

## Zwischen Schema F und Innovation: Weiterführung einer politikfeldübergreifenden Diskussion zu methodischen Standards der Evaluation

*Lisa Ringhofer*<sup>1</sup>

Zum dritten Mal in Folge führte der Arbeitskreis (AK) Methoden in der Evaluation der DeGEval im Rahmen der Jahrestagung 2018 in Dresden eine Session zum methodischen Erfahrungsaustausch durch. Den Startschuss dieses Formats lieferte die Auftaktveranstaltung im Zuge der DeGEval-Jahrestagung 2016 in Salzburg, wo Vertreterinnen der Arbeitskreise Kultur und Kulturpolitik, Forschungs-, Technologie- und Innovationspolitik und Entwicklungspolitik Erfahrungen hinsichtlich der Anwendung von Evaluationsmethoden und -standards in ihren jeweiligen Arbeitsfeldern diskutierten (vgl. Altenburg 2017). Fortgesetzt wurde der begonnene Erfahrungsaustausch während der DeGEval-Jahrestagung 2017 in Mainz, diesmal unter Beteiligung der Arbeitskreise Soziale Dienstleistungen, Hochschulen und Gesundheitswesen (vgl. Heinze/Reiter 2018).

Auch die dritte Session hatte zum Ziel, mehr über aktuelle methodische Trends, Herausforderungen, Gemeinsamkeiten und Unterschiede in den jeweiligen Arbeitskreisen der DeGEval herauszuarbeiten und mögliche Synergieeffekte und deren Innovationspotenzial auszuloten. In diesem Sinne mochte die Session zur Nutzung der Potenziale der methodischen Cross-Fertilisation über die Grenzen von Politikfeldern hinweg beitragen.

Im altbewährten Format der Fishbowl-Diskussion (Innen-/Außenkreis-Methode) stellten dabei folgende drei Diskussionsteilnehmer im inneren Kreis exemplarisch die Evaluationserfahrungen aus ihren jeweiligen Politikfeldern vor und diskutierten gemeinsam mit dem Publikum methodische Trends:

- Michael Fraiss (Pädagogische Hochschule Zürich) – aktuell Sprecher AK Schulen
- Manfred Rolfes (Universität Potsdam) – aktuell Sprecher AK Stadt- und Regionalentwicklung
- Dominik Jessing (ifeu: Institut für Energie- und Umweltforschung Heidelberg GmbH) – aktuell Sprecher AK Umwelt

---

<sup>1</sup> TripleMinds Consultancy Network, Wien

Durch das Anbieten von freien Plätzen im inneren Kreis wurden die Seminarteilnehmerinnen und Seminarteilnehmer aktiv in die Methodendiskussion eingebunden und konnten somit ihre individuellen Sichtweisen und Erfahrungen schildern. Moderiert wurde die Diskussion von Jan Tobias Polak – Sprecher des AK Methoden – und anhand folgender drei Leitfragen strukturiert:

- Was sind ‚typische‘ Evaluationsmethoden bzw. methodische Herangehensweisen innerhalb der Politikfelder der DeGEval? Was sind Best-Practice-Beispiele?
- Was sind Mindestanforderungen an methodische/methodologische Qualität?
- Welche methodischen/methodologischen Innovationen lassen sich innerhalb des jeweiligen Politikfelds ausmachen?

## Typische und innovative Evaluationsmethoden – eine Bestandsaufnahme

Auf die Frage nach den ‚typischen‘ Evaluationsmethoden innerhalb des Politikfeldes Schule betonte Michael Frais in seinem Eingangsstatement, dass der Arbeitskreis einen gemeinsamen Orientierungsrahmen zu Qualitätskriterien habe, der den Ausgangspunkt hin zu den Methoden bildet. Die Methodenvielfalt im schulischen Bereich ist relativ groß, sowohl für die interne Evaluation (wie Lernstandserhebungen und Vergleichsarbeiten als Steuerungsmodus), als auch für externe Evaluationsverfahren (wie Dokumentenanalyse, Beobachtungsbögen, Rating-Konferenzen, Schulhausrundgang, Peer-Ansätze). Eine besondere Rolle spielen schriftliche Onlinebefragungen, weil durch die Massendatenverarbeitung interessante Datensätze generiert werden können. Die Verknüpfung quantitativer und qualitativer Daten im Sinne von Mixed-Methods-Designs bilden auch in diesem Politikfeld einen zentralen methodischen Ansatz. Im Feld der Schule – so Michael Frais – beobachtete er eine Art zeitliche Wellenbewegung in Bezug auf die Rolle von Evaluation: „Irgendwo werden wieder Sachen aufgebaut, Evaluationen abgeschafft, modifiziert und wieder ausgesetzt“.

Auf die Frage nach den Mindestanforderungen an methodischer Qualität unterstrich er einerseits den methodischen Anspruch auf sozialwissenschaftliche Gütekriterien, verwies jedoch gleichzeitig auf einen aktuellen Trend – der Abkehr von einer methodisch exakten Arbeitsweise basierend auf wissenschaftlicher Genauigkeit hin zur Betrachtung von Qualität als soziales Konstrukt. Denn auch die beste Methode nütze nichts, wenn eine Lücke zwischen Evaluierung und Umsetzung bestehe.

Im Handlungsfeld Umwelt betonte Dominik Jessing die Schwierigkeit einer einheitlichen Definition aufgrund seines fragmentierten Charakters. Aus diesem Grund sei eine Differenzierung zwischen einer naturwissenschaftlichen und einer menschlichen Umwelt unabdingbar, da sich beide aufgrund unterschiedlicher Erkenntnisinteressen verschiedener Methodensätze bedienen. Auf naturwissenschaftlicher Ebene ist die Methodenbasis zunächst recht einfach. Hier existiert eine relativ ausgeprägte Umwelt-Indikatorik mit teilweise standardisierten Monitoring-Instrumenten und Benchmarks (Was wird als gute oder schlechte Umwelt bezeichnet?). Auf der

menschlichen Ebene hingegen erstreckt sich ein breiter Sektor, den Dominik Jessing als „technische und organisatorische Maßnahmen“ im Umweltbereich subsumiert – von Bilanzierungen, Umweltverträglichkeitsprüfungen, Regulierungsfolgenabschätzungen bis hin zu Fragen von Umwelt als soziales Konstrukt. Je nach Blickwinkel bediene man sich folglich unterschiedlicher Methodenpaletten.

Bei Programmevaluationen gilt die Faustregel: Je mehr öffentliches Geld im Spiel, desto eher werden Evaluationen methodisch getrieben (Rechenschaftspflicht). Entsprechend sind die gesetzten Standards nicht unbedingt naturwissenschaftlich, sondern tendenziell im wirtschaftswissenschaftlich-mathematischen Bereich angesiedelt. Diese unterschiedlichen Standards können durchaus Spannungen erzeugen. Als Fazit betonte Dominik Jessing erneut, dass es also kaum übergreifende Standards gäbe, auf die man sich im Umweltbereich einigen könne.

Im Politikfeld Stadt- und Regionalentwicklung führte Manfred Rolfes ein paar typische methodische Merkmale an. Allem voran geht es bei größeren Städtebauförderungsprogrammen mehr um formative – also prozessbegleitende Evaluation – als um eine summative Bewertung der Maßnahmen. Gleichzeitig verfestigt sich die Perspektive, dass bei Evaluation der Stadt- und Regionalentwicklung drei Aspekte der Evaluation im Vordergrund stehen – die Kontrollperspektive, die Legitimationsperspektive und die Symbolperspektive. Der Kontroll- und Legitimationsaspekt hat vielerorts auch durchgesetzt, dass Evaluation grundsätzlich eher ‚skeptisch beäugt‘ wird. Im Rahmen von Evaluation kommt grundsätzlich ein sehr breites Methodenspektrum aus der Sozialforschung zum Einsatz. Da sich die Stadt- und Regionalentwicklung mit raumbezogenen Veränderungsprozessen beschäftigt, werden ferner vielfach Methoden aus dem Bereich der Regionalstatistik und der empirischen Regionalforschung eingesetzt. Zur Beantwortung der Frage: Wie verändern sich Räume? kommen immer wieder (quasi-)kontrafaktische Analysen zum Einsatz – also Vergleichsstudien zwischen Interventions- und Vergleichsgebieten. Diese Analysen stoßen jedoch schnell an ihre Grenzen, da es de facto keine zwei gleichen Gebiete gibt.

Aktuell zeichnen sich in der Stadt- und Regionalentwicklung zwei Diskussions-trends ab: Einerseits gilt es, die vorhandenen Methoden zu optimieren und bessere Messinstrumente (Frauen-Monitoring, Geographische Informationssysteme) zu entwickeln. Andererseits beobachtet er ein wachsendes Bewusstsein darüber, dass Projekte nicht mehr linear ablaufen, sondern diskursiv und prozesshaft sind. Hier gilt es, diese Prozesshaftigkeit besser zu planen und in der Evaluation dementsprechend einzufangen. Dies wiederum würde eine produktive Feedbackkultur erzeugen und der Wahrnehmung von Evaluationen als Kontrollinstrumente entgegenwirken. Inspiration dafür bieten systemische Ansätze, die mehr Perspektivität, Strukturierung und Systematik bieten. Methodisch sieht er eher eine Flexibilisierung der Methoden und die Tendenz, schlanke und schnelle Evaluierungsmethoden im Feld einzusetzen (z.B. bei Ad-Hoc- und Peer-Gruppen).

## Politikfeldübergreifende Diskussion zu methodischen Standards

In der anschließenden Diskussion wurden verschiedene Aspekte vertieft. Zunächst wurden die methodischen Mindestanforderungen an die Datenerhebung betrachtet – dabei kristallisierte sich das Thema der Transkription als zentraler Diskussionspunkt heraus.

Michael Fraiss zeigte das oftmalige Dilemma zwischen wissenschaftlich rigorosen Daten und knappen Zeit- und Finanzressourcen auf. Daher müsse seiner Meinung nach pragmatisch vorgegangen werden, denn im Gegensatz zu universitären Forschungsarbeiten können bei Evaluationen nicht alle qualitativen Interviews akribisch transkribiert werden. Als Rechtfertigung führte er an, dass qualitative Methoden oft nur eine Ergänzung zum Gesamtbild darstellen. Im schulischen Bereich gehe es vielmehr darum, Evaluation als Impulsgeber für Schulentwicklung zu betrachten und ein Bild einzufangen, das auf allgemeine Akzeptanz stoßt. Daher gehe es weniger um „ganz hohe methodische Ansprüche“, denn die Methodik als reiner Selbstzweck sei überflüssig. Vielmehr gehe es darum, kontextspezifisch zu überlegen, welchen Aufwand es braucht, um in die Tiefe des Datenmaterials vorzudringen. Wichtig sei dabei, „dass ich meine Vorgehensweise vertreten, rechtfertigen und plausibilisieren kann“.

Ähnlich im Falle der Stadt- und Regionalentwicklung. Hier geht es laut Manfred Rolfes primär darum, dass Evaluation aufzeigen soll, dass etwas Vernünftiges gemacht wird. Die inhaltlich-methodische Debatte sieht er eher als sekundär. „Da mag ich mich auch nicht der Illusion hingeben, dass da was transkribiert wird“ – lautet sein genereller Eindruck von größeren Rahmenevaluationen. Generell gebe es seitens der Auftraggeberinnen und Auftraggeber viel Vertrauen in die Evaluierenden, hier ohnehin methodisch sauber zu arbeiten. Auf die Frage von Jan Tobias Polak, ob dieses nicht genaue Hinsehen eventuell ein Problem für die Evaluierung darstellen könnte, konterte Dominik Jessing mit der Aussage „Wo kein Kläger, da kein Richter und verwies darauf, dass es zur Beurteilung der methodischen Qualität von Evaluationen vielerorts an Kompetenz und manchmal auch an Interesse seitens der Auftraggeberin bzw. des Auftraggebers fehle.

Im weiteren Verlauf der Diskussion stellte Martin Bruder (DEval) die Frage, inwieweit die beiden Pole – partizipative Prozesse und die soziale Konstruktion von Qualität auf der einen Seite, und das Streben nach Standards und Rigorosität auf der anderen Seite – denn miteinander kombinierbar wären? Oder gibt es da vielleicht auch Grenzen in die eine oder andere Richtung? Die Frage stelle er vor dem Hintergrund seines eigenen Politikfeldes – der Entwicklungszusammenarbeit – wo Debatten im Gange sind, inwieweit man denn überhaupt als Sprachrohr für die von den Maßnahmen Betroffenen agieren kann.

Für den Umweltbereich sei dies extrem kontextabhängig, führte Dominik Jessing an. Hier gibt es einerseits den technischen Bereich der Indikatorik, wo man aber auch ohne Fachwissen erkennen kann, dass etwas nicht stimmt – wie etwa am Beispiel vom Dieselskandal. Der andere große Bereich beschäftigt sich damit, wie

Umweltfragen im Alltag umgesetzt werden (Wie gestalte ich Trassenplanung? Was passiert, wenn tatsächlich ein Windrad in unserer Stadt gebaut wird?). Dieser letztere Bereich, in dem es um diskursive Prozesse geht, bedient sich ganz anderer, meist qualitativer Methodensätze. Bei diesen Fragen seien auch die Qualitätsanforderungen gar nicht so sehr methodischer Natur, sondern rein ‚gefühlte‘. Wie wird das Gefühl des Interessensausgleichs wahrgenommen? Dieses Wahrnehmungsgefühl muss nicht unbedingt mit methodischer Qualität zusammenhängen.

In der Stadt- und Regionalentwicklung, hob Manfred Rolfes nochmals die Wichtigkeit von Methoden hervor, die partizipative Prozesse unterstützen und Diskursivität einfangen. Ad-Hoc-Auswertungen von Interviews, die dann an die Leute zurückgespielt werden, bilden dabei einen wichtigen Bestandteil. Er betont ferner, dass es methodisch noch „sehr viel Luft nach oben“ gebe, um Evaluationsergebnisse zu transportieren, Feedback einzuholen und daraus kohärente Rückschlüsse zu ziehen.

Im weiteren Diskussionsverlauf wurde der oftmalige Gap zwischen Wissenschaft und Praxis debattiert. Im Umweltbereich verweist Dominik Jessing auf die Problematik, Umweltphänomene von wissenschaftlicher Seite so darzustellen, dass sie für die Normalbevölkerung auch praktisch nutzbar sind (z.B. werden Wetter und Klima oft synonym wahrgenommen). Dies sei insbesondere bei komplexen naturwissenschaftlichen Wirkungszusammenhängen der Fall, wo eine Vereinfachung sehr fehleranfällig sei (da Folgeeffekte oftmals ausgeklammert werden). Darüber hinaus sei es auch schwierig, diese gefühlte Diskrepanz durch Standards zu entschärfen.

Daher wurde die Frage der Kommunikation von zuverlässigem und gültigem Wissen als zentraler Aspekt aufgegriffen. Manfred Rolfes plädiert für eine andere Form von Kommunikation. Nur durch eine enge Zusammenarbeit und einen offenen Dialog- und Kommunikationsprozess zwischen Wissenschaft und Praxis sei es möglich, Methoden zu entwickeln, die auch das Diskursive einfangen. Und da, so sein Fazit, „sind wir überhaupt nicht weit“. Kommunikation sei jedoch auch fehlerbehaftet, meint Martin Bruder und verweist auf die Gefahr von Fake News. Eine Vertreterin des Bundesministeriums für Wirtschaft und Energie warf ein, dass Kommunikation als Brückenschlag zur Politik nicht einzig und allein die Aufgabe von Evaluierenden sei, da Politikberatung schließlich auf vielfältiger Ebene passiere. Dabei verweist sie auf Expertengremien, die Ethikkommission und Ressortforschungseinrichtungen an Universitäten, die genau diesen Auftrag haben.

Insgesamt gebe es offensichtlich systemische Probleme, die im Spannungsverhältnis zur methodischen Güte stehen, fasste Jan Tobias Polak die unterschiedlichen Sichtweisen zu methodischen Qualitätsansprüchen zusammen. Auf der einen Seite brachte Tanja Schmutz (Pädagogische Hochschule Graubünden) ihre Kritik zum Ausdruck, bei der gewählten Methodik Abstriche zu machen. Schließlich sei es Aufgabe der Wissenschaft, gut fundierte Daten zu liefern – wenn die Validität und Reliabilität der Daten nicht mehr stimme, dann leide auch die Aussagekraft darunter. Je partizipativer man in der Evaluation vorgeht – so ihre Erfahrung – desto höher sollten auch die Ansprüche in der methodischen Umsetzung sein. Dem gegenüber stehen jedoch oftmals andere Evaluierungswirklichkeiten, konterte Dominik Jessing, und zeigte mögliche Dilemmata auf. Zum einen könnte seitens der Auftraggeber

rin bzw. des Auftraggebers eine komplexe Methodik nicht verstanden bzw. erst gar nicht gewünscht sein (weil dies eventuell bedeuten würde, dass die Programmzahlen schlechter werden). Zum anderen gibt es sehr viele konkurrierende Programmwirklichkeiten. Er veranschaulichte dies anhand eines Beispiels: „wenn ich Klimawandel im Heizungsbereich vermeiden will, dann ist es unsinnig, dass ich Ölheizungen fördere. Aber es werden sowohl Ölheizungen gefördert als auch Gesetze erlassen, die Ölheizungen verbieten. Und da komme ich mit Methodik an Realitätsgrenzen, die ich einfach nicht übersetzen kann“.

Ein wichtiger Punkt, den Johannes Schmitt (DEval) aus Sicht des Evaluierenden anführte, ist die Bedeutung der stärkeren Zusammenarbeit mit den Auftraggebern, um gemeinsam die Vor- und Nachteile der methodischen Optionen, welche im gegebenen Umfang von Zeit und Geld möglich sind, durchzudiskutieren. Dies wurde auch von Seite des Auftraggebers bestätigt: man verwies darauf, dass man immer transparent machen sollte, welche Methoden man verwendet habe. Es sei ebenfalls wichtig, immer klar zu kommunizieren, in welchem Falle methodisch weniger tief gegangen wurde. Nicht immer sei die volle Bandbreite des Methodenspektrums möglich bzw. auch nötig, denn dies wäre schließlich auch gegen den Aspekt der Wirtschaftlichkeit, teilte eine Vertreterin des Bundesministeriums für Wirtschaft und Energie in diesem Zusammenhang mit.

Johannes Schmitt brachte zudem einen interessanten Aspekt in die Runde, indem er das Referenzgruppenmodell des DEvals vorstellte. In diesem Referenzrahmen werden Themen vorgeschlagen, diskutiert, die Fragestellungen mit den Beteiligten zusammen besprochen und basierend darauf die methodischen Optionen überlegt. Insbesondere dient dieses Referenzgruppenmodell zur methodischen Absicherung, was vor allem bei politisch brisanten Themen von größter Bedeutung ist.

Im Rahmen der Diskussion kam auch der Aspekt der Nützlichkeit von Evaluation auf, einem wesentlichen Qualitätskriterium in der DeGEval. Es sollte eine enge Verbindung zwischen Evaluationsqualität und Evaluationsnutzen geben, betonte Dominik Jessing, einen automatischen Zusammenhang gebe es jedoch nicht. Michael Frais bestätigt das wichtige Kriterium der Nützlichkeit auch für den Schulbereich: „Wenn da der Sinn nicht erkannt wird, ist es ganz schnell vorbei mit Evaluierungen“.

Manfred Rolfes machte ferner auf einen interessanten Aspekt aufmerksam. In seiner Erfahrung sei es wichtig, zwischen Programm- und kleineren Projektevaluierungen zu unterscheiden. Programmevaluierungen laufen meist „sehr geölt“ und zeigen erfahrungsgemäß keine großen Überraschungen auf. Denn die Entscheidung, ob ein Programm gut läuft, ist meist schon vorher gefallen. Die Evaluierung muss dies nur „irgendwie nachfühlen“ und macht das in der Regel auch methodisch gut. Anders sehe er das bei kleineren Projektevaluierungen, die sie auch häufig selbst durchführen. Da stehe eben ganz stark das Interesse im Vordergrund, offen und ehrlich von den Ergebnissen zu lernen und diese auch dementsprechend an die Befragten rückzuspiegeln. Bei dieser Art der Evaluation brauche es nicht unbedingt das große methodische Besteck der empirischen Sozialforschung, „denn es müsse ja schnell gehen und bestimmte Punkte auf den Punkt gebracht werden“. Daher genüge oftmals eine qualitative Validierung dessen, was man in den Interviews gehört

hat. Dazu merkte Jan Tobias Polak an, dass Evaluationen in der Entwicklungszusammenarbeit unabhängig von deren Umfang und Größe immer die fünf DAC-Kriterien (Effizienz, Effektivität, Relevanz, Nachhaltigkeit, Impact) abdecken müssen. Hier sollte man über mögliche Stoßrichtungen nachdenken, diese Ansprüche zu reduzieren, ohne jedoch die methodische Güte zu beeinträchtigen.

Im Anschluss an die Diskussion resümierte Jan Tobias Polak drei Aspekte, um eine bessere methodische Qualität von Evaluation sicherzustellen: (1) es bedarf einem kritischeren methodischen Blick auf wissenschaftliche Studien seitens der Öffentlichkeit; (2) Auftraggebende müssen mehr Verständnis für und Wissen über methodische Güte haben; und (3) es braucht einen eigenen Anspruch als Evaluierende, was die Güte unserer methodischen Arbeit angeht. Dies würde zudem bedeuten, dass der/die Evaluierende immer kontextspezifisch den Aufwand und Ertrag kalkuliert und seine/ihre methodische Vorgehensweise dementsprechend abstimmt.

## Wohin bewegen sich Ansätze und Methoden in der Evaluation – ein Ausblick

Mit Blick in die methodische Zukunft meinte Dominik Jessing, dass im Politikfeld Umwelt vor allem im quantitativen Bereich erhebliche Fortschritte zu verzeichnen sind, während er den Aufholbedarf eher im qualitativen Bereich und in der Frage der Übersetzung und Nutzbarmachung der Ergebnisse sieht. Für das Politikfeld Schule verfestigt sich laut Michael Fraiss der Begriff der Evaluationskultur, welcher auch mit einem höheren methodischen Anspruch einhergeht. Im Politikfeld der Stadt- und Regionalentwicklung beobachtet Manfred Rolfes einen Trend hin zu partizipativen und iterativen Prozessen und Methoden, bei denen abwechselnd Daten erhoben, analysiert und kommuniziert werden. Auf die Frage der ‚Drivers‘ für diese methodischen Veränderungen wurden unterschiedliche Aspekte eingebracht. Im Umweltbereich habe dies vor allem mit der Rechenschaftspflicht zu tun (Wie sinnvoll werden die Gelder eingesetzt?). Im Schulbereich geht die methodische Innovation mit einem gewissen Sättigungsgrad an monotonen Befragungsmethoden einher. Man ist schlichtweg auf der Suche nach Methoden, die mehr Potenzial zu Inspiration und Motivation bieten. Die methodische Veränderung in der Stadt- und Regionalentwicklung ist eng mit einem wachsenden Bewusstsein und Bedürfnis verbunden, komplexe und komplizierte Zusammenhänge besser zu verstehen und einordnen zu können. Daher auch der Trend hin zu prozessorientierten Methoden, die zum Ziel haben, Komplexität besser abbilden zu können.

Aufgrund des regen Interesses wird der Arbeitskreis Methoden im Rahmen der kommenden 22. Jahrestagung der DeGEval in Bonn mit gleichem Format drei weitere Arbeitskreise näher betrachten.

## Literatur

- Altenburg, Thomas (2017): Zwischen Schema F und Innovation. Eine politikfeldübergreifende Diskussion zu methodischen Standards. In: Zeitschrift für Evaluation, 17 (1), S. 210-217.
- Heinze, Franziska/Reiter, Stefanie (2018): Methodische Standards der Evaluation zwischen Schema F und Innovation – Weiterführung einer politikübergreifenden Diskussion. In: Zeitschrift für Evaluation, 18 (1), S. 188-197.