

Wie weit können, dürfen und müssen Evaluierende gehen, um die Welt zu retten? Zum Rollenverständnis von Evaluierenden

Frühjahrstagung des Arbeitskreises Aus- und Weiterbildung in
der Evaluation am 21. März 2014 in Saarbrücken

*Maria Gutknecht-Gmeiner*¹

Das Rollenverständnis von Evaluierenden war das Thema der Frühjahrstagung des Arbeitskreises Aus- und Weiterbildung in der Evaluation am 21. März 2014 in Saarbrücken. Gastgeber waren das Centrum für Evaluation (CEval) in Saarbrücken und der Masterstudiengang Evaluation der Universität des Saarlandes und der Hochschule für Technik und Wirtschaft des Saarlandes.²

Ausgehend von der Frage der Nutzung von Evaluationen sollte die Tagung mit dem provokativen, von Bob Stake (2004) entlehnten Titel „Wie weit können, dürfen und müssen Evaluierende gehen, um die Welt zu retten?“ den etwa 20 Teilnehmenden aus verschiedenen Institutionen und Evaluationskontexten eine eingehende Auseinandersetzung mit der Rollenfrage ermöglichen. Ziel war es, verschiedene Perspektiven auf das Rollenverständnis von Evaluierenden explorativ zu ergründen, sie an inhaltlich-theoretischem Wissen zu messen sowie Herausforderungen für die Theorie und Praxis der Evaluation zu thematisieren. Die Fachtagung war sehr interaktiv, streckenweise auch workshopartig konzipiert. Sie diente dem Austausch, der Diskussion und Meinungsbildung zu diesem Thema, angeregt durch unterschiedliche Methoden der Auseinandersetzung.³

In einer ersten aktivierenden Einstiegsphase, in der die Teilnehmenden gebeten wurden, sich als ‚lebende Antwortskala‘ im Raum zu verschiedenen Aussagen zu ihrem evaluatorischen Arbeiten zu positionieren, zeigte sich, dass den Teilnehmenden

1 IMPULSE – Evaluation und Organisationsberatung, Wien

2 Der Tagungsbericht fußt auf der Dokumentation der Autorin sowie der Organisatorinnen/Organisatoren der Tagung. Ein spezieller Dank geht an Jan Hense (Universität Gießen), Sandra Schopper (Masterstudiengang Evaluation), Heike Steckhan (DEval) und Stefan Rädiker (VERBI GmbH), die mit Kurzbericht, Fotodokumentation, Notizen und einem Mitschnitt der Podiumsdiskussion diesen Bericht unterfüttert haben.

3 Die Tagungsdokumentation sowie Fotos finden sich unter <http://www.degeval.de/arbeitskreise/aus-und-weiterbildung-in-der-evaluation/aktuelles/>

den der Einbezug von Stakeholdern ein Anliegen ist und sie durch ihre Tätigkeit als Evaluator/-in zu Verbesserungen der evaluierten Programme und/oder Institutionen beitragen wollen. Evaluation als ‚Advocacy‘, d.h. als Beitrag zur Förderung bestimmter gesellschaftlicher Ziele und Zielgruppen, wurde jedoch eher skeptisch gesehen: Ein direkter Beitrag zu Empowerment, Selbstbestimmung und Emanzipation der Beteiligten kann da und dort möglich und erwünscht sein (vor allem wenn der Evaluationsgegenstand darauf abzielt), meist ist dies jedoch nicht Teil des Evaluationsauftrags. Problematisiert wurde in diesem Zusammenhang auch die Frage der Unabhängigkeit und Unparteilichkeit von Evaluierenden. Trotzdem kann Evaluation allein durch die Einbeziehung und die Art des Umgangs mit Zielgruppen positiv auf diese wirken, in dem ihre Stimme wertgeschätzt und zu Gehör gebracht wird. Evaluierende können dazu beitragen, die Welt zu einem besseren Ort zu machen, so der einhellige Tenor, sie können jedoch nicht verantwortlich gemacht werden, die Welt zu retten, da dies nicht in ihrer Macht liegt.

*Weihnachtsmänner und Dompteurinnen –
was haben Evaluierende mit anderen Berufen gemeinsam?*

In einer von Maria Gutknecht-Gmeiner (IMPULSE – Evaluation und Organisationsberatung) moderierten Assoziations- und Reflexionsübung (vgl. DeGEval 2010; Gutknecht-Gmeiner 2015) wurden daraufhin die Rollen und Aufgaben von Evaluierenden spielerisch mit denen anderer Berufe verglichen. Dabei wurden verschiedene – auch ambivalente – Aspekte der Tätigkeit von Evaluierenden sichtbar:

- Angesprochen wurde immer wieder die Aufgabe von Evaluierenden, Daten zu erheben und dabei Neues und Verborgenes zu entdecken – z.B. wie die „Tiefseetaucher, weil sie in die Tiefe gehen und Verborgenes hervorheben“ oder wie „Aliens, weil sie viel Neues entdecken“. Wie „Schatzgräber“ finden sie Ergebnisse, die bislang „ungehobene Schätze sind“. Sie bringen damit wie die „Gärtner“, die „Praxis zum Blühen“, d.h. tragen zur Weiterentwicklung des Evaluationsgegenstandes bei.
- Nicht immer ist dieses Heben und Beisteuern von Erkenntnissen durch die Evaluierenden ausschließlich positiv. Wie die „Angler fischen sie nach Ergebnissen, [...] manchmal in trüben Teichen“. Oder es werden Erwartungen an sie herangetragen, wie an den „Weihnachtsmann“, der „Überraschungen bringt und Wünsche erfüllen soll“.
- Evaluierende haben manchmal das ambivalente Image von „Zahnärzten“, „weil jeder Angst hat und weil Vorsorge besser ist als bohren [...] und manchmal ist der Auftraggeber danach zahnlos“. Oder sie sind auf der anderen Seite wie „Häftlinge gefangen in politischen Konstellationen“ und sollen wie „Friseure die Ergebnisse so frisieren, dass sie nett erscheinen“.
- Die Kommunikation mit Auftraggebenden und Stakeholdern ist das Um und Auf: Evaluierende müssen wie die „Pfarrer viel ‚predigen‘ und Überzeugungsarbeit leisten“. Ihre Arbeit gleicht der von „Regisseurinnen/Regisseuren, weil sie viele Akteure unter ein ‚Set‘ bringen müssen“, bisweilen sogar der von „Dompteurinnen/Dompteuren, die den Stakeholderzirkus bändigen“.

- Nicht immer ist Evaluation eine aufregende und kreative Tätigkeit: Evaluierende sind manchmal wie „Roboter, die fremdgesteuert und auf Knopfdruck immer dieselben Routinen abspulen“. Evaluierende haben auch viel mit „Hausfrauen“ gemeinsam, „weil sie mit kleinem Budget viele glücklich machen“ müssen. Schließlich sind Evaluierende wie „Fußballspieler/-innen“ ständig gefordert, denn „nach der Evaluation ist vor der Evaluation“.

*Beraten, Richten, Lehren, Forschen:
Rollenverständnis in der Evaluationsforschung*

Nach dieser Aufwärmübung stellte Jan Hense (Justus-Liebig-Universität Gießen) die Ergebnisse einer von Anna von Werthern (Ludwig-Maximilians-Universität München) und ihm selbst durchgeführten Literaturrecherche zum „Rollenverständnis von Evaluatorinnen und Evaluatoren“ vor. Die Ausgangslage wurde in diesem Kurz-Input von Jan Hense folgendermaßen skizziert: Die Person und Rolle von Evaluierenden werden im Rahmen von Forschung zu Evaluation immer wieder aufgegriffen und spielen für die Professionalisierung von Evaluation eine wichtige Rolle. Trotzdem gibt es dazu wenig Theorie und kaum empirische Befunde. Das Thema wird auch in der (internationalen) ‚Community of Practice‘ wenig diskutiert.

Dann wurden vorhandene theoretische Positionen zum Rollenverständnis vorgestellt. Diese beziehen sich auf unterschiedliche Aspekte einer Evaluation: So können sich Rolle und Aufgabe von Evaluatorinnen/Evaluatoren aus der Wahl des Evaluationsmodells und der Evaluationsmethoden ergeben oder aus der Beziehung zu den Stakeholdern. Sie können aber auch als prinzipiell situationsabhängig und flexibel verstanden werden oder in Abhängigkeit von wechselnden Rollen je nach Phasen im Evaluationsprozess. Vorgestellt wird das Modell von King und Stevhan (2002), die je nach Evaluationsansatz – traditionell, partizipatorisch oder Aktionsforschung – unterschiedliche Rollen und Aufgaben festmachen. Dabei entsteht ein Modell einer doppelten Ausrichtung („Dual Concerns Model“) in einem Matrixfeld, das zwischen sachlicher Zielorientierung und der Orientierung an sozialen Beziehungen aufgespannt ist. Folgende Rollen schließlich werden als die relevantesten präsentiert:

- Judge (Richter/-in),
- Scientist/Researcher (Wissenschaftler/-in, Forscher/-in),
- Consultant (Berater/-in),
- Educator/Teacher (Erzieher/-in, Lehrer/-in),
- Detective (Detektiv/-in),
- Reporter (Reporter/-in),
- Facilitator (Moderator/-in, Prozessbegleiter/-in).

Vier dieser Rollenbilder – Berater/-in, Lehrer/-in, Wissenschaftler/-in und Richter/-in – wurden anschließend in Kleingruppen diskutiert. Basierend auf den Erfahrungen der Teilnehmer/-innen mit dieser Rolle wurde erörtert, in welchen Kontexten und für welche Evaluationszwecke diese Rolle besonders wichtig ist und inwiefern sie als „Leitrolle für die Evaluation“ taugt. Die Gruppen kamen zu folgenden Ergebnissen:

Für die Rolle *des Beraters/der Beraterin* wurde unterschieden zwischen Prozessberatung und Experten-/Expertinnenberatung, also fachlich-inhaltliche Beratung. Prozessberatung wurde als – manchmal wenig beachteter, aber für die Praxis umso bedeutsamerer – Teil des Evaluationsgeschehens bewertet; sie findet vor allem in der Anfangs- und Endphase von Evaluationen statt, d.h. wenn es z.B. darum geht, zu Beginn eine Evaluation aufzusetzen, Evaluationsfragestellungen zu klären etc. oder gegen Ende Ergebnisse zu kommunizieren und nutzbar zu machen. Auftraggeber/-innen erwarten sich zusätzlich oft auch eine Experten-/Expertinnenberatung, d.h. eine fachlich-inhaltliche Beratung in Bezug auf das weitere Vorgehen. Dafür spricht, dass auf diese Weise einschlägige Kenntnisse und Erfahrungen der Evaluierenden genutzt werden können; dagegen spricht allerdings eine als problematisch bewertete Rollenkonfusion. Auftraggeber/-innen kaufen unter dem Titel der Evaluation auch Beratung ein, oft wird dies aber nicht explizit gemacht. Evaluierende, die in die Berater-/Beraterinnenrolle schlüpfen, verlieren ihre formale Unabhängigkeit und können sich dadurch von weiteren Evaluationsaufträgen ausschließen. In einem Fall jedoch passt Experten-/Expertinnenberatung zu Rolle und Aufgaben von Evaluierenden, nämlich wenn es um das Vorgehen in der Evaluation und um ‚Evaluation Capacity Building‘ geht. Als Leitrolle taugt die Beratung insgesamt nur bedingt und wenn, dann vornehmlich als Prozessberatung.

Evaluationen sollen vielfach Best Practice eruieren und Lernprozesse anstoßen. Evaluierende sind daher immer wieder in der Rolle von *Lehrern/Lehrerinnen*. Ähnlich wie die Beratung setzt auch die Funktion als Lehrer/-in in der Anfangs- und Endphase an, d.h. zum Beispiel bei der Zielklärung oder dann bei der Präsentation von Ergebnissen, vor allem wenn es sich um eine formative und/oder wissensgenerierende Evaluation handelt. Dazu braucht es eine pädagogische Leistung in der Kommunikation und Vermittlung. Wichtig ist der direkte Kontakt zu den Stakeholdern der Evaluation; wie in der Lehre ist für Evaluierende die Zusammenarbeit eine Voraussetzung für eine erfolgreiche Weitergabe von Wissen und Kenntnissen. Wie die Berater-/Beraterinnenrolle ist auch die Lehrer-/Lehrerinnenrolle ein wichtiger Aspekt, stellt jedoch keine Leitrolle dar.

Evaluation als ‚wissenschaftliche Dienstleistung‘ hat eine enge Beziehung zu Forschung und Wissenschaft, die Rolle als *Wissenschaftler/-in* liegt also nahe und sie eignet sich am stärksten als Leitrolle. Trotz vielfältiger Überlappungen und breiten Grenzbereichen unterscheiden sich die beiden Rollen jedoch, insbesondere wenn es um die Differenz zwischen Grundlagenforschung und Evaluation geht. Thematisiert wurden z.B. das Spannungsfeld zwischen wissenschaftlichem Anspruch der Evaluierenden bei Fragestellungen und Methodik und den Ansprüchen der Auftraggebenden sowie den Rahmenbedingungen und Ressourcen. In der Praxis stoßen Evaluierende meist auf Grenzen der Wissenschaftlichkeit. Sie sind auch oft in der Dolmetscher-/Dolmetscherinnenfunktion, wenn es darum geht, wissenschaftliche Vorgehensweisen und Ergebnisse Auftraggebenden und Stakeholdern verständlich zu vermitteln. Angesprochen wurde auf der anderen Seite die Bedeutung der Einbettung von Evaluationen in die wissenschaftliche Forschung sowie die intrinsische forschersische Motivation von Evaluierenden: Diese erhöhen die Qualität von Evaluationen und tragen vor allem auch zum Erkenntnisgewinn aus Evaluationen

und zur Nachhaltigkeit bei. Die Verallgemeinerbarkeit von Ergebnissen ist dafür Voraussetzung.

Die Bewertung als zentrale Leistung der Evaluation führt zur Rolle *des Richters/der Richterin*. Allerdings ist die Ermittlung des Wertes eines Evaluationsgegenstands noch kein richterliches Urteil – für Entscheidungen sind Evaluierende ja nicht verantwortlich. Zusätzlich sind Evaluierende nicht alleine für die Bewertung zuständig, für gewöhnlich sind verschiedene andere Personengruppen darin eingebunden, wie z.B. Datengebende, Expertinnen/Experten, Rezipientinnen/Rezipienten und Auftraggebende; Evaluierende vergleichen und synthetisieren diese Bewertungen der Stakeholder. Weiterhin sind verschiedene Formen der Bewertung möglich: Sie kann *ex-ante*, im Verlauf (*Ongoing*) oder *ex-post* erfolgen. Kriterien der Bewertung können von Auftraggebenden, Evaluierenden und Stakeholdern gemeinsam aufgestellt werden oder sie werden ‚*autoritär*‘ von einer Partei – meist Auftraggebenden oder Evaluierenden – festgelegt. Sie können explizit sein oder implizit, als Schwellenwerte wirksam werden oder durch Skalierungen. Die Schlussfolgerungen aus Evaluationen sind daher selten so eindeutig wie Richtersprüche, sondern lassen meist Interpretationsspielraum zu. Auch die Richter-/Richterinnenrolle taugt daher nicht als Leitrolle für Evaluierende.

Podiumsdiskussion:

Wie weit können, dürfen und müssen Evaluierende gehen, um die Welt zu retten?

Die Fachtagung wurde durch eine Podiumsdiskussion zum Thema „Wie weit können, dürfen und müssen Evaluierende gehen, um die Welt zu retten?“ abgerundet. Sie wurde von Jörg Rech (CEval) moderiert. Am Podium saßen Christoph Emminghaus (Rambøll Management Consulting GmbH), Maria Gutknecht-Gmeiner (IMPULSE – Evaluation und Organisationsberatung) und Stefan Silvestrini (CEval), aber auch die anderen Teilnehmenden beteiligten sich rege an der Diskussion, die tief in die Frage der Bedeutung von eigenen Zielen und Werthaltungen von Evaluierenden führt.

Die Diskutantinnen/Diskutanten sehen in der Unterstützung der Praxis ein wichtiges Ziel von Evaluation; daraus ergibt sich auch der Sinn der Tätigkeit von Evaluierenden. Die Welt zu retten, geht ihnen aber zu weit, dies ist nur mittelbar möglich. Unterschiedliche Positionen gibt es in der Frage, welche Personen durch Evaluation unterstützt werden sollen und wie stark Evaluation aus Verbesserung und Weiterentwicklung ausgerichtet sein soll.

So kann Evaluation als Management-Support gesehen werden, d.h. sich vor allem an die Programmverantwortlichen wenden. Andere fassen die Aufgabe der Evaluierenden weiter: Sie erkennen in der Unterstützung des Programmmanagements einen wichtigen Teil des Evaluationsauftrags, sehen sich aber auch in der Pflicht, anderen Stakeholdern, insbesondere den weniger stark gehörten Betroffenen und Beteiligten, eine Stimme zu geben. In besonderer Weise sind Evaluierende den Zielgruppen eines Programms verpflichtet, aufgeworfen wird aber auch die Frage nach der Verantwortung für Programmumsetzung. Die Palette von Beteiligten und Betroffenen, denen Evaluation dienen kann und soll, ist also breit.

Auch in Bezug auf das Ziel, durch Evaluation Programme zu verbessern, zeigen sich Abstufungen: Evaluation kann so verstanden werden, dass sie vor allem Potenziale aufdecken und Programme entwickeln soll, die Bewertung dabei in den Hintergrund tritt. Eine etwas gemäßigte Sicht sieht in den Handlungsspielräumen, die Evaluation in Programmen eröffnet, den Reiz von Evaluation, variiert die verschiedenen Evaluationszwecke jedoch stärker aus. Schließlich wird die Frage in den Raum gestellt, wie es sich vermeiden lässt, dass Evaluierende implizit die Welt retten wollen. Das ‚Weltrettertum‘, dem sich Evaluierende oft nur schwer entziehen könne, birgt ja auch die Gefahr, dass Programmen die eigenen Weltsicht ‚aufgepfropft‘ wird. Dies kann gerade im interkulturellen Kontext zu Verzerrungen führen. Schließlich beinhaltet auch nicht jeder Evaluationsauftrag das Ziel der Programmverbesserung, es kann auch um Rechenschaftslegung gehen.

Es schließt sich eine Diskussion an, ob Evaluierende jedes Programm und jedes Management mit ihrer Arbeit unterstützen wollen und wie mit eigenen Werthaltungen umgegangen wird. Es gibt einige Übereinstimmung unter den Teilnehmenden, dass bereits der eigenen Positionierung und dem Expertiseaufbau als Evaluator/-in eine Wertentscheidung zugrunde liegt, die dann zu bestimmten Arten von Evaluationsaufträgen hinführt. Oder man versucht im Vorfeld zu eruieren, ob mögliche Evaluationsprojekte zu den eigenen Zielen und Werthaltungen passen, und bewirbt sich für unpassende Aufträge erst gar nicht. Ungelöst bleibt die Frage nach dem hinter diesen Erwägungen stehenden starken Spannungsfeld zwischen der von Evaluierenden geforderten Neutralität in der Datenerhebung und der Tätigkeit der Bewertung, in der man sich als Evaluator/-in einer Wertepositionierung nicht entziehen kann.

Diskutiert wird auch der Umgang mit Zielkonflikten und ethischen Dilemmata, wenn z.B. Maßnahmen für die Zielgruppen nicht (ausreichend) wirksam oder ineffizient sind, das Programm aber Arbeitsplätze geschaffen hat. Hier sollte Evaluation entsprechend ihrem Auftrag die Datenlage sprechen lassen. Betont wird in diesem Zusammenhang die Trennung von Evaluation und Entscheidungen, d.h. es obliegt den Programmverantwortlichen, in konfliktären Situationen Entscheidungen zu treffen.

Immer wieder kreist die Diskussion um zwei Themen, die in Kernbereiche der Evaluationstätigkeit führen:

Erstens wird die Interpretation und Bewertung von empirischen Daten als Grundlage der Evaluation thematisiert, d.h. wie der Prozess gestaltet ist, der zu Bewertungen führt. Es stellt sich die Frage nach Einflüssen auf Schlussfolgerungen und Empfehlungen. So können unterschiedliche Werthaltungen dazu führen, dass gleiche Daten zu verschiedenen Empfehlungen führen. Hier besteht Klärungs- und Forschungsbedarf.

Zweitens wird von Evaluierenden die Generierung von passenden Empfehlungen erwartet. Dieses Element des Evaluationsprozesses wird als Schwachpunkt der aktuellen Evaluationspraxis und auch der Evaluationstheorie ausgemacht. Je nach Arbeitsweise und Ansatz gibt es hier unterschiedliche Praktiken – von den konkreten, mit der Organisation ausgearbeiteten Handlungsempfehlungen bis zu eher formal gehaltenen Schlussfolgerungen. Wird nicht mit den Betroffenen und Beteiligten

direkt gearbeitet, so kann sich gerade in professionellen Kontexten eine gute Beschreibung, so die Erfahrung, manchmal besser eignen, um bei Programmumsetzenden Handlungsimpulse auszulösen, da sie stärker zum eigenen Denken anregt und weniger Widerstände erzeugt.

Schließlich wird das Spannungsfeld zwischen Evaluation und Beratung noch einmal aufgegriffen. Es herrscht weitgehend Einigkeit darüber, dass das Ableiten von Empfehlungen ein Handlungsfeld von Evaluation ist; wie weit Evaluierende dabei gehen sollen, ist jedoch umstritten: Wenn Handlungsempfehlungen gemeinsam mit der Organisation entwickelt werden und sie bei Bedarf in der Umsetzung unterstützt wird, erhöht dies die Nutzung von Evaluation. Unterschiedliche Meinungen gab es in Bezug auf die Frage, ob die Umsetzungstätigkeit noch Teil der Evaluation ist/sein sollen. Ein formaler Rollenwechsel zwischen Evaluation und Beratung kann stattfinden; es ist nach einer der diskutierten Positionen aber auch möglich beides zusammenzuführen: Die Grenzen sind hier fließend bzw. oft nicht genau zu erkennen. Jedenfalls brauchen Evaluierende Beratungskompetenzen, wenn sie sich so stark in die Umsetzung involvieren. Andere sehen in dieser Vorgangsweise eine ‚prozessbegleitende Beratung mit evaluatorischen Elementen‘ und grenzen Evaluation stärker von Beratung ab. Unterschieden wird auch zwischen der Moderation des Prozesses der Generierung von Empfehlungen und der konzeptuellen und operativen Mitarbeit. Dieser Graubereich der Aktivität von Evaluierenden zwischen Evaluation und Umsetzungspraxis sollte im Sinne der Professionalisierung auf jeden Fall noch weiter ausgeleuchtet und transparent gemacht werden. Dazu gehört auch, dass verstärkt thematisiert werden sollte, dass für eine sorgfältige Bewertung und Ableitung von Empfehlungen Zeit und Ressourcen eingeplant werden müssen. Dies ist aus Sicht der Teilnehmenden nur selten der Fall.

Fazit

Die Veranstaltung hat eine erste Reflexion des Rollenverständnisses von Evaluierenden ermöglicht. Sie hat aufgezeigt, dass Evaluierende im Evaluationsverlauf verschiedene Rollen annehmen und diese flexibel auf den situativen Kontext anpassen (müssen). Spannungsfelder zwischen verschiedenen Rollen wurden deutlich. Diese sind nicht ein für allemal aufzulösen, sondern in der Evaluationspraxis immer wieder zu reflektieren und gegebenenfalls auszuhandeln.

Offenkundig wurden auch die Konsequenzen eines breiteren Rollenverständnisses für das Kompetenzprofil und die Aus- und Weiterbildung von Evaluierenden. Insbesondere in den Handlungsfeldern außerhalb von methodischen Fragen der Datenerhebung und -auswertung, d.h. vor allem in den Anfangs- und Endphasen der Evaluation, kommt eine Reihe von Aufgaben und Rollen auf Evaluierende zu, die bislang wenig thematisiert wurden, für eine anspruchsvolle Evaluationspraxis jedoch hoch bedeutsam sind. Besonders die Tätigkeit des Bewertens als Grundlage der Evaluation ist wenig erforscht, ebenso wurde bislang wenig Augenmerk auf den Prozess der Ableitung von (Handlungs-)Empfehlungen gelegt. Es fehlt auch eine Auseinandersetzung mit dem Verhältnis von Evaluation und Beratung, wenn es um die Verbesserung und Entwicklung von Programmen und Organisationen geht.

Die Tagung hat große theoretische Defizite in der Evaluationsforschung rund um das Rollenverständnis deutlich gemacht. Die Teilnehmenden regten daher an, innerhalb der Gesellschaft für Evaluation Forschung zu Aufgaben und Rollen von Evaluierenden zu unterstützen. Das Thema Rollenverständnis sollte auch weiterhin auf der Agenda stehen – sei es auf Tagungen, wo es bislang meist nur implizit mitdiskutiert wurde, sei es innerhalb von Weiterbildung zu Evaluation. Dabei wäre es wichtig, sowohl Evaluierende aus anderen Kontexten (z.B. stark methodisch verankerten Evaluationsfeldern) einzubeziehen als auch Auftraggebende, um auch deren Perspektiven und Erwartungen an die Aufgaben und Rollen von Evaluierenden einzufangen.

Literatur

- DeGEval – Gesellschaft für Evaluation e.V. (Hg.) (2010): Session D 1 – Methoden der Evaluationsdidaktik – Aktivierende Teaching Tools für Evaluations-Workshops und -Seminare, Jahrestagung Luxembourg. Verfügbar unter: <http://www.degeval.de/veranstaltungen/jahrestagungen/luxemburg-2010/block-d/view/action/d-1-methoden-der-evaluationsdidaktik-aktivierende-teaching-tools-fuer-evaluations-workshops/> [11.7.2014].
- DeGEval – Gesellschaft für Evaluation e.V. (2014): Tagungsdokumentation der Frühjahrstagung des Arbeitskreises Aus- und Weiterbildung in der Evaluation. Verfügbar unter: <http://www.degeval.de/arbeitskreise/aus-und-weiterbildung-in-der-evaluation/aktuelles/> [11.7.2014].
- Gutknecht-Gmeiner, Maria (erscheint 2015): The Occupation Metaphor Exercise: Using Metaphors to Reflect upon and Clarify Tasks and Roles of Evaluators. In: Preskill, Hallie S./Russ-Eft, Darlene F. (Hg.): Building Evaluation Capacity. Thousand Oaks, CA u.a.: Sage (2. Aufl.).
- King, Jean A./Stevahn, Laurie (2002): Three frameworks for considering evaluator role. In: Ryan, Katherine E./Schwandt, Thomas A. (Hg.): Exploring evaluator role and identity, Greenwich, CT: Information Age Publishing, S. 1-16.
- Stake, Bob (2004): How Far Dare an Evaluator Go Toward Saving the World. In: American Journal of Evaluation, 25 (1), S. 103-107.

Wirkungsanalysen nach § 7, Abs. 2 Bundeshaushaltsordnung – Zwischenbilanz zur Auswirkung veränderter Anforderungen auf die Praxis der Evaluation

Frühjahrstreffen des Arbeitskreises Forschungs-,
Technologie- und Innovationspolitik der Gesellschaft
für Evaluation e.V. am 5. Juni 2014 in Berlin

Sonja Kind,¹ Marianne Kulicke,² Jan Wessels¹

Das diesjährige Frühjahrstreffen des Arbeitskreises Forschungs-, Technologie- und Innovationspolitik (FTI) widmete sich dem Thema Wirkungsanalysen nach §7, Abs. 2 der Bundeshaushaltsordnung (BHO) und VV Ziffer 2.2. In jüngster Zeit hat gerade der Bundesrechnungshof verstärkt auf diese verpflichtende Anforderung der BHO verwiesen. Die Bundesministerien befinden sich in unterschiedlicher Form in der Anpassung ihrer Vergabepaxis. Methodisch stellt die Vorgabe alle Evaluationsstudien im FTI-Bereich vor erhebliche Herausforderungen. Echte Kontrollgruppenansätze lassen sich gerade in Technologieförderprogrammen mit kleinen Zielgruppen nur eingeschränkt umsetzen, Wirtschaftlichkeitsuntersuchungen erfordern eine nur bedingt leistbare Monetarisierung von Programmwirkungen.

In der Diskussion mit Auftraggebenden von Evaluationen und Evaluationseinrichtungen wurde eine Zwischenbilanz gezogen, wie sich die veränderten Anforderungen auf die Praxis der Evaluation auswirken, welche methodischen Ansätze diesen Anforderungen gerecht werden können und welche weiteren Konsequenzen für die Vergabe und Durchführung von Evaluationen zu erwarten sind. Ein intensiver Diskurs erfolgte durch eine Podiumsdiskussion und eine vertiefte Diskussion in Gruppenarbeit, dazwischen wurden drei Praxisbeispiele von kürzlich abgeschlossenen oder noch in Bearbeitung befindlichen Evaluationen vorgestellt.

In ihrer Einführung betonte Marianne Kulicke, Fraunhofer ISI, dass Wirkungsanalysen von Fördermaßnahmen mittlerweile als notwendiges Element der Ordnungspolitik allgemein anerkannt sind, um die Relevanz der Ziele/des Förderbedarfs

1 Institut für Innovation und Technik (iit) in der VDI/VDE Innovation + Technik GmbH, Berlin

2 Fraunhofer Institut für System- und Innovationsforschung (ISI), Karlsruhe

zu messen, die Effektivität des Förderinstrumentariums und die Wirtschaftlichkeit des Mitteleinsatzes zu bewerten sowie Hinweise für die Anpassung des Programmsatzes an Veränderungen im Förderfeld bis hin zur Einstellung einer Maßnahme zu erhalten.

Die meisten Wirkungsanalysen mit geänderten Anforderungen sind derzeit vom Bundesministerium für Wirtschaft und Energie (BMWi) beauftragt worden. Hierauf ging Ute Hörrmann (BMWi) in einem Überblicksvortrag ein. Sie führte aus, welche inhaltlichen Implikationen sich aus der Bundeshaushaltsordnung, den Prüfungen des Bundesrechnungshofes sowie Anforderungen aus dem Bundestag und der EU-Kommission für die Inhalte von Evaluationen bzw. Erfolgskontrollen ergeben. Letztere umfassen eine Kontrolle der Zielerreichung, der Wirkungen und der Wirtschaftlichkeit. In zwei umfänglichen Querschnittsprüfungen konstatierte der Bundesrechnungshof erhebliche Defizite bei Erfolgskontrollen und mahnte Änderungen in ihrer Durchführung an.

Ergänzend dazu zeigte Bernd Hausberg, VDI Technologiezentrum GmbH, die hohen Ansprüche auf, die sich aus den EU-Leitlinien 2014 für die Bereiche Monitoring und Evaluierung im Rahmen der Innovationsförderung mittels vielfältiger Maßnahmen auf EU-Ebene ergeben. Er verwies darauf, dass die EU-Kommission immer stärker auf ein differenziertes Evaluierungsschema setzt, mit unterschiedlichen Formen im Lebenszyklus einer Fördermaßnahme und einem breiten Einbezug von Betrachtungsebenen und Stakeholder-Gruppen. Ziel ist die Wirkungsevaluierung als *ein* integriertes Instrument zu verstehen und einzusetzen. Die dabei postulierte Notwendigkeit eines Kontrollgruppenansatzes sieht er jedoch kritisch, da die Interventionslogik vieler EU- oder BMBF-Maßnahmen dessen Umsetzung eigentlich ausschließt, da es an nicht geförderten, aber vergleichbaren Einheiten für eine Kontrollgruppe mangelt.

Die Podiumsdiskussion ging zunächst auf die Implikationen geänderter Anforderungen an Evaluationen für die Fördergeber und Projektträger ein. Sie betreffen (1) das laufende Programm-Monitoring mit einer systematischen Datenerfassung für die spätere Erfolgskontrolle, (2) die Festlegung komplexer Zielwerte vor Programmstart sowie (3) Lernprozesse bei der Beauftragung von Erfolgskontrollen und deren Umsetzung. Diskutiert wurde danach über die Herausforderungen variierender Interventionslogiken angesichts der Bandbreite an Fördermaßnahmen. Durch hohen Druck zum Nachweis der Wirksamkeit steigt für jedes einzelne Programm die Notwendigkeit, Erfolgskontrollen zur Legitimation der Fördermittel durchzuführen. Bislang fehlen meist Leistungsindikatoren zu den Zielen vieler Maßnahmen der FTI-Politik. Doch sollten solche Indikatoren zumindest intern beim Fördergeber vorhanden sein, damit eine Wirkungskontrolle durchführbar ist. Dazu sollte ein Kernset an Indikatoren gehören, das auf die Ausrichtung des jeweiligen Ministeriums abstellt und weitere programmspezifische Indikatoren enthält. Ausführlich wurde auch über die Grenzen und Möglichkeiten eines Kontrollgruppenansatzes diskutiert sowie über die Übertragbarkeit des Evaluationssystems der EU. Dieses macht bei Programmstart relativ detaillierte Vorgaben für Ex-ante-Betrachtungen, ein laufendes Monitoring und Ex-post-Untersuchungen. Es handelt sich dabei aber häufig um lang laufende und großvolumige EU-Maßnahmen mit weit besseren Möglichkei-

ten der Wirkungsmessung, als dies bei typischen Innovationsförderprogrammen der Fall ist.

Stefan Ekert und Tim Grebe, InterVal, stellten als Projektbeispiel 1 die Evaluation des Programms MNPQ-Transfer zur Unterstützung kleiner und mittlerer Unternehmen bei der Umsetzung von Innovationen in den Bereichen Messen, Normen, Prüfen und Qualitätssicherung vor. Das Programm fördert den Technologietransfer aus drei Bundesanstalten mit FuE-Aufgaben und weist ein niedriges Programmvolumen und nur 116 Projekte seit 2006 auf. Vertieft wurden die Themen Wirkungsmessung und Wirtschaftlichkeitsuntersuchung angesichts dieser Programmkonstellation sowie begrenzter finanzieller und zeitlicher Spielräume. Die Vortragenden stellten mögliche Zugänge zu einer Zielerreichungs- und Wirksamkeitsuntersuchung mit ihren Grenzen und Möglichkeiten vor und zogen den Schluss, dass auch mit zeitlichen und finanziellen Beschränkungen eine Bewertung der Wirksamkeit und Wirtschaftlichkeit kleiner Programme möglich ist. Dies könne insbesondere durch Kombination von qualitativen und quantitativen Erhebungsschritten sowie durch Vergleich/Benchmarking mit anderen Programmen erfolgen.

Marianne Kulicke, Fraunhofer ISI, skizzierte im Projektbeispiel 2 den Evaluationsansatz und die Datenerhebungsverfahren zur Erfolgskontrolle von SIGNO – Schutz von Ideen für die Gewerbliche Nutzung des BMWi. Das Programm stellt einen Mix aus Förderportfolio (Angebote für unterschiedliche Zielgruppen) und Instrumentenportfolio (verschiedene Instrumente für gleiches Förderziel) dar. Es gab vier separat zu evaluierende Förderbereiche. Dazu wurden im sechsmonatigen Bearbeitungszeitraum zunächst für jeden Bereich ein Wirkungsmodell und ein daraus abgeleitetes indikatorgestütztes Evaluationssystem erarbeitet. Letzteres bildete die Basis für die Ex-post-Analyse und Wirtschaftlichkeitsbetrachtung. Im Mittelpunkt der Studie stehen fünf Online-Befragungen von Hochschulen/außeruniversitären Forschungseinrichtungen, von sogenannten patentrelevanten Wissenschaftler(inne)n, Patent- und Verwertungsagenturen (PVAs) sowie geförderten KMU, nicht geförderten KMU mit Patenterfahrung und ohne Patenterfahrung. Es wurde ein dezidierter Vergleichsgruppenansatz umgesetzt. Ausgangspunkt zur Ableitung des Wirkungsmodells von SIGNO waren die Programmziele, die jedoch keine direkt messbaren, quantitativen Zielgrößen beinhalten. Auch gab es weder Leistungsindikatoren zu den Zielen noch eine Vorgängerevaluation im Sinne einer Erfolgskontrolle. Zu bestimmten Outputgrößen lag eine differenzierte Datenbasis aus dem Programm-Monitoring des Projektträgers vor. Da beim Programmstart kein Konzept für die Erfolgskontrolle erstellt wurde, gab es ansonsten keine, durch ein laufendes Monitoring erhobene Daten. Es musste nachträglich ein Evaluationssystem entwickelt und eine Datenerhebung zu früheren Ereignissen durchgeführt werden. Die besondere Herausforderung bestand jedoch darin, dass zum Teil schwer zu untersuchende und mit deutlichem Zeitverzug auftretende Wirkzusammenhänge evaluiert werden mussten und komplexe, vielfältige Wirkungen durch den Förderansatz möglich waren. Das Fraunhofer ISI verfolgte dazu einen Mehrsichtenansatz, setzte unterschiedliche quantitative und qualitative Methoden ein und untersuchte eine große Bandbreite an Wirkungsdimensionen. Eine besondere Anforderung stellt die Zusammenführung zu

einer Gesamtaussage zur Wirtschaftlichkeit des Programms mit dem Versuch einer Monetarisierung der Programmwirkungen dar.

Britta Oertel und Lydia Illge, IZT – Institut für Zukunftsstudien und Technologiebewertung, berichteten im Projektbeispiel 3 über ihre Vorgehensweise bei der Evaluation der Projektförderung des BMWi in der Energieforschung, Fachbereich Energieeffizienz in Industrie, Gewerbe, Handel und Dienstleistung (IGHD) im Rahmen des 5. Energieforschungsprogramms. Zum Einsatz kam ein Methodenmix zur Wirkungsanalyse, der aus der Auswertung der Vorhabendaten mithilfe inhaltsanalytischer Verfahren und statistischer Analysen, aus einer standardisierten Befragung, einem Expertenworkshop sowie aus Fallstudien zu Schlüsselprojekten bestand. Untersucht wurden zwischen 2003 und 2012 beendete Projekte. Dies ermöglichte die Beobachtung von Langzeitwirkungen, erschwerte aber den Zugang zu den früheren Fördernehmern und damit belastbare Aussagen zu lange zurückliegenden Projekten. Die Evaluation konzentrierte sich bei der Wirkungsanalyse insbesondere auf Soll-Ist- und Vorher-Nachher-Vergleiche. Nach Einschätzung der Vortragenden war der eingesetzte Methodenmix (Datenanalysen, Online-Befragung als Vollerhebung, Fallstudien) auf die Evaluationsfragestellungen des BMWi ausgerichtet und hat sich bewährt.

Der Workshop schloss mit einer Arbeitsphase in drei Kleingruppen zu den Themen Standardisierung des Indikatorensets, Kontrollgruppen und Datensätze. Bei der Diskussion des ersten Aspekts wurde zunächst dessen Sinnhaftigkeit infrage gestellt und konstatiert, dass lediglich eine Festlegung auf einige wenige Indikatoren sinnvoll ist. Diese sollten bereits bei der Programmplanung bestimmt und die Interventionslogik der FTI-Maßnahmen im Vorfeld durchdacht werden. Anschließend sollte diese Logik im Kontext des Innovationssystems analysiert sowie Wirkungsmodelle erarbeitet werden. Mit ausgeprägten Zweifeln an der prinzipiellen Aussagekraft von Kontrollgruppen startete diese Arbeitsgruppe. Anschließend wurden konkrete Optionen und Rahmenbedingungen für einen solchen Ansatz diskutiert. Als zentrale Herausforderung zum dritten Thema wurde eine Klärung des Zugangs zu und der Verfügbarkeit von Daten für Evaluationsvorhaben benannt, idealerweise bereits vor Vergabe einer Evaluation. Einen breiten Raum nahmen die Erfahrungen über Grenzen und Potenziale vorhandener Sekundärdatenbestände bei Fördergebern, der amtlichen Statistik und weiterer Quellen ein.

Standards in der Evaluation beruflicher Bildung: Anspruch und Wirklichkeit

Frühjahrstagung des Arbeitskreises Berufliche Bildung im Mai 2014

*Michael Kalman*¹

Am 16. Mai 2014 fand die Frühjahrstagung des Arbeitskreises Berufliche Bildung an der Humboldt Universität (HU) zu Berlin statt. Knapp 20 Expertinnen und Experten tauschten sich zum Thema „Standards in der Evaluation beruflicher Bildung: Anspruch und Wirklichkeit“ aus.

Das Thema ist aktuell, befindet sich die Gesellschaft für Evaluation e.V. (DeGEval) doch in einem Revisionsprozess ihrer Standards, die inzwischen ein wenig ‚in die Jahre gekommen‘ sind. Im Oktober 2001 wurden die „Standards für Evaluation“ von der Mitgliederversammlung einstimmig angenommen, 2002 erfolgte dann die Veröffentlichung. Obwohl die Revision seinerzeit schon mitgedacht worden war, dauerte es doch 11 Jahre bis zum Mitgliederbeschluss von 2013, bis der Revisionsprozess tatsächlich angestoßen wurde. Die Tagung verstand sich so auch als Inputgeber für die Revision.

Diese Inputs sollten aus der Praxis kommen. Daher wurden die Referent(inn)en Rüdiger Preißer und Tülin Engin gebeten, bereits durchgeführte und abgeschlossene Evaluationsprojekte vorzustellen und retrospektiv hinsichtlich der Relevanz und des Erfüllungsgrades der Standards für Evaluation zu bewerten.

Michael Kalman gab zu Beginn eine kurze Einstimmung in den „generischen Anforderungskatalog“ der 25 Standards, die in die vier Gruppen Nützlichkeit, Durchführbarkeit, Fairness und Genauigkeit aufgeteilt sind, und schlussfolgerte, dass „der generische Charakter der Standards [...] mitunter hohe Herausforderungen an die Evaluierenden“ stellt, „um die Anforderungen auf ihr konkretes Evaluationsprojekt zu beziehen.“ Schließlich beanspruchen die Standards Geltung für viele, wenn nicht für alle Felder, Akteure, Zwecke, Evaluationsgegenstände und Evaluationsarten. Und: Sie schreiben keine bestimmte Methode bzw. einen spezifischen Methodenmix vor.

1 Kalman Consult, Berlin

Zur Adressierung dieser Herausforderungen stellt die Standards-Broschüre Interpretations- und Anwendungshilfen bereit, nämlich Einleitung/Vorwort, Standards-Erklärungstexte, ein funktionales Inhaltsverzeichnis und eine Transformationstabelle zu den US-amerikanischen Joint Committee on Standards und den Standards der Schweizerischen Evaluationsgesellschaft (SEVAL). Die Brückenfunktion zur Evaluationspraxis dieser Interpretations- und Anwendungshilfen – so verdienstvoll sie sind – wurde kritisch hinterfragt. Die zuweilen schwierige Aufgabe der Abwägung bei der Auslegung und Verwendung der Standards kann dem Evaluationsteam freilich nicht abgenommen werden. Dies umso mehr, als sich einige Standards durchaus widersprechen können, z.B. aus den Gruppen Durchführbarkeit und Genauigkeit. Es wird stets angesichts begrenzter Ressourcen und bezogen auf Evaluationsgegenstand und Fragestellung abzuwägen sein, wie genau das Evaluationsteam sein muss, um auf der einen Seite die Gütekriterien der empirischen Sozialforschung zu erfüllen und auf der anderen Seite die Durchführbarkeit des Evaluationsprojekts angesichts häufig knapper Ressourcen nicht zu gefährden. Auch ist zu fragen, ob in jedem Fall eine Evidenzbasierung erforderlich ist, um die Nützlichkeit der Evaluationsergebnisse für den Auftraggeber sicherzustellen.

Die Entlastung in diesem schwierigen Geschäft der Abwägung und Auslegung ist gleichwohl in Sicht, weil die Nichtanwendung bestimmter Standards ausdrücklich möglich ist; auch darf und muss die Art und Weise der Erfüllung von Standards variieren und schließlich: Die Standards sind aktuell keine Grundlage für Dienstleistungszertifizierungen, was jedoch auch den Vorteil hat, dass kein verengender Formalismus vorliegt.

Insgesamt sah Kalman in der Ableitung von handlungsleitenden Checklisten aus dem funktionalen Inhaltsverzeichnis eine Möglichkeit, die „Brücke zur Praxis“ flankierend zu den Standards zu stärken.

Für die Tagung wurden zwei Perspektiven aufgezeigt, nämlich die Überprüfung des praktischen Umgangs mit den Standards und die Überprüfung der Standards in ihrer Relevanz für die Praxis (Revision). Den beiden Referent(inn)en wurden die folgenden Fragen an die Hand gegeben:

- Hat der Auftraggeber die Berücksichtigung der Standards vorgegeben?
- Wenn ja, wie hat der Auftraggeber die Einhaltung der Standards kommuniziert/kontrolliert?
- Welche Relevanz/welchen Stellenwert hatten die Standards für die verschiedenen Projektphasen (Konzept, Einstiegsphase, Durchführungsphase, Berichterstattung, Nachprojektphase)?
- Welche expliziten und impliziten Rekurse auf die Standards wurden im Evaluationsprojekt vorgenommen?
- In der Retrospektive: Welche Standards wurden erfüllt, welche nicht und warum?

Der Erziehungs- und Sozialwissenschaftler *Rüdiger Preißer* präsentierte „Schlaglichter auf die Standards für Evaluation aus der Perspektive ausgewählter Evaluationsprojekte“. Preißer stellte folgende Projekte vor:

- Qualität der Lehre und Studienbedingungen an der HU Berlin (1990-93);
- ProfilPass (DIE, DIPF, 2004/05);

– Berufsvorbereitende Bildungsmaßnahmen (BvB, 2009/10).

Bereits zu Beginn stellte er klar, dass er Standards für Evaluation nie zur Grundlage genommen habe. Stattdessen waren für ihn die Regeln der empirischen Sozialforschung handlungsleitend.

Im Rahmen des Projekts Qualität der Lehre und Studienbedingungen an der HU Berlin (1990-93) wurden Studierende aus 14 Studiengängen mit insgesamt ca. 8000 Bewertungsbögen zu ihrer Lehrveranstaltung befragt. Auf diese Weise wurden ca. 300 Lehrveranstaltungen (LV) bewertet. Zu den erhobenen Merkmalen gehörte der Input (Infrastruktur, Rahmenbedingungen der LV), der Prozess (Durchführung) und der Outcome (Lernfortschritt, Studienverhalten). Aus den Befragungsergebnissen wurden verschiedene Auswertungen vorgenommen, z.B. zu Strukturvergleichen zwischen LV-Typen, zu Studiengängen, Studienabschnitten und Studierenden-Typen im Sinne soziodemografischer Merkmale. Ferner wurden die Bewertungsdimensionen Input und Durchführung sowie der Outcome im Sinne des subjektiven Lernerfolgs ermittelt.

Preißer setzte sich kritisch mit dem Untersuchungsdesign auseinander – zunächst immanent. So kritisierte er die zu starke Konzentration auf die Outcome-Dimension bei Vernachlässigung des Inputs. So hätte weder ein Curriculum noch ein didaktisches Konzept existiert; die Qualifikation der Lehrenden sei ebenfalls nicht vorhanden gewesen. Zudem sei die Outcome-Dimension nur aus der Zielgruppenperspektive im Sinne des subjektiv bewerteten Lernerfolgs erhoben worden; dasselbe gelte für die Prozessqualität, wo „veraltete Indikatoren“ verwendet worden seien.

Auf einer grundsätzlicheren Ebene lautete seine Kritik: „Wieso konnte das Ergebnis nicht lauten, dass Lehre prinzipiell falsch angelegt ist: weitgehend dem traditionellen Vermittlungsmodell von Wissen folgt, anstatt auf Grundlage konstruktivistischer Lerntheorie („Shift from Teaching to learning“) zu operieren?“ Sein skeptischer Dreiklang bezogen auf die Standards für Evaluation – die es freilich im deutschsprachigen Raum in den 1990er Jahren so noch nicht gegeben hatte:

- Sind die Standards verwendet worden?
- Hätten die Standards etwas geändert (am Resultat, an der Durchführung)?
- Hätten andere Standards etwas geändert?

Der Referent sprach damit die Steuerungsmöglichkeiten von Standards in Bezug auf die Qualität konkreter Evaluationsprojekte an. Insgesamt – so kann implizit aus dem Referat von Preißer geschlossen werden – würde die Anwendung der Standards nicht bewirkt haben, dass im Evaluationsprojekt die Input-Dimension stärker beachtet oder die Grundanlage des Evaluationsgegenstandes prinzipiell hinterfragt worden wäre.

In einem weiteren Projekt, die Evaluation des ProfilPasses im Jahr 2004/05, sollten Entscheidungsgrundlagen über die bundesweite Einführung dieses Instrumentes erbracht werden. Zudem sollten Anhaltspunkte für Modifizierungen und Verbesserungen generiert werden. Der ProfilPass ist ein Portfolio-Instrument zur Exploration und Dokumentation der individuellen Kompetenzen. Es wurde im Rahmen der Evaluation eine standardisierte schriftliche Befragung der Nutzer(innen), teilstandardisierte schriftliche Befragung der Teilnehmenden an der Qualifizierung zur ProfilPass-Beratung und leitfadengestützte Interviews mit Berater(inne)n durchge-

führt. Die Evaluation erbrachte verschiedene Ergebnisse zur Outcome-Dimension insgesamt und für Personengruppen, ermöglichte z.B. Strukturvergleiche zwischen Outcome-Dimensionen und Input-/Durchführungs-Dimensionen und gewährleistete Hypothesengenerierung aufgrund der Bewertungen der Berater(innen). Für Preißer war der Evaluationsgegenstand, ein „übliches“ Verfahren, „richtig“ entwickelt und „richtig“ evaluiert; aber wenn das Verfahren losgelöst von Zielen eingesetzt werde, drohe eine Verselbstständigung. Der Referent bezweifelte zudem, ob Schlussfolgerungen aus der Evaluation für die weitere Entwicklung/die endgültige Einführung des ProfilPasses gezogen wurden. Letzteres wird ja durch den Standard N8 „Nutzung und Nutzen der Evaluation“ angesprochen.

Im Jahr 2009/10 evaluierte Preißer Berufsvorbereitende Bildungsmaßnahmen (BvB) durch die Erfassung (Rating) der „erworbenen Kompetenzen“ bei rund 300 Jugendlichen („Wirkungskontrolle“). Dabei erfolgten kein Vergleich mit den Anfangskompetenzen und auch keine direkte Befragung der Zielgruppe. Zusätzlich wurden Fallstudien mit Experteninterviews und Dokumentenanalysen bei 10 Bildungsträgern zu Maßnahmenkonzeption, -durchführung und -resultaten durchgeführt. Offiziell sollte im Rahmen einer „Wirkungskontrolle“ evaluiert werden, ob es bei den Teilnehmer(inne)n einen Kompetenzzuwachs gegeben hat und ob ein Ausgleich der Kompetenzdefizite erfolgte. Inoffiziell ging es auch um die Input- und Durchführungsqualität. Preißer warf in diesem Zusammenhang das Verhältnis von Wirkungsevaluation zur Input-Evaluation auf und konstatierte, dass eine Wirkungskontrolle, die den Namen verdient, in diesem Projekt nicht möglich war. Er bemängelte, dass es kaum Evaluationen des Übergangssystems gebe und zeigte sich skeptisch, ob bisherige Evaluationen irgendetwas bewirkt haben.

Insgesamt beschlich den Referenten bei allen drei Projekten ein Unwohlsein, obwohl diese Vorhaben methodisch korrekt durchgeführt wurden. Und er fragte sich, inwieweit die Anwendung von Standards diese Projekte auf ein ihm angemessenes erscheinendes Qualitätslevel hätten heben können.

Anhand seiner Projekte legte Preißer dar, dass der Zuschnitt des Evaluationsgegenstandes unzureichend oder gar falsch sein kann – etwa, wenn die Input-Dimension ausgenommen wird. Der dafür vorgesehene Standard G 1 „Beschreibung des Evaluationsgegenstandes“ – so möchte der Verfasser hinzufügen – entlässt das Evaluationsteam ja nicht aus der Verantwortung, den Evaluationsgegenstand entsprechend der selbst aufgestellten oder vorgegebenen Kriterien zu bestimmen. Dieser Standard geht jedoch nicht so weit, dass er Detailvorgaben macht – etwa im Sinne: Die Input-Dimension müsse bei jedem Evaluationsgegenstand mit bedacht werden.

Preißer bemängelte, dass das Schema Input – Aktivitäten – Output – Outcome, wie es etwa der Programmbaum der Firma Univation impliziere, so von den Standards für Evaluation nicht nahegelegt werde. Aus seiner Sicht seien die Standards auch zu vage formuliert. Er empfahl, die Standards besser zu operationalisieren. Ferner sollte man Mindeststandards/Gütekriterien formulieren. Dabei sollte man sich vom Prinzip der Gleichgewichtigkeit der Standards verabschieden. Insgesamt unterstütze er aber eher eine Initiative für die Erarbeitung von Grundlagen für bessere Evaluationskonzeptionen anstatt die Revision der Standards.

In der Summe diskutierte Preißer hochrelevante Aspekte, welche die Qualität von konkreten Evaluationsprojekten betreffen, die aber durch das gröbere Raster der Standards teilweise durchzufallen scheinen.

Die Psychologin *Tülin Engin* vom *uzbonn* – Gesellschaft für empirische Sozialforschung und Evaluation referierte zu „Nutzung und Nutzen der DeGEval-Standards zur Meta-Evaluation – Ein Erfahrungsbericht“. Frau Engin stellte dabei eine vom Bundesinstitut für Berufsbildung (BIBB) in Bonn beauftragte Meta-Evaluation vor. Dabei sollte eine abgeschlossene Evaluation noch einmal bewertet werden. Letztendlich handelte es sich um eine kriteriengeleitete Dokumentenanalyse, denn die Informationsbasis für die Meta-Evaluation bestand aus dem Abschlussbericht der Evaluation und dem entsprechenden Kurzbericht sowie der Leistungsbeschreibung der damaligen Ausschreibung des BIBB. Diese Informationsbasis wurde ergänzt durch Abstimmungsgespräche mit dem BIBB als Auftraggeber. Die Aufgabenstellung der Meta-Evaluation umfasste:

- kritische Reflexion der Vorgehensweise und Schlussfolgerungen im Hinblick auf die definierten Fragestellungen der Evaluation;
- Fokus auf Methoden/Vorgehensweise der Evaluation mit der Frage: Inwiefern waren die Methoden geeignet, um eine valide Informationsbasis zu erarbeiten und auf Basis der Ergebnisse begründete Empfehlungen auszusprechen;
- darüber hinaus Bewertung, inwieweit die vorgelegten Dokumente aktuellen Standards und Methoden der empirischen Sozialforschung entsprechen.

Zunächst machte die Referentin deutlich, dass die Einhaltung der Standards für Evaluation in der Ausschreibung nicht eingefordert wurde. Engin reflektierte dann anhand sämtlicher 25 DeGEval-Standards für Evaluation, inwieweit diese für ihr Projekt der Meta-Evaluation geeignet oder handlungsleitend gewesen wären. Diese Herangehensweise ist sehr relevant, enthalten die Standards für Evaluation selbst doch den Standard G9 „Meta-Evaluation“, wo es im Erläuterungsteil heißt: „Im Rahmen einer Meta-Evaluation können die hier vorliegenden Standards eingesetzt werden [...].“ (DeGEval 2004: 36). Die Standardformulierung selbst legt implizit als wichtigste Informationsbasis für Meta-Evaluationen den Evaluationsbericht und begleitende Dokumente fest: „Um Meta-Evaluationen zu ermöglichen, sollen Evaluationen in geeigneter Form dokumentiert werden“ (DeGEval 2004: 35). So stellte die Referentin heraus, dass insbesondere die Nützlichkeitsstandards und die Genauigkeitsstandards gut für die Meta-Evaluation nutzbar sind. Dies verwundert nicht, sind in diesen beiden Standardgruppen doch auch die meisten Vorgaben für die Evaluationsberichterstattung enthalten. Insbesondere N6 „Vollständigkeit und Klarheit der Berichterstattung“ ist für Engin im Kontext der Meta-Evaluation zentral. Schließlich ist der Evaluationsbericht doch faktisch der zentrale Evaluationsgegenstand. Gerade der angesprochene Standard G9 „Meta-Evaluation“ erfordere die Orientierung am Standard N6. Dasselbe gelte für die gesamte Gruppe Genauigkeit, insbesondere aber für die Standards G6 „Systematische Fehlerprüfung“ und G7 „Analyse qualitativer und quantitativer Informationen“. Die Referentin konstatierte überhaupt deutliche Überschneidungen zwischen den Standardgruppen N und G, die auch als Leitlinien bei der Erstellung von Evaluationsberichten dienen könnten.

Auf der anderen Seite eigneten sich verschiedene andere Standards kaum zur qualitativen Bewertung eines Evaluationsberichts – und könnten dementsprechend auch nicht eingehalten werden. So sei Standard D2 „Diplomatisches Vorgehen“ kaum aus einem Bericht heraus zu beurteilen. Dasselbe gelte für Standard F2 „Schutz individueller Rechte“, N3 „Glaubwürdigkeit und Kompetenz der Evaluierenden“ und N5 „Transparenz von Werten“. In der Anforderung von N5 „Die Perspektiven und Annahmen der Beteiligten und Betroffenen, auf denen die Evaluation und die Interpretation der Ergebnisse beruhen, sollen so beschrieben werden, dass die Grundlagen der Bewertungen klar ersichtlich sind“ sieht Engin zwar eine gute Grundlage für die Meta-Evaluation, würde diesen Standard aber eher N6 „Vollständigkeit und Klarheit der Berichterstattung“ zuschlagen. Ansonsten seien Perspektiven, Werthaltungen und implizite Annahmen aus einem Evaluationsbericht schwer herauszulesen.

Die Referentin zog weitere Schlussfolgerungen, die teilweise auch für den aktuellen DeGEval-Revisionsprozess der Standards von Nutzen sein könnten. So konstatierte sie, dass die definierten Einzelstandards nicht vollständig trennscharf zueinander seien und führte als Beispiele N4 „Auswahl und Umfang der Informationen“ und D1 „Angemessene Verfahren“ an. Eingedenk der Tatsache, dass die Standards sich gegenseitig bedingen und teilweise in einem hierarchischen Verhältnis zueinander stehen, schlug sie vor, die Standards in ein logisches Modell oder ein Zielsystem zu überführen. Als Beispiel führte sie auf: Leitziel N8 „Nutzung und Nutzen der Evaluation“, Mittlerziel N6 „Vollständigkeit der Berichterstattung“ und als Handlungsziele die Genauigkeitsstandards.

Die regen Diskussionen reflektierten die Möglichkeiten und Grenzen der Standards für Evaluation und zeigten Verbesserungspotenziale auf. Hierzu in der gebotenen Kürze einige Schlaglichter:

Die Standards wurden als Reflexionsinstrument für eigene Evaluationsprojekte von den Referent(inn)en durchaus gewürdigt. Ferner stand das Thema Ziele im Fokus, zum einen die Beachtung der Ziele des zu evaluierenden Gegenstands, wie sie im Standard G1 „Beschreibung des Evaluationsgegenstandes“ nicht angesprochen ist. Zum anderen wurde gefordert, dass die Standards selbst mehr als Ziele formuliert werden sollten, siehe z.B. N2 „Klärung der Evaluationszwecke“ und G3 „Beschreibung von Zwecken und Vorgehen“.

Das Thema „Auftraggebende – Auftragnehmende“ wurde verschiedentlich diskutiert: Hier bestehe ein „Ungleichgewicht“, da vor allem die Auftragnehmenden angesprochen werden, obwohl die Standards auch für die Auftraggebenden gelten. Es wurde für den Revisionsprozess angeregt, den Erläuterungsteil zu ergänzen im Sinne: Was bedeutet dieser Standard für den Auftragnehmenden, was bedeutet er für den Auftraggebenden? Weitere Forderungen waren, dass Auftraggebende und Auftragnehmende im Vorfeld eines Evaluationsprojekts die Standards durchgehen sollten. Hier wurde auch ein Verfahren, um zu einem Konsens zu kommen, propagiert.

Hinsichtlich des Standards G1 „Beschreibung des Evaluationsgegenstandes“ sollte – so ein(e) Teilnehmer(in) – auch dargelegt werden, was alles nicht beschrieben wird.

Die von Engin beschriebene Hierarchie innerhalb der Standards – obwohl diese eigentlich nicht intendiert ist – wurde als sinnvoll erachtet. Ein anderer Teilnehmer regte an, dass Formulierungen wie „soweit wie möglich“ ersetzt werden sollten durch „angemessen“. Schließlich wurde das funktionale Inhaltsverzeichnis als nicht sehr hilfreich bezeichnet.

Literatur

DeGEval – Gesellschaft für Evaluation (Hg.) (2004): Standards für Evaluation. Alfter (3. Aufl.).

Mit neun disziplinübergreifenden Beiträgen bietet der Band ein breites Panorama zum Thema Verantwortung und vereint Perspektiven aus Wissenschaft und Praxis. Im ersten Teil werden Begriffe und Betrachtungsweisen diskutiert. Im zweiten Teil geht es um Verantwortung, Risiko und Innovation in Organisationen. Im dritten Teil thematisieren die Autorinnen und Autoren die Übernahme von Verantwortung und Eigenverantwortung in und durch Unternehmen.



Nino Tomaschek,
Andreas Streinzer (Hrsg.)

Verantwortung

Über das Handeln in einer
komplexen Welt

University – Society – Industry, Band 3
2014, 158 Seiten, br., 24,90 €
ISBN 978-3-8309-3163-8
E-Book: 21,99 €;
ISBN: 978-3-8309-8163-3



WAXMANN

Evaluierung von Politikberatung – Advocacy, Policy Advice und Good-Governance- Vorhaben staatlicher und nicht staatlicher Organisationen

Bericht über die Frühjahrstagung des Arbeitskreises
Entwicklungspolitik 2014 am 5. und 6. Juni 2014 in
der Physikalisch-Technischen Bundesanstalt (PTB) in
Braunschweig

Susanne Neubert,¹ Jens Koy,¹ Stefan Silvestrini,¹ Daniela Richter²

Hintergrund und Tagungsprogramm

Welche Methoden eignen sich zur Evaluation von Politikberatung? Anhand welcher Kriterien kann Advocacy-Arbeit evaluiert werden? Welche Daten sind erforderlich, um die Wirksamkeit von Good-Governance-Vorhaben zu bewerten? Diesen und weiteren Fragen zum Thema Evaluierung von Politikberatung widmete sich der AK Entwicklungspolitik auf seiner diesjährigen Frühjahrstagung. Präsentationsformate wurden mit interaktiven Werkstattgesprächen kombiniert, bei denen politische Stiftungen ihre Evaluierungsarbeit zur Debatte stellten. Neben dem offiziellen Tagungsprogramm bot die PTB als Gastgeberin Führungen durch ihre Labore an und einen abendlichen Stadtrundgang mit gemütlicher Einkehr zum Abendessen, das Gelegenheit zu weiteren Gesprächen bot.

Der erste Tag des Fachprogramms begann mit einem Vortrag von Franziska Krisch (DEval) und Bernward Causemann (FAKT) über Evaluierung von Politikbeeinflussung durch zivilgesellschaftliche Organisationen und wurde mit vier parallelen Werkstattgesprächen weitergeführt. In den zweiten Tag stieg Susanne von Jan (CEval) mit einem Vortrag über eine Wirkungsevaluierung der Advocacy-Arbeit für Kinderrechte ein. Anschließend erörterte Frau Dr. Justine Hunter (GIZ)

1 Sprecherteam des Arbeitskreises Entwicklungspolitik: Susanne Neubert (SLE – Seminar für Ländliche Entwicklung, Berlin), Jens Koy (Freier Gutachter und Evaluator), Stefan Silvestrini (Centrum für Evaluation – CEval, Saarbrücken)

2 SLE – Seminar für Ländliche Entwicklung, Berlin; Protokollantin der Veranstaltung

Überlegungen zur Indikatorenformulierung für Good-Governance-Vorhaben. Zum Schluss folgte eine Plenumsdiskussion nach Fishbowl-Methodik.³

Evaluierung von Politikbeeinflussung durch zivilgesellschaftliche Organisationen

Frau Krisch und Herr Causemann stellten eine *Querschnittsanalyse* von Evaluationen im Bereich Politikbeeinflussung vor: Advocacy-Arbeit, Lobbying, Politische Bildung, Kampagnen. Die Evaluation von Politikbeeinflussung sei eine große Herausforderung, so die Redner/-innen, da Politik kaum vorhersagbar sei. Die Beeinflussung durch Initiativen und Projekte strebe vergleichsweise langfristige Ziele an, die selten vollständig erreicht würden. Nicht zuletzt sei die Erfolgsbewertung durch die Beteiligten subjektiv, die Zuordnung und die Definition des jeweiligen Beitrags daher noch schwieriger als üblich. Weiterhin würde die Arbeit der Evaluatoren/Evaluatorinnen durch teils mangelnde Gesprächsbereitschaft der Gesprächspartner/-innen erschwert, die sich oft in dem Konflikt befänden, dass eine Veröffentlichung erfolgreicher Strategien eine Gefahr für die Betroffenen darstellt oder dass mit der Evaluation Dinge gemessen würden, die die politischen Akteure nicht unbedingt gemessen haben möchten. Zentral bei solchen Evaluationen sei eine Darstellung des wechselnden Beziehungsgeflechts zwischen den Akteuren und deren Einflusskanälen. Sozialwissenschaftliche Theorien des Wandels (wie die der *Power Politics* oder des *Window of Opportunity*) könnten Veränderungen erklären und zeigen, ob sich die Organisation auf einen angemessenen Wirkungszusammenhang bezieht und auf welche Bereiche sie sich fokussiert. Eine begleitende Prozessdokumentation helfe, Erfolgsfaktoren herauszuarbeiten. Die Autoren/Autorinnen betonten, dass zur Datenerhebung die Sicht möglichst vieler Stakeholder-Gruppen dialogisch und mit Reflexionsschleifen eingeholt werden sollte. Einen Hinweis, ob sich ein Politiker/eine Politikerin oder ein Gremium durch eine zivilgesellschaftliche Organisation beeinflussen ließe, gebe schon deren Interviewbereitschaft. Meist gäbe es nur eine Sprechbereitschaft, wenn die evaluierende Organisation besonders bekannt und geschätzt ist. Die Botschaft der Organisation an die Politik solle so formuliert werden, dass ein konkreter Nutzen erwartet werden kann.

Werkstattgespräche: Evaluieren im Rahmen der Arbeit politischer Stiftungen

Eingangs stellte Frau Köhler-Renfordt von der Friedrich-Naumann-Stiftung die Arbeit der politischen Stiftungen kurz vor. Die sechs politischen Stiftungen in Deutschland seien zumeist privatrechtlich organisierte Vereine und im In- und Ausland tätig. Die Auslandsarbeit würde durch die Ministerien finanziert, für die Arbeit in Deutschland gäbe es Stipendien. Die Stiftungen suchten sich ihre Projekte und Partner eigenständig aus, seien nach der Distanzregel nicht weisungs-, aber wertgebunden an die Parteien und damit in der politischen und Informationsarbeit im Inland nicht als Kaderschmiede der Parteien zu sehen. Sie betrieben vielmehr Bil-

3 In der *Fishbowl*-Methode werden Publikumsbeiträge systematisch in Podien einbezogen. Vor oder innerhalb des Publikums wird ein kleiner Stuhlkreis errichtet. Möchte jemand etwas sagen, nimmt er/sie entweder auf einem der freien Stühle Platz oder löst einen vorangegangenen Sprecher/eine Sprecherin auf einem anderen Stuhl ab.

dungsarbeit im Interesse der öffentlichen Hand und des politischen Pluralismus. Im Ausland hingegen würde gezielt mit politischen Partnern zusammengearbeitet. In diesem Sinne würde eine langfristige, wertorientierte Entwicklungsrichtung in Projektländern angestrebt.

Herr Dilger von der Heinrich-Böll-Stiftung stellte eine Querschnittsevaluierung der Stiftung dar, die eine Strategiesteuerung unterstützen sollte. Leitfragen waren die Bewertung der derzeitigen Strategie und ihre zukünftige Gestaltung, einbezogene Akteure und auftretende Schwierigkeiten. Dabei wurden in einer ersten Phase der interne Strategieprozess und in einer zweiten Phase die Umsetzung im Ausland untersucht und kartiert.

Frau Scherkus und Herr Vossebrecher von der Rosa-Luxemburg-Stiftung stellten in einer parallelen Werkstattgruppe eine Evaluation der Schnittstelle zwischen Zivilgesellschaft und Nationalversammlung in Vietnam dar. Die Debatte über geeignete Methoden zeigte, dass hier die Sensitivität für (politische) Kontextbesonderheiten wichtiger ist als in vielen anderen Sektoren. Einige Informationen seien nicht zugänglich oder könnten nicht im Evaluierungsbericht veröffentlicht werden. Daher sollte eine systemische Perspektive eingenommen und eine Kontributionsanalyse vorgenommen werden. Diese könne zeigen, wo Veränderungen stattgefunden und welche Aktivitäten wahrscheinlich dazu beigetragen haben.

Herr Heinle von der Hanns-Seidel-Stiftung stellte in der dritten Werkstattgruppe eine interne Evaluation einer Delegationsreise sechs afrikanischer Projektpartnervertreter ins entwicklungspolitische Büro in Brüssel dar. Diese Evaluierung war als deskriptive, explorative Studie angelegt worden. Leitfragen waren, welche Akteure bei Delegationsreisen mitwirkten, welche Informationsflüsse davor und danach entstanden, welche Veränderungen sich im Haus und im Partnerland sowie in der Zentrale oder auch in Brüssel ergaben. Die Ziele müssten auf mehreren Ebenen gesucht werden: Was will die Stiftung mit Dialogreisen generell erreichen und was speziell mit dieser Reise? Daraus könne eine wirkungsorientierte Planung künftiger Reisen und Veranstaltungen abgeleitet werden.

Frau Köhler-Renfordt präsentierte in der vierten Gruppe die Evaluation der Zusammenarbeit mit Parteien im Ausland. Die Werkstattgruppe arbeitete die Besonderheiten der Förderung von Parteien heraus:

- langfristige Unterstützung von Organisationen, statt der sonst üblichen kürzeren Programme.
- statt klarer Definition von Zielen, hohe Bedeutung der Beziehungsarbeit.
- besonders hohe Bedeutung des politischen Kontextes: demokratisches versus autoritäres Umfeld.
- spezielle *Windows of Opportunity*, wie z.B. ein Regierungswechsel, werden genutzt.
- Die Parteiförderung setzt, im Gegensatz zu anderen EZ-Feldern, aus Sicherheitsgründen oft bewusst auf geringe Sichtbarkeit.
- Entstehen vieler ungeplanter Wirkungen, Unterstützung von Organisationen, kein eigenes Programm.

Als Besonderheiten der Evaluierung von Parteiförderung einigte sich die Werkstattgruppe auf die induktive Vorgehensweise bei der Entwicklung von Kriterien, die

Betonung qualitativer Methoden (der Präsident/die Präsidentin wird nicht per Standard-Fragebogen befragt), die Notwendigkeit der Beibehaltung eines Feingefühls der Evaluatoren/Evaluatorinnen für politische Beziehungsgeflechte sowie die Notwendigkeit der Vertrauensbildung und der Akzeptanz unklarer Systemgrenzen für die Zuordnung von Wirkungen. Offen blieb, ob es die Möglichkeit von allgemeinen Effektivitätskriterien für Parteienförderung geben sollte, ob diese gemeinsam von allen Stiftungen entwickelt werden sollten und ob die Förderung demokratischer Grundprinzipien als gemeinsames Ziel formuliert werden könne.

Die Arbeit des costa-ricanischen Opferschutzbüros und dessen Wirkungen auf verschiedene Zielgruppen – Design, Instrumente und Lessons Learned einer Wirkungsevaluation

Frau von Jan (CEval) erörterte eine Hochschulkooperation mit der Universidad de Costa Rica als Hintergrund dieser Evaluation. Evaluiert wurden die Servicequalität des Beratungsprogramms des costa-ricanischen Opferschutzbüros und dessen Wirkung auf die Opfer. Dieses Programm strebe durch Interventionen in den Bereichen der sozialen Arbeit, durch psychologische Betreuung und eine Beratung zu Rechtsfragen eine ganzheitliche Betreuung bis zum Urteil an. Bei der Evaluation der Servicequalität stehe der Prozess im Vordergrund. Dafür sei der organisationstheoretische Ansatz des CEval-Modells herangezogen worden, der verschiedene Aspekte wie die Qualifikation der Mitarbeiter/-innen oder eine geeignete Infrastruktur thematisiere. Die Sprecherin wertete die Verwendung eines Quellen- und Methodenmix aus Beobachtung, Dokumentenanalyse, Workshops und Befragungen zur Triangulierung der Daten positiv. Die Opferbefragung lege im Einklang mit dem Programm großen Wert darauf, die Opfer nicht erneut zu viktimisieren. Daher beschränkte sie sich auf Fragen zur Betreuung statt zum Verbrechen. Trotzdem könne festgehalten werden, dass die Interviewer/-innen in einem derart schwierigen Umfeld leicht an Grenzen stießen. Eine ‚Mini-Ausbildung‘ in psychologischer Erster Hilfe könne hier die Kompetenz der Evaluierenden erhöhen. Zweck der Evaluation wäre das Schaffen von Lerneffekten. Dazu würden die Mitarbeiter/-innen durch Workshops sensibilisiert – sie sollten nicht kontrolliert werden, sondern Empfehlungen erhalten, die ihnen wirklich weiterhelfen. Diese Sensibilisierungsmaßnahme, zusammen mit der starken Einbeziehung des Büros und dessen hoher Ownership, wären sehr erfolgreich verlaufen.

Die Definition von Indikatoren für Vorhaben im Bereich Good Governance

Frau Dr. Hunter präsentierte in ihrem Vortrag zentrale Punkte der GIZ-Arbeitshilfe „Indikatoren“. Ausgangspunkt der Indikatorenbildung war das GIZ-Auftragsmanagement. Für die Angebotserstellung wurden Indikatoren definiert, in der Durchführung wurden diese in das wirkungsorientierte Monitoring-System übertragen und in Meilenstein-Intervallen überprüft. In der Berichterstattung und Evaluierung lieferten sie zentrale Daten. Wirkungsorientierte Monitoringsysteme in der GIZ verfolg-

ten dabei nicht nur einen indikatorenbasierten Ansatz, darüber hinaus wurden die Perspektiven von Stakeholdern mit qualitativen Methoden erfasst.

Für die Wirkungsmessung sei es entscheidend, in der Indikatorenformulierung verschiedene Dimensionen des Veränderungsprozesses und unterschiedliche Akteursperspektiven abzubilden. Dabei sei die Indikatorenauswahl ein politischer Prozess, denn die Präferenzen aller wichtigen Akteure fließen ein. Zu den wichtigsten Prinzipien gehörten die Anlehnung an die Partnersysteme und die Verhältnismäßigkeit der Erhebungskosten zum Gesamtvolumen der Maßnahme. Bei der Indikatorenbildung sollten M&E-Systeme und nationale statistische Systeme im Partnerland hinsichtlich des Vorhandenseins, der Verlässlichkeit und der Nutzbarkeit von Daten untersucht werden. Die Belastbarkeit der Daten, der Erfassungsbereich und die Terminierung von Messintervallen sollten jedoch einer kritischen Analyse unterzogen werden, um die Aussagekraft des Wirkungsnachweises nicht zu beeinträchtigen.

Nach der Geberdefinition mäßen quantitative Indikatoren Messgrößen und qualitative Indikatoren die Güte oder Qualität von Veränderungen oder sie informieren zu Einschätzungen. Qualitative Indikatoren könnten in diesem Sinne auch quantitativ gemessen werden. Die Vorstellung, dass keine Quantifizierung der qualitativen Indikatoren möglich ist, sei weit verbreitet. Die in der internationalen Gebergemeinschaft verbreite Definition, die sich auf Dimensionen des Veränderungsprozesses stützt, würde in der wissenschaftlichen Praxis nicht immer geteilt, so dass im Plenum unterschiedliche Sichtweisen auf diese Interpretation zum Ausdruck kämen.

Abschlussdiskussion

Zum Abschluss der Tagung fand im Plenum eine Fishbowl-Diskussion statt. Die Diskussion konzentrierte sich auf die während der Tagung aufgetauchten Fragen:

- Wie viel Komplexität verträgt eine Evaluierung, wie kann man sie reduzieren?
- Politik als hoch ‚kontaminiertes‘ Feld mit vielen Akteuren, wie kann eine eindeutige Zuschreibung oder Definition des Beitrags (Kontribution) erreicht werden?
- Welche Zielformulierung, welche Wirkung wird untersucht?

Es wurden erneut die Herausforderungen erwähnt, die die Politikberatung an die Evaluierung stellt: Die Evaluierenden sahen sich mit diversen *hidden agendas* konfrontiert, die sowohl hier mitbedacht werden müssen. Dieser Umstand erschwere die Bestimmung der Zielformulierung. Als eine der größten Herausforderungen für die Evaluierung wurde die Komplexität sozialer Veränderungen gesehen. Diese benötigen viel Zeit, auch seien nicht alle gesellschaftlichen Aspekte veränderbar. Ebenso komplex seien *Advocacy*-Vorgänge, um selbige zu beeinflussen. Ähnlich wie sich die Vermarktung kommerzieller Produkte wie Coca-Cola ohne nennenswerte Veränderung am Produkt alle Jahre neu inszeniere, müsse auch die *Advocacy*-Arbeit kontinuierlich angepasst werden. Eine authentische Einschätzung der Sachverhalte sei wichtig, um die Komplexität zu erfassen. Das Beispiel Costa Rica zeige, dass dazu eine genaue Definition des Betrachtungsgegenstands gehöre.



„Evaluation und Wissensgesellschaft“

18. Jahrestagung der DeGEval – Gesellschaft für Evaluation e.V. vom 16. bis 18. September 2015 am Deutschen Forschungsinstitut für öffentliche Verwaltung Speyer

Call for Proposals

Der Wandel zur ‚Wissensgesellschaft‘ hat spätestens mit Daniel Bells Arbeiten in den frühen 1970 Jahren breite öffentliche und wissenschaftliche Aufmerksamkeit erfahren. Wissen und seine Rolle in der Gesellschaft sind im Zusammenhang mit Wissensökonomie und Wissenspolitik vielfältig analysiert worden. Dabei wurde die Qualität des Wissens selbst als mehr oder weniger ‚wissenschaftliches‘ Wissen hinterfragt, aber auch die Rolle von Nichtwissen thematisiert. Die technologische Entwicklung der Datenverarbeitung und Digitalisierung ermöglicht neue Zugänge zu Wissen. Koordinations- und Managementprozesse bedienen sich immer stärker Instrumenten der Wissensgenerierung und -aufbereitung. Dies gilt sowohl auf Ebene von Organisationen als auch für gesellschaftliche Teilbereiche oder ganze Gesellschaften. Modernisierungskonzepte wie ‚Better bzw. Smarter Regulation‘ sowie ‚Open Government‘ sind ohne eine spezifische Form der Wissensgenerierung, -aufbereitung und -nutzung kaum denkbar.

Nicht zufällig fällt die Entwicklung der Wissensgesellschaft mit der steigenden Bedeutung der Evaluierung zusammen. Evaluierung ist für sich genommen ein Verfahren zur Generierung von Wissen. Die Jahrestagung 2015 will das Verhältnis von Evaluation und Wissen aus unterschiedlichen Perspektiven beleuchten. Dabei stehen folgende Schwerpunkte im Vordergrund:

- Welche *Rolle* spielt *Evaluation in der Wissensgesellschaft* und wo sind ihre Grenzen? Welche Arten von Wissen steuern unterschiedliche Evaluationsdesigns und -methoden für welche Zwecke bei? Welche Rolle spielt Evaluation in unterschiedlichen Politikfeldern und organisatorischen Kontexten?
- Welches *Nichtwissen* ist mit *Evaluierung* verbunden bzw. muss in Kauf genommen werden („blinde Flecken der Evaluation“)? Wie wird mit Wissenslücken umgegangen?
- Wie kann die *Nutzung von Evaluationsergebnissen* gestärkt werden? Welche Voraussetzungen erleichtern die Nutzung von evaluationsgeneriertem Wissen? Welchen Beitrag leistet Evaluation für eine evidenzbasierte Governance und Steuerung? Welche Rolle spielt Evaluation für Koordinationsprozesse innerhalb von Organisationen und Politikfeldern?
- Welche Veränderungen ergeben sich für Evaluation durch die neuen Möglichkeiten der Wissensgesellschaft? Welche *ethischen Fragen* entstehen für die Evaluation? Welche *methodischen Herausforderungen* ergeben sich einerseits durch verfügbare ‚Datenfluten‘ und Wissenslücken andererseits?
- Wie kann systematisch *Wissen über Evaluation* gewonnen und zur Weiterentwicklung von Evaluation genutzt werden? Für welche Bereiche liegen Metaevalationen vor und wie werden deren Ergebnisse genutzt?

Diese Fragen stehen im Mittelpunkt der 18. Jahrestagung der DeGEval, die vom 16. bis 18. September 2015 in Speyer am Deutschen Forschungsinstitut für öffentliche Verwaltung stattfindet.

Wir laden Sie herzlich ein, an der Diskussion in Speyer teilzunehmen und Ihre Arbeit auf der Jahrestagung vorzustellen. Eingeladen sind Beiträge, die sich mit den oben genannten Fragen wissenschaftlich, reflexiv oder praktisch auseinandersetzen. Darüber hinaus ist 2015 das Internationale Jahr der Evaluation. Besonders interessant sind deshalb auch Beiträge, die eine solche Perspektive reflektieren können. Die Tagung bietet folgende Beitragsmöglichkeiten:

- **Blitzvortrag:** In einem Blitzvortrag wird ein Forschungs- oder Praxisprojekt mit seinen innovativen Aspekten oder neuartigen Fragestellungen auf möglichst allgemein verständliche Weise dargestellt. Eine Blitzvortragssession besteht aus zehn fünfminütigen Blitzvorträgen, an die sich je drei Minuten für Fragen anschließen. Die Möglichkeit des Blitzvortrags richtet sich insbesondere an Studierende, Doktorand(inn)en und Berufseinsteiger(innen), aber auch an Evaluationspraktiker(innen), welche ihre Arbeit schlaglichtartig einem möglichst breiten Publikum präsentieren möchten.
- **Einzelvortrag:** In Einzelvorträgen wird das Thema Evaluation entweder praktisch, reflexiv oder wissenschaftlich betrachtet: (a) Evaluationen als Projekte mit praktischen Herausforderungen (Planung, Steuerung, Durchführung etc.), (b) Evaluationen als Grundlage zur Reflexion und zum Generieren von (Meta-)Wissen über Evaluation (methodische, theoretische, methodologische Aspekte) oder (c) Forschungsergebnisse über Evaluation (Evaluation als Forschungsgegenstand) präsentiert. Für den Vortrag stehen 20 Minuten und 10 Minuten Diskussionszeit zur Verfügung. Einzelvorträge werden zu thematischen Sessions zusammengestellt (drei Vorträge pro Session).
- **Positionsvortrag:** Positionsvorträge dienen der fundierten Darstellung einer eigenen wissenschaftlichen Position zum Thema Evaluation. Die Vortragszeit beträgt 30 Minuten mit einer anschließenden Diskussion von 15 Minuten.
- **Komplette Session:** Komplette Sessions (90 Minuten) können als Vortragssession, als Podiumsdiskussion, Workshop oder in anderer Form konzipiert sein. Abstracts umfassen die Beschreibung der geplanten Session sowie aller beteiligter Beiträge und Personen. Bei einer Vortragssession sind zusätzlich Abstracts für jeden Einzelvortrag einzureichen und ein(e) Moderator(in)/Diskutant(in) zu benennen. Komplette Sessions können in deutscher und englischer Sprache stattfinden.
- **Poster:** Poster präsentieren entweder innovative Evaluationsvorhaben oder Forschungsarbeiten zu Fragestellungen, welche empirisch, konzeptionell, theoretisch oder methodisch behandelt werden. Es wird eine Posterführung bzw. eine moderierte Postersession stattfinden.

Bitte reichen Sie Ihr Abstract mit maximal 2.200 Zeichen bis zum **31. Januar 2015** über unsere Konferenzhomepage (<https://www.conftool.com/degeval2015/>) ein und beachten Sie bitte die dort angeführten Hinweise zur Einreichung. Alle Beiträge werden in einem Review-Verfahren begutachtet. Sie werden bis zum 29. Mai 2015 über die Beitragsannahme informiert. Nähere Informationen zur Tagung finden Sie unter www.degeval.de/veranstaltungen/jahrestagungen/speyer-2015/.

Kontakt:

DeGEval – Gesellschaft für Evaluation e.V.
Johannes Gutenberg-Universität Mainz
c/o Zentrum für Qualitätssicherung
und -entwicklung (ZQ)
Colonel-Kleinmann-Weg 2, SBII, 04-543
55099 Mainz

Natalie Salf

Tel: +49-(0)6131- 39 2 68 69
Fax: +49-(0)6131- 39 2 68 68

E-Mail: info@degeval.de
<http://www.degeval.de>

Ukrainian Evaluation Association (UEA)

*Iryna Kravchuk,¹ Tetiana Narchynska,¹ Kateryna Stolyarenko,¹
Natalya Yakymchuk¹*

UEA Development, Membership and Management

The Ukrainian Evaluation Association (UEA) is a non-governmental professional association in the field of evaluation of projects, programs, policies and organizational development in Ukraine. It was created in July 2011 as an informal network of professionals, until in June 2013 it was officially registered. UEA has developed from a Facebook group to a wide network of people sharing one common mission and common goals.

The mission of the Ukrainian Evaluation Association is to support development, dissemination and use of evaluation in Ukraine, to improve decision-making in public, non-governmental and private sectors and to contribute to the development of evaluation culture. UEA operates based on the principles of independence, transparency and legitimacy, ethics and corporate social responsibility, professionalism and efficiency, using the best international practices in the evaluation area. UEA acts on the basis of its statutes, which stipulate the principles of functioning, management and structure, membership procedures, decision-making mechanisms and the strategic plan for 2013-2015.

According to the UEA statutes, membership in UEA is individual and voluntary. Members are accepted on the basis of: personal application, completion of a membership form, and payment of membership fees. The vast majority of activities are done on voluntary basis. Official membership and the membership fee that amounts to 25 Dollar per year were introduced just at the end of 2013. In September 2013, the UEA had 13 members, a Facebook-group with 203 participants and an email contact list with 142 contacts. At the moment, the UEA has 32 official members and 256 followers on Facebook.

The UEA has a board and secretariat. Each member of the board is responsible for different aspects of the UEA development, including strategic planning, fundraising and communication. The UEA holds regular meetings with its members to exchange M&E experiences, and it has formed seven Topical Interests Groups (TIGs) (Evaluation of Social Programs, Evaluation in Public Sector, M&E Ter-

¹ Ukrainian Evaluation Association (UEA)

minology, Evaluation of Environmental Programs, Training in M&E, Evaluation Standards, National M&E System) to facilitate discussions in particular fields of interest of its members.

UEA Activities and Achievements.

Due to the need of fostering common understanding of the usage of the monitoring and evaluation terminology in third, private and governmental sectors, from February to October 2012, the Ukrainian Evaluation Association conducted *the baseline study* “State of Evaluation Capacity Development in Ukraine: demand, supply, institutionalization.” In addition, the UEA organized, in partnership with the National Academy of Public Administration (NAPA), round-table and presentation of this baseline study in November 2012.

The study showed that the evaluation capacity in Ukraine currently is on quite a low level, and there are different interpretations of the key M&E terms in Ukrainian language by local representatives of business, donors and NGO sectors. Therefore, UAE members established a working group, and in April 2013 they started, with grant support by the Swiss Cooperation Office in Ukraine, a joint project on the development and adjustment of the key M&E terminology in Ukrainian language, with equivalents provided in English and Russian.

For July 2014, the team has already completed the final draft of the M&E glossary in Ukrainian language, which includes recommendations of four peer reviewers. In addition, it conducted 27 team meetings, prepared and revised 13 glossary drafts in total, and organized an open discussion of the M&E glossary on May 27, 2014 in Kyiv. Project updates were made public regularly and presented to experts of the field in the context of the UEA’s monthly meetings in 2013-2014. The completed draft of the M&E glossary is now in the final stages of editorial proofreading and is to be disseminated for the wide audience in autumn 2014, in order to contribute to reaching an agreement and common understanding of evaluation terminology with the aim of standardizing evaluation practice in Ukraine.

As a result of starting the activities on the M&E glossary development and the contact with international communities during a set of international European and US evaluation conferences, the Ukrainian Evaluation Association came into a process of identifying and codifying its own national system of evaluation standards.

In June 2013, the UAE has joined the international peer-to-peer support program called “The joint process on revisiting the DeGEval Standards for Evaluation and the SEVAL Standards for Evaluation: Lessons learned on using standards and recommendations for the implementation of standards in a young evaluation Society.” The project involved the DeGEval – Gesellschaft für Evaluation e.V. (Germany and Austria, DeGEval), the Schweizerische Evaluationsgesellschaft/Société suisse de l’évaluation, Switzerland (SEVAL), and the Ukrainian Evaluation Association (Ukraine, UEA); it includes discussing the revision of the DeGEval and SEVAL standards, exchanging documents, and conducting a workshop held in Berne, Switzerland on November 1, 2013.

As a result of attending the workshop in Berne by the Ukrainian side of the project, information obtained on challenges and advantages of the usage of evaluation standards for professional evaluation were presented to all members of the Ukrainian Association of Evaluation in November 2014. The discussion of lessons learned covered questions like: are there any national aspects of evaluation standards and a quality definition in Ukraine? Does the young democratic society have enough qualified experts in evaluation to spread evaluation culture in Ukraine? And what are the best practices of societies experienced in developing a system of professional evaluation (e.g. in Canada, Germany, France, Switzerland, USA, taking into account the workshop in Bern with the DeGEval and the SEVAL)?

The UEA working discussion resulted in establishing a working group on UEA evaluation standards in late November 2013. The first draft of UEA standards was prepared and processed by the working group from December 2013 to January 2014. The second draft of UEA standards with additional edits and comments was to be reviewed by UEA members in early February 2014. Due to lessons learned during the workshop discussion in Berne, the UEA members decided to start the process of preparing the UEA evaluation standards by formulating its key principles, including a clarification of basic terms of evaluation as well as national specifics of evaluation practice in Ukraine.

Moreover, based on experiences of the DeGEval and the SEVAL, presented in shared documents, protocols and in detailed discussions in the workshop on revising evaluation standards in Berne, the UEA understands that the basic steps for young societies in developing evaluation standards are:

- scrupulous clarification of a glossary for evaluation key terms with taking into account the national context and the audience to whom this information will be presented;
- studying best practices and preparing a checklist to develop a specific, systematic, concise, and understandable set of evaluation standards that meets quality standards;
- understanding the applicable meaning of evaluation standards as an instrument for professional evaluation and adapting evaluation principles from theory to practice;
- language of evaluation standards should be understandable and clear;
- process of elaborating evaluation standards should be as open and inclusive as possible to include all possible stakeholders to build a shared understanding of the key principles;
- consideration of the need to spread the culture of evaluation and evaluation standards in wider collaboration with academic institutions, governmental and non-governmental organizations through a number of public trainings, conferences and presentations as well as the application of a creative approach to introducing evaluation standards to evaluators and the general public, promoting evaluation culture as a part of democratic values and freedom.

Starting the development of evaluation standards, the association should decide on a general approach to evaluation standards (general principles or detailed standards),

the structure of the standards, the process of elaboration (working group, discussions, procedures) and a dissemination strategy.

From June 2011 to September 2013, the UEA held 19 working meetings. Due to the political situation in Ukraine, the UEA held only two meetings in 2014.

The UEA organized and conducted a Skype conference with the Slovak Evaluation Association; conducted an assessment of the needs and interests of UEA members; developed – in partnership with the Ukrainian Philanthropists Forum – a training course in M&E called the “Alphabet for Evaluation of Charitable Projects and Programs” and conducted two series of trainings for representatives of Ukrainian Civil Society Organizations (CSOs) in 2013 and 2014; developed a strategic plan for 2013 to 2015; became a member of the Ukrainian National Platform of the Civil Society Forum of the Eastern Partnership and prepared two public annual reports.

In May 2014, the NAPA annual conference included the special section “Enhancement of M&E system of public policies, programs and projects.” The UEA invited the international expert Alexey Kyzmin to give a presentation in this section.

The Ukrainian Evaluation Association is a member of the Network of Evaluation Societies in Europe (NESE), the Eurasian Alliance in Evaluation and signed the “Chiang Mai declaration of civil society on uniting efforts aimed at improving evaluation” at the International Forum by the EvalPartners initiative (December 3-6, 2012, Thailand).

UEA future

Currently, the UEA works under political and financial constraints in Ukraine. Many activities were slowed down in winter and spring 2014. Ukraine passed through Euromaidan revolution and the election of a new president. Presently, there is an armed conflict in the Ukrainian regions of Lugansk and Donetsk in the East. Civil society is mainly focused on the assistance to the Ukrainian army, the injured people in hospitals and displaced persons from unsafe regions. On the one hand, due to increased budget expenses on military forces and economic crises, the state does not have any extra financial resources for program evaluation. The state budget has been cut considerably. On the other hand, public resource management has to be more efficient and effective. The current government seems to be more reform-oriented than the previous one. It promises new opportunities for evaluation development in the public sector. In spite of the difficult times in country and society, the UEA has survived and pursues its mission in new conditions.

The UEA’s main objectives for the coming two years are the following ones:

- organization of the National Evaluation Conference and Evaluation Year 2014;
- development of the UEA website;
- development of a fundraising plan and communication strategy;
- development of evaluation standards and an evaluators competencies framework;
- development of the concept of the National M&E system and its advocacy among representatives of the public sector;
- formation of an evaluators database and of an M&E Library.

UEA basic information:

UEA - Facebook page: <https://www.facebook.com/groups/Ukr.evaluationsociety/>

E-mail: ukrevaluation@gmail.com

*UEA BOARD**Iryna Kravchuk*

Head of UEA Board

irene.kravchuk2010@gmail.com

Olga Schetinina

Deputy Head of the Board

oschetinina@mail.ru

Kateryna Stolyarenko

Coordinator on Strategic Planning and fundraising

katya.stolyarenko@gmail.com

Larysa Pylgun

Coordinator of Educational Programmes

lara_dc@yahoo.com

Iryna Ozymok

Coordinator on information and communications

ozzymok@gmail.com

Natalya Yakymchuk

Administrative Coordinator

natalya.yakymchuk@gmail.com

Österreichische Plattform für Forschungs- und Technologiepolitikevaluierung (fteval)

*Klaus Schuch*¹

Entstehung und Motivation

Die Österreichische Plattform für Forschungs- und Technologiepolitikevaluierung (fteval) wurde 1996 von einer Gruppe engagierter Personen aus der Forschungs- und Technologiepolitikverwaltung, der Forschungsförderung sowie der FTI-Politikanalyse in Form einer losen Kooperation gegründet, um die Evaluierungskultur und Evaluierungspraxis in Österreich in einem umfassenden Sinn weiterzuentwickeln und um damit einen Beitrag für eine effizientere Planung der Forschungs-, Technologie- und Innovationspolitik in Österreich zu leisten.

Angestoßen wurde dieses Ansinnen durch den österreichischen EU-Beitritt, der zwei Jahre zuvor erfolgte, und dem sich daraus ergebenden Modernisierungsschub im F&E-Bereich mitsamt seinen Anpassungserfordernissen sowie einer insgesamt zunehmenden Bedeutung von Wissenschaft, Forschung und Technologie in Österreich. Diese Bedeutungszunahme wurde von der raschen Einführung neuer Maßnahmen, Verfahren und Institutionen begleitet, was – wie auch anderswo in Europa – zu einer Reihe von F&E-politischen Experimenten geführt hat, die durch den Einsatz von Evaluierungen begleitet und ‚kontrolliert‘ wurden.

Seit 2006 ist die Plattform gesellschaftsrechtlich ein Verein. Anders als die DeGEval – Gesellschaft für Evaluation e.V. hat die fteval bis dato ausschließlich institutionelle Mitglieder. Sie vereinigt nahezu alle wesentlichen österreichischen Institutionen der Forschungs- und Technologiepolitik und umfasst dabei Institutionen der Politikgestaltung, der Forschungs-, Technologie- und Innovationsförderung sowie Einrichtungen, die Evaluierungen durchführen.

Auf Basis dieser Mitgliedsstruktur werden gemeinsam Grundlagen für transparentere Evaluierungsverfahren erarbeitet sowie jährliche Schwerpunktsetzungen durchgeführt. Konkret ausgeschriebene oder laufende Evaluierungen werden hingegen im Rahmen der Plattform weder besprochen noch sonst irgendwie kommentiert oder zur Diskussion gestellt.

¹ Österreichische Plattform für Forschungs- und Technologiepolitikevaluierung (fteval), Wien

Mitglieder

Zur Zeit umfasst die Plattform fteval 19 Organisationen, nämlich das Bundesministerium für Verkehr, Innovation und Technologie, das Bundesministerium für Wissenschaft, Forschung und Wirtschaft, das Austrian Institute of Technology GmbH, die AQ Austria, die Austria Wirtschaftsservice Gesellschaft mbH, die Christian Doppler Forschungsgesellschaft, die convelop – cooperative knowledge design GmbH, die Österreichische Forschungsförderungsgesellschaft (FFG), den Fonds zur Förderung der wissenschaftlichen Forschung (FWF), das IHS – Institut für Höhere Studien, die KMU Forschung Austria, die Joanneum Research Forschungsgesellschaft mbH, die Ludwig Boltzmann Gesellschaft, das Österreichische Institut für Wirtschaftsforschung, den Rat für Forschungs- und Technologieentwicklung, Technopolis Austria, den Wiener Wirtschafts-, Forschungs- und Technologiefonds (WWTF), die Technologieagentur der Stadt Wien GmbH (ZIT) und das Zentrum für Soziale Innovation (ZSI).

Leistungen

Unter Einbindung von nationalen und internationalen Expert(inn)en wird im Rahmen von Veranstaltungen und einem regelmäßig erscheinendem Journal sowie eines Newsletters der Ideenaustausch auf dem Gebiet der Forschungs- und Technologiepolitikevaluierung gefördert. Aktuelle Evaluierungsansätze und -methoden werden hierbei präsentiert und kritisch erörtert. Die Diskussion und der Einsatz guter Methoden, Verfahren, professioneller Vorgangsweisen und Rahmenbedingungen sind Kernaufgabe der Plattform und waren wesentlicher Bestandteil in der Entstehung der fteval-Standards, die 2012 überarbeitet wurden.

Die Standards der fteval sind als ‚Code of Conduct‘ zu verstehen. Sie informieren über die Notwendigkeit, Bedeutung und Funktion von Evaluierungen in der Forschungs- und Technologiepolitik, positionieren Evaluierung im Politikzyklus, informieren über methodologische Aspekte und weisen Spielregeln für Evaluatord(inn)en und Auftraggebende von Evaluationen aus.

Die Plattform bietet des Weiteren regelmäßig Workshops sowie Trainings an, welche sich an den Interessen und Bedürfnissen der Mitgliedsorganisationen orientieren. Hinzu kommen individuelle Beratungsdienstleistungen für Mitglieder der Plattform fteval, z.B. über Evaluierungsdesigns für bestimmte Vorhaben oder die Gestaltung von Terms of Reference. Dass die die Beratung durchführenden fteval-Expert(inn)en dann an den Evaluierungsausschreibungen nicht teilnehmen, versteht sich dabei von selbst.

Die Plattform fteval verfasst Stellungnahmen zu evaluierungsrelevanten Kommunikationen wie dem Vorschlag der Europäischen Kommission zur „Evaluation in the field of State aid – Draft Methodological Guidance Paper“ sowie anderen relevanten Veröffentlichungen und Vorhaben, um den österreichischen und internationalen Diskurs zu bereichern. Die internationale Vernetzung wird von verschiedenen Mitgliedern der fteval anlassbezogen wahrgenommen, wobei der Geschäftsführer die fteval Österreich auch im European RTD Evaluation Network vertritt.

Die Plattform informiert ihre Mitglieder regelmäßig über interessante wissenschaftliche Artikel oder international durchgeführte Evaluierungsvorhaben und Studien und publiziert selbst ein einschlägiges Journal, das „*fteval Journal for Research and Technology Policy Evaluation*“, für das immer wieder auch Beiträge von außerhalb Österreichs willkommen sind. Auf der Homepage www.fteval.at werden alle relevanten Evaluierungsstudien gesammelt und sind frei zugänglich. Auskunft und Kontakt zu österreichischen Evaluator(inn)en gibt die Evaluator(inn)endatenbank der Plattform [fteval](http://fteval.at).

Finanzierung

Die Plattform [fteval](http://fteval.at) finanziert sich einerseits aus den Mitgliedsbeiträgen, die es erlauben eine kleine Sekretariatsinfrastruktur mit einer halbtagsbeschäftigten Fachkraft zu betreiben, die die Mitgliedsverwaltung, die Buchhaltungsvorbereitung und Rechnungsabschlussvorbereitung, die Homepage-Betreuung sowie allgemeine Informationsaktivitäten und Sekretariatsarbeiten durchführt. Andererseits finanzieren auf freiwilliger Basis einige Einrichtungen bestimmte Schwerpunktaktivitäten der [fteval](http://fteval.at). So wurde beispielsweise die große internationale [fteval](http://fteval.at)-Konferenz „*Evaluation of STI policies, instruments and organisations: new horizons, new challenges*“, die im November 2013 in Wien stattgefunden hat, vom Innovationsministerium, dem Wissenschaftsministerium, dem Wirtschaftsministerium sowie den Agenturen FWF, FFG und WWTF kofinanziert. Die Büroinfrastruktur wird vom ZSI bereitgestellt, das zurzeit auch den Geschäftsführer der [fteval](http://fteval.at) stellt.

Kontakt

Österreichische Plattform für Forschungs- und Technologiepolitikevaluierung
c/o ZSI
Linke Wienzeile 246, A-1150 Wien
www.fteval.at
office@fteval.at

Gegen PISA. Für Bildung.

Wolfgang Böttcher¹

Die wohl bedeutendste Maßnahme zur Überprüfung der Leistungsfähigkeit von Schulsystemen ist erneut in die Kritik geraten. Heinz-Dieter Meyer, Professor an der New York State University in Albany, hat einen offenen Brief an den Chefkoordinator von PISA (Programme for International Student Assessment) verfasst.² Hunderte Bildungsforscher(innen) haben sich seinem Aufruf zu einem Moratorium des Tests angeschlossen. Der Adressierte, Dr. Andreas Schleicher, hat die Kritik zurückgewiesen.

PISA verenge den Blick auf das schulische Curriculum auf gut Messbares. Mit PISA habe die OECD, so die Bildungsforscher(innen), nicht nur Curricula einseitig ökonomisch ausgerichtet, sie habe auch massiv zur Eskalation der quantitativen Vermessung der Schulsysteme beigetragen – und dies trotz aller Unsicherheit hinsichtlich der Validität und Reliabilität standardisierter Tests. Die Leitidee, Leistungen von Schüler(inne)n, Lehrer(inne)n und Systemen in Leistungstabellen abzubilden, habe auch dazu geführt, dass kurzschrittige schulpolitische Interventionen zu nichts anderem dienen als dazu, in den Tabellen aufzusteigen.

Die wettbewerbliche Ausrichtung und die inhaltliche Verengung, die die Bedeutung gesundheitlicher, ethischer, ästhetischer und politischer Aspekte der Bildung negiert, beschädigen Bildung in weltweitem Ausmaß. Dass Bildung, ihre Prozesse und Wirkungen begutachtet werden müssen, steht allerdings auch bei den Kritikern außer Frage. Aber in ihrem Aufruf erwarten sie, dass über aussagekräftigere und nützlichere Formen der Bewertung nachgedacht werden müsse.

Man könnte diese Forderung ergänzen und eine *Evaluation* der Schule fordern, statt eines Testregiments. Evaluation muss nicht nur genau sein, das heißt, sie wird sich nicht – wie es PISA vorgeworfen wird – auf wenige einseitige Indikatoren reduzieren. Und sie hat den Anspruch, nützlich zu sein und den evaluierten Gegenstand zu verbessern – und nicht nur zu vermessen. PISA, so die Kritik, ist mehr als das: Das Programm sei schädlich und zerstöre die Freude am Lernen.

1 Westfälische Wilhelms-Universität Münster und Vorstandsvorsitzender der DeGEval

2 Der Brief wurde am 6. Mai 2014 in der britischen Zeitung *The Guardian* veröffentlicht und ist online verfügbar unter: <http://www.theguardian.com/education/2014/may/06/oecd-pisa-tests-damaging-education-academics>.