

## Evaluation – Evidenz – Effekte

Bericht zur 15. Jahrestagung der DeGEval – Gesellschaft für Evaluation e.V. 2012 in Potsdam

*Philipp Pohlenz*<sup>1</sup>

Die 15. DeGEval-Jahrestagung fand vom 19.-21. September 2012 an der Universität Potsdam statt. Dort wurde sie vom Zentrum für Qualitätsentwicklung in Lehre und Studium (ZfQ) ausgerichtet.<sup>2</sup> Die Tagung in Potsdam bleibt nicht zuletzt wegen eines aufgestellten Rekordes im Gedächtnis. So konnten die Organisatorinnen und Organisatoren eine bislang nicht da gewesene Nachfrage verzeichnen. Mit mehr als 350 Anmeldungen und nahezu ebenso vielen Teilnehmerinnen und Teilnehmern war die Tagung die bislang größte Jahrestagung der DeGEval.

In seinem eröffnenden Grußwort wies der Präsident der Universität Potsdam, Prof. Oliver Günther, auf die Relevanz des Tagungsthemas sowie des Tagungsprogramms für eine evaluationsbasierte und entsprechend wissenschaftsadäquate Qualitätsentwicklung hin.

Beide stellten einen augenfälligen Anschluss an die Tagungen der letzten Jahre her, die sich ihrerseits mit Evaluationsmethoden und Fragen der Partizipation an Evaluationen beschäftigten. Beides – robuste Evaluationsmethoden und Evaluationsdesigns (DeGEval-Jahrestagung 2010 in Luxemburg) einerseits sowie Beteiligungschancen der an Evaluationsprozessen Beteiligten (DeGEval-Jahrestagung 2011 in Linz) andererseits – spielt eine wichtige Rolle bei der Erzeugung von Evidenz als Grundlage für evaluationsbasierte Entwicklungsstrategien bzw. Adjustierungen bereits durchgeführter Handlungsprogramme.

### *Weiterbildungsseminare*

Einer Tradition des Tagungsformats folgend wurden der eigentlichen Tagung drei Weiterbildungsseminare vorangestellt. Diese standen unter den Titeln (1) Erfolgreiche Gestaltung und Durchführung von Online-Befragungen (*Jan Hense*, München und *Stefan Rädiker*, Marburg), (2) Selbstevaluationen anleiten, um das Evaluations-

---

1 Zentrum für Qualitätsentwicklung in Lehre und Studium – ZfQ an der Universität Potsdam

2 Dieses ist an der Universität Potsdam zuständig für die Koordination der Evaluation von Lehre und Studium sowie für Prozesse der Akkreditierung von Studiengängen bzw. – im Rahmen der Systemakkreditierung – der Akkreditierung des Qualitätsmanagementsystems der Hochschule als Ganzem.

vermögen von Organisationen zu stärken (*Wolfgang Beywl*, Köln) und (3) Effektive Datengraphiken für die Evaluation – theoretische Grundlagen und praktische Tipps (*Frederik Malter*, München). Zielgruppe waren Evaluationspraktikerinnen und -praktiker, die sich im Bereich der quantitativen Datenanalyseverfahren (Seminare 1 und 3) bzw. der Evaluationstheorie (Seminar 2) weiterbilden wollten. Ziel des ersten Seminars bestand darin, den Blick für methodische Besonderheiten online-gestützter Datenerhebungsverfahren, wie etwa Anonymität, Rücklaufquote bzw. die Datenqualität allgemein zu schärfen und so dazu beizutragen, dass die Datenqualität online durchgeführter Befragungen in den Anwendungskontexten der Teilnehmerinnen und Teilnehmer gesteigert werden kann. Das zweite Seminar adressierte Selbstevaluationen als „interventionsintegrierten“ Evaluationsansatz. Ziel war es, die professionelle Handlungssicherheit von Akteurinnen und Akteuren in Situationen zu steigern, in denen sie im Spannungsfeld eigener Evaluationsinteressen und externer Vorgaben (bspw. durch gesetzliche Vorgaben zur Evaluation im Schul- oder Hochschulbereich) agieren müssen. Im dritten Seminar beschäftigte die Teilnehmerinnen und Teilnehmer, wie sich die Nützlichkeit von Datengraphiken für die Verstehbarkeit von Texten steigern lässt. Anhand praktischer Beispiele und theoretischer Grundlegungen wurden Verfahren zur Erstellung aussagekräftiger Graphiken geübt. Wie die anderen Seminare auch stand der Bezug der Inhalte zu konkreten Anwendungskontexten der Teilnehmerinnen und Teilnehmer im Vordergrund der Gestaltung der Veranstaltung.

### *Keynote*

Gert Biesta war der Redner der einführenden „Keynote“. Biesta, der seit Beginn 2013 an der Universität Luxemburg eine Professur für Educational Theory and Policy inne hat (zum Zeitpunkt der Tagung war er Professor für Erziehungswissenschaft an der University of Stirling, UK), beschäftigt sich mit der Theorie der Erziehungswissenschaften und insbesondere mit dem Verhältnis zwischen Bildung, demokratischen Institutionen und der Zivilgesellschaft.

In seinem einleitenden Vortrag reflektierte er über den Stand der Diskussion zu evidenzbasierter politischer Praxis in verschiedenen Feldern. Die entsprechende Landschaft stellt sich durchaus divers dar. Einerseits gibt es kaum Argumente, die sich als Ablehnung einer evidenzbasierten Praxis ins Feld führen ließen, weshalb sie als Anforderung kaum in Zweifel gezogen wird. Andererseits gibt es Politikfelder, die sich gegen die Forderung evidenzbasierten Entscheidens und Handelns positionieren, nicht zuletzt, weil die Umsetzung der Forderung nach evidenzbasiertem Handeln in ein Programm entsprechender Evaluationen nicht trivial ist und insbesondere durch methodische Limitationen erschwert wird.

Biesta ging auf die Implikationen ein, die sich für die (Evaluations-)Forschung aus der erhobenen (politischen) Forderung nach evidenzbasierten Entscheidungsgrundlagen ergeben. Zu diesem Zweck diskutierte er einerseits Prämissen über die Professionalität von Forschenden sowie hinsichtlich des Verhältnisses von Wissen und Werten, die in der Diskussion zu evidenzbasierter Praxis gesetzt werden. Andererseits ging er auf den Einfluss ein, den „wissenschaftlich gesichertes Wissen“ auf

politisches Handeln und Entscheiden nimmt bzw. nehmen kann. Dazu zog er das Beispiel der Medizin sowie der Agrarökonomie heran. Beides sind Wissensbereiche, die landläufig als Musterbeispiele für ein evidenzbasiertes Zustandekommen von Steuerungsentscheidungen herangezogen werden. Biesta zeigte, dass dies weniger auf eine evidenzbasierte bzw. auf wissenschaftliches Wissen gestützte Praxis zurückzuführen ist als auf die Tatsache, dass Anforderungen evidenzbasierten Handelns in der Forschungslogik in diesen Bereichen ‚vorweggenommen‘ werden.

Aus diesem Zusammenhang leitete Biesta schließlich die Notwendigkeit ab, das Verständnis von Evidenz in der Evaluationsforschung (sowie der Forschung allgemeiner) zu überdenken und bei der Beurteilung, was als evidenzbasiertes Handeln sowie Evidenz produzierendes Forschen akzeptiert wird, eine größere Bandbreite an Perspektiven – methodisch, wie auch theoretisch – zuzulassen.

### *Treffen des Wissenschaftlichen Nachwuchses in der Evaluation*

Dank gebührt Stefan Rädiker, Heike Steckhan, Dörte Schott und Susanne Mäder, die ein Vernetzungstreffen des wissenschaftlichen Nachwuchses in der Evaluation organisiert haben. Ziel war es, Nachwuchsforscherinnen und -forscher in einen Austausch zu bringen. Relevante Themen dabei sind beispielweise die im Rahmen von Qualifikationsarbeiten adressierten Forschungsfelder. Durch das Treffen ist ein wichtiger Beitrag zur Identifikation von Kontaktmöglichkeiten geleistet worden. Eine weitere (von vielen) Initiative(n) der Gruppe ist die Sammlung von Rezensionen zu Lehrbüchern im Bereich der Evaluation. Diese sollen aus dem Kreis der in der Community versammelten Evaluatorinnen und Evaluatoren selbst entstehen und so einen direkten Bezug zur Praxis und zum Stand der Diskussion haben. Alle Interessierten sind aufgerufen und eingeladen, sich an dieser Initiative zu beteiligen.

### *Postersession*

Die parallelen Sessions der Tagung wurden durch eine Ausstellung von Postern sowie der Präsentation der Tagungssponsoren und -sponsorinnen begleitet. Die insgesamt rund 25 Poster vertraten eine enorme thematische Bandbreite und wurden im Rahmen von insgesamt zwei zeitlich dafür vorgesehenen Sitzungen von den Teilnehmerinnen und Teilnehmern diskutiert.

### *Parallele Sessions*

Die parallel angebotenen Sessions der Tagung griffen das Rahmenthema des Verhältnisses von Evaluationspraxis zu der an Evaluation gerichteten Anforderung, evidenzbasiertes Handeln zu ermöglichen, aus verschiedenen Perspektiven auf. Ein Schwerpunkt lag dabei auf methodischen Aspekten bzw. der Frage, welche Art von Evaluationsergebnissen als Evidenz gelten kann. In diesem Zusammenhang wurden die verschiedenen Verständnisse, die zu angemessenen Methoden von Wirkungsmessungen und Evidenz produzierenden Evaluationsansätzen bestehen, sowie die zu diesen Themen vorherrschenden Diskussionslogiken reflektiert.

Daneben stand aber auch die Nutzungs- oder Verwertungslogik von Evaluationen für evidenzbasiertes politisches Handeln im Vordergrund: Welche Kompetenzen werden in Organisationen für die Durchführung und/oder Nutzung von evidenzproduzierenden Evaluationen gebraucht? Welche intendierten und nicht intendierte Effekte können Evaluationen auf das Handeln von Organisationen entfalten? Nachfolgend werden die Themen einiger der Sessions, stellvertretend für alle 24 parallelen Veranstaltungen, kurz zusammengefasst. Weitere Informationen sowie die Dokumentation der Vorträge finden sich auch auf der Homepage der DeGEval (<http://www.degeval.de/veranstaltungen/jahrestagungen/potsdam-2012>).

Die Session „Evaluation von Qualitätsmanagement im Bildungsbereich. Welche Evidenz ist erreichbar?“, durchgeführt vom Arbeitskreis Berufliche Bildung (vertreten durch *Jürgen Götzhaber*, *Michael Kalman*, *Peter Jablonka*), widmete sich der entscheidenden Frage des Qualitätsmanagements im Bildungsbereich: Unter welchen Bedingungen entfalten Qualitätsmanagementsysteme die angezielte Wirkung, durch den Einsatz von Evaluation Wissen für eine rationale Steuerung von Bildungsprozessen zu sorgen? Ausgangspunkt war der Befund, dass die Wirksamkeit von Qualitätsmanagementsystemen im Bildungsbereich bislang vergleichsweise wenig erforscht ist, was vermutlich auf fehlende Zielbestimmungen sowohl hinsichtlich der Wirkungsweise dieser Systeme als auch des gemanagten Gegenstandes selbst zurückzuführen ist. Seit Langem wird als eines der schwierigsten Probleme von Evaluationen im Bildungsbereich die Uneinheitlichkeit der Qualitätsvorstellungen von Bildungsprozessen angeführt. Die Session thematisierte dementsprechend verschiedene Modelle zur Analyse von Qualitätsmanagementsystemen und die in ihnen zum Einsatz kommenden Evaluationsdesigns bezüglich ihrer Wirksamkeit für eine handlungsleitende Qualitätsbeurteilung der jeweiligen Lehr-/Lernprozesse.

Mit methodischen Herausforderungen bei der Evaluation von Programmen, die sich im Bereich der Egalisierung der beruflichen Karrierechancen von Männern und Frauen stellen, beschäftigte sich die Session „Evidenzbasierte Evaluation von Maßnahmen zur Unterstützung der Karriereplanung von Frauen und Männern“, durchgeführt vom Arbeitskreis Gender Mainstreaming (vertreten durch *Angela Wroblewski* und *Maria Gutknecht-Gmeiner*). Dabei wurden Themen angesprochen, die virulent werden, wenn von Evaluationen eine evidenzbasierte Steuerung oder Beurteilung dieser Programme erwartet wird. Probleme ergeben sich daraus, dass die Zielstellungen der zu evaluierenden Förderprogramme nicht eindeutig oder sogar widersprüchlich sind oder dass die an die Programme geknüpften Erwartungen hinsichtlich der zu erzielenden Effekte zu hoch sind. Auf der Ebene der in Evaluationen eingesetzten Untersuchungsdesigns spielen eine unsichere oder ungenügende Datenlage, die Auswahl geeigneter Kontrollgruppen sowie zeitlich verzögerte Wirkungen und damit einhergehende Probleme bei der Modellierung von Ursache-Wirkungsrelationen eine wichtige Rolle. Gegenstand der Session war es zu diskutieren, wie Evaluatorinnen und Evaluatoren die beschriebenen Probleme im Sinne einer Herstellung von Evidenz zur Beurteilung der evaluierten Programme lösen können und welche Methoden(triangulationen) in unterschiedlichen Settings die am meisten Erfolg versprechenden sind. Mit den Potenzialen von Evaluationen, Ma-

nagemententscheidungen durch die Herstellung eines umfassenden Informationspools zu erleichtern oder zu ermöglichen, beschäftigte sich die Session „Evidenzbasierte Organisations- und Programmentwicklung durch Evaluation“ des Arbeitskreises Wirtschaft (vertreten durch *Sonja Kind*, *Martin Rost* und *Barbara Schneider*). Anhand von Praxisbeispielen für evaluationsbasierte Organisationsentwicklungsprozesse wurden verschiedene Evaluationsansätze als Instrument der ganzheitlichen Informationsbeschaffung diskutiert. Dabei wurde nicht zuletzt auch auf Fragen des evaluationsmethodischen Designs eingegangen und es wurden Beispiele für Mixed-Method-Ansätze präsentiert, in denen sowohl Dokumentenanalysen, wie auch halbstandardisierte Interviews und statistische Kennzahlen, zum Einsatz kommen.

In der gemeinsamen Session des Arbeitskreises Methoden und des Arbeitskreises in Gründung Ethik (vertreten durch *Udo Kelle*, *Christian Erzberger*, *Reinhard Stockmann* und *Wolfgang Meyer*) wurden Metaevaluationen als Lösungsweg für das Problem diskutiert, dass in einem landläufigen Verständnis ausschließlich Wirkungen messende Evaluationen als „Evidenz produzierende“ Evaluationen anerkannt und für politisches Steuerungshandeln eingesetzt werden. Metaevaluationen können gegenüber Einzelevaluationen als deren „externe Validierung“ verstanden werden und durch die Akkumulation von Einzelergebnissen zur Erstellung umfassender Beurteilungsgrundlagen von politischen Programmen genutzt werden. Am Beispiel eines (Meta-)Evaluationsverfahrens der Gesellschaft für Internationale Zusammenarbeit (GIZ) wurden entsprechende Verfahren und methodische Implikationen diskutiert. Ziel der Session war es, die Relevanz von Querschnittsauswertungen von Evaluierungen und Meta-Evaluierungen für eine Evidenzbasierung von Entscheidungsfindungen zu reflektieren.

In einer Session des Vorstandes (vertreten durch *Wolfgang Böttcher*, *Christiane Kerlen* und *Thomas Weith*) wurde die Debatte innerhalb der DeGEval über die Standards der Evaluation geführt. Anlässe für eine offene Diskussion zu deren Überprüfung und ggf. ihrer Revision bestehen in der Weiterentwicklung der Evaluationspraxis oder auch in neuen Erkenntnissen und Entwicklungspfaden der Evaluationsforschung. Zudem stellen sich Fragen der Professionalisierung des Umgangs von Evaluationspraktikerinnen und -praktikern mit erhöhten Anforderungen der Auftraggeber und mit veränderten politischen Rahmenbedingungen für evidenzinformierte Prozess- oder Wirkungsevaluationen. Diese Entwicklungen legen eine Reflexion der Qualität und Effekte der bisherigen Standards nahe und nicht zuletzt durch die Standards selbst, die eine kritische Überprüfung der Standards festschreiben, wird die entsprechende Diskussion verlangt. In der Session wurde diese Herausforderung durch die Teilnehmerinnen und Teilnehmer angenommen. Eingeleitet wurde das Gespräch zu Inputs zu aktuellen Entwicklungen in den USA sowie zum Stand der Implementierung der Standards.

Neben methodischen Voraussetzungen von Evaluationen als Informationsgrundlage und Voraussetzung für evidenzbasiertes Handeln wurde auch dieses Handeln selbst zum Gegenstand der Diskussion gemacht. In der Session „Evaluation Capacity Building als Ansatz zur Förderung evidenzbasierten Handelns in Organisationen und Institutionen“ des Arbeitskreises Aus- und Weiterbildung in der Evaluation (vertreten durch *Jan Hense*, *Heike Steckhan* und *Stefan Rädiker*)

wurden unter der Bezeichnung Evaluation Capacity die Kompetenzen einer Organisation bei der Durchführung von Evaluationen sowie bei der Umsetzung von Evaluationsergebnissen diskutiert. Das entsprechende Capacity Building bezieht sich auf die Professionalisierung von Mitgliedern in Organisationen hinsichtlich des Einsatzes der entsprechenden Kompetenzen. Anhand von Praxisbeispielen aus unterschiedlichen Politikfeldern wurden Maßnahmen des Capacity Building im Evaluationsbereich sowie daraus entstehende Evaluationskulturen diskutiert.

Auch in Politikfeldern wie der Gesundheitspolitik oder auch der Regionalentwicklung wird die Frage nach der Angemessenheit von Evaluationsansätzen vor dem Hintergrund des Ziels, durch Evaluationen Evidenz als Grundlage des Steuerungshandelns herzustellen, diskutiert. Im Fokus der Session „Evidente Komplexität?! Eine besondere Herausforderung für Evaluationen in der Stadt- und Regionalentwicklung“ des Arbeitskreises Stadt- und Regionalentwicklung (vertreten durch *Manfred Rolfes* und *Brigitte Wotha*) standen demgemäß komplexe Projektstrukturen und Mehrebenenproblematiken. Diese bilden die Rahmenbedingungen für die Durchführung von Evaluationen zur Bewertung der politischen Praxis in diesem Feld. Auch hier haben Evaluatoredinnen und Evaluatoren mit methodischen Problemen wie einer langen Zeitspanne zwischen Maßnahme und (intendierter oder nicht intendierter) Wirkung zu kämpfen. Durch multikausale Beziehungs- und Wirkungsflechte zwischen dem Input und den Outputs/Outcomes einer Maßnahme werden hohe Anforderungen an die Durchführung der Evaluation gerichtet. In der Session wurden Praxisbeispiele zum Umgang mit diesen Rahmenbedingungen zur Diskussion gestellt.

### *Abschlussgespräch*

An einem die Tagung abschließenden Gespräch nahmen *Helmut Asche*, Leiter des Instituts für Unabhängige Evaluierung der Entwicklungszusammenarbeit, *Wolfgang Böttcher*, Vorstandsvorsitzender der DeGEval – Gesellschaft für Evaluation e.V., *Gerd Heyer*, ehemaliger Leiter des Referates Grundsatzfragen der Arbeitsmarktpolitik im Bundesministerium für Arbeit und Soziales, *Ute Hörrmann*, Leiterin des Aufbaustabes Fördercontrolling und Evaluation im Bundesministerium für Wirtschaft und Technologie sowie *Rupert Pichler*, Leiter der Abteilung für Forschungs- und Technologieförderung im Bundesministerium für Verkehr, Innovation und Technologie Österreich teil. Die Zusammensetzung der Runde folgte der Idee, die an der Tagung teilnehmenden Evaluationspraktikerinnen und -praktiker mit Vertreterinnen und Vertretern der „Auftraggeber(innen)seite“ von Evaluationen ins Gespräch zu bringen. Leitende Fragestellungen waren diejenigen nach der Bedeutung, die der Evaluation zukünftig als ein politisches Steuerungsinstrument zukommt, welche Art Evaluationen zukünftig nachgefragt werden, und welche Rolle dabei das Thema Evidenzbasierung spielen wird.

Einhellige Meinung der Diskutantinnen und Diskutanten war dabei, dass Evaluation aus dem Alltag politischer Steuerungspraxis nicht mehr wegzudenken und als Entscheidungsinstrument fest etabliert ist. Mit Blick auf die Evidenzbasierung politischen Handelns mittels Evaluationen bleibt festzuhalten, dass von Evidenz

produzierenden Evaluationsansätzen nicht per se nur dann zu reden ist, wenn sie einem experimentallogischen Design folgen, sondern dass durchaus auch andere Entwicklungslinien in der Evaluationsforschung im Sinne ihrer Eignung, politische Entscheidungen zu begründen, innovativ und respektiert sind.

### *Fazit*

Die während der Tagung geführten Diskussionen zeigten die Bedeutung des Themas für die derzeitige und zukünftige Evaluationspraxis in allen Politikfeldern.

Beispielsweise im Bereich der Hochschullehre ist Evaluation bzw. deren Durchführung seit Langem Gegenstand einer kontroversen Diskussion. Diese bezieht sich auf die Eignung der eingesetzten Verfahren, Evidenzen zu produzieren, die ihrerseits methodisch belastbare Rückschlüsse auf die Qualität von Lehre und Studium zulassen. Vielfach wird diese Eignung in Zweifel gezogen: Studentische Befragungen (als landläufig am meisten herangezogene Informationsquelle) seien subjektiv verzerrt, die kausale Verknüpfung von Input und Output/Outcomes der Lehre sei nicht hinreichend theoretisch (und empirisch) geklärt, Leistungsindikatoren wie die Absolventenquote seien Artefakte der Messungen und so fort. Evaluation droht so als Instrument der akademischen Selbstreflexion insgesamt an Akzeptanz zu verlieren. Insbesondere ist dies dadurch bedingt, dass die Effekte, die Evaluation für die Qualitätsentwicklung erzeugen kann, durch die bestehenden Akzeptanzprobleme geringer ausfallen, als es für eine auf rational basierten Entscheidungen beruhende Selbststeuerung der Hochschulen wünschenswert wäre.

Evaluationsergebnisse kämpfen vielfach mit Akzeptanzproblemen. Insbesondere, wenn diese für Auftraggeber oder allgemein die „Evaluationssubjekte“ wenig wünschenswert ausfallen, wird die Eignung der eingesetzten Methoden in Zweifel gezogen, valide Ergebnisse zum evaluierten Gegenstand zu produzieren. Dementsprechend ist das Anliegen der Teilnehmerinnen und Teilnehmer, sich argumentativ zu munitionieren, mehr als plausibel.

Die Tagung hat in einer großen thematischen Breite methodische und evaluationstheoretische Probleme im Umfeld der Thematik „Evaluation und Evidenz“ adressiert. Auf diese Weise konnten sich Evaluatorinnen und Evaluatoren aus verschiedenen Perspektiven auf den Stand der Diskussion der Evaluation und der Evaluationsforschung bringen.

Den lokalen Organisatorinnen und Organisatoren bleibt übrig, dem Vorstand und der Geschäftsstelle der DeGEval für die vertrauensvolle Zusammenarbeit bei der Tagungsorganisation zu danken und sich zu wünschen, dass die Teilnehmerinnen und Teilnehmer in ebenso großem Maße von der Tagung profitiert haben.