt nach der Forderung von Guba/Lincoln eine wichtige Rolle bei der Realitätskonstruktion zu.⁵⁶ Die Diskussion um partizipative Methoden ist damit auch eine Diskussion um unterschiedliche Wissenschafts- und Politikverständnisse: Was ist Wahrheit? Was ist das Erkenntnisinteresse? Wie wird Realität konstruiert oder erfasst? Welche Verpflichtung besteht gegenüber den Beteiligten?⁵⁷

⁵⁵ Laut Guba und Lincoln (1987) wechselte die Rolle des Evaluierers vom Techniker (1. Generation), zum Beschreibenden (2. Generation), über den Urteilenden (3. Generation) zum Mediator (4. Generation).

⁵⁶ House (1980) hat sich bereits früher diesem Punkt in Bezug auf Evaluierungen Bedeutung beigemessen.

⁵⁷ Der Kritische Rationalismus postuliert die Existenz einer objektiven, erfahrbaren Realität, wohingegen Guba/Lincoln (1987) im Sinne des Konstruktivismus erklären, Realität könne nicht entdeckt werden, sondern entstehe jeden Moment durch die Teilnehmer und aus Sicht der Teilnehmer neu.

Partizipative Verfahren gehen von der Annahme aus, dass die Wirkung von Maßnahmen nicht vollständig und ausreichend erfasst werden kann, wenn die Wahrnehmung der Beteiligten nicht aufgenommen und analysiert wird.

Definition

Partizipative Evaluationsmethoden gehören zu den qualitativen Verfahren und zeichnen sich dadurch aus, dass sie unter Mitwirkung der Programmbeteiligten nicht nur die Informationen erheben, sondern auch Interpretationen vornehmen. Zentrales Merkmal partizipativer Verfahren ist daher die Mitwirkung der an einer Maßnahme beteiligten Personen bzw. Zielgruppen. Dies beinhaltet sowohl die Mitarbeiter einer durchführenden Organisation als noch in stärkerem Maße die Begünstigten und Betroffenen eines Programms. Während die konventionelle Evaluierungspraxis die Projektbeteiligten weitgehend ignoriert oder sogar als hinderlich für die Ermittlung objektiver Ergebnisse ansieht, bedeuten partizipative Evaluierungen einen völligen Wechsel der Perspektive der Evaluierer. Dem Wissen und der Wahrnehmung der Projektbeteiligten wird entscheidende Bedeutung zugemessen (vgl. Guba/Lincoln 1987, Krimmel et al. 1990: 19, UNDP 1996 und Gavai 1996).

Das grundsätzliche Schema einer Evaluierung bleibt erhalten, doch Daten und Interpretation kommen anders zustande:

„Participatory evaluation incorporates the basic components of conventional evaluation – planning, data-gathering and analysis, and dissemination – and infuses them with three essential elements: participation, negotiation and action“ (UNDP 1996: 34).

Vorgehen

Zu welchem Zeitpunkt, in welchem Maße und mit welchem Instrumentarium die Partizipation ermöglicht wird, wird sehr unterschiedlich gehandhabt. Grundsätzlich sollte Partizipation in allen Evaluierungsphasen ermöglicht werden, in der Praxis beschränkt sich die Beteiligung oftmals jedoch lediglich auf die Endphase einer Evaluierung.

Charakteristisch für partizipative Verfahren sind (vgl. z. B. Marsden et al. 1994, Neubert 1999, Dolzer et al. 1998, USAID 1996a und UNDP 1996: 29):

- Nutzung lokaler Ressourcen,
- Anerkennung des Wissens der Projektbeteiligten;
- Demonstration der Fähigkeiten der Projektbeteiligten;
- Einbeziehung der Beteiligten in den Entscheidungsprozess;
- Lernprozess;
- Flexibilität;
- Meinungsvielfalt wird widerspiegelt.

Im Unterschied zu traditionellen Evaluationsverfahren bezieht sich *ownership* damit nicht nur auf die Geber, sondern ebenfalls die Beteiligten und Betroffenen eines Projektes/Programms. Partizipative Evaluationsverfahren legen das Schwergewicht nicht auf die Überprüfung der *accountability*, sondern auf die Unterstützung eines Lernprozesses.

Vorteile

Ein zentrales Argument für partizipative Evaluationsmethoden ist die hohe Integrität und demokratische Basis in Bezug auf die Beteiligten. Vertreter der partizipativen Verfahren postulieren für ihren Ansatz eine Übereinstimmung zwischen dem Entwicklungskonzept und der Evaluierungspraxis, die bei traditionellen Verfahren nicht gewährleistet sei (vgl. Bamberger 1991: 331ff., Marsden et al. 1994 und Neubert 1999: 62f.).

Partizipativen Evaluierungsmethoden wird auch eine wichtige Rolle für den Lernprozess in Programmen zugeschrieben, Belege dazu stehen bislang noch aus. Sowohl die Mitarbeiter als auch die Zielgruppen sind an einem Lernprozess beteiligt, der im Idealfall die weitere Programmgestaltung positiv beeinflusst. Diese qualitative Dimension bezieht sich nicht nur auf die Evaluierungsergebnisse, sondern die Teilhabe an dem Prozess fördert die Fähigkeiten aller Beteiligten und trägt somit zum *capacity building* bei.

Des Weiteren stellt sich bei Evaluierung wiederholt nicht nur die Frage nach der Qualität der Datensammlungen in Entwicklungsländern, sondern auch nach der Legitimität der Informationen, die mitunter von undemokratischen Regierungen erhoben wurden.⁵⁸ Dieser Widerspruch kann durch Partizipation der Betroffenen teilweise aufgelöst werden.

Eine Vergrößerung der Akzeptanz der Evaluierungsergebnisse kann durch eine Beteiligung der Betroffenen an Evaluierungen selbst realisiert werden; gleichzeitig wird dadurch der Lernprozess intensiviert und die Kohärenz partizipativer Projektplanungs-, Durchführungs- und Evaluierungsansätze hergestellt.⁵⁹

Nachteile

Der Hauptvorwurf an partizipative Verfahren ist die mangelnde Validität der erhobenen Daten, bei einer Untersuchung, in der der Forschungsgegenstand zum Subjekt wird (vgl. z.B. USAID 1996a).⁶⁰

Bislang liegen noch keine umfassenden Untersuchungen zur Validität partizipativer Evaluierungen vor. Entsprechende Untersuchungen in den nächsten Jahren werden aber auch die paradigmatischen Differenzen nicht beilegen können. Vertreter partizipativer Ansätze verwerfen diesen Einwand heute bereits auf der Grundlage eines konstruktivistischen Realitätskonzeptes, das Subjektivität nicht als Hindernis, sondern notwendigen Bestandteil der Wahrheitsfindung versteht (so z.B. House 1980, Guba/Lincoln 1987, Marsden et al. 1994 und Neubert 1999).

Im Rahmen partizipativer Evaluierungen ist von den Zielgruppen keine Beurteilung der Kosten eines Vorhabens zu erwarten. Solange es nicht um die alternative Ressourcenverteilung geht, an der die Zielgruppen beteiligt werden können, in dem sie den alternativen Nutzen für sie bewerten sollen, können mit partizipativen Methoden keine Aussagen über die Effizienz von Maßnahmen getroffen werden.

⁵⁸ Dieser Aspekt findet viel zu selten Erwähnung, eine der Ausnahmen findet sich bei Rist (1995: 167f.).

⁵⁹ In diesem Sinne sind Partizipation und Nachhaltigkeit eng miteinander verknüpft, so z.B. bei Schuster/Pinger (1998: 160).

⁶⁰ Verschiedene Kritikpunkte werden angeführt in Marsden et al. (1994: 112ff.).

Zusätzliche Kritikpunkte sind der höhere personelle und zeitliche Aufwand und der höhere Stellenwert von Einzelinteressen. Partizipative Verfahren stellen auch hohe Anforderungen an die Methoden, wenn dem Einwand mangelnder Validität begegnet werden soll.

Anwendung in Entwicklungsländern

In der Entwicklungszusammenarbeit entstand aus der aufkommenden Diskussion um partizipative Verfahren in der ländlichen Entwicklung in den 80er Jahren auch die Idee der Fortführung dieses Ansatzes in Evaluierungen. Vor allem bei entwicklungspolitischen NRO und im UN-Bereich entwickelte sich das Bedürfnis, partizipative Elemente in die Evaluierungen einzubeziehen. Wenn größere Nachhaltigkeit erzielt werden soll, ist Beteiligung und Stärkung der Managementfähigkeiten der Zielgruppe ein wichtiges Ziel. Die Ausweitung der Partizipation auf den Evaluationsprozess schließt sich unmittelbar daran an. Hinter dieser Position steht eine grundsätzliche Kritik an dem vorherrschenden Entwicklungsparadigma (vgl. z.B. Marsden et al. 1994: 31f.).

Gerade für Evaluierungen in Entwicklungsländern stellt sich die besondere Problematik des kulturellen *bias* bzw. des Machtgefälles zwischen Gebern und Nehmern, Projektdurchführung und den Zielgruppen sowie anderen Beteiligten. Auch Evaluierungen, die in der Praxis im Extremfall über Fortführung einer Maßnahme entscheiden, stellen ein Machtinstrument dar.

Gleichzeitig stellen gerade die kulturellen Divergenzen zwischen Evaluierern und Beteiligten in Projekten ein Problem für partizipative Verfahren dar. Es besteht die Gefahr, dass westliche Kommunikationsmuster auf die Zielgruppen übertragen werden oder traditionelle, einer Partizipation aller Beteiligten entgegenstehende Strukturen, gestützt werden (vgl. Neubert 1999: 62f.). Diese Unterschiede können aber auch hilfreich sein, wenn es darum geht, Partizipation trotz traditioneller Machtverhältnisse (z.B. *gender*) zu ermöglichen.

Trotz der höheren Beteiligung verändern auch partizipative Verfahren nicht die einseitige Planungshoheit auf Seiten der Geber. Mutter weist daher darauf hin, dass partizipative Evaluierungen das methodische Machtgefälle nicht aufheben (1998: 145).⁶¹ Erfasst werden sollen die konkreten Wirkungen von Maßnahmen, eine Überprüfung der Prioritätensetzung der Geber aus Sicht der Beteiligten ist darin nicht enthalten.

Dazu gehört, dass die Forderungen nach einer stärkeren Partizipation der Projektbeteiligten nicht nur auf die Ebene der Wirkungserfassung zielt, sondern teilweise einen instrumentellen Charakter hat: durch partizipative Ansätze soll der Projekterfolg verbessert werden (vgl. World Bank 1994b).

⁶¹ Siehe dazu auch die Diskussion um *pseudoempowerment* bei Mark/Shotland (1987: 143).

Übersicht 5: Bewertung partizipativer Verfahren im Hinblick auf die Kernfragen	
Entwicklungsbegriff	Ein Hauptkriterium der Entwicklungszusammenarbeit, tatsächlich die Zielgruppen zu erreichen, kann hierdurch überprüft werden – vor allem wenn es sich um schwer erreichbare Zielgruppen z.B. im Rahmen der Armutsbekämpfung geht. Entwicklung wird als ein Prozess begriffen, der eine demokratische Basis haben muss und positive Veränderungen der Lebensverhältnisse der armen Bevölkerungsgruppen als Folge haben sollte.
Vollständigkeit / Breite des Ansatzes	Schwerpunkt dieses Ansatzes sind Aspekte unmittelbarer, sozialer Wirkungen auf die Betroffenen. Andere Ebenen wie z.B. ökologische oder ökonomische Wirkungen werden ausschließlich in ihrem Bezug auf die Lebensverhältnisse der Betroffenen erfasst. Ein Problem ist dabei die Eingrenzung auf die Beteiligten. Wenn dies zu eng gefasst ist, können darüber hinaus gehende Wirkungen übersehen werden.
<i>Input-</i> oder <i>output</i> -Orientierung	Der <i>input</i> eines Programms wird von den Beteiligten überhaupt nicht erfasst oder bewertet. Bislang wird von den Beteiligten lediglich die <i>output</i> -Seite berücksichtigt, während die Evaluierer dies in Bezug zu dem Aufwand eines Programms setzen. Auf der <i>output</i> -Seite werden mit partizipativen Verfahren gezielt langfristige Wirkungen erfasst und keine Beschränkung auf die Ergebnisse vorgenommen.
Aussageebene (Mikro/Makro)	Aussagen beziehen sich auf die Mikro-Ebene, die unmittelbaren Lebensverhältnisse der Betroffenen.
Aussagekraft über Ursache-Wirkungszusammenhänge	Die Herstellung eines Vorher-Nachher-Vergleichs mittels partizipativer Verfahren ist eher schwach. Ausnahmen stellen vielleicht SWAP und <i>beneficiary assessment</i> dar.
Demokratische Basis	Hohe demokratische Legitimation der Evaluationsergebnisse.
Verständlichkeit / Anwendbarkeit	Verfahren setzen hohe methodische, interkulturelle Kompetenz voraus. Dieser Auf-

	wand kann dadurch legitimiert werden, dass die Evaluierung selbst Entwicklungsprozesse fortführen kann. Entscheidungsträger sind mehrheitlich skeptisch gegenüber den Verfahren und der Vernachlässigung der <i>input</i>-Seite.
--	--

1.3.6 Methodenüberblick

Übersicht 6: Vergleichende Bewertung der Evaluationsmethoden im Hinblick auf die Kernfragen					
	Kosten-Nutzen-Analyse	<i>Environmental impact assessment</i>	<i>Impact analysis</i>	Soziale Wirkungs-analyse	Partizipative Evaluationsmethoden
Qualitativ oder quantitativ	Quantitativ	Überwiegend quantitativ	Quantitativ und qualitativ	Überwiegend qualitativ	Qualitativ
Entwicklungsbegriff	Ökonomisch geprägt	Nicht Entwicklung sondern Umweltwirkungen sollen erfasst werden	Abhängig vom Forschungsansatz, überwiegend breiter Entwicklungsbegriff	Breit, armutsorientiert	Breit, armutsorientiert
Vollständigkeit, Breite des Ansatzes	Fokus auf ökonomische, quantifizierbare Faktoren	Ökologischer Fokus	Alle Wirkungsebenen	Fokus auf sozialer Dimension	Starke soziale Komponente, Armutsorientierung
Mikro- oder Makroebene	Überwiegend mikro	Mikro	Mikro- und Makroebene	Überwiegend Mikroebene	Mikro
Aussagekraft Kausalzusammenhang	hohe Validität, geringe Erklärungskraft	Orientiert auf Vorhersagen	Hohe qualitative Aussagekraft	Hohe Erklärungskraft	Zusammenhänge aus Sicht der Betroffenen erklärt
Anwendbarkeit in Entwicklungsländern	Bislang unzureichende Anpassung	Bislang unzureichende Anpassung	Noch Anpassungsprobleme	Methodische Probleme, aber Instrumentarium für Entwicklungsländer entwickelt	Einige Instrumente im entwicklungspolitischen Kontext entwickelt
Demokratische Basis	Gering	Gering, abhängig von Methode	Abhängig von Methoden	Eher hoch	hoch
Verständlichkeit	Mittel	Mittel	Mittel	Mittel	Mittel
Anwendbarkeit	Mittel	Mittel	Gering	Gering	Gering

Aus dem Überblick über relevante Ansätze in der Evaluationsforschung ergibt sich ein relativ uneinheitliches Bild. Mehrheitlich wird zwar inzwischen ein Methodenpluralismus

propagiert (z.B. Picciotto 1994, Baker 1999, Lee 2000), andererseits stehen sich manche Positionen sehr kritisch und scheinbar unvereinbar gegenüber.⁶²

Den Hintergrund dieser Auseinandersetzungen bilden ideologische Kontroversen, die sich auf ein unterschiedliches Wissenschaftsverständnis und – in entwicklungspolitischer Hinsicht – auf unterschiedliche Entwicklungsbegriffe beziehen:

- die Begriffe von *Wahrheit*, *Objektivität* bzw. *Subjektivität*, *Realität* sowie die Aufgabe der Wissenschaft werden ganz unterschiedlich gewichtet und führen daher auch zu unterschiedlichen Beurteilungen der jeweiligen Methode;
- ökonomisch-technische und soziale und ökologische Ansätze gehen von unterschiedlichen Entwicklungszielen und Erklärungen für Entwicklungsprozessen aus. Daher unterscheiden sich auch die untersuchten Wirkungsebenen wesentlich. Hinzu kommen kontroverse Auffassungen über die Beziehungen und Verpflichtungen zwischen Gebern und Projektbegünstigten in Entwicklungsländern, die ebenfalls Konsequenzen für Evaluierungen haben.

Kausalzusammenhang/Validität

Kausalzusammenhänge methodisch einwandfrei zu belegen, erweist sich nicht unerwartet als Schwäche der Evaluationsforschung. So einhellig alle der Meinung sind, nur Experimente würden den höchsten wissenschaftlichen Ansprüchen genügen, so einig ist man sich, dass das in der Regel nicht machbar und wünschenswert ist.

Ohne eine fundierte Verbindung von quantitativen und qualitativen Verfahren, bleibt die Wahl zwischen statistisch abgesicherten Ergebnissen und qualitativen Aussagen, deren Verallgemeinerungsgrad ungesichert ist, deren Erklärungskraft für komplexe soziale Prozesse und deren demokratische Basis aber deutlich größer ist.

Aussageebene

Die Aussageebene einer Evaluierung bestimmt sich hauptsächlich durch den jeweiligen Gegenstand. Gleichwohl gibt es Methoden wie z.B. partizipative Verfahren und soziale Wirkungsanalysen, die auf regionaler, sektoraler oder gar nationaler Ebene in vertretbarem Umfang und Aufwand nicht anwendbar sind. Vergleicht man die Tragweite der Aussagen aller Evaluationsansätze auf Projektebene, kann am ehesten der Kosten-Nutzen-Analyse, dem EIA und eingeschränkt der *impact analysis* eine Aussagekraft auf Makroebene zugesprochen werden.

Anwendbarkeit für Entwicklungsländer

Die Methodenentwicklung in der Evaluationsforschung ist bislang fast ausschließlich auf die Anforderungen und Realitäten der Industrieländer zugeschnitten. Gerade bei den beiden Verfahren, die am weitesten verbreitet sind, der Kosten-Nutzen-Analyse und dem EIA, ist dies sehr deutlich und es gab bislang nur unzureichende Anpassungsleistungen auf die spezifischen Situationen in Entwicklungsländern. Da die Weiterentwicklung partizipativer Evaluationsverfahren und sozialer Wirkungsanalysen gerade auch im entwicklungspolitischen Zusammenhang vollzogen wurde, gibt es hier größere

⁶² Vgl. z.B. Marsden et al. (1994) für partizipative, soziale Wirkungserfassung und Squire (1995) als Plädoyer für experimentelle bzw. quasi-experimentelle Verfahren.

Anstrengungen, die Methode dem jeweiligen Untersuchungsgegenstand anzupassen. Die kulturelle Dimension wird aber in allen Ansätzen noch unzureichend berücksichtigt.

Einen wichtigen Beitrag zur Evaluationsforschung könnte aus den Ländern des Südens selbst kommen, wenn finanzielle Ressourcen, *ownership* und auch demokratische Basis die Entwicklung eigener Evaluationskapazitäten ermöglichen.

Entwicklungsbegriff

Die Beschränkungen und methodischen Defizite der Kosten-Nutzen-Analyse bei der Erfassung von Entwicklungsprozessen konnten auch durch Modifizierungen bislang nicht beseitigt werden. Folgt man dem Ziel der Armutsbekämpfung bzw. der menschlichen Entwicklung, sind *impact*-Analysen, partizipative und soziale Wirkungsanalysen geeigneter, die Komplexität und das Prozesshafte von Entwicklung zu erfassen.

Die vorliegende Arbeit unternimmt den Versuch einer Wirkungsanalyse, die alle Dimensionen des Entwicklungsbegriffs erfassen will. Da ein experimentelles Design außer Frage steht und der Betrachtungsgegenstand der vorliegenden Untersuchung über ein Einzelprojekt hinausgeht, werde ich für die Sektoranalyse sowohl quantitative als auch qualitative Sekundäranalysen aufgreifen. Zudem werden durch Interviews qualitative Primärdaten herangezogen.

Unabhängig von der jeweiligen Methode gilt für alle Verfahren, dass sinnvolle Antworten zu finden voraussetzt, die richtigen Fragen gestellt zu haben. Ein qualitatives Vorgehen, das die relevanten Fragestellungen in Bezug auf eine Maßnahme herausarbeitet, ist nach dem Verständnis der Autorin rein quantitativen Analysen vorzuziehen. Die Verwendung eines breiten Entwicklungsbegriffs, der verschiedene Wirkungsebenen betrachtet, ist wichtig, um den komplexen Entwicklungsprozess verstehen und analysieren zu können. Die Einbeziehung der Wahrnehmung der Zielgruppen ist dabei ein relevantes Kriterium. Quantitativ valide Daten über beispielsweise Einkommenszuwächse in Dorf X durch Aquakulturen sind nicht ausreichend für die Bewertung der sozio-ökonomischen Dimension, wenn etwa die Wahrnehmung der Dorfbewohner ist, dass sich ihre Lebenssituation verschlechtert hat (z.B. durch Trennung der Familien). Ziel dieser Arbeit ist daher eine Verbindung verschiedener Bewertungsansätzen.

2 Entwicklungspolitische Evaluationspraxis

Stellenwert

Die Entwicklungszusammenarbeit ist ein Politikfeld, in dem zahlreiche Evaluierungen jährlich durchgeführt werden und das erste Politikfeld, in dem in Europa Evaluierungen begonnen wurden.⁶³ Bei einer Befragung im Auftrag des DAC ergaben sich für den Zeitraum 1991 bis 1998 eine Zahl von 21.500 Evaluierungen, die von den beteiligten bilateralen oder multilateralen Gebern durchgeführt wurden.⁶⁴ Gleichwohl wird nur ein geringer Prozentsatz (10-20% nach DAC-Schätzungen) der Gesamtvorhaben evaluiert.

Begründung

Evaluierungen erfüllen auch in der Entwicklungszusammenarbeit den Zweck, Legitimation zu schaffen und die sinnvolle Verteilung knapper Ressourcen und Lerneffekte zu ermöglichen. Die hohe Bedeutung von Evaluierungen in diesem Politikfeld resultiert jedoch vor allem daher, dass die Durchführung der Projekte und Programme bzw. die Zielgruppe der Maßnahmen sich in einem anderen Land befinden. Nationale Ressourcen werden für Projekte und Programme in anderen Ländern verwandt, die gerade in Zeiten knapper Haushaltsmittel nachweisen müssen, dass sie positive Wirkungen erzielen. Dies verstärkt den Legitimationszwang und stärkt bei den staatlichen Institutionen das Bedürfnis, die Ergebnisse der Programme genauer zu überprüfen (vgl. Lachenmann 1977, Bolay 1988, Berlage/Stokke 1992, Stockmann 1996, DAC 1998b).

Die Komplexität des Vorhabens, Entwicklungsprozesse in anderen Ländern zu initiieren oder zu unterstützen, ist zudem ein starker Motor für Evaluierungen:

„Evaluation provides an essential validity check for the assumptions, theories, and forecasts that underlie development activities“ (Serageldin 1995: 23).

Mittels Evaluierungen soll eine Qualitätsverbesserung erzielt werden, indem erfolgreiche Maßnahmen ausgebaut bzw. problematische Aktivitäten verändert oder beendet werden können. Gerade die Berichte über zahlreiche, erfolglos verlaufende Projekte in den 60er Jahren, gaben den Anlass für stärkere Evaluierungsaktivitäten (vgl. Coleman 1992: 498).

In der breiten öffentlichen Diskussion werden Evaluierungen kaum wahrgenommen. Ein Grund hierfür ist die teilweise zurückhaltende Umgangsweise der zuständigen Institutionen mit den Evaluationsergebnissen. Oftmals leiden allerdings auch die Ergebnisse darunter, dass sie stark unter dem Legitimationszwang stehen und die Gutachter ent-

⁶³ Die Anfänge der Etablierung von Evaluierungen in der Entwicklungszusammenarbeit lagen in den 60er Jahren, mit einer Zunahme der Projektaktivitäten in den 80er Jahren, kam es zu einer starken Ausweitung der Evaluierungsaktivitäten (vgl. Cracknell 1988, Stokke 1991b, Berlage/Stokke 1992). Neben der Entwicklungspolitik sind als weitere Politikbereiche, in denen Evaluierungen stärker genutzt werden, die Bildungs- und Agrarpolitik, die Arbeitsmarkt- und Sozialpolitik sowie die Forschungs- und Technologiepolitik zu nennen.

⁶⁴ Zu den beteiligten multilateralen Gebern gehörten u.a. die Weltbank und UNDP. NRO-Evaluierungen gingen nicht in diese Zahl ein, gleichzeitig geben die Autoren selber die genannte Zahl aufgrund von Doppelnennungen als sehr wahrscheinlich etwas zu hoch an (vgl. DAC 1998: 21).

weder direkt der Institution angehören oder von Anschlussaufträgen abhängig sind.⁶⁵ Außerhalb der USA und Kanadas bestehen zudem starke Defizite, was die institutionelle Verankerung der Evaluationsforschung in Politik und Wissenschaft angeht (vgl. Bolay 1988: 90f. und Stockmann 1996).

Spezifika des Anwendungsgebietes

Das Politikfeld Entwicklungszusammenarbeit stellt besondere Anforderung an Evaluationsmethoden (siehe oben). Die besonderen Schwierigkeiten, die an die Entwicklungszusammenarbeit insgesamt gestellt werden, d.h. unter anderen sozio-kulturellen Bedingungen zu arbeiten, gelten selbstverständlich auch für Evaluierungen. Als zentrale Punkte sind die großen kulturellen Unterschiede zu nennen, eine schlechtere Datenlage und ethische Probleme aufgrund extremer Armutssituationen.

Daran schließt sich an, dass das Machtgefälle zwischen Gebern und Nehmern auch bei den Evaluierungen besteht. Evaluierungen werden fast ausschließlich von Gebern veranlasst und mehrheitlich von Gutachtern aus den Geberländern durchgeführt.

Erscheinungsformen

Die Unterschiede zwischen den verschiedenen Institutionen der Entwicklungszusammenarbeit hinsichtlich ihrer Evaluationspraxis sind sehr groß. Vor allem bestehen Unterschiede hinsichtlich

- der Zahl von Evaluierungen,
- der institutionellen Verankerung von Evaluierungen,
- der zugrundeliegenden Prinzipien,
- der verwendeten Methoden.

In der Praxis relativieren sich einige dieser kategorischen Unterschiede, da die jeweiligen spezifischen Richtlinien nur bedingt verfolgt werden. Mehrheitlich werden in Evaluierungen nur Mitteleinsatz, Aktivitäten und Ergebnisse erfasst, weil die Zeit und die Mittel nicht zur weitergehenden Wirkungsanalysen ausreichen.

Eine genauere Betrachtung der zentralen Akteure und der verwendeten Methoden der Evaluierungen in der Entwicklungszusammenarbeit soll Hinweise geben, die im Weiteren sowohl die Erarbeitung der methodischen Grundlage dieser Arbeit als auch die Bewertung der vorliegenden Evaluationsergebnisse zum Thema Aquakultur ermöglichen sollen.

2.1 Akteure

Überblick

Ob entwicklungspolitische Akteure ihre Aktivitäten mit staatlichen Geldern oder Spenden finanzieren – sie wollen und müssen den Nachweis über die positiven Effekte ihrer Maßnahmen leisten. Bis auf wenige Ausnahmen, d.h. im Wesentlichen Nicht-

⁶⁵ Die Problematik der Unabhängigkeit und Unparteilichkeit von Evaluierern wird daher regelmäßig thematisiert (vgl. z.B. DAC 1991, Stokke 1991b, Chelimsky 1995, Stockmann 1996, Caracelli 2000, Widmer 2000).

regierungsorganisationen mit einem geringen Mittelaufkommen, führen alle entwicklungspolitischen Akteure Evaluierungen durch.

Finanzierungsform und –höhe, nationale Traditionen, aber auch politische Einstellungen bestimmen unterschiedliche Evaluierungsschwerpunkte bei multilateralen und bilateralen Gebern bzw. Nichtregierungsorganisationen. Zwischen den unterschiedlichen Akteuren gibt es zwar eine gemeinsame Diskussion und große Übereinstimmung bei den Leitlinien für Evaluierungen, doch gleichzeitig auch teilweise Unterschiede in der Evaluierungspraxis (vgl. z.B. Mosley 1983, Beiträge im Band von Stokke 1991a, DAC 1998b, Brüne 1998b). Zu wenige Versuche gibt es bislang, geberübergreifend zu evaluieren.⁶⁶

Die Kapazität für eigene Evaluierungsaktivitäten der Entwicklungsländern ist überwiegend unzureichend. Einige Länder wie Indien, aber auch die asiatischen Länder insgesamt, haben eine Evaluierungstradition staatlicher Aktivitäten und entsprechende Strukturen und Forschungsaktivitäten.⁶⁷ In der Mehrzahl der Länder fehlen diese jedoch. Ein Grund ist die Ressourcenknappheit, die eine Mittelverwendung für Evaluierungen nicht möglich erscheinen lässt. Gleichzeitig sind Evaluierungen ein mögliches Kontrollinstrument und haben daher in Ländern mit geringer demokratischer Kultur keinen hohen Stellenwert.⁶⁸

In den 90er Jahren ist eine Debatte über eine Stärkung der eigenen Evaluierungskapazitäten von Entwicklungsländern aufgekommen (vgl. Bamberger 1991, Wiesner 1993). Vor allem die Weltbank möchte durch Eigenanstrengungen der Entwicklungsländern in diesem Bereich den Ansatz staatlicher Reformen verbessern.

Übergeordnete Richtlinien

Die im DAC⁶⁹ vertretenen Länder haben sich 1991, auf einheitliche Evaluierungsrichtlinien geeinigt, die *Principles for Evaluation of Development Assistance* (DAC 1991). Das Schwergewicht dieser Prinzipien liegt auf der Betonung von Unvoreingenommenheit, Unabhängigkeit, Glaubwürdigkeit sowie der Institutionalisierung einer zentralen Evaluationsplanung. Zudem wird die Notwendigkeit der Verwendbarkeit von Evaluierungen, der Partizipation und der Geberkoordination betont. Die Evaluationsmethoden spielen in den Richtlinien kaum eine Rolle, es erfolgt lediglich der Hinweis, „that in most cases evaluation will involve a combination of quantitative and qualitative techniques“ (DAC 1991: 10).

Formale Gesichtspunkte stehen bei diesen Richtlinien im Vordergrund. Eine Einigung über inhaltliche Kriterien, die Vorgaben über Fragestellungen, Vorgehensweise und Verwendung der Ergebnisse bedeutet hätten, war nicht zu erzielen. In der vorliegenden

⁶⁶ Ein wichtiger Versuch in diese Richtung ist die angestrebte Entwicklung eines *Performance Measurement System*, das anhand quantitativer Indikatoren die Qualität der Durchführung sektorweit erfassen soll (vgl. Kumar 1995: 82).

⁶⁷ Nähere Informationen über die Entstehung des indischen Evaluationssystems bei Shah (1984).

⁶⁸ Teilweise abweichend dazu ist die Entwicklung in Asien zu sehen, wo staatliche Planung und damit auch Monitoring eine größere Rolle spielen und Evaluationssysteme weniger ein Ausdruck demokratischer Kontrolle sind (ausführlicher dazu Bamberger 1989).

⁶⁹ *Development Assistance Committee* (DAC) der OECD, 1961 gegründet mit Zweck der Koordination der Entwicklungszusammenarbeit der OECD-Staaten.

Form wurden die Richtlinien von allen wichtigen bilateralen Gebern und multilateralen Institutionen wie Weltbank, UNDP und Europäischer Kommission angenommen. Demgegenüber bleibt die Praxis der Geber sehr vielfältig.⁷⁰

Tabelle 1: Öffentliche Entwicklungszusammenarbeit (<i>Official Development Assistance</i>, ODA) verschiedener Gebergruppen (in Mrd. US-\$)				
	1996	1997	1998	1999
Bilaterale EZ	39,1	32,4	35,2	37,9
Multilaterale EZ	16,7	15,3	14,5	13,4
EZ der NRO	5,6	5,2	5,6	6,7
Quelle: DAC 2001a				

Multilaterale, bilaterale Geber und Nichtregierungsorganisationen finanzieren die Gesamtheit aller entwicklungspolitischen Maßnahmen. Die Finanzmittel, die von NRO aufgebracht werden, sind deutlich geringer, als die Beiträge anderer Institutionen (vgl. Tabelle 1). Da NRO aber auch im Hinblick auf Evaluationsmethoden einen spezifischen Standpunkt vertreten, soll auf sie im Weiteren noch einmal genauer eingegangen werden. Das gilt ebenso für multilaterale und bilaterale Geber, die nicht nur von ihrem Finanzvolumen, sondern auch von ihren Evaluationskapazitäten deutlich das Feld dominieren. Der zentrale Akteur ist in dieser Hinsicht die Weltbank, die daher auch im Mittelpunkt der Ausführungen zu multilateralen Gebern stehen soll.

2.1.1 Multilaterale Akteure

Stellenwert

Multilaterale Geber sind nicht nur was das Finanzvolumen der durchgeführten Maßnahmen, sondern ebenfalls im Hinblick auf die Zahl der veranlassten Evaluierungen, ein zentraler Akteur der Entwicklungszusammenarbeit.

In der jährlichen Zahl der Evaluierungen⁷¹ ist die Weltbank⁷² die absolut führende Institution in diesem Bereich. Neben der zahlenmäßigen Dominanz hat die Weltbank bei der institutionellen Verankerung und Methodenentwicklung teilweise eine bestimmende Funktion in der entwicklungspolitischen Evaluationsforschung eingenommen (vgl. Picciotto/Rist 1995, Stockmann 1996: 40). Indien ist zudem ein Land, dass von der Weltbank seit Ende der 90er Jahre verstärkt im Hinblick auf sektorale und sektor-über-

⁷⁰ Hinweise dazu gibt auch die vom DAC (1998) selbst veranlasste Überprüfung der Einhaltung der Richtlinien, auf die in Abschnitt II. 2.1.2 noch einmal näher eingegangen wird.

⁷¹ Für den jährlichen Bericht des *Operations Evaluation Department* (OED) der Weltbank wurden 2002 allein 348 Projektevaluationen (Zeitraum 9/2000 bis 10/2001) berücksichtigt. Davon bestanden 279 in *evaluative summaries* und 69 aus *project performance audit reports* (vgl. World Bank 2002a).

⁷² Zur Weltbankgruppe zählen die *International Bank for Reconstruction and Development* (IBRD) (die eigentliche Weltbank und als solche im Weiteren bezeichnet), die *International Development Association* (IDA), die *International Finance Corporation* (IFC) und die *Multilateral Investment Guarantee Agency* (MIGA).

greifende Fragestellungen evaluiert wird.⁷³ Zudem sind wesentliche Mittel der Aquakulturförderung in Indien von der Weltbank zur Verfügung gestellt worden (siehe Kapitel III.4.3.2.2). Aus diesem Grund soll an dieser Stelle näher vor allem auf diesen Geber eingegangen werden.

Evaluationssystem der Weltbank⁷⁴

Die Weltbank baute ihr Evaluationssystem Mitte der 70er Jahre auf und hat gerade mit der institutionellen Verankerung eine Vorbildfunktion übernommen. (vgl. z.B. Carvalho/White 1994, Sandström 1995, Thumm 1998). Die Weltbank versucht, die Unabhängigkeit der Evaluationsabteilung (*Operations Evaluation Department* (OED)) durch die direkte Zuordnung zum Vorstand und die Sonderstellung des Direktors des OED zu gewährleisten. Der Direktor kann nach fünf Jahren Amtszeit für eine weitere Amtsperiode wiedergewählt werden und muss danach die Weltbank verlassen. Diese institutionellen Eckpfeiler des Evaluationssystems der Weltbank sind wesentlich eingeflossen in die DAC Richtlinien für Evaluierungen in der Entwicklungszusammenarbeit (vgl. DAC 1991).

Mosley verglich 1983 die Evaluierungen von Weltbank und der (damaligen) *United Kingdom Overseas Development Administration* (UKODA) und stellte „certain systematic differences“ fest (Mosley 1983: 594). Seine Erklärung für die größere Zahl von Evaluierungen, mehr ex-post Evaluierungen und Angabe von ex-post *rate of return* bei der Weltbank, war die größere Unabhängigkeit der zuständigen Abteilung in der Weltbank, die direkt dem Präsidenten unterstellt ist und außerhalb der Organisation steht (Mosley 1983: 599f.). Die größere Unabhängigkeit der Evaluierer kann auch die Ergebnisse beeinflussen: Seitdem das OED der Weltbank die Ergebnisse erfasst, hat sich die Bewertung der Maßnahmen verschlechtert (vgl. Picciotto 1996: 12).

Die öffentliche Diskussion und Kritik an Maßnahmen der Weltbank hat dazu geführt, dass Sozial- und Umweltwirkungen stärker berücksichtigt werden sollen. Gleichwohl bleiben im Zentrum der Evaluierungen der Weltbank die Kosteneffektivität und *financial integrity*. Die Kosten-Nutzen-Analyse ist daher die dominante Evaluationsmethode der Weltbank. Vorrangig wird sie in der Projektprüfung eingesetzt, aber ebenfalls für ex-post Evaluierungen. Die *economic rate of return* ist der zentrale Bewertungsmaßstab für Weltbankevaluierungen.

Die Weltbank beteiligte sich seit den 70er Jahren an der Diskussion über die Kosten-Nutzen-Analyse und versuchte, auf die Kritik zu reagieren. Die rein quantitative und ökonomische Sicht wurde in Frage gestellt und das Instrumentarium der *social cost-benefit analysis* entwickelt, das stärker Verteilungsaspekte mit einbeziehen sollte.⁷⁵

⁷³ Vgl. verschiedene Publikationen zu Armut, *gender* und Umweltwirkungen der Weltbankmaßnahmen in Indien, World Bank (2000a, 2000b, 2000d) und Überblick über verschiedene Sektorpapiere in Zanini (2001).

⁷⁴ Der Standardablauf bei der Weltbank sieht vor, dass nach Abschluss des Projektes ein *Project Completion Report* durch die Mitarbeiter des Nehmerlandes verfasst wird, dazu gibt dann das OED mit einem *Project Performance Audit Report* eine Stellungnahme ab. Bei zehn Prozent aller Projekte gibt es fünf oder mehr Jahre nach Projektende *Impact Evaluation Reports*. Im Vergleich zu anderen Gebern ist das eine hohe Quote. Dennoch werden damit die langfristigen Wirkungen von 90% aller Projekte nicht erfasst.

⁷⁵ Dies erfolgte auf der Grundlage der Arbeit von Squire/van der Tak (1975).

Nicht zuletzt aufgrund starker Kritik an den ökologischen und sozialen Folgen von Weltbank-Projekten hat die Weltbank auch in den 80er Jahren versucht, die Methoden zu erweitern: EIA (vgl. Serageldin 1995) und SIA (vgl. Ingersoll 1990b) sollen mögliche negative Folgen rechtzeitig erfassen und in die Planung einbeziehen.

Der Wandel der Projektpolitik und -prüfung der Weltbank führte dazu, dass Mitte der 90er Jahre nur noch bei einem Drittel aller Projektprüfungen die *economic rate of return* ermittelt wurde bzw. werden konnte (vgl. Picciotto 1994: 226). Ökonomische und quantitative Verfahren dominieren jedoch weiterhin das methodische Handwerkszeug der Weltbank. Auch *impact*-Evaluierungen stützen sich auf quantitative, statistische Verfahren (vgl., Cameron 1993: 94, Davies 1996, Thumm 1998).

Kritiker sehen die Reformmaßnahmen bislang als halbherzig und unzureichend an (vgl. von Amsberg 1993, Marsden et al. 1994, Winpenny 1995: 108ff., Rich 1998). Die *social cost-benefit analysis* findet trotz der Weiterentwicklung durch Weltbank-Mitarbeiter auch zwanzig Jahre später nicht allgemeine Anwendung (vgl. Brent 1998: 288).

Die institutionelle Verankerung des Evaluationssystems in der Weltbank ist beispielgebend.⁷⁶ Sehr viel zurückhaltender fiel die Anpassung des methodischen Instrumentariums an ökologische oder soziale Herausforderungen aus. Die ökonomische Sichtweise auf den Projekt- bzw. Programmerfolg dominiert weiterhin. Die ökonomisch geprägte Philosophie der Institution schlägt sich auch in der Evaluationspolitik nieder.

Interne Auswertungen der Evaluierungen von Seiten der Weltbank selbst haben Anhaltspunkte ergeben, dass es eine Tendenz gibt, die positiven Wirkungen von Maßnahmen zu überschätzen und die Analyse zu sehr auf die *economic rate of return* zu beschränken (vgl. z.B. Little/Mirrlees 1991).

Die Weltbank gehört jedoch zu den Institutionen, die – auch aufgrund ihrer finanziellen Stärke – den Fokus bei Evaluierungen auf andere Bereiche als das Einzelprojekt ausgedehnt und thematische, Länder- und Sektoranalysen angefertigt hat. Diese projektübergreifenden Evaluierungen sind aber noch zahlenmäßig gering. Sektoranalysen liegen zum Beispiel für den Energie-, Gesundheits- Bildungs- oder Forstsektor vor (vgl. z.B. für den Energiesektor in Indien World Bank 1999b).

Evaluierungen im UN-System

Das Entwicklungsprogramm der Vereinten Nationen, UNDP, ist für das UN-System der maßgebliche Akteur für Evaluierungen der Entwicklungszusammenarbeit. Alle Vorhaben mit einem Budget ab 1 Mio. US-\$ müssen evaluiert werden, darüber hinaus werden vor allem innovative oder strategisch bedeutsame Projekte evaluiert.

Das UNDP setzt bei Evaluierungen deutlich stärkere Akzente im Hinblick auf Partizipation der Entwicklungsländer bzw. der Zielgruppen. 1995 wurde ein Rahmen für einen partizipativen Evaluierungsansatz formuliert (vgl. UNDP 1996: 30ff.). Diese Politik korrespondiert mit der stärkeren Betonung der nationalen Verantwortung, des *ownership*, im Programmbereich.

⁷⁶ Der *International Monetary Fund* (IMF) hat zum 10.4.2000 eine ähnliche, unabhängige Evaluationsabteilung gegründet.

Die FAO ist im UN-System die Institution, die für die Entwicklung des Fischerei- und Aquakultursektors eine hohe Relevanz hat. Die Evaluationspraxis der FAO ist daher auch zur Bewertung der im Weiteren Verlauf der Arbeit benutzten Analysen relevant. Aufgrund mangelnder Kapazitäten (finanziell, personell) sind die Evaluierungen der FAO quantitativ und auch qualitativ schlechter als bei der Weltbank aber auch einigen bilateralen Gebern. Es gibt eine kleine (8 Mitarbeiter, Stand 2001) Evaluationsabteilung, die bei Evaluierungen mit den Programm- und Projektzuständigen zusammenarbeitet. Im *Programme Evaluation Report* (FAO 2000b) werden die Probleme der Unabhängigkeit und begrenzten Expertise, die daraus entstehen können, direkt angesprochen. Da die FAO ebenfalls unter zunehmenden Legitimationsdruck steht, gibt es eine Diskussion nicht nur über die Qualität der Programme, sondern auch über die der Evaluierungen. Als erste Konsequenz werden seit Ende der 90er Jahre verstärkt externe Sachverständige hinzugezogen.

Methodisch basieren FAO-Evaluierungen häufig auf einem an den *logical framework* Ansatz angelehnten Verfahren. Das starke Engagement der FAO bei ländlicher Entwicklung führte zu einer frühen und häufigen Anwendung partizipativer Elemente wie dem *participatory rural assessment*.⁷⁷ Diese Ansätze wurden aber weder systematisch evaluiert noch hatten sie stärkere Auswirkungen auf die Evaluationspraxis.

Fazit

Die Evaluationskonzepte auf multilateraler Ebene werden von der jeweiligen Aufgabe und Stellung der Institution geprägt. Insgesamt ist der Ansatz der UN-Institutionen breiter und stärker auf Partizipation angelegt. Der quantitative und ökonomische Fokus der Weltbank-Evaluierungen dominiert jedoch weiterhin die multilateralen Evaluierungen.

2.1.2 Bilaterale Akteure

Bilaterale Akteure finanzieren jährlich ca. 74 % (für 1999 nach DAC 2001a: 72) der weltweiten öffentlichen Entwicklungshilfe (ODA). Diese Mittel der nationalen Haushalte stehen ebenso wie die Beiträge für multilaterale Organisationen unter Rechtfertigungsdruck. Der seit den 70er Jahren angestrebte nationale Anteil der ODA am Bruttosozialprodukt von 0,7 % wurde nicht nur nicht erreicht, sondern ging auch zurück und lag 1999 durchschnittlich bei 0,24% für alle DAC-Länder (DAC 2001a: 199). Evaluierungen sind für bilaterale Akteure ein Instrument, die Wirkungen ihrer Projekte, Programme und Politiken nachzuweisen.

Zwar haben sich die im DAC vertretenen Länder 1991 auf gemeinsame Evaluierungsrichtlinien verständigt (siehe oben), die Praxis der Länder ist jedoch teilweise sehr unterschiedlich und weit von der Erfüllung dieser Prinzipien entfernt (vgl. DAC 1998b, Hoebink 1998).

⁷⁷ 1992 legte die FAO einen *Plan of Action on People's Participation* vor und erweiterte die Abteilung *Women and Participation*.

Insgesamt haben die europäischen Länder mit dem Aufbau ihrer Evaluierungssysteme in den 70er und Anfang der 80er Jahre deutlich nach dem Vorbild in den USA und Kanada begonnen (vgl. Hoebink 1998). Insbesondere die Erfahrungen der Weltbank und von USAID dienten den europäischen Ländern im entwicklungspolitischen Bereich dabei als Vorbild. Dies ist begründet in der reinen Zahl der durchgeführten Evaluierungen, aber auch in der langjährigen Tradition und Erfahrung mit Evaluierungen in dem politischen System der USA.

In den USA spielt in den entwicklungspolitischen Evaluierungen von USAID die Kosten-Nutzen-Analyse eine herausragende Rolle. Sie wird für alle Projektformen angewendet. USAID betont die Vorzüge der KNA (vgl. USAID 2000):

- wenn Ergebnisse monetär erfasst werden können, können so unterschiedliche Maßnahmen verglichen werden;
- wenn Ergebnisse zwar quantifiziert, aber nicht direkt monetär ausgedrückt werden können, verhilft die KNA (mit Schattenpreisen) zu vergleichbaren Kennziffern;
- wenn Ergebnisse nicht quantifiziert werden können, ermöglicht es die KNA, unterschiedliche Modelle anhand der verursachten Kosten zu bewerten.

Ende der 60er Jahre musste sich auch USAID der Kritik an zu quantitativ und ökonomisch ausgerichteten Methoden stellen. USAID begegnete der Kritik, dass ökonomische Ziffern für Kosten und Nutzen nicht zwangsläufig die Ziele eines Projektes oder Programms widerspiegeln, 1969/70 mit der Entwicklung des *logical framework* (oder *Logframe*)⁷⁸, das in der Planungsphase die spezifischen Entwicklungsziele einer Maßnahme erfassen sollte (vgl. Binnendijk 1990: 167f.).

Dieser ökonomische Ansatz hat sich in Europa nicht in gleichem Maße durchgesetzt. Für die europäischen Länder kann man einen deutlich höheren Anteil qualitativer Evaluierungen verzeichnen.

Die kleineren europäischen Staaten, insbesondere die Niederlande und die skandinavischen Länder, zeichnen sich nicht nur durch einen höheren Anteil der ODA am Bruttosozialprodukt aus (z.B. liegt der Anteil bei den Niederlande, Schweden und Dänemark noch über 0,7%), sondern setzen auch bei Evaluierungen deutliche Akzente in Bezug auf die für Evaluierungen bereitgestellten Mittel⁷⁹, die Methodenentwicklung und –anwendung und die Geberkoordination bei Evaluierungen.

Für Deutschland gilt anders als in den USA:

„Das Politikfeld Entwicklungszusammenarbeit gehört zu den wenigen, in denen sich auf der Ebene der Bundesressorts bisher überhaupt eine systematische Erfolgskontrolle entwickelt hat“ (Borrmann et al. 1999: 324).

In Deutschland ist die Evaluierungspraxis der verschiedenen Akteure sehr unterschiedlich und von der Tradition und Funktion des jeweiligen Akteurs im Politikfeld abhängig (siehe die ausführliche Studie von Borrmann et al. 1999 sowie Musto 1991, Dolzer et al.

⁷⁸ Dieses Konzept wurde von der GTZ zu ZOPP weiterentwickelt und 1983 verbindlich eingeführt.

⁷⁹ Innerhalb der EU-Staaten fällt z.B. auf, dass Dänemark mit 0,1 % der ODA einen verhältnismäßig großen Anteil für Evaluierungen vorsieht und qualitativ hochwertige Studien vorlegt (Hoebink 1998).

1998: 50ff.). So führt die für die Durchführung zuständige Institution der *Deutschen Gesellschaft für Technische Zusammenarbeit* (GTZ) nur exemplarisch ex-post Evaluierungen durch und ist stärker am Monitoring interessiert, wohingegen die *Kreditanstalt für Wiederaufbau* (KfW), die finanzielle Mittel zur Verfügung stellt, ihren Schwerpunkt bei ex-post-Analysen hat und dabei vorrangig auf ökonomische Kriterien zurückgreift (vgl. Stockmann 1996: 49ff., Raschen 1998, Neubert 1999 31ff., Kreditanstalt für Wiederaufbau 1999). Das BMZ⁸⁰ führt jährlich ca. 50-60 Evaluierungen nach einem Evaluierungsraster durch, d.h. 1,5 – 2 % aller laufenden Projekte werden überprüft. Die methodische Grundlage der Evaluierungen ist angelehnt an ZOPP, aber in den Evaluierungen existieren unterschiedliche methodischen Ansätze (vgl. Stockmann 1996: 49ff., Breier 1998, Neubert 1999: 30). Bis auf wenige Querschnittsevaluierungen⁸¹ gibt es eher fallbezogene Evaluierungen, deren Vergleichbarkeit stark eingeschränkt ist. Bewertungsmaßstab ist weitgehend die Erfüllung der jeweiligen Projektvorgaben und nicht sektorale oder nationale entwicklungspolitische Zusammenhänge.

Fazit

Die Praxis der bilateralen Geber ist von den Erkenntnissen der Evaluationsforschung, aber auch den selbst aufgestellten Richtlinien sehr weit entfernt.⁸²

Während in den USA ökonomisch Verfahren dominieren, bietet sich in den anderen Ländern ein buntes, kaum methodengeleitetes Bild:

„the methods and techniques used have varied substantially depending on the type of evaluation and, for the most part, have been qualitative or a mix of qualitative and quantitative. Desk studies with site verification have been a standard practice among members“ (DAC 1998b: 58).

Über ein Drittel der dem DAC gemeldeten 6.500 Evaluierungen der zentralen Evaluierungseinheiten waren Projektabschlussberichte. Immerhin 1405 entfielen auf *impact evaluations* und 168 auf *sector impact evaluations* (vgl. DAC 1998b: 21). In der Selbsteinschätzung der an Evaluierungen Beteiligten, werden die verwendeten Methoden und Techniken sehr positiv bewertet (vgl. DAC 1998b: 34). Gerade vermeintliche *impact evaluations* beziehen sich dennoch fast ausschließlich auf *output* bzw. *outcome* einer Maßnahme (vgl. z.B. Kohnert 1998). Hoebink, bezeichnet die meisten Evaluierungen eher als Projektprüfungen, weil sie von Methode und Inhalt nicht Evaluierungen entsprechen würden (vgl. Hoebink 1998).

Zu begrüßen ist die Tendenz, den Betrachtungsgegenstand von der reinen Projektenebene auf sektorale oder thematische Evaluierungen auszudehnen (vgl. Binnendijk 1990: 177f., DAC 1998b: 14).

⁸⁰ Das BMZ richtete bereits 1970 ein Referat „Inspektion und Revision der Entwicklungshilfe“ ein (vgl. Lachenmann 1977: 70).

⁸¹ Jährlich werden nur zwei bis sechs Querschnittsevaluierungen durchgeführt.

⁸² Der DAC führte hierzu 1996 und 1998 eine Untersuchung zum Thema durch, inwieweit die 1991 verabschiedeten Richtlinien umgesetzt worden sind. Die Untersuchung fand in zwei Phasen statt, einem Fragebogen an die Evaluierungseinheiten und persönlichen Interviews mit Evaluierern und Nutzern von Evaluierungen (vgl. DAC 1998b).

Bis auf wenige Ansätze der *like-minded countries* spielen partizipative Verfahren bis heute kaum eine Rolle bei bilateralen Evaluierungen (vgl. DAC 1998b: 11, 22). Nach den Angaben der Verantwortlichen in den Geberländern liegt der Hauptgrund hierfür an der "recipient unwillingness to participate" (DAC 1998b: 27). Mangelndes *ownership* bzw. das bestehende Machtgefälle bei Planung und Durchführung von Maßnahmen können hierfür eine Rolle spielen. Partizipative Verfahren erfordern ebenfalls ein anderes Selbstverständnis und auch Herangehen nicht nur bei Evaluierungen, sondern auch bei der Projektplanung und –durchführung.

Evaluierungen bilateraler Zusammenarbeit finden nach wie vor fast ausschließlich auf der jeweiligen bilateralen Ebene statt. Das Bestreben der DAC-Richtlinien von 1991, eine stärkere Geberkoordination bei Evaluierungen zu initiieren, war nicht erfolgreich. Aus entwicklungspolitischer Sicht wären aber *multi-donor*-Wirkungsanalysen als Schritt zu einer stärkeren Koordination weiterhin zu begrüßen.

Bilaterale Geber stehen im Hinblick auf Evaluierungen vergleichbaren Anforderungen und Problemen gegenüber. Die unterschiedlichen entwicklungspolitischen Haltungen der Geber (siehe z.B. *like-minded countries*) manifestieren sich ebenfalls in inhaltlichen und methodischen Unterschieden in den Evaluationssystemen.⁸³ Vereinzelt werden Akzente gesetzt und neue Verfahren propagiert. Mehrheitlich wird die Praxis der bilateralen Geber dennoch von einfachen, qualitativen Untersuchungen geprägt.

2.1.3 Nichtregierungsorganisationen

Die wenigsten entwicklungspolitischen NRO verfügen über etablierte Evaluationssysteme und –kapazitäten. Rein quantitativ, von der Zahl der durchgeführten Evaluierungen, sind die NRO für die Betrachtung der Evaluationspraxis von geringer Bedeutung. Ihr Stellenwert ergibt sich aufgrund der methodischen und inhaltlichen Anstöße, die von Seiten der NRO seit den 80er Jahren an die entwicklungspolitische Evaluationsforschung herangetragen wurden.

NRO führen im Vergleich zu bi- und multilateralen Gebern wenige Evaluierungen durch. Dies hat mehrere Gründe:

- Spendengelder sollen direkt für die Projekte ausgegeben werden; das Verständnis dafür, dass Evaluierungen zur Qualitätsverbesserung beitragen und damit dazu, dass die Spende „ankommt“, entwickelt sich bei NRO und Spendern erst langsam;
- NRO fühlen sich als zumeist Mitgliedsorganisationen stärker interner Rechtfertigung verpflichtet, nicht aber der öffentlichen Diskussion (vgl. Mutter 2000: 351);
- Sorge, durch Evaluierungen, die kritisch berichten oder als kritisch bewertet werden, öffentliche, finanzielle Unterstützung zu verlieren (vgl. Dolzer et al. 1998: 66ff., Roche 2000: 544);

⁸³ Die Untersuchung des DAC zeigt zudem zum einen eine Stärkung der Evaluationskapazitäten in einigen Geberländern auf und beschreibt zum anderen eine Gegenentwicklung. Denn die Evaluationssysteme der bilateralen Geber leiden unter einem zunehmenden institutionellen Druck. In verschiedenen Ländern (z.B. in England) wurde der Bereich Entwicklungspolitik dem Außenministerium zugeordnet und verlor seine Eigenständigkeit. Dies hat die Evaluierungskapazitäten eindeutig geschwächt (vgl. DAC 1998b).

- die Selbsteinschätzung und Wahrnehmung der Öffentlichkeit, dass NRO direkt mit den Betroffenen arbeiten, scheint Evaluierungen, die nur die Wirkung auf der Basisebene erfassen sollen, überflüssig zu machen;
- Evaluierungen erfordern Geld, Zeit und qualifiziertes Personal, all das haben NRO noch weniger als staatliche Institutionen;
- Evaluierungen werden von NRO mitunter als Kontrollinstrument verstanden, dass der angestrebten Partnerschaft mit den Projekten in den Entwicklungsländern, widerspricht.

In den 90er Jahren haben sich NRO verstärkt für Evaluierungen geöffnet.⁸⁴ Gerade NRO die (auch) staatliche Gelder erhalten, müssen die Wirksamkeit ihrer Aktivitäten nachweisen. Zunehmend werden in der öffentlichen Diskussion die positiven Vorurteile zugunsten des Basisbezugs der NRO in Frage gestellt. Eine stärkere Evaluierung der NRO Projekte trägt somit zu einer Versachlichung der Debatte bei.

Der entwicklungspolitische Ansatz der NRO – Armutsorientierung, Basisbezug, soziale, ökologische und partizipative Ausrichtung – hat auch die Evaluierungsdiskussion beeinflusst:

- eine Vielzahl der NRO-Projekte zielt explizit auf nicht-ökonomische Ebenen; die Kosten-Nutzen-Analyse wurde daher als ungeeignet abgelehnt und neue Verfahren wurden gesucht;
- NRO betonen den Prozesscharakter von Entwicklung, der auch in Evaluierungen erfasst werden müsse;
- der partnerschaftliche, partizipative Ansatz in der Projektarbeit müsse seine Entsprechung in Evaluierungen finden.

Das zentrale Ziel und damit Meßlatte für die Beurteilung der Wirksamkeit der Maßnahmen ist die Armutsbekämpfung. Gerade britische NRO haben die Diskussion um Theorie und Praxis partizipativer Verfahren vorangetrieben (vgl. Marsden/Oakley 1990 und Marsden et al. 1994, aber auch für deutsche NRO Dolzer et al. 1998, VENRO 2000). NRO greifen stärker auf einheimische Gutachter zurück. Die Erfahrungen sind bislang jedoch zu vereinzelt und methodisch unsystematisch, um dabei von einer etablierten Evaluationspraxis sprechen zu können

Eine Analyse von NRO-Evaluierungen kam daher zu einem zurückhaltenden Urteil. Im Auftrag des DAC (Riddell et al. 1997) waren Evaluierungen analysiert worden, um die Wirkungen von NRO-Projekten zu erfassen. Die meisten Evaluierungen, die der Studie zugrunde lagen, beruhten auf qualitativen Analysen, weil es nur unzureichende Ausgangsdaten gab. Riddell et al. (1997: 65) haben als einheitliche Methode den Projektbesuch, ein paar Gespräche, ein wenig Datenanalyse und die Berichterstellung festgehalten. Wirkungen wurden von NRO im Vergleich zu den gesteckten Zielen (z.B. Armutsbekämpfung, Partizipation, Nachhaltigkeit, Innovation, Flexibilität, *gender*, Umwelt, Demokratieförderung) bewertet. Insgesamt wurden bei den Evaluierungen

⁸⁴ Zur Unterstützung einer stärkeren Professionalisierung der NRO auch durch Evaluierungen hat der Verband Entwicklungspolitik deutscher Nichtregierungsorganisationen (VENRO) eine Handreichung herausgegeben (vgl. VENRO 2000).

festgestellt, dass „none has provided detailed information on precisely how these have been assessed and judged“ (Riddell et al. 1997: x).

Fazit

Aufgrund mangelnder Ressourcen und teilweise fehlender Selbstkritik waren NRO bislang lediglich in der Lage, wichtige inhaltliche Anstöße für die Methodendiskussion zu geben, aber es fehlt an der systematischen Erprobung und Umsetzung dieser Überlegungen in der eigenen Arbeit.

2.2 Evaluationsebenen

Wirkungen können auf unterschiedlichen Ebenen erfasst werden. Nicht nur die institutionelle Verankerung, die Zahl oder die verwendeten Methoden prägen das Bild der entwicklungspolitischen Evaluationspraxis, sondern auch der Fokus der Untersuchungen. Werden gesamtwirtschaftliche Indikatoren betrachtet oder Einzelprojekte, um Wirkungen nachzuweisen? Sollen Wirkungen sektoral oder thematisch überprüft werden?

In der entwicklungspolitischen Evaluationspraxis ist seit den 60er Jahren auch ein Wandel des Untersuchungsgegenstandes zu verzeichnen. Zu Beginn der Entwicklungszusammenarbeit entsprach dem wachstumsorientierten Entwicklungsmodell eine Überprüfung dieser Strategie im Wesentlichen durch ökonomische Indikatoren auf nationaler, volkswirtschaftlicher Ebene.

Die Diskussion um die Grundbedürfnisorientierung der Entwicklungszusammenarbeit verstärkte den Fokus zur Einzelprojektbetrachtung. Seit den 70er Jahren ist daher die Mehrzahl der Evaluierungen auf das individuelle Projekt vor Ort bezogen. Das löste jedoch das Problem der Wirkungszuordnung nicht:

„Die pragmatische Fixierung auf Einzelmaßnahmen bei der Untersuchung und Messung von Ergebnissen oder Zielerreichung birgt zweifellos das Problem in sich, den Gesamtzusammenhang aus dem Auge zu verlieren oder das vorgegebene Ziel zu eng zu definieren“ (Mutter 1998: 134).

In den 90er Jahren wurde daher verstärkt die Relevanz projektübergreifender, thematischer, nationaler oder sektoraler Evaluierungen erkannt (vgl. z.B. Thin et al. 1997, Mutter 1998):

„Länderbezogene, komparative Sektor-, Programm- und Projektanalysen scheinen besser als übergreifende Makroanalysen geeignet zu sein, um die Wirkungen und die Nachhaltigkeit der Entwicklungszusammenarbeit beurteilen zu können“ (Stockmann 1996: 10).⁸⁵

Dies führte zu einer Ausweitung dieses Evaluierungsansatzes, der jedoch bislang nur einen geringen Anteil an den gesamten Evaluierungen hat (vgl. Stockmann 1996, DAC

⁸⁵ Die Planung des BMZ sieht einen stärkeren Akzent bei thematischen, Sektor- und Instrument-Evaluierungen vor (vgl. Breier 1998: 129).

1998b, Donecker/Green 1998, Borrmann et al. 1999). Evaluierungen von Sektorpolitiken sind bislang noch in der Minderheit.⁸⁶ Auch bei der Evaluierung von Einzelmaßnahmen werden sektorale Fragestellungen meist nicht einbezogen. Die Weltbank ist federführend bei Evaluierungen auf Sektorebene, in kleinerem Umfang spiegelt sich dies bei den Institutionen der Finanziellen Zusammenarbeit (FZ) der bilateralen Geber.⁸⁷

Bislang stehen einer weiteren zahlenmäßigen Ausweitung projektübergreifender Evaluierungen vor allem der höhere Aufwand (Zeit, Kosten, Personal) und die parallel fortbestehende Notwendigkeit, die Ergebnisse des Einzelprojektes nachzuweisen, gegenüber.⁸⁸ Gerade auch Evaluierungen von Länderprogrammen sind sehr aufwendig und werden daher nur begrenzt durchgeführt.⁸⁹ Damit bleibt ein Manko bestehen, da „die Evaluierung von Projekten eine Hilfskonstruktion darstellt, mit der man sich der Gesamtausgabe nähern will, indem man das Puzzle der Einzelprojekte zu Programmen und Sektorkonzepten und dann zu einer konsistenten Politik zusammensetzen will“ (Mutter 1998: 137).

Bei projektübergreifenden Evaluierungen stellt sich ebenfalls die Frage einer stärkeren Geberkoordination bei Evaluierungen, d.h. vor allem geberübergreifende Evaluierungen. Hinderungsgründe für ein gemeinsames Vorgehen der Geber sind wiederum der höhere Aufwand bei Einbeziehung mehrerer Geber, Abgrenzungstendenzen der einzelnen Institutionen und der fehlende unmittelbare Nutzen für die Managementebene. Geberübergreifende Analysen für Länder, Sektoren oder Themen würden allerdings entwicklungspolitisch weitreichende Erkenntnisse liefern und nicht nur die wissenschaftliche Diskussion über Entwicklungsstrategien voranbringen, sondern sich auch für die operative Planung und Nutzung einsetzen lassen.

2.3 Evaluationsmethoden in der entwicklungspolitischen Praxis

Kontroversen

Im entwicklungspolitischen Kontext fanden eine ähnliche Entwicklung und vergleichbare Kontroversen wie in der gesamten Evaluationsforschung statt. Zwei Kontroversen prägen die Debatte der Evaluationsforschung in der Entwicklungszusammenarbeit:

- inhaltlich: unterschiedliche Entwicklungsmodelle bzw. -theorien führen zu divergierenden Evaluationsansätzen, da Ziele und Mittel unterschiedlich beurteilt werden;

⁸⁶ Vor allem der Energiesektor ist bevorzugter Gegenstand von Sektorevaluierungen, daneben gibt es beispielsweise Untersuchungen zu Bildung oder Forstwirtschaft. Teilweise sind die Analysen eines Sektors eher als Serienevaluierungen anzusehen und beinhalten keinen konzeptionellen Sektorzugang. So nahm 1995 UNDP eine Evaluierung von zwanzig Projekten im Energiesektor in Form einer *desk study* vor. Die Analyse stieß jedoch auf grundsätzliche Schwierigkeiten, da es anhand der Unterlagen kaum Indikatoren zur Messung der *performance* und der langfristigen Wirkungen gab (vgl. UNDP 1996: 47ff.).

⁸⁷ Für die Akteure der deutschen staatlichen Entwicklungszusammenarbeit hat so lediglich die KfW sektorspezifische Prüflinien (vgl. Kreditanstalt für Wiederaufbau (KfW) 1999).

⁸⁸ Stockmann nennt darüber hinaus als Gründe für die Präferenz von Einzelfallstudien die fortbestehende Dominanz der *input*-orientierten Erfolgsmessung und der Planungs- und Implementationsperspektive bei den Durchführungsorganisationen sowie ein mangelhaftes Methodeninstrumentarium für die Wirkungsmessung (vgl. Stockmann 1996: 29ff.).

⁸⁹ Vgl. dazu ausführlicher auch mit Hinweisen zu methodischen Problemen DAC (1998b).

- methodisch: die sozio-kulturellen Besonderheit der Arbeit in Entwicklungsländern prägt die Debatte zwischen dem Anspruch auf methodischen Perfektionismus und dem Streben nach Vereinfachung, die für eine größere Partizipation erforderlich ist.

Die Diskrepanz zwischen der theoretischen Debatte und der entwicklungspolitischen Praxis ist sehr hoch. Die Anforderungen der Auftraggeber in Bezug auf möglichen Aufwand (Zeit, Kosten) stehen methodisch fundierten Untersuchungen entgegen.

Von einigen, seltenen Forschungsprojekten abgesehen, die einen methodischen Anspruch verfolgen, dominiert weiterhin die ökonomische Analyse mit ein paar quantitativen Indikatoren⁹⁰ oder das *Gutachtermodell*. Hauptmethoden sind Sekundärquellenstudium, halbstrukturierte Interviews und Beobachtungen (vgl. Bolay 1988: 94, 316, Berlage/Stokke 1992: 25, Stockmann 1996, Neubert 1999). In der Praxis gibt es weder experimentelle Ansätze noch valide Vorher-Nachher-Vergleiche. Neubert attestiert das Fehlen eines Evaluationskonzepts, „die erhobenen Daten werden dementsprechend eher intuitiv als systematisch verarbeitet“ (Neubert 1999: 29).

Mutter sieht die Probleme hingegen stärker in der schematischen Nutzung von Methoden wie ZOPP. Der Methodenboom habe zwar Instrumente zur Verfügung gestellt, die aber nicht flexibel eingesetzt würden (Mutter 1998: 139).⁹¹

Das Methodenproblem der sozialwissenschaftlichen Evaluationsforschung, valide Wirkungen zu erfassen, kann auch im entwicklungspolitischen Kontext nicht gelöst werden; die Divergenzen über Realitäts- und Wissenschaftsverständnisse werden nahtlos fortgesetzt. Die Erfahrungen mit unzureichenden Ergebnissen entwicklungspolitischer quasi-experimenteller Evaluationsverfahren in den 70er Jahren, haben die methodische Diskussion geprägt (vgl. Binnendijk 1990: 173ff.) und in der Praxis zu einer stillschweigenden Nutzung qualitativer Verfahren geführt.

Wandel entwicklungspolitischer Evaluierungen

Trotz der methodischen Zurückhaltung in der Praxis hat die theoretische Diskussion und Differenzierung sich in der Anwendung niedergeschlagen. Die stärkere Berücksichtigung sozialer, kultureller und ökologischer Faktoren in den Evaluierungen ist trotz aller Mängel und berechtigten Kritik deutlich.

Dazu gehören sowohl die Überarbeitung der Kosten-Nutzen-Analyse für den entwicklungspolitischen Kontext (vgl. Little/Mirrlees 1974, Dasgupta et al. 1972, Squire/van der Tak 1975), als auch die Anwendung von EIA und SIA. Innerhalb der Diskussion um tatsächliche Verbesserung der Lebensverhältnisse der Menschen durch Entwicklungszusammenarbeit in den 70er Jahren rückten auch die Wirkungen der Entwicklungszusammenarbeit auf Frauen, ihre Rolle im Entwicklungsprozess und – später – das Geschlechterverhältnis/*gender* in den Blickpunkt (vgl. Boserup 1970, Bruchhaus et al. 1979, Wichterich 1984, von Braunmühl 1991 und 1998, Baud et al. 1992). Die Integra-

⁹⁰ Für die Evaluierungsaktivitäten der KfW weist Stockmann z.B. auf eine Überbewertung der ökonomischen und technischen Faktoren und eine entsprechende Vernachlässigung sozialer, ökologischer, politischer und kultureller Fragen hin (Stockmann 1996b: 208). Bei den Evaluationen der Weltbank kommt diese Kritik z.B. von Winpenny (1995), Picciotto (1996) und Brent (1998).

⁹¹ Vgl. dazu auch die ausführlichere Kritik von Kohnert (1998).

tion von *gender*-Aspekten in Evaluierungen hat bislang aber nur unzureichend stattgefunden:

„Even today, a large number of evaluations are not ‘peopled’, and those that are do not necessarily distinguish people by gender“ (Moser 1995: 105).

Auch das Ziel, Partizipation der Zielgruppen zu gewährleisten, fand ausgehend von Logframe und ZOPP bei der Projektplanung, teilweise Eingang in die Evaluierungen (vgl. Cameron 1993: 93f., Neubert 1999: 29). Die Argumentation für die stärkere Nutzung partizipativer Verfahren wird sowohl instrumentell als auch inhaltlich geführt, zur verbreiteten Anwendung hat dies aber noch nicht geführt. Die wenigen, fundiert partizipativ gestalteten Evaluierungen ermöglichen noch keinen Nachweis, über die Validität der Methode. Bislang fehlen systematische Auswertungen partizipativer Evaluierungen, um zu klären, ob die methodischen Ansprüche erfüllt werden.

Ein weiterer Gesichtspunkt, der in den Evaluierungen meist nicht berücksichtigt wird, obwohl er einhellig als wichtig anerkannt ist (vgl. z.B. Stockmann 1996, BMZ 1997, Mutter 1998, Neubert 1999), ist die Nachhaltigkeit, verstanden als die Aufrechterhaltung des Projekterfolges über das Projektende hinaus. In der deutschen Diskussion hat sich Stockmann intensiv mit dieser Ebene befasst (vgl. Stockmann 1992, 1996). Sein Modell der Erfassung von Nachhaltigkeit anhand eines mehr als 200 Fragen langen Fragebogens, scheint aber wenig praxistauglich zu sein.

Die beschriebenen Defizite der methodischen Ansätze (siehe Kapitel II.1.3.6) wurden auch im Rahmen der Entwicklungszusammenarbeit nicht gelöst und die Anpassung an den Untersuchungsgegenstand ist bislang nur unzureichend gelungen. Aufgrund der praktischen Probleme und inhaltlichen Kontroversen hatten die DAC-Richtlinien Methodenempfehlungen für Evaluierungen fast vollständig ausgespart.

Propagierte Methodenvielfalt

In Abgrenzung zu eher paradigmatischen Methodenauseinandersetzungen argumentiert die Mehrzahl der entwicklungspolitischen Evaluationsforscher und –praktiker heute für einen Methodenmix, für ein flexibles Nebeneinander unterschiedlichster Verfahren (vgl. z.B. Chelimsky 1995, Picciotto 1996, Borrmann et al. 1999, Baker 1999).⁹²

Auch Borrmann et al. (1999: 37f.) kommen zu dem Schluss, dass kein Mangel an Methoden herrscht, sondern viele Methoden einfach nicht angewendet werden. Oftmals liegen die Voraussetzungen nicht vor, um bestimmte Methoden anwenden zu können, z.B. vorherige Operationalisierung der Ziele und Festlegung von Erfolgsindikatoren, vertretbarer personeller und finanzieller Aufwand der Evaluierung, Akzeptanz der Methode (vgl. auch Kevenhörster/Stockmann 2001).

Die Argumentation für die Methodenvielfalt bei entwicklungspolitischen Evaluierungen kann den Widerspruch nicht lösen, warum in der Praxis, die bislang kaum in der Lage war, eine Methode sauber anzuwenden, plötzlich Zeit, Know-how und Kapazität für einen Mix vorhanden sein sollen.

⁹² Ein Beispiel: Gavai (1996) plädiert für einen Mix aus qualitativen und quantitativen Methoden und wendet dieses Verfahren bei der Evaluierung eines Kindergesundheitsprojektes in Indien an.

2.4 Schlussfolgerungen

2.4.1 Bewertung der Evaluationsforschung und -praxis

Der Überblick über die Evaluationsforschung allgemein und speziell im Hinblick auf die Entwicklungszusammenarbeit hat verschiedene Punkte deutlich werden lassen, die für den Zusammenhang der Arbeit relevant sind:

- (a) es liegen verschiedene methodische Ansätze vor, die ein breites methodisches Spektrum darstellen;
- (b) es bestehen weiterhin ungelöste methodische bzw. wissenschaftstheoretische Probleme;
- (c) konzeptionell wurden Evaluationsmethoden bislang nur mangelhaft an die Anwendung im entwicklungspolitischen Konzept angepasst und es besteht eine eindeutige Geberdominanz;
- (d) die Methodenentwicklung hat die entwicklungspolitische Evaluationspraxis mehrheitlich nicht erreicht (Theorie-Praxis-Trennung);
- (e) Evaluierungen sind nicht wertfrei und können es auch nicht sein.

zu

(a) Unterschiedliche methodische Ansätze

Die Evaluationsforschung ist eine junge Disziplin und hat in den vergangenen Jahrzehnten interessante Ansätze entwickelt, diskutiert und variiert. Es ist ein Forschungsgebiet, wo die Debatte kontrovers geführt wird. Die Methodendiskussion ist bislang zu keinem befriedigendem Ergebnis gekommen. Die Entwicklung der „einen“ Methode kann auch nicht das Ziel sein. Für die komplexe Aufgabe, Realität darzustellen, ist eine Methodenvielfalt angemessen.

Offene Forschungsfragen bestehen und beinhalten u.a. eine weitergehende Öffnung der Kosten-Nutzen-Analyse für Verteilungsaspekte sowie soziale und ökologische Faktoren. Partizipative Evaluationsverfahren, die nicht nur Prozesse bewerten, sondern Teil der Prozesse sind, müssen weiterentwickelt werden und mit konservativeren, ökonomisch orientierten Verfahren verbunden werden.

(b) Ungelöste methodische - wissenschaftstheoretische Probleme

Die Evaluationsforschung ist eng verknüpft mit der methodischen Diskussion in den empirischen Sozialwissenschaften generell. Methodische und wissenschaftstheoretische Kontroversen bestehen daher zwingend auch in der Evaluationsforschung fort und können nicht aufgelöst werden. In der Praxis wird ein undogmatischer, aber kaum theoriegeleiteter Mix verschiedener Methoden benutzt, der ungeachtet der wissenschaftstheoretischen Ausrichtung stark auf Subjektivität beruht. Auch die Evaluationsforschung sollte sich stärker um die systematische Verknüpfung von qualitativen und quantitativen Methoden bemühen. Der geforderte Methodenmix muss theoretisch begleitet werden.

(c) Mangelhafte entwicklungspolitische Anpassung - Geberdominanz

Die Evaluationsforschung ist in den USA entstanden und wird auch bislang von dort geprägt. Die Entwicklungszusammenarbeit war seit den 60er Jahren ein wesentliches

Anwendungsgebiet für Evaluierungen, gerade in Europa, und daraus ist auch eine Methodendiskussion entstanden. Diese befindet sich aber immer noch in den Anfängen. Fast vollkommen fehlen Überlegungen aus den Entwicklungsländern selbst. Die gängigen Evaluationsverfahren in der Entwicklungszusammenarbeit beruhen weiterhin auf Übertragungen der für die Industrieländer entwickelten Verfahren und werden in der Regel von Gutachtern aus den Geberländern angewendet.

Der geforderte Ausbau von Evaluierungskapazitäten und –aktivitäten in Entwicklungsländern sollte auch methodische Impulse bringen.

(d) Theorie-Praxis-Trennung

Die Unterschiede zwischen den diskutierten Evaluationsansätzen und den in der Praxis angewendeten Evaluationsverfahren sind sehr groß.

Großangelegte quantitative Untersuchungen gibt es vorrangig in den USA oder mittels staatlicher Forschungsmittel. Darüber hinaus überwiegen methodisch wenig verankerte qualitative Analysen, die höchstens Teilaspekte einzelner Forschungsansätze aufgreifen und weitgehend auf subjektive Urteile und Erfahrungen der Evaluierer vertrauen.

Diese Diskrepanz ist wesentlich in den äußeren Anforderungen an Evaluierungen begründet; fehlende Daten und mangelnde Vorbereitungen bereits während der Projektplanung und –durchführung sowie vorrangig Zeit- und Geldmangel bestimmen die verwendeten Methoden.

(f) Evaluierungen sind nicht wertfrei

Selbst bei der Annahme, dass Evaluierungen in völliger Unabhängigkeit vom Auftraggeber und Unparteilichkeit durchgeführt werden, sind sie nicht neutral. In welchem Kontext entwicklungspolitischer Ziele die Ergebnisse einer Maßnahme bewertet werden und in welchem Erkenntnisinteresse die Methodenwahl erfolgt, beruhen nicht zuletzt auf normativen Entscheidungen (vgl. Mark/Shotland 1987, Bamberger 1991: 326, Berlage/Stokke 1992).

2.4.2 Konsequenzen für die Wirkungsanalyse der Aquakultur in Indien

Für den Kontext dieser Arbeit ergeben sich daraus verschiedene Anknüpfungspunkte für eine Wirkungsanalyse. Sie beziehen sich auf

- a) die Zieldefinition,
- b) die Bewertungskriterien ,
- c) die Wirkungsdimensionen,
- d) den Evaluationsgegenstand und
- e) die Methode

der Analyse von Aquakulturen in Indien.

a) Zieldefinition

Im Grundsatz stellen Evaluierungen einen Soll-Ist-Vergleich dar. Doch schon die Bestimmung des Ausgangspunktes, der Zielebene bzw. des Bewertungsmaßstabes ist problematisch:

- in der Mehrzahl der Projekte und Programme liegt keine oder nur eine mangelhafte Zielformulierung zu Beginn einer Maßnahme vor;⁹³
- es existieren unterschiedliche theoretische und politische Vorstellungen über Entwicklungsmodelle und –wege;
- die Ziele der Zielgruppen sowie der Regierung des Projektlandes können sich untereinander und von den Zielen der Geberinstitutionen unterscheiden;
- im sozialen, gesellschaftlichen Kontext sind Ziele nicht statisch, sondern können ebenfalls einem Wandel unterliegen.

Wenn Ziele überprüft werden, welche sind das? Das Bündel der allgemeinen, konsensualen Entwicklungsziele? Die Ziele eines bestimmten Akteurs oder die konkreten, ggf. isolierten Ziele eines bestimmten Programms oder Projektes? Insofern bestehen für unterschiedliche Ansätze Spielräume, bei der Überprüfung der Zielerreichung.

Im Kontext dieser Arbeit sollen auf diesem Hintergrund alle Zieldimensionen in die Analyse einbezogen werden. Dies bedeutet die Verwendung eines breiten Entwicklungsbegriffs; Aquakulturprojekte und -maßnahmen sollen innerhalb ihres politischen, sozialen, ökologischen und makroökonomischen Kontextes bewertet werden. Dies beinhaltet auch eine kritische Prüfung der Ziele, die den Maßnahmen zugrunde liegen. Dabei müssen nicht nur die Werte der betroffenen Geber berücksichtigt werden, vielmehr werden Ziele wie Effektivität, Umweltschutz oder Gleichheit auch kulturell unterschiedlich beurteilt.

b) Bewertungskriterien

Die verschiedenen, möglichen Kriterien für die Bewertung von Maßnahmen (z.B. Effizienz, Effektivität, Signifikanz, Relevanz oder *impact* (vgl. u.a. DAC 1991, Breier 1998) werden nicht in gleichem Maße bei jeder Evaluierung zugrunde gelegt. Vielmehr werden Akzente gesetzt, die

- die betriebswirtschaftliche Rentabilität,
- die wirtschaftspolitische Bewertung oder
- die entwicklungspolitische Relevanz

in den Vordergrund stellen.

Die betriebswirtschaftliche Rentabilität ist für jede wirtschaftliche Aktivität und daher auch für den Aquakultursektor eine Grundvoraussetzung. Übergreifende Bedeutung für die vorliegende Analyse hat jedoch die Bewertung der entwicklungspolitischen Relevanz.

⁹³ Bolay betont, die methodisch-systematischen Defizite von Evaluierungen, wenn der Evaluierungsgegenstand nicht von vornherein auf die „Evaluierbarkeit“ angelegt war. Zwar wären trotzdem Evaluierungen zulässig, aber eben mit methodischen Abstrichen (vgl. Bolay 1988: 319, ebenso Binnendijk 1990: 169ff.).

c) Wirkungsdimensionen

Auch nach eher konservativen Definitionen, umfasst der Begriff Entwicklung mehrere Dimensionen. Die – weiterhin – zentrale ökonomische Dimension, wurde um soziale, ökologische und politische Dimensionen erweitert. In der Diskussion um unterschiedliche Entwicklungsmodelle differieren die Gewichte, die den unterschiedlichen Ebenen zugeschrieben werden; die Notwendigkeit einer Entwicklung auf allen Ebenen und die Vernetzung aller Dimensionen wird jedoch immer stärker Konsens.

Die Praxis der entwicklungspolitischen Evaluierungen insgesamt hat jedoch weiterhin ein Schwergewicht bei der Erfassung ökonomische Wirkungen. Vor allem eine zusammenhängende Betrachtung verschiedener Wirkungsdimensionen ist eher die Ausnahme. Soziale und ökologische Fragestellungen werden nicht selbstverständlich berücksichtigt, noch weniger trifft dies für die kulturelle oder gesellschaftspolitische Dimension zu.

Der Versuch, alle Ebenen einzubeziehen, ist aber notwendig, um die Wirkung einer Maßnahme bewerten zu können – auch wenn nicht jede einzelne Wirkung erfasst werden kann.

d) Evaluationsgegenstand

Die Überprüfung einzelner, laufender Maßnahmen bleibt das vorrangige Interesse der beteiligten entwicklungspolitischen Institutionen. Evaluierungen von Einzelprojekten bilden daher weiterhin die Mehrheit. Projektübergreifende Evaluierung haben in den 90er Jahren allerdings zugenommen, thematisch sowie sektoral.

Die vorliegende Analyse des Aquakultursektors in Indien bezieht sich auf die Aktivitäten unterschiedlicher nationaler sowie internationaler, privater sowie staatlicher Akteure. Die Bewertung von Einzelprojekten und Programmen wird genutzt für die Analyse des gesamten Sektors.

e) Methode

Evaluierungen in der Entwicklungszusammenarbeit sind aufgrund der sozio-kulturellen Unterschiede und des Geber-Nehmer-Verhältnisses mit zu dem schwierigsten Evaluationsbereich zu rechnen. Dem trägt die Methodenentwicklung und vor allem – anwendung kaum Rechnung. In der entwicklungspolitischen Praxis beruhen die Evaluierungsergebnisse zu weiten Teilen auf den subjektiven Urteilen der Gutachter nach Aktenlage und kurzen Gesprächen. Die Qualität ist damit unmittelbar abhängig von den Fähigkeiten des jeweiligen Gutachters.

Aus der Komplexität des Entwicklungsprozesse und der Methodendiskussion ergibt sich der Ruf nach einer Methodenvielfalt. In der Realität entwicklungspolitischer Evaluierungen wird die konsequente Anwendung verschiedener Methoden die Ausnahme bleiben – allein aus finanziellen und zeitlichen Begrenzungen. Zentral ist daher sowohl die Offenheit unterschiedlichen Methoden gegenüber als auch eine kritische Distanz zu ihren Ergebnissen.

Aus der Kenntnis unterschiedlicher Methoden der Wirkungserfassung und ihrer spezifischen Eigenschaften müssen vorliegende Analysen zu Aquakulturen kritisch

betrachtet werden (siehe das Beispiel in Kasten 4). Die vorliegende Analyse bezieht daher unterschiedliche quantitative und qualitative Sekundäranalysen mit ein.

Kasten 4: Verschiedene Methoden, verschiedene Bewertungen - ein Beispiel

Anhand der Kritik an ökonomischen Evaluationsverfahren, zeigt Brent (1998) am Beispiel einer Fallstudie von Tyler über die Einführung von Traktoren in Pakistan, wie Evaluierungen zu unterschiedlichen Ergebnissen kommen können. Zuerst macht er drei Modellrechnungen auf: eine betriebswirtschaftliche, eine konservativ-volkswirtschaftliche und eine Bewertung, die sich an sozialer Entwicklung orientiert. Aus der Sicht der wohlhabenderen pakistanischen Farmer, die sich Traktoren anschaffen, steht am Ende ein Nettonutzen, der den Kauf und damit das Projekt gerechtfertigt hat. Aus der Sicht des traditionellen Ökonomen, der in Schattenpreisen den wahren Wert des Projektes ausrechnet, steht am Ende ebenfalls ein Nettogewinn des Projektes. Unter sozialen Gesichtspunkten allerdings, also angenommen aus der Sicht der pakistanischen Regierung die am Wohlergehen der Gesamtbevölkerung interessiert sein muss und daher auch die Einkommensverteilung bzw. die Verlierer dieses Projektes in die Bewertung einbeziehen muss, steht am Ende ein negatives Resümee. Zu hoch ist die Zahl von Kleinbauern, die durch das Projekt arbeitslos werden und zu hoch sind die Einkommens- und Konsumeinbußen. (Brent 1998:12ff.)

Schließlich untersucht Brent das Traktor-Projekt zurück anhand der von Weltbankmitarbeitern entwickelten S & T Methode (nach Squire/van der Tak 1975). Obwohl dieser Ansatz soziale Kriterien mit einbezieht, steht am Ende ein positiver Nutzen, weil – so die Kritik von Brent – Verteilungsaspekte nur ungenügend berücksichtigt werden. Brents Fazit ist daher kritisch:

„It would seem that using S & T's methods, which do allow for distributional considerations, could nonetheless justify projects that many people would judge to be clearly socially undesirable. It would certainly conflict with the World Bank's stated objective of promoting the welfare of the poor“ (Brent 1998: 275).

Selbst die Verwendung aller vorgestellten Methoden würde nicht die Erfassung der gesamten Wirkungen des Aquakultursektors in Indien garantieren. Es müssen die richtigen Fragen gestellt werden bzw. Zusammenhänge analysiert werden. Die Suche nach Indikatoren zur Abbildung ökonomischer, sozialer, ökologischer, kultureller sowie gesellschaftspolitischer Dimensionen ist daher von zentraler Bedeutung.