



Evaluation von Klimaschutz

Stand und Qualität der Evaluation von Klimaschutzprojekten/-strategien

Präsentation auf der 10. Jahrestagung DeGEval
in Dresden 10.-12.10.07

Sebastian Bamberg
Universität Gießen



Übersicht

- Die DeGEval Genauigkeitsstandards
- I. Teil: Kritik der vorherrschenden Evaluationspraxis
 - Genauigkeit = interne Validität?
 - Stand der momentanen Evaluationspraxis
 - Die Vorteile echter experimenteller Designs
- II. Teil: Kritik der momentanen Forschungssynthese
 - Forschungssynthese und politische Entscheidung
 - Kritik der narrativen Forschungssynthese
 - Die Logik der quantitativen Meta-Analyse
- Fazit



Die DeGEval-Genauigkeitsstandards

- Sie sollen sicherstellen, dass eine Evaluation gültige Informationen und Ergebnisse zu dem jeweiligen Evaluationsgegenstand und den Evaluationsfragestellungen hervorbringt und vermittelt.




Die DeGEval-Genauigkeitsstandards

- **G5 Valide und reliable Informationen** - Orientierung an den Gütekriterien quantitativer und qualitativer Sozialforschung.
- **G6 Systematische Fehlerprüfung** -Die Informationen sollen systematisch auf Fehler geprüft werden.
- **G7 Analyse qualitativer und quantitativer Informationen** - Orientierung an fachlichen Maßstäben.
- **G8 Begründete Schlussfolgerungen**- Begründung der Folgerungen, damit die Adressaten und Adressatinnen diese einschätzen können.
- **G9 Meta-Evaluation** - Um Meta-Evaluationen zu ermöglichen, sollen Evaluationen in geeigneter Form dokumentiert und archiviert werden.



Teil I

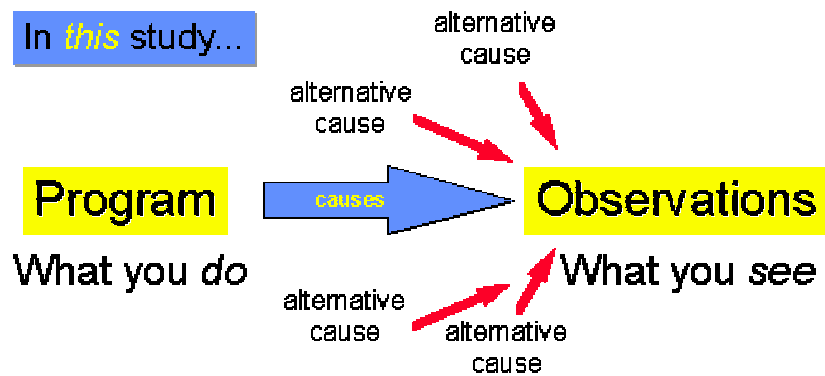
Kritik der vorherrschenden Evaluationspraxis



Kritik: Was bedeutet ‚reliable‘ und ‚valide‘ Evaluationsinformation?

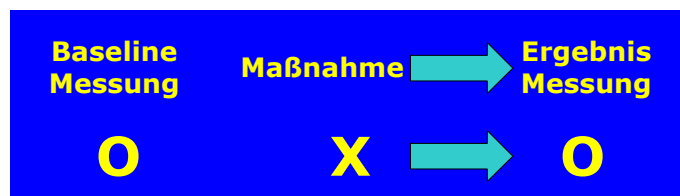
- Die Kernfrage bei der Evaluation von Maßnahmen zum Klimaschutz lautet:
- Können gemessene Veränderungen (CO₂-Reduktion) KAUSAL auf die implementierte Maßnahme zurückgeführt werden und NICHT auf andere mögliche Ursachen (sog. alternative Erklärungen)?
- Fachterminus: INTERNALE VALIDITÄT

Internale Validität



Internale Validität und Evaluationspraxis im Verkehr

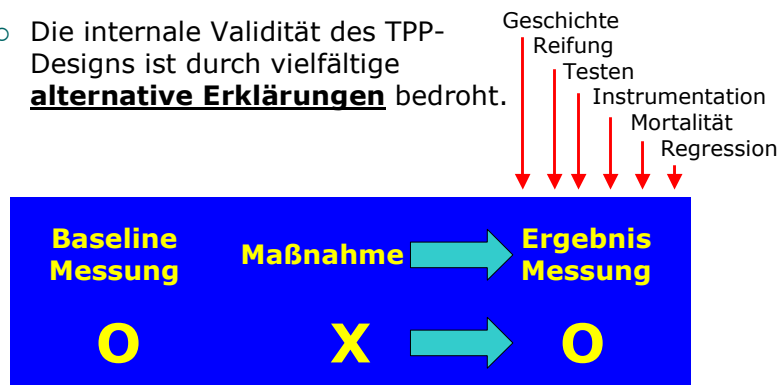
- Wenn im Verkehrsbereich überhaupt die kausale Wirkung einer Maßnahmen evaluiert wird, dann wird in der Regel ein ‚Treatment-Pre-Post-Test‘-Design (sog. Vorher-Nachher-Messung) verwendet.



- Besitzt das Vorher-Nachher-Design hohe interne Validität, d.h. ermöglicht es KAUSALSCHLÜSSE?

Internale Valitität und Evaluationspraxis im Verkehr

- Die interne Validität mittels Vorher-Nachher-Messung gewonnener Evaluationsdaten ist gering, gesicherte Kausalschlüsse nicht möglich.
- Die interne Validität des TPP-Designs ist durch vielfältige **alternative Erklärungen** bedroht.



Konsequenzen für die Evaluationpraxis

- Um zu wirklich belastbaren Evaluationsbefunden zu kommen müssen anstatt schwacher quasi-experimenteller Designs wie dem TPP echte experimentelle Designs mit hoher interner Validität benutzt werden („Gold-Standard“).
- Merkmale ‚echter‘ experimenteller Designs:
 - Kontrollgruppe
 - Randomisierte Zuweisung der Teilnehmer
- Merkmale ‚quasi‘-experimenteller Designs:
 - Vergleichsgruppe
 - Keine randomisierte Zuweisung der Teilnehmer

Alternative, experimentelle Evaluationsdesigns

Nur-Post-Test-Randomisiertes Design

R	X	O
R		O

Pre-Post-Test-Randomisiertes Design

R	O	X	O
R	O		O


Konsequenzen für die Evaluationpraxis

- Besonders das Nur-Post-Test randomisierte Design ist nicht nur viel mächtiger als das häufig benutzte TPP-Design, es ist auch billiger!
- Fazit: Vielen Verkehrswissenschaftler/innen sind zu wenig über die Konsequenzen nicht adäquater Evaluationsdesigns informiert.
- DeGEval sollte hier präzisere Standards formulieren und Praxisleitfäden für gute Evaluationsdesigns erstellen



Teil II

Kritik der momentanten Forschungssynthese



Forschungssynthese: Unverzichtbar für valide Politikberatung

- Eine Evaluationsstudie allein, selbst wenn sie eine hohe interne Validität hat, stellt keine belastbare Basis für evidenz-basierte politische Entscheidungen dar.
- Belastbar sind generalisierbare Evaluationsbefunde, d.h. vielfach replizierte Befunde mit einem klaren Trend.
- Aufgabe der Forschungssynthese ist es, reliable und valide Aussagen über den Trend einer Reihe von Evaluationsstudien zur gleichen Fragestellung zu machen.

Narrative Forschungssynthese problematisch

- Im Kontext summative Evaluation bezieht sich systematische Forschungssynthese auf die adäquate Zusammenfassung aller vorliegenden quantitativen Evaluationsbefunde.
- In der Praxis ist immer noch die narrative Forschungssynthese (Auflistung und verbale Diskussion) die Regel.
- Für die Synthese quantitativer Evaluationsbefunde ist die narrative Praxis der Forschungssynthese wissenschaftlich kaum verteidigbar.

Meta-Analyse als statistischer Ansatz der Forschungssynthese

- Ist die quantitative Analyse von quantitativen Studien.
- In einer Meta-Analyse wird ein Durchschnittseffekt (\overline{ES}) über alle verfügbaren einzelnen Interventionsstudien geschätzt.
- Bei der Berechnung des Durchschnittseffekts wird die unterschiedliche Präzision der Einzelbefunde berücksichtigt (Gewichtung über Inverse der Studienvarianz (Hedges & Olkin, 1985).

$$\overline{ES} = \frac{\sum_{j=1}^k w_j ES_j}{\sum_{j=1}^k w_j}$$

Meta-Analyse als statistischer Ansatz der Forschungssynthese

- So wird für das diskutierte Nur-Post-Test-Design die in einer Einzelstudie gemessene standardisierte Effektstärke ES wie folgt berechnet:

$$ES^{OPC_noB}_j = (M^{after,E}_j - M^{after,C}_j) / SD^P_{j..}$$

Meta-Analyse als statistischer Ansatz der Forschungssynthese

- Die Varianz einer Einzelstudie läßt sich ermitteln über:

$$SD^{OPC}_j = \sqrt{\left(\frac{ne_j + nc_j}{ne_j nc_j}\right) + \frac{(ES^{OPC}_j)^2}{2(ne_j + nc_j)}}$$

- Das Studiengewicht ist

$$w^{OPC}_j = 1 / (SD^{OPC}_j)^2.$$

Meta-Analyse als statistischer Ansatz der Forschungssynthese

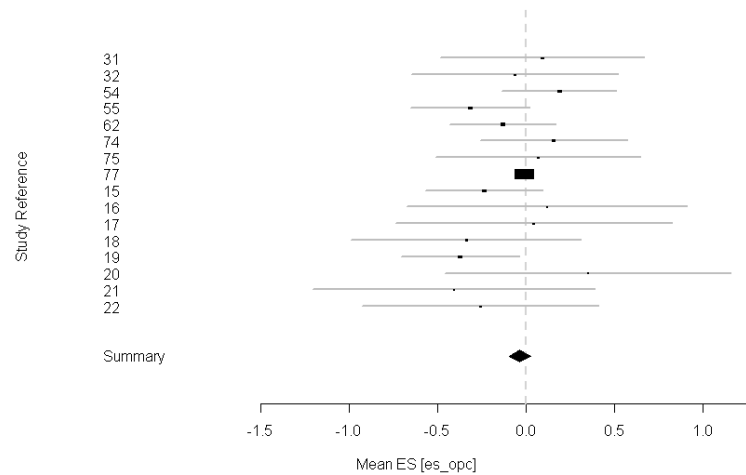
- Ferner läßt sich in einer Meta-Analyse die Wahrscheinlichkeit abschätzen, dass der berechnete gewichtete Durchschnittseffekt nur zufällige Fluktuationen widerspiegelt.
- Meta-Analyse liefert auch ein formales Modell zur Abschätzung und Modellierung der Studienheterogenität (within- und between-Study Varianz)

Meta-Analyse: Ein Beispiel

16 japanische Individualisierte Marketing-Kampagnen


TFP cases	N (exp. group)	N (control group)	M (exp. pretest)	M (exp. posttest)	M (control pretest)	M (control posttest)
2003 Sapporo GIS-based TFP	26	21	6.28	5.56	5.43	5.05
2003 Sapporo Paper-based TFP	24	21	5.21	4.75	5.43	5.05
2005 Ryugasaki TFP with feedback comments	83	67	6.32	5.83	5.41	5.31
2005 Ryugasaki TFP without feedback comments	70	67	4.27	4.04	5.41	5.31
2005 Fukuoka home-visit TFP	103	72	8.33	6.93	7.56	8.82
2005 Takasaki new comer TFP	108	28	6.08	5.58	6.15	5.13
2005 Ryugasaki new comer TFP	21	25	4.95	5.3	3.68	5.08
2003 Tokio TFP without behavioral feedback	2507	2865	7.87	8.40	6.97	8.50
2003 Kawanishi pt user TFP without behavioral feedback	108	52	7.61	7.09	6.81	8.75
2003 Kawanishi non pt user TFP without behavioral feedback	16	10	9.33	11.08	9.1	9.57
2003 Kawanishi non pt user TFP with ticket and without behavioral feedback	17	10	12.34	10.15	9.1	9.57
2003 Kawanishi TFP without non behavior change intention	18	19	9.71	9.45	10.8	12.9
2003 Kawanishi pt user TFP with behavioral feedback	106	52	6.28	6.07	6.81	8.75
2003 Kawanishi non pt user TFP with behavioral feedback	15	10	12.9	13.53	9.1	9.57
2003 Kawanishi non pt user TFP with ticket & behavioral feedback	16	10	8.89	6.56	9.1	9.57
2003 Kawanishi TFP with behavioral feedback without non behavior change intention	16	19	10.94	10.22	10.8	12.9

Meta-Analyse: Forest-Plot der 16 gewichteten Einzeleffekte und des gewichteten Durchschnittseffekts (OPC-Design)



Meta-Analyse: Gewichtete durchschnittliche Effektgröße berechnet nach 3 Forschungsdesigns

	\overline{ES}	$SE(\overline{ES})$	$z(\overline{ES})$	Q		Original Metrik			
						Verher (Pkw-Fahrten/Woche)	Nachher (Pkw-Fahrten/Woche)	Reduktion s-rate	
<u>Treatment Gruppe Pre-Post Test (TPP) Design</u>									
	-0.121	0.021	-5.68	(p<.001)	22.92	(p= .061)	6.91	6.03	12.7%
<u>Only Post-Test Kontrol (OPC) Design</u>									
	-0.109	0.06	-1.82	(p = .069)	13.01	(p = .525)	6.91	6.12	11.4%
<u>Pre-Post-Test Kontrol (PPC) Design</u>									
	-0.165	0.085	-1.95	(p = .052)	4.48	(p = .991)	6.91	5.72	17.2%



Fazit: Evidenzbasierte Entscheidung über Klimaschutzprojekte braucht Meta-Analysen

- Meta-Analyse ist ein sehr transparenter und mächtiger Ansatz zur Synthese quantitative Evaluationsbefunde.
- Kompetente Meta-Analyse eine Serie von Evaluationstudien, die Designs mit hoher interner Validität besitzen markiert den methodischen ‚Goldstandard‘ für evidenzbasierte Politikberatung
- DeGEval sollte so einen Ansatz für die Evaluation von Klimaschutzprojekten/-strategien promoten.