

Veranstaltungsbericht: Frühjahrstreffen des AK FTI in Berlin am 12. Mai 2016

Jan Wessels,¹ Sonja Kind,¹ Leo Wangler¹

Am 12. Mai 2016 fand in den Räumen des Instituts für Innovation und Technik der VDI/VDE-IT in Berlin das diesjährige Frühjahrstreffen des Arbeitskreises Forschungs-, Technologie- und Innovationspolitik (AK FTI) statt. Schwerpunkt des diesjährigen Treffens war das Thema „Projekträger und Evaluation“.

Evaluation hat sich im Politikfeld FTI als Standardverfahren der Bewertung von Fördermaßnahmen sowie als Steuerungsinstrument für Ministerien und Projekträger etabliert. Der Druck – z.B. durch den Bundrechnungshof – in Richtung einer weiteren Standardisierung von Verfahren und eines kontinuierlichen Monitorings wächst allerdings. Damit ändert sich auch die Rolle von Projekträgern (PTs) in Deutschland, die bislang eher reaktiv in Evaluationen eingebunden waren. Sie nehmen nun zunehmend eine aktivere Rolle ein, müssen die Evaluierbarkeit von Programmen von Anfang an mitdenken, müssen entsprechende Monitoring-Daten bereithalten, die Evaluierende bei der Umsetzung unterstützen und die Implementierung von Evaluationsergebnissen sicherstellen. Zudem übernehmen PTs zunehmend häufiger selbst die Rolle von Evaluatoren, dies sowohl bei der Selbstevaluation eigener Maßnahmen wie auch bei der Evaluation ‚fremder‘ Programme. Stärker zentralisierte Projekträger in anderen europäischen Ländern, auch in Österreich, sind hingegen schon länger in eine aktivere Rolle innerhalb von Evaluationen hineingewachsen und haben wichtige Erfahrungen für die aktuelle deutsche Diskussion beizusteuern.

Das Frühjahrstreffen sollte die veränderte Rolle von PTs aus zwei möglichen Perspektiven beleuchten:

- PTs und ihre sich verändernde Rolle bei Evaluationen ihrer eigenen Fördermaßnahmen
- PTs als ‚Evaluatoren‘ fremder Maßnahmen

In seinem einführenden Vortrag rief Jan Wessels zunächst die Vorgeschichte der Themenauswahl, also die zunehmende Zahl und aktive Rolle von Projekträgervertretern im Arbeitskreis in Erinnerung. Um die potenzielle Breite des Themas zu unterstreichen, kontrastierte er dann die deutschen und österreichischen Rollenmuster und Funktionen von Projekträgern mit der Situation in anderen europäischen Län-

¹ Institut für Innovation und Technik, Berlin

dern, in denen Projektträger zum Teil sehr systematisch und weitgehend ihre eigenen Maßnahmen evaluieren. Als Einstimmung in die weitere Agenda des Tages skizzierte Herr Wessels typische Herausforderungen dieser neuen Funktionen für die Projektträger selbst, aber auch für die externe Evaluationseinrichtung. Spannungsgeladen sind beispielsweise die unterschiedlichen Kompetenzerwartungen bzw. vermutete Kompetenzdefizite von PT und Evaluatoren. Dies kann sich sowohl auf methodisches Wissen über Evaluationen wie auch auf das Untersuchungsobjekt beziehen, also auf ein spezifisches Technologiefeld oder auf die internen Strukturen und Prozesse der PT-Arbeit. Ein weiteres typisches Konfliktfeld umschreibt die Qualität und Lesbarkeit der Evaluationsberichte und die Brauchbarkeit von Handlungsempfehlungen. Nehmen Projektträger selbst eine aktive Rolle in der Evaluierung ein, so ergeben sich Rollenkonflikte, neue Konkurrenzsituationen und Herausforderungen hinsichtlich des notwendigen methodischen Know-hows.

Sabine Mayer von der Österreichischen Forschungsförderungsgesellschaft (FFG) verdeutlichte einige dieser eingangs eingeführten Problemstellungen am Beispiel der Evaluationspraxis der FFG. Die Evaluierung hat eine klar umrissene Position in der Prozesslandschaft der FFG. Die FFG setzt ausschließlich auf externe Evaluationen, stellt aber über die klare Positionierung innerhalb des Gesamtprozesses der FFG sicher, dass die Organisation ein Höchstmaß an Nutzen durch Evaluationen erfährt. Sie stellt damit z.B. auch sicher, dass Evaluationsausschreibungen die Evaluierbarkeit bestimmter Fragestellungen vorab berücksichtigen und den Zugang zu den vorhandenen Daten absichern. Am Ende einer Evaluierung wird zudem darauf geachtet, dass die Ergebnisse in eine Optimierung des eigenen Handelns einfließen und die tatsächliche Umsetzung mit einem gewissen Zeitabstand überprüft wird. Als eine mehrerer Herausforderungen sprach Frau Mayer die Maßnahmen-/ Programmorientierung von Evaluationen an, die systemische Aspekte und übergreifende Perspektiven in der Regel weitgehend unberücksichtigt lässt.

Norbert Knoll berichtete über Möglichkeiten und Grenzen agenturinterner Evaluationen am Beispiel von Austria Wirtschaftsservice (aws). Die aws ist als staatliche Förderbank Österreichs auf die Finanzierung von kleinen und mittelständischen Unternehmen spezialisiert. Die Evaluation der aws-Programme wird häufig intern in Form eines Wirkungsmonitorings umgesetzt. In der aws selbst stieß dieses Vorgehen zu Beginn auf Skepsis. Mittlerweile ist die Anerkennung jedoch gewachsen, die Umsetzung interner Evaluationen wird heute nicht mehr grundsätzlich infrage gestellt. Die Akzeptanz gegenüber den Evaluationsergebnissen ist dann am größten, wenn aus dem Programm heraus ein Wille für Veränderungen existiert. Besonders der enge Kontakt zum Programmmanagement eröffnet den internen Evaluierenden Vorteile. In der Praxis erzielen die internen Evaluationen positive Effekte. Häufig gelingt es, bestehende Widersprüche aufzulösen und die Programme strategisch weiter zu entwickeln. Intern ausgeführte Evaluation erfüllen nach der Erfahrung der aws insbesondere eine Erkenntnis- und eine Lernfunktion. Im Bereich der Kontroll- und Legitimitätsfunktion sind externe Evaluationen hingegen klar im Vorteil. Daher bewähren sich die internen Evaluationen der aws insbesondere für die strategische Weiterentwicklung von Programmen. Hier profitiert die Agentur besonders durch die Nähe und den intensiven Austausch mit den Programmverantwortlichen. Durch

die in der Regel äußerst begrenzten Budgets interner Evaluationen sind die adressierten Fragestellungen allerdings sehr auf das Wesentliche reduziert, d.h. sie haben einen indikativen bzw. pilothaften Charakter.

Frau Gesa Koglin vom Projektträger Jülich (PtJ) stellte in ihrem Vortrag die Evaluations- und Monitoringaktivitäten zur Programmfamilie ‚Unternehmen Region‘ vor. Die Innovationsinitiative des Bundesministeriums für Bildung und Forschung (BMBF) unterstützt bereits seit 1999 mit einer Reihe von Programmlinien regionale Bündnisse in den Neuen Ländern. Den Programmlinien gemeinsam ist das Ziel, die Wettbewerbsfähigkeit von Wirtschaft und Wissenschaft in den Regionen zu stärken. Je nach Programmlinie steht eher die wissenschaftliche oder die wirtschaftliche Profilbildung der Regionen im Mittelpunkt. Zudem unterscheiden sich die Programmlinien mit Blick auf die Fördervolumina und Gestaltung der Auswahlverfahren. Die einzelnen Programmlinien wurden bis auf das Modul WK Potenzial von externen Evaluierenden evaluiert. Darunter waren Strategieberatungen, wissenschaftliche Einrichtungen sowie ‚typische‘ Evaluationseinrichtungen. Ein Überblick der bei den Evaluationen eingesetzten Methoden zeigt eine große Vielfalt. Die Methoden umfassten SWOT-Analysen, Beschreibung von Kennzahlen, Netzwerkanalysen sowie Fallstudien und Soll-Ist-Vergleiche. Die Strategieberatungen tendierten dabei eher zu einem Instrumentarium, das auch in Beratungsprojekten von Unternehmen verwendet wird wie z.B. SWOT-Analysen oder eine Betrachtung des gesamten Managementsystems. In der Regel wurden die Daten mit ähnlichen Methoden wie schriftlichen Befragungen, Vor-Ort-Besuchen oder Interviews erhoben. Nur in einem Fall erfolgte eine Regressionsanalyse. Im zweiten Teil des Vortrags standen die Monitoringaktivitäten von PtJ im Mittelpunkt, die am Beispiel der Programmlinie ‚Innovative regionale Wachstumskerne‘ vorgestellt wurden. Während Evaluationen eher punktuell, extern und mit einem größeren Fokus durchgeführt werden, findet das Monitoring regelmäßig, intern und zu bestimmten Schlüsselindikatoren statt. Beim Monitoring der Förderlinie Wachstumskerne werden z.B. die Anpassung der Strategie, die unternehmerische Entwicklung sowie Ausgründungen/Ansiedlungen als Indikatoren erhoben. Zusammenfassend wurde festgehalten, dass das Programmmonitoring für Projektträger eine zunehmende Bedeutung bekommt, um projektspezifische Entwicklungen möglichst zeitnah zu dokumentieren. Der im Vortrag vorgenommene Vergleich von Monitoring- und Evaluationsaktivitäten zeigte, dass beide Vorgehensweisen sowohl Gemeinsamkeiten als auch Unterschiede aufweisen.

Michael Astor von der Prognos AG berichtete von der Entwicklung eines Evaluationssystems zur Evaluation von Förderprogrammen im Bereich Informations- und Kommunikationstechnologien (IKT). Die Studie wurde zusammen mit den Projektpartnern Fraunhofer-Institut für Graphische Datenverarbeitung und dem Zentrum für Informations-, Wissens- und Dienstleistungsmanagement (ZIWD) im Auftrag des Bundesministeriums für Wirtschaft und Energie (BMWi) erarbeitet. Das zu entwickelnde Evaluationssystem musste mehrere Voraussetzungen erfüllen: Die Fördergegenstände der IKT-Förderung zeichnen sich durch hohe Entwicklungsdynamiken und komplexe Wirkungsketten mit sich überlagernden Effekten aus. Die Fördermaßnahmen sind häufig auch durch umfassende Wirkungsansprüche gekennzeich-

net. Ferner sind die adressierten Zielgruppen der Fördermaßnahmen nicht notwendigerweise identisch mit den Akteuren, bei denen Wirkungen zu beobachten sind. Oftmals können die hohen Entwicklungsdynamiken die Wirkungen überlagern, verstärken oder gar konterkarieren. Das Evaluationssystem sollte hierauf besondere Rücksicht nehmen und entsprechend passende und praktikable Indikatoren bieten, die über die Beobachtung des geförderten Akteurskreises hinausgeht. Dabei sollte das Evaluationssystem so flexibel gestaltet sein, dass es an zukünftige Erfordernisse im Sinne einer lernenden Evaluation angepasst werden kann. Eine weitere Voraussetzung bestand darin, das Evaluationssystem im Einklang mit den Anforderungen der Erfolgskontrolle durch §7 Abs. 2 der Bundeshaushaltsordnung (BHO) zu gestalten.

Zur Realisierung der genannten Voraussetzungen wurde ein erweitertes Methodenspektrum gewählt, das über jenes von klassischen Evaluierungen hinausgeht und vier Elemente enthält. Im Rahmen einer Basisanalyse werden Informationen gesammelt, die helfen, die Ursächlichkeit der Förderprogramme besser abzuschätzen, indem Strukturmerkmale, Konjunkturen/Entwicklungsbrüche sowie die generelle Impulsstärke der Förderung in Betracht gezogen werden. Ein zweites Element ist ein dynamisierter Zielkatalog, mit dem die Ziele der Förderprogramme und deren Fortentwicklung datenbankgestützt dokumentiert werden. Das dritte Element besteht aus einer Ergebnisdatenbank, die den im Zielkatalog definierten Zielen zugeordnete Indikatoren enthält. Ein besonderer Schwerpunkt wird dabei auf nicht intendierte Effekte gelegt, auch fließen semantische Webanalysen als Indikatoren ein. Das vierte Element des Evaluationssystems besteht in der Strukturierung der Indikatoren, indem Effekte dahin gehend unterschieden werden, ob sie auf den Kreis der Förderteilnehmer, Nichtförderteilnehmer oder beide gleichzeitig wirken. Im Ergebnis wurden Indikatortabellen erarbeitet, die zeigen, welche geförderten Akteure (z.B. Unternehmen, Forschungseinrichtungen) oder nicht geförderte Akteure (z.B. IKT-Branche, Endanwender, Gesamtwirtschaft) mit dem Indikator erreicht werden. Die Indikatoren müssen an die jeweiligen Maßnahmen angepasst werden. In der Zusammenfassung wies Herr Astor u.a. darauf hin, dass Teildaten für die Erhebung durch ein kontinuierliches Monitoring seitens des PT oder durch eine Begleitforschung erhoben werden können. Die Rolle externer Evaluierender liege in der neutralen Erhebung und Bewertung der aggregierten Daten.

In der durch Jan Wessels und Sonja Kind moderierten Plenumsdiskussion zu individuellen Erfahrungen und Strategien der Teilnehmenden wurden Vorteile, Nachteile und Konfliktpotenziale diskutiert, die sich aus den veränderten Rollen von Projektträgern im Innovationsprozess ergeben können.

Hinsichtlich der Konflikte wurde zunächst angesprochen, dass PTs gegenüber dem Auftraggeber (in der Regel Ministerien) oftmals als ‚Gatekeeper‘ agieren. Es sei daher die Frage zu stellen, wie neutral PTs in der Praxis als Evaluator auftreten können. Gerade aus der Konkurrenz zwischen den PT-Einrichtungen ergibt sich ein Konfliktpotenzial, wenn sich konkurrierende PTs nicht ‚in die Karten‘ schauen lassen wollen. Eine weitere Herausforderung liegt darin, dass interne Evaluatoren gegebenenfalls nicht die notwendige Deutungshoheit für weitgehende Empfehlungen haben. Der PT hat in der Regel ein Interesse an einer guten Bewertung,

daher stellt sich immer die Frage, wie unabhängig interne Evaluationsteams tatsächlich arbeiten können. Projektmanager(innen) könnten versuchen, die internen Evaluationsergebnisse so zu beeinflussen, dass die sensiblen Punkte der Evaluation nicht nach außen treten. Bei PTs sind auch hinsichtlich des Datenschutzes Konflikte denkbar, z.B. indem Evaluators(inn)en Zugang zu vertraulichen Daten haben, diese jedoch nicht für die Evaluation einsetzen dürfen. Auch die Evaluation der Vollzugswirtschaftlichkeit birgt sowohl intern als auch extern Konfliktpotenzial.

Mit Blick auf eine aktive Rolle der PTs wurde hervorgehoben, dass der Aufbau eines systematischen Programmmonitorings umfassende Vorteile mit sich bringen kann. Ein solches Monitoring könnte dazu führen, dass das Berichtswesen zur Projektförderung so zugeschnitten wird, dass es sehr viel besser den Evaluationsanforderungen gerecht wird. Klare Vorteile haben PTs hinsichtlich der Kenntnis der Programminterna und Komplexität der Projektförderung. Die Evaluationsergebnisse lassen sich direkt zur strategischen Steuerung bzw. Weiterentwicklung der Programme einsetzen. Auch die Hoffnung auf ein besseres ziel- und wirkungsorientiertes Forschungsmanagement wird zum Teil an Evaluationen durch PTs geknüpft. Profitieren könnten Evaluationen auch von der Nähe zu ihren Auftraggebern. Die umfassende Kenntnis der Spezifika der Projektförderung und der Programmlandschaft macht die Ableitung strategischer Handlungsempfehlungen gegebenenfalls einfacher.

Die Nachteile einer aktiven Rolle der PTs bei Evaluationen sahen einige Diskutant(inn)en darin, dass PTs weniger gut in der Lage sind, einen neutralen Vergleich zu Programmen anderer PTs herzustellen. Auch wäre es notwendig, dass PTs im Bereich Evaluation erst Methodenkompetenzen aufbauen müssen. Die Mitarbeiterinnen und Mitarbeiter des PTs könnten eine stärkere Pfadabhängigkeit bei der Bewertung von Projektträgerarbeit an den Tag legen oder deutlicher durch den Fördergeber beeinflusst sein. Fraglich ist auch, ob PTs nicht zu sehr in Programmstrukturen verhaftet sind und daher gegebenenfalls kaum in der Lage sein werden, ein Programmportfolio zu evaluieren. Die Ergebnisse könnten von dem Wunsch geleitet sein, positive Evaluationsergebnisse zu produzieren. Es würde PTs zudem schwerer fallen, interne Evaluationsergebnisse glaubhaft extern zu vermitteln.

Als Fazit lässt sich zusammenfassen, dass eine aktivere Rolle von Projektträgern insbesondere beim Monitoring, aber auch bei der aktiven Begleitung von externen Evaluatoren von einer Mehrheit der Teilnehmenden sehr positiv gesehen wurde, ein Ineinanderfallen der Rollen von PTs und Evaluatoren in einer Institution gleichwohl weiterhin skeptisch kommentiert wurde. Angesichts der veränderten Erwartungen der Ministerien an die PTs wird dieses Thema aber auch in den kommenden Jahren noch an Relevanz zunehmen. Die Diskussion geht also sicherlich weiter.

Die Nutzung von Wissensressourcen und von Zusammenarbeit für die Weiterentwicklung der Evaluationspraxis

Tagungsbericht zur Frühjahrstagung 2016
des AK Professionalisierung

Dörte Schott,¹ Christine Nolte,² Manuel Merz³

Thema

*„If you and I swap a dollar, you and I still each have a dollar.
If you and I swap an idea, you and I have two ideas each.“
(Gloor 2006)*

Am Freitag, den 08. April 2016, fand die Frühjahrstagung des AK Professionalisierung in Berlin beim Gastgeber Wikimedia Deutschland – Gesellschaft zur Förderung Freien Wissens e.V. statt. Das Transdisziplinäre Kolloquium stand ganz im Zeichen der Professionalisierung der Evaluationspraxis. Beitragende und Gäste haben sich gemeinsam auf die Suche nach Ressourcen begeben, welche *jenseits von etablierten* Aus- und Weiterbildungsmöglichkeiten, Zeitschriften oder Fachtagungen zur Verbesserung und *Weiterentwicklung der Evaluationspraxis* beitragen können. Zu derartigen Ressourcen sind unter anderem zu zählen: freie Enzyklopädien oder Glossare der Evaluationsbegriffe nach dem Beispiel von Wikipedia, Forschungs- und Literaturdatenbanken oder andere Internetportale, auf denen Informationen und Multimediaangebote zum Thema Evaluation vorgehalten werden, aber auch verschiedene Formate und Möglichkeiten des zwischenmenschlichen Austausches sowie gezielter Zusammenarbeit.

Ziel des Kolloquiums war es, in einen Austausch darüber zu kommen, welche frei oder leicht zugänglichen Wissensressourcen über Evaluation zur Weiterentwicklung der Evaluationspraxis genutzt und möglicherweise noch gewünscht werden und

1 Freie Evaluatorin, Bonn
2 Centrum für Evaluation (CEval), Saarbrücken
3 Wikimedia Deutschland, Berlin

welche Potenziale im zwischenmenschlichen Austausch oder in der gezielten Zusammenarbeit liegen.

Zu Beginn ihres kurzen **Einführungsvortrages** ging Dörte Schott auf die nicht zufällige Nähe zwischen Tagungsthema und Tagungsort ein. Nachdem sie mögliche Formen von (freien) Wissensressourcen und zwischenmenschlichem Austausch angerissen hatte, schilderte sie ausgewählte wissenschaftliche Erkenntnisse zur Intelligenz und Kreativität von sogenannten ‚Schwärmen‘ sowie ihre Erfahrungen damit, welche in Form einer Fallstudie vorliegen (vgl. Schott/Beywl 2015). Abschließend skizzierte sie vier Spannungsfelder, welche sich aus ihrer Sicht bei der zwischenmenschlichen Kollaboration, wie im beschriebenen Fallbeispiel, ergeben können: Neben der fachlichen und menschlichen Bereicherung birgt Zusammenarbeit gegebenenfalls zusätzliches Konfliktpotenzial, welches sich in den Abstimmungen und Interaktionen offenbaren kann (Reichtum versus Konflikt/Störung). Aus der Kollaboration entstehende Synergien werden möglicherweise durch den Zusatzaufwand für erforderliche Abstimmungen oder Überarbeitungen geschmälert (Synergien versus Zusatzaufwand). Fachlicher Dissens kann statt der angestrebten inhaltlich hohen Qualität, zumindest aus der Sicht Einzelner, eventuell eine Einigung auf ‚kleinstem gemeinsamem Nenner‘ zur Folge haben. Nicht zuletzt wird in der wissenschaftlichen Literatur im Kontext unterschiedlicher Fächer eine Debatte darüber geführt, ob Kollektive im Einzelfall tatsächlich intelligent oder doch eher ‚dumm‘ sind, das heißt einer durch die soziale Interaktion herbeigeführten systematischen Fehleranfälligkeit unterliegen (kollektive Intelligenz versus kollektive ‚Dummheit‘/group bias) (vgl. Mollick/Nanda 2015; Surowiecki 2004). Vor diesem Hintergrund könnten folgende Fragen diskutiert werden: Was sind Voraussetzungen und Erfolgsbedingungen für kollektive Intelligenz? Wie können förderliche Bedingungen beispielsweise in Evaluationsprojekten unterstützt werden? Welche Möglichkeiten und Herausforderungen bestehen im Hinblick auf das Ergebnis der Zusammenarbeit in Form von Produkten wie Evaluationsberichten?

In der anschließenden **interaktiven Einstiegsphase** gab es für alle Tagungsgäste die Möglichkeit, sich in das Thema hineinzudenken und sich mit ihren Sitznachbar(inne)n vorwiegend zu drei Leitfragen auszutauschen. Im Plenum wurden die mit den Sitznachbar(inne)n diskutierten Inhalte und Fragen anschließend zusammengetragen:

1) **Auf welche Art von Wissensressourcen greifen Sie im beruflichen Alltag gerne und wiederholt zurück?** Hier wurden jenseits der etablierten Ressourcen in Form von Büchern, Fachliteratur, Onlinepublikationen, Literaturdatenbanken, Bibliotheken (z.B. an Hochschulen) sowie Austausch auf Tagungen, Konferenzen und innerhalb der eigenen Organisation folgende im Bereich der Evaluation weniger verbreitete Ressourcen benannt:

...als Informationsquellen:

- a. Internet allgemein, freie Suche im Internet via Suchmaschine/ ‚Googeln‘
- b. wikipedia.org/eval-wiki.org, u.a. um die Literaturangaben als Ausgangspunkt für weitere Recherchen zu nutzen
- c. Videos in Onlineportalen, z.B. youtube.com

...sowie folgende Ressourcen zum Mitlesen und Austauschen:

- d. Mailingliste (z.B. forum-evaluation)
 - e. Austausch mit Kolleg(inn)en außerhalb der eigenen Organisation, beispielsweise über Stammtische oder Regionalgruppen
 - f. Blogs oder soziale Netzwerke (z.B. LinkedIn)
- 2) **Welche Situationen in Ihrem beruflichen Alltag haben in der Vergangenheit am ehesten Lern- oder Aha-Effekte ausgelöst?** Auf diese Frage wurden mit breiter Zustimmung fachliche oder berufliche Herausforderungen, Schwierigkeiten/ ‚etwas klappt nicht‘ oder auch ‚Schocks‘ genannt, welche dann jeweils eine Suche nach sich zogen, wie man etwas anders oder besser machen könne. Nicht nur in solchen Situationen wird von einigen Tagungsteilnehmenden Feedback in einem geschützten Rahmen (z.B. innerhalb des Teams) zwecks Weiterentwicklung eingeholt. Andere begeben sich stetig auf die Suche nach neuen Situationen zum Lernen, da sie Lernen als motivierend empfinden.
- 3) **Welche Fragen bewegen Sie persönlich in Bezug auf das Tagungsthema?** Es gab ganz unterschiedliche Beweggründe und Erkenntnisinteressen unter den Teilnehmenden. Mehrmals wurde beispielsweise die Problematik der fehlenden übergeordneten Qualitätskontrolle im Internet genannt – ein Aspekt, welcher im Verlauf des Kolloquiums mehrfach aufgegriffen und diskutiert wurde. Einige Teilnehmende wiesen darauf hin, dass es für Recherchierende teilweise schwierig ist, fundierte Informationsquellen sicher zu identifizieren. In diesem Kontext tauchte die Frage auf, welche Möglichkeiten des Umgangs mit Informationsüberflutung es gibt, welche eine Recherche im Internet beispielsweise mit sich bringen kann. Des Weiteren wurden Zweifel formuliert, inwiefern ein offener Informationsaustausch auf einem umkämpften Markt, wie dem Evaluationsmarkt, überhaupt möglich ist.

Auf einige dieser Fragen und Anmerkungen konnten im weiteren Tagungsverlauf Antworten oder Hinweise gegeben werden. Im Rahmen des unmittelbaren Austausches im Plenum wurden darüber hinaus weitere mit dem Thema verwandte Fragen und Probleme angesprochen: Wie kann man das erworbene Wissen privat archivieren? Führt direkter persönlicher Austausch zu einem besonders nachhaltigen Lerneffekt? Warum funktioniert persönlicher Austausch zum Teil sehr gut und zum Teil gar nicht? Es wurde festgestellt, dass es in der Evaluations-Community nicht ‚die‘ für alle eindeutige Fachsprache gibt und nicht jeder die Begriffe mit dem gleichen Verständnis benutzt. Über gute Erfahrungen wurde mit der Plattform stackoverflow.com berichtet, welche eine Mischung aus Forum und Wiki zum Thema ‚Softwareentwicklung‘ ist. Dort werden viele fachliche Fragen gestellt und beantwortet. Die Antworten werden von den Plattformnutzer(inne)n bewertet, wonach die hilfreichen Antworten sich nach oben sortieren und dadurch bei weiteren Recherchen schneller gefunden werden können. Laut Erfahrungsbericht tritt in diesem Zuge erstaunlich viel Fachwissen zutage, so dass sich die Frage stellt, inwiefern diese Form des Austausches und der Wissensvermittlung auch in Bezug auf andere Themenbereiche wie Evaluation nützlich wäre.

Es folgten drei **Tagungsbeiträge**, welche das Tagungsthema aus unterschiedlichen Perspektiven aufgriffen: Inspirationen und Herausforderungen aus der offenen Zusammenarbeit innerhalb der Wikimedia-Bewegung (Manuel Merz, Wikime-

dia Deutschland, Berlin), Entstehung einer frei zugänglichen Ressourcensammlung im Evaluationsbereich (Katharina Klockgether, Univation GmbH, Köln) sowie Erfahrungen mit dem fachlichen Austausch im Rahmen eines sogenannten „Evaluationsstammtisches“, vorgestellt von Jutta Wolff (Institut für Bildungsmonitoring und Qualitätsentwicklung, Hamburg), Ralf Schattschneider und Edith Halves (Hochschule für Angewandte Wissenschaften, Hamburg).

Tagungsbeitrag 1: Die Wikimedia-Bewegung als ‚Community of Practice‘ im Evaluationsbereich

Manuel Merz stellte eingangs die gemeinnützige Wikimedia Foundation als Träger verschiedener Projekte vor, die sich für die Erstellung, Sammlung und Verbreitung freier Inhalte einsetzen (z.B. die freie Online-Enzyklopädie Wikipedia als bekanntestes Projekt). Die tragende Vision von Wikimedia ist es, das gesammelte Wissen der Menschheit jedem frei zugänglich zu machen. Mit Blick auf das Tagungsthema ging er auf fünf Inspirationen und nachfolgend auf fünf Spannungsfelder ein. Die **Inspirationen** lehnen sich an Prinzipien von Wikimedia an und lassen sich wie folgt charakterisieren:

1. Das Wiki-Prinzip: Jeder hat Änderungsrechte. Das funktioniert, weil jede Änderung für alle transparent protokolliert wird und stets von jedermann wieder rückgängig gemacht werden kann. Es gibt wenige (technisch) vorgegebene Strukturen; die Regeln der Zusammenarbeit werden im Verlauf der Zusammenarbeit ausgehandelt. Hierdurch entstehen zahlreiche Varianten gemeinsamer Arbeit.
2. Freie Lizenzierung: Alle Inhalte und Ressourcen stehen bei Wikimedia unter einer freien Creative Commons Lizenz (vgl. https://de.wikipedia.org/wiki/Creative_Commons). Diese Lizenz erlaubt es anderen, die Inhalte nach Belieben zu ändern und – auch kommerziell – zu verbreiten, sofern die Bedingungen der Lizenzen eingehalten und die Inhalte wieder unter gleichen Lizenzen veröffentlicht werden. Persönlichkeits- und Markenrechte bleiben davon unberührt.
3. Freier Zugang: Niedrige Zugangshürden gelten nicht nur für die in den Wikimedia-Projekten erstellten Inhalte, sondern auch für verschiedene Logdaten wie z.B. anonymisierte Nutzungsdaten. Basierend auf diesen Daten haben Freiwillige aus der Wikimedia-Bewegung Software-Tools programmiert, die eine datenbasierte Bewertung der eigenen Arbeit ermöglicht (z.B. zum Erfolg von thematischen Fotowettbewerben). Die Software-Tools werden wiederum frei zur Verfügung gestellt und können von Dritten für eigene Forschungsfragen genutzt werden.
4. Transparenz: Die Wikimedia-Projekte werden von gemeinsamen, starken Werten getragen, darunter Transparenz, Partizipation und Wirkungsorientierung. Nahezu alle Diskussionen und Entscheidungen finden öffentlich statt oder werden öffentlich dokumentiert. Die allgemeine Grundorientierung „etwas zu bewegen“ (Impact) trägt zu einem relativ hohen Stellenwert von Evaluation und Analyse bei.

5. Diversität: Ziel von Wikimedia ist eine weltweite Bewegung, auch wenn zurzeit noch ‚weiße Flecken‘ auf der Weltkarte sind (z.B. auf dem afrikanischen Kontinent). Die daraus folgende Diversität erfordert ständige Kommunikation, unter anderem in Form von internationalen Konferenzen. Auch beim gemeinschaftlichen Aushandeln von Evaluationsstandards zeigen sich unterschiedliche Bedürfnisse und Perspektiven.

Bei der geschilderten Art und Weise der Zusammenarbeit entstehen folgende **Spannungsfelder**:

1. Die lebhafte und transparente Diskussionskultur erfordert einen hohen Aufwand; es kommt nicht selten vor, dass Entscheidungen offen bleiben. Im positiven Fall können Widersprüche für Weiterentwicklung und gemeinsames Lernen genutzt werden.
2. Ein weltweites Netzwerk von motivierten Freiwilligen erschafft große Dinge; es besteht jedoch die Gefahr, dass besonders aktive Freiwillige wegen steigender Anforderungen irgendwann ausfallen (Volunteer-Burnout).
3. Wikimedia-Organisationen tauschen sich untereinander aus und lernen voneinander, auch zum Thema Evaluation. Gleichzeitig konkurrieren sie jedoch um Gelder und Status; was grundsätzlich Konfliktpotenzial birgt.
4. Mit der Möglichkeit, Gelder bei der Wikimedia Foundation zu beantragen, gewann auch die Evaluation der Wikimedia-Bewegung an Bedeutung. Durch die Verbindung von Finanzierung und Evaluationsstrategien ‚in einer Hand‘ entsteht ein gewisses Konfliktpotenzial, unter anderem dadurch, dass die Evaluationskriterien schließlich Top-Down gesetzt wurden und einen einseitigen Charakter haben. Kritisch betrachtet stellen die gewählten Kriterien den vermeintlich ‚kleinsten gemeinsamen Nenner‘ aller Beteiligten dar und erfassen eher ‚leicht Messbares‘ als ‚eigentlich Relevantes‘.
5. Die dezentrale Erstellung von vielfältigen Ressourcen durch eine weltweite Community führt zur Unübersichtlichkeit. Das *Auffinden* von Ressourcen ist folglich oft die größte aller Schwierigkeiten bei deren Nutzung.

Ein Fazit des Vortrages von Merz war, dass die Wikimedia-Bewegung sich hochgesteckte, idealistische Ziele setzt, die in der Praxis regelmäßig zum ‚Scheitern‘ führen, was an sich jedoch nicht unbedingt als Problem gesehen wird und werden sollte. Scheitern sei erlaubt.

An den Vortrag schloss sich eine angeregte Diskussion vor allem zu den Fragen der Qualitätssicherung bei Wikipedia und dem Schutz vor Missbrauch und Manipulation freier Inhalte an. Manuel Merz erklärte, dass Qualitätssicherung stattfindet, indem Autor(inn)en sich gegenseitig kontrollieren und stets Quellenangaben machen müssen. Denn Wikipedia sammelt ausschließlich Wissen, das es in der Außenwelt bereits gibt. Alle Beiträge können darüber hinaus in den Diskussionen zu den jeweiligen Artikeln hinterfragt werden. Missbrauch oder Manipulation kann allerdings nicht vollständig verhindert werden, da nicht alle Mitwirkenden jederzeit und ausschließlich gute Absichten hätten. Gleichwohl haben sich hilfreiche Regeln herausgebildet und gewählte Administrator(inn)en übernehmen diesbezüglich viel Verantwortung. Ein Fazit aus der Diskussion war, dass dem Umgang mit Information und

entsprechenden Kompetenzen bei dieser Art der Kollaboration eine hohe Bedeutung zukommt.

Tagungsbeitrag 2: Das Eval-Wiki heute, gestern, morgen

Katharina Klockgether berichtete über die Entstehung des Evaluations-Wikis (eval-wiki.org) auf Grundlage des Glossars der Evaluationsbegriffe von Univation, welches unter Nutzung der freien MediaWiki-Software (ebenfalls ein Wikimedia-Projekt) kostenlos nutzbar im Internet angeboten und von Univation gepflegt wird. In der dazugehörigen Glossar-Werkstatt ist es registrierten Nutzenden möglich, neue Begriffe vorzuschlagen und Werkstattbegriffe zu bearbeiten.

In den letzten Jahren entwickelte sich das Wiki zu einer Ressourcensammlung, welche als Informationsquelle und wachsender Wissensspeicher für die Evaluations-Community gedacht ist. In diesem Zuge sind weitere Evaluationswerkzeuge hinzugefügt worden: jüngst eine Quellensammlung, einige Evaluationschecklisten sowie der Programmbaum von Univation als logisches Modell und generisches Programmmodell für den Einsatz in Programmevaluationen. In die Website eingebunden wurde des Weiteren das forum-evaluation, welches als Mailingliste bereits seit 1997 von Univation in Kooperation mit der DeGEval betrieben wird.

Abschließend äußerte Frau Klockgether den Wunsch nach Anregungen, wie das Portal eval-wiki.org gestaltet werden bzw. was es beinhalten sollte, damit potenzielle Nutzer(innen) einen hohen Nutzen daraus ziehen könnten.

Tagungsbeitrag 3: Stammtisch – aber nicht Stammtisch-Niveau! Nutzen und Gelingensbedingungen des Hamburger Evaluationsstammtisches

Jutta Wolff, Ralf Schattschneider und Edith Halves stellten ihre Befragung unter Beteiligten und Gästen des Hamburger Evaluationsstammtisches zu erzieltm Nutzen und Gelingensbedingungen vor. Als zentraler Nutzen des Hamburger Stammtisches für die Befragten ergeben sich die Erweiterung von Fachwissen und Perspektiven, der Erfahrungsaustausch und die Reflexion des Evaluationsverständnisses.

Die ‚Kümmerer‘ des Stammtisches werden von den Autor(inn)en als wichtige Gelingensbedingung vorgestellt: Diese laden regelmäßig zu Treffen ein, kontaktieren potenzielle Interessent(inn)en oder Beitragende, koordinieren die inhaltliche Ausgestaltung und denken mit einem erweiterten Planungshorizont voraus. Den Befragten sind im Hinblick auf ihre künftige Teilnahme vor allem die Arbeitsatmosphäre und die behandelten Inhalte wichtig. An dieser Stelle identifizieren die Autor(inn)en zugleich zwei Baustellen, an welchen sie in Zukunft weiter arbeiten möchten: Wie kann die Vertrauensgrundlage für den offenen Austausch gestärkt werden (z.B. über Regelungen zur Vertraulichkeit)? Wie kann die Relevanz des Austausches für den Arbeitsalltag der Beteiligten stärker herausgearbeitet werden?

Fragen, die perspektivisch auf den Stammtisch zukommen, sind unter anderem, wie die wachsende Teilnehmendenzahl mit einer gewissen Kontinuität, einem ‚verlässlichen‘ Rahmen des Stammtisches vereinbar ist oder in welchem Verhältnis vorbereitete Beiträge und offener Austausch bzw. kollegiale Beratung stehen sollten. Es schließt sich die Frage an, inwiefern Kontinuität per se in diesem Kontext überhaupt erstrebenswert ist bzw. welche Bedingungen stabil bleiben sollten, während andere sich ändern können.

Die weitere Entwicklung des Stammtisches, und vielleicht das Entstehen weiterer Stammtische, bleibt mit Spannung zu erwarten.

Nach einer Pause tauschten sich alle Tagungsteilnehmenden an zwei Thementischen zu von ihnen mitgebrachten sowie auf der Tagung gewonnenen Ideen, Erfahrungen und Fragen aus.

Thementisch A – Zur Nutzung und Weiterentwicklung von Wissensressourcen:

Als frei zugängliche Wissensressource wurde eine Datenbank mit Evaluationsberichten und Verschlagwortung gewünscht, in welcher man nach Lösungen für Problemstellungen suchen kann. Dem Vernehmen nach existiert bereits ein solcher Versuch, welcher jedoch derzeit (noch) nicht öffentlich gemacht wurde bzw. nicht frei zugänglich ist. In diesem Zusammenhang wurde diskutiert, inwiefern Problemlösungen eher durch Reflexion, z.B. im zwischenmenschlichen Austausch, oder das Einnehmen einer Metaebene in der Betrachtung, unterstützt wird als durch die reine Lektüre von Beispielberichten. Daraus entstand die Idee, typische Herausforderungen in der Evaluation und deren Lösungen in einer Art Datenbank zusammenzutragen. Hierbei könne auch das Lernen aus Fehlern gefördert werden. Aus dem medizinischen Bereich ist beispielsweise das Sammeln anonymer Fehlerberichte/Fehlereingeständnisse bekannt, mit dem Zweck aus diesen – teilweise folgenschweren – Fehlern lernen zu können. Als Voraussetzung dafür wird eine ‚Fehlerkultur‘ (d.h. konstruktiver Umgang mit Fehlern) gesehen, welche ein solches Lernen aus Fehlern ermögliche – leider jedoch nicht unbedingt vorausgesetzt werden könne, wie in E-Mail-Foren o.ä. teilweise spürbar sei.

Als weiterer Punkt wurde diskutiert, dass – insbesondere für den Start einer Recherche – strukturierte oder geordnete Information besonders nützlich ist. Kommentierte Literaturlisten beispielsweise, welche thematisch sortiert und mit Empfehlungen versehen sind, erscheinen nützlicher als lose oder alphabetische Aufzählungen. Weitere Hilfe bei der Einordnung von Literatur bieten beispielsweise qualifizierte Rezensionen. Wenn die gesuchten Eigenschaften einer Quelle sich nicht nachschlagen lassen, wie bei bestimmten Eigenschaften von Audio- oder Videoquellen, können nur Praxistipps von Erfahrenen weiterhelfen. In E-Mail-Foren findet man schneller die gesuchte Lösung, wenn die Antworten von den Nutzenden nach Nützlichkeit (z.B. „Wie hilfreich finden Sie diesen Beitrag?“) bewertet und entsprechend sortiert werden.

Es wurde festgestellt, dass verschiedene Formen von Informationsquellen und Austauschmöglichkeiten sich gut ergänzen lassen zur (fachlichen) Weiterentwicklung: Literatur und schriftliche Quellen würden sich gut mit zwischenmenschlichem

Austausch ergänzen; für die Problemlösung sei oft eine Kombination aus Sachwissen und Erfahrungswissen optimal. Zwischenmenschlicher Dialog werde dann unentbehrlich, wenn eine gemeinsame Gesprächsbasis erst noch hergestellt werden müsse – wie es bei der Evaluation als interdisziplinärem Bereich des Öfteren der Fall sei.

Thementisch B – Zu Chancen und Herausforderungen der Zusammenarbeit: Diskutiert wurden die Chancen und Herausforderungen der Zusammenarbeit für das Lernen, zum einen hinsichtlich des direkten Austauschs, z.B. in Form von Stammtischen, zum anderen hinsichtlich eines Austauschs im digitalen Raum.

Als größter Vorteil für den direkten persönlichen Austausch wurde das dabei entstehende Vertrauensverhältnis genannt. Gerade die Stammtische bieten hier Vorteile, da Probleme in offiziellen Kontexten eher nicht angesprochen würden. Viele Praxisprobleme der Evaluierenden ließen sich zudem nicht auf Tagungen besprechen, da die dort besprochenen Inhalte meist zu theoretisch oder abstrakt sind. Stammtische stellen eine gute Ergänzung zu Tagungen dar, um zum Beispiel praktische Probleme, etwa das Verhältnis zum Auftraggeber, besprechen zu können. Zudem wurden die Stammtische als niedrighschwelliges Angebot beschrieben, welches sich auch gut für Einsteiger(innen) eignet. Ein letzter Vorteil des Stammtisches wird in der Tatsache gesehen, dass die Themenbereiche aufgrund der Teilnehmenden aus unterschiedlichen Bereichen (EZ, Bildung, Soziale Dienstleistungen...) sehr divers sind und sich somit neue Verbindungen finden lassen und interdisziplinäres Lernen ermöglicht wird.

Als Nachteil wurde diskutiert, dass diese Art des niedrighschwelligem persönlichen Austausches zumeist nur lokal möglich ist. Ein weiterer Schwachpunkt sei das Wissensmanagement, da bei den – bisher – ehrenamtlich koordinierten Stammtischen zu wenig Kapazität vorhanden ist, um Ergebnisse ausreichend zu dokumentieren.

Hinsichtlich des Austausches im digitalen Raum wurden Möglichkeiten der digitalen Unterstützung der Stammtische diskutiert, z.B. eine Mailingliste. Auch hier wurde eine Fehlerdatenbank angesprochen, in der Fehler, welche von Evaluierenden gemacht wurden, dokumentiert werden, so dass diese Fehler zum Lernen genutzt und nicht noch einmal wiederholt werden. Als kritisch wurde jedoch die Diskussionskultur im digitalen Raum gesehen; mit bestimmten Fragen möchte man sich daher nicht unbedingt im Internet ‚outen‘. Spezifische Fragen schriftlich zu formulieren wird zudem als sehr zeitaufwändig und generell schwierig angesehen. Viele Teilnehmende wünschen sich darüber hinaus eher einen direkten persönlichen Austausch als Ausgleich zum digitalisierten Büroalltag.

Ein Fazit der Diskussion war, dass digitale Angebote, etwa Wikis, optimal für die Bereitstellung von Informationen sind, wohingegen direkte Zusammenarbeit besser für den Austausch von Erfahrungen geeignet ist.

Mit einem kurzen **Resümee aus dem gemeinsamen Austausch** schloss Dörte Schott die Tagung. Ein interessantes Ergebnis des gemeinsamen Austausches sei, dass Lernprozesse am ehesten beim Bewältigen von Herausforderungen oder ‚Schocks‘ und der damit einhergehenden Lösungssuche auftreten oder bewusst wer-

den. Die Reflexion von Fachinhalten auf der Metaebene – sei es durch die Bewertung von Informationen in einem Onlineportal oder im niedrigschwelligen Austausch unter Kolleg(inn)en – scheint ein wichtiger Faktor für Lernen und Weiterentwicklung zu sein. Die gezielte Nutzung einer Vielfalt von Ressourcen (Informationsquellen und Austauschmöglichkeiten) erweist sich für die Weiterentwicklung besonders fruchtbar, da unterschiedliche Ressourcen unterschiedliche Stärken und Schwächen haben, welche sich gegenseitig ergänzen und ausgleichen können. Eine Kompetenz im Umgang mit vielfältigen Informationen ist dabei sicherlich nicht nur hilfreich, sondern des Öfteren auch ausbaufähig.

Gemessen an den Diskussionen, Gesprächen und Rückmeldungen war es für viele Teilnehmende eine inspirierende Tagung, aus welcher sie einige Anregungen für sich mitnehmen konnten.

Literatur

- Gloor, Peter A. (2006): *Swarm Creativity*. Oxford: Oxford University Press.
- Mollick, Ethan R./Nanda, Ramana (2015): *Wisdom or Madness? Comparing Crowds with Expert Evaluation in Funding the Arts*. Working Paper 14-116. Harvard Business School.
- Schott, Dörte/Beywl, Wolfgang (2015): *Schwarm-Evaluation als Steuerungsansatz*. In: Giel, Susanne/Klockgether, Katharina/Mäder, Susanne (Hg.): *Evaluationspraxis: Professionalisierung – Ansätze – Methoden*. Münster: Waxmann.
- Surowiecki, James (2004): *Die Weisheit der Vielen: Warum Gruppen klüger sind als Einzelne und wie wir das kollektive Wissen für unser wirtschaftliches, soziales und politisches Handeln nutzen können*. München: Bertelsmann.

Veranstalter

Arbeitskreis Professionalisierung in der DeGEval – Gesellschaft für Evaluation e.V. (www.degeval.de/arbeitskreise/professionalisierung-ehemals-aus-und-weiterbildung-in-der-evaluation/)

Kontakt

Dr. Dörte Schott (info@doerte-schott.de)

Evaluation und politische Gestaltungs- und Entscheidungsprozesse

Frühjahrstagung 2016 des AK Soziale Dienstleistungen in der DeGEval in Kooperation mit dem Forschungsinstitut für gesellschaftliche Weiterentwicklung (FGW)

Stefanie Reiter,¹ Stefan Schmidt,² Rainer Strobl³

Tagungsthema und Format

Evaluationen können wichtige Informationen für politische Entscheidungsträger liefern, die deutlich über die Legitimationsfunktion der Evaluation hinausgehen. Eine Diskussion über diese umfassende Form der Evaluation war das Ziel einer gemeinsamen Frühjahrstagung des Arbeitskreises (AK) Soziale Dienstleistungen in der DeGEval in Kooperation mit dem Forschungsinstitut für gesellschaftliche Weiterentwicklung (FGW). Die Fachtagung zum Thema „*Evaluation und politische Gestaltungs- und Entscheidungsprozesse*“ fand am 15./16. März 2016 in Düsseldorf statt.⁴ Bei der Tagung stand die Fragestellung im Zentrum, wie Evaluation zur gesellschaftlichen Weiterentwicklung beitragen kann. Dabei wurden die Möglichkeiten und Grenzen einer Verzahnung zwischen Evaluations- und politischen Entscheidungsprozessen beleuchtet und die Frage behandelt, welche Rationalitätsgewinne die Evaluation bei politischen Entscheidungen verspricht und was daraus für die Zusammenarbeit von Programmentscheidern aus Politik und Verwaltung auf der einen Seite und Evaluationsexpertinnen und -experten auf der anderen Seite folgt. Das Tagungsformat beinhaltete Impulsreferate von Expertinnen und Experten aus diesen unterschiedlichen Bereichen sowie ausführliche Diskussionen zu den aufgeworfenen Themen im Plenum und in Kleingruppen. Insgesamt nahmen rund 80 Vertreterinnen und Vertreter aus den Bereichen Politik, Verwaltung, Zivilgesellschaft und Evaluation an der Tagung teil. Referierende aus Österreich und der Schweiz gaben Einblicke in aktuelle Entwicklungen und Diskurse der Nachbarländer und eröffneten neue Perspektiven auf die Thematik. Die Fachtagung bot somit Vertreterinnen und Vertretern

1 Deutsches Jugendinstitut e.V., Halle an der Saale

2 schmidt evaluation, Köln

3 proVal, Gesellschaft für Sozialwissenschaftliche Analyse – Beratung – Evaluation, Hannover

4 Die Frühjahrstagung des AK Soziale Dienstleistungen wurde von Olaf Lobermeier (proVal) und dem Sprecherteam des AK Soziale Dienstleistungen, bestehend aus Rainer Strobl (proVal), Stefanie Reiter (DJI) und Stefan Schmidt (schmidt evaluation), in Kooperation mit Wilhelm Heitmeyer (Universität Bielefeld) und dem FGW organisiert.

aus Politik, Verwaltung, Zivilgesellschaft und Evaluation eine Gelegenheit, Fragestellungen zum Zusammenwirken von Evaluation und politischen Gestaltungs- und Entscheidungsprozessen aus unterschiedlichen Perspektiven zu diskutieren und gemeinsam über Grenzen sowie Potenziale nachzudenken.

Impulsreferate und Berichte aus der Evaluationspraxis sowie aus Politik und Verwaltung

Nach dem Grußwort von Svenja Schulze (Ministerin für Innovation, Wissenschaft und Forschung des Landes Nordrhein-Westfalen), welche die Veranstaltung als Chance eines moderierten Dialogs zwischen Wissenschaft, Politik und Zivilgesellschaft bei der Evaluation von Programmen lobend hervorhob, eröffnete Wilhelm Heitmeyer (Universität Bielefeld) mit dem Vortrag „*Gesellschaftliche Weiterentwicklung, die Bedeutung von Evaluation und Problembereiche*“ die Tagung. Durch die Aufzeichnung von Diskrepanzen einer gesellschaftlichen Problementwicklung, die aus seiner Sicht von Zielkonflikten in der pluralen Gesellschaft, daraus resultierenden Dilemmata und einer Richtungslosigkeit von Entwicklungen gekennzeichnet ist, und dem produzierten wissenschaftlichen Wissen, warf er die tagungsleitenden Fragen auf: Wie kann Erkenntniswissen in relevantes Handlungswissen transferiert werden? Welche Rolle kann die Evaluation bei diesen Gestaltungsprozessen einnehmen? Der zunehmende Bedarf an Evaluationsergebnissen zeigt sich nach Heitmeyer in einem Paradigmenwechsel in der Politik, die sich stärker einer Wirkungsorientierung zuwendet, und wirkt sich zudem in andere gesellschaftliche Bereiche aus. Problematisch könne sich hierbei eine Tendenz zur Risikovermeidung aufseiten von Auftraggebenden aus dem politischen Feld sowie von Evaluierenden für die gesellschaftliche Weiterentwicklung auswirken, die unter anderem auf Zielkonflikte zwischen dem wissenschaftlichen System (Wahrheit) und der Politik (Machterhalt) zurückgeht. Es bedarf jedoch nach Heitmeyer einer gewissen Risikobereitschaft und der Entwicklung einer Fehlerkultur im Sinne eines Lernens aus Erfahrungen mit Rückschlägen oder einem Scheitern, um innovative Programme und gesellschaftliche Entwicklungen voranzutreiben. Evaluation müsse sich in diesem Kontext der Gefahr einer Instrumentalisierung entziehen und der Politik Expertenwissen für rationale Entscheidungen zur Verfügung stellen. Wichtig ist nach Heitmeyer in diesem Zusammenhang, dieses Expertenwissen auch der Öffentlichkeit zur Verfügung zu stellen, es für diese zu übersetzen und so aufzubereiten, dass Ergebnisse vor dem Hintergrund von Zielkonflikten, Dilemmata und Entwicklungen transparent und verständlich werden. Heitmeyer spannte mit seinem Eröffnungsvortrag somit einen breiten Rahmen, an den die nachfolgenden Beiträge und Diskussionen anknüpfen konnten.

Eine Erwartungsabfrage des Moderators zeigte in diesem Zusammenhang, dass insbesondere der Austausch mit Vertreterinnen und Vertretern aus den jeweils anderen Bereichen für die Teilnehmenden ein wichtiger Aspekt der Fachtagung war.

Thomas Widmer (Universität Zürich) trug zu diesem Erfahrungsaustausch mit seinem Impulsreferat⁵ „*Evaluation in der Politik*“ bei, in dem er das Zusammenspiel von Evaluation und Politik am Beispiel der Schweiz veranschaulichte und folgende Thesen zur Genese und *Nutzung von Evaluation in der Politik* zur kritischen Diskussion stellte:

1. Evaluation soll zielgerichtet und selektiv mit klarem funktionalem Fokus eingesetzt werden.
2. Evaluationen werden von der Politik oft bestellt, aber selten ‚abgeholt‘, was mitunter darauf zurückzuführen ist, dass Evaluationsergebnisse nicht schnell genug zur Verfügung gestellt werden können.
3. Evaluationsbefunde stehen bei der politischen Entscheidungsfindung in Konkurrenz zu anderen entscheidungsmächtigeren Faktoren (z.B. Ideologie, Interessen, institutionalisierte Werte und Praktiken sowie bereits zuvor existierende Informationen).
4. Politik nutzt Evaluationen und Evaluationsbefunde opportunistisch (z.B. zur Sicherung der eigenen Positionen und zu Legitimationszwecken).
5. Evaluation ist kein explizites Steuerungsinstrument, sondern dient primär der Bewertung und eignet sich nicht für einen flächendeckenden sowie permanenten Einsatz. Sie erfordert Handlungsfreiheit und fügt sich entsprechend schlecht in Strukturen hierarchischer Steuerung ein.

Die anschließende Diskussion im Plenum über Möglichkeiten und Grenzen für die Nutzung von Evaluationsergebnissen in politischen Gestaltungs- und Entscheidungsprozessen setzte sich vor allem mit der letztgenannten These zum Einsatz von Evaluationen zu Steuerungszwecken konstruktiv auseinander. Dabei wurde auch thematisiert, welche Voraussetzungen und Prozesse sich aufseiten von Evaluation und Politik dabei als förderlich auswirken können.

Die nachfolgenden Vorträge widmeten sich der Frage, welche Rolle *Wirkungsorientierung* bei Evaluationen spielt. So unterstrich Rainer Strobl (proVal, Hannover) in seinem Vortrag „*Möglichkeiten und Konsequenzen eines wirkungszentrierten Vorgehens bei der Planung, Umsetzung und Evaluation von politischen Programmen und Initiativen*“ zunächst die Bedeutung einer stärkeren Fokussierung auf Wirkungen und wies darauf hin, dass in diesem Fall auch der Bedarf an Evaluationsergebnissen steige. Im Anschluss nannte Strobl vier Hürden (bestimmte Formen der Verwaltungssteuerung; einseitige Fokussierung auf Maßnahmen und Aktivitäten; sehr allgemeine und abstrakte Wirkungsformulierungen; Überforderung bei der Planung, Umsetzung und Überprüfung bei einer Fokussierung auf Wirkungen), die einer stärkeren Ausrichtung auf Wirkungen entgegenstehen. Zur Überwindung dieser Hürden empfahl Strobl, Evaluationsexpert(inn)en frühzeitig einzubinden, um Programme und Projekte schon in der Konzeptionsphase auf Wirkungen auszurichten. Ferner komme es darauf an, dass auch bei der Umsetzung die Frage der Wirksamkeit im Mittelpunkt stehe. Strobl empfahl hier formative Evaluation und Monitoring als Maßnahmen zur Verbesserung der Prozessqualität. Zum Abschluss skizzierte er einige Möglichkeiten zur Durchführung von Wirkungsevaluationen.

5 Folien zu einzelnen Vorträgen sind der Homepage des AK Soziale Dienstleistungen in der DeGEval unter <http://www.degeval.de/arbeitskreise/soziale-dienstleistungen/> zu entnehmen.

Im Anschluss verdeutlichte Olaf Lobermeier (proVal, Hannover) diese Aspekte anhand konkreter *Beispiele für ein wirkungszentriertes Vorgehen bei der Evaluation von politischen Programmen und geförderten Projekten*. Er lieferte somit insbesondere den Teilnehmenden aus Politik und Zivilgesellschaft Einblicke in die Vorgehensweisen während der unterschiedlichen Evaluationsphasen. So erläuterte er, wie Ziele in der Konzeptionsphase konkretisiert und auf Wirkungen ausgerichtet werden können. Lobermeier plädierte in diesem Zusammenhang dafür, solche Wirkungsziele auch auf der Programmebene zu entwickeln und als Rahmen für die Programmsteuerung und für die Projektförderung zu nutzen. Im Hinblick auf die Umsetzungsphase zeigte er, wie Evaluationsergebnisse genutzt werden können, um z.B. Förderrichtlinien weiterzuentwickeln, die Beurteilung von Projektanträgen zu optimieren oder die Implementation von bestimmten Maßnahmen zu verbessern. Lobermeier erläuterte zum Schluss anhand verschiedener Beispiele, wie Wirkungen mit unterschiedlichen Evaluationsdesigns nachgewiesen werden können.

Die Einblicke aus dem Feld der praktischen Evaluationstätigkeit wurden durch den Vortrag von Ursula Rosenbichler (Bundeskanzleramt Österreich) ergänzt, welche aus Sicht der Verwaltung Chancen und Grenzen einer Wirkungsorientierung anhand der kürzlich gesammelten Erfahrungen mit der Einführung einer wirkungsorientierten Haushaltsführung in Österreich darstellte. In ihrem Vortrag *„Monitoring: Voraussetzung für wirkungsorientiertes Steuern – Evaluation: Basis für nachhaltiges Entscheiden? Erfahrungen aus der österreichischen Bundesverwaltung“* verdeutlichte sie neben positiven Erfahrungen und Potenzialen auch Herausforderungen aus der praktischen Umsetzung. Unter anderem betonte Ursula Rosenbichler, dass auch bei sogenannten evidenzbasierten Entscheidungen nicht die Erwartung an Kennzahlensysteme gestellt werden könne, automatische Entscheidungsmechanismen zu generieren. Die Daten sollen vielmehr als Informationsgrundlage in einer lernenden Organisation Reflexionsprozesse herbeiführen. Dies stelle eine komplexe Herausforderung für alle Seiten dar, die jedoch Weiterentwicklungspotenziale und Chancen mit sich bringe.

Im Anschluss an die drei Beiträge zur Wirkungsorientierung hatten die Teilnehmenden die Möglichkeit, interessante Punkte im Rahmen einer *Podiumsdiskussion mit der Referentin und den Referenten* zu diskutieren und zu vertiefen.

Dieter Filsinger (Hochschule für Technik und Wirtschaft des Saarlandes) setzte sich am Ende des ersten Tagungstages mit seinem Vortrag unter dem Titel *„Politikberatung und/oder Evaluation“* abschließend mit der Frage des Verhältnisses von Monitoring, Evaluation und (kritischer) Politikberatung am Beispiel der Integrationspolitik auseinander, um Entwicklungen nachzuzeichnen, den Bedeutungsgewinn kritischer Politikbegleitung aufzuzeigen und um eine stärkere Kooperation von Policy Analyse, Politikberatung und Evaluation, insbesondere in Bezug auf Integrationspolitik als einem Querschnittsfeld mit wachsender Bedeutung, zu werben. Er merkte kritisch zu diesem Feld an, dass es trotz einer Vielzahl von Evaluationen in der Gesamtschau an Ex-ante-Evaluationen, ‚echten‘ Wirkungsuntersuchungen und einer ressortübergreifenden Zusammenarbeit bei der Zusammenfassung von Ergebnissen fehle. Den Teilnehmenden gab er für den gemeinsamen Ausklang des Tages die Anregung mit auf den Weg, neue Formate für eine systematische Integration von Poli-

tikberatung und Evaluation zu entwickeln. Hierbei solle sich die Evaluations-Community stärker um Policy Analyse kümmern und den Blick für den gesamten Policy Circle weiten, anstatt den Fokus nur auf ausgewählte (Modell-)Programme zu richten. Dies erfordere ein Umdenken auf allen Seiten, auch auf der der Ressourcengeber.

Die Teilnehmenden nutzten nach der anschließenden lebhaften Diskussion über Potenziale und Hürden eines solchen Vorgehens die Gelegenheit zum weiteren fachlichen und kollegialen Austausch am Buffet.

Beiträge des zweiten Tagungstages

Mit einem gemeinsamen Vortrag *aus Sicht von Auftraggeberin und Evaluator* als Auftragnehmer führten Cornelia Stern (Ministerium für Schule und Weiterbildung NRW, Referat 13) und Stefan Schmidt (schmidt evaluation, Köln) in den zweiten Tagungstag ein. Sie veranschaulichten die Möglichkeiten und Grenzen einer systematischen Verzahnung zwischen einem Evaluationsprozess und einem politischen Entscheidungsprozess am Beispiel der Evaluation der regionalen Bildungnetzwerke in NRW. Dabei stellten sie das Vorgehen des Evaluationsverfahrens, Ergebnisse und folgende zentrale Erfolgsfaktoren der Evaluation vor:

1. Einbindung der relevanten Entscheidungsträger (Land, Bezirk, Kommune),
2. konsequente Ausrichtung der Evaluation auf den vordringlichen Zweck,
3. Belastbarkeit und Akzeptanz der Evaluationsergebnisse,
4. leistbarer Aufwand und Anspruch für die Verantwortlichen vor Ort,
5. Sichtbarkeit der Evaluation auf politischer Ebene,
6. Evaluationsergebnisse nutzen.

Der Beitrag illustrierte eine Form der gelungenen Zusammenarbeit von Politik und Evaluation im Sinne der Nutzung von Evaluationsergebnissen, sowohl auf Landes- als auch auf kommunaler Ebene.

Anschließend beschäftigte sich Reinhard Stockmann (Universität des Saarlandes, Centrum für Evaluation (CEval)) damit, welche *Rationalitätsgewinne die Evaluation bei politischen Entscheidungen* verspricht und was daraus für die Zusammenarbeit von Programmentscheidern und Evaluationsexpertinnen und -experten folgen kann. Ausgehend von Überlegungen zu verschiedenen Formen rationalen Handelns griff er dabei die Frage auf, wie Evaluation zu Rationalitätsgewinnen (im Sinne eines zweckrational und sachlogisch-orientierten Verhaltens), die auch der Verbesserung von politischen Programmen dienen können, beitragen kann. Nach Stockmann können Rationalitätsgewinne durch Evaluation auf drei Ebenen auftreten (dreifache Zweckbestimmung von Evaluation):

1. bei der Umsetzung von Programmen, indem entscheidungsrelevante Informationen für Programmmanager(innen)/Programmverwalter(innen) bereitgestellt werden,
2. in der Politik, wobei Daten beispielsweise für Entscheidungen zu politischen Strategien herangezogen werden und somit einen Beitrag zur Kontrolle nach in-

nen bzw. – bei Veröffentlichung der Ergebnisse – zur Legitimation leisten können sowie

3. als Element der gesellschaftlichen Aufklärung, die es erlaubt, Politik zu bewerten.

Am weitesten verbreitet sei die „Managementfunktion“ von Evaluation (Punkt 1). Die Legitimations- und Rechtfertigungsfunktion wachse ebenfalls zusehends (Punkt 2). Die Aufklärungsfunktion sei – sofern sie eingenommen wird – eng mit der Legitimationsfunktion verknüpft, da Evaluation nur in den Bereichen erfolge, in denen Aufträge dafür erteilt werden. Dieses in Deutschland vorherrschende Vorgehen liefere keine Garantie, dass wichtige Felder des gesellschaftlichen Lebens evaluiert werden. Stockmann sprach sich in diesem Kontext für von der Politik unabhängige Institutionen aus, die solche Evaluationen übernehmen können und lieferte damit wiederum Denkanstöße für neue Formate der Zusammenarbeit zwischen Evaluation und Politik. Abschließend stellte er Faktoren heraus, welche die Nutzung von bereitgestellten Ergebnissen erhöhen können:

1. Timing einer Evaluation,
2. Relevanz,
3. Rechtzeitigkeit,
4. Prägnanz (kurz und knackig),
5. Transfer über Beratende (Ergebnisse stärker in Beraterstäbe der Politik bringen),
6. Glaubwürdigkeit,
7. Advocacy/Promotion,
8. angepasste Lösungen,
9. Gesetze und Verordnungen,
10. Partizipation.

Thematische Arbeitsgruppen

Die Vorträge der beiden Veranstaltungstage lieferten Impulse für eine vertiefende Diskussion im Rahmen von Arbeitsgruppen, die von den Referierenden geleitet wurden. Es wurden drei thematische Arbeitsgruppen zu folgenden Aspekten durchgeführt:

1. Steuerung und Evaluation – Möglichkeiten und Grenzen
2. Rationalitätsgewinne durch Evaluationen
3. Herausforderungen bei Wirkungsanalysen

Dieses interaktive Element der drei Arbeitsgruppen wurde von den Teilnehmenden sehr positiv aufgenommen und als gute Gelegenheit gesehen, mit den Referierenden nochmals beitragsübergreifend zu interessanten Fragestellungen ins Gespräch zu kommen. Auch aus Rückmeldungen von Referierenden ging hervor, dass dieses Vorgehen als gewinnbringendes Format für die Nachbereitung der eigenen Vorträge und zum Kennenlernen anderer Perspektiven empfunden wurde. Entsprechend wird das Sprecherteam des AK Soziale Dienstleistungen in der DeGEval ähnliche Tagungselemente auch bei zukünftigen Planungen berücksichtigen.

Fazit und abschließende Bemerkungen

Der AK Soziale Dienstleistungen versteht sich als Austauschforum für Personen, die im Kontext Sozialpolitik und Soziale Arbeit mit dem Thema Evaluation befasst sind. Ein wichtiges Thema des AK Soziale Dienstleistungen ist von jeher, wie die Nutzung von Evaluationsergebnissen durch die Adressatinnen und Adressaten, auch aus Politik und Verwaltung, sowie der Wissenstransfer zwischen Wissenschaft und Praxis in beiden Richtungen gefördert und intensiviert werden kann. Deshalb ist das Sprecherteam der Kooperation mit dem Forschungsinstitut für gesellschaftliche Weiterentwicklung sofort offen und gerne begegnet, da sich dadurch die Möglichkeit eröffnete, gemeinsam hochrangige Akteure aus den Bereichen Politik, Verwaltung, Evaluation und Fachpraxis in einen konstruktiven Austausch zu bringen. Unsere Erwartung, das Verständnis für die Möglichkeiten und Grenzen der Evaluation in politischen Gestaltungs- und Entscheidungsprozessen zu vertiefen und Wege zu einer effektiveren Zusammenarbeit auszuloten, wurde dabei erfüllt, wenngleich noch viele Fragen offenblieben. Wir sehen daher den Bedarf, auch in Zukunft die unterschiedlichen Perspektiven aus Politik, Verwaltung, Wissenschaft und Praxis zusammenzuführen und die begonnenen Diskussionen fortzusetzen.

Danksagung

Herzlich danken möchten wir insbesondere Frau Ramona Liedtke und Frau Olivia Mackowiak (FGW) für die organisatorische Unterstützung der Tagung und der Handwerkskammer Düsseldorf, die uns die Tagungsräume zur Verfügung gestellt hat. Ein besonderer Dank geht natürlich an die Referierenden für ihre kenntnisreichen und spannenden Beiträge.

„Zukunft gestalten – Methoden der strategischen Vorausschau und Praxisbeispiele“

Veranstaltungsbericht vom Frühjahrstreffen des Arbeitskreises Evaluation in der Wirtschaft am 12. Mai 2016

Sonja Kind,¹ Roman Noetzel,² Oliver Rohde²

Der Arbeitskreis Evaluation in der Wirtschaft (AK Wirtschaft) führte am 17. März 2016 sein Frühjahrstreffen in den Räumlichkeiten des DLR Projektträgers in Bonn durch. Im Mittelpunkt des Workshops stand das Thema „Zukunft gestalten – Methoden der strategischen Vorausschau und Praxisbeispiele“.

Teilgenommen haben 40 Multiplikatoren aus Verbänden, Forschungseinrichtungen, Beratungsunternehmen, Clustern und Stiftungen, also Organisationen, die selbst wirtschaftlich arbeiten oder mit Wirtschaftsunternehmen interagieren.

Ziel der Veranstaltung bestand darin, den Nutzen einer strategischen Vorausschau als Teil von Ex-ante-Evaluationsprozessen in Unternehmen, Clustern und Netzwerken darzustellen. Hierzu wurde zunächst in das Thema Vorausschau eingeleitet und im Anschluss anhand von Praxisbeispielen veranschaulicht, für welche Fragestellungen und in welchen Settings Vorausschaumethoden in der unternehmerischen Praxis zum Einsatz kommen. Abschließend wurde zusammen mit den Workshop-Teilnehmer(inne)n eine kreative Methode der Innovationsvorausschau ‚live‘ durchgeführt.

Strategische Vorausschau und Ex-ante-Evaluation in Unternehmen

Im Einführungsvortrag erläuterte Frau Dr. Sonja Kind vom iit – Institut für Innovation und Technik die Aktivitäten des AK Wirtschaft und beantwortete typische Fragen, was Evaluation in der Wirtschaft ist, für welche Fragestellungen sich eine Evaluation in Unternehmen eignet und welcher Nutzen von einer Evaluation in unternehmerischen Kontexten zu erwarten ist. Der Arbeitskreis hat hierzu in der Vergangenheit praxisnahe Broschüren erarbeitet, wie den „Leitfaden Evaluation in der Wirtschaft“ oder die Broschüre „Praxisbeispiele: Wie Evaluation gewinnbringend in

1 Institut für Innovation und Technik (iit), Berlin

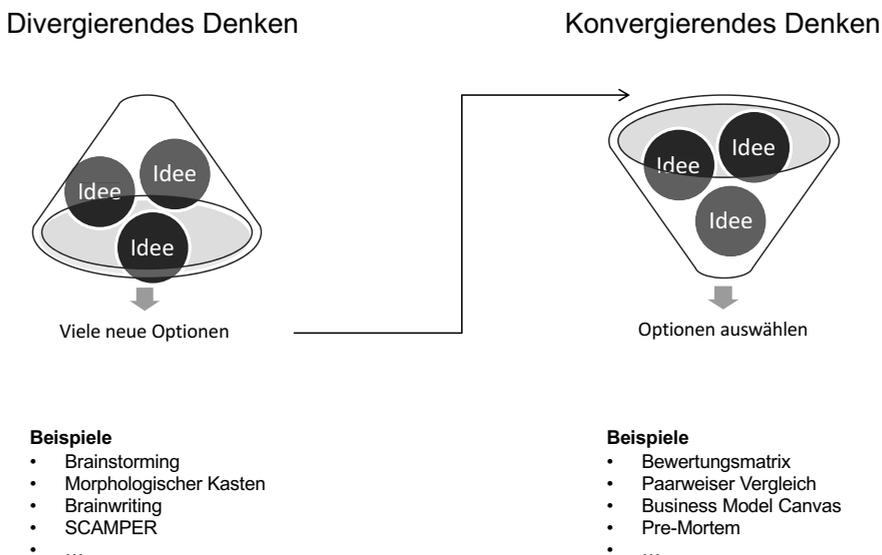
2 DLR Projektträger, Bonn

der unternehmerischen Praxis angewandt wird.“ Diese Broschüren stehen über die DeGEval-Webseite zum Download zur Verfügung.

Der AK Wirtschaft hatte auf der DeGEval-Konferenz in Speyer 2015 beschlossen, das Thema Ex-ante-Evaluation in Bezug auf unternehmerische Fragestellungen aufzugreifen. Dies umfasst die Vorausschau von möglichen Wirkungen, genauso wie die Zielfindung und die Ausarbeitung von Lösungsmöglichkeiten. In der Regel werden hierzu verschiedenen kreative Denkwerkzeuge eingesetzt. Im zweiten Teil des Vortrags stand deshalb die Einführung in kreative Denkkarten und Denkprozesse im Mittelpunkt.

Systematisches kreatives Denken besteht primär aus zwei Denkphasen, die getrennt voneinander erfolgen sollten. In der Phase des divergierenden Denkens geht es um eine breite Suche nach vielen Möglichkeiten und Alternativen, wie z.B. Ideen, Informationen, Problemformulierungen oder Handlungsschritte. In dieser Phase sollte noch nicht gewertet werden. Typische Methoden für ein divergierendes Denken sind das Brainstorming oder der Morphologische Kasten. Dem divergierenden Denken gegenüber steht das konvergierende Denken. In dieser Denkphase werden die zuvor erdachten Alternativen positiv evaluiert und bewertet. Als methodische Vorgehensweise hierfür eignen sich zum Beispiel einfache Bewertungsmatrices, um Optionen nach verschiedenen Kriterien zu bewerten, ein paarweiser direkter Vergleich von jeweils zwei Optionen oder die „Pre-Mortem“-Methode, bei der danach gefragt wird, was alles schiefgehen kann.

Abbildung 1: Methoden der Kreativität



Das besondere an Kreativprozessen besteht darin, dass das Neue per se anders ist, als das, was es früher schon gegeben hat, d.h. es lässt sich vorher nicht wissen, wie das Neue aussieht. Im Nachhinein erscheint das Ergebnis eines Kreativprozesses jedoch meistens logisch, einfach und offensichtlich, während es im Vorhinein alles andere als klar ist. Diese Art der Voreingenommenheit gegenüber den Ergebnissen von Kreativprozessen nennt sich auch „Hindsight Bias“ und kristallisiert beispielsweise in Aussagen wie „Da hätte ja jeder draufkommen können!“.

Auch wenn Kreativprozesse immer sehr individuell ablaufen, können diese grob in vier Phasen strukturiert werden: In der ersten Phase wird die Situation geklärt, in der zweiten werden Ideen entwickelt, in der dritten Phase Lösungen entworfen und in der vierten Phase erfolgt die Umsetzung. Der Prozess der Kreativität verläuft in der Praxis jedoch nicht linear, vielmehr werden zahlreiche Schleifen gedreht, Ideen wieder verworfen und viel Zeit eingesetzt. Zwei sehr verbreitete Modelle für Kreativprozesse sind das „Creative Problem Solving Model“ und „Design Thinking“.

Das Creative Problem Solving Model hat seine Anfänge in den 1950er Jahren in den USA. Ausgang war das Brainstorming, mit dem Ideen entwickelt werden. Das Modell beschreibt den gesamten Prozess der Problemlösung – d.h. auch Schritte vor und nach der Ideenentwicklung. Jeder Schritt umfasst dabei divergierendes und konvergierendes Denken.

Das Design-Thinking-Modell kommt aus dem Designbereich. Es ist von den Abläufen ähnlich wie das Creative Problem Solving Model und beginnt ebenfalls mit dem Verstehen einer Situation und dem Sammeln von Daten und Fakten. Beim Design Thinking stehen jedoch die Nutzer einer angestrebten Lösung und deren Bedürfnisse im Vordergrund. Es handelt sich um eine stark qualitativ und partizipativ orientierte Vorgehensweise, in deren Verlauf Prototypen erstellt und getestet werden. Das Design-Thinking-Modell ist aufgrund der starken Ausrichtung auf das Testen und Erproben von Prototypen weniger breit anwendbar als das Creative Problem Solving Model.

Anwendung von Foresight-Methoden in der Praxis

Die nachfolgend zusammengefassten Praxisbeispiele zeigen, dass Methoden der Vorausschau unabhängig von Branchen oder Sektoren eingesetzt werden können. Die Methoden erwiesen sich dabei nicht nur relevant für die Strategieentwicklung einzelner Unternehmen, sondern auch für öffentliche Institutionen, die auf eine strategische Entwicklung von Wirtschaftsregionen oder -sektoren zielen. Auch Forschungseinrichtungen greifen auf Methoden der Vorausschau zurück, um ihre Aktivitäten im Transfer von Wissen und Technologie zu unterstützen.

Wie KMU und Cluster aus Nordrhein-Westfalen Unterstützung bei der Erstellung von Technologie-Roadmaps erhalten und wie sie diese nutzen präsentierte Dr. Herbert Rath von der Zenit GmbH. Roadmaps sollen die Entscheidungsfindung im Unternehmen unterstützen, Hindernisse benennen und den effizientesten Weg zum Erreichen der Unternehmensziele identifizieren. Roadmaps sind ein Instrument, um diese Wechselbeziehungen zwischen Technologien, Produkten und Märkten in

Verbindung mit einer Zeitachse grafisch abzubilden. Interne und externe Einflussfaktoren können so bestimmt werden. Damit ermöglichen sie das frühzeitige Erkennen von Trends, die bei der Entwicklung neuer Technologien berücksichtigt werden können. Alternative Lösungswege können aufgezeigt werden, die ebenfalls der Nachfrage der Kunden entsprechen. Die zielgerichtete Suche nach verfügbaren technologischen Lösungen hilft damit auch, Kosten und Zeit zu sparen. Typische Entscheidungen, die durch Roadmaps unterstützt werden, sind die Suche nach vorhandenen Technologien, die Strukturierung eigener FuE-Aktivitäten, der Eintritt in neue Märkte sowie die Absicherung bestehender Marktpositionen. So können auch neue Anwendungsmöglichkeiten für bestehende Technologien identifiziert werden. Neben einzelnen Unternehmen wird die Methode von Zenit auch bei Clustern genutzt, die gemeinsame Innovationsaktivitäten befördern wollen. Eine Herausforderung bei der Erstellung von Roadmaps ist die Identifizierung der Anforderungen am Markt und die Definition der Unternehmensziele. Die Erarbeitung der Roadmaps in einer Serie von Workshops mit Vertreterinnen und Vertretern aller beteiligten Unternehmensbereiche ermöglicht die Nutzung einer breiten Wissensbasis. Vorteilhaft ist auch, dass die im Roadmapping-Prozess entwickelten Maßnahmen in der Regel schnell im Tagesgeschäft umsetzbar sind.

Wie Szenarioprozesse genutzt werden können, um das Marktumfeld von Unternehmen abzubilden, präsentierte Holger Bornemann von der Prognos AG. Szenarien zum Marktumfeld unterstützen Unternehmen dabei, Märkte, Branchen, Wettbewerber und die globale Situation besser einzuschätzen. Die Umfeldfaktoren können dabei in branchenspezifische und allgemein bedeutsame Faktoren unterteilt werden. Die Analyse erfolgt dabei in vier Schritten: Ausgehend von der Festlegung der Rahmenbedingungen und der Trendanalyse werden in Workshops die Einflussfaktoren bestimmt und Szenarien entwickelt. Beispielhaft wurde die Analyse des Marktumfelds einer regional orientierten Bank dargestellt. Als Einflussfaktoren wurden dabei die demographische Entwicklung in der Region, neue Technologien, neue Wettbewerber, die Entwicklung der Nachfrage sowie neue Märkte identifiziert. Aus diesen Einflussfaktoren werden Trends entwickelt und im Hinblick auf Relevanz und Eintrittswahrscheinlichkeit gefiltert. Als weitere Dimension kommt die Trendqualität hinzu: Dabei wird betrachtet, ob es sich um disruptive Entwicklungen handelt oder ob der Trend spezifisch für den betrachteten Sektor ist. Daraus werden unter Nutzung unterschiedlicher Bewertungsskalen die jeweiligen Wirkungen auf den Auftraggeber abgeleitet. In Workshops werden unter Einbezug des Auftraggebers und weiterer Expertinnen und Experten Projektionen zur Entwicklung der unterschiedlichen Trends und Einflussfaktoren entwickelt. Üblicherweise werden bei der Marktumfeldanalyse eine Reihe von Methoden angewandt: Für die Ex-ante-Bewertung werden Projektionen für einen Zeitraum von fünf bis zehn Jahren entwickelt, die auf quantitativen und qualitativen Daten, Beratungen und Workshops basieren. Diese Vorausschau ist ein kontinuierlicher Prozess, bei dem die Zielsetzungen der Roadmap jährlich evaluiert werden. Idealerweise werden die Szenarien zur Entwicklung des Marktumfelds alle fünf Jahre angepasst.

Frau Dr. Sonja Kind und Herr Dr. Marc Bovenschulte vom iit stellten in ihrem Vortrag die Methode Visual Roadmap vor, die für die Strategieentwicklung des For-

schungscampus Hannover-Garbsen eingesetzt worden ist. Ziel einer Visual Roadmap ist es, (technologische) Entwicklungen vorauszudenken, und zwar in ihrem sozialen, rechtlichen und ökonomischen Umfeld. Dabei werden wichtige Meilensteine auf dem Weg vom ‚Jetzt‘ in die Zukunft identifiziert. In einem moderierten Dialog werden die Erkenntnisse von Expertinnen und Experten visualisiert und versucht, ein Konsens in Bezug auf die dargestellten Prozesse und Faktoren zu finden. Die gemeinsam erarbeitete Visual Roadmap dient als Ausgangspunkt für weiterführende Diskussionen und Handlungsempfehlungen. Besonders eignet sich die Visual Roadmap für die Strukturierung komplexer Themen und Prozesse. Das Zusammenspiel verschiedener Zusammenhänge und Abhängigkeiten kann visualisiert werden und durch die Einbindung verschiedener interdisziplinärer Expertengruppen kann ein umfassendes Bild erzielt werden. Herausforderungen bestehen in einer möglicherweise zu starken Vereinfachungen sowie einer Akzeptanzproblematik (Wissen wird aus den Köpfen geholt...) und dem durchaus hohen Zeitaufwand. Die Durchführung dauert circa zwei Stunden. Der Roadmapprozess umfasst aber zusätzlich entsprechende vor- und nachbereitende Arbeiten. Im vorgetragenen Beispiel ging es um die Entwicklung des Forschungscampus Hannover-Garbsen. Bis zum Jahr 2018 sollen die 17 produktionstechnischen Institute der Universität Hannover an einem neuen Standort, dem Forschungscampus Hannover-Garbsen, zusammengefasst werden. Mithilfe der Visual Roadmap wurden die wissenschaftlich-technischen und sozio-ökonomischen Entwicklungen mit Blick auf die globalen versus regionalen Trends im Bereich Maschinenbau diskutiert. Als Ergebnis der Visual Roadmap zeichneten sich einige Schwerpunkte für den Standort ab, wie zum Beispiel die bleibende bzw. zunehmende Bedeutung von Maschinenbau und Produktionstechnik für den Standort, das Thema Energie(-effizienz), Flexibilität und Wandlungsfähigkeit von Maschinen und Anlagen, sowie das Potenzial in der transdisziplinären Zusammenarbeit im Bereich (Bio-)Medizintechnik. Als Maßnahmen zur Realisierung der Potenziale in der Region Hannover wurde ein Masterplan entwickelt, der als Elemente u.a. die Etablierung eines Steuergremiums mit verschiedenen Akteuren beinhaltete, um die Erarbeitung eines Gesamtkonzepts zu koordinieren. Zudem wurde vorgeschlagen, eine zentrale Managementstruktur am Campus zu etablieren, einen Fokus auf Infrastrukturen zu legen und ein Marketingkonzept zu erarbeiten.

In der Einführung zu den hiernach folgenden Beiträgen der Forschungseinrichtungen gab Herr Roman Noetzel vom DLR Projektträger zunächst einen allgemeinen Überblick zu den Methoden, die sowohl im DLR als auch im DLR Projektträger Anwendung finden. Beispielhaft skizzierte er das Zusammenwirken unterschiedlicher methodischer Ansätze bei der Erarbeitung einer Regionalstrategie durch den DLR Projektträger, die u.a. Methoden wie die SWOT-Analyse und eine Logical-Framework-Analyse kombiniert, mit dem Ziel Zukunftsthemen der internationalen Kooperation herauszuarbeiten. Die dargestellten Praxisbeispiele aus dem DLR und der Fraunhofer-Gesellschaft zeigten, wie Foresight-Methoden von Forschungseinrichtungen in Strategieprozessen angewandt werden, um – in Kooperation mit Unternehmen – marktnahe ‚Zukunftsprodukte‘ zu identifizieren.

Dr. Dietmar Heyland vom Technologiemarketing des DLR erläuterte, wie mithilfe der Analyse gesellschaftlicher Megatrends die Forschungsangebote des DLR

und die gesellschaftliche Nachfrage nach Lösungen drängender Probleme in Einklang gebracht werden können. Ziele sind dabei der rasche Transfer von Forschungsergebnissen und Technologien in die Wirtschaft, das Schließen strategischer Partnerschaften von Wissenschaft, Wirtschaft und Politik sowie generell das Generieren eines sichtbaren Nutzens von Forschung für Industrie und Gesellschaft. Das Technologiemarketing agiert dabei an der Schnittstelle von Forschung und Markt und begleitet alle Schritte des Innovationsprozesses bis zur Verwertung mit Partnern aus der Industrie oder alternativ auch über eigene Ausgründungen. Alle Mitarbeiterinnen und Mitarbeiter können eigene Projektideen in einem mehrstufigen Prozess in einem Intranetportal einstellen und durch das Technologiemarketing begleiten und bewerten lassen. Die Megatrendanalyse flankiert diesen Bottom-up-Ansatz durch die Analyse zukünftiger gesellschaftlicher Bedarfe und eine Zuordnung dieser Bedarfe zu den Forschungskompetenzen des DLR. Die DLR-Megatrendanalysen werden alle zwei Jahre durchgeführt und bestehen aus vier Elementen: Auf die Identifikation der Megatrends in unterschiedlichen Bereichen folgt die Zuordnung und Bewertung der Bedürfnisse bis hin zur Identifikation von Zukunftsprodukten mithilfe von Mind Mapping. Im nächsten Schritt werden die Geschäftsfelder und Forschungsschwerpunkte des DLR hinsichtlich des Angebots an Know-how und Technologien zur Deckung des zukünftigen Bedarfs von Wirtschaft und Gesellschaft analysiert. Daraus werden Ideen für neue Produkte, Verfahren und Dienstleistungen entwickelt.

Dr. Antonino Ardilio vom Fraunhofer Institut für Arbeit und Organisation (IAO) berichtete über die Chancen der Nutzung von Big Data beim Technologiemanagement. Semantische Anwendungen werden hierzu sowohl genutzt, um chancenträchtige Technologien weltweit identifizieren zu können als auch, um potenzielle Anwendungen zu erkennen. Ziel ist, an Stelle zufälliger Prozesse des Transfers von Technologie in neue Anwendungen, einen systematischen Ansatz zu verfolgen. Die Technologien und Anwendungsfelder werden dabei in einzelne Attribute und Funktionen zerlegt, um unabhängig von einzelnen Märkten oder Produkten neue Möglichkeiten zu erkennen. Dabei können neue Technologien alte ersetzen, ergänzen oder gemeinsam mit diesen neue Anwendungsmöglichkeiten erschließen. Die vom IAO genutzten Werkzeuge „Technology Radar“ und „Market Explorer“ nutzen nicht nur offen zugängliche Quellen im Internet, sondern greifen auf eine Reihe von Publikations- und Patentdatenbanken zurück. So können beispielsweise bei der Marktanalyse zuerst alle potenziellen Märkte identifiziert und in weiteren Schritten die attraktivsten Felder herausgefiltert werden. Suchprozesse werden verkürzt und günstiger. Die Dienstleistungen des Technologiemanagements mithilfe von Big Data stehen dabei Instituten der Fraunhofer-Gesellschaft ebenso wie externen Kunden zur Verfügung.

Zum Abschluss führte Johannes Gabriel von Foresight Intelligence einen kurzen experimentellen Szenarioprozess zum Arbeitsmarkt und der Rolle digitaler Infrastruktur im Jahr 2030 durch. Die teilnehmenden Expertinnen und Experten diskutierten in vier Schritten, welche Faktoren die Entwicklung des Arbeitsmarktes bis 2030 am stärksten beeinflussen und sich am wenigsten abschätzen lassen, welche zwei plausiblen Extremausprägungen des Unsicherheitsfaktors für das Jahr 2030 zu

erwarten sind und wie aus der Kombination dieser Faktoren Szenarien für die Zukunft gebildet werden können. Anschließend wurden die Konsequenzen und Eintrittswahrscheinlichkeiten der Szenarien sowie deren jeweilige Chancen und Risiken im Plenum diskutiert.

Fazit und Ausblick

Die Praxisbeispiele und die Diskussion stellten die große Bedeutung der strategischen Vorausschau für Unternehmen, Cluster und Netzwerke dar. Eine strategische Vorausschau ist ein wichtiges Element ihres Handelns, um dauerhaft die eigene Wettbewerbsfähigkeit zu erhalten und/oder auszubauen. Die Gemeinsamkeiten der vorgestellten Ansätze liegen u.a. in der Anwendung eines Methoden-Mix bzw. der Triangulation, dem partizipativen Ansatz sowie der inter- und transdisziplinären Perspektive.

Wie sich auf dem Workshop gezeigt hat, entsteht genau hier eine Nähe zwischen strategischer Vorausschau und Evaluation: Ein Technologie- und Innovations-scouting, expertenbasierte Trendanalysen, die Erstellung von Zukunftsvisionen und -szenarien oder ganz allgemein Innovations- und Strategieprozesse basieren häufig auf der Nutzung von kreativen Methoden der strategischen Vorausschau, die nach Einschätzung des Sprecherinnen- und Sprecherteams im weiteren Sinne auch als Ex-ante-Evaluationsmethoden bezeichnet werden können.

Der AK Evaluation in der Wirtschaft muss sich der Herausforderung stellen, dass die gesamte Thematik Evaluation in Unternehmen nach wie vor recht unbekannt ist, wenngleich typische Evaluationsmethoden wie Befragungen, Interviews oder Audits im Rahmen des Qualitätsmanagements schon lange selbstverständlich von diesen eingesetzt werden.

Für den Arbeitskreis Wirtschaft ist der Workshop ein besonderer Erfolg. Seit seiner Gründung vor über zehn Jahren hat der Arbeitskreis mit rund 40 Teilnehmenden einen so großen Zuspruch für sein Frühjahrstreffen erzielt wie noch nie zuvor. An der Veranstaltung nahmen zahlreiche Beratungsunternehmen und Intermediäre teil, die Dienstleistungen für Unternehmen und Wirtschaftsregionen anbieten. Diese können als Multiplikatoren für das Konzept von Evaluation in Unternehmen dienen.

Das Sprecherteam sieht sich darin bestärkt, das Thema Ex-ante-Evaluation verbunden mit einer Nutzung von kreativen Methoden zur strategischen Vorausschau für unternehmerische Fragestellungen weiterzuverfolgen. Das Ziel von Unternehmen ist es u.a. ihre Produkte und Leistungen zukunftsorientiert auszurichten. Es ist die Aufgabe von Unternehmen, stetig innovative Produkte, moderne Geschäftsmodelle und neue Märkte zu erschließen. Auch sind Unternehmen gefordert, sich mit Fragen des zukünftigen Wandels zu beschäftigen, wie mit den Folgen der Digitalisierung oder des demographischen Wandels. Zur Bewältigung dieser Fragen können die Ex-ante-Evaluations-orientierten Methoden der strategischen Vorausschau helfen.

Der Arbeitskreis Evaluation in der Wirtschaft möchte hier ansetzen und über diesen Zugang Unternehmen und deren Multiplikatoren für das Thema Evaluation

in der Wirtschaft im Allgemeinen sensibilisieren. Hierzu wird sich auch die Jahreskonferenz 2017 eignen, die unter dem Motto „Evaluation (in) der Zukunft“ ganz im Zeichen der Zukunft stehen wird.

Zur Qualität qualitativer Sozialforschung in der Evaluation – Möglichkeiten und Grenzen

Frühjahrstagung des AK „Methoden in der Evaluation“
der DeGEval in Bonn

Lucas Jacobs,¹ Jessica Schendzielorz,¹ Dorothea Giesen-Thole¹

Der Arbeitskreis (AK) Methoden in der Evaluation der DeGEval – Gesellschaft für Evaluation lud seine Mitglieder und weitere Interessierte am 20. und 21. Mai zur Frühjahrstagung 2016 nach Bonn ein, um mit ihnen die Möglichkeiten und Grenzen qualitativer Sozialforschung in der Evaluation zu erörtern. Gastgeberin der Veranstaltung war die Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ). Mehr als 100 Teilnehmerinnen und Teilnehmer folgten der diesjährigen Einladung, rund die Hälfte davon DeGEval-Mitglieder.

Das von der AK-Sprecherin Prof. Dr. Alexandra Caspari, Professorin für Evaluationsforschung, Methoden der empirischen Sozialforschung und Statistik der Frankfurt University of Applied Sciences, und dem AK-Sprecher Dr. Jan Tobias Polak, Evaluator am Deutschen Evaluierungsinstitut der Entwicklungszusammenarbeit (DEval), sorgfältig zusammengestellte Programm wartete mit hochkarätigen Referentinnen und Referenten aus Wissenschaft und Praxis auf, die sich in ihren Vorträgen sowohl den theoretischen Ansprüchen qualitativer Methoden der Sozialforschung als auch deren praktischer Umsetzung in der Evaluation widmeten.

In Ihrer Begrüßung zu Beginn der Tagung richtete Dr. Annette Backhaus, Gruppenleiterin in der Stabsstelle Evaluierung der GIZ, den Fokus auf das Spannungsfeld zwischen praktischem Nutzen von qualitativ hochwertigen Evaluationen und dem dafür erforderlichen Aufwand. Entsprechend sei es notwendig nach dem möglichen abnehmenden Grenznutzen bei der Erfüllung von Qualitätssteigerungen zu fragen. Was ist „good enough“? Es gebe zwar Qualitätsstandards aber keine standardisierten „Messlatten“. Kann man sich perspektivisch auf praxisorientierte Mindeststandards verständigen? Und warum werden überhaupt nur so wenige Metaevaluerungen gemacht, und die wenigen adressieren fast ausschließlich die methodische Qualität, aber nicht die Erfüllung der Qualitätsstandards der Prozesse und des Nutzens der Evaluierungen. Auch hier gäbe es weiteren Diskussionsbedarf.

¹ Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) GmbH, Bonn

Mindestanforderungen an qualitative Sozialforschung – das Spannungsverhältnis von Theorie und Praxis

Zunächst stellte Prof. Dr. Philipp Mayring, Professor für Psychologische Methodenlehre der Alpen-Adria Universität und Vorsitzender der DeGEval, im ersten Vortrag des zweitägigen Programms die Notwendigkeit zur Aufstellung methodischer Mindeststandards, Gütekriterien und Ablaufmodelle dar. Er führte aus, dass die Qualitative Inhaltsanalyse eine Mischform qualitativer und quantitativer Analyse darstelle und stellte die Überwindung des zum Teil noch immer herrschenden Gegensatzes zwischen qualitativen und quantitativen Methodenstandards in das Zentrum seines Vortrags.

In der sich anschließenden Diskussion wurde lebhaft erörtert, ob die zum Teil dogmatische Trennung von qualitativen und quantitativen Methoden nicht auch eine Generationenfrage sei und insbesondere bei jüngeren Evaluatorinnen und Evaluatoren diese strikte Unterscheidung im Sinne eines immer häufiger verwendeten Mixed-Methods-Ansatzes immer weniger eine Rolle spiele. Angesprochen auf die Operationalisierung von Mindeststandards in der Durchführung von Interviews, hob Prof. Dr. Philipp Mayring hervor, dass die Transkription ein selbstverständliches Gütekriterium sei. Wenn notwendig, sei sie auch in reduzierter Form möglich. Als unerlässlich stellte er jedoch das Mitschneiden von Interviews in Audioform heraus. Angesprochen auf die Frage, wie in fragilen Kontexten ein den methodischen Qualitätskriterien entsprechendes Sampling möglich sei, hob er die Vorbereitung in Form eines angemessenen Stichprobenplans hervor.

Potenziale qualitativer Sozialforschung in Evaluationen

Als zweiter Referent des Tages hielt Prof. Dr. Dieter Filsinger, Professor für Sozialwissenschaftliche Grundlagen, Sozialpolitik und Evaluation der Hochschule für Technik und Wirtschaft des Saarlandes und (Co-)Studiengangsleiter des Masterstudiengangs Evaluation an der Universität des Saarlandes, den gemeinsam mit Dr. Marianne Lück-Filsinger, Leiterin der Forschungsgruppe Bildungs-, Evaluations- und Sozialstudien (ForBES) der Hochschule für Technik und Wirtschaft des Saarlandes vorbereiteten Vortrag über die „Potenziale qualitativer Sozialforschung in Evaluationen“. Die Auswahl quantitativer und qualitativer Methoden hänge primär von der Fragestellung und dem zu evaluierenden Gegenstand ab. Hieraus folgert Prof. Dr. Dieter Filsinger, dass bestimmte Fragestellungen bestmöglich mit qualitativen Methoden zu beantworten seien. Insbesondere das Fallstudienkonzept erlaube es, evaluierende Handlungseinheiten (Fälle), wie etwa Programme, als komplexe, regelgeleitete Interaktionszusammenhänge in spezifischen Kontexten zu analysieren. Über die Rekonstruktion von Prozessstrukturen ließen sich zudem Erkenntnisse über Wirkungszusammenhänge gewinnen. Weiter stellte er die Reichweite von Aussagen und Ergebnissen in den Fokus seiner Ausführungen und unterstrich in diesem Zusammenhang die Bedeutung einer geeigneten Auswahl des zu untersuchenden Samples.

Im Anschluss an den Vortrag entstand eine angeregte Diskussion, wie auf praxistaugliche Art und Weise mit bekannten Herausforderungen im Sampling umgegangen werden könne. Aus dem Publikum wurde auf das Problem von beabsichtigten und unbeabsichtigten Verzerrungen in den Aussagen von Interviewten aufgrund von persönlichen Interessen hingewiesen. Dazu führte Prof. Dr. Dieter Filsinger aus, diesem Sachverhalt könne durch eine transparente Rekonstruktion und Darlegung der Interessen und zugrundeliegenden Strukturen angemessen begegnet werden. Angesprochen auf die Vorteile und Funktionen von qualitativen Evaluierungsmethoden bekräftigte er die Notwendigkeit der Identifikation von Wirkungsmechanismen, entscheidend sei hier das ‚Wie‘.

Qualitative Methoden im Mixed-Methods-Ansatz – Möglichkeiten und Grenzen qualitativer Methoden

Dezidiert aufgegriffen wurden die – bereits von Prof. Dr. Philipp Mayring und Prof. Dr. Dieter Filsinger angeführten – Mischformen von qualitativen und quantitativen Ansätzen von Prof. Dr. Udo Kelle, Professor für Sozialforschung und Statistik an der Helmut-Schmidt-Universität/Universität der Bundeswehr in Hamburg, mit seinem Vortrag „Qualitative Methoden im Mixed-Methods-Ansatz – Möglichkeiten und Grenzen qualitativer Methoden“. Prof. Dr. Udo Kelle erkennt den Methodenmix als wechselseitigen Ausgleich von Limitationen der rein qualitativen oder quantitativen Ansätze. In seiner Präsentation zeichnete er die grundlegenden Problemstellungen des dichotomen Verständnisses von rein qualitativen und quantitativen Methoden nach und referierte über mögliche Ansatzpunkte des Mixed-Methods-Ansatzes, um die jeweiligen Schwächen der gewählten Methoden gezielt auszugleichen. Es folgte ein Plädoyer für eine systematische Aufarbeitung und Beschreibung von Methodenproblemen und -stärken quantitativer und qualitativer Forschung, die durch Methodenkombination entdeckt und bearbeitet werden könnten. Neben der systematischen Herausarbeitung von bekannten Stärken und Schwächen der jeweiligen Methoden ergänzte Prof. Dr. Udo Kelle, dass eine wertschätzende Diskussionskultur zwischen den Disziplinen von grundlegender Bedeutung sei.

In der folgenden Diskussion wurde deutlich, dass seine Ausführung zur Überbrückung der Gräben zwischen den beiden klassischen Schulen auf breiten Konsens unter den Zuhörer(inne)n stieß. Es wurde zudem angemerkt, dass der Mixed-Methods-Ansatz beispielweise in der kommerziellen Marktforschung bereits gänzlich etabliert sei. Als weitere wichtige Erfolgsfaktoren wurden in der Diskussion vor allem das kluge Kombinieren und das entsprechende Sequenzieren von geeigneten Methoden herausgearbeitet. Den Mixed-Methods-Ansatz zeichnet demnach mehr als das Mixen von Methoden aus.

Minimumstandards und Voraussetzungen für gute qualitative Methoden

Im Anschluss an die drei Vorträge wurde in Arbeitsgruppen anhand von Leitfragen intensiv über Mindeststandards und erforderliche Voraussetzungen für die Umsetzung guter qualitativer Methoden weiter diskutiert. Der Austausch erfolgte in Gruppen: Auftraggeber(innen), Auftragnehmer(innen) und Wissenschaftler(innen).

Nach dem Austausch in den Kleingruppen kam die gesamte Teilnehmerschaft zur Podiumsdiskussion zusammen. Neben Vertreterinnen und Vertretern der Arbeitsgruppen war das Podium besetzt mit den Referentinnen und Referenten des Vormittags sowie den Vortragenden des Folgetages: Barbara Befani, PhD, Research Fellow, University of Surrey, Centre for the Evaluation of Complexity and Research Associate der University of East Anglia, School of International Development und Dr. Vera Hundt, Fachkonzeptionistin in der Stabsstelle Evaluierung der GIZ. Eröffnet wurde die Podiumsdiskussion, indem die Vertreterinnen und Vertreter der Arbeitsgruppen in einem kurzen Statement den Diskussionsstand aus ihrer Gruppe wiedergaben. Die unterschiedlichen Perspektiven zeichneten hinsichtlich der Frage nach Mindeststandards und Voraussetzungen für gute qualitative Methoden ein rundes Bild: Wichtige Voraussetzungen seien neben Methodenkenntnissen und -erfahrungen sowie Fachwissen jedenfalls gute Leistungsbeschreibungen, ein angemessenes Mengengerüst sowie das Evaluierungsmanagement. Als Mindeststandards wurden unter anderem genannt: ein dem Evaluierungsgegenstand angemessenes Design, Triangulation und transparente Dokumentation. Grundlegend sei, dass zu Beginn eines Projektes eine später erfolgende Wirkungsmessung mitgedacht und ermöglicht werde (Indikatoren, Baseline, Monitoring, etc.).

In der anschließenden Diskussion, unter reger Beteiligung der Teilnehmerschaft, wurde die Bedeutung eines realistischen Erwartungsmanagements herausgestellt. Zum Thema Ausschreibung von Evaluationen wurde diskutiert, dass bei unrealistischen Leistungsbeschreibungen nicht mit einem unrealistischen Angebot reagiert werden sollte. Es sollte klar benannt werden, was unter welchen Bedingungen möglich und machbar ist und was nicht. Bei limitierten Ressourcen sollte fokussiert werden, um eine gute methodische Qualität zu gewährleisten. Hinsichtlich der erforderlichen Kompetenzen von Evaluatorinnen und Evaluatoren wurde auf das DeGEval-Papier „Empfehlungen für die Aus- und Weiterbildung in der Evaluation“⁴² verwiesen. Auch Nutzen und Nutzung von Evaluationen wurden im Spannungsfeld mit methodischer Qualität und erforderlichen Mindeststandards thematisiert.

Der zweite Tagungstag zeichnete sich insbesondere durch seine Praxisnähe aus: Neben einer Vorstellung der Qualitative Comparative Analysis (QCA) durch Barbara Befani, PhD referierte Dr. Vera Hundt anhand einer Meta-Evaluierung über die Erfahrungen mit qualitativer Sozialforschung aus Perspektive der GIZ.

2 Verfügbar unter: http://www.degeval.de/fileadmin/Publikationen/Publikationen_Homepage/DeGEval_-_Empfehlungen_Aus-_und_Weiterbildung.pdf [08.06.2016].

Generalising Case-Based Findings with Qualitative Comparative Analysis

Angereist aus Großbritannien legte Barbara Befani, PhD, mit einem engagierten Vortrag den Fokus auf die Potenziale und Limitationen der Methode Qualitative Comparative Analysis (QCA). Angereichert mit Beispielen aus erfolgten Evaluationen stellte sie das ‚Herzstück‘ der QCA, das Boolean Dataset vor. Es zeichne sich durch die Formalisierung von qualitativen Ergebnissen aus, welche in einer Binärcodierung in eben jenes Dataset eingetragen würden. Das sich ergebende Hybridmodell aus qualitativen und quantitativen Methoden bietet die Möglichkeit, formalisierte Muster deutlich zu machen, ohne dabei an die zum Teil hinderliche Strenge einiger quantitativer Methoden gebunden zu sein. Auf diese Weise können auf unterschiedlichen Ebenen der Abstraktion und Betrachtung verschieden generalisierbare Aussagen gemacht werden.

Die sich anschließende Diskussion zeigte deutlich auf, dass die QCA eine vielversprechende Möglichkeit bietet, im Sinne eines Mixed-Methods-Ansatzes die Vorteile verschiedener Evaluierungsinstrumente miteinander zu kombinieren. Unbestritten wurden jedoch auch die sich logischerweise ergebenden Nachteile einer hybriden Methode deutlich: Einige Stärken der qualitativen Sozialforschung – wie das Abbilden komplexer Sachverhalte – gingen, so wurde angemerkt, durch die binäre Einteilung der Ergebnisse verloren. Umgekehrt wurde deutlich, dass das Herausarbeiten von Mustern nicht der Robustheit im klassischen quantitativen Sinne entspräche. Legt man die vorangegangenen Diskussionen eher theoretischer Natur zugrunde, zeigt sich an der QCA jedoch, wie ein gangbarer Weg im Sinne eines Mixed-Methods-Ansatzes in der Praxis aussehen könnte.

Qualitative Sozialforschung in der Evaluationspraxis – Erfahrungen und Perspektiven der GIZ

Der Abschlussvortrag der Frühjahrstagung kam von Dr. Vera Hundt zum Thema: „Qualitative Sozialforschung in der Evaluierungspraxis – Erfahrungen und Perspektiven der GIZ“. Sie berichtete zum Standard-Evaluierungsinstrument der GIZ, der Projektevaluierung (PEV)³. Neben den methodischen Anforderungen der PEV ging es in dem Vortrag um Ansatzpunkte und Herausforderungen zur Verbesserung der methodischen Qualität auf Basis der Ergebnisse einer aktuell durchgeführten Meta-Evaluierung. Im Rahmen dieser Meta-Evaluierung wurde neben der Bewertung der methodischen Qualität (Genauigkeit) auch die Prozessqualität (Durchführbarkeit,

3 Eine Projektevaluierung ist obligatorisch in allen vom Bundesministerium für wirtschaftliche Zusammenarbeit und Entwicklung (BMZ) beauftragten Projekten einer Laufzeit von drei Jahren und einem Auftragswert von mindestens 3.000.000 €. Näheres zur Projektevaluierung siehe „WIRKUNG MESSEN – ZU WIRKUNG BEITRAGEN“. Verfügbar unter: https://www.giz.de/de/downloads/giz2015-de-evaluierungsbericht_2012-2014.pdf [08.06.2016] und www.giz.de/evaluierung

Fairness) sowie Nützlichkeit und Nutzung der PEV anhand von Dokumentenanalysen und einer Befragung von Auftragsverantwortlichen untersucht.

Während die Prozessqualität der PEV gute Ergebnisse erzielt und auch die Nutzung als hoch bewertet wird, stellt die methodische Qualität die schwächste Qualitätsdimension dar. Die Meta-Evaluierung zeigt dabei auf, dass bei der PEV vorwiegend qualitative Methoden eingesetzt werden, meist Interviews und Dokumentenauswertung. Gleichzeitig zeigen sich Schwächen in der Darstellung der Anwendung dieser Methoden, wodurch die Bewertung der Qualität dieser Methoden erschwert wird. Somit stellt die Darstellung der Qualität der Durchführung von Methoden eine weitere Herausforderung im Rahmen von Evaluierungen dar. Regelmäßig durchgeführte Meta-Evaluierungen der PEV sollen der GIZ als ein kontinuierliches Monitoring der Evaluierungspraxis dienen. Auf Basis der Ergebnisse der aktuellen Meta-Evaluierung setzt sich die GIZ nun mit möglichen ‚Stellschrauben‘ zur Verbesserung der methodischen Qualität der PEV auseinander. Hierbei geht es beispielsweise um die Diskussion erforderlicher Evaluierungskompetenzen, aber auch um die Formulierung von Mindeststandards.

Auch mit Blick auf das Spannungsverhältnis von Aufwand und Qualität stellt sich die Frage: Wann ist eine Evaluierung ‚gut genug‘? Leider gibt es bisher erst wenige Meta-Evaluierungen, die zudem sehr unterschiedliche Analyseraster verwenden. Allgemeingültige Operationalisierungen und Maßstäbe, zu welchem Grad Evaluierungsstandards erfüllt sein müssen, damit eine ‚gute‘ Evaluierung vorliegt, fehlen. Hier besteht weiterer Diskussions- und Orientierungsbedarf in der Evaluierungscommunity, wie sich auch in diversen Diskussionen der Tagung zeigte.

Fazit

Nach der finalen Diskussionsrunde umriss AK-Sprecher Dr. Tobias Polak die Charakteristika der Vorträge. Abschließend, nach Diskussionen über das Potenzial qualitativer Methoden sowie Vorschlägen für Gütekriterien, der Bedeutung der Komplementarität von Methoden, sei festzuhalten, dass es tatsächlich schwer sei, gute qualitative Methoden einzusetzen, diese nicht umsonst zu haben seien und hier noch ein langer Weg vor uns liege. Eine nächste Etappe auf diesem Weg, bei der die Möglichkeit besteht, sich hierzu weiter auszutauschen, ist die DeGEval-Jahrestagung in Salzburg, vom 21. bis 23. September 2016 zum Thema „Zwischen Nutzung, Einfluss und Nachhaltigkeit – Wie wirken Evaluationen in unterschiedlichen Systemen?“ Vielleicht begegnen wir hier auch dem von Prof. Dr. Udo Kelle zitierten „Elefanten im Wohnzimmer“ wieder.

„Evaluation (in) der Zukunft“

20. Jahrestagung der DeGEval – Gesellschaft für Evaluation e.V. vom 20. bis 22. September 2017 an der Johannes Gutenberg-Universität in Mainz

Call for Proposals

Zum Anlass des 20. Jubiläums wird die DeGEval-Jahrestagung **den Blick zurück mit einem Blick in die Zukunft verknüpfen**. Wie hat sich seit Gründung der Gesellschaft das Handlungsfeld Evaluation in Praxis, Theorie und Forschung verändert? Welche Entwicklungen sind in den kommenden Jahren absehbar, zu befürchten oder zu erhoffen? Welche Rolle hat die DeGEval bei diesen Entwicklungen gespielt und welche wird sie in Zukunft einnehmen?

Unter Bezug auf einige pointierte Vorhersagen des gegenwärtigen Vorsitzenden unserer Schwestergesellschaft American Evaluation Society,¹ John Gargani, wollen wir in diesem Zusammenhang unter anderem folgende Fragen diskutieren:

- Wie verändert sich das **Verhältnis von interner zu externer Evaluation**? Werden Organisationen und Verwaltungen die Evaluation zunehmend internalisieren (Stichwort ‚Evaluation Capacity Development‘) oder wird der Bedarf nach Objektivität und Unabhängigkeit die Nachfrage nach externen Evaluationen hochhalten?
- Wie verändern sich die **Rolle von Evaluierenden** und die an sie herangetragenen Rollenerwartungen? Wie stark wird etwa von Evaluierenden verlangt werden, zunehmend gestaltend auf der Gegenstandsebene mitzuwirken, wie es etwa Pattons Ansatz der Developmental Evaluation vorsieht?
- Wie verändern die **sozialen Medien** die Evaluation und den sozialen Kontext, in dem sie agiert? Welche Möglichkeiten ergeben sich etwa für die Kommunikation in Evaluationen z.B. bei partizipativ angelegten Verfahren? Welche neuen Risiken ergeben sich in der digitalen Öffentlichkeit?
- Wie verändert sich unter den Bedingungen einer zunehmenden Digitalisierung das **Berichtswesen in Evaluation** hinsichtlich der medialen Aufbereitung (z.B. interaktive Online-Berichte, Data Visualisation), aber auch hinsichtlich der zeitlichen Taktung (‚just-in-time reporting‘)?
- Wie werden sich **Datenerhebung und Datenanalyse in Evaluationen** verändern? Welche Rolle werden Big Data, Data Mining oder zunehmend intelligente Auswertungsverfahren v.a. für qualitative Daten spielen?
- Welche weiteren Schritte zur **Professionalisierung der Evaluation** werden wir erleben? Wird es zu einer Vereinheitlichung der Fachsprache kommen? Wird es Schließungsanstrengungen geben, um Evaluation vor unqualifizierten Anbietern zu schützen? Welche Folgen haben internationale Ansätze zu standardisierten Anerkennungsverfahren für Evaluierende?

1 Vgl. <https://evalblog.com/2012/01/30/the-future-of-evaluation-10-predictions/>

- Und schließlich: Wie verändert sich insgesamt die **Nachfrage nach Evaluationen**? Erleben wir nach 20 Jahren einer vielfach konstatierten steigenden Nachfrage (die sich auch in einem kontinuierlichen Wachstum der DeGEval spiegelt) derzeit eine Sättigung oder gar einen Gipfel vor dem Abstieg?

Neben diesen Fragen zur Zukunft der Evaluation wird ein zweiter Schwerpunkt der Tagung die **Zukunft als Gegenstand in der Evaluation** sein. Im Mittelpunkt steht dabei die Frage: Inwiefern kann und soll Evaluation valide Aussagen über die Zukunft treffen, obwohl sie sich als empirisch gestütztes Verfahren immer nur auf Vergangenes beziehen kann? Diese Frage berührt vielfache Aspekte der Evaluationstätigkeit:

- Aussagen über die Zukunft treffen Evaluationen immer dann, wenn sie die **Generalisierbarkeit von Befunden** feststellen, da damit eine Prognose hinsichtlich zukünftiger Wirkungen einer evaluierten Maßnahme getätigt wird. Unter welchen Bedingungen sind solche Prognosen tatsächlich tragbar? Warum scheitern positiv evaluierte Innovationen dennoch oft nach Einführung im Feld?
- Auch **Empfehlungen**, die in Evaluationsberichten ausgesprochen werden, enthalten immer eine Vorhersage in Form von prognostizierten Verbesserungen für die Zukunft. Auf welcher Basis macht Evaluation im Einzelfall solche Vorhersagen, vor allem, wenn sich Empfehlungen nicht zu 100% aus den empirischen Befunden ergeben? Wie geht Evaluation mit den dabei inhärenten Unwägbarkeiten um?
- Bei verschiedenen Evaluationsansätzen ist die Zukunft darüber hinaus genuine Perspektive der Betrachtung, indem sie Pläne, Anträge und sonstige in die Zukunft gerichtete Konzepte bewerten. Zu ihnen gehören die **Ex-ante-Evaluation**, die **präformative Evaluation** oder die **Evaluierbarkeitsanalyse**. Sie befassen sich genuin mit Fragen, die die Zukunft betreffen, etwa nach Bedarfslagen, Realisierbarkeit oder Erfolgchancen. Wie und in welchen Kontexten haben sich entsprechende Ansätze bewährt und wo besteht brachliegendes Potenzial? Wie verhalten sich die Ansätze untereinander und zu ähnlichen Verfahren wie der Zukunftsforschung, Foresight-Prozessen oder der Gesetzesfolgenabschätzung?

Wir laden Sie herzlich ein, an der Diskussion in Mainz teilzunehmen und Ihre Arbeit auf der Jahrestagung vorzustellen. Eingeladen sind Beiträge, die sich mit den obengenannten Fragen wissenschaftlich, reflexiv oder praktisch auseinandersetzen. Darüber hinaus sind auch alle anderen Beiträge mit Evaluationsbezug willkommen, wenn sie zumindest in der Diskussion das Tagungsthema aufgreifen.

Die Beitragseinreichung wird ab dem 1. November 2016 über das Tagungsmanagementsystem der DeGEval (<https://www.conftool.com/degeval2017>) möglich sein. Die Tagung bietet folgende Beitragsmöglichkeiten:

- **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). Für den Vortrag stehen 20 Minuten und 10 Minuten Diskussionszeit zur Verfügung. Einzelvorträge werden zu thematischen Sessions zusammengestellt (max. drei Vorträge pro Session).
- **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 fünfminütigen Blitzvorträgen, an die sich je drei Minuten 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.

- **Komplette Session:** Komplette Sessions (90 Minuten) können als Vortragsession, als Podiumsdiskussion, Workshop oder in anderer Form konzipiert sein. Abstracts umfassen die Beschreibung der geplanten Session sowie aller beteiligter Beiträge und Personen. Bei einer Vortragsession sind kurze Inhaltsangaben der einzelnen Vorträge mit einzureichen sowie die/der Moderator(in) und ggf. Diskutant(inn)en zu benennen. Bitte reichen Sie diese dann gesammelt in einem Dokument ein.
- **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 max. 3.000 Zeichen inklusive Leerzeichen (für komplette Sessions mit max. 6.000 Zeichen) bis zum **11. März 2017** über unser Tagungsmanagementsystem (<https://www.conftool.com/degeval2017>) ein und beachten Sie bitte die dort angeführten Hinweise zur Einreichung. Alle Beiträge werden in einem anonymen Review-Verfahren begutachtet. Die Kriterien der Begutachtung finden Sie auf der Tagungshomepage. Sie werden bis zum 31. Mai 2017 über die Beitragsannahme informiert. Die Tagungshomepage mit weiteren Informationen zur Tagung finden Sie unter www.degeval.de/veranstaltungen/jahrestagungen/mainz-2017/.

Kontakt:

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

Bericht über das Forschungs- und Praxiskolloquium am 21. Mai 2016 im Rahmen der Frühjahrstagung des AK Methoden der DeGEval

*Susanne Mäder,¹ Miriam Burfeind,² Annekatriin Gehre,³
Chiara Pierobon,⁴ Angela Ulrich⁵*

Am 21.05.2016 fand bei der Gesellschaft für Internationale Zusammenarbeit (GIZ) das Forschungs- und Praxiskolloquium des Nachwuchsnetzwerks statt. Die Veranstaltung schloss sich an die Frühjahrstagung des AK Methoden zum Thema „Zur Qualität qualitativer Sozialforschung in der Evaluation – Möglichkeiten und Grenzen“ an. Das Forschungs- und Praxiskolloquium wurde zum fünften Mal in Folge angeboten und war mit circa 25 Teilnehmenden sehr gut besucht. Zielgruppe des Forschungs- und Praxiskolloquiums sind vorrangig Nachwuchswissenschaftlerinnen und -wissenschaftler sowie Berufseinsteigerinnen und -einsteiger im Bereich der Evaluation. Evaluation ist in der Regel interdisziplinär und daher an den Hochschulen eher selten durch einschlägige Fachbereiche, Institute oder Lehrstühle repräsentiert. Ziel des Kolloquiums ist es daher, Studierenden und Promovierenden eine evaluationsfachliche Unterstützung für ihre Qualifizierungsarbeiten bereitzustellen. Seit 2014 können auf Wunsch des Nachwuchses im Rahmen des Kolloquiums auch Praxisprojekte vorgestellt und diskutiert werden. Novizinnen und Novizen erhalten damit einen Raum für die Diskussion von Fragen im Zusammenhang mit konkreten Evaluationsprojekten. Das Format ist als Peer-to-Peer-Ansatz angelegt, wodurch eine offene, wertschätzende und niedrigschwellige Form der Unterstützung ermöglicht werden soll. Zudem soll das Kolloquium dem Nachwuchs die Gelegenheit zur Vernetzung bieten. Hierzu fand zunächst im Rahmen des Kolloquiums eine intensivere Kennenlernphase statt, bei der die Teilnehmenden sich gegenseitig interviewten und entlang von persönlichen Steckbriefen miteinander ins Gespräch kamen. In diesem Jahr wurden im Anschluss ein Praxisprojekt, zwei Masterarbeiten und ein Habilitationsprojekt vorgestellt und diskutiert. Im Folgenden werden die Beiträge

1 Univation – Institut für Evaluation Dr. Beywl & Associates GmbH, Köln

2 Universität Osnabrück

3 Universität Augsburg

4 Universität Bielefeld

5 Hochschule der Bundesagentur für Arbeit, Mannheim

inhaltlich kurz vorgestellt und zentrale Themen der anschließenden Diskussionen nachgezeichnet. Im Anschluss beschreiben drei Referentinnen, wie sie das Kolloquium erlebt haben und was sie für ihre Arbeit mitgenommen haben.

Als Praxisprojekt wurde das Projekt Praelab (Praevention von Lehrabbrüchen) von Angela Ulrich vorgestellt, das seit 2010 in Kooperation mit der Bundesagentur für Arbeit von der Hochschule der Bundesagentur für Arbeit (HdBA) durchgeführt wird. Praelab hat das Ziel, durch ein ‚Frühwarnsystem‘ Ausbildungsabbruchtendenzen von Auszubildenden zu erkennen. Dies geschieht, indem Azubis nach dem Ende der Probezeit (etwa nach vier Monaten) in ihren Berufsschulen eine Online-Befragung zu überfachlichen Kompetenzen und zur Ausbildungszufriedenheit durchlaufen. Diese Befragung misst nach Selbsteinschätzung Sozial-, Methoden- und Personalkompetenzen (überfachliche Kompetenzen). Von diesen überfachlichen Kompetenzen wird, durch Forschungsergebnisse gestützt, angenommen, dass sie in bestimmter Ausprägung für den erfolgreichen Abschluss einer Ausbildung notwendig sind. In Kombination mit den Angaben zur Ausbildungszufriedenheit werden so Azubis mit Abbruchrisiko identifiziert. Diesen Jugendlichen wird eine speziell entwickelte Beratung angeboten. Im Rahmen einer Evaluation soll nun überprüft werden, ob Praelab wirkt und das Abbruchrisiko bei Azubis, die Praelab durchlaufen und nach der Methode beraten wurden, geringer ist. Als zentrale Schwierigkeit bei der Befragung der Auszubildenden stellte sich die Erreichbarkeit der Jugendlichen heraus. Diskutiert wurde daher im Rahmen des Kolloquiums, welche Formen der Ansprache wie online, SMS oder E-Mail sich bei der Zielgruppe eignen würden. Außerdem ergab die Diskussion, dass neben diesen individuellen Ansprachestrategien auch Befragungen während der Schulzeit (aufgrund der besseren Erreichbarkeit) eine sinnvolle Strategie wären. Generell bestanden Bedenken seitens der Teilnehmenden, die Befragung auf lediglich eine Frage („Besteht die Ausbildung noch?“) zu reduzieren, um so die Rücklaufquote zu erhöhen. Um zu erfahren, ob die Ausbildung fortbesteht, könnten alternativ auch die Ausbildungsbetriebe angefragt werden. Insgesamt wurde die erreichte Rücklaufquote von ca. 23% für diese Zielgruppe und Art der Befragung als gar nicht so schlecht eingeschätzt.

Kommentar von Angela Ulrich: „Die Veranstaltung war für mich eine gute Möglichkeit, zahlreiche Fragen, die sich während meiner Arbeit ergeben haben, mit fachkundigen Teilnehmenden zu besprechen. Die Atmosphäre habe ich als sehr wertschätzend erlebt. Es war hilfreich, dass gerade kein Bericht über ein abgeschlossenes Projekt erwartet wurde. So bot sich mir die Möglichkeit, Inputs aus verschiedenen Fachrichtungen noch in mein Projekt integrieren zu können. Die Teilnehmenden habe ich als sehr interessiert wahrgenommen. Es wurde ernsthaft mit überlegt und von eigenen Erfahrungen berichtet, wie Jugendliche als längerfristige Studienteilnehmende gewonnen werden könnten.“

Nachfolgend präsentierte Annekatrin Gehre von der Universität Augsburg das methodische Vorgehen und ausgewählte Ergebnisse einer Wissenssoziologischen Diskursanalyse zur Erfolgseinschätzung von Entwicklungsprojekten in Evaluationsberichten der GIZ und der KfW. In Rahmen ihrer Masterarbeit unternimmt sie damit den Versuch, von einer kritischen Analyse der Erfolgs- und Wirksamkeitsbewertung innerhalb von Ex-post-Evaluationsberichten zu einem konstruktiven Beitrag für die

deutsche Evaluationspraxis bzw. Entwicklungszusammenarbeit zu kommen. Die Arbeit untersucht das Wissen, das durch Evaluationsberichte an die Gesellschaft kommuniziert wird. Im Fokus steht dabei, wie Erfolgs- und Misserfolgseinschätzungen sprachlich vermittelt werden und welche narrative Struktur sich als Verständnisangebot aus den analysierten Ex-post-Kurzberichten ergibt. Analysiert wird hierbei auch, inwieweit Erklärungen, Entschuldigungen und Schuldzuweisungen im Zusammenhang mit den Bewertungen vorgenommen werden. Im Ergebnis zeigt sich in Verbindung mit der Erfolgseinschätzung der Projekte eine Anpassungs- und Legitimierungspraxis als Deutungsmuster. Erfolg wird durch die gelungene Anpassung – in Form der Abbildbarkeit eines Projektes mittels vorab festgelegter Indikatoren – an übergeordnete Bildungs- und Armutreduktionsstrategien (beispielsweise EFA⁶, PRSPS⁷) attestiert. Legitimiert wird das Projekt final mit eben jenen Strategiepapieren, die nicht weiter begründungsbedürftig sind. Die Möglichkeit, eine Referenz herstellen zu können, führt bereits zu einer positiven Erfolgsbewertung. Außerdem fällt bei der Betrachtung der Erfolgseinschätzungen entlang der Kriterien des Development Assistance Committee (DAC) auf, dass beispielsweise bei der Bewertung der Effektivität eine Verdrängungspraxis vorherrscht. Dies meint, dass in den Berichten zum Beispiel nicht hinterfragt wird, warum das Projekt gut oder eben nicht gut funktioniert hat. Aus Sicht der Teilnehmenden wird deutlich, dass die DAC-Kriterien (Relevanz, Effektivität, Entwicklungspolitische Wirkungen, Effizienz und Nachhaltigkeit) in den Berichten offensichtlich anders als intendiert verstanden und genutzt werden. Außerdem wird angemerkt, dass die Ex-post-Kurzberichte vor allem der Rechenschaftslegung dienen und daher weniger Erklärungen und vertiefende Informationen enthalten. Ein Einbezug weiterer Berichte würde demgemäß ein verändertes Ergebnis ergeben. Die Referentin verweist an dieser Stelle nochmals darauf, dass es bei einer wissenssoziologischen Analyse nicht darum geht, das richtige Verständnis zu suchen, sondern die zur Verfügung stehenden Informationen zu analysieren. Die Ex-post-Berichte werden dabei als ein öffentliches Wissensprodukt verstanden, das sich an jede Bürgerin/jeden Bürger richtet und somit für sich steht und damit auch isoliert betrachtet werden kann.

Kommentar von Annekatrin Gehre: „Das Praxiskolloquium hat mir in sehr vertrauter und wohlwollender Atmosphäre die Möglichkeit geboten, einen interdisziplinären Austausch und Diskurs zu erleben. Die Diskussionen, Fragen und Ratschläge habe ich als sehr unterstützend, interessiert und vor allem fruchtbar wahrgenommen. Persönlich nehme ich aus dieser Veranstaltung die Wichtigkeit und Bedeutung einer verstärkten Forschung über Forschung mit. In meinem Fall folglich die Konfrontation der Wissenssoziologie mit einer angewandten Sozialforschung wie der Evaluation, wodurch erstere punktuell aus dem Elfenbeinturm als Grundlagenforschung heraustreten kann und selbst zu einer Art angewandten Forschung werden kann. Für die Evaluationsforschung kann ein derartiges Aufeinandertreffen die Konfrontation mit einer extern hergestellten Konstruktion evaluativen Arbeitens bedeuten, die das Nachdenken über das eigene wissenschaftliche Vorgehen herausfordert und in konstruktiver Weise motiviert.“

6 Education for All – EFA

7 Poverty Reduction Strategy Paper – PRSP

Dr. Chiara Pierobon von der Universität Bielefeld präsentierte ihr Forschungsprojekt über die Bedeutung der EU-Förderung für die Stärkung der Zivilgesellschaft in Kasachstan, Kirgistan und Tadschikistan. Im Besonderen werden die Auswirkungen der Förderung zivilgesellschaftlicher Organisationen analysiert, welche im Rahmen des „European Instrument for Democracy and Human Rights“ (EIDHR) und des thematischen Programmes „Non-State Actors and Local Authorities in Development“ (NSA-LA) seit 2007 stattfindet. Die Besonderheit dieser Forschung besteht darin, die Auswirkungen der EU-Maßnahmen im Hinblick auf Sozialkapital und dessen drei Dimensionen – die kognitive, die relationale und die strukturelle – zu bemessen. Dabei liegt der Fokus der Analyse auf den Werthaltungen und politischen Orientierungen, die die Projekte charakterisierten (kognitives Kapital), auf den hergestellten Verbindungen zwischen Akteuren auf individueller, zivilgesellschaftlicher und Regierungsebene (relationales Kapital); und auf den Eigenschaften dieser Beziehungen vor allem im Hinblick auf den Aufbau von Vertrauen (strukturelles Kapital). Die Präsentation konzentrierte sich auf die Ergebnisse der ersten Erhebungswelle, die 2014 in Kasachstan durchgeführt wurde und stellte die Verwendung des Sozialkapitals als analytisches Instrument für die Analyse zivilgesellschaftlicher Förderungsmaßnahmen zur Diskussion. Seitens der Teilnehmenden wurde darauf hingewiesen, dass sich die theoretische Konzeption von Sozialkapital innerhalb der Arbeit nicht mit den Vorstellungen der EU deckt. Zudem wurde diskutiert, inwieweit es sich bei diesem Projekt um Forschung oder Evaluation handelt. Seitens der Wissenschaft wird das Projekt laut der Referentin mehr der Evaluation zugeordnet, weshalb sie die Einschätzung der Evaluationscommunity zu dieser Frage interessiert. In der Diskussion zeigte sich, dass die Anwesenden bei einer stärkeren Orientierung an Evaluation als Disziplin eine systematische Erfassung der Wirkungen, die Gestaltung von Bewertungsprozessen und die Auseinandersetzung mit den Zielvorstellungen der Programmverantwortlichen als wichtig erachten.

Kommentar von Dr. Chiara Pierobon: „Die Präsentation ermöglichte mir, wichtige methodologische Fragen zu diskutieren, die im Laufe meiner Pilotstudie auftauchten, und Anregungen zu sammeln, wie die „Standards für Evaluation“ in meinem Projekt besser berücksichtigt werden könnten. Diskutiert wurde die Nützlichkeit, die Auswirkungen der EU-Maßnahmen im Hinblick auf Sozialkapital zu bemessen. Gleichmaßen wurde die Bedeutung der kognitiven, strukturellen und relationalen Dimensionen des Sozialkapitals für die Stärkung der Zivilgesellschaft erörtert. Ein fruchtbares Ergebnis meiner Teilnahme am Nachwuchskolloquium, das im wechselseitigen Austausch generiert wurde, war der Vorschlag, konkrete Hypothesen zu entwickeln, die die Theorie des Sozialkapitals mit dem von der EU verwendeten Begriff von *Civil Society Empowerment* verknüpfen.“

In der zweiten Masterarbeit, die von Miriam Burfeind von der Universität Osnabrück vorgestellt wurde, wird der Effekt spielerischer Auseinandersetzung mit Lerninhalten auf den Lernerfolg untersucht. Hierfür wurde die Lernsimulation von Simdustry® genutzt. Im Fokus der Arbeit steht, inwieweit der haptische Aspekt wie „Anfassen und Bewegen“ die Lernprozesse verbessert. Dies wird mittels eines einfaktoriellen Designs mit Experimental- und Kontrollgruppe untersucht. Dabei wurden 60 Versuchspersonen einer Kontroll- und einer Experimentalgruppe zugewiesen.

In der Kontrollgruppe besteht die Arbeitsaufgabe darin, einen Lehrtext zu lesen, die Handlungen in ihrer Vorstellung nachzuvollziehen und im Anschluss Bilanzen und eine Gewinn- und Verlustrechnung zu erstellen. In der Experimentalgruppe erhalten die Versuchspersonen den gleichen Text und die gleiche Aufgabe und zusätzlich die Aufforderung, die Handlungen auf dem Spielbrett nachzuvollziehen. Erhoben werden im Anschluss die Akzeptanz, die Benutzerfreundlichkeit, der Spaß, der Flow und der subjektive Lernerfolg (zur Erfassung der Reaktion der Teilnehmenden) sowie mittels eines Kurztests der Lernerfolg aufseiten der Versuchspersonen. Auch das Transferwissen (Verhalten) wird mittels eines kurzen Tests abgefragt. Hypothetisch wird davon ausgegangen, dass Akzeptanz, Lernerfolg und Transferwissen bei der Gruppe mit Simulation der Lerninhalte höher ist als bei der Gruppe ohne Simulation. Seitens der Referentin besteht vor allem das Interesse, alternative Erklärungen für den Fall zu eruieren, dass sich die Hypothesen nicht bestätigen. Im Kolloquium werden folgende Ideen gesammelt: So könnten die Probandinnen und Probanden durch das Spiel auch von der eigentlichen Aufgabe abgelenkt werden oder gegebenenfalls zu viel Zeit mit dem Spiel verbringen. Weiterhin könnte es unterschiedliche Vorerfahrungen und Affinitäten gegenüber Brettspielen geben und gegebenenfalls könnten auch die unterschiedlichen Lerntypen aufseiten der Versuchspersonen das Ergebnis beeinflussen. Interessant wäre auch zu prüfen, ob Personen aus der Kontrollgruppe Zeichnungen anfertigen (ebenfalls haptischer Reiz).

Als nächstes wird sich das Nachwuchsnetzwerk auf der Jahrestagung 2016 in Salzburg treffen. Dort kann sich der Nachwuchs mit Autorinnen und Autoren aktueller Publikationen über praktische und theoretische Fragen austauschen („Meet the Authors“). Weitere Hinweise zu den Aktivitäten des Nachwuchsnetzwerks finden sich auf der Internetseite <http://www.degeval.de/arbeitskreise/nachwuchsnetzwerk/>. Interessentinnen und Interessenten können zudem im E-Mail-Verteiler https://de.groups.yahoo.com/neo/groups/evaluation_nachwuchs/info oder in der LinkedIn-Gruppe des Nachwuchsnetzwerks mitdiskutieren.

Aktuelle Formate der beruflichen Aus- und Weiterbildung – Evaluation in einem weiten Feld

Frühjahrstagung des Arbeitskreises Berufliche Bildung im Juni 2016

Michael Kalman¹, Anja Lietzmann²

Am 16.06.2016 fand die Frühjahrstagung 2016 des Arbeitskreises Berufliche Bildung am Bundesinstitut für Berufsbildung (BIBB) in Bonn statt. Rund 30 Expertinnen und Experten tauschten sich zum Thema „Aktuelle Formate der beruflichen Aus- und Weiterbildung. Evaluation in einem weiten Feld“ aus.

Die berufliche Aus- und Weiterbildung ist in vielerlei Hinsicht von interessanten (Neu-)Entwicklungen geprägt. Die Frühjahrstagung nahm solche Entwicklungen zum Anlass, um die Evaluationen relevanter aktueller Formate der beruflichen Aus- und Weiterbildung aufzugreifen. Hierbei wurde eine Bandbreite unterschiedlicher Themen abgebildet: Ansätze des Distance Learning sowie ein Mobile Learning-Tool kamen ebenso zur Sprache wie die Mobilität von Auszubildenden innerhalb Europas. Mit den Themen variierten auch Evaluationsgegenstände und -methoden, die im Rahmen der Fachtagung diskutiert wurden.

Einleitend stimmte *Frau Dr. Anja Lietzmann* vom Sprecherteam des Arbeitskreises die Anwesenden mit Ihrem Input „Aktuelle Formate der beruflichen Aus- und Weiterbildung“ auf das Tagungsthema ein. Sie skizzierte mit Themen wie digitales Lernen, Mobilität sowie Qualitätssicherung Megatrends, von der die berufliche Bildung aktuell wesentlich bestimmt ist. Dr. Lietzmann nahm die Leitmotive auf, die im Folgenden die Tagung bestimmten und weiter elaboriert wurden.

Danach stellten drei Experten ihre Evaluationsansätze im Kontext einer großen Fernschule, eines Projektes für neuartige mobile Lernangebote sowie eines Sonderprogramms des Bundes vor.

Als erster Referent sprach *Herr Andreas Vollmer*, Mitglied der Geschäftsleitung der Studiengemeinschaft Darmstadt (SGD) und der Wilhelm Büchner Hochschule Darmstadt über „Integration von Evaluation in das Qualitätsmanagement einer großen Fernschule“.

1 Mitglied Sprecherteam des AK Berufliche Bildung, Kalman Consult, Berlin

2 Mitglied Sprecherteam des AK Berufliche Bildung, k.o.s. GmbH, Berlin

Vollmer stellte zunächst die Studiengemeinschaft Darmstadt (SGD) vor, die zu den großen Fernschulen in Deutschland gehört. Rund 150 festangestellte Mitarbeiter(innen) am Standort Pfungstadt bei Darmstadt und über 500 freie Mitarbeiter(innen) in ganz Deutschland betreuen circa 60.000 Teilnehmende in über 200 Fernkursen jährlich. Die SGD ist zugelassen bzw. zertifiziert nach den Qualitätsmanagementmodellen ISO 29990, AZAV und der Staatlichen Zentralstelle für Fernunterricht (ZFU).

Vollmer stellte verschiedene Praxisbeispiele vor, wie Evaluationsinstrumente eingesetzt werden. Ein Schwerpunkt bildet aus naheliegenden Gründen die Wahrnehmung der Bildungsdienstleistung durch die Kund(inn)en mit Blick auf deren Zufriedenheit – für Marktteilnehmer wie SGD eine essenzielle Kategorie. Über eine Online-Befragung mit Versand eines Links über den firmeneigenen Online-Campus waveLearn werden alle aktiven Kund(inn)en, d.h. Teilnehmende der Fernkurse, erreicht. Schwerpunkt dieser Befragung sind die wahrgenommenen Leistungen und die Qualität der Fernlehrer(innen) und Tutor(inn)en. Die ausgefüllten Fragebögen werden automatisch und anonymisiert über Evasys ausgewertet, die Fernlehrer(innen) erhalten ihr individuelles Ergebnis und das Gesamtergebnis per E-Mail zugesandt.

Weitere Instrumente, um Kunden-Feedback einzuholen, sind das Beschwerdemanagement über das CRM-System demsy. Die zentral einkommenden Beschwerden werden an die Fachbereiche weitergeleitet; es erfolgt ein regelmäßiger, übergreifender Report zur Ursachenanalyse und Maßnahmenermittlung.

Auch über externe Bewertungsportale wird die Kundenzufriedenheit erfasst. Dabei werden negative Bewertungen vom Service zusammen mit den Fachbereichen geprüft und kommentiert; Erkenntnisse fließen dann in die Entwicklungsprozesse mit ein.

Zu den internen Bewertungsaktivitäten, die eher dem Monitoring oder Controlling zuordenbar sind, gehört die Analyse von Kennzahlen wie Rücktrittsquoten, Kündigungsquoten und Bestehensquoten, die in Masterreports zusammengefasst werden.

Ein Beispiel beauftragter, externer Evaluation ist die Personalerbefragung zu Trends in der Weiterbildung durch die Firma TNS Infratest. Dabei wird u.a. die Akzeptanz von Fernunterricht ermittelt. Die Ergebnisse dieser regelmäßig durchgeführten Befragungen gehen in die Weiterentwicklung der Bildungsangebote ein.

Vollmer bot interessante Innenansichten zur Evaluation und zum Qualitätsmanagement eines großen, innovativen Bildungsträgers. Dabei geht der Referent von einem relativ weiten Evaluationsbegriff aus.

Im Rahmen der Diskussion und der Nachfragen aus dem Teilnehmendenkreis wurden verschiedene Aspekte angesprochen. So wurde die Frage nach den genauen Evaluationsgegenständen gestellt. Gegenstand seien in erster Linie die Maßnahmen, dabei vor allem die Betreuung durch Lehrkräfte und Tutor(inn)en, sowie die Kursinhalte. Daneben werde auch die Organisation selbst evaluiert, einschließlich des Managements.

Auch die Impulse des Anforderungskataloges der DIN ISO 29990, einer weltweit gültigen Weiterbildungsnorm für die Evaluation und die Betriebsorganisation

des Bildungsträgers, war Thema. Prozesse würden dadurch von A bis Z angeschaut. Man betrachte den gesamten Kundenfall, von Anfang bis Ende – nicht nur partielle Ereignisse. Umgreifende Einblicke würden so möglich. Auf Führungsebene würden Mitarbeitergespräche stärker formalisiert. Es gebe zudem bessere Aufgaben- und Prozessbeschreibungen. Schließlich würde das Verbesserungsmanagement etabliert und optimiert.

Der personelle Aufwand, das QM-System aufrechtzuerhalten, sei wegen des hohen Automatisierungsgrads gering. Personal könne QM-relevante Aufgaben neben ihrem normalen Job versehen. Es gäbe aber Potenzial, die Daten mit zusätzlichen Ressourcen noch besser auszuwerten.

Zu diesen Daten gehören auch solche zur Integration in den Arbeitsmarkt, insbesondere bei den arbeitssuchenden Teilnehmenden. Diese müssen insbesondere gemäß der Akkreditierungs- und Zulassungsverordnung Arbeitsförderung (AZAV) erhoben werden. In diesem Kontext würden Zielgruppen über die Vermittlung in Arbeit und nach dem beruflichen Fortgang befragt. Oft wird ein beruflicher Aufstieg schon während der Weiterbildung beobachtet.

Der Referent vertiefte im Diskussionsteil auch die Funktion des Beschwerdemanagements. Es gebe Einzelfälle, die betrachtet werden, einerseits, und das Gesamtbild auf Kursebene andererseits, einschließlich der Erstellung von Quoten. Auch hier wird die Rolle von qualitativem Feedback betont. Im Beschwerdemanagement gäbe es zehn vorstandardisierte Beschwerdetypen, unter denen ausgewählt werden könne. Die meisten Beschwerden beträfen Lernmaterialien und die Aktualität der Lerninhalte. Nicht immer könne aber etwa die Aktualität verbessert werden, da sich die Kursinhalte an Prüfungsinhalten ausrichten müssten. Man träte dann in Kontakt mit Prüfungsinstanzen, um Aktualisierungen zu erwirken.

Bei Abbruch eines Kurses würden systematische Befragungen nach Kündigungsgründen durchgeführt. Gegebenenfalls würden auf Basis der erhobenen Daten Verbesserungsmaßnahmen eingeleitet.

Als zweiter Referent behandelte *Herr Prof. Dr. Marc Beutner*, Lehrstuhlinhaber für Wirtschaftspädagogik II, Wirtschaftspädagogik und Evaluationsforschung an der Fakultät für Wirtschaftswissenschaften der Universität Paderborn das Thema „Evaluation eines Mobile Learning-Tools im Projekt ‚NetEnquiry‘“.

Als Evaluationsgegenstände nennt Beutner die Evaluation eines entwickelten und eingesetzten Mobile Learning-Tools und das Projekt ‚NetEnquiry‘ selbst. Als weitere spezifizierende Evaluationsgegenstände werden genannt: die Lernenden, die Akzeptanz und Nutzerfreundlichkeit des Tools und Nachhaltigkeit.

Ziel des Projekts NetEnquiry ist die Entwicklung, Erprobung und Evaluation neuartiger mobiler Lernangebote in Kombination mit dem dazugehörigen Softwaretool. Damit soll die Nutzung der Potenziale des mobilen Lernens im Rahmen der betrieblichen Aus- und Weiterbildung ermöglicht werden.

Im Vorhaben wird realitätsnahes Lernen im exemplarischen Bereich des Bankensektors (Privatbanken, öffentlich-rechtliche Kreditinstitute und Genossenschaftsbanken) mit dem Einsatz mobiler Endgeräte verknüpft, Anforderungen des Arbeitsumfeldes berücksichtigt und dazu beigetragen, komplexe betriebliche Abläufe und

Systeme in ihrer Gesamtheit zu erlernen und zu verstehen. Das Setting bezieht auch berufsbildende Schulen als erweiterte Erprobungspartner mit ein.

Sehr anschaulich legte Beutner den Einsatz des Mobile Learning-Tools anhand des Szenarios ‚Privatkredit‘ dar. Darin sollen sich die Lernenden mit fachlichen, sozial-kommunikativen sowie moralisch-ethischen Herausforderungen im Rahmen einer Privatkreditvergabe auseinandersetzen. Dazu gehören u.a. Fragestellungen der Gesprächsführung, der Besicherung, der Kapitaldienstfähigkeitsberechnung sowie der Kreditarten. Das Szenario enthält Rollen aus dem realen Bankenbetrieb wie Kundenberater(innen), Kreditmanager(innen) und Filialleiter(innen). Es sind Abwägungen sowohl im kooperativen (Zusammenarbeit der genannten Rollen in einem Team) als auch im wettbewerblichen Modus – die gleiche Aufgabe wird von mehreren Teams bearbeitet – zu treffen. Sämtliche Komponenten des Szenarios sind aus den realen Geschäftsprozessen der beteiligten Projektpartner aus dem Bankensektor in das Mobile Learning Tool einprogrammiert.

Als Evaluationsgrundlagen greift Beutner auf das ‚Dreischalenmodell der systematisch-analytischen Berufsbildungsevaluation‘ zurück. Hier werden unterschiedliche Dimensionen, wie ‚Kontext‘, ‚Lehr-/Lernprozess‘, ‚Outcome‘ etc. in einer Mehrebenenbetrachtung berücksichtigt. Ferner bildet Kromreys Zielkatalog von Evaluation einen weiteren theoretischen Hintergrund. Danach sind Werturteile zu fällen über Wirkung, Umsetzung/Implementation und Akzeptanz. Das bedingt das Herangehen der Wirkungsanalyse hinsichtlich der Effekte, der Implementationsforschung hinsichtlich der systematischen Planung, Durchsetzung und Umsetzung sowie der Akzeptanzforschung hinsichtlich der Zustimmung oder Ablehnung durch Personen bzw. Gruppen. Als Erhebungsformen werden schriftliche und mündliche Interviews und eine Tool-Analyse eingesetzt.

Mündliche Einzelinterviews wurden mit den involvierten Lehrkräften, Softwareentwickler(inne)n und Bankenvertreter(inne)n geführt. Die gleichen Akteure wurden gemeinsam in qualitativen Gruppeninterviews zu ihren Einstellungen hinsichtlich Authentizität, Nutzen, Didaktisches Design sowie Chancen und Herausforderungen befragt. Daneben wurden die beteiligten Berufskollegs um Stellungnahmen u.a. zum Transferfokus und den Umsetzungspotenzialen in Schulen, zur Usability und Verständlichkeit, aber auch zu den Herausforderungen im Schulkontext gebeten.

Quantitative schriftliche Befragungen verschiedener Akteure wurden u.a. zur Akzeptanz und Zukunftsperspektiven von Mobile Learning durchgeführt. Weitere quantitative schriftliche Befragungen zur Dozentenschulung und für Lernende komplettieren den vielseitigen Einsatz der Evaluationsmethode Befragung. Die Lernenden sollten sich vor allem zu ihrer Zufriedenheit, Tool-Einschätzung, Potenzialen, didaktischem und technischem Design äußern.

Die durchgeführte Tool-Analyse gliedert sich in die Usability-Untersuchung, zum einen quantitativ, ferner mittels des TAM-Modells,³ und zum anderen in das

3 TAM bedeutet Technology Acceptance Model. In der Wirtschaftsinformatik bezeichnet TAM ein Modell, welches Aussagen darüber trifft, warum Personen eine Technologie nutzen oder nicht nutzen (vgl. Wikipedia Technology Acceptance Model, https://de.wikipedia.org/wiki/Technology_Acceptance_Model, 27.06.2016).

qualitative Szenario-Testing. Hier wurden die Bereiche technische und didaktische Umsetzung, Vollständigkeit, Richtigkeit, Prozessabbildung und Szenario-Gestaltung untersucht.

Die Erhebungen werden je nach Zeitpunkt zum einen für die formative Evaluation verwendet, um das laufende Projekt bzw. das eingesetzte M-Learning-Tool zu verbessern, zum anderen auch für die summative Evaluation, die nach dem intendierten Projektabschluss Ende Juli 2016 fertiggestellt werden soll.

Obwohl die Evaluation noch nicht abgeschlossen ist, referiert Beutner verschiedene Zwischenergebnisse. Die Ergebnisse des TAM-Bogens waren positiv bis sehr positiv: so stimmten über 85 % der User der Aussage stark oder vollkommen zu, dass das M-Learning Tool leicht zu benutzen sei. Die halbstrukturierten, qualitativen Interviews mit den verschiedenen Endnutzer(inne)n ergaben überwiegend positive Aussagen und in Grenzen auch Verbesserungspotenziale. So wird das kompetitive Design als unterschiedlich wichtig angesehen. Ferner wird darauf verwiesen, dass mobiles Lernen bislang nur sehr begrenzt umgesetzt wird.

In der Diskussion wurde die Frage aufgeworfen, ob NetEnquiry auch noch in fünf Jahren genutzt werden wird. Für Beutner bestehen hier aufgrund der Einbindung der Partner gute Hoffnungen. Weitere Aspekte waren die Offenheit von NetEnquiry, die u.a. Probleme der Informationssicherheit aufwerfen. Dafür gebe es zwar Richtlinien, aber die würden von unterschiedlichen Akteuren unterschiedlich ausgelegt. Es bliebe eine Herausforderung, Mobile Learning an den Arbeitsplatz zu bringen.

Die Wettbewerbsdimension des Tools wird durch ein Systemscoring bewertet; hierbei werden die Beratungsergebnisse der User-Teams miteinander verglichen. Ferner werde ein Dozentenscoring durchgeführt.

Vergleichsgruppen von Personen, die auf herkömmliche Weise lernen, könnten nicht herangezogen werden, weil NetEnquiry ganz neue Wissensfelder abbilde, die auf herkömmliche Weise nicht oder zumindest weniger anschaulich vermittelt werden. Die Inhalte von NetEnquiry gingen über das normale Curriculum hinaus. Die Lernergebnisse der User würden aber gemessen.

Den Abschluss der Referate bildete *Herr Marco Puxi*, Leiter des Büros Berlin des Instituts für Sozialforschung und Gesellschaftspolitik GmbH (ISG) mit Hauptsitz in Köln. Sein Thema lautete „Evaluation des Sonderprogramms des Bundes ‚Förderung der beruflichen Mobilität von ausbildungsinteressierten Jugendlichen aus Europa (MobiPro-EU)‘“.

Dieses Sonderprogramm zielt auf eine Förderung junger Erwachsener aus der EU ab, mit dem Ziel, Mobilitätshemmnisse abzubauen, junge Europäerinnen und Europäer auf eine duale Ausbildung in Deutschland vorzubereiten und Betriebe bei der erfolgreichen Besetzung freier Ausbildungsplätze mit jungen Menschen aus dem europäischen Ausland zu unterstützen. Das Programm setzt sich aus unterschiedlichen Förderbausteinen zusammen, mit denen vom Sprachkurs im Herkunftsland über Zuschüsse zu Reisekosten für Bewerbungsgespräche in Deutschland bis hin zur Sicherung des Lebensunterhalts und ausbildungsbegleitenden Hilfen alle Phasen der Vorbereitung auf eine Ausbildung als auch die Ausbildung selbst fördertechnisch abgedeckt werden.

Ziel des dargestellten Evaluationsvorhabens in Zusammenarbeit mit dem Institut für Angewandte Wirtschaftsforschung e.V. (IAW) und dem SOKO-Institut ist es, das Sonderprogramm im Hinblick auf Umsetzung, Zielerreichung und kausale Wirkungen zu bewerten. Darüber hinaus sollen von einer laufenden Berichterstattung Impulse zur Steuerung des Programms ausgehen.

Das Evaluationsdesign besteht aus den fünf Bausteinen Monitoring, Qualitative Implementationsanalysen, Ergebnisanalyse, Wirkungsanalyse und Laufende Rückkopplung der Ergebnisse in die Umsetzung. Dies bedingt den kombinierten Einsatz unterschiedlicher Methoden, insbesondere

- Analyse von Monitoringdaten zum Programmverlauf,
- Interviews und vertiefende Fallstudien im In- und Ausland,
- Standardisierte telefonische Befragungen bei Teilnehmenden und Nichtteilnehmenden, bei Betrieben und bei Projektträgern,
- Verzahnung von Monitoring, qualitativen Fallstudien und Interviews sowie quantitativer Zielerreichungs- und Wirkungsanalyse.

Die quantitative Wirkungsanalyse identifiziert den kausalen Effekt als Unterschied zwischen gemessenem Ergebnis und kontrafaktischem Ergebnis (Erfolg der Teilnehmenden, wenn sie nicht an MobiPro-EU teilgenommen hätten). Dazu wurde eine Vergleichsgruppe aus den Integrierten Erwerbsbiographien (IEB)⁴ gebildet.

Die Evaluation hat damit sowohl einen formativen als auch einen summativen Charakter. Als eine mögliche Folge des formativen Evaluationsteils fand während der Programmlaufzeit ab 2015 ein Wechsel von der Individualförderung zur Projektförderung statt. Die Individualförderung der Teilnehmenden von 2013 bis 2014 sah bei aufeinander aufbauenden Förderbausteinen zur Vorbereitung und während der Ausbildung/Beschäftigung eine Antragsabwicklung über die Zentrale Auslands- und Fachvermittlung (ZAV) vor; die Programmsteuerung lag bei der Bundesagentur für Arbeit (BA).

Die Evaluation hatte u.a. einen immensen Antragsbearbeitungsaufwand bei allen Beteiligten identifiziert. Ferner wurde u.a. eine frühzeitige Betreuung der interessierten Teilnehmenden (sogenannte ‚Kümmerer‘), eine Ausweitung und Verbesserung der Sprachkursförderung und eine Streichung der Vorleistungen durch Teilnehmende empfohlen.

Mit der Neuauflage des Programms im Sinne der Projektförderung konnte die Evaluation eine effektive Finanzsteuerung, klare Rahmenbedingungen für Projektträger und ein verbessertes Qualitätsmanagement feststellen. Weitere Handlungsempfehlungen konnte die fortgeführte Evaluation auch in der neuen Programmphase identifizieren. Als besondere Herausforderungen für die Evaluation nennt Puxi u.a. die Umstellung des Programms von Individual- auf Projektförderung und in der Folge die Notwendigkeit der Anpassung des Untersuchungskonzepts ‚on the run‘; ferner die Organisation und Durchführung der Feldphasen in den Herkunftsländern der Teilnehmenden, die Bildung und Identifikation der Vergleichsgruppen für die

4 Die Integrierten Erwerbsbiographien (IEB) sind ein Datenangebot des IAB (Institut für Arbeitsmarkt- und Berufsforschung), welches seit 2004 erzeugt wird. Es umfasst vier verschiedene Quellen und ist damit die umfangreichste Datenmenge zu erwerbshistorischen Meldungen auf Personenebene, die vom Forschungsdatenzentrum der BA am IAB (FDZ) angeboten wird.

kontrafaktische Wirkungsanalyse sowie der diskontinuierliche und unterschiedlich intensive Austausch mit dem Auftraggeber zu den Ergebnissen und Schlussfolgerungen der Evaluation

In der Diskussion wurden verschiedene Fragen aufgeworfen und Aspekte angesprochen. Eine Empfehlung aus der Evaluation an den Auftraggeber sei gewesen, dass ein Programm wie MobiPro-EU besser vorbereitet werden müsse. Werden – so wurde gefragt – solche Empfehlungen überhaupt angenommen? Gemäß Puxi hätte der Auftraggeber absichtlich bestimmte Entscheidungen (Individualförderung und nicht Trägerförderung) getroffen und diese für angemessen gehalten. Er fühlte sich daher gut vorbereitet. Eine andere Frage betraf den Matching-Prozess zwischen Jugendlichen und Unternehmen. Die ZAV der EURES-Länder⁵ hätten die Jugendlichen gesucht. Meist seien es aber nicht Betriebe gewesen, die Bedarfe bei den ZAV anmeldeten, sondern Kammern. Diese gaben Auskunft, was genau in welcher Region gebraucht werde. Zu den Gründen für Ausbildungsabbrüche sagte Puxi, dass nicht selten dann doch Arbeitsstellen im Herkunftsland gefunden wurden. Weitere Gründe seien Heimweh gewesen, fehlende oder ungenügende Informationen über das Berufsbildungssystem des Ziellandes und damit zusammenhängend falsche Erwartungen der Teilnehmer(innen). Die Berufsschulen seien z.T. wegen Sprachproblemen überfordert gewesen. Ferner hätte es Wohnungsprobleme zu Beginn des Programms gegeben.

Zur Frage, ob auch die nachhaltigen Wirkungen des Programms in die Evaluation miteinbezogen werden, wurde ausgeführt, dass dies wegen der zeitlichen Begrenzung des Evaluationsprojekts nicht möglich sei.

Die Intention des Sonderprogramms, ausbildungsinteressierten ausländischen Jugendlichen den Weg nach Deutschland zu bahnen, könnte – so ein Teilnehmer – auch den Braindrain im abgebenden Land befördern. So wurde die Frage gestellt, wie die abgebenden Länder zu diesem Aspekt stehen. Einige Aspekte hierzu – so Puxi – konnten untersucht werden, z.B. das Unterstützungsverhalten in den einzelnen Herkunftsländern. Spanien z.B. unterstütze das Programm stark, Polen und Ungarn hingegen nicht – aus Furcht vor Braindrain.

Zum Abschluss wurde vom Referent festgehalten, dass nicht ermittelt werden konnte, wieviel Geld wirklich bei den Jugendlichen angekommen sei, da die Höhe der Verwaltungskosten unbekannt sei.

In der Abschlussdiskussion wurden anhand vorgegebener Vergleichskriterien Gemeinsamkeiten, Unterschiede und Spezifika der Referatsthemen herausgearbeitet. Ergebnisse wurden auf einer vorbereiteten Matrix an der Pinnwand stichwortartig festgehalten (siehe nachfolgende Tabelle).

5 EURES bedeutet EUROpean Employment Services, ein im Jahr 1993 gegründetes, europaweites Netzwerk, das Arbeitssuchenden dabei hilft, ins Ausland zu gehen und eine Beschäftigung in Europa zu finden.

Tabelle 1: Wandzeitung: Komparative Aspekte

Kriterium	Evaluation und QM (Bildungsträger SGD)	Mobile Learning-Tool (Projekt NetEnquiry)	Sonderprogramm Bund MobiPro-EU
Veranlassung der Ev.	Geschäftsführung SGD	Universität Paderborn; Anfrage aus der Wirt- schaft	Beauftragung durch öffentliche Hand (Bund)
Laufzeit des Ev. Gegenstandes	Permanent	01.08.2013 bis 31.07.2016	Dezember 2013 bis Dezember 2018
Ev. Gegenstand	Organisation, Produkt/ Maßnahmen, TN-Be- treuung, Konzept	Produkt, Projekt	Verschiedene Ebenen eines Programms
Normative Ver- ankerung der Ev.	DIN ISO 29990, AZAV, ZFU	Standards für Evaluation, Curricula	Standards für Evalu- ation
Ev. Art	Formativ/summativ	Formativ/summativ	Formativ/summativ
Ev. Ansatz	Verschränkung QM - Evaluation	Design-based Research, responsive Evaluation	„Handlungsorientierter Ansatz“
Methodeneinsatz	TN-Befragung mündlich/schriftlich, Inhaltsanalyse	TAM2-Analyse, Frage- bogen, Gruppen- und Einzelinterviews, ARIS	Dokumentenanalyse, Auswertung Monitoring- Daten, Fallstudien, Interviews, Sonderaus- wertung IEB
Ev. und Kausalität	Plausibilitäten	Empirische Evidenz, Fal- sifizierbarkeit, Lerntheo- retische Fundierung	Theoriebasiertes Wir- kungsmodell, Kontroll- gruppendesign
Nachhaltigkeit der Ev.	Liegt vor, da in ein permanentes QMS integriert	Universitäre Forschungs- kontinuität ja; hinsichtlich der zukünftigen Anwen- dung ungewiss	Ungewiss, abhängig von politischen Ent- scheidungen

Anmerkung: Das Kriterium „Normative Verankerung der Evaluation“ bezieht sich z.B. auf Normen und Standards, ggf. auch auf den ‚Stand der Technik‘. Die Kriterien „Herausforderungen der Evaluation“ und „Wirkungen und Konsequenzen der Evaluation“ konnten aus Zeitgründen nicht mehr berücksichtigt werden.

Abkürzungen:

ARIS = Architektur integrierter Informationssysteme

AZAV = Akkreditierungs- und Zulassungsverordnung Arbeitsförderung

IEB = Integrierte Erwerbsbiografien

QM = Qualitätsmanagement

QMS = Qualitätsmanagementsystem

TAM = Technology Acceptance Model

TN = Teilnehmer(innen)

ZFU = Staatliche Zentralstelle für Fernunterricht

Die Tabelle illustriert das ‚weite Feld‘ beruflicher Bildung mit seinem breiten Spektrum von möglichen Evaluationsgegenständen und -kontexten.

Auf der einen Seite findet sich das Bildungsunternehmen SGD und sein dauerhaftes Evaluationsgeschehen eingebettet in den Rahmen der Organisation und seines QM-Systems. Auf der anderen Seite stehen zeitlich terminierte Projekte oder Programme mit unterschiedlichen Implikationen für nachhaltige Effekte.

In der Tat findet beim Bildungsunternehmen SGD im Rahmen seines angewendeten Qualitätsmanagementsystems eine Art ‚permanente Evaluation‘ statt; daher kann hier durchaus Nachhaltigkeit konstatiert werden. Voraussetzung für diese Feststellung ist in Teilen sicher ein recht weiter Evaluationsbegriff, wie ihn der Referent, Herr Vollmer, verwendete.⁶

Bei den anderen beiden Projekten haben die Auftragnehmer nicht die nötige Kontrolle über die zukünftige Nutzung der Evaluationsergebnisse, Nachhaltigkeit ist aber auch hier angestrebt oder wünschenswert. Beim M-Learning-Projekt NetEnquiry ist die nachhaltige Nutzung des entwickelten Tools deswegen offen, da diese von der Finanzierung durch die involvierten Banken abhängt. Aber als Universität wird an dem Thema weitergearbeitet; die Forscherkontinuität dürfte immerhin gesichert sein.

Bei MobiPro-EU gestaltet sich die Nachhaltigkeit ohnehin schwierig, weil es sich hier um ein zeitlich limitiertes Bund-Sonderprogramm handelt. Man hofft auf einen nachhaltigen Lerneffekt, was die offenkundigen Defizite der Individualförderung in der ersten Förderphase betrifft. Aber: Eine Gewissheit, dass die Evaluationsergebnisse von den Stakeholdern genutzt werden, gibt es nicht – und ist vom Auftragnehmenden nur bedingt beeinflussbar.

Als interessante zu behandelte Aspekte für die Zukunft wurden von den Teilnehmenden feldspezifische und übergreifende Themen genannt:

Feldspezifische Themen:

- Sprachbildung von Menschen mit Migrationshintergrund und deren Einbindung in die berufsbezogene Bildung
- Inklusion im berufsbildenden Schulwesen
- Transferevaluation
- Einsatz von Mobile Learning in Entwicklungsländern

Übergreifende Themen:

- Umgang mit Schwierigkeiten in Evaluationsvorhaben, z.B. bei unklaren oder wechselnden Zielen (des Evaluationsgegenstandes)
- Lernen aus Evaluation
- Wie werden Ergebnisse aus Programmevaluationen in nachfolgenden Programmen berücksichtigt?
- Theorie-Praxis-Kommunikation
- Kritische Würdigung von Evaluationsansätzen

Das abschließende mündliche Feedback der Teilnehmer(innen) verriet eine sehr hohe Zufriedenheit mit der Tagung. Ein großer Dank gebührt einmal mehr dem Bundesinstitut für Berufsbildung, dieses Mal mit den Ansprechpartnern Herrn Dr. Lehmpfuhl und Frau Knöller für die großzügige und unbürokratische Unterstützung der Tagung.

6 Einige der präsentierten Praxisbeispiele der SGD könnten auch im Monitoring oder Controlling angesiedelt werden.