

Zum Verhältnis von Steuerung und Evaluation – am Beispiel der Arbeitsmarktpolitik
About the relation of governance and evaluation – the case of labor market policy

Eröffnungsvortrag der 11. Jahrestagung der DeGEval – Gesellschaft für Evaluation: “Zum Verhältnis von Steuerung und Evaluation”, 24.-26. September 2008, Universität Klagenfurt

Lorenz Lassnigg
Institut für Höhere Studien (IHS)
Stumpergasse 56
A-1060 Wien
lassnigg@ihs.ac.at

Inhalt

1. Evaluierung von Arbeitsmarktpolitik: Ansätze, Ergebnisse, Probleme	2
1.1. Ansätze und Methoden der Evaluierung	4
1.2. Probleme der Umsetzung des 'State-of-the-art': Erweiterung des Bezugsrahmens	5
1.3. Stilisierte Ergebnisse der Maßnahmen- und Programmevaluierungen	7
1.4. Eine stilisierte Einordnung der AMP im Rahmen der Evaluierungsansätze und -praktiken	9
1.5. österreichische Erfahrungen	11
2. Steuerung von Arbeitsmarktpolitik: Neue Governance-Modelle und Politik-Paradigmen jenseits der Evaluierung	13
2.1. Hintergrund und Phänomenologie: Österreich im internationalen Kontext.....	13
2.2. Governance-Modelle: Bürokratie, Markt und 'New Public Management'	16
2.3. Politik-Paradigmen: Jobs Study, Flexicurity und Life-Course-Policy	17
2.3.1. Die OECD Jobs Study	17
2.3.2. Flexicurity	18
2.3.3. Life-Course-Policy	18
2.3.4. Governance und AMP: Diskurse und Realität	19
3. Evaluierung und Governance	20
3.1. Ein normatives Modell und seine Komplikationen: Evaluierung im Qualitätszyklus	21
3.2. Wie passt Evaluierung in Governance-Systeme? 'New public management (NPM)' und 'Evidence based policy and practice (EBPP)'	25
3.2.1 'New Public Management': Begrenzte Aufgabenzuweisung	27
3.2.2. 'Evidence-based Policy and Practice': Evaluierung als Umsetzungsforschung?	28
3.2.3. Thesen und Erfahrungen zur Beziehung von Evaluierung und Governance	29
4. Zusammenfassung	32
Literatur	33
ANHANG	37
Übersicht 1: Hauptmerkmale von Evaluierung und Steuerung im Bereich Arbeitsmarktpolitik	37
Übersicht 2: Evaluierung von AMP in Österreich.....	38

Im vorliegenden Beitrag wird das Verhältnis von Steuerung und Evaluierung am Beispiel der Arbeitsmarktpolitik (AMP) behandelt, wobei ein gewisser Fokus auf die österreichischen Erfahrungen gelegt wird, in denen der Autor von Beginn an involviert war. Es handelt sich um einen Versuch der Reflexion dieser Erfahrungen, der empirische Entwicklungen berücksichtigt und v.a. die theoretisch-konzeptionell-methodischen Entwicklungen der letzten Jahrzehnte in diesem Teilgebiet der Evaluierungspraxis als Hintergrund nutzt. Die Reflexion findet in einem vergleichenden Rahmen statt, in dem der 'State-of-the-art' der AMP-Evaluierung als Referenz für die nationalen Erfahrungen herangezogen wird. Die Evaluierungsthematik wird mit den neueren Diskursen über die politische Steuerung und über die Nutzung von Forschung und Entwicklung (F&E) in der Politik verknüpft. Es handelt sich um eine interpretative Überblicksarbeit, die in drei Schritten vorgeht:

- Erstens wird die Entwicklung der Evaluierung von Arbeitsmarktpolitik (Methodik und Ergebnisse) skizziert, wobei die institutionellen Besonderheiten in diesem Bereich herausgearbeitet werden;
- zweitens werden die Grundstrukturen und Entwicklungstendenzen von Governance in der Arbeitsmarktpolitik auf Basis der theoretisch-konzeptionellen Diskussionen und Forschungsergebnisse behandelt, um die institutionellen Bedingungen für das Verhältnis von Steuerung und Evaluierung in den Blick zu bekommen;
- drittens wird in einem eher spekulativen Abschnitt versucht den Zusammenhang von Evaluierung und Governance in einer realistischen Weise zu fassen, um die normativen Vorstellungen herauszuarbeiten und in Frage zu stellen, die oft der Behandlung des Verhältnisses von Evaluierung und Governance zugrunde gelegt werden.

In diesem Sinne ist der Beitrag auch evaluativ angelegt, indem eine Bilanzierung dessen angestrebt wird, was die bisherige Evaluierungspraxis im Bereich der Steuerung der Arbeitsmarktpolitik am Beispiel Österreichs erreicht hat.

1. Evaluierung von Arbeitsmarktpolitik: Ansätze, Ergebnisse, Probleme

Die Evaluierung von Arbeitsmarktpolitik ist im Vergleich zu vielen anderen Politikbereichen ein stark – wenn auch sehr einseitig – entwickeltes Feld. In Österreich wurde seitens der damaligen Arbeitsmarktverwaltung, die eine nachgeordnete Dienststelle der Bundesverwaltung mit regionalen Organisationsstrukturen war, bereits in den frühen 1980ern mit größeren nationalen Evaluierungsprojekten begonnen. Auslöser waren die Probleme mit steigender Arbeitslosigkeit, die auch Österreich erreicht hatten. Zu dieser Zeit war sonst von Evaluierung noch keine Rede – im Hochschulwesen begannen diese Diskussionen etwa zehn Jahre später eine größere Verbreitung zu erlangen, und im Schulwesen entsteht eine Renaissance noch einmal zehn Jahre später, nachdem die Schulversuchsevaluierungen der 1970er für einige Zeit vergessen und verdrängt worden waren.

Es gab bereits jahrzehntelange Erfahrungen seit den 1960ern in den USA, die von den Wirtschaftswissenschaften und der Ökonometrie geprägt waren und teilweise bereits sehr anspruchsvolle methodischen Instrumentarien¹ entwickelt und verwendet hatten (vgl. Smith 2000). In Deutschland waren in den 1980ern im Gefolge der neuen empirischen Forschungen und theoretischen Konzeptionalisierungen über Arbeitsmarktfragen (z.B. Sengenberger 1987) erste größere Evaluierungen z.B. durch das Wissenschaftszentrum Berlin (WZB) durchgeführt worden.

¹ Eine Schlüsselperiode waren die Evaluierungen des „Job Training Partnership Act (JTPA)“ in den 1980ern (vgl. z.B. Breslau 1998 für einen Überblick).

Spätestens in den 1980ern hat sich eine internationale Community der Evaluierung von AMP zu formen begonnen, die auch von internationalen Organisationen (v.a. OECD, aber auch ILO) stark beeinflusst war und es ist ein 'body of knowledge' entstanden, der Methoden und einen Bestand an stilisierten Ergebnissen umfasst. Die Ergebnisse wurden in umfassenden Reviews zusammengefasst und ansatzweise in international vergleichenden Meta-Studien gesichert. Vor allem seit der Entwicklung der Beschäftigungsstrategie in den 1990ern hat auch die Europäische Union die Evaluierungsaktivitäten verstärkt und gefördert, die sich insbesondere im Bereich der Regionalpolitik und des Europäischen Sozialfonds (ESF) entwickelt hatten. Evaluierungen in Form von Begleitstudien wurden in diesem Kontext zu einem administrativen Erfordernis, wobei aber trotz der umfangreichen Ausarbeitungen und Toolboxen² die Praktiken und Möglichkeiten de facto – aufgrund der zeitlichen und politischen Rahmenbedingungen – häufig auf einem eher niedrigen Niveau der methodischen Entwicklung angesiedelt waren bzw. sind.

Der hauptsächliche Evaluierungsgegenstand ist die sogenannte 'aktive Arbeitsmarktpolitik', die im Unterschied zu den gesetzlich abgesicherten Ansprüchen auf die sogenannten 'passiven' Maßnahmen der Einkommenssicherung in Form von Versicherungsleistungen bzw. teilweise auch Sozialleistungen ('Pflichtleistungen') ein breites Spektrum von diskretionär einsetzbaren Maßnahmen umfasst ('Ermessensleistungen'). Dieses Maßnahmespektrum wurde im Laufe der Jahre zunehmend strukturiert und kodifiziert. Die Taxonomie der Maßnahmen der AMP, die in den 1990ern von der OECD etabliert wurde, exemplifiziert dieses Spektrum an Maßnahmen (siehe Abb.1).

Abbildung 1: OECD-Taxonomie der Maßnahmen der Arbeitsmarktpolitik

"Aktive" Programme

1. Öffentliches Arbeitsmarktservice und Administration
2. Arbeitsmarktausbildung
 - a) Ausbildung für arbeitslose Erwachsene und Risikogruppen
 - b) Ausbildung für beschäftigte Erwachsene
3. Maßnahmen für Jugendliche
 - a) Maßnahmen für arbeitslose oder benachteiligte Jugendliche
 - b) Unterstützung für Lehrlinge oder ähnliche Formen allgemeiner Ausbildung
4. Förderung von Beschäftigung
 - a) Förderung von regulärer Beschäftigung im privaten Sektor
 - b) Unterstützung von Arbeitslosen für Unternehmensgründung und
 - c) Direkte Schaffung von Beschäftigung (öffentlich oder non-profit)
5. Maßnahmen für Behinderte
 - a) Berufliche Rehabilitation
 - b) Beschäftigung

"Passive" Programme (Einkommenssicherung)

6. Arbeitslosenunterstützung
7. Frühpensionierung aus arbeitsmarktpolitischen Gründen

Quelle: OECD, Employment Outlook 1993

Aus dem vergleichsweise starken Entwicklungsstand der Evaluierung in diesem Feld folgt jedoch bei weitem nicht, dass das generierte Wissen und die Erfahrungen politisch und praktisch stark

² Siehe die "MEANS Collection", die ein umfassendes Instrumentarium für die Evaluierung sozio-ökonomischer Programme entwickelt hat (http://ec.europa.eu/regional_policy/sources/docgener/evaluation/means_en.htm)

genutzt werden. Weder folgt die Evaluierungspraxis dem 'State-of-the-art', noch werden die Ergebnisse merkbar umgesetzt, teilweise sind sie auch aufgrund von vielen von vorneherein eingebauten Hemmnissen nicht besonders leicht umsetzbar.

Es gibt sehr unterschiedliche Hinweise auf die reale Evaluierungskultur. John P. Martin, ein langjähriger führender OECD-Beamter und einer der Hauptautoren der OECD Jobs Study bemerkt am Ende der 1990er, dass nur wenige europäische Staaten bis dahin rigorose Evaluierungen im Bereich der AMP durchgeführt haben, und benennt Norwegen, Schweden und das U.K. als diese Ausnahmen. „One explanation for the lack of evaluations in Europe was put to me frankly about ten years ago (also am Ende der ersten Dekade von Erfahrungen mit AMP-Evaluierung; d.Verf.) by a leading policy-maker who shall remain nameless. If I can paraphrase his explanation, it would run as follows: 'Most of our programs are lousy! They were dreamed up quickly to give the Minister some good news to announce at a time when unemployment is rising. We do not want evaluations revealing to the general public how bad our programs are; we know this already!'.“ (Martin 1998, 285) In der Zwischenzeit haben sich die Aktivitäten deutlich verstärkt. Als ambivalentes Beispiel können die deutschen Hartz-Reformen genannt werden, eine stärkere aktive Nutzung von Evaluierung im Zuge der AMP-Reformaktivitäten kann für Finnland konstatiert werden (vgl. den gemeinsam mit 14 Ländern durchgeführten Peer Review der Reform der aktiven AMP 2004: <http://www.mutual-learning-employment.net/peerreviews/2004/05/13-14>).

1.1. Ansätze und Methoden der Evaluierung

Ausgangspunkt der Evaluierungen war eine Ja/Nein-Wirkungsevaluation der Maßnahmen, wobei methodisch, aber nicht unbedingt praktisch, die Identifikation von "Netto"-Wirkungen zunehmend in das Zentrum der Aufmerksamkeit rückte. Das Hauptproblem dabei ist die Definition des 'Counterfactual', also des 'Normal'-Zustandes gegenüber dem eine Wirkung identifizierbar ist. Das randomisierte Experiment oder quasi-experimentelle Methoden auf Basis von Vergleichsgruppen und Propensity-Score Matching haben sich als Hauptmethoden dafür herausgebildet. In den letzten Jahren werden aufgrund komplexerer Datenbasen mit Längsschnittinformationen auch einige neue statistische Verfahren angewendet, die auf Gruppen-, Maßnahmen- und zeitliche Heterogenität abstellen (Instrumentvariablen, Difference-in-Difference-, sowie Timing-of-Event-Ansätze). Als wesentliche Kriterien für die Wirkung gelten Beschäftigungs- oder Einkommenswirkungen. Als zusätzliche Dimension, die jedoch nicht immer eine Rolle spielt, werden auch Kosten-Nutzen-Relationen, also Effizienzwirkungen im engeren Sinne, herangezogen.

Historisch geht die Entwicklung zurück auf die großen US-Programme gegen die Armut in den 1960ern. In den 1980ern wurden große und methodisch sehr anspruchsvolle Evaluierungen durchgeführt. In manchen Programmen wurden die Erhebungen so angelegt, dass Vergleiche von Experiment und Quasi-Experiment möglich wurden, die oft zu sehr unterschiedlichen Ergebnissen führten und bedeutende Anstöße für methodische Entwicklungen auf beiden Seiten gegeben haben. In den 1990ern wurden im Rahmen der OECD mehrere Reviews der Ländererfahrungen durchgeführt, die zu einer ersten Kodifizierung eines 'State-of-the-art' (OECD 1991) geführt haben, das Handbuch von Schmid/O'Reilly/Schömann (1996) arbeitet die verschiedenen Facetten der Praxis aus und kann als Weiterentwicklung in dieser Richtung gesehen werden.

Wesentliche Aspekte der Entwicklung waren:

- Es wurde jedenfalls die Verwendung eines 'Counterfactuals' empfohlen (auch wenn dieses nicht in allzu elaborierter Form möglich war)
- Es wurden drei Kategorien von Evaluierung unterschieden, die gemeinsam berücksichtigt werden sollten: Wirkungs-, Politik-, und Implementationsevaluation (darunter Monitoring, Performance Indicators, Implementation i.e.S.)
- Es erfolgte der Aufbau einer internationalen Datenbasis zur AMP, auf deren Grundlage dann auch Makrostudien mit international vergleichenden Daten durchgeführt werden konnten, die auch Gleichgewichtseffekte für die Wirtschaft erfassen können.
- Als stilisiertes verallgemeinertes Ergebnis dieser Studien wurden von Lars Calmfors (1994) 'crucial design features' für die Gestaltung arbeitsmarktpolitischer Maßnahmen formuliert, aus denen der spätere Imperativ der sogenannten 'Aktivierung' hervorgegangen ist (die Maßnahmen sollen so gestaltet sein, dass sie gleichzeitig Anreize für die Aufnahme einer Beschäftigung enthalten)
- In den letzten Jahren wurden auf Basis der skizzierten konzeptuellen Elemente über einige breit angelegte nationale Evaluierungen hinaus auch international vergleichende Meta-Evaluierungen durchgeführt, die ein bestimmtes Ergebnismuster der verschiedenen Maßnahmetypen ergeben haben.

Um eine entsprechende Nutzung der Evaluierung zu ermöglichen wurde bereits in OECD (1991) ein gleichgewichtiger Einsatz der drei Kategorien von Wirkungsevaluation, Politik(ziel)-Evaluation und Implementationsanalyse gefordert. „The three elements need to be *equally well developed* and co-ordinated. Deficiencies in any of the three would limit the overall contribution to policy development. (...) However, it appears that *implementation evaluation* is a weak link in most countries, as is the coverage of issues within each of the three components (particularly *economic and social effects* in impact studies).”(OECD 1991, 10)

1.2. Probleme der Umsetzung des 'State-of-the-art': Erweiterung des Bezugsrahmens

Als ein Grundproblem der Umsetzung der skizzierten Vorstellungen von Evaluierung stellen sich die vergleichsweise strengen Anforderungen dar, die sich aus einer methodisch anspruchsvollen Evaluierung an die Politik ergeben. Von der Evaluierung werden Bedingungen an die Politik gestellt, die der Logik der Politik nicht entsprechen. Man kann davon sprechen, dass idealen Evaluierungsstrategien ein '*naives Politik-Konzept*' zugrunde liegt:

- Es müssen evaluierbare Politik-Ziele formuliert werden,
- der Evaluierung muss eine genaue Zielklärung zugrunde gelegt werden,
- und die Evaluierung muss idealerweise in Form einer vorauslaufenden Konzeption und Datenerhebung in die Politik-Maßnahmen eingebaut sein.

Diese Bedingungen können zwar in Einzelfällen möglicherweise erfüllt werden, als allgemeine Regel erweist sich dies aber als undurchführbar (wie man z.B. an den Problemen der Formulierung von evaluierbaren Zielen in der Europäischen Politik sehen kann).

Die Problematik oder 'Naivität' liegt dabei nicht nur in den politischen Widersprüchen und der Unwilligkeit, derart strenge Auflagen an die Politikformulierung zu akzeptieren, sondern durchaus auch in den technischen Problemen der Operationalisierung und Messung. Die Außerstreitstellung von bestimmten 'Evidenzen' folgt zumeist nicht zwingend aus den Ergebnissen, und erfordert daher letztlich immer einen bestimmten Anteil an voluntaristischer Entscheidung und Akzeptanz. Auch wird ein Grad an Zielklärung zwischen Auftraggeber/innen

und Evaluierung vorausgesetzt, der in der Realität selten erreicht werden kann, da dabei die politische Dynamik nicht ausreichend berücksichtigt wird.

In der Maßnahmen- bzw. Programmevaluierung stellen sich für die Feststellung von Ergebnissen drei allgemeine Aufgaben: Erstens die einwandfreie Erfassung von Wirkungen für die TeilnehmerInnen ('treatment-of-the-treated-effekt'); zweitens, die Wirkungen auf Andere bzw. Nicht-TeilnehmerInnen (insbesondere Verdrängung und Substitution, sowie Mitnahme- und Droh-Effekte); und drittens die v.a. bei größeren Programmen eigentlich arbeitsmarktpolitisch relevante Frage nach dem sogenannten Gleichgewichtseffekt, nämlich, ob der Maßnahmeneinsatz auch Wirkungen auf die Arbeitsmarktlage bzw. das Beschäftigungsvolumen hat, oder nur Umverteilungen der Risiken und Vorteile bewirkt.

Im Hinblick auf die Umsetzung von AMP Evaluierungsergebnissen hat sich auch etwa zur gleichen Zeit, wo sich ein gewisser Konsens über stilisierte Ergebnisse herausbildete, die politische Grundkonstellation geändert, indem sich die Aufmerksamkeit von den Maßnahmen und Programmen weg und zum Gesamtrahmen der Beschäftigungspolitik hinwendete (vgl. dazu mehr im Abschnitt 2 über Steuerung und Governance). Ein wichtiger Schritt in diese Richtung waren die ersten makroökonomischen Evaluierungen, die in den erwähnten 'design features' mündeten, und dann die OECD Jobs Study (1994) und die Europäische Beschäftigungsstrategie (1998).

Die angesprochenen 'crucial design features' für Arbeitsmarktpolitik können als generische Regeln für den Einsatz aktiver AMP gesehen werden, die in die Funktionsweise des Arbeitsmarktes eingebettet sind und grundsätzlich die Wirksamkeit beeinflussen sollen:

- die finanzielle Kompensation in 'aktiven Maßnahmen' soll die 'passiven' Versicherungsleistungen/Sozialleistungen nicht übersteigen;
- es sollen gezielte Maßnahmen für bestimmte Personengruppen bzw. angestrebte Wirkungen eingesetzt werden (keine flächendeckenden 'Gießkannen-Maßnahmen');
- ein ausgewogenes Portfolio von Maßnahmen soll verhindern dass bei zu breit eingesetzten Maßnahmen sinkende Skalenerträge eintreten;
- 'passive' Maßnahmen und 'aktive' Maßnahmen sind als zusammenhängendes Gesamtsystem zu sehen, es soll kein zyklischer Wechsel von Personen zwischen diesen (reine Maßnahmekarrieren) stattfinden (Calmfors 1994).

Schon damals wurden zwei wesentliche Probleme bei der Umsetzung hervorgehoben, die nach wie vor gelten: Erstens die Risiken von Mitnahmeeffekten sowie der Verdrängung von anderen Personen oder der Substitution von Arbeitsplätzen durch die Maßnahmen, die im Zuge von Evaluierungen abgeschätzt werden sollten (aber oft nicht werden, und insbesondere in Monitoring-Systemen in der Regel nicht erfasst werden können); zweitens die Gefahr von rein 'kosmetischen Wirkungen' durch die statistische Umdefinition von Arbeitslosen in MaßnahmeteilnehmerInnen (Lars Calmfors hat vorgeschlagen, neben der standardmäßigen auch eine entsprechend bereinigte Arbeitslosen-Rate zu publizieren, um diesen Effekt zumindest sichtbar zu machen).

Die OECD Jobs Study als nächster Schritt der Formulierung eines umfassenden Rahmens für die AMP wird oft als neoliberales Projekt gesehen. Die Aufmerksamkeit wurde damit auf die Gestaltung des Beschäftigungssystems und der Arbeitsmarktinstitutionen insgesamt gelenkt, und die Bedeutung der makroökonomischen Steuerung mit Kontrolle der Inflation und der Staatshaushalte als Kernpunkte wurden betont. Mit stärker operativem Anspruch, und durchaus eigenständig konzipiert, wurde die EU-Beschäftigungsstrategie (implizit) oft im Kontext der einflussreicheren Jobs Study interpretiert.

Wichtig unter Gesichtspunkten der Evaluierung war die Erweiterung des methodischen Spektrums durch die Einführung von Peer Review Verfahren sowohl in der OECD Strategie als auch in der Europäischen Politik. Dabei evaluieren sich im Prinzip verschiedene Länder unter starker Mitwirkung von Expert/innen gegenseitig nach dem Muster Selbstevaluierung (Hintergrundbericht) - Site visit - Review (Prüfer/innenbericht; siehe die Gesamtdarstellung der OECD Strategie bei Cook 2008; Watts 2008; sowie für die EU Beschäftigungsstrategie: ec.europa.eu/employment_social/employment_strategy, und für das EU Peer Review Projekt: www.mutual-learning-employment.net). Diese Evaluierungen überschreiten im Spektrum ihrer Bewertungen bei Weitem die Praktiken der Maßnahmen- und Programmevaluierung, der Methodik wird jedoch wenig Aufmerksamkeit geschenkt, so dass der Unterschied zwischen systematischer Evaluierung und politischer Aushandlung nicht immer erkennbar ist.

1.3. Stilisierte Ergebnisse der Maßnahmen- und Programmevaluierungen

Zu den Wirkungen von arbeitsmarktpolitischen Maßnahmen hat sich seit dem Ende der 1990er aufgrund von einer Reihe von übergreifenden und länderspezifische Reviews ein Grundbestand an stilisierten Ergebnissen herausgebildet, der insgesamt kein besonders optimistisches Bild ergibt (vgl. v.a. den einflussreichen Review von Heckman/LaLonde/Smith 1999, die Aufarbeitung für Europa in Kluve 2006, sowie die Übersichten über Deutschland, die Schweiz, Schweden, U.K. und Österreich in Lassnigg 2007 und die dort angegebene Literatur). Es gibt auch international vergleichende Metastudien zu den Wirkungen der Maßnahmen auf die Teilnehmer und Teilnehmerinnen (treatment-of-the-treated-effect), aus denen sich stilisierte Ergebnisse zu diesen Wirkungen ableiten lassen. Verdrängungs- und Substitutionseffekte auf die Nicht-Teilnehmer und Nicht-Teilnehmerinnen sind damit nicht erfasst (deren Risiko wird v.a. bei der Förderung von regulärer Beschäftigung im privaten Sektor sehr hoch angesetzt, manche Schätzungen gehen bis zu 90%).

Ebenso sind Effekte auf die allgemeine Arbeitsmarktsituation sehr schwer zu erfassen, da bei festgestellten Zusammenhängen zwischen Maßnahmen und Arbeitslosigkeit oft die Richtung der Kausalität nicht klar ist (in Situationen, wo die Arbeitslosigkeit steigt, werden auch mehr Maßnahmen eingesetzt, und wo die Nachfrage konjunkturell bedingt steigt, sind logischerweise bessere Maßnahmewirkungen zu erwarten).

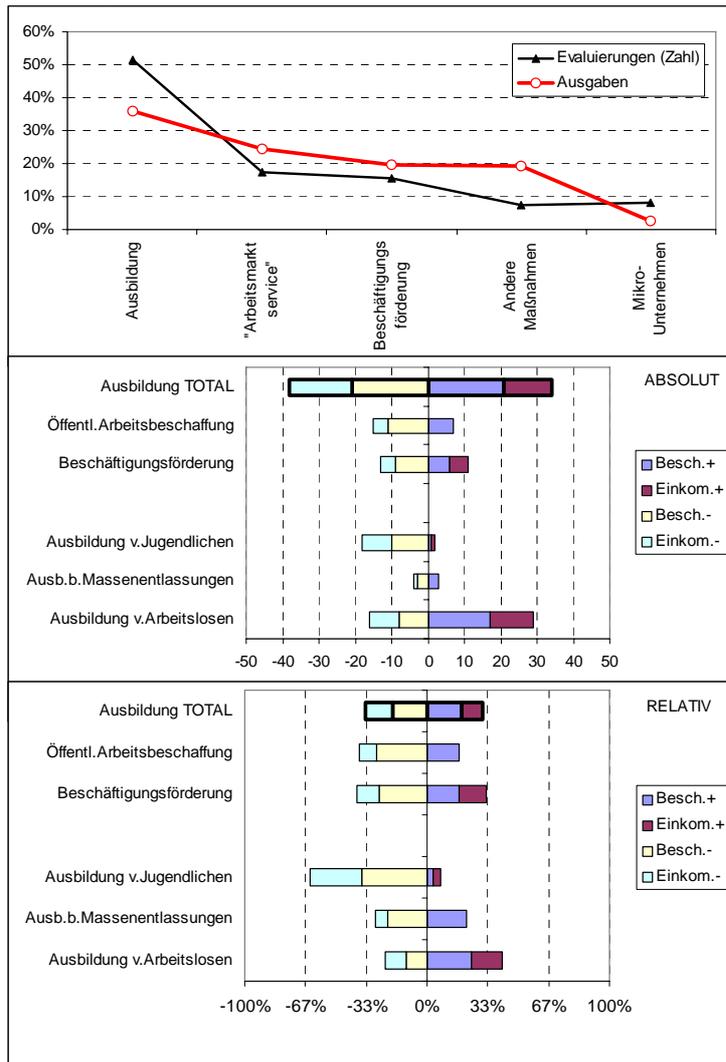
Die Metastudie von Betcherman/Olivas/Dar (2004) bezieht sich auf 109 Evaluierungsstudien die methodisch zumindest ein Counterfactual verwendet haben und wertet diese deskriptiv aus, indem Studien mit signifikanten positiven/negativen Wirkungen im Hinblick auf Beschäftigung und Einkommen gezählt werden (es werden also immer Ergebnisse im Vergleich zu einer Referenzgruppe dargestellt). Abb.2 fasst die Ergebnisse in stilisierter Form zusammen.

Man sieht, dass Ausbildungsmaßnahmen sowohl im Hinblick auf die Ausgaben als auch noch stärker im Hinblick auf die Evaluierung die größte Kategorie der AMP darstellen. Die Ergebnisse sind gemischt und nicht sehr erfolversprechend. Ausbildungsmaßnahmen haben etwa zur Hälfte positive Wirkungen, die anderen Maßnahmetypen haben überwiegend keine positiven Wirkungen. Wenn man unter den Ausbildungsmaßnahmen weiter unterscheidet, so zeigen v.a. Maßnahmen für Jugendliche ein sehr ungünstiges Bild, Ausbildung von Arbeitslosen zeigt das beste Ergebnis, mit einem Drittel der signifikanten Ergebnisse auf der positiven Seite.

Eine zweite Meta-Evaluierung (Kluve 2006) bezieht sich auf 137 Studien im EU-Raum und erfasst nur Beschäftigungswirkungen. Diese Studie geht über die deskriptive Zählung hinaus und versucht in einer multivariaten Auswertung die Wahrscheinlichkeit, ein positives Ergebnis zu erzielen, auch für eine Reihe von Hintergrundvariablen zu kontrollieren: experimentelles

Studiendesign, institutionelle Variablen (v.a. Regulierung der Beschäftigung), wirtschaftliche Lage und Länder. Auch hier haben Ausbildungsmaßnahmen moderate Ergebnisse, Maßnahmen für Jugendliche schneiden sehr ungünstig ab, Beschäftigungsförderung im privaten Sektor und Sanktionen bei fehlender Arbeitsbereitschaft haben aber bessere Wirkungen als Ausbildung. Die Hintergrundvariablen und auch die Länder sind meistens insignifikant, letzteres unterstreicht in der Sicht der Autor/innen die Universalität der Ergebnisse.

Abbildung 2: Stilisierte Darstellung der Verteilung der Ausgaben, Evaluierungsstudien* und Wirkungen in der Meta-Studie von Betcherman/Olivas/Dar (2004)



* Studien, welche Beschäftigungs- und Einkommenswirkungen erfassen, sind zweimal gezählt, negative Ergebnisse für die Darstellung mit (-1) multipliziert, insignifikante Ergebnisse sind nicht gezählt (diese machen die Differenz auf 100% aus).

Quelle: eigene Berechnung und Darstellung aufgrund von Betcherman/Olivas/Dar 2004

Ein wichtiges stilisiertes Faktum, das sich durch viele Evaluierungen und Reviews zieht, besteht in geschlechtsspezifischen Unterschieden, die auf tendenziell positive Wirkungen für Frauen, aber keine oder negative Wirkungen für Männer deuten (vgl. Bergemann/van den Berg 2007).

Diese Ergebnisse werden von Lechner/Wiehler (2007) möglicherweise relativiert, die aufgrund umfassenderer Daten v.a. Unterschiede zwischen Teilnehmerinnen und Nicht-Teilnehmerinnen betreffend ihrer Geburten und Kinderbetreuungsphasen herausarbeiten, so dass die geschlechtsspezifischen Wirkungen der Teilnahme verschwinden.

Wenn man die große Variation der Ergebnisse betrachtet, stellt sich die Frage, woher die Unterschiede kommen? Hier kann man eine Reihe von Faktoren benennen, von denen viele auch auf den noch immer unzureichenden Stand der Evaluierungspraxis verweisen:

- Methodische Probleme (Konstruktion des Counterfactual, Modellierung)
- Datenprobleme (kleine Ns, Besonderheiten im erfassten Zeitraum, fehlende Informationen über wichtige Merkmale von Personen und Maßnahmen)
- Erfassung der Maßnahmen als “black box”, keine Unterscheidung unterschiedlicher Maßnahmentypen
- (Unbeobachtete) Heterogenität der Effekte für verschiedene Gruppen
- (zu kurzer) Betrachtungszeitraum für Wirkungen (es gibt unterschiedliche Effekte kurz- und längerfristig)
- Bei Ausbildungsmaßnahmen spielt der „Lock-in“-Effekt eine wesentliche Rolle, da während der Ausbildung definitionsgemäß nicht nach Beschäftigung gesucht wird, die Vergleichspopulationen aber suchen.

Zusammenfassend kann man drei Gruppen von Herausforderungen identifizieren, mit denen die Evaluierung von Arbeitsmarktpolitik generell konfrontiert ist: Erstens, auf der Ebene der *Evaluierung* selbst bestehen schwache Daten, methodische Probleme und enge Evaluierungskriterien; zweitens auf der Ebene der *Maßnahmen* spielen teilweise hohe Kosten, eine schwache Intensität der Interventionen und große Unterschiede zwischen Maßnahmen eine Rolle; schließlich sind der oft *geringe Interventionsgrad* im Vergleich zur Problemlage und insbesondere auch der *isolierte und punktuelle Charakter* der Maßnahmen im Lebenslauf der betroffenen Individuen auf der Ebene der Interventionen als noch mehr oder weniger unerforschte aber wesentliche Aspekte zu nennen (zu all diesen Aspekten gibt es in den letzten Jahren Fortschritte und auch bereits Befunde; siehe z.B. das umfassende Material auf www.iza.org).

1.4. Eine stilisierte Einordnung der AMP im Rahmen der Evaluierungsansätze und -praktiken

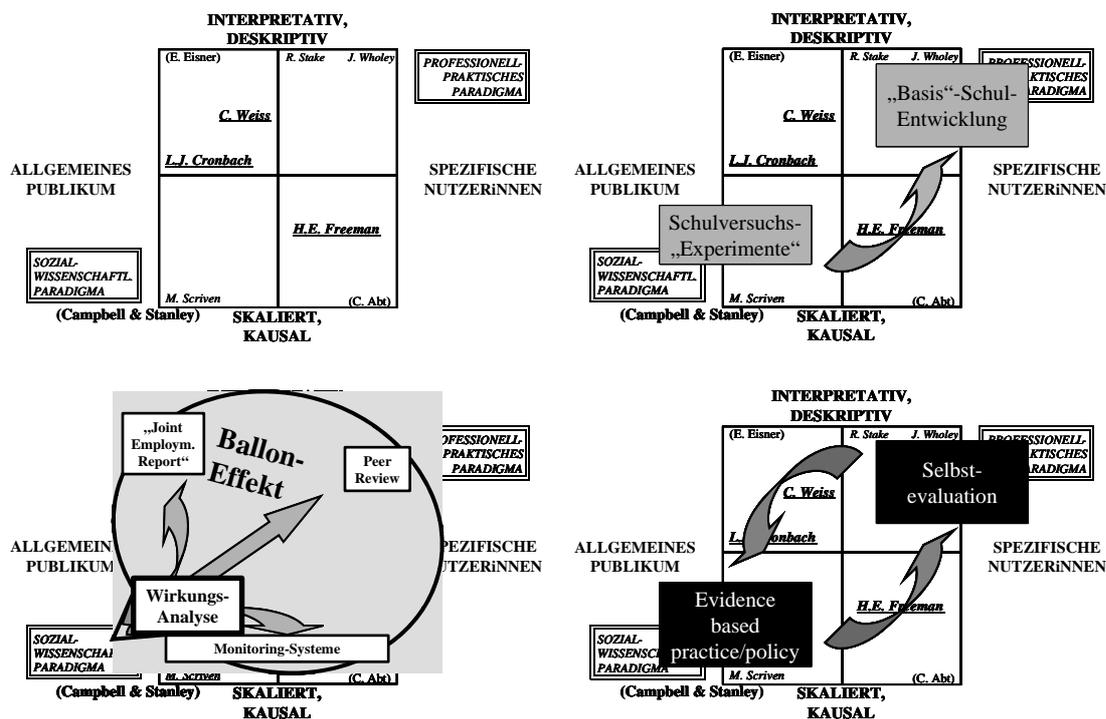
Versucht man die Evaluierungspraxis im Bereich AMP in das Gesamtspektrum der Evaluierungsansätze einzuordnen, so kann ein heuristisches Schema aufgrund von Analysen der Ansätze der US-„Klassiker/innen“ herangezogen werden, das dieses Spektrum in einem zweidimensionalen Raum von Methodik und Nutzung einordnet (Abb.3).

In diesem Schema ergeben sich zwei gegenüberliegende Muster eines sozialwissenschaftlichen Paradigmas im Sinne des bei weitem einflussreichsten Werkes von Campbell/Stanley (skaliert-kausale Methodik für das allgemeine Publikum) auf der einen Seite und eines professionell-praktischen Paradigmas (interpretativ-deskriptive Methodik für spezifische Nutzer/innen), das mit dem Werk von Robert Stake und noch deutlicher von Joseph Wholey in Verbindung gebracht werden kann, auf der anderen (zu diesem Quadranten vgl auch Guba/Lincoln 1987). In diesem vereinfachenden Schema zeigt sich die starke Ausdifferenzierung der Praktiken der Evaluierung, die in der Zwischenzeit noch durch viele Ansätze bereichert worden ist, die oft konkurrierend gegeneinander antreten.

Man kann aber auch die Entwicklung in der Weise sehen, dass sich diese gegenüberliegenden Ansätze eigentlich ergänzen müssen, und dass eine erfolgreiche Umsetzung letztlich darauf beruht, dass die Kompetenz der Evaluierung in Form der systematischen Selbstevaluierung bei den Akteur/innen entwickelt wird, die evaluiert werden sollen, ergänzt durch eine systematische Außenkontrolle und Spiegelung.

Versucht man die Entwicklung der AMP Evaluierung in dieses Schema einzubringen, so kann man eine Art 'Ballon-Effekt' konstatieren, eine Entwicklung dergestalt, dass in der Ecke des sozialwissenschaftlichen Paradigmas hineingeblasen wird, und sich die Aktivitäten in alle Richtungen ausbreiten, das Kraft- oder Energiezentrum bleibt jedoch beim sozialwissenschaftlichen Paradigma. Drei prominente Beispiele für die Ausbreitung der Aktivitäten über das Feld sind die Entwicklung der Monitoring-Systeme (quantitativ für spezifische Nutzung), die Peer Reviews (qualitativ für spezifische Nutzung), und die politischen Bewertungen in der EU-Beschäftigungsstrategie (qualitativ für das allgemeine Publikum). In der AMP ist diese Ausdifferenzierung zwar vom skizzierten normativen 'State-of-the-Art'-Konzept her auch empfohlen, die Praxis ist jedoch stark in der linken unteren Ecke konzentriert geblieben. Dies gilt selbst für die Methodik der in den OECD Peer Reviews herangezogenen Materialien. Stärker qualitative Methoden der Evaluierung und Selbstevaluierung werden in den EU-Aktivitäten angewendet.

Abbildung 3: Schematisches Mapping der Evaluierungsansätze und Entwicklungen



Quelle: eigene Darstellung auf Basis von Williams 1989, Shadish/Epstein 1987

Vergleicht man die Bewegung im Bereich der AMP mit der Entwicklung im Bereich des Bildungswesens, so kann dort eher eine zyklische Bewegung beobachtet werden, die mit den (in Österreich dilettantischen, verzerrten und schief gegangenen) 'Experimenten' der 1970er beim

sozialwissenschaftlichen Paradigma beginnt, dann eine starke Wendung in Richtung qualitativer (Schul)-Entwicklung an der Basis genommen hat, und nun mit den vergleichenden 'Large-Scale-Assessments (LSAs)' wieder auf höherer Stufe beim Anfangspunkt gelandet ist – mit dem Unterschied, dass heute großes Einverständnis über die Notwendigkeit des Zusammenspiels der beiden Paradigmen besteht.

Versucht man die drei Elemente von Politik(ziel)-, Wirkungs- und Implementationsevaluierung in diesem Feld zu verorten, so impliziert der Kernbereich der AMP Wirkungsevaluierung ein Operationalisierung von Zielen, die aufgrund der Modelle sehr technisch ausfallen muss. Eine nähere Zieldiskussion wird meistens nicht angestellt. Es ist jedoch auch zu konstatieren, dass aufgrund der vorhandenen Ergebnisse durchaus auch verallgemeinerte Zieldiskussionen geführt werden.

John P. Martin (1998, 300) in den Schlussfolgerungen zu seinem eher pessimistischen Überblick über die Wirkungen von aktiver AMP bereits betont, dass „...one should not oversell the case for active labour market policies. More effective active labour market policies (...) are not a magic bullet on their own to solve the unemployment problem. (...) they require a reasonably buoyant supply of job vacancies in order to be effective. If an economy is generating few vacancies, one should not be surprised if active measures prove to be relatively ineffective. Aggregate demand matters too.“

Über die grundlegenden Zielsetzungen der Vollbeschäftigung gibt es heftige Debatten. Während manche Beobachter davon ausgehen, dass die aktuellen neo-liberal inspirierten Konzepte dieses Ziel grundsätzlich verabschiedet haben, wird dieses Ziel z.B. in der offiziellen österreichischen arbeitsmarktpolitischen Programmatik nach wie vor in den Vordergrund gerückt.

Schneider/Eichhorst/Zimmermann (2006, 1) zweifeln an den Möglichkeiten der politischen Beeinflussung der Arbeitsmarktlage: „Die Arbeitsmarktpolitik steht vor großen Schwierigkeiten. Zunächst ist es ein Problem, dass die Politik weiter an der Fiktion festhält, sie könne und müsse die letzte Verantwortung für die Arbeitslosigkeit übernehmen.“

Besonders pointiert hat James Heckman immer wieder zu s.E. sinnvollen oder effektiven politischen Zielsetzungen aus der Sicht seiner Ergebnisse Stellung genommen. Dabei ist auch bemerkenswert, dass dazu nicht nur die Ergebnisse aus den sozial- und arbeitsmarktpolitischen Evaluierungen herangezogen werden, sondern auch Ergebnisse aus bildungs- und neuerdings verteilungspolitischen Analysen (diese Ergebnisse finden manchmal auch ihren Weg in politische Dokumente der Europäischen Union). Insbesondere die Förderung der Bildung von sozial Benachteiligten in der frühen Kindheit wird dabei sehr stark betont, während Weiterbildung in späteren Phasen als wirkungsloses soziales Instrument herausgearbeitet wird.

1.5. österreichische Erfahrungen

Wenn man auf diesem Hintergrund die österreichischen Erfahrungen mit der Evaluierung von Arbeitsmarktpolitik betrachtet, so bestätigen sich viele der bisher herausgearbeiteten Merkmale und Probleme (vgl. Wroblewski 2009, Hofer/Weber 2006). Es gibt inzwischen eine Fülle an Evaluierungsstudien, die meistens vom zuständigen Arbeitsmarktservice (AMS) oder Arbeitsministerium in einem 'Markt' von einigen außer-universitären Instituten in Auftrag gegeben und durchgeführt werden, und auch häufig der Öffentlichkeit zugänglich sind (vgl. Übersicht 2 im Anhang, sowie www.ams-forschungsnetzwerk.at). Seit dem Beginn der Evaluierungsaktivitäten, und angefeuert durch die EU-Förderungen wird parallel am Aufbau einer umfassenden Datenbasis für das Monitoring der Maßnahmen gearbeitet. Diese kombiniert Daten der sozialversicherungspflichtigen Beschäftigten und des AMS und ist mittlerweile auch

öffentlich für Forschungszwecke zugänglich. Es gibt jedoch außerhalb der Auftragnehmer-Auftraggeberbeziehungen sehr wenig Austausch und Auseinandersetzung über die methodischen Entwicklungen und die Qualität der Evaluierungen, und die akademische Anbindung und Publikationstätigkeit ist sehr gering ausgeprägt und stellt in der 'Szene' auch keinen Wert dar.³

Tabelle 1: Typologie von ausgewählten Evaluierungen in Österreich

Jahr	Maßnahme	Region	Implementation	Verbleib	Netto-Wirkung	Fiskalanalyse
2008	Jugendliche (JASG-Lehrgänge)	REG	X	X	-	-
	Berufsqualifizierung von MigrantInnen	REG	X	X	-	-
2007	Jugendliche (JASG-Lehrgänge)	REG	X	X	-	-
2006	Unternehmensgründung (UGP)	NAT	-	X	-	-
2003	Territorialer Beschäftigungspakt (TBP)	REG	X	-	X	X
2002	Stiftungsähnliche Maßnahme	REG	X	X	-	X
	Karenz plus	REG	X	-	-	-
	Bildungsmaßnahmen	REG	X	X	-	-
2001	EDV- und kaufmännischen Kurse	REG	X	-	X	-
	Qualifizierungsmaßnahmen (TBP)	REG	X	-	-	-
	Sozialökonomischer Betrieb	REG	X	-	-	-
	Lehrstellenförderung Maßnahmen Frauenförderung	REG	-	X	-	X
2000	Stiftungsähnliche Maßnahme	REG	X	X	-	X
	Sozialökonomische Betriebe	NAT	X	-	X	X
1999	Trainingsmaßnahmen	NAT	X	X	-	-
	Qualifizierungsverbund	REG	X	-	-	-
	Wiedereinsteigerinnen	NAT	X	-	X	-
1997	Betriebliche Eingliederungsbeihilfe	REG	X	X	-	X
	experimentelle Arbeitsmarktpolitik	NAT	X	-	X	X
Summe	20	REG: 15	18	10	5	7
Prozent	100%	75%	90%	50%	25%	35%

Quelle: Eigene Auswertung der in AMS-Infos 1997-2009 im Titel als Evaluierung gekennzeichneten Einträge (es kann noch weitere Evaluierungen geben, die nicht im Titel erkennbar sind; und es sind auch nicht alle in diesem Medium publiziert).

Die Evaluierungstätigkeit von speziellen Maßnahmen hatte einen 'Boom' um 2000, die meisten wurden in Wien durchgeführt. Zwischen 2003 und 2005 in der Zeit der ESF-Evaluierungen gibt es keine anderen Publikationen, anschließend erfolgte eine Prüfung durch den Rechnungshof (RH 2007) und die erwähnte internationale Wirkungsevaluierung durch Lechner et al (2007). Tabelle 1 gibt einen Überblick über 20 von AMS publizierte spezielle Evaluierungen.

Implementationsaspekte werden in fast allen Studien behandelt, in der Regel in Form von deskriptiven Erhebungen von Teilnehmer/innenstrukturen, Merkmalen der

³ Ausnahmen bestätigen die Regel: In den letzten Jahren sind einige akademische Publikationen entstanden, die v.a. mit zwei Forscher/innen verbunden sind, die bezeichnenderweise nicht zur 'Szene' gehören (Rudolf Winter-Ebmer und Andrea Weber) und auch nur sehr bedingt auf diese zurückwirken. Auch aus einer rezenten Evaluierung durch eine Forscher/innengruppe aus St.Gallen (geleitet durch Michael Lechner) sind erste akademische Publikationen auf Basis österreichischer Daten verfügbar.

Maßnahmedurchführung und Einschätzungen bzw. Bewertungen seitens der Beteiligten. In der Hälfte der Evaluierungen werden auch deskriptive Verbleibsanalysen durchgeführt, so dass der typische Fall eine Kombination aus Implementations- und Verbleibsanalyse ist (eigentlich Fragestellungen, die durch Monitoring auch – weniger detailliert – erfasst werden können). In einem Viertel der Evaluierungen werden Netto-Wirkungen erfasst und in einem Drittel fiskalische Aspekte (diese werden jedoch meistens auf die Brutto-Wirkungen bezogen). In einem kleinen Teil der Studien wird also mit einem Counterfactual gearbeitet, meistens mit gematchten Vergleichsgruppen (die neueren Verfahren werden bisher wenig angewendet), Kosten-Nutzen-Aspekte werden in Form von (Brutto)-Fiskalanalysen berücksichtigt.

Der Tendenz nach finden die Studien (sehr) geringe positive Effekte, was im wesentlichen der Legitimierung dessen dient, was getan wird. Die Evaluierungen können zu (kleinen) Modifikationen im Maßnahmenangebot führen, die wesentlichen politischen Entscheidungen werden jedoch abseits der Evaluierung getroffen, und meistens auch nicht evaluiert. So war beispielsweise das verfügbare Finanzvolumen in den letzten Jahren immer wieder durch sehr kurzfristig getroffene Regierungsentscheidungen über massive Mittelverschiebungen bestimmt. Auch wenn diese Mittel stoßweise ausgeweitet werden, kann das System der Maßnahmenbereitstellung durch die entsprechend kurzfristigen Dispositionen in Schwierigkeiten kommen, die dann seinerseits mit einem hohen Anteil an prekärer Beschäftigung gemeistert werden.

In jüngerer Zeit sind drei breite Ergebnisevaluierungen zum österreichischen Maßnahmenpektrum durchgeführt worden (WIFO 2005, Lechner et al 2007, Synthesis 2006), die durchwegs wesentliche daten- oder auswertungsbedingte Lücken für die grundlegenden Politikentscheidungen aufweisen. Obwohl Österreich einen prononcierten Schwerpunkt im Bereich der Ausbildung legt, sind die Förderinformationen nach wie vor nicht ausreichend, um zwischen unterschiedlichen Ausbildungsmaßnahmen unterscheiden zu können. Auch sind die wichtigen Maßnahmen der Förderung von regulärer Beschäftigung im privaten Sektor entweder aufgrund fehlender Informationen nicht einbezogen worden (Lechner et al. 2007) oder aufgrund der gewählten Klassifikation nicht von Ausbildungsmaßnahmen unterscheidbar (Synthesis 2006). Gemessen an den drei eingangs unterschiedenen Basis-Kategorien der Evaluierung spielt die Ziel- oder Politik-Evaluierung so gut wie keine Rolle. In Abweichung vom weiter oben charakterisierten Gesamtbild, umfasst der größte Teil der Studien – wenn auch oft mit weichen Methoden und deskriptiven Fragestellungen – eine überwiegende Komponente von (ausgewählten) Fragen der Implementation. Oft ist auch eine kleinere Komponente mit der (deskriptiven) Analyse von Wirkungen enthalten.

2. Steuerung von Arbeitsmarktpolitik: Neue Governance-Modelle und Politik-Paradigmen jenseits der Evaluierung

2.1. Hintergrund und Phänomenologie: Österreich im internationalen Kontext

Der Kernbereich der arbeitsmarktpolitischen Evaluierungsaktivität, die 'aktive Arbeitsmarktpolitik' ist im Zuge der wohlfahrtstaatlichen Entwicklung v.a. seit den 1960er Jahren als neuer Politikbereich aufgetreten, und fungierte als Ergänzung der traditionellen Funktionen von Arbeitslosenversicherung und Arbeitsvermittlung, und auch der (Keynesianischen) Wirtschaftspolitik. Als Arbeitsmarktförderung war die 'aktive Arbeitsmarktpolitik' immer eine flexible und zu den traditionellen Funktionen 'zusätzliche' entscheidungsabhängige Aktivität, die teilweise auch unterschiedliche Zwecke verfolgte (in den

USA eher Armutsbekämpfung, in Europa eher zusätzliche Beschäftigungseingliederung) und deren Wert immer wieder in Frage gestellt wird.

Die Durchführung erfolgt v.a. in Form von Maßnahmen und Programmen, ist diskretionär und wegen der tendenziell eher nicht hoch angesehenen Klientel oft politisch umstritten, und im Umfang relativ klein im Vergleich zu den passiven Maßnahmen. Die Kritik am Wohlfahrtsstaat seit den 1980ern hat auch zu einem Wandel der Arbeitsmarktpolitik geführt, und die neoliberale Ideologie hat die Aufmerksamkeit stärker auf die institutionelle Gestaltung der Beschäftigung und der Lohnfindung gelegt, gleichzeitig haben sich aber auch immer wieder – bis zur gegenwärtigen in ihren Auswirkungen noch unabsehbaren Krise – die Beschäftigungsprobleme verschärft.

Mit der EU-Beschäftigungsstrategie am Ende der 1990er, die dann noch durch die Inklusionsstrategie ergänzt wurde, ist insofern eine neue Situation eingetreten, als ein gewisser politischer Verpflichtungsgrad für diesen Politikbereich hergestellt. Dafür wurde eine damals ganz neue spezielle Governance-Methode entwickelt (Rhodes 1996) wurde, die 'Open Method of Coordination (OMC)' als freiwillige und weiche Regulierung mit schwacher oder nicht vorhandener gesetzlicher Basis. Es wurde damit ein erweitertes arbeitsmarktpolitisches Instrumentarium in Form von 'Leitlinien' kodifiziert, und EU-weit auch einem losen qualitativ-politischen Beurteilungsverfahren in Form des 'Joint Employment Report (JER)' unterzogen. Evaluierung hat damit eine gewisse Verankerung gefunden, jedoch – nicht zuletzt wegen der Kurzfristigkeit und raschen Abfolge der Aktionen – auf schwacher methodischer Basis. Wesentliche Fragen müssen oft auf die Zukunft verschoben werden, wenn die Programme aber bereits beendet sind.⁴

In die politischen Konflikte zwischen Wohlfahrtsstaat und Neoliberalismus wurde die Arbeitsmarktpolitik einbezogen, indem sie zunächst gewissermaßen als Konkurrenzinstitution zur Wirtschaftspolitik etabliert wurde. Die traditionelle „Nachfragesteuerung“ auf nationaler Ebene wurde wegen Maastricht mehr oder weniger unmöglich, was in Österreich aufgrund der noch frischen Tradition des „Austrokeynesianismus“ einen besonders starken ideologischen Kontrast erzeugte, mit verbreitetem Zweifel an der potentiellen Wirksamkeit der Maßnahmen (vgl. näher Lassnigg 2004). Es wurde argumentiert, dass als die Beschäftigungsstrategie kein adäquater Ersatz für wirtschaftspolitische Nachfragesteuerung sein könne, weil sie bestenfalls zur Umverteilung von Beschäftigungsmöglichkeiten dienen könnte.

Für die politische Steuerung kommt der AMP aber eine gewisse Vorreiterrolle zu. Schon in den 1980er Jahren wurde in Österreich im Bereich der Arbeitsmarktpolitik das Modell der 'Programmbudgetierung' des US-Verteidigungsministers McNamara als strategische Orientierung herangezogen, das als Vorläufer von 'New Public Management' gesehen werden kann. In diesem Sinne wurde die 'aktive Arbeitsmarktpolitik' als neuer Politikbereich von vorneherein auf Basis eines technokratischen 'Steuerungskonzeptes' der Zielsteuerung entwickelt. In den 1990ern erfolgte eine Ausgliederung des Arbeitsmarktservice aus der öffentlichen Verwaltung mit dem Arbeitsministerium als Aufsichtsbehörde und eine (teilweise)

⁴ Das in seinen Zielen äußerst ambitionierte EQUAL-Programm sollte im Wege der Aktionsforschung wesentliche neue Aspekte der sozialen Ungleichheit, Benachteiligung und Diskriminierung und Lösungsmöglichkeiten dafür herausfinden – im österreichischen Evaluierungsbericht zeigt sich aber vorwiegend eine Darstellung der umgesetzten Maßnahmen und der kurzfristigen Wirkungen (die stark auf Cream skinning hindeuten), sowie eine deskriptive Darstellung vielfältigster vorwiegend inkrementeller Maßnahmen, deren innovativer Charakter zweifelhaft ist. Eine vertiefte Auswertung des innovativen Charakters wird mit Vorschlägen wie der Errichtung einer Informationsdrehscheibe, einer Forschungsplattform und einer Nachbetrachtung auf die Zukunft verschoben (www.equal-esf.at; www.kmuforschung.ac.at/de/Projekte/Equal/equal.htm).

Privatisierung dieser Funktionen durch die Zulassung privater Arbeitsvermittler. Es wurde auch ein loses System der Zielsteuerung der Arbeitsmarktpolitik mittels Indikatoren eingerichtet (siehe www.bmask.gv.at/cms/site/liste.html?channel=CH0700; sowie: http://www.ams.at/14169_18574.html).

Zunehmend hat sich die Aufmerksamkeit – nicht zuletzt wegen der geringen Wirkungen der 'aktiven Arbeitsmarktpolitik' – aber auf die breitere wirtschaftspolitische Einbindung der Arbeitsmarktpolitik (Ablösung der 'Nationalen Aktionspläne' durch die 'Nationalen Reformprogramme') und auf das gesamte Regulierungsregime des Arbeitsmarktes und der sozialen Sicherung verlagert. Als Stichworte fungieren hier „Flexicurity“ (EU) und „Life-course-policy“ (OECD). Die Arbeitsmarktpolitik und in Verbindung damit auch die Evaluierung bekommt damit konzeptionell einen neuen politischen Stellenwert.

Die Europäische OMC, wurde teilweise geradezu zu einem Paradigma des Governancebegriffs. In diesem Ansatz der gegenseitigen (Selbst)-Verpflichtung ohne gesetzlicher Sanktionsbasis besteht die politische Steuerung *aus Leitlinien und Zielen in Kombination mit Berichterstattung und Evaluierung*. Jonathan Zeitlin (2005, 2006) hat diese Governance-Methode als paradigmatischen Fall von 'Policy Learning' analysiert und hat drei Hauptelemente dieser Methode herausgearbeitet“:

- Heuristic. Es werden neue *Konzepte* etabliert (z.B. 'Employability')
- Capacity building. Es werden neue *Informationsbasen* eingerichtet (z.B. EU-Strukturindikatoren und Benchmarks)
- Maieutic. Es werden neue Instrumente der *Reflexion* aufgebaut (z.B. Evaluierung der 'Nationalen Aktionspläne' und ESF-Begleitausschüsse).

In der Beschäftigungsstrategie wurden beispielsweise die Konzepte der Beschäftigungsfähigkeit und der Flexibilität (später auch Flexicurity) formell in den politischen Prozess eingeführt. In den Förderstrategien des Europäischen Sozialfonds spielen operationelle Ziele und standardisierte Monitoring-Mechanismen als neue Informationsbasen eine wesentliche Rolle. Wenn auch die Nationalen Aktionspläne wie auch die Begleitausschüsse der Strukturfonds als Reflexionsmechanismen gesehen werden können, so spielt die Evaluierung in den politischen Aushandlungsprozessen eine untergeordnete Rolle (Die Wirkungsevaluierung der Arbeitsmarktpolitik wurde beispielsweise erst lange nach Programmende öffentlich verfügbar). Evaluierung ist in diese Prozesse zwar explizit (dem Namen nach) eingebaut, aber weder die Konzepte noch die Praxis entsprechen dem skizzierten 'State-of-the-Art'. Als Hemmnisse einer rigorosen Evaluierung stellen sich die folgenden Aspekte heraus:

- *politische Rücksichten*. Es geht um die Lukrierung bedeutender Fördermittel, die eigentlich Resultat eines politischen Verhandlungsprozesses sind, daher liegt es in der Natur der Sache, dass aus nationaler Perspektive die Vergabe dieser Mittel durch widerständige Evaluierungsergebnisse nicht gefährdet werden darf. Dies stellt ein Hindernis für eine ernsthafte *Policy- oder Ziel-Evaluierung* dar, da die Ziele eigentlich nicht mehr hinterfragt werden dürfen.
- *zu enger Zeithorizont*. Die Förderprogramme folgen einer Logik, in der ambitionierte Ziele in einem sehr engen Zeithorizont erreicht werden sollen, wobei der zentrale Evaluierungsschritt der 'Halbzeitbewertung' de facto bereits zu einem Zeitpunkt erfolgt, wo Wirkungen der Programme noch gar nicht erfasst werden können (obwohl nach den offiziellen Evaluierungskonzepten Ergebnisse eine wichtige Dimension darstellen sollen). *Wirkungsevaluierung* kann also eigentlich nicht stattfinden, zumindest nicht so, dass sie operativ wirksam werden könnte.

- *begrenzte Mittel*. Im Vergleich zu den vielfältigen und umfangreichen Programmen wurden – zumindest nach den österreichischen Erfahrungen – für die Evaluierung viel zu begrenzte Mittel aufgewendet, um eine 'State-of-the-Art'-Evaluierung durchzuführen.

Wesentliche Fortschritte sind v.a. beim bereits vor Beginn der EU-Programme begonnen Aufbau von öffentlich nutzbaren Monitoring- und Informationssystemen im Bereich der Arbeitsmarktpolitik zu verzeichnen. So ist mittlerweile ein im Vergleich zu anderen Bereichen sehr ausführliches und gut nutzbares System der Arbeitsmarktstatistik verfügbar, das auch Informationen über Förderungen enthält.⁵ Darüber hinaus befindet sich ein für Forschungszwecke verfügbares System des Arbeitsmarktmonitoring im Aufbau, das Auswertungen zu vielen Fragestellungen der Evaluierung ermöglicht, aber auch immer noch fundamentale Lücken enthält (siehe dazu Lechner et al. 2007, Kap.8).⁶

2.2. Governance-Modelle: Bürokratie, Markt und 'New Public Management'

In verdichteter Form kann die skizzierte Phänomenologie durch die Begrifflichkeit der neuen Steuerungsmodelle bzw. von Governance dargestellt werden, wobei der Begriff Governance insofern besser geeignet erscheint, als er die politische Komponente direkt einschließt, während Steuerung eher den technisch-kybernetischen Aspekt in den Vordergrund rückt. Governance betrifft die Art der Zusammengehörigkeit der verschiedenen AkteurInnen in einem Politikfeld und schließt im Unterschied zu Government auch die vielfältigen informellen Mechanismen und AkteurInnen ein. Als paradigmatische Formen von Governance fungieren die *Bürokratie* und der *Markt*, die auch in den neoliberalen Diskursen diametral gegenübergestellt wurden. Im Lauf der Diskussion hat sich eine dritte Zwischenform herausgebildet, die man als Konzept des '*New Public Management (NPM)*' bezeichnen kann. Dabei wird insbesondere der Abbau der Bürokratie durch die Etablierung von Steuerungsformen aus dem privaten Sektor verfolgt, ohne jedoch unbedingt eine volle Marktsteuerung anzustreben. Es soll ein neues Verhältnis zwischen Politik im engeren Sinne und Praxis etabliert werden, das anstelle der inputseitigen Vorschriften eine Ergebniskontrolle auf Basis von Zielvorgaben und eine ausgeprägte Autonomie der durchführenden Stellen in der Umsetzung der Zielvorgaben impliziert. Damit entsteht im Vergleich zur Bürokratie eine relative Entkoppelung von Zielen, Umsetzung und Ergebniskontrolle, die auch für die Evaluierung Konsequenzen hat. Vor allem die Ergebniskontrolle wird zu einem vordringlichen Problem, während die Zielfestlegung de facto stärker in das politische System driftet und die Umsetzung stärker Aufgabe der PraktikerInnen wird, die größere Diversität annehmen und direkt an die Ergebnisse gebunden werden soll. Dieses System impliziert für die Evaluierung zwei gegenläufige Tendenzen: Einerseits entsteht eine Tendenz zur Fragmentierung, indem das 'State-of-the-Art'-Modell von Politik(ziel)-Wirkungs-Implementations-Evaluierung vordergründig leichter in seine Teilaspekte zerlegbar wird, die von unterschiedlichen Akteur/innen vertreten werden – andererseits wird aber aufgrund der steigenden Komplexität des Gesamtsystems die Anforderung einer Gesamtevaluierung wichtiger, nicht zuletzt weil eben bestimmte Anforderungen an eine sinnvolle Ergebniskontrolle bestehen (s.o. das 'naive' Politik-Konzept). Eine Instanz für diese Aufgabe der Gesamtevaluierung tritt aber tendenziell in den Hintergrund, insbesondere wenn man auf Seite der Politik eine Tendenz zum Opportunismus unterstellt (was ziemlich allgemein der Fall ist).

⁵ Vgl. http://www.ams.at/ueber_ams/14172.html

⁶ http://www.bmwa.gv.at/NR/rdonlyres/B4A8868C-B4F8-4036-B2CD-0DFA13E24C3D/0/Arbeitsmarktmonitoring_Homepage.pdf

Aus der Phänomenologie kann man auch entnehmen, dass von der Entwicklung her gesehen die Evaluierung auf dem Hintergrund des bürokratischen Systems entstanden ist, und nun gewissermaßen einem neuen Governance-Modell gegenübersteht. Im Bereich der Arbeitsmarktpolitik besteht auch die Besonderheit, dass sie direkt an der Schnittstelle zwischen dem Governance-System des (Arbeits)-Marktes und der Politik liegt, und in ihrem Erfolg letztlich vom Markt abhängig ist. In der Phänomenologie wurden für das österreichische Governance-System der AMP Entwicklungslinien gezeichnet, die von einer einwandfreien Bürokratie zu einem Modell ähnlich der (alten) deutschen Bundesanstalt für Arbeit umgestaltet wurde. In Deutschland verstärken die Hartz-Reformen diesen Übergang, wobei auch eine typische Verbindung zur Evaluierung hergestellt werden kann: Die Reformen wurden ohne viel Evaluierung durchgeführt, werden aber dann im politischen Auftrag einem umfassenden Evaluierungsprozess unterzogen, der sehr viel – auch umstrittenes – Wissen erzeugt (einen guten Überblick geben die Publikationen auf www.iza.org, z.B. Eichhorst/Zimmermann 2007, Schneider/Eichhorst/Zimmermann 2006; vgl. zur Diskussion und Widerständigkeit auch www.monapoli.de). Ob dieses Wissen der Evaluator/innen aber in die weiteren politischen Prozesse einfließt bzw. einfließen wird, kann durchaus angezweifelt werden.

2.3. Politik-Paradigmen: Jobs Study, Flexicurity und Life-Course-Policy

Es befinden sich aber nicht nur die Steuerungsstrukturen in Veränderung, sondern es sind auch neue Politik-Paradigmen für die Arbeitsmarktpolitik entstanden, die die AMP-Instrumente in neue Kontexte stellen.

2.3.1. Die OECD Jobs Study

Die OECD Jobs Study als konzeptioneller Hintergrund für die OECD Beschäftigungsstrategie war in den 1990ern ein wesentlicher Faktor bei der Verbreitung der heute vertrauten Themen von Deregulierung, Globalisierung und technologischem Wandel als Determinanten der Wirtschafts- und Beschäftigungsentwicklung (Watts 2008, 5). Die hohe Arbeitslosigkeit wurde nun darauf zurückgeführt, dass die Länder mit ihren Institutionen nicht ausreichend auf diese Determinanten reagieren würden. Ein breites Programm von Strukturreformen wurde entwickelt, in dem die aktive Arbeitsmarktpolitik nur ein (kleines) Element darstellt. In Stichworten kann das Gesamtprogramm durch die folgenden Elemente charakterisiert werden:

- wachstumsfördernde und nachhaltige makroökonomische Politik (v.a. Reduzierung der Budgetdefizite)
- Flexibilisierung der Arbeitszeiten
- Flexibilisierung der Löhne und Anpassung der Lohnniveaus an lokale und qualifikatorische Voraussetzungen
- Reform der Beschäftigungssicherheit (Abbau von Kündigungsschutz)
- Stärkung der aktiven AMP
- Verbesserung der Qualifikationen und Kompetenzen
- Reform der sozialen Sicherung (Arbeitslosengeld, Sozialhilfe und Steuersysteme)
- Verstärkung des Wettbewerbs auf den Produktmärkten (gegen Monopolisierungstendenzen).

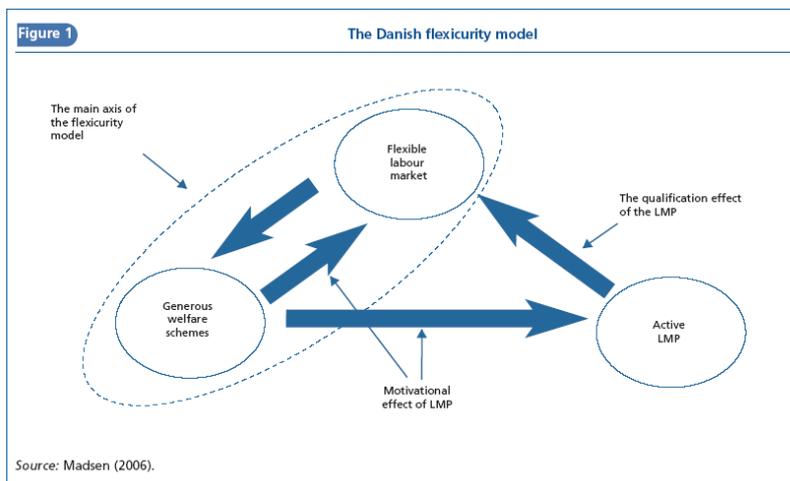
Dieses Gesamtprogramm wurde in eine im Laufe der Zeit immer wieder angepasste Strategie umgesetzt, die in Form von Peer Reviews in den Mitgliedsländern evaluiert wird. Die Annahme von Kritik und Empfehlungen ist freiwillig. Wesentlich ist dieses Programm auf die Förderung von Wachstum und Beschäftigung ausgerichtet, wobei die maximale Ausschöpfung nicht nur der

Gegebenheiten sondern der wirtschaftlichen Potentiale in den Mittelpunkt gestellt wird ('Potentialwachstum' im Vergleich zum tatsächlichen Wachstum als neues Maß). Die beiden weiteren Konzepte können ihre Verwandtschaft mit dem skizzierten Programm kaum verleugnen.

2.3.2. Flexicurity

Mit dem Konzept der 'Flexicurity' wurde ein neuer Ansatz der Beschäftigungspolitik konzipiert, in der die *aktive AMP* einen Eckpunkt des 'Goldenen Dreiecks', gemeinsam mit den beiden anderen Eckpunkten von *Beschäftigungsflexibilität* und *generöser Existenzsicherung* darstellt (vgl. die Schematisierung in Abb.4). Die Bewertungsmaßstäbe für die Realisierung dieses Konzeptes sind Kündigungsschutz, Ausgaben für aktive Arbeitsmarktpolitik und Weiterbildung und die Ersatzrate der Arbeitslosenversicherung. Empirische Auswertungen von (groben) Indikatoren ergeben verschiedene Cluster von EU-Ländern mit Dänemark und weniger ausgeprägt Schweden und Finnland mit hohem Grad an Flexicurity, Österreich und Deutschland liegen bei Flexibilität und Sicherheit gemeinsam mit Frankreich und Belgien auf einem mittleren Niveau. Dieser Ansatz ist jedoch auch in verschiedener Hinsicht umkämpft, indem er in operativen politischen Vorschlägen häufig auf Abbau von Kündigungsschutz (Flexibilität) kombiniert mit Weiterbildung (Sicherheit) verkürzt wird (man könnte sagen 'Flexification' anstatt 'Flexicurity'). Insbesondere die dritte Säule der sozialen Absicherung wird gerne vergessen, bzw. auch als so stark traditionsgebunden gesehen dass daraus Politikinvarianz abgeleitet wird. Auch auf Forschungsebene bestehen Zweifel an der Tragfähigkeit dieses Konzeptes für eine wirksame Reform der Arbeitsbeziehungen (Abu Sharkh 2008).

Abbildung 4: Schematische Darstellung des Flexicurity-Konzeptes



Quelle: EC 2006, Ch.2

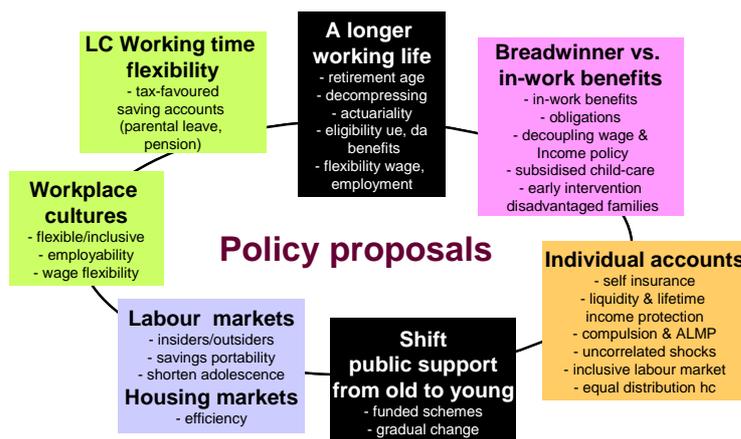
2.3.3. Life-Course-Policy

Ein zweiter Ansatz, aus dem sich Schlussfolgerungen für die Positionierung und Konzeption der Arbeitsmarktpolitik ergeben ist als 'Life-course-policy' in die Diskussion gebracht worden (vgl. Bovenberg 2007 für eine umfassende Darstellung). In diesem Ansatz wird vor allem auf den

Zusammenhang von sozialer Sicherung und Beschäftigung abgestellt und herausgearbeitet, dass die Grundstrukturen der sozialen Sicherungssysteme oft größeren indirekten Einfluss auf die Beschäftigungs- und Qualifizierungsanreize haben, als die Beschäftigungs- und Qualifizierungspolitik wettmachen kann.

Insbesondere die Pensionssysteme und die industriellen Beziehungen wirken sich auf die Verteilung von Chancen zwischen Jüngeren und Älteren aus, indem sie den Zugang für Jüngere erschweren und durch die Förderung des frühen Abganges aus Beschäftigung bei den Älteren zu Qualifikationsvernichtung führen. Ebenso wirkt sich ein System, das die soziale Sicherung auf dem männlichen Breadwinner-Modell aufbaut, derart negativ auf die Qualifikationsentwicklung und –nutzung der Frauen aus, dass diese Wirkungen durch Arbeitsmarktpolitik nicht aufgewogen werden können. Life-course-policy strebt daher einen umfassenden Policy-mix an, in dem Sozialleistungen direkt an Beschäftigung gebunden werden (in-work-benefits) und vielfältige Formen der Flexibilität steuerlich und durch individuelle Ansparkonten gefördert werden. Inwieweit diese Methoden nicht nur zu einer Schlechterstellung der ‘Insider’ sondern auch tatsächlich zu einer Besserstellung der ‘Outsider’ des Beschäftigungssystems führen können, ist weithin umstritten.

Abbildung 5: Schematische Darstellung von Life-Course-Policy



Quelle: Eigene Darstellung, zusammengestellt nach Bovenberg 2007

2.3.4. Governance und AMP: Diskurse und Realität

Mit den skizzierten Ansätzen sollen sich die Umsetzungsmechanismen wie auch der Stellenwert der AMP ändern. Marktelemente und individuelle Anreize und Verantwortlichkeiten sollen gegenüber den traditionellen bürokratischen und auf detaillierten gesetzlichen Ansprüchen beruhenden Systemen gestärkt werden. Bei den Inhalten geht es um vielfältige Umverteilungen von vorhandenen (ineffizienten) Ansprüchen zu potentiellen neuen (effizienten) Vorteilen, die auf Basis wissenschaftlicher Ergebnisse evidenz-gestützt konzipiert wurden.

Vergleicht man die verschiedenen Ansätze, so besteht eine Homologie zwischen den Vorstellungen über die Governance-Mechanismen und die inhaltlichen Vorschläge, die beispielsweise beim Umriss der 'Life-Course-Policy' ganz deutlich werden. Es geht hier um sehr umfassende und komplexe Programmatiken, deren Konzeption und Analyse viele Fragen aufwirft: Sind die Wirkungsmechanismen dieser Programmatiken empirisch noch fassbar? Sind

derartige komplexe Programmatiken überhaupt umsetzbar? Ist ihre Umsetzung überhaupt analysierbar, oder handelt es sich um derart allgemeine Konzepte, dass darunter sehr viele unterschiedliche Realisierungen fallen können?

Zu diesen Fragen gibt es sehr unterschiedliche Thesen und Analysen. Es würde diesen Text überschreiten, diese Thematik zu vertiefen, klar erscheint aber, dass die 'State-of-the-Art'-Methoden der Evaluierung von AMP für diese Ebene von Fragen nicht greifen. Wenn aber die Arbeitsmarktpolitik als systemisch in diese Konzepte eingebunden begriffen wird, muss sich dies auf ihre Wirksamkeit auswirken, und damit sind die Ansätze und Methoden der Evaluierung auch von diesen Veränderungen betroffen.

Es wurde bereits im Abschnitt 1 herausgearbeitet, dass im Bereich der Evaluierung von AMP die Kontextfaktoren essentiell sind – hier ergibt sich, dass die skizzierten Veränderungen im Bereich Governance gewissermaßen einen zweiten Ring von politischen Kontextfaktoren beisteuern, die ebenfalls zu berücksichtigen wären. Fragen in dieser Richtung werden im nächsten Abschnitt behandelt.

3. Evaluierung und Governance

Der Zusammenhang zwischen Evaluierung und Governance betrifft letztlich die Frage der Umsetzung von Evaluierungsergebnissen. Neben den methodischen und technischen Problemen der Evaluierung hat diese Frage aber eher eine untergeordnete Rolle in der einschlägigen Forschung gespielt. Es geht dabei um den Meta-Rahmen, in dem sich die Evaluierung abspielt, und Aussagen dazu sind eher anekdotisch oder normativ.

Nach der ersten 'Evaluierungseuphorie' während der Reformphase der 1960er in den USA erfolgte eine starke Ernüchterung, als begonnen wurde, Fragen der Implementation zu behandeln. Aus dieser Phase stammt das als 'Wilensky's law' bekannt gewordene Diktum, das eine substitutive Beziehung zwischen Evaluierung und Umsetzung aufstellt: „*the more evaluation, the less program development, the more demonstration projects, the less follow through*“ (Wilensky 1985, 9). Zwei weitere sehr bekannte Evaluationsforscher haben eine ähnlich radikale Position vertreten: „...*if you advocate a particular policy reform or innovation, do not press to have it tested*“ (Gary Burtless und Robert Haveman zit.n. OECD 1991, 49). In politikwissenschaftlichen Theorien und Analysen wird die Nutzung von formalisiertem Wissen im Politikprozess eher ambivalent eingeschätzt (vgl. z.B. Parsons 2005, ch. 2.8).

In der heutigen Diskussion spielt Evaluierung im Rahmen von Modellen der Qualitätssicherung eine große Rolle, wobei diese Modelle aber vorwiegend normativ die erforderlichen Voraussetzungen für Evaluierung und Nutzung aufstellen. Eine breitere über Evaluierung hinausgehende – im Prinzip ebenfalls normative – Auffassung wird in den Konzepten der 'Evidence-based policy and practice' entwickelt, die teilweise von der Politik gefordert und anscheinend auch vorangetrieben werden. Inwieweit sich hier eine Verbindung von Evaluierung und Governance abzeichnet, soll im Folgenden auf Basis der bisherigen Überlegungen diskutiert werden.

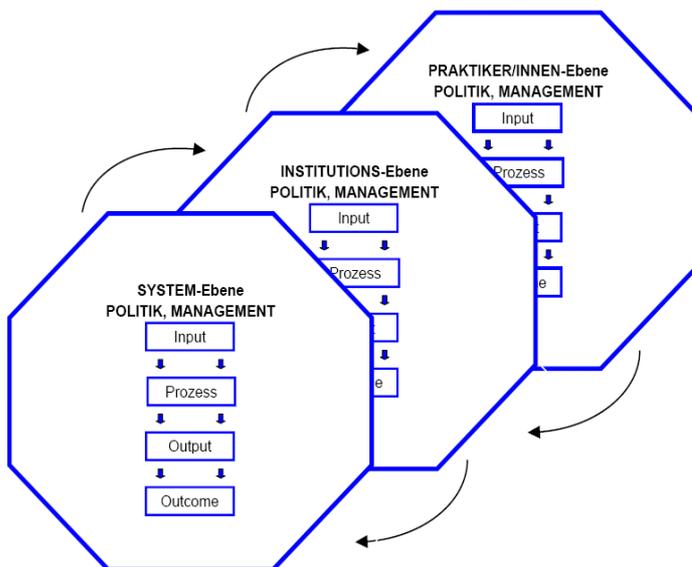
3.1. Ein normatives Modell und seine Komplikationen: Evaluierung im Qualitätszyklus

Es wurde bereits dargelegt, dass eine Umsetzung des 'State-of-the-art' die drei Kategorien von Wirkungsevaluation, Politik-Evaluation und Implementationsevaluation einschließt. Am Beispiel der 'zielorientierten Evaluation' (Schmid et al 1996) erfordert dies vier notwendige Schritte:

1. Definition der Ziele auf der Basis der Analyse von Politik- und Marktversagen
2. Monitoring der Umsetzung auf der Basis von finanziellen und physischen Indikatoren
3. Monitoring der Ergebnisse und Erfassung der Netto-Effekte durch experimentelle oder nicht-experimentelle Forschungstechniken
4. Schätzung der Netto-Kosten und des Netto-Nutzens unter Berücksichtigung der Opportunitätskosten.

Man kann die wesentlichen Dimensionen des Qualitätszyklus in einigen Schritten beschreiben, und damit die Komplexität des Unterfangens sichtbar machen, insbesondere wenn es sich um breitere Maßnahmen oder Programme handelt (vgl. Abbildungen 6-10).

Abbildung 6: Anwendungsbereich und Betrachtungsebenen des Qualitätszyklus



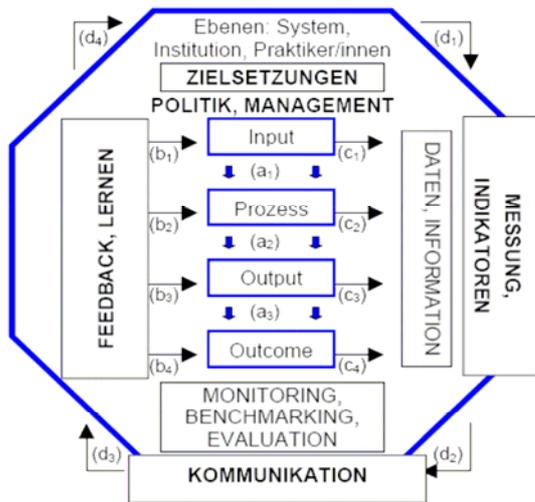
Quelle: Eigene Darstellung

Abb.6 schlüsselt die Anwendungsbereiche in doppelter Hinsicht auf, einerseits nach dem Produktionsverlauf einer Maßnahme oder Politik (Input-Prozess-Output-Outcome), andererseits nach Betrachtungsebenen (System, Institution/Organisation, Praktiker/innen). Hier kann im Prinzip eine Auswahl getroffen werden (z.B. Outputs auf Systemebene), die jedoch im Hinblick auf die Umsetzung jeweils gravierende Einschränkungen mit sich bringt.

In Abb.7 werden die wesentlichen Elemente eines Qualitätszyklus dargestellt: (1) Daten und Messung, (2) Kommunikationssysteme (Monitoring, Benchmarks, Evaluation im engeren Sinne) und (3) Rückmeldung. Evaluierung wird hier als ein spezieller Kommunikationsmechanismus betrachtet, der tiefergehend ist als die beiden alternativen Kommunikationsmechanismen Monitoring oder Benchmarking. Diese können sich wiederum mehr oder weniger umfassend auf die Stadien des Prozesses beziehen (typischerweise gibt es diese Mechanismen im traditionellen bürokratischen System für das Input-Stadium, weniger jedoch für die weiteren Stadien).

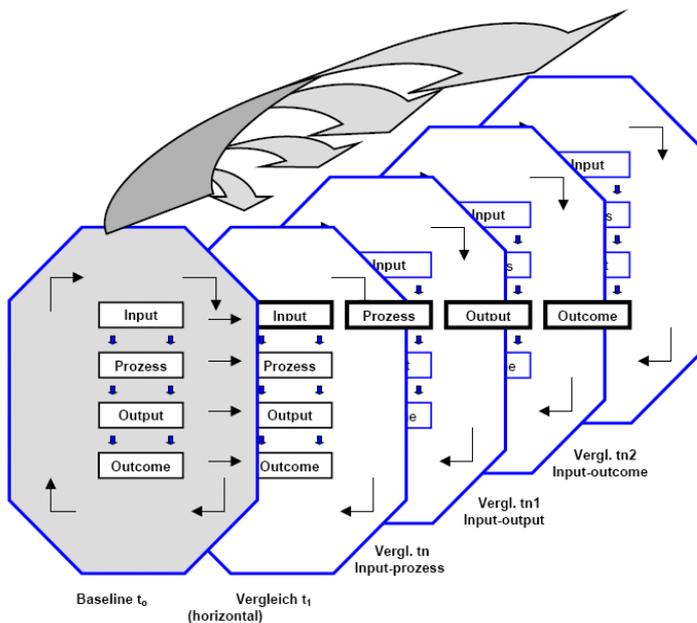
Gegenstand einer Evaluierung können im Prinzip alle in der Abb. angedeuteten Pfeile sein: (a_i) Übergänge zwischen den Stadien, (b_i) Wirkungen von Feedback auf die Stadien, (c_i) Messung und Daten in den Stadien, (d_i) Umsetzung der Elemente des Qualitätszyklus. Wesentlich für jede Beurteilung der Qualität ist die Formulierung von Zielsetzungen, deren Erreichung überprüfbar und für die Beurteilung der Qualität sinnvoll und aussagekräftig ist (Problem des 'Teaching to the Test').

Abbildung 7: Dimensionen eines vollen Qualitätszyklus (Querschnittsbetrachtung)



Quelle: eigene Darstellung

Abbildung 8: Zeitdimension und Anwendungsbereiche (Längsschnitt)

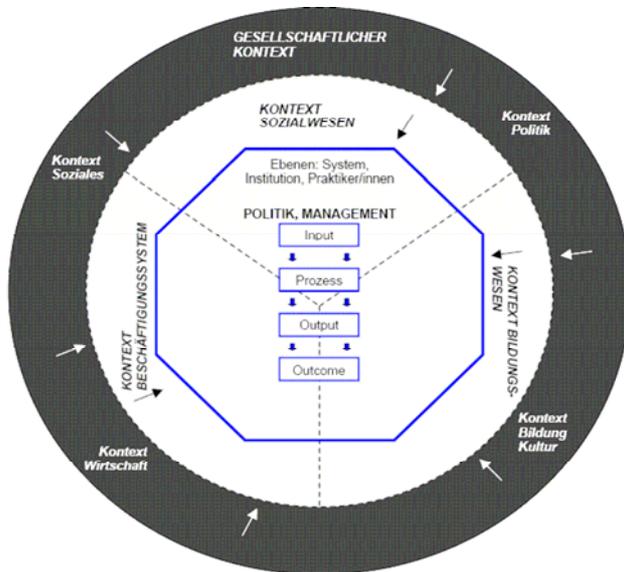


Quelle: eigene Darstellung

Was bisher im Querschnitt bzw. in einem stilisierten Produktionszyklus betrachtet wurde, spielt sich normalerweise in fortlaufenden Zyklen über mehr oder weniger lange Zeiträume ab. Zu einem fixierten Zeitpunkt (Querschnitt) wird man die verschiedenen Stadien von unterschiedlichen Perioden beobachten, für die Evaluierung eines wirklichen Produktionszyklus wird man eine Längsschnittbetrachtung benötigen (Abb.8).

Ein wesentliches Problem besteht auch darin, dass die evaluierten Vorgänge in Kontextbedingungen eingebettet sind, die über die Zeit variieren und die gesamte Maßnahme mehr oder weniger gravierend beeinflussen können. Im Bereich der Arbeitsmarktpolitik ist dies grundsätzlich etwa die wirtschaftliche Entwicklung oder auch die Fiskalpolitik, aber wie wir gesehen haben spielt auch die Sozialpolitik eine wesentliche Rolle und aufgrund der großen Bedeutung der Qualifikationen ist auch das Bildungswesen zu berücksichtigen (Abb.9).

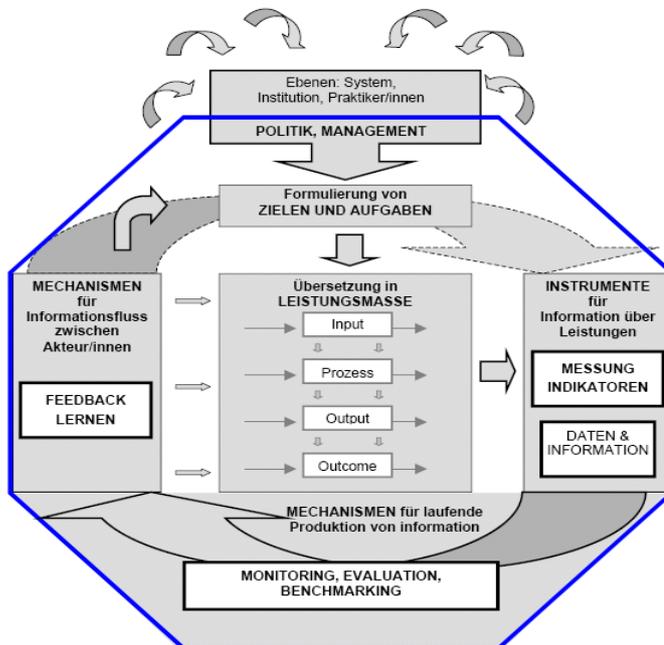
Abbildung 9: Berücksichtigung der Kontexte



Quelle: eigene Darstellung

Abb.10 bettet die Evaluierung als speziellen Mechanismus in die verschiedenen Dimensionen des Qualitätszyklus ein. Je nachdem, wie ausgebildet diese Dimensionen in einem Governance-System sind, wird Evaluierung unterschiedliche Möglichkeiten und Grenzen haben. Wo der Qualitätszyklus überhaupt nicht entwickelt ist, es keine Daten, Monitoring- und Feedbackmechanismen gibt, wird es sich um 'Pioniertätigkeiten' handeln, die überhaupt erst erste objektivierte Informationen gewinnen. Wenn der Qualitätszyklus zwar input-seitig entwickelt ist, über Ergebnisse jedoch keine Informationen vorhanden sind, können Evaluierungen die Auseinandersetzung mit den Zielen auf eine neue Ebene heben. Wenn es einen entwickelten Qualitätszyklus mit Benchmarks und Feedback gibt, wird Evaluierung mit höheren Ansprüchen konfrontiert sein, sich möglicherweise auch stärker mit konkurrierenden Situationsdefinitionen auseinandersetzen müssen, aber wichtige zusätzliche Tiefeninformationen liefern können.

Abbildung 10:



Quelle: eigene Darstellung

Die Abbildung verweist auch auf die zentrale Bedeutung der Formulierung von Zielsetzungen für den gesamten Prozess, da Qualität normalerweise in Relation zu den Zielen definiert wird. Daher hat die 'State-of-the-art'-Auffassung von Evaluierung auch die Politik- bzw. Ziel-Evaluierung als integralen Bestandteil aufgenommen.

Wenn man nun aber die Umsetzung des „State-of-the-art“ betrachtet, so zeigt sich gerade die große Schwäche im Bereich der Politikevaluierung. Nimmt man als Beispiel die großen EU-Programme (ESF, EQUAL, NAP, Reformprogramm etc.), so sind die politisch ausgehandelten Ziele im Prinzip gegeben und ihre Formulierung entspricht oft nicht den Anforderungen einer „State-of-the-art“-Evaluierung. Die Evaluierer/innen stehen in einer Sandwichposition zwischen den EU-Autoritäten, die die Europäischen Ziele vertreten, und den nationalen Regierungen, die die Programme oft als Kompromisse zwischen nationalen Praktiken und den EU-Vorgaben formulieren. Da es um die Vergabe beträchtlicher Mittel geht, kann die Evaluierung in beträchtliche Konfliktsituationen kommen, wenn diese Kompromisse nicht gut durchdacht und formuliert sind („Landesverräter“).

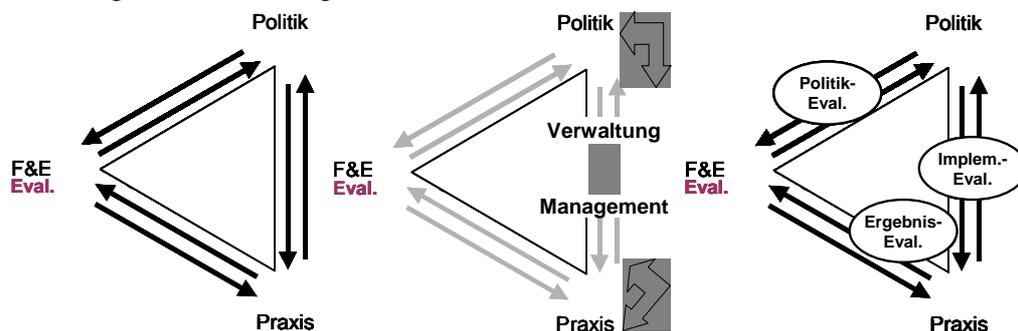
Auch der Zeithorizont entspricht nicht der Logik der Evaluierung, insbesondere ist nicht genügend Zeit für die Realisierung der Ergebnisse. 'State-of-the-art'-Ergebnisse können daher erst vorliegen, wenn die Programmperiode bereits einige Zeit vorbei ist. In der nächsten Programmperiode – so es diese gibt – stehen dann aber oft andere Aspekte im Vordergrund, und die Vorbereitung erfolgt bereits bevor die evaluierte Periode endet. Ein Lern- und Feedbackprozess ist genau genommen auf Basis der Evaluierung nicht möglich ist, es kann bestenfalls mit – mehr oder weniger (un)-ausgegorenen Zwischenergebnissen gearbeitet werden. Grundsätzlich kann man die folgenden Probleme festhalten, wenn man Evaluierung im Modell eines umfassenden Qualitätszyklus betrachtet:

- Eine ‚State-of-the-Art‘-Evaluierung wie beispielsweise die angesprochene ‚zielorientierte Evaluierung‘ erfordert die Berücksichtigung des gesamten skizzierten Qualitätszyklus, und stellt damit hohe Anforderungen an alle Beteiligten
- Wenn eine derartige Evaluierung durchgeführt wird, kommen die Evaluierer/innen in eine Position, wo sie den gesamten Prozess auf Basis objektivierten Wissens überblicken, damit entsteht auf dieser Ebene eine gewisse Konkurrenzposition zu den Vertreter/innen des Governance-Systems (an der letztere eigentlich nicht interessiert sein können)
- Wenn nur spezifische Teile oder Ausschnitte evaluiert werden, entstehen spezifische Bedingungen, indem die Evaluierer/innen einen begrenzten Ausschnitt zu Gesicht bekommen, und daher auch nur begrenzte Aussagen machen können; die Verwendbarkeit der Ergebnisse hängt letztlich von der Kenntnis seitens der Vertreter/innen des Governance-Systems über das Gesamtsystem ab, woher soll diese Kenntnis aber ohne umfassende Evaluierung kommen?
- Letztlich ergibt sich hier ein grundlegender Widerspruch zwischen den hohen und prinzipiell schwer erfüllbaren Anforderungen, die für eine ‚State-of-the-Art‘-Evaluierung gegeben sind, und den Interessen der Vertreter/innen des Governance-Systems, die eher in Richtung Begrenzung der Aufträge gehen werden, was jedoch zwar die (demagogische oder legitimatorische) Verwendbarkeit der Ergebnisse erhöht, nicht jedoch das Wissen über die in Frage stehenden Politiken – das Wissensmonopol der Auftraggeber/innen bleibt bestehen, aber inwieweit es sich um zutreffendes Wissen handelt, bleibt dennoch offen.

3.2. Wie passt Evaluierung in Governance-Systeme? ‚New public management (NPM)‘ und ‚Evidence based policy and practice (EBPP)‘

Um den Zusammenhang zwischen Evaluierung und Governance zu explizieren kann das seit längerem verwendete OECD-Dreieck zwischen Politik, Praxis und Forschung & Entwicklung (F&E) herangezogen werden. Abb.11 zeigt die Grundstruktur, wobei grundsätzlich Politik und Praxis durch das Governance-System verbunden sind, und Verwaltung bzw. Management je nach Konfiguration zwischen diesen Polen angesiedelt sind. Evaluierung kann grundsätzlich als Bestandteil von F&E gesehen werden, wobei hier wiederum je nach Ansatz stärker der Aspekt der Forschung oder Entwicklung betont werden kann.

Abbildung 11: Evaluierung im Dreieck von F&E, Politik und Praxis



Quelle: Eigene Darstellung

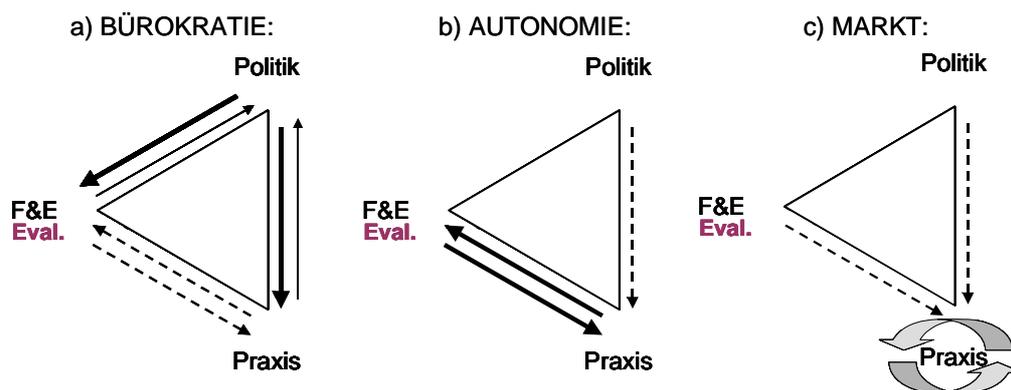
Fragt man nach den Beziehungen in diesem Dreieck, so ist erstens ersichtlich, dass zwei unterschiedliche Pfade zur Politik bzw. zur Praxis gehen. Dies drückt die unterschiedlichen Ansprüche und Interessen dieser beiden, oder bei Einbeziehung von Verwaltung bzw.

Management bis zu vier Parteien aus, die grundsätzlich unterschiedliche Ansätze erfordern werden. Zweitens ist ersichtlich, dass Evaluierung nicht nur den unterschiedlichen Parteien gegenübersteht, sondern eigentlich den Beziehungen zwischen diesen, was die Position de facto kompliziert, weil im realen Evaluierungsproblem Politik und Praxis eben meistens nicht säuberlich getrennt auftreten werden. Vielmehr wird man es mit der (gegebenenfalls auch konflikthaften) Beziehung zwischen diesen zu tun haben.

In der Bürokratie, und oft auch im NPM, wird eine typische Aufgabe sein, im Auftrag der Politik die Praxis zu evaluieren, wobei Verwaltung bzw. Management variabel auf der Seite der Auftraggeber/innen oder der zu Evaluierenden (oder schlimmer: irgendwo dazwischen) eingebunden sein können. In dem Maß, in dem die Umsetzung der Praxis autonom erfolgt, können hier mehr oder weniger große Variationen auftreten, die im Prinzip auch im Auftrag der Praxis evaluiert werden können. Dabei kann auch Selbstevaluierung eine mehr oder weniger große Rolle spielen.

Eine mehr systematische Betrachtung dieser Beziehungen kann die Position der Evaluierung mit den bereits diskutierten Typen von Governance-Modellen verknüpfen: Bürokratie, Markt und Autonomie im NPM.

Abbildung 12: Stilisierte Beziehungen im Dreieck von F&E, Politik und Praxis nach unterschiedlichen Governance-Modellen



Quelle: Eigene Darstellung

Im *bürokratischen Modell* ist die Politik der entscheidende Gatekeeper und kontrolliert F&E und Praxis, sowie auch den Fluss der Ergebnisse von Evaluierung in die Praxis. Die Verbindung zwischen Evaluierung und Praxis ist dementsprechend schwach ausgeprägt, und wenn es in diesem System Evaluierung gibt, sind Nutzungsprobleme zu erwarten. Im *Marktmodell* geht die Entwicklung vorwiegend inkrementell aufgrund der Praxiserfahrungen vor sich, der Einfluss der Politik ist schwach, und F&E wird entsprechend der Nachfrage und der Ressourcen eingesetzt werden. Es ist wenig Evaluierungsaktivität zu erwarten, wenn sie stattfindet wird sie vermutlich genutzt werden. Im *Autonomie-Modell (NPM)* ist Möglichkeiten für Nutzung von F&E durch die Praxis im Prinzip gegeben, es stellt sich aber die Frage der Verfügbarkeit von Ressourcen, was

von der Politik als Instanz der Ressourcenverteilung abhängen wird. Wenn es in diesem Modell die Mittel gibt, ist wirksame Evaluierung zu erwarten.⁷

In der österreichischen Arbeitsmarktpolitik sind diese Modelle – im Unterschied etwa zur Bildungspolitik – nicht rein ausgeprägt, sondern es besteht ein hybrides System, das auch einen beträchtlichen Wandel durchgemacht hat. In den 1980ern, zu Beginn der Evaluierungsaktivität bestand ein intaktes bürokratisches System. In den 1990ern wurden Reformen in Richtung NPM und Zielsteuerung durch die Etablierung des Arbeitsmarktservice (AMS) als einer teilweise regionalisierten aus der Verwaltung ausgegliederten Körperschaft durchgeführt, in der den Sozialpartnern beträchtlicher Einfluss zukommt. Im Zuge der EU-Integration fand auch eine beträchtliche Repolitisierung durch die Beschäftigungsstrategie und die ESF-Programme statt (wobei die Beteiligung dem AMS durchaus auch politisch auferlegt wurde). Die aktive Arbeitsmarktpolitik wird im Bereich der Ausbildungsmaßnahmen in Form von Interventionen in den Weiterbildungsmarkt bzw. in die Erwachsenenbildungsinstitutionen und im Bereich der Förderung von Betrieben durch Interventionen in den Markt abgewickelt. In diesem hybriden System ist durchaus rege Evaluierungsaktivität zu erwarten und findet auch statt, wobei aufgrund der komplexen Strukturen für die Nutzung der Ergebnisse keine definitive Voraussage formuliert werden kann – gemischte Nutzung kann erwartet werden.

3.2.1 'New Public Management': Begrenzte Aufgabenzuweisung

Die traditionelle Einordnung von Evaluierung besteht darin, dass die übergeordnete Instanz die untergeordnete Instanz evaluiert oder evaluieren lässt. In diesem Zusammenhang ist das Diktum des Überganges von der Input- zur Output-Steuerung verbreitet. In der Bürokratie regulieren die Vorschriften, im neuen Governance-System entstehen Prinzipal-Agent-Verhältnisse mit Informationsvorsprung der Agents. Die Prinzipale versuchen, dies in den Griff zu bekommen. Im NPM bestehen zwei unterschiedliche Grundorientierungen gegenüber den Agent/innen: Anreize ('*make the managers manage*') vs. Motivation ('*let the managers manage*'; Kettl 1998, 13-14).

- Anreize werden gesetzt durch Mechanismen der rigiden radikalen Zielsteuerung, die der Logik des Marktmodells folgt, und Quasi-Märkte verwendet. Die Instrumente sind eng definierte Ziele, Monitoring und Leistungsindikatoren, und klare berechenbare Belohnungen/Sanktionen bei Erreichung/Nicht-Erreichung der Ziele (dieses Modell wurde am klarsten in Neuseeland verfolgt)
- Motivation setzt auf Empowerment der Agent/innen, glaubt an deren grundsätzlich positive Absicht die Ziele zu verwirklichen, und erwartet eine bessere Zielerreichung von der Erhöhung des Handlungsspielraumes und der Verbesserung der Bedingungen. In diesem Fall ist Evaluierung ein wichtiges Instrument zur Feststellung der Erfolgsbedingungen (dieses Modell wird beispielsweise Australien zugeschrieben).

Evaluierung passt in verschiedener Weise in das System von Prinzipal und Agent:

- indem sie Informationen über die 'Agents' an die 'PrinzipalInnen' liefert, und damit deren Handlungsspielraum verbessern kann,
- indem sie unter den Bedingungen ihrer verstärkten Handlungsautonomie den 'Agents' bei der Verbesserung ihrer Leistungen bzw. Aufgabenerfüllung helfen kann.

⁷ Im österreichischen Bildungswesen sind diese drei Governance-Modelle nebeneinander vertreten (Schule: Bürokratie; Hochschule: Autonomie; Berufs- und Erwachsenenbildung: Markt), und die Erfahrungen mit der Evaluierungsaktivität sind mit den Erwartungen in die verschiedenen Konstellationen konsistent.

Wenn man zwischen den beiden Ausformungen von NPM unterscheidet, so ist Evaluierung (etwa in Verbindung mit Zielvereinbarungen als Außenkontrolle) besser für das 'Empowerment'-Modell geeignet. In der rigiden Zielsteuerung sind Monitoring-Systeme besser geeignet als Evaluierungen. Die Bedingungen werden in dieser Struktur von den AuftraggeberInnen bestimmt, und das gesamte Setting sowie insbesondere die Zielsetzungen werden nicht unbedingt Gegenstand der Evaluierung sein.

In der österreichischen Arbeitsmarktpolitik ist diese Konkurrenz zwischen internen Monitoring-Systemen und den Evaluierungspraktiken bereits deutlich geworden. Der Rechnungshof – dessen Strategieformulierungen über weite Strecken im Einklang mit den Anforderungen an die 4th Generation It. Guba/Lincoln 1987 sind, und einem qualitativen Review-Verfahren entsprechen – hat beispielsweise in einer Prüfung des Arbeitsmarktservice die Ausgaben für externe Evaluierung den Ausgaben für den Aufbau des Monitoring-Systems gegenübergestellt, und eine klare Präferenz für die Verwendung des Monitoring ausgesprochen.⁸ Das AMS hat demgegenüber die begrenzte Aussagekraft des Monitoring ins Treffen geführt.

3.2.2. 'Evidence-based Policy and Practice': Evaluierung als Umsetzungsforschung?

Seit einiger Zeit haben sich Forderungen nach, wie auch Absichtserklärungen für stärkere 'Evidenzbasierung (EB)' der Politik weit verbreitet (besonders entwickelte Diskussionen gibt es im Bereich der Bildungspolitik, vgl. OECD 2007, EC 2007). Was bedeutet dies für die Evaluierung? Hier kann man zwei Fragen näher beleuchten, erstens die Frage nach der Beziehung von Forschung und Evaluierung, und zweitens die früher aufgeworfene Frage nach der Evaluierung bzw. Beurteilung von Politik.

Hier ist zunächst die Unterscheidung zwischen EB-Policy und EB-Practice zu unterstreichen. Im Bereich der Policy gibt es eine von OECD und EU geteilte Definition derzufolge EB-Policy "can be defined as the conscientious and explicit use of current best evidence in making decisions and choosing between policy options" (EC 2007, 3; vgl. OECD 2007, 16). Im Bereich der EB-Practice ist eine geteilte Definition weniger leicht zu finden. Hier stehen die innerwissenschaftlichen methodologischen Standards im Zentrum der Diskussionen, die stark disziplinspezifisch geführt werden. Der wichtigste Punkt auf der Ebene der Forschung betrifft Fragen der Identifikation kausaler Effekte vs. dem Verharren in deskriptiver Analyse. Thomas Cook and Stephen Gorard haben in OECD (2007) diese Frage anhand eines Modells von Gorard/Taylor (2004) aufgespannt. Es werden sieben Phasen eines vollen Forschungszyklus unterschieden (vgl. Abb.13), der noch einmal in 2 Subzyklen zerfällt, der erste Zyklus betrifft deskriptive und konzeptive Forschung, der zweite die kausalanalytische Forschung. In diesem Bereich stehen sich sehr unterschiedliche Positionen gegenüber, von der mehr oder weniger ausgeprägten Leugnung (der Erkennbarkeit) von Kausalität in sozialen Zusammenhängen (z.B. durch Spielarten der Systemtheorie oder des Konstruktivismus), über das multi-method-Paradigma bis zu den Vertreter/innen der Randomisierung als Ideal-Methode.

⁸ "Der RH wies ... darauf hin, dass das AMS aufgrund des Fehlens eines operativen Finanzcontrollings die Kosten-Nutzen-Relation der arbeitsmarktpolitischen Förderungen nicht messen und daher auch nicht steuern konnte. ... Das AMS startete im Jänner 2001 ein IT-Projekt ... Damit war es möglich, die Amortisation der eingesetzten finanziellen Mittel bei den durch die Förderungsmaßnahmen erreichten Beschäftigungsaufnahmen darzustellen. ... Die gesamten Ausgaben für das Installieren des IT-Projekts betrugen 1,22 Mill EUR. Das AMS vergab vergleichsweise allein in den Jahren 2000 und 2001 Evaluierungsstudien im Ausmaß von 1,11 Mill EUR. ... Der RH empfahl, in Hinkunft das zur Verfügung stehende Datenmaterial des internen IT-Systems zur Steuerung des Einsatzes von Förderungsmitteln entsprechend zu nutzen." (Der Rechnungshof 2007, 163-4)

Als siebente abschließende Phase fungiert die Umsetzungsforschung. Hier könnte man den Stellenwert der Evaluierung in diesem Gesamtmodell ansetzen und man könnte diese als einen wesentlichen Bestandteil der Erreichung von EB-Practice ansehen. Es wäre hier auch möglich, das Selbstverständnis der Evaluierung zwischen Forschung und professioneller Praxis variieren zu lassen, soweit sich die Aktivitäten als Forschung verstehen, stellt sich die Frage der Einordnung in die Forschungsdiskurse und –bewertungen. In manchen Disziplinen gibt es bereits lebendige Diskussionen darüber welche Kriterien für jene Forschungstätigkeiten sinnvoll sind, die auf den Transfer von Wissen in die Praxis abstellen (vgl. z.B. Spiel 2009, Kanning et al. 2007).

Abbildung 13: Voller Forschungszyklus für die Generierung von Evidenz

Subzyklus I:

- Phase 1: Evidence synthesis
- Phase 2: Development of idea or artifact
- Phase 3: Feasibility study

Subzyklus II:

- Phase 4: Prototyping and trialling
- Phase 5: Field studies and design stage
- Phase 6: Definitive testing

Umsetzungsforschung

- Phase 7: Dissemination impact and monitoring

Quelle: OECD 2007, nach Gorard and Taylor (2004).

3.2.3. Thesen und Erfahrungen zur Beziehung von Evaluierung und Governance

Versucht man aus der Entwicklung im Bereich von Steuerung und Governance nachfrageseitige Konsequenzen für die Evaluierung abzuleiten, so können diese in folgenden Punkten zusammengefasst werden:

- Die *Qualitätssicherung* verdrängt tendenziell Evaluierung durch Verstärkung von Monitoring-Systemen. Evaluierung kann jedoch wichtige ergänzende Beiträge in Form von vertiefenden Analysen leisten und auch vor allem durch das Zusammenspiel von Selbstevaluierung mit externer Evaluierung die Qualitätssicherung mit Leben und Bedeutung erfüllen. Dies gilt insbesondere in der Empowerment-version von NPM.
- Die Etablierung des *NPM-Governance Modells* schafft jedoch Widersprüche zwischen den verschiedenen Elementen der 'State-of-the-Art'-Evaluierung, indem sich der Schwerpunkt auf die Ergebnisevaluierung der Praxis-Einheiten (AgentInnen) im Auftrag von Politik und/oder Management (PrinzipalInnen) konzentriert, und Fragen der Politik(ziel)-Evaluierung und der Implementationsanalyse tendenziell abgekoppelt werden. Es könnte sich eine neue Struktur zwischen den Akteur/innen und den Evaluator/innen etablieren, wenn Evaluierungspraktiken zu einem integralen Element der Praxis werden, bzw. wenn sich eine allgemeine Instanz zur Politikevaluierung (evtl. im Parlament) etabliert.
- Bei den inhaltlichen Belangen der AMP wurde durch den breiteren Ansatz der *Jobs Study* die Rolle der AMP relativiert. Die Fragestellungen für Evaluierung wurden stark erweitert, so dass das 'State-of-the-Art'-Instrumentarium in seiner Reichweite die Fragestellungen nicht mehr

ausreichend erfassen kann. Grundlegende ökonomische Kompetenz wird damit für die Evaluierung erforderlich, wenn die Wirkung der Politik erfasst werden soll. Die Praxis hat darauf mit der Methodik von Peer Reviews reagiert, die jedoch unzureichend ausgearbeitet ist.

- Der 'Flexicurity'-Ansatz verstärkt die Rolle der aktiven AMP wiederum deutlich und die Frage nach dem Einsatz und den Wirkungen der auszuweitenden Maßnahmen wird als solche wieder relevant. Es sind aber auch breitere Analyse-Ansätze nötig, um Wirkungen des Gesamtansatzes und des Zusammenspiels zwischen den Eckpfeilern der Flexibilisierung, der sozialen Sicherung und der aktiven AMP zu erfassen. Angesichts der Debatten um das Profil von 'Flexicurity' vs. 'Flexification' ist auch die Politik-Evaluierung wichtig.
- Der Ansatz der 'Life-Course-Policy' relativiert die Bedeutung der AMP noch stärker als der allgemeine Ansatz der Jobs Study, und bietet mit dem Umbau der sozialen Sicherungssysteme eher eine Alternative zur AMP. Auch in diesem Ansatz ist die Politik-Evaluierung, und aufgrund der Vielzahl an Ansatzpunkten und Maßnahmen, die oft eher auf dem Reißbrett konzipiert sind und wenig empirische Erfahrungen umfassen, auch die Implementationsanalyse zu betonen.

Betrachtet man die Erfahrungen von der Entwicklung des Evaluierungsangebotes her, so ist im Bereich der Evaluierung von Arbeitsmarktpolitik die kausalanalytische Forschung stark verankert, und stellt insbesondere in der Ergebnisevaluierung bereits den Standard dar. In diesem Bereich ist die Generierung von „Evidenz“ für Politikentscheidungen stark ausgeprägt und dominiert den Aspekt der Generierung von Evidenzen für die Praxis. Aber, wie schon gesagt, stellt sich die Frage, inwieweit sich die Politik tatsächlich an diesen Evidenzen orientiert, bzw. orientieren kann. Die Schwierigkeiten hier werden in einem Diktum eines der politisch einflussreichsten Forscher auf den Punkt gebracht: *"There is nothing a government hates more than to be well informed; for it makes the process of arriving at decisions much more complicated and difficult"* (J.M. Keynes, zit. nach EC 2007, 8).

Möglicherweise wären hier stärkere Differenzierungen zwischen der Wirkungsevaluierung und den beiden anderen Elementen des „State-of-the-Art“-Modells angebracht. Aus den skizzierten Beobachtungen insbesondere der Dynamik im Evaluationsfeld würde sich die stärkere Ausdifferenzierung der Implementationsanalyse im Zusammenspiel mit Selbstevaluierung als Aufgabe ergeben. Das österreichische AMS orientiert sich jedoch eher am Anreiz-Modell des NPM als am Empowerment-Modell. Die folgende Interviewpassage indiziert eine starke Dominanz der Kennziffern und ihre Wirkung in Richtung von 'creaming off'-Effekten gegenüber dem Betreuungsbedarf: „Manager: '...alle probieren ... die Pflegeleichten unterzubringen, weil von der Statistik her und von den Kennziffern her ist es ja ziemlich wurst, ist es ein pflegeleichter oder ein schwieriger Fall.' - Interviewer: 'Das weiß dann keiner' - Manager: 'Das weiß dann kein Mensch, und die Zahlen sind hervorragend, und das ist halt wichtig...'“ (Zilian/Lasnigg/Wroblewski 1999, 144). Dieses Beispiel repräsentiert die alte Debatte im Zusammenhang mit Performance-Management aufgrund von formalen Zielen. Im Zusammenhang mit den Hartz-Reformen werden Entwicklungen in Richtung von NPM-Governance ganz klar verfolgt, und manche Evaluierer/innen schließen aus den Ergebnissen, dass dies zu wenig deutlich der Fall wäre, um die Wirksamkeit nachhaltig zu erhöhen. Die Empfehlungen können in Richtung verstärkter Aufmerksamkeit für Implementationsfragen verstanden werden.

Eine weitere Ausdifferenzierung betrifft die Frage der Politikanalyse. Es wurde angedeutet, dass diese bisher eher als Nebenprodukt oder als eine Art von 'spinn-off' aus der Wirkungsevaluierung resultiert. Im Bereich der Ziele wurde aber bereits im OECD 1991

(Review) festgestellt, dass soziale Zielsetzungen gegenüber den wirtschaftlichen Zielen in der Evaluierung vernachlässigt werden (s.o.). Diese Debatte hat sich ausdifferenziert und betrifft insbesondere den Aspekt der positiven bzw. negativen Anreize für Beschäftigungsaufnahmen in der Einkommenssicherung. Eine nähere Diskussion dieser Fragen würde den Rahmen sprengen, aber wir haben gesehen, dass eben die Wirkung von Arbeitsmarktpolitik letztlich von der Nachfrage nach Beschäftigung abhängig ist.

Die Jobs Study geht davon aus, dass diese eben mit allen Mitteln gefördert werden muss – aber, was soll geschehen, wenn bzw. solange dies nicht gelingt? Als alternative Ansätze werden hier Formen des Grundeinkommens und auch von direkter Schaffung von öffentlicher oder non-profit Beschäftigung diskutiert, wobei jedoch bei der Frage nach den Wirkungen stärker soziale als ökonomische Kriterien verankert werden müssen. Diese Formen von Evaluierung spielen jedoch im 'State-of-the-Art' eine sehr geringe Rolle, insbesondere sind sie nicht in einen gemeinsamen Diskussionszusammenhang integriert.

Ein Beispiel für eine umfassende multi-method Evaluierung eines regionalen Schulungssystems in Österreich (Zilian/Lassnigg/Wroblewski 1999) hat zu diesen Fragen teilweise bemerkenswerte Ergebnisse zutage gefördert, die an sich immer noch aktuell sind, aber nicht weiterverfolgt wurden. Es wird herausgearbeitet, dass eine dezidiert beschäftigungsorientierte und geschäftsmäßige, das Kundenmodell forcierende Arbeitsmarktpolitik unweigerlich mit fundamentalen Widersprüchen konfrontiert ist, die zwangsläufig zu einem Unterlaufen dieser Strategie – und auch an ihr orientierter Evaluierungen – führen. Aufgrund der Ergebnisse dieser Studie ist die Qualifizierungspolitik als wichtigste AMP-Maßnahme in einer Art zweiseitiger Symbiose eingebunden: (a) die Unternehmen flexibilisieren weniger durch prekäre Beschäftigung als vielmehr durch 'Temporary layoffs', also quasi durch 'Zwischenlagerung' regulär Beschäftigter in der Arbeitslosenversicherung (der Anteil von 'Flexibilitätsarbeitslosen', die zu ihrem vorhergehenden Betrieb zurückkehren liegt nach den detaillierten Befunden dieser Studie bei bis zu 50% und übersteigt bei Weitem das allgemein bekannte hohe Niveau der österreichischen Saison-Arbeitslosigkeit). Diese Gruppe ist im Prinzip eher nicht interessiert an 'aktiven Maßnahmen' und wird auch aufgrund der Einstellungszusagen nicht in diese vermittelt – sie stellt die Erfolgskomponente der Vermittlungsaktivitäten dar, ohne eine Vermittlungsaktivität zu erfordern. (b) für das Schulungssystem bleibt die Gruppe der 'Sucharbeitslosen', die ihre Beschäftigung ohne Einstellungsusage verloren hat, in der die 'marginalisierten' Langzeitarbeitslosen (die in nicht unbeträchtlichem Ausmaß durch Einweisung in Schulungsaktivitäten statistisch in Kurzzeitarbeitslose verwandelt werden) einen beträchtlichen Anteil ausmachen. Diese Gruppe ist ebenfalls in die Flexibilitätsprozesse eingebunden, indem sie am stärksten vom Auf und Ab des Konjunkturzyklus betroffen ist. Die zweite Symbiose wird hier dergestalt gesehen, dass die Ausbildung quasi ein 'Überlaufbecken des Beschäftigungspools' bildet, indem Geschulte und Schulende, die vorwiegend mit prekärer Beschäftigung arbeiten, sich finden. In diesem System mischen sich die offiziellen Ziele der Vermittlung und Beschäftigungsfähigkeit mit de-facto- Funktionen der sozialen Absicherung. Die Vielgleisigkeit von offiziellen und inoffiziellen Zielen, garniert mit Diskriminierung und Konfliktspielen, unterläuft einerseits die vorgegebene 'geschäftsmäßige' Strategie der AMP, erfüllt aber andererseits auch implizit starke soziale Funktionen. Seit den 1990ern ist ein stärkeres 'Profiling' vorangetrieben worden, was auch in Deutschland als wesentliches Lösungselement effizienter AMP gesehen wird. Die Frage bleibt jedoch, inwieweit dieses zu stärkerer Diskriminierung unter den Geförderten führt, ohne – bei nach wie vor mangelnden Beschäftigungsmöglichkeiten für die marginalisierten Sucharbeitslosen – die Probleme zu lösen.

4. Zusammenfassung

Im vorliegenden Beitrag wird versucht, das Verhältnis von Evaluierung und Steuerung am Beispiel der Arbeitsmarktpolitik (AMP) zu reflektieren. Es wird weithin davon ausgegangen, dass Evaluierung in einem mit verstärkter ergebnisorientierter Zielsteuerung arbeitenden `neuen Steuerungskonzept´ als korrekatives Element eingepasst werden soll und kann, in der Erwartung, damit zu einer evidenzbasierten Umsetzung und Weiterentwicklung von Praxis und Politik zu kommen. Eine nähere Betrachtung dieser Beziehungen zeigt jedoch, dass dies nicht so reibungslos funktioniert wie es zunächst scheinen mag.

Das zentrale Argument besteht darin, dass das Versprechen einer Verbesserung von Politik und Praxis durch Evaluierung Anforderungen impliziert, die in der Realität meistens nicht erfüllt werden können – dann entstehen jedoch uneingelöste Versprechen einer immer weiter in die Zukunft verschobenen „Rationalisierung“ von Politik, die de facto oft zu einer vorwiegend legitimatorischen Funktion von Evaluierung führen. Diese Anforderungen beziehen sich auf eine Reihe von Ingredienzien:

- die Methodik der Evaluierung
- die Strukturen und Mechanismen von Praxis und Politik
- die Kompetenzen der Beteiligten und
- die Reichweite der zu evaluierenden Probleme.

Wichtig für das Verständnis der Möglichkeiten von Evaluierung ist ihre Einordnung in die Praxis der Wissensproduktion. Wenn man Evaluierung als *systematisierte Bewertungspraxis* versteht, so knüpft diese an die täglich unter allen Beteiligten vor sich gehenden Bewertungsprozesse und die daraus resultierenden Bewertungen an, und muss diesen etwas hinzufügen. Dieses `Etwas´ kann Verschiedenes sein: eine bessere Informationsbasis, eine bessere methodische Basis, einer neue Sichtweise der Vorgänge, ein besser begründetes Verständnis der Vorgänge, etc. Unter Gesichtspunkten der Wissensproduktion steht Evaluierung also irgendwo zwischen dem impliziten und informellen Wissen der Beteiligten und dem formalisierten Wissen, das im System von Forschung und Entwicklung (F&E) produziert wird. Implizit gibt es im Rahmen der Evaluierungsdiskurse eine große Variationsbreite von Einordnungen in diesem Spektrum, zwischen wissenschaftlicher Forschung und professioneller Praxis.

Die Anforderungen an wirksame Evaluierung sind in diesem Spektrum sehr hoch anzusetzen: Es muss ein Mehrwert gegenüber dem *informellem Wissen* der Beteiligten produziert werden, und dieses zusätzliche Wissen muss diesen nachvollziehbar und in ihre Praxis integrierbar sein. Wenn die Evaluierer/innen auch im vorhandenen substantiellen formalisierten *Wissen aus den F&E-Quellen* auf dem aktuellen Stand sein sollten, so wird die Aufgabe umso anspruchsvoller. Es fragt sich angesichts dieser Einordnung, ob nicht eine derart große Diskrepanz zwischen dem Status (und der Remuneration) der Evaluierung und den skizzierten Anforderungen besteht, dass diese de facto schwer zu schließen ist. Jedenfalls muss man die Frage stellen, wie mit dieser Diskrepanz produktiv umgegangen werden kann. Die Antwort scheint am ehesten darin zu bestehen, die Evaluierungsaufgaben in Form von Kompetenzen zur Selbstevaluierung an die Beteiligten zurückzuspielen.

Im Zusammenhang mit dem Konzept von `Evidence-based policy and practice (EBPP)´ und der Frage nach Innovation und Lernen in der Politik sind Diskussionen über die Generierung und Verbreitung von Wissen für Politik und Praxis entstanden, die auch für die Standortbestimmung von Evaluierung von Bedeutung erscheinen. Eine wichtige Rolle spielt dabei das Konzept von `User-oriented basic research´ (Stokes 1997) sowie die Idee aus der Innovationsforschung, dass die traditionelle Abfolge der Wissensformen von Grundlagenforschung, angewandter Forschung

und praktischer Entwicklung durch ein interaktives System von vielfältigen Quellen der Wissensgenerierung abgelöst worden sei, die durch entsprechende Vernetzungen der Akteur/innen fruchtbar gemacht werden müssen. In diesen Diskursen kann man zwei Themenbereiche ausmachen, die in unterschiedliche Richtungen gehen, und für Evaluierung wichtig sind: einerseits wird für die Lösung von praktischen Problemen 'Transdisziplinarität' eingefordert, d.h. die *Zusammenarbeit von Wissenschaft und Praxis* bei der Wissensgenerierung, andererseits gibt es innerhalb der Wissenschaft eine Bewegung in Richtung *verstärkter rigoroser Forschung* mit kausalanalytischer Orientierung und akademischer Qualitätssicherung, die sich auch auf praktische Probleme erstrecken soll.

Im Zusammenspiel von Evaluierung und Steuerung ergibt sich als wesentliche Frage die nach der Funktionsfähigkeit der Steuerungsmechanismen. Hier stellt sich eine Art hermeneutischer Zirkel: Die Umsetzung und Wirksamkeit von Maßnahmen, Programmen oder Politiken hängt von der Funktionsfähigkeit der Steuerung ab, die daher als Ausgangsbedingung bekannt sein müsste. Wenn sich Evaluierung in ein System einpasst, ohne diese Frage zu stellen, dann wird sie zu keinen relevanten Ergebnissen kommen. Die Grundstruktur von NPM ist jedoch darauf angelegt, gerade diese Frage zu vermeiden, indem die politischen Fragen von den (technischen) Fragen der Umsetzung getrennt werden.

Innerhalb der Evaluierung besteht eine Tendenz die zu evaluierenden Gegenstände zu begrenzen, um sich methodisch und ressourcenmäßig nicht zu überfordern, die Problemlagen erfordern jedoch eher eine Expansion des Auftrages, um relevante Ergebnisse liefern zu können. Im Bereich der AMP zeigt sich eine eminente Entwicklung im Bereich der Methodik der Wirkungsevaluierung, ohne jedoch zu den materialen Fragen der Beschäftigungspolitik Antworten geben zu können. Dennoch hat die Evaluierungsaktivität wesentlich zu kumulativen Wissensbeständen beigetragen.

Auf der anderen Seite haben jene Aktivitäten, die stärker zur Beantwortung der Fragen der Beschäftigungspolitik beitragen können, Politikevaluierung durch 'Peer Reviews', sowie der Aufbau von Monitoring-Systemen und Leistungsindikatoren viel weniger Aufmerksamkeit der methodischen Entwicklung erfahren. Entscheidend um diese Probleme zu lösen erscheint die Stärkung der Selbstevaluierung und die Verbreitung von Kompetenzen und Instrumenten der systematischen Bewertung unter den Praktiker/innen, Manager/innen, Administrator/innen und nicht zuletzt insbesondere den Politiker/innen.

Literatur

- Abu Sharkh, Miriam (2008): Are there optimal global configurations of labour market flexibility and security? Tackling the "flexicurity" oxymoron. Employment Working Paper No. 15 2008. Genf: ILO. <http://www.ilo.org/public/english/employment/download/wpaper/wp15.pdf>
- Bergemann, A. and Van den Berg, G. J. (2007) 'Active labor market policy effects for women in Europe - a survey', IFAU Working Paper 2007:3, IFAU, Uppsala.
- Betcherman, G / Olivas, K / Dar A (2004): Impacts of Active Labor Market Programs: New Evidence from Evaluations with Particular Attention to Developing and Transition Countries. Social Protection Discussion Paper Series No. 0402. The World Bank. Internet: <http://siteresources.worldbank.org/SOCIALPROTECTION/Resources/SP-Discussion-papers/Labor-Market-DP/0402.pdf>
- Bovenberg, Ary Lans (2007): The life-course perspective and social policies: an overview of the issues. DELSA/ELSA/WP1(2007)8. Paris: OECD.

- Breslau, Daniel (1998) *In Search of the Unequivocal: The Political Economy of Measurement in U.S. Labor Market Policy*. Westport Conn.: Praeger.
- Calmfors, Lars (1994): *Active Labour Market Policy and Unemployment - A Framework for the Analysis of Crucial Design Features*. In: *OECD Economic Studies* 22, S.7-47.
- Campbell, Donald T./M. Jean Russo (1999), *Social Experimentation*, Thousand Oaks/London/New Delhi: Sage.
- Cook, Beth (2008): *Active Labour Market Policies in the Neo-Liberal Era*. Working Paper No. 08-03. Centre of full employment and equity CofFEE, University of Newcastle. Australia.
<http://e1.newcastle.edu.au/coffee/pubs/wp/2008/08-03.pdf>
- Der Rechnungshof (2007) *Prüfbericht Arbeitsmarktservice*. Berichte Bund 2007/4. Wien, 151-176.
http://www.rechnungshof.gv.at/fileadmin/downloads/Teilberichte/Bund/Bund_2007_04/Bund_2007_04_Bd2_12.pdf
- developmental psychology./ *European Journal of Developmental Psychology/ European Journal of Developmental Psychology*, Volume 6, Issue 1 January 2009 , pages 11 - 33.
- EC (2006): *Employment in Europe 2006*. Ch.2 Flexibility and security in the EU labour markets. Luxembourg: Office for Official Publications of the European Communities, S.75-118
http://ec.europa.eu/employment_social/employment_analysis/eie/eie2006_chap2_en.pdf
- EC (2007): *Towards more knowledge-based policy and practice in education and training*. Commission Staff Working Document. 28.08.2007, SEC(2007) 1098. Brussels
http://ec.europa.eu/education/policies/2010/doc/sec1098_en.pdf
- Eichhorst, Werner / Zimmermann, Klaus F. (2007) *Dann waren's nur noch vier... Wie viele (und welche) Maßnahmen der aktiven Arbeitsmarktpolitik brauchen wir noch? Eine Bilanz nach der Evaluation der Hartz-Reformen*. IZA DP No. 2605. Forschungsinstitut zur Zukunft der Arbeit. Bonn.
- Eichhorst, Werner / Zimmermann, Klaus F. (2007): *Dann waren's nur noch vier... Wie viele (und welche) Maßnahmen der aktiven Arbeitsmarktpolitik brauchen wir noch? Eine Bilanz nach der Evaluation der Hartz-Reformen*. IZA DP No. 2605. Forschungsinstitut zur Zukunft der Arbeit. Bonn.
- EQUAL Evaluierung
- Gorard, S./ Taylor, C. (2004): *Combining Methods in Educational and Social Research*. London: Open University Press.
- Guba Egon G. / Lincoln Yvonna S. (1987): *The countenances of fourth-generation evaluation: Description, judgment, and negotiation*. In: *Evaluation Studies Review Annual*, 11, S.70-88.
- Heckman, James J. (1999): *Doing It Right: Job Training and Education*. In: *Public Interest*, 135 (Spring), S. 86-107.
- Heckman, James J. (2006): *Skill Formation and the Economics of Investing in Disadvantaged Children*. In: *Science*, 312, S. 1900-1902.
- Heckman, James J./Lalonde, Robert J./Smith, Jeffrey A. (1999): *The Economics and Econometrics of Active Labor Market Programs*. In O. Ashenfelter & D. Card (Hg.): *Handbook of Labor Economics*, Volume 3A, Elsevier, Amsterdam, S. 1865-2097.
- Hofer, Helmut/ Weber, Andrea (2006) *Active Labor Market Policy in Austria : Practice and Evaluation Results*. *Vierteljahrshefte zur Wirtschaftsforschung*, 75 (3), 155-167
- Kanning et al (Uwe Peter Kanning, Lutz von Rosenstiel, Heinz Schuler, Franz Petermann, Friedemann Nerdinger, Bernad Batinic, Lutz Hornke, Martin Kersting, Reinhold Jäger, Rüdiger Manfred Trimpop, Christiane Spiel, Christian Korunka, Erich Kirchler, Werner Sarges, Manfred Bornewasser) (2007) *Angewandte Psychologie im Spannungsfeld zwischen*

- Grundlagenforschung und Praxis – Plädoyer für mehr Pluralismus. *Psychologische Rundschau*, 58(4), 238-248.
- Keep, Ewart (2007): Skills and Economic and Social Justice. SKOPE Issues paper 16. Oxford & Cardiff Universities. <http://www.skope.ox.ac.uk/WorkingPapers/Issuespaper16.pdf>
- Kettl, Donald F. (1998): Reinventing Government: A fifth-year report card. Washington D.C.: Brookings.
- Kluge, Jochen (2006): The Effectiveness of European Active Labor Market Policy. IZA DP No. 2018. Internet: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=892341#
- Lassnigg, L (1997): Evaluation: Aufdecken, zudecken, oder was sonst ... In: Zilian H.G. / Flecker J. (Hg.): Pathologien und Paradoxien der Arbeitswelt. Schriftenreihe Bd 2 Forum Sozialforschung. Wien: Forum Sozialforschung, 227-262. Internet: <http://www.equi.at/dateien/graz-denkw-evaluation96.pdf>
- Lassnigg, Lorenz (2004), National Social Dialogue on Employment Policies: Case Study Austria, IHS, Vienna. <http://www.equi.at/dateien/ilo-rilsa-report-jan-04final.pdf>
- Lassnigg, Lorenz (2007): Qualität und Wirkungen von arbeitsmarktpolitischen Trainingsmaßnahmen. In: Arbeiterkammer Wien (Hg.): Weiterbildung und Lebensbegleitendes Lernen. Vergleichende Analysen und Strategievorschläge für Österreich. Materialien für Wirtschaft und Gesellschaft Nr. 102, Wien. S. 69-93. <http://wien.arbeiterkammer.at/pictures/d60/MaterialienWuG1021.pdf>
- Lechner, Michael / Wiehler, Stephan (2007) Kids or Courses? Gender Differences in the Effects of Active Labor Market Policies. IZA DP No. 2740. Forschungsinstitut zur Zukunft der Arbeit
- Lechner, Michael / Miquel, Ruth / Werner, Stephan / Wiehler, Stephan (2007) Mikroökonomische Evaluierung der Instrumente der Aktiven Arbeitsmarktpolitik in Österreich. Projektbericht des Schweizer Instituts für Aussenwirtschaft und Angewandte Wirtschaftsforschung (SIAW). St. Gallen.
- Martin John P. (1998): What Works Among Active Labour Market Policies: Evidence from OECD Countries. Reserve Bank of Australia 1998 Conference “Unemployment and the Australian Labour Market”, 9-10 June 1998. <http://www.rba.gov.au/PublicationsandResearch/Conferences/1998/Martin.pdf>
- Mitchell, William (2007): Why public sector job creation should be fashionable. Working Paper No. 07-03. Centre of full employment and equity CoffEE, University of Newcastle. Australia. <http://e1.newcastle.edu.au/coffee/pubs/wp/2007/07-03.pdf>
- OECD (1991): Evaluating Labour Market and Social Programmes. The State of a Complex Art, Paris: OECD.
- OECD (1995): Local Responses to Industrial Restructuring in Austria. Paris: OECD
- OECD (2007): Evidence in Education. Linking research and policy. Paris: OECD.
- Parsons, Wayne (2005): Public policy. An introduction to the theory and practice of policy analysis. Cheltenham: Edward Elgar.
- Patton, Michael Q. (2002): A Vision of Evaluation that Strengthens Democracy. In: *Evaluation*, 8 (1), S. 125–139.
- Rhodes, Rod A.W. (1996) “The new governance: governing without government”, in *Political Studies*, Vol. 44, page 652. 652 - 667
- Schmid, G./ O’Reilly J./ Schömann K. (Hg.) (1996): International Handbook of Labour Market Policy and Evaluation. Cheltenham: Elgar.
- Schneider, Hilmar / Eichhorst, Werner / Zimmermann, Klaus F. (2006): Konzentration statt Verzettelung: Die deutsche Arbeitsmarktpolitik am Scheideweg. IZA DP No. 2055. Forschungsinstitut zur Zukunft der Arbeit. Bonn.

- Schneider, Hilmar / Eichhorst, Werner / Zimmermann, Klaus F. (2006) Konzentration statt Verzettlung: Die deutsche Arbeitsmarktpolitik am Scheideweg. IZA DP No. 2055. Forschungsinstitut zur Zukunft der Arbeit. Bonn.
- Sengenberger, Werner (1987): Struktur und Funktionsweise von Arbeitsmärkten. Die Bundesrepublik Deutschland im internationalen Vergleich. Frankfurt am Main: Campus.
- Shadish, W R Jr. / Epstein, R (1987), Patterns of program evaluation practice among members of the Evaluation Research Society and Evaluation Network. In: Evaluation Review, 11 (5) (October), S.555-590.
- Smith, Jeffrey (2000): Evaluation aktiver Arbeitsmarktpolitik: Erfahrungen aus Nordamerika. In: MittAB, 33, S. 345-356.
- Spiel, C. (2009). Evidence-based practice: A challenge for European developmental psychology. European Journal of Developmental Psychology/ European Journal of Developmental Psychology, Volume 6, Issue 1 January 2009 , pages 11 - 33.
- Stokes, Donald E. (1997): Pasteur's Quadrant. Basic Science and Technological Innovation. Washington, D.C.: Brookings Institution Press.
- Synthesis (2006) Gesicherte Beschäftigungsintegration? Arbeitslosigkeit und AMS-Maßnahmen in einer Mehrjahresperspektive 2000–2004. Forschungsbericht für das AMS-Österreich. Wien. <http://www.ams-forschungsnetzwerk.at/deutsch/publikationen/BibShow.asp?id=1728&sid=657919537&look=2&jahr=2006>
- Watts, Martin (2008): Labour market policy convergence? A case study of the OECD Jobs Study. Working Paper No. 08-07. Centre of full employment and equity CoffEE, University of Newcastle. Australia. <http://e1.newcastle.edu.au/coffee/pubs/wp/2008/08-07.pdf>
- Wilensky, Harold L. (1985): Nothing Fails Like Success: the Evaluation-Research Industry and Labour Market Policy In: Industrial Relations, 24 (Winter), S. 1-19.
- Williams, Janice E. (1989): A numerically developed taxonomy of evaluation theory and practice. In: Evaluation Review, 13 (1) (February), S.18-31
- Wroblewski, Angela (2009) Evaluation von arbeitsmarktpolitischen Programmen und Maßnahmen in Österreich. Thomas Widmer/Wolfgang Beywl/Carlo Fabian (Hrsg.) Evaluation. Ein systematisches Handbuch. VS Verlag für Sozialwissenschaften, Wiesbaden, S. 129-136.
- Zeitlin, Jonathan (2005): Social Europe and experimentalist governance: towards a new constitutional compromise? European Governance Papers (EUROGOV) No.C-05-04 [Internet: <http://connex-network.org/eurogov/pdf/egp-connex-C-05-04.pdf>
- Zeitlin, Jonathan (2006): When do policy innovations spread? Lessons for advocates of lesson-drawing. Harvard Law Review, 119, S.1466-1487.
- Zilian, Hans-Georg /Lassnigg, Lorenz /Wroblewski, Angela (1999): Arbeitslosenschulung in der flexibilisierten Wirtschaft: Am Beispiel der Evaluierung und der Analyse eines Schulungssystems. München: Hampp Verlag.

ANHANG

Übersicht 1: Hauptmerkmale von Evaluierung und Steuerung im Bereich Arbeitsmarktpolitik

<p>Stilisierte Meilensteine der AMP-Evaluierung</p> <ul style="list-style-type: none"> - Beginn der AMP-Evaluierung in den 1960ern in den USA, etwas später auch in Kanada, in Europa verstärkt seit 2000 - Anfang 1990er <i>internationaler 'State-of-the-Art'</i> (OECD 1991; Schmid/O'Reilly/Schömann 1996), umfasst 3 Dimensionen (Politik-, Wirkungs- und Implementationsevaluierung), sowie Qualitätsanforderungen für Wirkungsevaluierung - Evaluierungspraxis begrenzt auf <i>Wirkungsevaluierung von 'aktiver' AMP</i> als Teil der Arbeitsmarktpolitik, starke spezialisierte und sophistische Methodenentwicklung und verallgemeinerte stilisierte Ergebnisse um 2000 aufgrund Reviews und Meta-Evaluierungen (Heckman/LaLonde/Smith 1999; Betcherman/Olivas/Dar 2004; Kluge 2006) - Ab Mitte der 1990er <i>Erweiterung der Politikparadigmen und Relativierung der 'aktiven' AMP</i> durch generische 'Design Faktoren' (Calmfors 1994), sowie OECD Jobs Study (1994) und EU-Beschäftigungsstrategie (1998) - Späte 1990er Politik-Evaluierung verstärkt durch neue Formen der Evaluierung wie ExpertInnenrating (EU Gemeinsamer Beschäftigungsbericht), Monitoring (EU Strukturindikatoren) und 'Peer Reviews' auf politischer Ebene (Watts 2008 für OECD und EU: www.mutual-learning-employment.net) - Jedoch wenig Aufmerksamkeit für methodische Entwicklung dieser neuen Evaluierungsverfahren (teilweise politische Aushandlung von 'Ergebnissen') - Monitoring und Leistungsindikatoren als Teile von Implementationsevaluierung ab 2000 stärker entwickelt, Datenbasis für Wirkungsevaluierung; vertiefende Analysen zur Implementation weiterhin schwach 	<p>Stilisierte Meilensteine im Bereich der Steuerung</p> <ul style="list-style-type: none"> - Paradigmatischer Hintergrund: <i>Marktversagen</i>, Politik als Korrektiv, 'Krieg gegen die Armut' auf prosperierendem Hintergrund - Entwicklung der AMP-Evaluierung eng verbunden mit Groß-Projekten <i>sozialer Experimentation</i> in den USA ab den 1960ern; dies kann als spezielle Steuerungsform gesehen werden, die sowohl für die Evaluierung als auch für die Politik und Praxis starke Vorgaben macht (Campbell/Russo 1999) - 1980er und 1990er Evolution von Arbeitsmarkt- und Beschäftigungspolitik parallel zur (neoliberalen) Kritik am Wohlfahrtsstaat, Promotion von Evaluierung als Hilfe zur Verbesserung der Politik durch internationale Organisationen (v.a. OECD), Steuerungsformen zunächst nicht im Fokus - Parallel tritt <i>Staatsversagen</i> in den Vordergrund, Marktmechanismen und 'Business-Praktiken' als Korrektiv, 'New Public Management', Reform der Arbeitsmarktverwaltungen und -institutionen (Watts 2008) - Ergebnisse statt Ressourcen: rigorosere Evaluierungsformen in USA (Revival von experimentellem Design in 1980ern), Projekt- und Programmorientierte Steuerungsformen in EU - Entstehung von '<i>New Governance</i>' mit der Beschäftigungsstrategie in Europa (Rhodes 1996), weiche Steuerung durch gemeinsame Ziele und Informationsaustausch über Zielerreichung und 'good practice', Möglichkeit für <i>Politiklernen</i> durch Verbreitung neuer Heuristiken (z.B. Employability), Entwicklung neuer Informationsbasen (z.B. EU-Strukturindikatoren) und neuer Reflexionsformen (z.B. Expert/inneneinschätzung im Gemeinsamen Beschäftigungsbericht; Zeitlin 2005) - Modelle und Methoden der Qualitätssicherung/entwicklung auf Praxisebene als Steuerungsmittel (Monitoring, Leistungsindikatoren, soziale Mechanismen)
<p>Zusammenwirken von Evaluierung und Steuerung in AMP</p> <ul style="list-style-type: none"> - Die Entwicklungen von Evaluierung und Steuerung können als sich <i>gegenseitig beeinflussende Kontexte</i> gesehen werden - Das '<i>State-of-the-Art'-Konzept für Evaluierung von AMP</i> (OECD 1991) umfasste implizit ein Steuerungskonzept, in dem Politik und Ziele, Implementation und Wirkungen gemeinsam evaluiert werden sollten; dies wurde nicht realisiert, Evaluierung konzentrierte sich auf Wirkungen - Steuerungsformen wurden auf politischer Ebene erst mit dem Fokus auf <i>Staatsversagen</i> und '<i>New Public Management</i>' problematisiert, Evaluierung kann in diesem 'neuen Steuerungsmodell' je nach Ausformung unterschiedliche Positionen einnehmen, als Haupttendenz wird sich die Evaluierung der Praxis im Auftrag der Politik ergeben; das Feld differenziert sich, aber der ursprünglich umfassende Auftrag wird eher fragmentiert (im anreiz-orientierten NPM liegt der Schwerpunkt auf Monitoring und Leistungsindikatoren, im empowerment-orientierten NPM kann Evaluierung eine größere Rolle spielen) - teilweise (USA, Kanada) wird Evaluierung zu einem Erfordernis (siehe z.B. www.evidencebasedpolicy.org), auch in <i>EU-Strukturpolitik</i> (z.B. ESF) wird Evaluierungserfordernis etabliert, aber randständig gegenüber nationalen Politiken (vgl. auch 'MEANS Collection' als Ressource: ec.europa.eu/regional_policy/sources/docgener/evaluation/means_en.htm) - Mit der Erweiterung der beschäftigungspolitischen Paradigmen sind auch die politikrelevanten Evaluierungsfragen weit über die ursprünglichen Fragen der Wirkung von 'aktiver' AMP hinausgewachsen, die mit dem entwickelten Instrumentarium nicht mehr erfasst werden können, in den neuen Methoden von 'Peer Review' besteht großer Bedarf in Methodenentwicklung, Fragen der Grundlagenforschung gewinnen an Gewicht - Es können aus der Interaktion von Evaluierung und Steuerung unterschiedliche Szenarien abgeleitet werden: (a) Fragmentierung und Spezialisierung: NPM scheint die Komplexität der Politik und Umsetzungsprozesse einfacher zu strukturieren und zu zerlegen, es entstehen vielfältige Spezialaufgaben für Evaluierung, die den verschiedenen Beteiligten (Detail)-Informationen liefern können, die Gesamtheit kommt jedoch aus dem Blick; (b) aufgrund der Zerlegung der Politik- und Umsetzungsprozesse entsteht ein Bedarf nach Übersicht über die jeweiligen Gesamtprozesse bzw. -systeme, die ebenfalls zum Gegenstand von Evaluierung werden können; (c) ein wesentliches Element von NPM ist die angestrebte Diversität der Praxis der 'Leistungserfüllung an der Front', die ebenfalls Gegenstand der Evaluierung werden kann (wobei die Frage wichtig erscheint: in welchem Auftrag?). - Das Zusammenwirken von 'externer' und 'interner' Evaluierung, das in Peer Review Verfahren notwendigerweise gegeben sein muss, könnte als eine wesentliche Entwicklungsfrage gesehen werden, damit wird auch Selbstevaluierung und damit auch die Verbreitung der Kompetenz zur Evaluierung in Politik und Praxis zu einem wesentlichen Erfordernis. Die Evaluierung würde in diesem Fall neben der Funktion der Wissensproduktion auch eine wesentliche Rolle in der Kommunikation über die vorhandenen Informations und Wissensquellen, sowohl die selbst generierten, als auch die in Monitoring-Systemen und F&E verfügbaren, zukommen. Dies wäre eine Entwicklung in der Richtung, wie sie Michael Patton als '<i>Democratic Evaluation</i>' vorschwebt (Patton 2002). 	

Quelle: Eigene Zusammenstellung

Übersicht 2: Evaluierung von AMP in Österreich

Internationale Papers/Publicationen (www.IZA.com)

- Hofer, Helmut/ Weber, Andrea (2006) Active Labor Market Policy in Austria : Practice and Evaluation Results. Vierteljahrshefte zur Wirtschaftsforschung, 75 (3), 155-167
IZA-DP 1075: Andrea Weber, Helmut Hofer, Are Job Search Programs a Promising Tool? A Microeconomic Evaluation for Austria (2004)
IZA-DP 1076: Andrea Weber Helmut Hofer, Employment Effects of Early Interventions on Job Search Programs (2004)
IZA-DP 2740: Michael Lechner Stephan Wiehler, Kids or Courses? Gender Differences in the Effects of Active Labor Market Policies (2007)

Nationale durch AMS veröffentlichte Evaluierungen (www.ams-forschungsnetzwerk.at)

- | | | |
|------|---------------|--|
| 2008 | AMS info 122 | Evaluierung der JASG-Lehrgänge in Niederösterreich |
| | AMS info 121 | Evaluierung der Angebote 2007 des AMS Oberösterreich zur Berufsqualifizierung von MigrantInnen |
| 2007 | AMS info 99 | Evaluierung der Wiener JASG-Lehrgänge |
| | AMS report 55 | Aktive Arbeitsmarktpolitik im Brennpunkt X: Evaluierung der Wiener JASG-Lehrgänge |
| 2006 | AMS info 81 | Evaluierung des Unternehmensgründungsprogrammes (UGP) des AMS |
| 2003 | AMS info 57 | Evaluation der arbeitsmarktpolitischen Maßnahmen des Territorialen Beschäftigungspaktes Wien für 1999 |
| | AMS report 36 | Aktive Arbeitsmarktpolitik im Brennpunkt IX: Arbeitsmarktpolitische Wirkungen des TBP Wien |
| 2002 | AMS info 53 | Evaluation der stiftungsähnlichen Maßnahme COMEBACK des AMS Wien |
| | AMS info 48 | Evaluierung der Maßnahme Karenz plus des abzwien |
| | AMS info 47 | Evaluierung ausgewählter Bildungsmaßnahmen – AMIMBOS |
| | AMS report 31 | Aktive Arbeitsmarktpolitik im Brennpunkt VIII: Aktivierung als Prävention vor Langzeitarbeitslosigkeit. Evaluierung des Programmes Jobcoaching 2000 des AMS Wien |
| | AMS report 30 | Aktive Arbeitsmarktpolitik im Brennpunkt VII: Evaluierung von Kursen im EDV- und kaufmännischen Bereich |
| | AMS report 29 | Aktive Arbeitsmarktpolitik im Brennpunkt VI: Evaluierung der Maßnahme karenzplus des abzwien |
| | AMS report 26 | Aktive Arbeitsmarktpolitik im Brennpunkt V: Chance für Wiedereinsteigerinnen mit längeren Berufsunterbrechungen |
| 2001 | AMS info 46 | Evaluierung von EDV- und kaufmännischen Kursen für das Burgenland |
| | AMS info 40 | Evaluierung von Qualifizierungsmaßnahmen im Rahmen des Territorialen Beschäftigungspaktes 1999 |
| | AMS info 39 | Evaluierung des sozialökonomischen Betriebes PA-RE |
| | AMS info 36 | Evaluierung der Lehrstellenförderung in Wien |
| | AMS info 35 | Evaluation ausgewählter Maßnahmen des abzwien |
| | AMS report 23 | Aktive Arbeitsmarktpolitik im Brennpunkt IV: Evaluierung der Lehrstellenförderung in Wien |
| | AMS report 22 | Aktive Arbeitsmarktpolitik im Brennpunkt III: Evaluation ausgewählter Maßnahmen des abzwien |
| | AMS report 21 | Aktive Arbeitsmarktpolitik im Brennpunkt II: Evaluation der stiftungsähnlichen Maßnahme DYNAMO für Langzeitarbeitslose |
| 2000 | AMS info 33 | Evaluierung der stiftungsähnlichen Maßnahme DYNAMO für Langzeitarbeitslose |
| | AMS info 30 | Evaluierung Sozialökonomischer Betriebe |
| 1999 | AMS info 26 | Evaluierung von Trainingsmaßnahmen des AMS |
| | AMS info 25 | Zukunftsmodell Qualifizierungsverbund: Evaluierung des QV Triestingtal |
| | AMS info 23 | Evaluation der Initiative für Wiedereinsteigerinnen des AMS |
| 1997 | AMS info 15 | Ergebnisse der Evaluierung der Betrieblichen Eingliederungsbeihilfe (BEB) 1993 in Wien und Niederösterreich |
| | AMS info 1 | Evaluierung von Instrumenten der experimentellen Arbeitsmarktpolitik in Österreich |

Evaluierung Europäischer Sozialfonds 2000-2006: Ziel 3 – Österreich

- Hedwig Lutz, Helmut Mahringer, Andrea Pöschl (Koord.), 2005, "Aktualisierung der Halbzeitbewertung. Programmbezogene Zusammenschau",
http://www.wifo.ac.at/www/jsp/index.jsp?fid=23923&id=26492&typeid=8&display_mode=2
- Hedwig Lutz, Helmut Mahringer, Andrea Pöschl (Koord.), 2005, "Aktualisierung der Halbzeitbewertung. Schwerpunktanalysen",
http://www.wifo.ac.at/www/jsp/index.jsp?fid=23923&id=26491&typeid=8&display_mode=2
- Hedwig Lutz, Helmut Mahringer, Andrea Pöschl (Koord.), 2005, "Evaluierung der österreichischen Arbeitsmarktförderung 2000-2003",
http://www.wifo.ac.at/www/jsp/index.jsp?fid=23923&id=28498&typeid=8&display_mode=2
- Martin Unger, Angela Wroblewski (EQUIIHS), 2004, "ESF-Ziel 3 aus der Sicht der Maßnahmenträger",
http://www.wifo.ac.at/www/jsp/index.jsp?fid=23923&id=25660&typeid=8&display_mode=2
- Ferdinand Lechner, Nicole Nemecek, Andreas Riesenfelder (L&R Sozialforschung), Andrea Pöschl (WIFO), 2005, "Programmschwerpunkt 4 aus der Sicht der teilnehmenden Beschäftigten und Betriebe"
http://www.wifo.ac.at/www/jsp/index.jsp?fid=23923&id=25659&typeid=8&display_mode=2
- Ferdinand Lechner, Nicole Nemecek, Andreas Riesenfelder (L&R Sozialforschung), Andrea Pöschl (WIFO), 2004, "Programmschwerpunkt 1 aus der Sicht der Teilnehmerinnen und Teilnehmer",
http://www.wifo.ac.at/www/jsp/index.jsp?fid=23923&id=25658&typeid=8&display_mode=2
- Hedwig Lutz, Helmut Mahringer (Koord.), 2004, "Halbzeitbewertung",
http://www.wifo.ac.at/www/jsp/index.jsp?fid=23923&id=25462&typeid=8&display_mode=2
- Peter Huber (WIFO), 2003, "Schwerpunkt 6: Territoriale Beschäftigungspakte"
[http://www.wifo.ac.at/www/servlet/wwa.upload.DownloadServlet/bdoc/S_2003_ESF_SCHWERPUNKT6_25461\\$.PDF](http://www.wifo.ac.at/www/servlet/wwa.upload.DownloadServlet/bdoc/S_2003_ESF_SCHWERPUNKT6_25461$.PDF)
- Lorenz Lassnigg, Elisabeth Scheibelhofer, Mario Steiner, Peter Steiner (EQUIIHS), 2003 "Schwerpunkt 3: Lebenslanges Lernen und Förderung des Beschäftigungspotentials in Forschung, Wissenschaft und Technologie",
http://www.wifo.ac.at/www/jsp/index.jsp?fid=23923&id=25460&typeid=8&display_mode=2