

Zum Verhältnis von Steuerung und Evaluation Am Beispiel der Arbeitsmarktpolitik

Lorenz Lassnigg
(lassnigg@ihs.ac.at; www.equi.at)

Beitrag zur
“**EB-Kommission**”
19.9.2008, Wien

Agenda

- **Evaluierung** von Arbeitsmarktpolitik: Ansätze, Ergebnisse, Probleme
- **Governance** von Arbeitsmarktpolitik: Entwicklung, Besonderheiten im Vergleich zu anderen Feldern
- Das **Verhältnis von E x G**: New Public Management, Evidence-Based Policy/Practice
- Zusammenfassung, Schlussfolgerungen: Kann man aus der Erfahrung im Bereich Arbeitsmarktpolitik lernen?

Agenda

- **Evaluierung** von Arbeitsmarktpolitik: Ansätze, Ergebnisse, Probleme

Evaluierung von Arbeitsmarktpolitik: Ein stark entwickeltes Feld

- in **Österreich** Beginn von AMP-Evaluierung in den frühen 1980ern, als sonst von Evaluierung noch keine Rede war
- bereits **jahrzehntelange Erfahrung** in den USA, geprägt von Wirtschaftswissenschaft und Ökonometrie
- **internationale Community**, stark beeinflusst von internationalen Organisationen (v.a. OECD, aber auch IIO)
- gibt einen „**body of knowledge**“ sowohl von Methoden als auch von Ergebnissen, auch mehrere Meta-Studien
- **EU** hat Evaluierungsaktivität verstärkt, aber auf einem niedrigen Niveau
- ... **daraus folgt bei weitem nicht, dass Wissen und Erfahrungen politisch und praktisch genutzt werden**

Evaluierungsgegenstand: Aktive Arbeitsmarktpolitik

- Gegenstand wurde zunehmend strukturiert und kodifiziert
 - OECD-Taxonomie der Maßnahmen der AMP
 - Aktive Programme**
 1. Öffentlicher 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. Frührentierung aus arbeitsmarktpolitischen Gründen

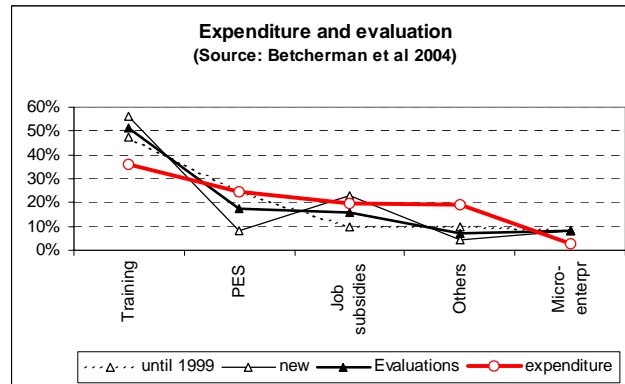
Quelle: OECD, Employment Outlook 1993

Evaluierung Ansätze und Methoden

- Ausgangspunkt der Evaluierungen: Ja/Nein-Wirkungsevaluation
 - Identifikation von "Netto"-Wirkungen
 - "Counterfactual": Experiment oder Quasi-Experiment
 - Kosten-Nutzen-Relation
- Historisch: US-Programme gegen die Armut in den 1960ern
 - In den 1980ern große methodisch sehr anspruchsvolle Evaluierungen
 - in manchen Programmen Vergleiche von Experiment und Quasi-Experiment
- OECD Reviews der Ländererfahrungen um 1990: „**State-of-the-art**“
 - Kodifizierung von Ansätzen: jedenfalls Counterfactual empfohlen
 - Drei Kategorien: **Wirkungs-, Politik-, und Implementationsevaluation** (darunter Monitoring, Performance Indicators, Implementation i.e.S.)
- Aufbau einer internationalen Datenbasis zur AMP
 - Makrostudien, Gleichgewichtseffekte
 - Calmfors: „**Crucial design features**“

Evaluierung Studien nach Maßnahmetypen

- Metastudie von Betcherman et al. 2004, 109 Evaluierungsstudien mit mindestens Counterfactual



Evaluierung „State-of-the-art“

- Gleichbedeutender Einsatz der drei Kategorien gefordert
 - Wirkungsevaluation
 - Politik-Evaluation
 - Implementationsanalyse

"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)

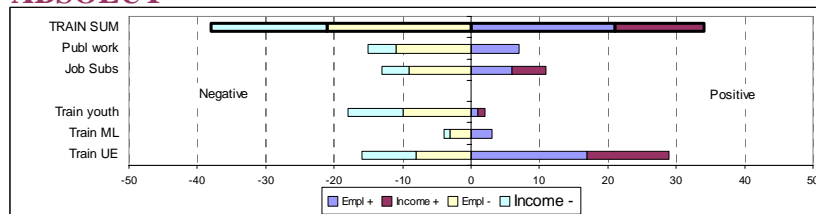
- Problem: Bedingungen für Politik von Evaluierung, die der Logik der Politik nicht entsprechen: „Naives Politik-Konzept“
 - Formulierung von Zielen; Einbau von Evaluierung in Politik-Maßnahmen „Naives Politik-Konzept“

Evaluierung „Crucial design features“

- Aus vorhandenen Ergebnissen und ökonomischen Makroanalysen:
 - finanzielle Kompensation in „aktiven Maßnahmen“ soll die Versicherungsleistungen/Sozialleistungen nicht übersteigen
 - gezielte Maßnahmen für Personengruppen und angestrebte Wirkungen
 - ausgewogenes Portfolio von Maßnahmen (sinkende Skalenerträge)
 - „passive“ Maßnahmen und „aktive“ Maßnahmen als Gesamtsystem, kein zyklischer Wechsel > Geburt der „Aktivierung“ (Calmfors 1994)
- Probleme hervorgehoben
 - Verdrängung und Substitution durch Maßnahmen
 - „kosmetische Wirkungen“ durch Umdefinition von Arbeitslosen in MaßnahmeteilnehmerInnen > Vorschlag von bereiniger AL-Rate

Stilisierte Ergebnisse nach MN-Typen Beschäftigungs-, Einkommenseffekte

ABSOLUT



RELATIV

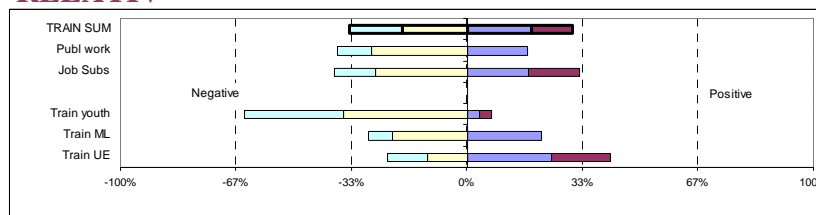


Table 6.2 Effectiveness of European ALMP: Quantitative Analysis, Specification 2

(a) Type of program and target group: Marginal Effect t-ratio

Ergebnisse einer EU-weiten Metaevaluation (Kluve 2006, 18)

(c) Institutional context on the labor market:

- Multivariate Meta-Evaluierung, 137 Studien, Beschäftigungseffekte
- Kontrolliert auch für Hintergrundvariablen:
 - Studiendesign
 - institutionelle Variablen (v.a. Regulierung Beschäftigung)
 - wirtschaftliche Lage
 - Länder
- Im Prinzip ähnliche Ergebnisse
 - Jugendliche sehr ungünstig
 - Beschäftigungsförderung im privaten Sektor und Sanktionen positiv
 - Hintergrundvariablen meistens insignifikant
 - Länder meistens insignifikant

Marginal effects printed in *italics* indicate marginal significance (10%-level), marginal effects printed in **boldface** indicate statistical significance (5%-level), and marginal effects printed in **boldface and italics** indicate high significance (1%-level). The underlying standard errors adjust for clustering by study.

Table 6.2 Effectiveness of European ALMP: Quantitative Analysis, Specification 2

(a) Type of program and target group: Marginal Effect t-ratio

(a) Type of program and target group:	Marginal Effect	t-ratio
Direct employment program	-0.338 **	-2.33
Private sector incentive scheme	0.309 **	2.34
Services and Sanctions	<i>0.346 *</i>	1.70
Young workers	-0.519 ***	-3.90
(b) Study design and time period:		
Experimental design	<i>-0.462 *</i>	-1.93
Program implemented in the 1990s	-0.211	-1.46
(c) Institutional context on the labor market:		
Index for dismissal protection regulation	-0.326	-1.64
Index for fixed-term contracts regulation	-0.166	-1.40
Index for temporary work regulation	0.085	1.43
Gross replacement rate	0.004	0.34
(d) Macroeconomic background:		
Unemployment rate	0.013	0.38
ALMP expenditure (% of GDP)	0.036	0.15
GDP growth	-0.030	-0.60
(e) Country dummies:		
	REF: Schweden	
Austria	0.299	0.69
Denmark	-0.308	-0.59
France	0.481	1.37
Germany	0.226	0.84
Ireland	0.367	1.04
Netherland	-0.087	-0.18
Norway	0.257	0.72
United Kingdom	-0.062	-0.09
Switzerland	-0.422	-0.79
Finland	<i>0.469 *</i>	1.71
Small country	0.256	0.57

Number of observations = 137. – Pseudo R² = 0.246.

Notes: The dependent variable is an indicator (1/0) variable, reflecting a positive estimate of the program effect. Table entries document the marginal effect (evaluated at the sample mean) in the corresponding probit regression, i.e. the difference in the predicted probability for achieving a positive treatment effect which arises from a marginal change in a continuous explanatory factor (such as the GDP growth rate) or which arises from changing an indicator among the explanatory factors (such as the indicator for an experimental study design) from 0 to 1. T-ratios of the marginal effects are reported in the third column. Marginal effects printed in *italics* indicate marginal significance (10%-level), marginal effects printed in **boldface** indicate statistical significance (5%-level), and marginal effects printed in **boldface and italics** indicate high significance (1%-level). The underlying standard errors adjust for clustering by study.

Probit-Regression mit abh.Variable: positiver Beschäftigungseffekt = 1

Diskussion

- Wenn man die große Variation der Ergebnisse der Ausbildung betrachtet, woher kommen die unterschiedlichen Ergebnisse?
 - Methodische Probleme (Konstruktion Counterfactual, Modellierung)
 - Datenprobleme (kleine Ns, Zeitraum)
 - Erfassung der Ausbildungsmaßnahmen als "black box", keine Unterscheidung unterschiedlicher Maßnahmentypen
 - Heterogenität der Effekte für verschiedene Gruppen
 - Betrachtungszeitraum (unterschiedliche Effekte kurz- und längerfristig)
 - Bei MN-Vergleich spielt der „Lock-in“-Effekt eine wesentliche Rolle
- Drei Gruppen von Herausforderungen:
 - Schwache Daten, methodische Probleme und enge Evaluierungskriterien
 - Hohe Kosten, schwache Intensität und große Unterschiede zwischen Maßnahmen
 - Geringer Interventionsgrad, isolierter und punktueller Charakter der Maßnahmen im Lebenslauf

Agenda

- **Evaluierung** von Arbeitsmarktpolitik: Ansätze, Ergebnisse, Probleme
- **Governance** von Arbeitsmarktpolitik: Entwicklung, Besonderheiten im Vergleich zu anderen Feldern

Governance von Arbeitsmarktpolitik: Zunehmende politische Integration?

- Arbeitsmarktpolitik ist ursprünglich ein neuer Politikbereich:
 - einerseits Versicherung und Vermittlung (Verwaltung)
 - andererseits Förderung („aktive Arbeitsmarktpolitik“)
- AM-Förderung war immer eine flexible „zusätzliche“ Aktivität
 - deren Wert in Frage gestellt wird
 - die stark entscheidungsabhängig ist
- Durchführung in Form von Maßnahmen und Programmen
 - kann gemacht werden oder nicht, politisch umstritten
 - ist relativ klein im Vergleich zu anderen Maßnahmen
- Neue Situation mit EU-Beschäftigungsstrategie
 - Spezielle Governance-Methode entwickelt: „Open Method of Coordination“:
Freiwillige und weiche Regulierung ohne gesetzliche Basis
- Zunehmend gesamtes Regulierungsregime im Zentrum:
„Flexicurity“ und „Life course“

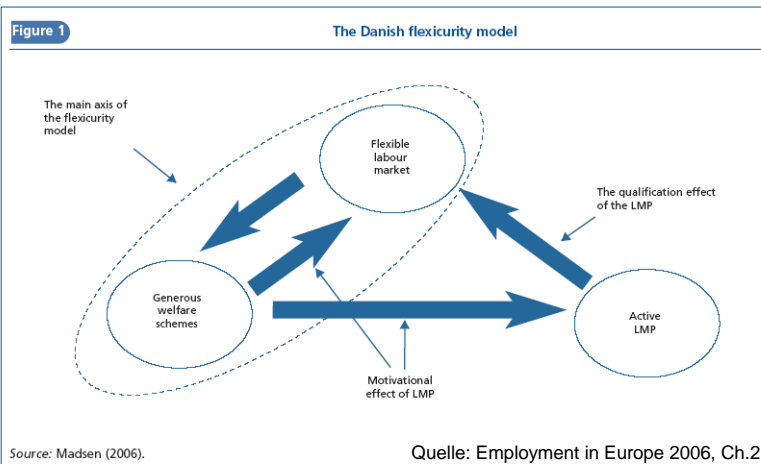
Politische Konflikte Wohlfahrtsstaat vs. Neoliberalismus

- Arbeitsmarktpolitik als Konkurrenz zur Wirtschaftspolitik
 - traditionelle „Nachfragesteuerung“ unmöglich wegen Maastricht
- Modell der Programmbudgetierung als Vorläufer von NPM
 - McNamara als technokratisches „Steuerungskonzept“
- Ausgliederung des Arbeitsmarktservice aus öffentlicher Verwaltung und (teilweise) Privatisierung
 - In Österreich besteht ein System der Zielsteuerung der Arbeitsmarktpolitik mittels Indikatoren

„OMC“ Paradigma des Governancebegriffs

- Angewendet erstmals in der EU Beschäftigungspolitik
 - gegenseitige Verpflichtung ohne gesetzliche Sanktionsbasis
 - Leitlinien und Ziele
 - Berichterstattung und „Evaluierung“
- Paradigmatischer Fall für „Policy Learning“ (J.Zeitlin)
 - Heuristic (new concepts)
 - Capacity building (new information base)
 - Maieutic (new mechanisms for reflection)
- „Evaluierung“ dem Namen nach eingebaut, aber weder Konzepte noch Praxis entsprechen dem „State-of-the-Art“
 - politische Rücksichten
 - zu enger Zeithorizont
 - begrenzte Mittel
 - Fortschritte bei Monitoring und Policy-Evaluierung

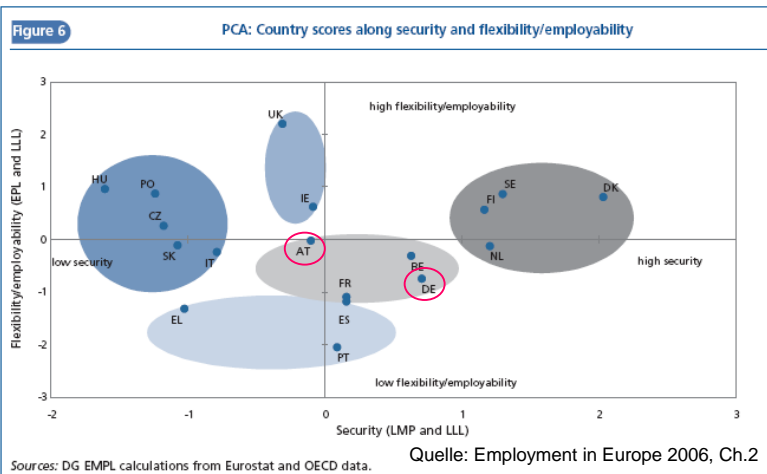
Flexicurity: das „Golden Triangle“



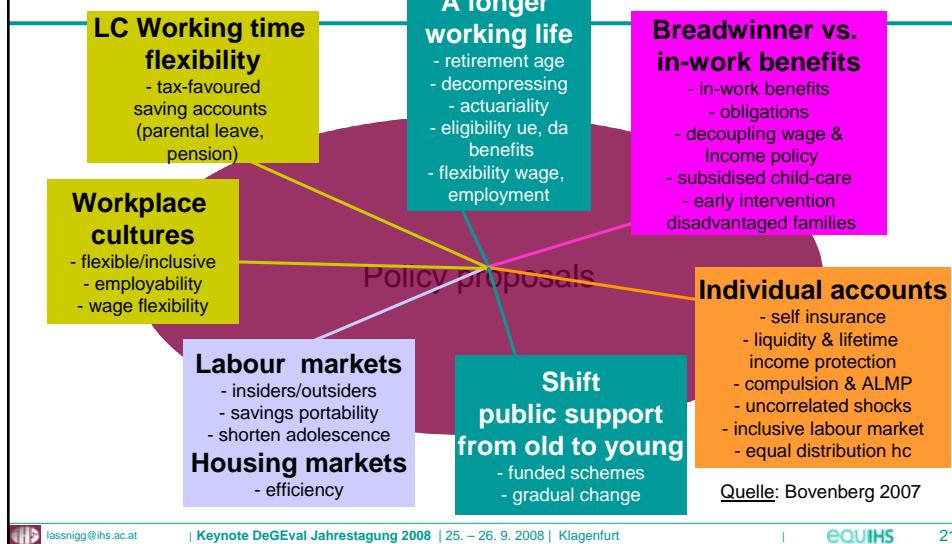
Flexicurity Bewertungsmaßstäbe

- Kündigungsschutz
- Ausgaben für aktive Arbeitsmarktpolitik und Weiterbildung
- Ersatzrate der Arbeitslosenversicherung

Flexicurity Bewertungsmaßstäbe



Life Course - policy als Rationale für Beschäftigungspolitik



Agenda

- **Evaluierung** von Arbeitsmarktpolitik: Ansätze, Ergebnisse, Probleme
- **Governance** von Arbeitsmarktpolitik: Entwicklung, Besonderheiten im Vergleich zu anderen Feldern
- Das **Verhältnis von E x G**: New Public Management, Evidence-Based Policy/Practice

E x G

Eine radikale Position

- *„...if you advocate a particular policy reform or innovation, do not press to have it tested“* (Gary Burtless & Robert Haveman)
- WILENSKY's LAW: *„the more evaluation, the less program development, the more demonstration projects, the less follow through“* (Harold Wilensky)

Nachweise: Lassnigg 1997

E x G

Umsetzung des „State-of-the-art“

- Voraussetzung für Verbindung von Evaluierung und Governance: Umsetzung der drei Kategorien
 - Wirkungsevaluation
 - Politik-Evaluation
 - Implementationsevaluation
- Beispiel „zielorientierte Evaluation mit vier notwendigen Schritten:
 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.

E x G

Umsetzung des „State-of-the-art“

- ABER: Grosse Schwäche Politikevaluation bleibt
- Beispiel EU-Programme (ESF, EQUAL, NAP, Reformprogramm...)
 - Formulierung der Programme erfolgt in politischem Aushandlungsverfahren...
 - ...und entspricht nicht den Anforderungen von „State-of-the-art“-Evaluierung
 - EvaluiererInnen stehen in Sandwichposition zwischen EU-Autoritäten und nationalen Regierungen („Landesverräter“)
 - Es geht um beträchtliche Mittel
 - Zeithorizont entspricht nicht der Logik der Evaluierung: „State-of-the-art“-Ergebnisse liegen erst vor, wenn Programmperiode vorbei ist

E x G

Überlegungen zum Zusammenhang

- Traditionelle Einordnung von Evaluierung: Die übergeordnete Instanz evaluiert die untergeordnete Instanz
 - Input > Output-Steuerung
 - in Bürokratie regulieren die Vorschriften
 - in neuem Governance-System entstehen Prinzipal-Agent-Verhältnisse mit Informationsvorsprung der Agents
- New Public Management: Anreize vs. Motivation
 - Rigide radikale Zielsteuerung (Monitoring, Performance Indicators) vs.
 - Empowerment (Evaluierung)
 - Einsatz und Ergebnisse umstritten
- Evidence-Based-Policy//Practice: eine wichtige Unterscheidung
 - verschiedene Logiken, verschiedene AdressatInnen, verschiedene Effekte und Voraussetzungen
 - Evaluierung eigentlich auf Praxis bezogen
 - Wer evaluiert Policy? Wer beauftragt das?

E x G

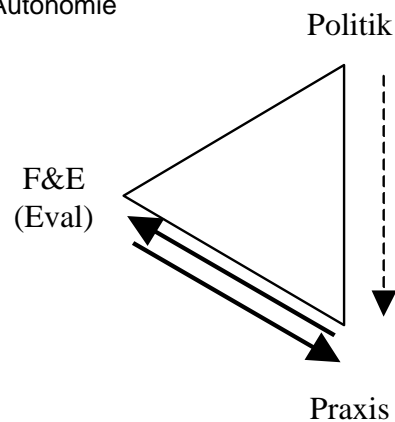
Überlegungen zum Zusammenhang

- Traditionelle Einordnung von Evaluierung: Die übergeordnete Instanz evaluiert die untergeordnete Instanz
 - Input > Output-Steuerung
 - in Bürokratie regulieren die Vorschriften
 - in neuem Governance-System entstehen Prinzipal-Agent-Verhältnisse mit Informationsvorsprung der Agents
- New Public Management: Anreize vs. Motivation
 - Rigide radikale Zielsteuerung (Monitoring, Performance Indicators) vs.
 - Empowerment (Evaluierung)
 - Einsatz und Ergebnisse umstritten
- Evidence-Based-Policy//Practice: eine wichtige Unterscheidung
 - verschiedene Logiken, verschiedene AdressatInnen, verschiedene Effekte und Voraussetzungen
 - Evaluierung eigentlich auf Praxis bezogen
 - Wer evaluiert Policy? Wer beauftragt das?

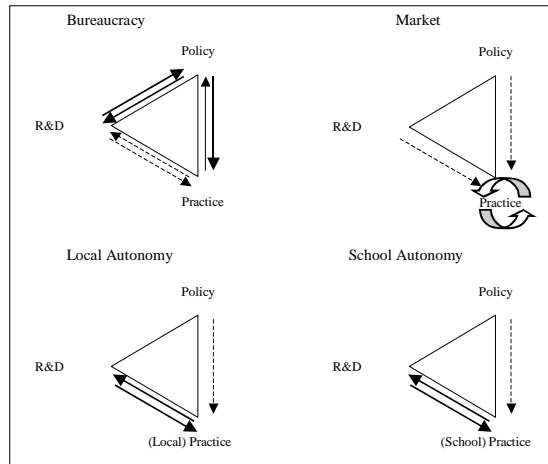
Das OECD Dreieck

F&E - Politik - Praxis

- Lokale Autonomie



Das OECD Dreieck F&E - Politik - Praxis & Governance



Evidence-Based-Policy

"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)

Agenda

- **Evaluierung** von Arbeitsmarktpolitik: Ansätze, Ergebnisse, Probleme
- **Governance** von Arbeitsmarktpolitik: Entwicklung, Besonderheiten im Vergleich zu anderen Feldern
- Das **Verhältnis von E x G**: New Public Management, Evidence-Based Policy/Practice
- Zusammenfassung, Schlussfolgerungen: Kann man aus der Erfahrung im Bereich Arbeitsmarktpolitik lernen?

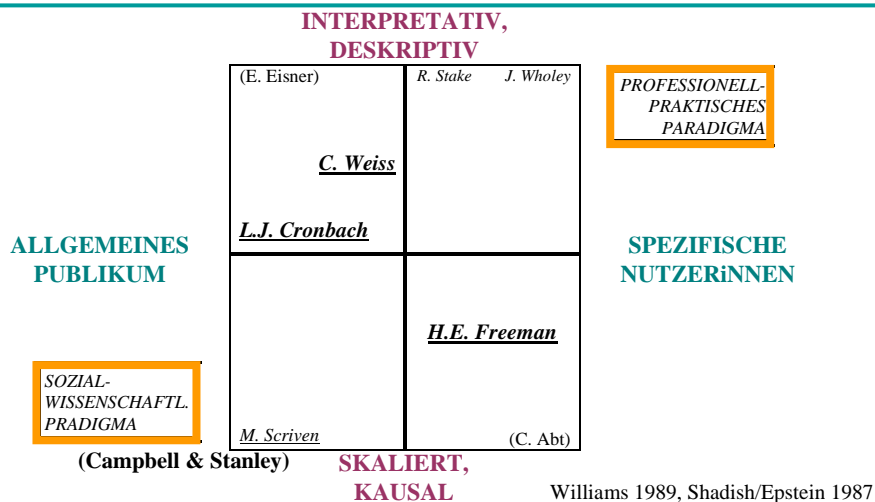
Zusammenfassende Punkte Kann man aus AMP was lernen?

- Der Bereich ist übersichtlicher als andere Bereiche (z.B. das Bildungswesen, oder die Wissenschaft), aber vielleicht stärker eingebettet (in Wirtschaft, Industrial Relations, Sozialsystem)
 - Aufgabenstellungen, Maßnahmen, etc.
- Trotz der längerfristigen internationalen Entwicklung eines State-of-the-Art im Bereich der Ergebnisevaluierung bleiben mehr Fragen offen als beantwortet werden
 - muss man sich wundern, dass die Methoden nicht angewendet werden?
- Die Politikentwicklung hat eigentlich die Fragestellungen und Methoden der Evaluierung transzendiert
 - Evaluierung ist eher für den Bereich der Evidence-Based-Practice verwendbar als für Evidence-Based-Policy
- Im Bereich des Policy Learning treten Probleme der Ziel- und Implementationsevaluierung in den Vordergrund

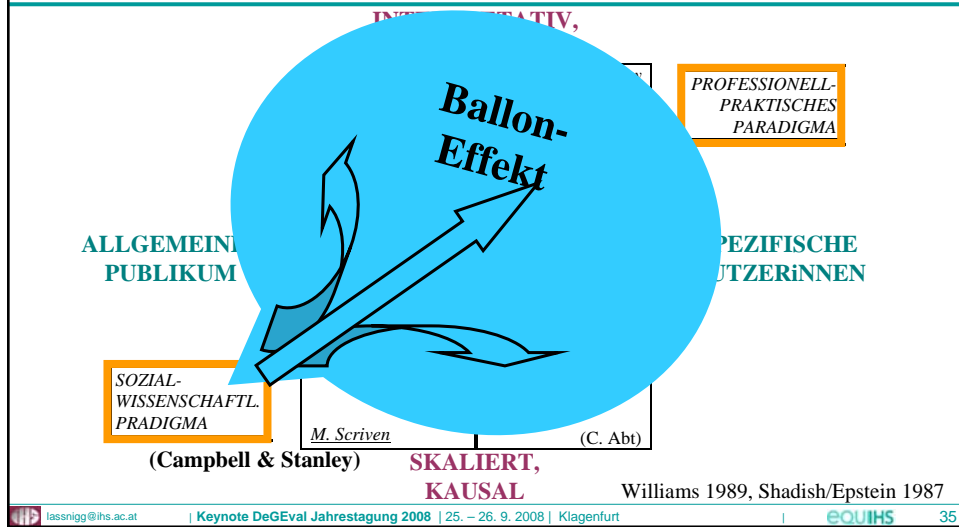
Verhältnis von Evaluierung und Governance oder Steuerung

- Evaluierung eignet sich als Instrument zwischen Prinzipals und Agents (zwischen Institutionen und Leistungserbringern) zur Verbesserung von Evidence-Based-Practice eher ...
- ... als zur Verbesserung von Evidence-Based-Policy: Gibt es aber ohne Evidence-Based-Policy einen sachgerechten Einsatz von Evaluierung?
- Betrachtet man die österreichische Arbeitsmarktpolitik, so gibt es bis heute fast keine Evaluierungen, die dem State-of-the-Art entsprechen würden ...
- ... 1995 hat ein OECD-Review-Team festgestellt, das AMS „does not seem to have a consistent programme of evaluation and only limited outcomes-testing policy to check the cost effectiveness of its varied programmes“ (OECD 1995, 104) - das hat sich nicht wesentlich geändert.

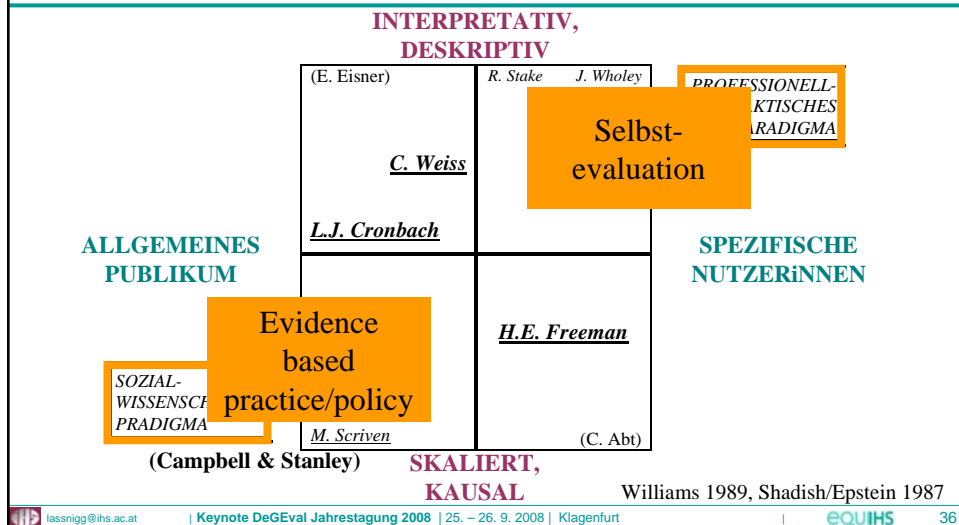
Evaluationspraxis: 2 Dimensionen: Methodologie, Nutzung



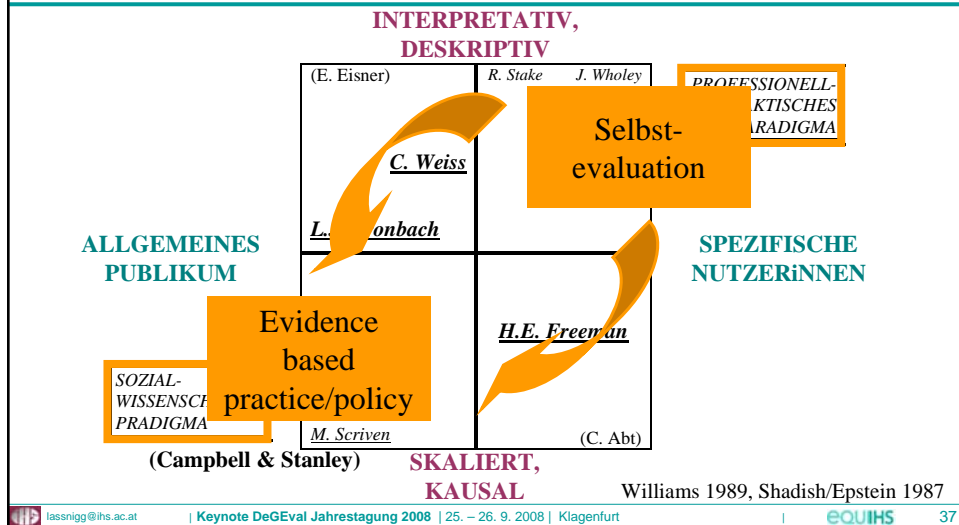
Evaluationspraxis: Entwicklung: „Ballon-Effekt“



Evaluationspraxis: 2 Dimensionen: Methodologie, Nutzung



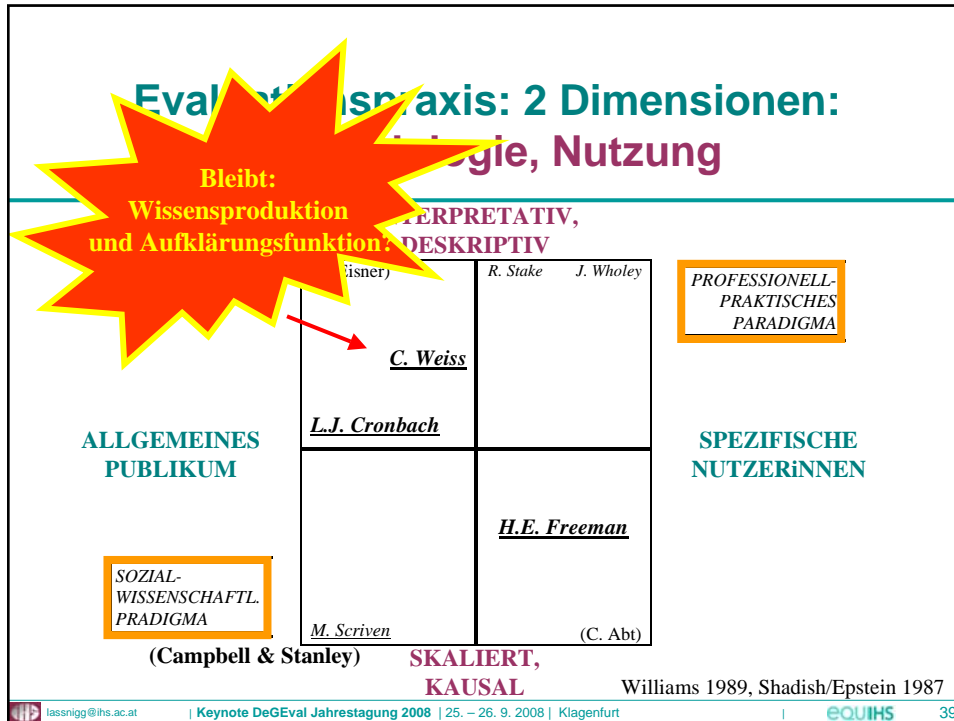
Evaluationspraxis: 2 Dimensionen: Methodologie, Nutzung



Prinzipien der “Vierten Generation”

<p>Axiome</p> <p>I. Werte-Pluralismus</p> <p>II. Beteiligte konstruieren ihre Wirklichkeiten</p> <p>III. Fairness</p> <p>IV. Leistung und Wert(schätzung)</p> <p>V. Aushandlung von Prozeß und Ergebnis</p>	<p>Prinzipien</p> <ol style="list-style-type: none"> 1. Eher Konflikt als Konsens bei den Werten der Beteiligten 2. Neue Methoden der Annäherung finden (trad. Evaluation reicht nicht aus) 3. Einbeziehung der Parteien mit unterschiedlichen Werten 4. Evaluator muß diese Konstruktionen erlernen 5. Evaluator klärt und vermittelt die Konstruktionen wechselseitig 6. Evaluator muß sich um alle Positionen bemühen 7. Evaluator muß alle relevante Information sammeln und zugänglich machen 8. Evaluator muß Information aufbereiten und "Fakten" für Gegner und Befürworter liefern 9. Befunde müssen laufend und offen zur Verfügung gestellt werden 10. Wert abhängig von Bedürfnissen und Kontext 11. Einschätzung von Bedürfnissen und Wert nur aufgrund von lokaler Feldarbeit 12. Übertragbarkeit dieser Einschätzung nur aufgrund empirischer Ähnlichkeit der Kontexte 13. Beteiligte müssen auf Ergebnisse reagieren können 14. Beteiligte müssen an allen Punkten der Evaluation gehört und ernstgenommen werden 15. Beteiligte müssen in die Formulierung der Schlußfolgerungen eingebunden werden 16. Zulassen von dauernden Meinungs-verschiedenheiten (Verstehen statt Konsens) 17. Evaluator ist Mittler und <i>change agent</i> in den Verhandlungen, Monitor von Aktionen <p style="text-align: right;">Quelle: Guba & Lincoln 1987, 74-79</p>
---	---

lassnigg@ihs.ac.at | Keynote DeGEval Jahrestagung 2008 | 25. – 26. 9. 2008 | Klagenfurt | EQUiHS 38



Material

Calmfors L (1994), Active Labour Market Policy and Unemployment - A Framework for the Analysis of Crucial Design Features, in: OECD Economic Studies 22, 7-47

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>

Kluve J (2006) The Effectiveness of European Active Labor Market Policy. IZA DP No. 2018 Internet: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=892341#

Bovenberg A L (2007) The life-course perspective and social policies: an overview of the issues. DELSA/ELSA/WP1(2007)8. Paris: OECD

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

Williams J E (1989), A numerically developed taxonomy of evaluation theory and practice, Evaluation Review 13 (1) (February), 18-31

Shadish W R Jr. / Epstein R (1987), Patterns of program evaluation practice among members of the Evaluation Research Society and Evaluation Network, Evaluation Review 11 (5) (October), 555-590.

Lassnigg L (1997), Evaluation: Aufdecken, zudecken, oder was sonst ... In: Zilian G.H. / Flecker J.. Hg. (1997): 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>

Guba E G / Lincoln Y S (1987), The countenances of fourth-generation evaluation: Description, judgment, and negotiation, Evaluation Studies Review Annual 11, 70-88

www.equi.at

lassnigg@ihs.ac.at | Keynote DeGEval Jahrestagung 2008 | 25. – 26. 9. 2008 | Klagenfurt | EQUIHS 40