

fteval

JOURNAL

for Research and Technology Policy Evaluation



ISSUE 42 | NOVEMBER 2016

GÜNTHER LANDSTEINER &
JOHANNES GADNER

PROZESSMERKMALE
UND INSTITUTIONELLE
RAHMENBEDINGUNGEN
EINER NÜTZLICHEN
EVALUATIONSPRAXIS IM
ÖSTERREICHISCHEN
FTI-BEREICH

GÜNTHER R. BURKERT, DAVID F. J.
CAMPBELL & THORSTEN D. BARTH

DIE VERMESSENHEIT
DER VERMESSUNG:
ÜBERLEGUNGEN ZUR
UNIVERSITÄREN
GOVERNANCE IN
DEN GEISTES- UND
SOZIALWISSENSCHAFTEN

ANKE REINHARDT & ANTON
GEYER

EVERYONE'S HAPPY? THE
CASE FOR EXPECTATION
MANAGEMENT IN
COMMISSIONED
EVALUATION PROJECTS

FEDERAL MINISTRY FOR SCIENCE,
RESEARCH AND ECONOMY

Minoritenplatz 5, 1014 Vienna

Mag.^a Irene Danler

E: irene.danler@bmwfw.gv.at

Mag.^a Simone Mesner

E: simone.mesner@bmwfw.gv.at

Stubenring 1, A - 1014 Wien

Mag.^a Sabine Pohoryles-Drexel

E: sabine.pohoryles-drexel@bmwfw.gv.at

FEDERAL MINISTRY OF TRANSPORT,
INNOVATION AND TECHNOLOGY

Radetzkystraße 2, 1030 Vienna

Dr. Rupert Pichler

E: rupert.pichler@bmvit.gv.at

Dr. Mario Steyer

E: mario.steyer@bmvit.gv.at

AQ AUSTRIA

Renngasse 5, 1010 Vienna

Dr.ⁱⁿ Elisabeth Froschauer-Neuhauser

E: elisabeth.froschauer-neuhauser@aq.ac.at

Dr.ⁱⁿ Eva Maria Freiberger

E: eva.maria.freiberger@aq.ac.at

AIT - AUSTRIAN INSTITUTE OF
TECHNOLOGY

Tech Gate Vienna,

Donau-City-Straße 1, 1220 Vienna

Mag. Michael Dinges

E: michael.dinges@ait.ac.at

Mag.^a Barbara Heller-Schuh

E: barbara.heller-schuh@ait.ac.at

AWS - AUSTRIA WIRTSCHAFTSSERVICE
GESELLSCHAFT MBH

Walcherstraße 11A, 1020 Vienna

Dr. Joachim Seipelt

E: j.seipelt@awsg.at

Mag. Norbert Knoll

E: n.knoll@awsg.at

CDG - CHRISTIAN DOPPLER RESEARCH
ASSOCIATION

Boltzmanngasse 20, 1090 Vienna

DIⁱⁿ Brigitte Müller

E: mueller@cdg.ac.at

CONVELOP COOPERATIVE KNOWLEDGE
DESIGN GMBH

Bürgergasse 8-10/I, 8010 Graz

Mag. Markus Gruber

E: markus.gruber@convelop.at

FFG - AUSTRIAN RESEARCH
PROMOTION AGENCY

Sensengasse 1, 1090 Vienna

DIⁱⁿ Dr.ⁱⁿ Sabine Mayer

E: sabine.mayer@ffg.at

Mag. Leonhard Jörg

E: leonhard.joerg@ffg.at

FWF - AUSTRIAN SCIENCE FUND

Sensengasse 1, 1090 Vienna

Dr. Falk Reckling

E: falk.reckling@fwf.ac.at

Dr.ⁱⁿ Dorothea Sturn

E: dorothea.sturn@fwf.ac.at

IHS - INSTITUTE FOR ADVANCED
STUDIES

Josefstädter Straße 39, 1080 Vienna

Mag. Richard Sellner

E: richard.sellner@ihs.ac.at

JOANNEUM RESEARCH
FORSCHUNGSGESELLSCHAFT MBH

Sensengasse 1, 1090 Vienna

Mag. Wolfgang Polt

E: wolfgang.polt@joanneum.at

Mag. Jürgen Streicher

E: juergen.streicher@joanneum.at

AUSTRIAN INSTITUTE FOR SME
RESEARCH

Gusshausstrasse 8, 1040 Vienna

Dr.ⁱⁿ Sonja Sheikh

E: s.sheikh@kmuforschung.ac.at

LUDWIG BOLTZMANN GESELLSCHAFT

Nußdorfer Str. 64, 1090 Vienna

Dr. Peter Mayrhofer

E: peter.mayrhofer@lbq.ac.at

AUSTRIAN COUNCIL FOR RESEARCH
AND TECHNOLOGY DEVELOPMENT

Pestalozziggasse 4/DG 1,

1010 Vienna

Dr. Johannes Gadner

E: j.gadner@rat-fte.at

Dr.ⁱⁿ Constanze Stockhammer

E: c.stockhammer@rat-fte.at

ÖAW - AUSTRIAN ACADEMY OF
SCIENCE

Dr. Ignaz Seipel-Platz 2, 1010 Vienna

Nikolaus Göth, MSc

E: nikolaus.goeth@oeaw.ac.at

TECHNOPOLIS
AUSTRIA

Rudolfsplatz 12/11, 1010 Vienna

Mag.^a Katharina Warta

E: warta@technopolis-group.com

WIFO - AUSTRIAN INSTITUTE OF
ECONOMIC RESEARCH

Arsenal, Objekt 20, PF 91, 1103 Vienna

Dr. Jürgen Janger

E: Juergen.Janger@wifo.ac.at

Dr. Andreas Reinstaller

E: andreas.reinstaller@wifo.ac.at

WWTF - VIENNA SCIENCE AND
TECHNOLOGY FUND

Schlickgasse 3/12, 1090 Vienna

Dr. Michael Stampfer

E: michael.stampfer@wwtf.at

Dr. Michael Strassnig

E: michael.strassnig@wwtf.at

WIRTSCHAFTSAGENTUR WIEN.
EIN FONDS DER STADT WIEN

Mariahilferstraße 20, 1070 Vienna

Robert Mayer-Unterholzer

E: mayer-unterholzner@wirtschaftsagentur.at

ZSI -
CENTRE FOR SOCIAL INNOVATION

Linke Wienzeile 246, 1150 Vienna

Dr. Klaus Schuch

E: schuch@zsi.at

MMag. Alexander Degelsegger

E: degelsegger@zsi.at

INHALT

ISSUE 42 | 2016

- 2 EDITORIAL
KLAUS SCHUCH UND STEFAN PHILIPP
- 6 PROZESSMERKMALE UND INSTITUTIONELLE RAHMEN-
BEDINGUNGEN EINER NÜTZLICHEN EVALUATIONSPRAXIS
IM ÖSTERREICHISCHEN FTI-BEREICH
GÜNTHER LANDSTEINER UND JOHANNES GADNER
- 17 DIE VERMESSENHEIT DER VERMESSUNG:
ÜBERLEGUNGEN ZUR UNIVERSITÄREN GOVERNANCE IN DEN
GEISTES- UND SOZIALWISSENSCHAFTEN
GÜNTHER R. BURKERT, DAVID F. J. CAMPBELL UND THORSTEN D. BARTH
- 25 EVERYONE'S HAPPY? THE CASE FOR EXPECTATION
MANAGEMENT IN COMMISSIONED EVALUATION PROJECTS
ANKE REINHARDT AND ANTON GEYER
- 31 LAUNCHING CAREERS AND BREAKING NEW GROUND:
RESULTS FROM THE EVALUATION OF THE AUSTRIAN SCIENCE
FUND'S FWF START & WITTGENSTEIN PROGRAMMES
SARAH SEUS, SUSANNE BÜHRER-TOPCU, NICLAS MEYER, EVA HECKL, MARIO MANDL AND
KLAUS ZINÖCKER
- 35 WIFAS
EIN MODELL ZUR ABSCHÄTZUNG SOZIALER WIRKUNGEN MISSION
SORIENTIERTER FORSCHUNGSFÖRDERPROGRAMME AM BEISPIEL
DER PERSONEN- UND GÜTERMÖBILITÄT
PETER KAUFMANN, LAURENZ WOLF, ALEX SCHUBERT UND ALEXANDER NEUMANN
- 42 BOOK REVIEW:
MARK ZACHARY TAYLOR - THE POLITICS OF INNOVATION
STEFAN PHILIPP
- 44 EVALUATION WITH SOUND JUDGEMENT: CONFERENCE ON
"EVALUATION OF RESEARCH FUNDING: REVIEW, INSIGHT,
OUTLOOK"
ANKE REINHARDT, UTA SASS AND RICHARD HEIDLER

EDITORIAL

ÜBER DEN RUF NACH GESELLSCHAFTLICHER WIRKUNGMESSUNG!

KLAUS SCHUCH, GESCHÄFTSFÜHRER DER FTEVAL UND STEFAN PHILIPP, ASSISTENZ DER GESCHÄFTSFÜHRUNG DER FTEVAL

LIEBE LESERIN, LIEBER LESER,

Dass sich die Wissenschaft und Forschung seit geraumer Zeit vor allem selbstreferenziell über Indikatoren wie die Anzahl der Publikationen und Zitationen oder die Anzahl kompetitiv eingeworbener Projekte und Drittmittel bemisst und bewertet, ist hinlänglich gleichermaßen bekannt wie unbefriedigend. Sie ist zum einen zu kurz greifend und zum anderen werden dadurch von Zeit zu Zeit Verhaltenseffekte stimuliert, die im einschlägigen Jargon als „pervers“ tituliert werden. Es erscheint daher nicht überraschend, wenn insbesondere von der Seite der Politik der Ruf nach der Wirkung der Wissenschaft und Forschung auf die Gesellschaft als Abstraktum (also „at large“), hierbei die Ökonomie implizierend, erfolgt, um die „accountability“ von F&E über die Grenzen des Selbst-referentiellen zu heben. Der Anspruch ist zwar als Anspruch nicht verkehrt, aber die Heilserwartungen, die an die Messung des „societal impact at large“ von F&E-Förderungen verknüpft werden, sind (a) nicht immer sinnvoll, weil zu simplifizierend, und (b) vor allem nicht realistisch.

Bezüglich Realisierbarkeit sei hier auf die in Fachkreisen seit langem breit rezipierten Argumente wie langer Zeithorizont, Attributionsproblematik, Selbstüberschätzung unterkritischer F&E-Programme, Zielüberfrachtung etc. verwiesen. Obwohl man diese Argumente als hinlänglich bekannt voraussetzen kann, werden sie insbesondere dann von Auftraggebern (Politik, Verwaltung, Agenturen) gerne mal zur Seite geschoben, wenn diese unter vermeintlichen Rechtfertigungsdruck stehen, und daher glauben, mit der Messung des hoffentlich überzeugenden gesellschaftlichen Impacts „ihrer“ Interventionen kritische Zauderer (z.B. aus dem Finanzministerium) überzeugen zu können.

Wir möchte hier nicht falsch verstanden werden: wir halten den Ruf nach gesellschaftlicher Wirkungsmessung für legitim, aber wir sehen diesen vor allem durch eine kritische und inklusiv aufgesetzte ex-ante Bewertung von neuen, sich in Entwicklung befindlichen Programmen oder solchen, die reformiert werden, immer noch am besten verankerbar. Bedauerlich finden wir hingegen, wenn Evaluationen gesellschaftliche Wirkungsmessung versprechen (oft weil es die Nachfrager so möchten), aber bei genauerer Betrachtung Etikettenschwindel betrieben wird. Auch erachten wir es für bedauerenswert, dass die notwendige Verknüpfung von Technologiefolgenabschätzung (sowohl ex-ante als auch ex-post) mit der Evaluation von F&E-Förderprogramme so wenig umgesetzt wird.

Bezüglich Sauberkeit in der Benennung von Wirkungsmessungen möchten wir vor allem auf den in der Evaluierungsliteratur leider auch nicht immer scharf definierten Unterschied in „outcome“ und „impact“ hinweisen. Unter Impact werden da in der Regel die beabsichtigten und nicht beabsichtigten, intendierten und nicht-intendierten, positiven und

negativen Wirkungen einer Intervention auf die „society at large“ verstanden. Diese sind, sofern sie nicht ohnehin verpuffen, in der Regel komplex, langfristig und methodisch nur äußerst schwer beherrschbar. Tatsächlich gibt es einige interessante Fachbücher, die sich mit den Wirkungen von Wissenschaft und Forschung auf die Gesellschaft „at large“ beschäftigen; diese sind aber keineswegs als Evaluationen bestimmter Interventionen zu charakterisieren. Stellvertretend dafür sei diesbezüglich auf die nicht umstrittenen Arbeiten von Jared Diamond verwiesen.

Dem gegenüber versteht man unter „outcome“ einer Intervention (im deutschen auch des Öfteren als „Resultat“ bezeichnet) den unmittelbaren *Nutzen*, der sich aus der *Nutzung* des „outputs“ („Produkts“), z.B. eines F&E-Projektes oder eines F&E-Programmes, auf die Teilnehmenden bzw. deren unmittelbare Zielgruppe einer Intervention beschränkt. Wichtig sind dabei zwei Aspekte: erstens, erst die *Nutzung* eines Outputs, Produktes, oder Prozesses einer Intervention generiert ein nützliches Outcome, weshalb so viel Wert auf „dissemination“, „valorisation“ und „capitalisation“ gelegt wird. Zweitens, die Generierung eines intendierten Outcomes liegt zumindest noch teilweise innerhalb der Einflussmöglichkeiten der projekt-durchführenden Teilnehmer und ihrer internen wie externen Netzwerke (z.B. teilnehmende Dachorganisationen und andere Transmissionsriemen). Bei der Messung des Outcomes ist der unmittelbare Bezugsrahmen der Intervention entscheidend und hier macht F&E-Evaluation zur Überprüfung und Messung von Effizienz, Effektivität, Accountability, Legitimation, Projekt- und Programmlernen Sinn. Der gesellschaftliche Impact („at large“), der über die richtungsgebende Beeinflussbarkeit, geschweige denn Kontrollierbarkeit der Intervention durch die Programm- oder Projektträger hinausgeht (sofern es diesen überhaupt gibt, was von Fall zu Fall bzw. von Projekt zu Projekt und Programm zu Programm zu entscheiden ist), ist eigentlich die Domäne sozialwissenschaftlicher Untersuchung unter Zuhilfenahme von Methoden und Theorien aus der Ethnologie, der Anthropologie, der Wirtschaftswissenschaften, der Soziologie, der Philosophie und vieler anderer Disziplinen, und wird in Österreich leider wenig beachtet (und gefördert).

Was den Etikettenschwindel anlangt, plädieren wir für mehr politische Akzeptanz von Wirkungsdimensionen, die sich zwischen selbstreferentieller Wissenschaftsindikatorik und dem Heilsversprechen der gesellschaftlichen Wirkungsmessung bewegen, wie z.B. strukturierende Wirkungen einer Intervention auf eine Zielgruppe, eine Region oder ein Politikfeld, oder Wirkungen auf die Innovationsfähigkeit von Unternehmen oder auch unmittelbare wirtschaftliche Effekte bei den teilnehmenden Organisationen. Würden wir zu diesen Dimensionen mehr wissen, könnten wir in Zukunft Programme besser designen, aber selbst die Messung dieser Wir-

kungsdimensionen ist weder trivial noch state-of-art (state-of-art ist vielmehr immer noch die Darstellung von Partizipationsstrukturen und deren Mobilisierung sowie die Messung wissenschaftlichen Outputs und dessen Rezeption innerhalb des Wissenschafts- und Forschungssystems).

Die ERAC-Arbeitsgruppe *“on Measuring the Impact of EU Framework Programmes for Research and Innovation at National Level”* versucht bis Ende des Jahres eine Art Katalog mit evaluationsleitenden Fragen und dazu passenden Indikatoren und Methoden zu erstellen, der zu den hier skizzierten unterschiedlichen Wirkungsdimensionen Anhaltspunkte geben soll, um in Zukunft die Messung der europäischen Rahmenprogramme auf die nationale Ebene über verschiedene EU-Mitgliedsländer vergleichbarer zu machen. Wenn das gelingt, könnte davon wahrscheinlich auch die Wirkungsmessung von nationalen F&E-Programmen profitieren. Wenn die Arbeit scheitert, dann ist das vielleicht ein Hinweis darauf, dass wir in Zukunft die Frage nach Wirkung redimensionieren müssen.

In der aktuellen Ausgabe des fteval-Journals finden Sie einen inspirierenden Artikel von Peter Kaufmann, Laurenz Wolf, Alex Schubert und Alexander Neumann, der beschreibt, wie versucht wird, soziale Wirkungen missionsorientierter Forschungsförderprogramme am Beispiel der Personen- und Gütermobilität zu operationalisieren, womit die in diesem Editorial versuchte abstrakte Betrachtungsweise in ein konkretes praktisches Politikfeld weiter entwickelt wird.

Wie immer seien Ihnen auch die anderen Beiträge, für die wir uns sehr herzlich bei allen Autorinnen und Autoren bedanken, wärmsten empfohlen!

Ihre

Klaus Schuch

Stefan Philipp

OPENEVALUATION2016

24 + 25 NOVEMBER 2016

VIENNA | AUSTRIA

www.openevaluation2016.eu



openevaluation2016

INTERNATIONAL RTI POLICY EVALUATION CONFERENCE

It is our pleasure to inform you, that the programme of the Open Evaluation 2016 conference is now available online. We received more papers than anticipated and were able to compose a high quality programme, which tackles today's pressing issues of RTI-policy evaluation.

The key notes from **Dr. Pierre-Benoit Joly** (LISIS - Laboratoire Interdisciplinaire Sciences Innovations Sociétés), **Dr. Steven Hill** (Higher Education Funding Council for England) and **Dr. Liz Allen** (Strategic Initiatives, F1000) will be accompanied by three panel sessions, 19 paper sessions and a poster session. The parallel paper sessions will allow for presentation of three to four papers in each session, leaving enough time for discussion.

Registration is open until the 11th of November – register as soon as possible! The registration fee includes all lunches and coffee breaks during the two conference days as well as the official dinner on November 24th 2016. Please check this section of our website for recommendations about accommodation and general information about Vienna. Please note that Vienna is a busy tourist location. Thus, plan your journey early enough.

On the evening of November 23rd 2016 the Wiener Wissenschafts-, Forschungs- und Technologiefonds (WWTF - Vienna Science and Technology Fund) invites participants to a **pre-conference reception at the Vienna Globe Museum**. As places for this unique venue are limited, the registration and allocation is following the first-come-first-serve principle. If you are interested in joining us at this event please register via mail to office@fteval.at.

The conference will be accompanied by several other gatherings, such as the **EU SPRI Doctorate seminar** (21st–23rd November), the meeting of the **Science Europe Working Group** on ex-post Evaluation (23rd November), as well as a **“Science for S3” Policy Workshop** organised by IPTS/Danube-INCO.NET (25th November). It is going to be a busy week for RTI policy evaluation stakeholders and the perfect opportunity for extending and refreshing your networks!

We are looking forward to seeing you in Vienna!

AUSTRIAN PLATFORM
for Research and Technology Policy Evaluation



IFRIS
Institute For Research
and Innovation in Society

MANCHESTER
1824
The University of Manchester

PROGRAM OVERVIEW

DAY1: NOVEMBER 2⁴th

08:15	Registration and warm-up coffee			
09:00	Welcome and impulse by Katharina Warta (fteval) & Jakob Edler (MIOIR)			
09:15	Key Note by Steven Hill (HEFCE): Research impact and its assessment - lessons form the UK experience			
10:00	PANEL 1: Trends and challenges in systematic impact evaluation in science and innovation funding agencies			
11:00	Coffee break			
11:30	1 Science system evaluations	2 RTI agency evaluation based policy learning and challenges	3 Societal impact of agricultural research	20 Advancing research policy & programme evaluation practice
13:00	Lunch break			
13:50	4 European challenges and large scale policy initiatives	5 Stakeholder involvement in evaluation processes	6 Evaluation of public research organisations	7 Innovation taxonomy, designs and assessment approaches (1)
15:20	Coffee break			
15:45	8 Participation in EU programmes	9 Measuring responsible research and innovation	10 Evaluation of mission oriented research	11 Innovation taxonomy, designs and assessment approaches (2)
17:45	PANEL 2: The changing challenges of RTI evaluation in Europe			
19:00	Bus leaves to Heuriger (dinner)			

DAY2: NOVEMBER 2⁵th

09:00	Key Note by Pierre-Benoit Joly (LISIS) - Challenges of Research Impact Assessment for addressing Societal Challenges			
09:45	PANEL 3: Radical ways to select risky research and create new programmes			
10:45	Coffee break and snacks			
11:10	12 Individuals and career trajectories	13 Concepts and approaches towards openness	14 New evaluation approaches, tools and sources	15 Evaluating smart specialisation and regional innovation
13:10	Lunch break			
14:00	Launch of the SIPER Evaluation Interactive STI Evaluation Database			
14:30	Key Note by Liz Allen (F1000 & King's College London): Accelerating science, understanding its impact - the promise of open science			
15:20	Coffee break			
15:45	19 Poster Session	16 Stakeholder involvement in research production & citizen science	17 Evaluating impacts of SSH	18 Evaluation of innovation regulations, programmes and instruments
17:10	Wrap-up and goodbye by Philippe Laredo (IFRIS) & Klaus Schuch (fteval)			

PROZESSMERKMALE UND INSTITUTIONELLE RAHMENBEDINGUNGEN EINER NÜTZLICHEN EVALUATIONSPRAXIS IM ÖSTERREICHISCHEN FTI-BEREICH

GÜNTHER LANDSTEINER UND JOHANNES GADNER

Durch eine Metaevaluation wurde die Evaluationspraxis im österreichischen FTI-politischen Bereich über einen zwölfjährigen Zeitraum anhand eines ausgewählten Sets an DeGEval-Standards auf ihre Nützlichkeit hin untersucht. Zugleich wurde tatsächlich nachvollziehbar entstandener Nutzen aus einzelnen Evaluationen und aus der längerfristigen Evaluationstätigkeit insgesamt analysiert. Die Daten zeigen, dass wertvolle Schritte beim Aufbau einer Evaluationskultur unternommen wurden, da Programmevaluationen durchaus Nutzen erzeugt haben. Dennoch zeigt sich die Entstehung von Nutzen aus Programmevaluationen auch mit verschiedenen Hindernissen konfrontiert. Die Ergebnisse ermöglichen eine Diagnose zum Entwicklungsstand der FTI-spezifischen Evaluationspraxis in Österreich mit dem Fokus auf Nutzen und Nützlichkeit von Programmevaluationen. Auf Basis dieser Evidenz werden Implikationen für eine Weiterentwicklung dieser Evaluationspraxis dargelegt.

EVIDENZ ZU NUTZEN UND NÜTZLICHKEIT VON FTI-PROGRAMMEVALUATIONEN

Unter die Agenden des Rates für Forschung und Technologieentwicklung (RFTE) fällt auch die Beobachtung der Evaluationspraxis im Bereich der nationalen Forschungs-, Technologie- und Innovationspolitiken. In seinem Arbeitsprogramm für 2014 hat der RFTE ein Projekt zur Klärung der Nützlichkeit der FTI-Evaluationen in Österreich verankert. Hintergrund dafür war die Tatsache, dass Österreich in Bezug auf die Anzahl von Evaluationen im FTI-Bereich seit Jahren im europäischen Spitzenfeld liegt, der RFTE in verschiedensten Kontexten seines Arbeitsbereichs je-

doch den Eindruck gewonnen hat, dass der Nutzen dieser Evaluationen nicht immer eindeutig nachvollziehbar ist (vgl. RFTE 2013, S.9f.). Die FTI-Community hat wiederholt ihre Skepsis gegenüber stattgefundenen Evaluationen artikuliert und deren Wirksamkeit in Frage gestellt. Des Öfteren wurden Enttäuschungen gegenüber ursprünglich hohen Erwartungen in Bezug auf die Effekte von Evaluationen thematisiert. Daher sollte eine evidenzbasierte Einschätzung erbracht werden, wie die Nützlichkeit der spezifischen Evaluationspraxis unter Bezugnahme auf internationale Wissensbestände über Evaluation beurteilt werden kann, und ob und inwiefern Verbesserungspotenzial für FTI-Evaluationen in Österreich besteht, das von Akteuren des Feldes in der Beauftragung und Durchführung von Evaluationen in Zukunft aufgegriffen werden kann. Mit dem Projekt „Nützlichkeit und Nutzen der Programmevaluationen im Bereich der österreichischen FTI-Politik. Metaevaluation der Programmevaluationen 2003-2014“ hat der RFTE die entsprechende Studie in Auftrag gegeben.

Ziel dieser Studie war es, Nutzungs- und Wirksamkeitsbedingungen von FTI-Evaluationen im unabhängigen Blick von außen anhand konkreter und eingegrenzter Fragestellungen zu untersuchen und in transparenter, kriteriengestützter Weise zu bewerten. Ein unabhängiger externer Metaevaluator, der im Beobachtungszeitraum nicht an österreichischen FTI-Evaluationen mitgewirkt hat, aber dennoch über gute Kenntnisse von System und Spezifika des Evaluierens im Politikfeld FTI verfügt, wurde mit dem erstgenannten Autor dieses Artikels gefunden. Das Design der Studie wurde sodann vom Metaevaluator in mehreren Schritten im Dialog mit dem Auftraggeber entwickelt.¹ In der Designphase des Projekts wurden zwei besonders ausgewiesene Evaluationsexperten des deutschsprachigen Raums im Rahmen eines eintägigen Workshops² einbezogen. Um einen ebenso aussagekräftigen wie relativ kurzfristig erzielbaren Zugriff auf die vielfältige Evaluationspraxis zu er-

1 Es wurde Wert darauf gelegt, die Vorteile einer schrittweisen und dialogischen Projektdefinition wahrnehmen zu können, die in einem klassischen, auf TORs basierten Ausschreibungsverfahren bekannten Einschränkungen unterliegt.

2 Prof. Dr. Wolfgang Beywl (Pädagogische Hochschule FHNW) und Prof. Dr. Thomas Widmer (Universität Zürich) haben die US-amerikanischen *Standards for Program Evaluation* ins Deutsche übertragen und mit kommentierenden Texten zu deren Verständnis und Einsatz versehen. Sie haben zentrale Rollen bei der Erarbeitung und Einführung der heutigen Evaluationsstandards in Deutschland und der Schweiz in den jeweiligen Evaluationsgesellschaften DeGEval und SEVAL inne gehabt, und haben international zu Fragen der Evaluationstandards und des Evaluationsnutzens publiziert. Thomas Widmer hat unter anderem bereits in den 1990er-Jahren eine richtungsweisende Metaevaluation anhand der Evaluationsstandards durchgeführt. Wolfgang Beywl war langjähriger Leiter des Studiengangs für Weiterbildung in der Evaluation an der Universität Bern. Beiden sei an dieser Stelle nochmals herzlich für das Einbringen ihrer Expertisen und ihr Engagement gedankt.

möglichen, wurde ein zwölfjähriger Beobachtungszeitraum von 2003 bis 2014 gewählt und auf externe Programmevaluationen im Auftrag zentraler Akteure der österreichischen Bundesverwaltung fokussiert.

Auf Bundesebene sind in Österreich drei ministerielle Verwaltungsbereiche, von denen zwei nunmehr in einem Ministerium zusammengelegt sind, sowie drei Agenturen mit unterschiedlichem Autonomiestatus für FTI-Agenden verantwortlich. Ihnen kommt auf Basis von Rechtsgrundlagen und Mittelzuordnungen das größte Gewicht in der FTI-Steuerung und -förderung zu, während im Gesamtbild eine Vielzahl an operativen Akteuren und verantwortlichen Organisationseinheiten sowie ein „Programmschub“ aus derzeit 136 Forschungsförderungsprogrammen von Bund und Ländern vorliegt (vgl. Rechnungshof 2016). Dass diese strukturelle Situation der FTI-Governance Herausforderungen für die korrespondierende Evaluationspraxis impliziert, ist wiederholt vermerkt worden (Pichler 2009, Zinöcker/Dinges 2009, RFTE 2013).

Um zu einer systematisch-gesamthaften Auffassung der Evaluationspraxis im FTI-Bereich unter dem spezifischen Gesichtspunkt der Entstehung von Evaluationsnutzen gelangen zu können, wurde die Metaevaluation anhand eines ausgewählten Sets von DeGEval-Standards durchgeführt und um Schritte der Nutzungsforschung zur Evaluation ergänzt. Nachdem bereits Standards für Programmevaluation auf FTI-Programme einer einzelnen Agentur angewendet wurden (Good 2006) und eine Untersuchung über Evaluationsnutzen aus mehr als 100 europäischen FTI-Evaluationen aus verschiedenen Ländern vorgelegt wurde (Edler et al. 2011), liegt mit der Metaevaluation nun die erste Untersuchung vor, die eine nationale Evaluationspraxis im FTI-Bereich gesamthaft-systematisch sowohl anhand von internationalen Evaluationsstandards als auch anhand fachwissenschaftlich eingeführter Kategorien der Nutzungsforschung zu Evaluation analysiert, unter Einbeziehung von Sichtweisen von EvaluatorenInnen und von AuftraggeberInnen.

Im vorliegenden Artikel berichten der Metaevaluator und der Vertreter der Auftraggeber-Institution, der für die Metaevaluation operativ zuständig war, gemeinsam über die in zahlreichen Schritten immer wieder abgestimmte Untersuchung und über deren Ergebnisse, die aus beider Sicht für eine Charakterisierung der spezifischen Evaluationspraxis als wesentlich gelten können und Hinweise für eine Weiterentwicklung der Evaluationspraxis bieten.³ Der Artikel gibt zunächst einen Überblick über die methodische Vorgehensweise der Untersuchung. Im Anschluss werden zentrale Ergebnisse präsentiert. Abschließend werden essentielle Ansatzpunkte für eine künftige Steigerung von Nützlichkeit und Nutzen von FTI-Programmevaluationen in Österreich benannt. Eine vollständige Darstellung der eingesetzten Methodik vor dem Hintergrund internationaler Fachdiskussionen zu Metaevaluation und zur Entstehung von Evaluationsnutzen, aller Teilergebnisse und der detaillierten Schlussfolgerungen und Empfehlungen findet sich im zweibändigen Bericht (Landsteiner 2015), der auf der Homepage des RFTE publiziert ist.⁴

METAEVALUATION IN KOMBINATION MIT INSTRUMENTEN DER NUTZUNGSFORSCHUNG

Metaevaluationen sind systematische Analysen von einer oder mehreren Evaluationen, die dazu dienen, Wert und Güte dieser Evaluationen zu beurteilen. Der Metaevaluation der Programmevaluationen im FTI-Bereich ging es also nicht um deren inhaltlichen Ergebnisse, sondern um Gestaltungsweisen dieser Evaluationen, die als essentielle Voraussetzungen und Bausteine der Nützlichkeit von Programmevaluationen zu erachten sind. In betreffenden Evaluationseigenschaften verkörpert sich das Nutzungspotenzial, das Evaluationen in sich tragen, und das als Grundlage für faktische Nutzungsweisen als solches auf Stärken und Schwächen hin analysiert werden kann.

Mit den *Standards für Evaluation* der deutsch-österreichischen Evaluationsgesellschaft DeGEval (DeGEval 2008) liegt ein hervorragender Bezugspunkt für eine neutrale und breit abgesicherte Thematisierung von Qualitätsaspekten einer Programmevaluation vor, die explizit mit deren Nützlichkeit verbunden sind. Die *Standards für Evaluation* verstehen sich als Anleitung für zielgerichtete professionelle Evaluation in allen Politikfeldern und als fachliches Dialoginstrument für einen Austausch über die Qualität von professionellen Evaluationen. Die DeGEval-Standards zählen dabei zu einer Standardfamilie, die auch die US-amerikanischen *Program Evaluation Standards* (Joint Committee & Sanders 2006 [1994]) und die von der schweizerischen SEVAL verwendeten Standards (Widmer/Landert/Bachmann 2001) umfasst. Die Wahl dieses international hoch relevanten, mit der übergreifenden Gesamtentwicklung von Evaluationstheorie und -forschung verbundenen Bezugspunktes mit seinen detaillierten Erläuterungen erfolgte in bewusster Abhebung von den Standards der Plattform fteval, die auf dem Feld der österreichischen FTI-Evaluation entstanden sind. 11 der insgesamt 25 DeGEval-Standards wurden unter Abwägung von theoretischen Überlegungen und in der Fachliteratur berichteten Erfahrungen zu Standard-basierten Metaevaluationen ausgewählt.⁵

Die DeGEval-Standards enthalten eine Gruppe von explizit als „Nützlichkeitsstandards“ bezeichneten Standards, die sämtlich herangezogen wurden, soweit Informationen zu ihrer Erfüllung aus Evaluationsberichten überhaupt erwartet werden konnten. Eine wesentliche Debatte der Evaluationstheorie und Nutzungsforschung kreist um die Frage, inwiefern Evaluationen über diese Gesichtspunkte hinaus Qualität auf allen Ebenen mitbringen sollen, um eine gute Grundlage für ihre Nutzung zu bilden. Neben Nützlichkeitsstandards wurden deshalb auch einige wei-

3 Die Position des Vertreters des RFTE versteht sich gemäß der Struktur der Institution vorbehaltlich aller Entscheidungen, die die amtierenden Räte des RFTE gefällt haben oder fällen werden.

4 http://www.rat-fte.at/tl_files/uploads/Studien/Metaevaluation-FTI-Endbericht-final-web.pdf,
http://www.rat-fte.at/tl_files/uploads/Studien/Metaevaluation-FTI-Anhaenge-web.pdf.

5 In den *fteval-Standards* werden zwar Fragen der Nutzung von Evaluationen angesprochen und die Position, die für Evaluationen im institutionellen Arrangement angestrebt wird. Eine eingehende Prüfung der *fteval-Standards* ergibt allerdings klare Einschränkungen ihrer Verwendbarkeit für eine Studie, die Fragen der Nützlichkeit von Evaluationen so gut wie möglich ausloten will. Zum Einen waren Nutzungsaspekte in den früheren Fassungen, wie sie während nahezu des gesamten Beobachtungszeitraums der vorliegenden Studie vorlagen, nur abrisshaft und nicht auf der Differenzierungsebene der *DeGEval-Standards* angesprochen. Zum Zweiten zeigen die *fteval-Standards* in der heutigen Fassung bei der vergleichsweise vertieften Einlassung auf Nutzungsaspekte eine klare Annäherung an die *DeGEval-Standards*. Sie bleiben jedoch auch dabei hinter deren Differenzierungsebene zurück, da sie weder die Erläuterungen noch die Literaturhinweise zu verfügbaren Forschungsgrundlagen enthalten, die die *DeGEval-Standards* vermitteln.

tere Standards ausgewählt, um einem holistischen Qualitätsverständnis von Evaluation besser gerecht zu werden. Die historisch hinter den 2001 eingeführten DeGEval-Standards stehenden *Program Evaluation Standards* bieten eine noch differenziertere Auffächerung von Ansatzpunkten für das Erreichen einer Evaluation, die beste Voraussetzungen für eine Entfaltung von Nutzen mitbringt. Sie bilden prinzipiell einen weiterführenden Interpretationshintergrund für die kürzer gefassten DeGEval-Standards (vgl. u.a. Beywl/Taut 2000), und dort vorfindliche Hinweise wurden im angewendeten rekursiven Verfahren aus dem Material heraus auch benötigt.⁶

Erster Ansatzpunkt zur Klärung der Nützlichkeit österreichischer Programmevaluationen im FTI-Bereich waren vorliegende Evaluationsberichte, die einer qualitativen Sekundäranalyse anhand der herangezogenen DeGEval-Standards unterzogen wurden. Um bei aller berechtigten Unterschiedlichkeit einzelner Evaluationen ein systematisch-gesamthafes Verständnis der österreichischen FTI-Evaluationspraxis zu ermöglichen, erfolgte eine Auswahl von 20 Berichten durch eine proportional geschichtete Zufallsstichprobe, der zentrale Merkmale der Evaluationspraxis zugrunde gelegt wurden. Die Berichte wurden vom Meaevaluator einer differenzierten und rekursiven Sekundäranalyse unterzogen. Zusätzlich zu dieser qualitativen Analyse wurde gleichzeitig zu Überwachungszwecken ein Grading der Standard-Erfüllung auf einer fünfstufigen Skala vom Meaevaluator vorgenommen.⁷

Evaluationsberichte stellen Produkte der hinter ihnen stehenden Evaluationsprozesse dar, die in Bezug auf angestrebte Erkenntnisse genauso wie auf Rahmenbedingungen des Evaluierens konfiguriert werden. Eine Metaevaluation wird daher bestrebt sein, nach Möglichkeit weitere Informationen heranzuziehen, um die Evaluationsprodukte eingehender als Ergebnisse von Prozessen verstehen zu können, die mit der Auslösung und Planung von Evaluationen beginnen und in einen konkreten Kontext eingebettet sind (Stufflebeam 2001, Cooksy/Caracelli 2005). Ergänzende Informationen über Gesichtspunkte der eingesetzten Standards, die typischerweise Berichten nicht zu entnehmen sind, und zu den hinter den Berichten stehenden Evaluationsprozessen wurden durch semistrukturierte Interviews im Auftraggeber-Bereich und durch eine strukturierte Online-Befragung von EvaluatorInnen eingeholt. In der EvaluatorInnen-Befragung wurde eine Vollerhebung unter allen im Beobachtungszeitraum tätigen FTI-EvaluatorInnen angestrebt. 37 EvaluatorInnen, die im Beobachtungszeitraum meist mehrfach tätig waren, haben die Umfrage beantwortet⁸. In 15 Interviews mit AuftraggeberInnen und HauptadressatInnen der Evaluationen wurde auf eine gleichmäßige Abdeckung der relevanten Systemstellen geachtet und auch die Geschichte der Plattform fteval reflektiert.

Wird unter der Perspektive der Nützlichkeit von Evaluationen thematisiert, welche Voraussetzungen für die Entstehung von Nutzen ge-

Set der herangezogenen DeGEval-Standards

N1 Identifizierung der Beteiligten und Betroffenen: Die am Evaluationsgegenstand beteiligten oder von ihm betroffenen Personen bzw. Personengruppen sollen identifiziert werden, damit deren Interessen geklärt und so weit wie möglich bei der Anlage der Evaluation berücksichtigt werden können.

N2 Klärung der Evaluationszwecke: Es soll deutlich bestimmt sein, welche Zwecke mit der Evaluation verfolgt werden, so dass die Beteiligten und Betroffenen Position dazu beziehen können und das Evaluationsteam einen klaren Arbeitsauftrag verfolgen kann.

N4 Auswahl und Umfang der Informationen: Auswahl und Umfang der erfassten Informationen sollen die Behandlung der zu untersuchenden Fragestellungen zum Evaluationsgegenstand ermöglichen und gleichzeitig den Informationsbedarf des Auftraggebers und anderer Adressaten und Adressatinnen berücksichtigen.

N5 Transparenz von Werten: Die Perspektiven und Annahmen der Beteiligten und Betroffenen, auf denen die Evaluation und die Interpretation der Ergebnisse beruhen, sollen so beschrieben werden, dass die Grundlagen der Bewertungen klar ersichtlich sind.

N6 Vollständigkeit und Klarheit der Berichterstattung: Evaluationsberichte sollen alle wesentlichen Informationen zur Verfügung stellen, leicht zu verstehen und nachvollziehbar sein.

N8 Nutzung und Nutzen der Evaluation: Planung, Durchführung und Berichterstattung einer Evaluation sollen die Beteiligten und Betroffenen dazu ermuntern, die Evaluation aufmerksam zur Kenntnis zu nehmen und ihre Ergebnisse zu nutzen.

F3 Vollständige und faire Überprüfung: Evaluationen sollen die Stärken und die Schwächen des Evaluationsgegenstandes möglichst vollständig und fair überprüfen und darstellen, so dass die Stärken weiter ausgebaut und die Schwachpunkte behandelt werden können.

F5 Offenlegung der Ergebnisse: Die Evaluationsergebnisse sollen allen Beteiligten und Betroffenen soweit wie möglich zugänglich gemacht werden.

G2 Kontextanalyse: Der Kontext des Evaluationsgegenstandes soll ausreichend detailliert untersucht und analysiert werden.

G3 Beschreibung von Zwecken und Vorgehen: Gegenstand, Zwecke, Fragestellungen und Vorgehen der Evaluation, einschließlich der angewandten Methoden, sollen genau dokumentiert und beschrieben werden, so dass sie identifiziert und eingeschätzt werden können.

G8 Begründete Schlussfolgerungen: Die in einer Evaluation gezogenen Folgerungen sollen ausdrücklich begründet werden, damit die Adressaten und Adressatinnen diese einschätzen können.

Abbildung 1: Set der ausgewählten DeGEval-Standards

6 Eine Zusammenstellung der 11 herangezogenen DeGEval-Standards im vollen Wortlaut und der Passagen aus den *Program Evaluation Standards*, die in der durchgeführten Metaevaluation im Umgang mit dem Material schlagend wurden, findet sich im Anhang zur Metaevaluation (Landsteiner 2015).

7 Die vergebenen Gradings sind in den Factsheets für jeden analysierten Evaluationsbericht nachvollziehbar gemacht, die im Anhang zum Metaevaluationsbericht (Landsteiner 2015) enthalten sind und die qualitativen Analyseergebnisse jeweils gemeinsam mit dem vergebenen Grading darstellen.

8 Wieviele EvaluatorInnen im österreichischen FTI-Bereich im Beobachtungszeitraum tatsächlich tätig waren, ist nicht bekannt und konnte auch im Zuge dieser Studie nicht geklärt werden. Verschiedentlich wurde in Diskussionen des Feldes sowie in einigen im Zuge dieser Studie geführten Interviews darauf hingewiesen, dass eine nicht näher bestimmbare Quote an nicht publizierten Evaluationen und relevanten, aber nicht explizit als Evaluation bezeichneten Studien existieren dürfte. In der durchgeführten Online-Befragung wurden insgesamt 273 EvaluatorInnen und MitarbeiterInnen von relevanten Instituten in Österreich und im deutsch- und englischsprachigen Ausland kontaktiert, die für die Durchführung einer FTI-Programmevaluation im Beobachtungszeitraum bekannt waren oder grundsätzlich in Frage kamen. Die auf der Homepage der Plattform fteval verfügbare Datenbasis wurde dabei einbezogen. Die beantworteten Fragebögen kamen sodann durch Selbsteinstufung der EvaluatorInnen in Kombination mit ihrer Antwortbereitschaft zustande. Eine Rücklaufquote im fachlichen Sinn kann angesichts fehlender gesicherter Information über die Grundgesamtheit nicht ausgewiesen werden.

schaffen wurden, so ist damit über faktisch entstandenen Nutzen noch nichts gesagt. Die beiden ergänzenden Erhebungsinstrumente der EvaluatordInnen-Befragung und der Auftraggeber-Interviews wurden daher herangezogen, um auch tatsächlich entstandenen Nutzen zu beleuchten und Einflussfaktoren zu identifizieren, die in den Wahrnehmungen der EvaluatordInnen und der AuftraggeberInnen für Entstehung oder Ausbleiben von Nutzen ausschlaggebend waren. Mit dieser Anlage und Vorgehensweise ermöglichte die im Kern als Metaevaluation definierte Studie eine gleichzeitige Anwendung von zwei Perspektiven auf die Nutzungsproblematik, die miteinander in der allgemeinen Evaluationstheorie aufs Engste verbunden sind und seit den 1970er-Jahren in einem Wechselspiel von empirischen Studien, einschlägigen Theoriearbeiten und Standard-Formulierungen einen kontinuierlichen Kreislauf der Ideen bilden. Die Auftraggeber-Interviews wurden ferner dafür genutzt, zusätzliche Informationen über die Rolle von evaluativer Information im FTI-Governancesystem (vgl. z.B. Rist/Stame 2006) einzuholen.

Alle angelegten Blickwinkel tragen von vornherein auch die Perspektive in sich, dass es unumgänglich ist, auch Begrenzungen der Möglichkeiten einer noch so optimalen Evaluation hinsichtlich ihres Einflusses auf Politikgestaltung anzuerkennen. So wertvoll die Standards bei der Erzielung guter Evaluationsprozesse und -produkte sind, indem sie eine gemeinsame Qualitätsperspektive und ein gemeinsames Verständnis von diesbezüglichen Anforderungen unter EvaluatordInnen und AuftraggeberInnen verankern, kann ihre Existenz und Beachtung doch nicht als solche für die Qualität von politisch-administrativen Entscheidungen eintreten, die unter stärkerer oder schwächerer Bezugnahme auf Evaluationen gefällt werden. Neuere konzeptuelle Arbeiten zur Entstehung von Evaluationsnutzen weisen darauf hin, dass sich Evaluationsnutzen in vielschichtiger Weise auf unterschiedlichen Ebenen entfalten kann („evaluation influence“), ohne dabei mit dem Vorliegen einer ersten Reaktion als abgeschlossen gelten zu können (Kirkhart 2000, Mark/Henry 2004).

ENTSTANDENER EVALUATIONSNUTZEN

Bisherige Programmevaluationen haben durchaus Nutzen generiert, wie die hochgradig miteinander konsistenten Auskünfte der AuftraggeberInnen und der EvaluatordInnen zeigen. Dabei stehen Nutzungsweisen im Vordergrund, die von der Evaluationsforschung als instrumenteller und konzeptueller Nutzen von Evaluation bezeichnet werden (vgl. etwa Alkin/Taut 2003, Patton 1997, Shulha & Cousins 1997, Weiss 1998).⁹ Seitens der Auftraggeber-Institutionen bzw. HauptadressatInnen der Evaluationen wurden Entscheidungen über Programme getroffen und neue Sichtweisen gewonnen, die zur Nachschärfung von Programmen oder zur Bereinigung von Zielkatalogen geführt haben. Immer wieder wurden Lerneffekte unter Programmeigentümern und mit Programmumsetzungen befassten Akteuren erzeugt, sodass durch Programmevaluationen Themen verankert und Annahmen über Funktionsweisen von Teilen des FTI-Systems und der darauf gerichteten Steuerungs- und Anreizsysteme

verändert wurden. Entscheidungen über evaluierte Programme betrafen vor allem Adjustierungen, sowohl hinsichtlich von Umsetzungsaspekten in den betrauten Agenturen, als auch in formalen Programmdokumenten für eine nachfolgende Programmphase. Hinzu kommen konzeptuelle Einsichten über Merkmale von FTI-Segmenten ebenso wie über das Management von Programmen, die durchaus auch da eintraten, wo nicht für außenstehende BeobachterInnen leicht erkennbare Entscheidungen gefällt wurden. Interessant erscheint speziell, dass auch Evaluationen von Programmen, die in der Folge nicht weitergeführt wurden, systemisch wertvolle Einsichten und Lerneffekte erbracht haben, sodass von schrittweisen Beiträgen zur kumulativem Entstehung von konzeptuellem Nutzen gesprochen werden kann. Beispiele für weiterreichende Evaluationfolgen sind unter anderem an neueren Programmgenerationen zur Wissenschafts-Wirtschafts-Kooperation oder an Anpassungen der Maßnahmenfamilie rund um den Innovationsscheck nachvollziehbar.

Zu den beobachtbaren Evaluationsfolgen zählen auch organisatorische Anpassungen im Auftraggeberbereich, die den Umgang mit Evaluationen und deren Ergebnissen unterstützen. An allen Systemstellen wurden Zuständigkeiten für die Führung der Vergabe-, Durchführungs- und Abnahmeprozesse geschaffen und Prozesse für die Weitergabe von Evaluationsergebnissen in der Hierarchie etabliert. Insbesondere haben zwei Agenturen in jüngster Zeit Prozesse eingerichtet, die die Evaluationsfunktion innerhalb der Organisation klar verankern, sie mit allen Organisationsteilen einschließlich der Entscheidungsgremien verknüpfen, und die gezielte Auseinandersetzung mit Evaluationsergebnissen

Typologie des Evaluationsnutzens (nach Patton 1997, Shulha & Cousins 1997, Weiss 1998)

Instrumenteller Nutzen: Evaluationsergebnisse und/oder Empfehlungen werden direkt zur Entscheidungsfindung über das evaluierte Programm genutzt.

Konzeptueller Nutzen: Evaluationsergebnisse helfen Programmbeteiligten, über den Evaluationsgegenstand zu lernen und neue Sichtweisen darauf zu entwickeln.

Symbolischer Nutzen: Das Vorliegen eines Evaluationsberichts oder die Tatsache, dass überhaupt evaluiert wird, dient zur Rechtfertigung bereits zuvor getroffener Entscheidungen, oder zur formalen Untermauerung, dass mit dem Programm rational umgegangen wird, unter Umständen um andere Akteure in der politischen Sphäre vom Programm zu überzeugen.

Aufklärung: Evaluationsergebnisse reichern das verfügbare Wissen an, das von Akteuren im Umfeld des Evaluationsgegenstands genutzt werden kann. Die Anreicherung verfügbaren Wissen kommt darüber hinaus auch Personen bzw. gesellschaftlichen Sphären zugute, die nicht unmittelbar am Programm beteiligt waren oder sind.

Prozessnutzen: Kognitive, verhaltensförmige oder organisatorische Veränderungen treten bereits im Laufe des Evaluationsprozesses ein, bereits vor Vorliegen der Ergebnisse bzw. unabhängig von diesen.

Abbildung 2: Typologie des Evaluationsnutzens

auf Strategieebene auch über das jeweils evaluierte Programm hinaus ermöglichen.

Die Programmevaluationen können in ihrer Rolle als FTI-politische Informationsmittel als konkurrenzlos gelten, wenn sie auch oft nicht die alleinigen Grundlagen FTI-politischer Entscheidungen über Einsätze und Mittelzuweisungen darstellen. Es handelt sich bei Umsetzungen von Erkenntnissen aus Evaluationen weder um Automatismen noch um eindimensionale Ereignisse. Alle Auftraggeber-Institutionen verstehen Evaluationsergebnisse (Datenlagen, Schlussfolgerungen und Empfehlungen) als Motivationsgrundlage und Ansatzpunkt für Reflexionen, die ihren Sinn gerade auch daraus gewinnen, dass sie nicht „sklavisch“ umgesetzt werden. Freilich dienten bisherige Programmevaluationen im FTI-politischen System auch durchaus dazu, symbolischen Nutzen zu generieren. Zweifellos ist von der grundsätzlichen Notwendigkeit einer fortlaufenden Überzeugungsarbeit in einem von differenzierten Principal-Agent-Verhältnissen, Hierarchien und Kleinteiligkeit geprägten System auszugehen. International wiederholt festgestellte Phänomene der „Beharrungskraft“ von administrativen Stellen sind nicht auszuschließen. Zudem fällt hier auch die Art der Verankerung der Evaluationsfunktion im rechtlich-institutionellen Rahmen der Bundesverwaltung ins Auge, auf die weiter unten nochmals zurückzukommen sein wird.

In zweierlei Hinsicht ist bislang Nutzen aus den Programmevaluationen zwar nicht völlig ausgeblieben, aber doch als eher gering einzuschätzen: Unter dem Nutzentyp „Aufklärung“ lassen sich Effekte der Verbreitungen und Diskussion von Evaluationsergebnissen im Governance-System verstehen, die über die enge Sphäre der unmittelbar für das evaluierte Programm Zuständigen hinausreichen, sowie solche, die in breiteren sozioökonomischen Sphären, Zielgruppen der Programme und interessierten Öffentlichkeiten eintreten. Beides stellt sich für die Vergangenheit als vom Einzelfall abhängig, unsystematisch und akzidentell dar. Das grundsätzliche Potenzial, von Programmevaluationen auftraggeberseitig bereits während der Evaluationsdurchführung und vor dem Vorliegen von Evaluationsergebnissen bzw. unabhängig von solchen unmittelbar zu profitieren („Prozessnutzen“), scheint bislang nur wenig ausgeschöpft worden zu sein.

QUALITÄTSDIMENSIONEN DER NÜTZLICHKEIT ALS GRUNDLAGEN KONSTRUKTIVER NUTZUNGSPROZESSE

Die analysierten Evaluationsberichte entsprechen den herangezogenen DeGEval-Standards auf einem im Großen und Ganzen mittleren Niveau, wie die Auswertung der vergebenen Gradings in Abbildung 3 zeigt. Der Schwerpunkt der Verteilungen zu allen Standards liegt im mittleren Bereich einer neutralen Einschätzung, dass Gesichtspunkte des Standards soweit erfüllt sind, dass weder eine klare Schwäche noch eine klare Stärke besteht. Die Einstufungen im mittleren Bereich kommen in etlichen Fällen auch dadurch zustande, dass feststellbare Stärken durch gleichzeitig feststellbare Schwächen aufgewogen werden. Verbesserungspotenzial ist demnach oftmals vorhanden, wenn es um bestmögliche Programmevaluationen geht, die hohe Nützlichkeit erzielen und das im „Unternehmen Programmevaluation“ angelegte Potenzial bestmöglich ausschöpfen. Eine sehr gute Erfüllung eines der 11 herangezo-

genen Standards konnte nur in einigen wenigen Fällen attestiert werden (für 5 Evaluationsberichte von 20 im Bereich von 5 verschiedenen Standards). Ebenso selten ist zugleich eine völlige Nichterfüllung eines der Standards, die auch in den letzten Jahren nicht mehr auftritt (insgesamt 5 Evaluationsberichte im Bereich von 3 der 11 herangezogenen Standards). Während zu allen herangezogenen Standards grundsätzlich noch Verbesserungspotenzial besteht, sollen hier einige hoch relevante Gesichtspunkte aufgegriffen werden.

Die analysierten Programmevaluationen waren mit Ausnahme einer ex post-Evaluation Zwischenevaluationen oder Teile von Begleitevaluationen. In den meisten Fällen waren sie sehr breit angelegte „Multi-Purpose Evaluationen“ und verfolgten in Einem Klärungen zur Relevanz, Effektivität und Wirkung der Programme. Dabei wurden Outputs, Outcomes und erste Wirkungen der Programme untersucht, sodass durchaus Erkenntnisse über die Programme erzielt wurden. Allerdings war die Beobachtbarkeit von Programmwirkungen wegen der früh gewählten Evaluationszeitpunkte fast immer deutlich eingeschränkt. Aber auch darüber hinaus kann in vielen Fällen nicht von einer umfassenden und gründlichen Aufarbeitung der Programme gesprochen werden.

Die umfangreichen Evaluationsvorhaben wurden anhand von erhältlichen Monitoringdaten und zusätzlichen, innerhalb der einzelnen Programmevaluationen jeweils erhobenen Primärdaten durchgeführt, die allerdings des Öfteren doch keine analytisch konsequente Ausleuchtung aller Programmkomponenten zuließen. Die Gesamtebenen aller relevanten Programmoutputs und outcomes, die schrittweise hin zur Erreichung der Programmziele führen sollten, und insbesondere die Verbindungen zwischen diesen Ebenen, wurden nur mit teils deutlichen Einschränkungen greifbar gemacht (Standard N4). Aufgrund dieser Ausschnitthaftigkeit haben die meisten der untersuchten Programmevaluationen letztlich doch den Charakter einer sogenannten „black box“-Evaluation, durch die die genaue Art und Weise, wie ein Programm die intendierten Wirkungen erzielt bzw. an der Erzielung dieser Wirkungen gehindert ist, nicht oder zumindest nicht vollständig erfasst wird. Zugleich ist eine tendenzielle Schematisierung der Evaluationsanlagen und Vorgehensweisen mit ihren Methodensets für einen nicht unbedeutlichen Teil der analysierten Evaluationen festzustellen. Lediglich zwei der analysierten Evaluationen haben innovative Methoden zum Einsatz gebracht, um eine spezifische Analysestrategie umsetzen zu können.

Häufig zeigten sich auch Schwierigkeiten mit einer konzipierten Gliederung von Programmkomponenten und Umsetzungsschritten zu Zielen unterschiedlicher logisch-hierarchischer Stellung (unmittelbare, intermediäre und übergeordnete Programmziele) sowie hinsichtlich der Art der Erreichung von direkten und indirekten Zielgruppen. Mehrfach wurden Fragestellungen behandelt, die für ein tieferes Verständnis des evaluierten Programms als nicht zentral zu erachten sind. Einige Evaluationsberichte tragen Züge eines „evaluability assessment“, in dem die Bedingungen für eine zielführende Evaluation des Programms überhaupt erst geklärt werden.

Wie essentiell die erbrachten Ergebnisse in Hinblick auf die Gesamtlogiken der evaluierten Programme jeweils tatsächlich sind (Standards G3, N4), ist anhand der Evaluationsberichte wegen ihrer Gestaltungsweise oft nicht gut beantwortbar. Bei allen vorhandenen Methodendarstellungen vernachlässigen die Evaluationsberichte oft doch Gesichtspunkte des konzeptuellen Bezugs zwischen Programmanlage und analytischer Vorgehensweise. In ähnlicher Weise erscheint die Frage virulent, wie Programmen insgesamt Wert zugemessen wurde (Standard N5). Hier offenbart sich ein „blinder Fleck“ eines stark datenorientierten und oft-

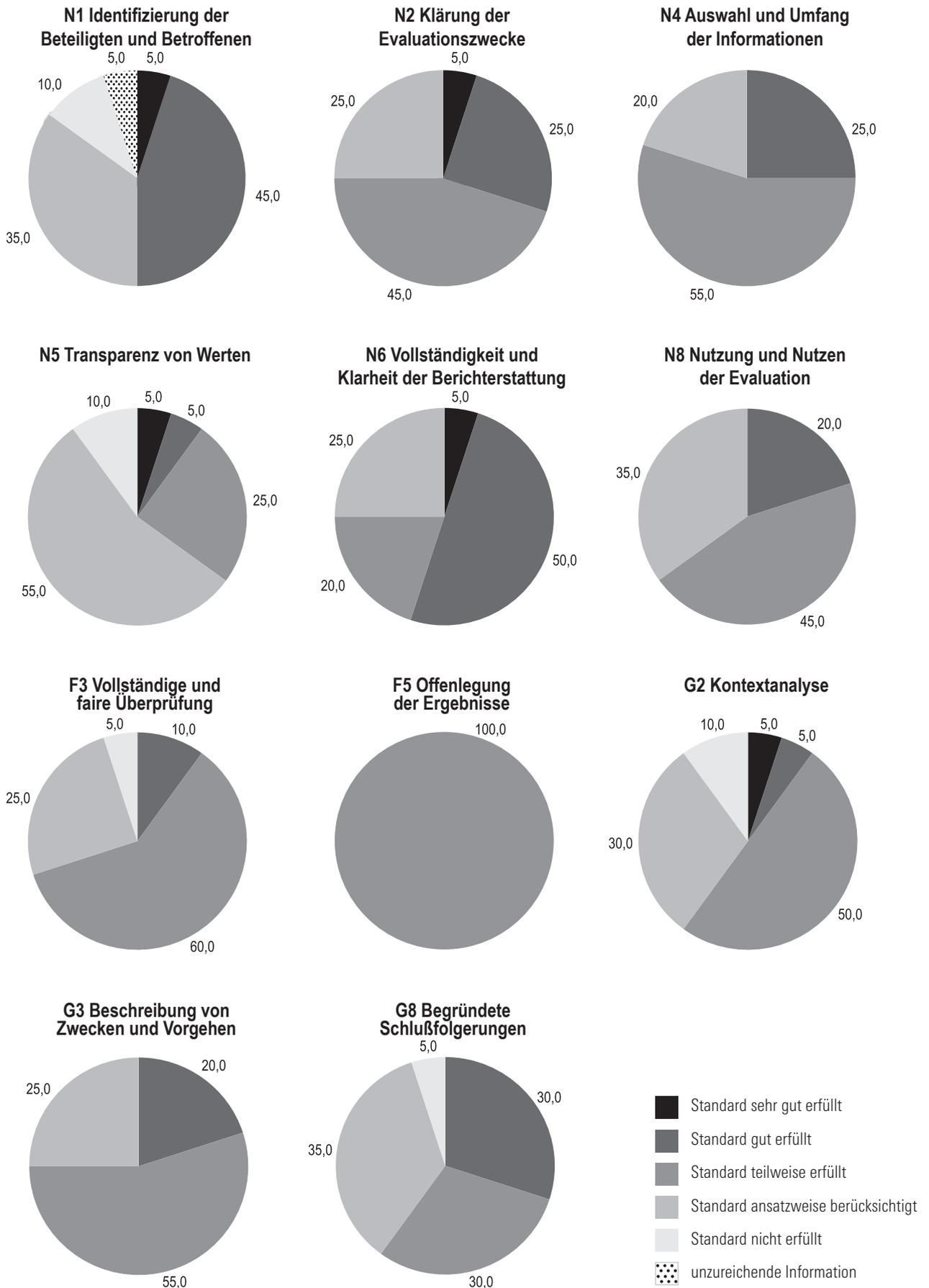


Abbildung 3: Erfüllungsggrad der Standards

mals eher unsystematischen Zugangs. Während manche Evaluationen konsequent nachvollziehbare Bewertungsmaßstäbe anwendeten, die in einer klaren Verbindung zu den Programmzielen standen, griffen andere eher auf Einzelbewertungen von singulären Beobachtungen zurück, ohne dass in der Kombination von „üblichen“ Betrachtungsweisen ein stringentes Gesamtkonzept greifbar wäre. Es wird in der internationalen Evaluationstheorie allerdings davon ausgegangen, dass die Wahl der Bewertungsmaßstäbe ebenso eine tragende Säule jedes Evaluationskonzepts darstellt wie ihre Wissenschaftlichkeit und ihre gezielte Auseinandersetzung mit dem intendierten Nutzen.

In Zusammenhang mit eingeschränkten Datenlagen waren die EvaluatorenInnen immer wieder bestrebt, Lücken durch ihr Hintergrundwissen über das FTI-System und Annahmen über dessen Funktionsweisen oder Eigenschaften von Akteursgruppen zu kompensieren (Standards F3, G8). Dies beeinflusste häufig den Charakter von Schlussfolgerungen und Empfehlungen, die in unterschiedlicher Weise, aber doch teils recht deutlich, einen Zug von ExpertInnengutachten tragen, in denen das persönliche Wissen der AutorInnen zur Geltung gebracht wird. Gemäß dem Grundansatz der DeGEval-Standards sollten freilich alle Aussagen einer Programmevaluation in transparenter Weise in von ihr herangezogenen Fakten und Quellen abgestützt sein.

Fast alle analysierten Programmevaluationen haben sich auch mit dem Kontext der evaluierten Programme auseinandergesetzt (Standard G2), allerdings in unterschiedlicher Intensität und mit unterschiedlichen Perspektiven. Vor allem auf Basis qualitativer Untersuchungsstrategien haben manche Evaluationen essentielle Bedingungen greifbar gemacht, unter denen das jeweilige Programm in seinen Zielgruppen Wirkungen entfalten konnte bzw. daran gehindert war. Etliche Kontextanalysen leiden jedoch darunter, dass zwar einige Faktoren untersucht und für skizzenhafte Bilder fruchtbar gemacht wurden, aber der systematische Stellenwert dieser untersuchten Faktoren unklar bleibt bzw. keinen expliziten Bezug zu einer strukturierten und gesamthaft verstandenen Wirklogik des jeweiligen Programms aufweist.

Was die Klärung der Evaluationszwecke angeht (Standard N2), wird aus den Angaben in rund drei Vierteln der Berichte ersichtlich, dass ein lernorientierter Ansatz verfolgt wurde, der einer Anpassung des evaluierten Programms oder der Ausrichtung zukünftiger Programme zugute kommen konnte. Ein erheblicher Teil der Evaluationsberichte benannte die jeweilige Zwecksetzung allerdings nur in einer recht oberflächlichen Weise (etwa „Reflexion des Programmverlaufs“, „kritische Würdigung des Programms“). Die vom Standard N2 empfohlene Priorisierung und Strukturierung von Evaluationszwecken angesichts der bereits erwähnten multiplen Evaluationsschwerpunkte ist dabei kaum zu entnehmen.

PROZESSMERKMALE VOR DEM HINTERGRUND DER DEGEVAL-STANDARDS

Die herangezogenen DeGEval-Standards befassen sich sowohl mit Eigenschaften der Evaluationsberichte als auch mit Prozessmerkmalen, die als förderlich für hohe Evaluationsqualität und bestmögliche Präsentation und Verwertung der Ergebnisse erachtet werden. Im Konnex mit den oben dargestellten „Problemzonen“ der Evaluationspraxis erscheinen unter anderem die folgenden Punkte wesentlich, ohne dass es sich bei diesen Aspekten um eine abschließende Liste handeln würde.

Alle die Befragung beantwortenden EvaluatorInnen sprechen davon, dass die Erwartungen der AuftraggeberInnen in Anbetracht der verfügbaren Ressourcen und der eingesetzten Methoden zumindest in manchen Fällen unrealistisch waren. In den Gesprächen mit den AuftraggeberInnen wurde eine zunehmende Erkenntnis zum Ausdruck gebracht, dass in früheren Entwicklungsphasen der Evaluationspraxis gehegte Vorstellungen über die Leistungskraft von Programmevaluationen im Verhältnis zu verfügbaren Ressourcen oft deutlich überzogen waren. Allerdings geben nur 29% der EvaluatorInnen an, dass in den Planungen stets auch gezielte Schritte unternommen wurden, um etwaige Unterschiede in der Erwartungshaltung und im Verständnis von AuftraggeberInnen und EvaluatorInnen zu klären.

Dass Auswahl und Umfang der Informationen so bestimmt wurden, dass alle vorgesehenen Evaluationsfragestellungen gut behandelt werden konnten, war aus Sicht von nur 23% der EvaluatorInnen in allen ihrer Projekte erfüllt. Einem Drittel der antwortenden EvaluatorInnen zufolge waren Programmevaluationen nur selten oder nie so angelegt, dass sie auch eine Auseinandersetzung mit nicht-intendierten Wirkungen oder unerwünschten Nebenwirkungen des evaluierten Programms ermöglichten. Eine schrittweise Projektgestaltung und von vornherein eingeplante Flexibilität, um während der Evaluationsdurchführung auf veränderte Umstände reagieren zu können (z.B. unerwartete Datenlage, überraschende Zwischenergebnisse, Veränderungen der Informationsbedürfnisse), wird lediglich von 18% der EvaluatorInnen allen von ihnen durchgeführten Evaluationen attestiert.

Welche Wertmaßstäbe für die Interpretation der Ergebnisse herangezogen werden sollten (N5), wurde 42% der antwortenden EvaluatorInnen zufolge immer oder zumindest häufig frühzeitig geklärt. Allerdings meinen nur 4%, dass ein solcher wesentlicher Klärungsschritt immer erfolgte. In den konkreten Bewertungsverfahren wurden dann 54% der EvaluatorInnen zufolge immer oder häufig Maßstäbe und Kriterien angewendet, die aus Sicht der EvaluatorInnen denjenigen ihrer jeweiligen AuftraggeberInnen entsprachen bzw. die diese vorgaben. 44% berichten hingegen, dass Bewertungsmaßstäbe, die bei AuftraggeberInnen oder anderen Stakeholdern des Programms möglicherweise existierten, für die konkrete Vorgehensweise der durchgeführten Evaluationen nicht wesentlich waren.

Das zentrale Anliegen, dass die EvaluatorInnen durch präzise Definitionen der übergreifenden Evaluationszwecke mit einem klaren Arbeitsauftrag ausgestattet sein sollen (N2), sehen lediglich 39% der antwortenden EvaluatorInnen stets erfüllt. Und lediglich 28% gaben an, dass bei den von ihnen durchgeführten Programmevaluationen immer ein Hauptzweck der Evaluation klar im Vordergrund stand bzw. dass die Evaluationszwecke gemeinsam mit den AuftraggeberInnen mit klaren Prioritäten ausgestattet wurden.

Mit der Identifikation und Einbeziehung der Beteiligten und Betroffenen (N1) wird ein Qualitätsaspekt als relevant für die Nützlichkeit einer Evaluation ausgewiesen, dem die analysierten Evaluationsberichte bislang so gut wie keine Aufmerksamkeit geschenkt haben. Relevante Angaben finden sich nahezu ausschließlich in Methodendarstellungen. Zwar existiert kaum eine Programmevaluation im untersuchten Sample, die nicht im Zuge ihrer Datenerhebungen auch Erfahrungen und Sichtweisen von Programmverantwortlichen und mit der Umsetzung des Programms Betrauten in ihre Analyse einbezogen hätte. Hinzu kommen oft Erfahrungen und Sichtweisen der Zielgruppen des evaluierten Programms bzw. von Ausschnitten dieser Zielgruppen, da in aller Regel auf Fördernehmer fokussiert wird. Allerdings wird in dieser Herangehensweise der Grundidee einer Einbeziehung aller Akteure in die Eva-

luationsplanung, für die eine Programmevaluation nützlich werden soll, nur bedingt Genüge getan. Dies betrifft insbesondere Akteure im FTI-politischen Governancesystem, die ähnliche Programme planen bzw. in ihren FTI-politischen Rollen Interesse am untersuchten Programm hatten bzw. haben können. Die EvaluatordInnen-Befragung bestätigt, dass es am ehesten die für die Evaluation unmittelbar zuständigen AnsprechpartnerInnen in den auftraggebenden Institutionen waren, die in die Evaluationsplanung einbezogen wurden – und auch dies nicht in allen Fällen.

In den Interviews mit den AuftraggeberInnen tritt ein Evaluationszweck der Rechenschaftslegung deutlich hervor, der in den analysierten Evaluationsberichten nicht explizit wird. Zugleich wird erkennbar, dass diese Evaluationsfunktion der Rechenschaftslegung, die im Rahmen der institutionell-rechtlichen Verankerung der Programmevaluationen immer schon mit einprogrammiert ist, auch zu Einschränkungen für die Lernfunktion der Programmevaluationen führte – also für die Erzielbarkeit von instrumentellem und konzeptuellem Nutzen sowie von Wissenszuwachs in breiteren Sphären („Aufklärung“). Von Auftraggeberseite wird beschrieben, dass durch Evaluationsvorgaben (Evaluationszeitpunkte, Verankerung von Evaluationsfragestellungen, Budgetierung) aktueller Erkenntnisbedarf des Öfteren nur bedingt befriedigt werden konnte.

Evaluationsberichte werden in der österreichischen FTI-Evaluationspraxis regelmäßig publiziert. Darüber hinaus reichende gezielte Schritte hin auf eine Nutzung waren bislang vor allem auf die direkten AuftraggeberInnen der Evaluation konzentriert. Wie Interviews und EvaluatordInnen-Befragung zeigen, wurde bislang eine Zirkulation evaluativer Erkenntnisse zu einer Auseinandersetzung mit ihnen im breiteren Governance-System oder hin zu Zielgruppen kaum gezielt unterstützt. Die Daten aus der EvaluatordInnen-Befragung räumen gezielten Planungen für die Nutzung und insbesondere gezielten Disseminationsaktivitäten nur untergeordnete Rollen bei der Erzeugung von Evaluationsnutzen ein. Chancen und Potenziale, die sich aus der Thematisierung von Programmintentionen und deren Verwirklichung mit systemreferentiellen und selbstreflexiven Akteuren des FTI-Systems ergeben können, wurden hier nicht wahrgenommen. Manche Akteure weisen schließlich darauf hin, dass relevante Wissensproduktion für Ausschnitte der FTI-Governance auch in nicht-publizierter Form erfolgt und damit nicht für andere Akteure im System greifbar wird.

EINFLUSSFAKTOREN AUF DIE ENTSTEHUNG VON EVALUATIONSNUTZEN

Wenn sich für die bisherige Evaluationspraxis in einem großzügigen Bild, das Besonderheiten einzelner Evaluationen zwangsläufig hintanstellen muss, zeigt, dass keineswegs alle Möglichkeiten im Sinne der DeGEval-Standards ausgeschöpft wurden, worauf ist dann der von den verschiedenen Akteuren attestierte Nutzen aus Programmevaluationen zurückzuführen? Zentrale Arbeiten der Nutzungsforschung zur Evaluation gehen davon aus, dass zwei Faktorengruppen bei der Entstehung von Evaluationsnutzen wirksam werden (Cousins/Leithwood 1986, Cooksy/Caracelli 2005, Fleischer/Christie 2009): Auf der einen Seite stehen

Faktoren, die durch die Gestaltung eines Evaluationsprojekts beeinflusst werden können und daher auch dem Verantwortungsbereich der EvaluatordInnen prinzipiell zugerechnet werden. Auf der anderen Seite stehen Kontextfaktoren, die sich einer Beeinflussung auf der Ebene der Projektgestaltung entziehen und definitiv außerhalb einer Verantwortung der EvaluatordInnen bleiben.

Auf Basis der EvaluatordInnen-Befragung und der AuftraggeberInnen-Interviews ergibt sich ein Bild, in dem evaluationsinhärente Faktoren und Kontextfaktoren gleichgewichtig auf die Entstehung von Evaluationsnutzen gewirkt haben. An erster Stelle unter denjenigen Faktoren, die sich im Verantwortungsbereich einer einzelnen Programmevaluation ansiedeln – und zugleich an der Spitze aller Einflussfaktoren – steht aus Sicht der EvaluatordInnen ihre Glaubwürdigkeit bei den AuftraggeberInnen.¹⁰ Dies spiegelt unmittelbar eine Situation wieder, in der einige spezialisierte Evaluationsinstitute über Jahre, wenn nicht Jahrzehnte hinweg wiederholt herangezogen wurden und, sofern es sich um österreichische EvaluatordInnen handelt, mit den AuftraggeberInnen in der Plattform ftval in ständigem Austausch standen bzw. stehen.

Unter den Kontextfaktoren rangiert die Erwartung der AuftraggeberInnen, dass ihnen die konkrete Programmevaluation von Nutzen sein wird, an erster Stelle. Hier wird augenscheinlich eine Spannung virulent, die sich einer grundsätzlichen Kombination von Lern- und Legitimationszwecken in der Art der institutionell-rechtlichen Verankerung der Programmevaluationen verdankt und jedes Evaluationsprojekt von vornherein durchzieht. Deutliche Unterschiede im Umgang mit verschiedenen Programmevaluationen sind zu bemerken. Immer wieder wurden Evaluationen von zuständigen AuftraggeberInnen von vornherein als notwendige Übungen betrachtet und kaum mit Nutzenerwartungen verbunden, was sich von der Evaluationsplanung bis zum Umgang mit den Ergebnissen niederschlug.

Ebenfalls bedeutend für Art und Ausmaß der Nutzung von Evaluationsergebnissen ist der Umstand, ob eine Programmevaluation in direktem Bezug zu einem aktuellen FTI-politischen Entscheidungsbedarf oder Problemdruck steht. Freilich wurden aktuelle FTI-politische Themenstellungen oder „Windows of Opportunity“ bislang kaum als solche zum Auslöser von direkt auf sie gemünzten Evaluationsaktivitäten. Viel eher dominierten vorgegebene Evaluations- und Verhandlungsrhythmen für Programmvereinbarungen der Ressorts mit dem Finanzministerium. Aktuelle Informationsbedürfnisse der Ressorts und Agenturen wurden innerhalb dieses Rahmens des Öfteren nur in eingeschränkter Weise befriedigt. Zudem besteht in einzelnen Systemsegmenten in den letzten Jahren Bedarf an Typen von Politikinformation, die mit den routinisierten Multi-Purpose-Evaluationen nicht gut abgedeckt werden können.

Neben institutionellen Rahmenbedingungen des Evaluierens tritt des Weiteren eine starke Personenabhängigkeit der genaueren Umgangsweise mit einzelnen Programmevaluationen und ihren Ergebnissen zu Tage. Die Wirksamkeit dieses „Human Factor“ in der Evaluationsnutzung betrifft persönliche Denkstile der jeweils Evaluationzuständigen, die Konsistenz von Evaluationsergebnissen mit ihren Sichtweisen und Erwartungen, ihre Erfahrung mit Evaluation, und ihre Rolle innerhalb der jeweiligen Institution. Nicht zu übersehen ist daneben ein erheblicher Stellenwert von organisatorischen Anpassungen, Ressourcen und Erfahrungen der auftraggebenden Institutionen. Aus den Darstel-

lungen der AuftraggeberInnen ergibt sich, dass der Umgang mit den institutionell-rechtlichen Rahmenbedingungen sowie Bemühungen um die Verbreitung der Evaluationsergebnisse stark von einem Engagement abhängig bleibt, das für eine einzelne Programmevaluation aufgebracht wird, und sich nicht zuletzt an den zum jeweiligen Zeitpunkt gegebenen Möglichkeiten der fachzuständigen Einzelpersonen im Rahmen verschiedener Agenden bemisst. Durchgehend wird dargestellt, dass im Rahmen der gegebenen Kapazitäten keine weiteren Spielräume mehr bestehen.

Evaluationsmethodische Gesichtspunkte wie die Angemessenheit der Evaluationskriterien, eine ausgewogene Darstellung von Stärken und Schwächen des untersuchten Programms oder die Art des Evaluationsansatzes sind in den 20 wesentlichsten Einflussfaktoren auf eine Nutzenentstehung aus der Sicht der EvaluatorInnen enthalten. Methoden Aspekte im engen Sinn wie die Anwendung eines Methodenmix, Triangulation oder die Finesse, mit der bestimmte Methoden eingesetzt werden, kommen unter den 20 wesentlichsten Einflussfaktoren auf Evaluationsnutzung, so wie die EvaluatorInnen sie einschätzen, nicht vor. AuftraggeberInnen haben gelegentlich auf Mängel im Einzugsbereich der Methodenanwendung hingewiesen, die in ihren Augen in der Vergangenheit die Entstehung von Nutzen aus einzelnen Programmevaluationen beeinträchtigt haben. Bei aller Wichtigkeit dieser Diagnose ist zugleich festzuhalten, dass Methodendiskussionen und -expertise alleine weitere wesentliche Aspekte der Entstehung von Evaluationsnutzen nicht abzudecken vermögen.

ANSATZPUNKTE EINER STEIGERUNG VON NÜTZLICHKEIT UND NUTZEN VON PROGRAMMEVALUATIONEN

Aus den Ergebnissen der Metaevaluation wird deutlich, dass Programmevaluationen im österreichischen FTI-Governancesystem augenscheinlich eine wichtige Funktion zukommt. Es wird jedoch auch evident, dass die Rolle der Programmevaluationen hinsichtlich ihrer Lernfunktion und der Schaffung von bestmöglichen Nutzungsvoraussetzungen durchaus noch weiter ausgebaut werden kann. Über die vergangenen beiden Jahrzehnte hinweg ist der Aufbau einer Evaluationskultur in diesem Bereich ein Stück weit gelungen, wobei der Plattform fteval unbezweifelbar eine bedeutende Rolle zukommt (vgl. Warta/Philipp 2016). Die bisher erreichte Evaluationspraxis im FTI-Bereich erweist sich als gleichermaßen durch Gestaltungsmerkmale einzelner Programmevaluationen wie durch Kontextfaktoren bedingt. Es sind Faktorenbündeln von hoher Komplexität, mit denen man es bei der Erreichung von Evaluationsnutzen und damit verbundenem Verbesserungspotenzial zu tun hat. Daher kann kein einzelner hauptverantwortlicher Faktor für eine entscheidende Veränderung namhaft gemacht werden. In der sowohl von den Standards als auch von der Nutzungsforschung eingenommenen Grundperspektive verteilt sich die Verantwortung für gelungene Evaluationsprojekte gleichermaßen auf die EvaluatorInnen wie auf deren AuftraggeberInnen. Damit werden für möglichst produktive und nutzbringende Evaluationen nicht nur Fragen der Erarbeitung von Evaluationsergebnissen mit Hilfe der eingesetzten Methoden relevant, sondern darüber hinaus Anbahnungsprozesse, Interaktionsmuster vor, während und nach der Evaluationsdurchführung, und Handlungsspielräume innerhalb struktureller

Rahmenbedingungen. Verbesserungsmöglichkeiten siedeln sich daher auf allen Ebenen der Evaluationsplanung, -durchführung, -präsentation und -verwertung an.

Ein Eintreten von Nutzen überhaupt kann keineswegs gleichgesetzt werden mit der Schaffung bestmöglicher Grundlagen für eine größtmögliche Entstehung von Nutzen. Evaluationsberichte und die dahinter stehenden Evaluationsprozesse gehen auf das, was die DeGEval-Standards als optimale Schritte hin zu hoher Nützlichkeit bezeichnen, bislang nur bedingt ein. Wenn es also zu Beiträgen der Evidenzproduktion gekommen ist, die die von den Standards empfohlenen bzw. als notwendig erachteten Evaluationseigenschaften nicht optimal verwirklichten, so geht es um die Realisierung von Evaluationsprozessen, die von den ersten Schritten der Evaluationsplanung weg in vielfältiger Weise beeinflusst und gesteuert werden. Und während die DeGEval-Standards vielfältige Hinweise zur Bewältigung von Problemen der Nutzenentstehung enthalten, so können sie doch Probleme nicht lösen, die außerhalb der Gestaltungsmacht eines konkreten Evaluationsprojekts liegen.

Die Art der Verankerung der Evaluationsfunktion im FTI-politischen Governancesystem hat in der Vergangenheit zum Aufbau der Evaluationskultur beigetragen. Es sind jedoch Limitierungen für die Gestaltung von Programmevaluationen und den Nutzen von Evaluationen aus Merkmalen des institutionellen Arrangements der Bundesverwaltung erkennbar, in dem die Programmevaluationen eine Schnittstellenfunktion zwischen den Ressorts und dem Finanzministerium erfüllen, in der auch die Legitimation der Mittelausgaben zur Debatte steht. Dies hat nachvollziehbar bei der Planung und Durchführung der einzelnen Evaluationsprojekte eine gewisse Schematisierung mit sich gebracht, wodurch der Verwirklichung von möglichst nützlichen und wirksamen Evaluationsprojekten auch Grenzen gesetzt wurden. Ressourcen- und Kapazitätsbeschränkungen schlagen sich hier ebenso nieder, genauso wie bei der Auseinandersetzung mit Evaluationsergebnissen, insbesondere in einem breiteren systemorientierten Rahmen.

Der Bericht (Landsteiner 2015) gibt 20 Empfehlungen zu konkreten Schritten hin zu einer noch nutzbringenderen Evaluationspraxis, die sich sowohl auf einer evaluationsmethodischen Ebene als auch auf der Ebene der institutionellen Einbettung der Evaluationsfunktion ansiedeln. Der Bericht empfiehlt, möglichst viele dieser Schritte in Kombination miteinander zu setzen. An dieser Stelle sollen grundlegende Linien aufgezeigt werden, entlang derer eine kompromisshafte Situation hin zu größeren Nutzenpotenzialen und effektiveren Nutzungen weiter entwickelt werden kann:

- Wie können Programmevaluationen im Rahmen der Vergabeverfahren und Auftragserteilungen sowie im Verständnis der EvaluatorInnenrollen so gestaltet werden, dass sie in bestmöglicher Weise Lernen unterstützen, systemische Erkenntnisse erbringen und hinsichtlich von Fragestellungen und herangezogenen Daten nicht durch frühe Vordefinitionen eingeschränkt werden?
- Wie kann ein übergreifendes, systemisch wirksames Qualitätsverständnis für die Planung, Durchführung und Verwertung von Evaluationen quer über verschiedene Systemstellen des politisch-administrativen Handelns im FTI-Bereich hinweg verankert und laufend gesichert werden?
- Wie kann ein solches systemisch tragendes Qualitätsverständnis angesichts der kleinteiligen Zuständigkeitsbereiche der österreichischen FTI-Governance und der beschränkten Zahl deutschsprachiger FTI-EvaluatorInnen gegenüber relevanten

evaluationsmethodischen Entwicklungen und Optionen offen gehalten werden?

- Wie kann es künftig besser gelingen, auf einer institutionell-organisatorischen Ebene auf Wegen der Einbeziehung und Koordination unterschiedlicher Akteure in Governancesystem und Zielgruppen bestmögliche Voraussetzungen für nutzbringende Evaluationsprojekte im systemischen Gesamtzusammenhang zu erreichen und zu einer umfassenderen Organisation von Reflexion zu gelangen?
- Können entscheidende Schritte zugunsten einer Lernorientierung und einer systematisch verlässlichen Schaffung von Lernmöglichkeiten alleine im Handeln der betroffenen Akteure unter gleichbleibenden Rahmenbedingungen gegründet werden?

Eine Flexibilisierung in mehreren Hinsichten erscheint notwendig, um tiefere Einsichten über FTI-politische Handlungsabschnitte im Gesamtsystem auf hohem Qualitätsniveau zu gewinnen und größere systemische Wirksamkeit der Evaluationseinsätze zu erzielen. Es bleibt zu sehen, ob vorhandener Spielraum für nützlichere und wirksamere Evaluationsprojekte durch die handelnden Akteure genutzt werden kann, um tragfähige Verbesserungen zu ermöglichen, oder ob zur Schaffung optimaler Voraussetzungen auch die Art der Verankerung der Evaluationsfunktion verändert werden sollte. Strukturell gegebene Potenziale für die Planung, Durchführung, Kommunikation und Nutzung von Programmevaluationen bleiben stets in den jeweiligen Projekten auf der Basis von Kapazitäten und Kompetenzen erst tatsächlich auszuschöpfen. Letztlich wird eine hoch entwickelte Evaluationskultur auch auf gesellschaftlich-kulturellen Faktoren wie der Offenheit für sachlich fundierte Kritik und der Bereitschaft zur offenen Diskussion beruhen, die sich freilich einer gezielten Beeinflussung entziehen.

LITERATUR

Alkin M.C. , Taut S.M. (2003), Unbundling Evaluation Use. *Studies in Educational Evaluation*, 29 (1), 1-12.

Beywl W., Taut S. (2000), Standards: Aktuelle Strategie zur Qualitätsentwicklung in der Evaluation. *Vierteljahrshefte zur Wirtschaftsforschung* 69. Jahrgang, Heft 3/2000, S. 358–370.

Cooksy L.J., Caracelli V.J. (2009), Metaevaluation in Practice. Selection and Application of Criteria, *Journal of MultiDisciplinary Evaluation*, Volume 6, Number 11: 1-15.

Cooksy L.J., Caracelli V.J. (2005), Quality, Context, and Use Issues in Achieving the Goals of Metaevaluation. *American Journal of Evaluation*, Vol. 26 No. 1, 31-42.

Cousins J. B., Leithwood K. A. (1986), Current empirical research on evaluation utilization. *Review of Educational Research*, 56, 331–364.

DeGEval – Gesellschaft für Evaluation (2008) Standards für Evaluation, 4. Aufl., Mainz: DeGEval.
www.degeval.at/publikationen/standards-fuer-evaluation/

Edler J., Berger M., Dinges M., Gök A. (2012), The practice of evaluation in innovation policy in Europe. *Research Evaluation* 21 (3), 167-182.

Fleischer D. N., Christie C. A. (2009), Evaluation Use - Results From a Survey of U.S. American Evaluation Association Members. *American Journal of Evaluation* Volume 30 (2), 158-175.

Good B. (2006), Technologie zwischen Markt und Staat. Die Kommission für Technologie und Innovation und die Wirksamkeit ihrer Förderung, Zürich-Chur:Rüegger.

Mark M.M., Henry G.T. (2004), The Mechanisms and Outcomes of Evaluation Influence, *Evaluation* Vol 10(1): 35–57.

Joint Committee on Standards for Educational Evaluation / James R. Sanders (Hg.) (2006), *Handbuch der Evaluationsstandards*. 3., erweiterte und aktualisierte Auflage, übersetzt und für die deutsche Ausgabe erweitert von Wolfgang Beywl und Thomas Widmer, Wiesbaden: VS Verlag für Sozialwissenschaften.

Johnson K., Greenside L.O., Toal S.A., King J.A., Lawrenz F. and Volkov B. (2009), Research on Evaluation Use: A Review of the Empirical Literature from 1986 to 2005. *American Journal of Evaluation*, 30(3): 377 - 410.

Kirkhart E.K. (2000), Reconceptualizing evaluation use: An integrated theory of influence. *New Directions for Evaluation Special Issue: The Expanding Scope of Evaluation Use*, Volume 2000, Issue 88, 5–23.

Landsteiner G. (2015): Nützlichkeit und Nutzen der Programmevaluationen im Bereich der österreichischen FTI-Politik. *Metaevaluation der*

Programmevaluationen 2003-2014. Studie im Auftrag des Rates für Forschung und Technologieentwicklung, Wien: RFTE.

Leviton L.L., Hughes E.F.X (1981), Research on the Utilization of Evaluations. *Evaluation Review* Vol 5, No 4, 525-548.

Patton M.Q. (1997), *Utilization-Focused Evaluation*. The New Century Text, Thousand Oaks: Jossey Bass.

Pichler R. (2009), Institutionelle Dimensionen von Evaluierung in Österreich. In: Widmer T., Beywl W., Fabian C. (Hrsg.), *Evaluation - ein systematisches Handbuch*, Wiesbaden: VS Verlag für Sozialwissenschaften, 40-51.

Rat für Forschung und Technologieentwicklung (RFTE, 2013): *Weißbuch zur Steuerung von Forschung, Technologie und Innovation in Österreich*, Wien: RFTE.

Rist R.C., Stame N. (eds.) (2006), *From Studies to Streams: Managing Evaluative Systems*, New Brunswick – NJ: Transaction Publishers.

Shulha, L. M., Cousins, J. B. (1997). Evaluation use: Theory, research, and practice since 1986. *Evaluation Practice*, 18, 195-208.

Stufflebeam D.L. (2001), The Metaevaluation Imperative, *American Journal of Evaluation* Vol. 22, No. 2, 2001, 183–209.

Warta K, Philipp S. (2016), Was bringt die Plattform fteval und zu was hat sie es gebracht? Rückblick und Ergebnisse einer Mitgliederbefragung. *fteval Journal for Research and Policy Evaluation*, Issue 41, March 2016, 5-14.

Weiss C.H. (1998), Have We Learned Anything New About the Use of Evaluation? *American Journal of Evaluation*, Vol. 19, No. 1, 21-33.

Widmer T., Landert C., Bachmann N. (2001), *Evaluations-Standards der Schweizerischen Evaluationsgesellschaft (SEVAL-Standards)*. Bern/Genève: SEVAL.

Zinöcker K, Dinges M. (2009), Evaluation von Forschungs- und Technologiepolitik in Österreich. In: Widmer T., Beywl W., Fabian C. (Hrsg.), *Evaluation - ein systematisches Handbuch*, Wiesbaden: VS Verlag für Sozialwissenschaften, 295-304.

AUTOREN

GÜNTHER LANDSTEINER

Unabhängiger Forscher-Berater und Evaluator

E: landsteiner-rce@chello.at

JOHANNES GADNER

Rat für Forschung und Technologieentwicklung

E: j.gadner@rat-fte.at

DIE VERMESSENHEIT DER VERMESSUNG: ÜBERLEGUNGEN ZUR UNIVERSITÄREN GOVERNANCE IN DEN GEISTES- UND SOZIALWISSENSCHAFTEN

GÜNTHER R. BURKERT, DAVID F. J. CAMPBELL UND THORSTEN D. BARTH

1. DIE BESTEHENDE GOVERNANCE

Die durch das UG 2002 mit weitreichenden Kompetenzen ausgestatteten Rektorate der österreichischen Universitäten haben die verstärkte Profilbildung – hier vor allem den gezielten Ausbau der Forschungsaktivitäten – vor allem durch eine gezielte Beru- fungs- und Personalpolitik sehr erfolgreich in die Wege geleitet. Die der New Public Management-Philosophie (NPM)¹ geschuldete Verteilung der Ressourcen innerhalb der Universitäten wurde mit Hilfe von Zielvereinbarungen (Outputsteuerung)² durchgeführt. Dafür war die Messbarkeit von Effizienz und Effektivität bei Forschungstätigkeiten notwendig, weil ja gerade diese als profilbildendes Element von entscheidender Bedeutung geworden sind⁴.

Große Teile dieser Forschungsaktivitäten wurden mit Hilfe von Drittmitteln finanziert, das für die immer zahlreicher werdenden Anträge notwendige Personal wurde größtenteils aus dem Grundbudget finanziert. Inzwischen kam es – nicht zuletzt durch Gesetzesänderungen – zu wesentlichen Verbesserungen von stabilen Forschungskarrieren, sodass in den Disziplinen und Fachbereichen der Naturwissenschaften und Technischen Wissenschaften ein ausgewogenes Verhältnis zwischen drittmittelangestellten und grundbudgetfinanzierten Forscherinnen und Forschern zu erwarten ist. Dass die Profilbildungen überwiegend in diesen Wissenschaftsbereichen erfolgten, hat einerseits mit dem leichteren Zugang zu Drittmitteln für diese Bereiche zu tun, spiegelt aber auch die unterschiedlichen Zugänge zu den Drittmitteln in den verschiedenen Fachkulturen wider. Geht es nämlich um die Vermessung der Disziplinen, so werden diese nach den Regeln der Naturwissenschaften und Technischen Wissenschaften durchgeführt, was zur „Vernaturwissenschaftlichung“ ehemals geistes- und sozialwissenschaftlicher Fächer beigetragen hat⁵.

2. DIE „UMSTRITTENE“ VERMESSUNG

Nimmt man die beiden Hauptindikatoren klassischer Vermessung von wissenschaftlichen Leistungen – sie müssen natürlich unter dem Aspekt der Vergleichbarkeit für das Controlling quantitativ sein – dann sind das Publikationen und Drittmittel. Gerade hier wird sehr schnell sichtbar, dass wir es in den verschiedenen Fachbereichen mit höchst unterschiedlichen Zugängen zu diesen beiden Indikatoren zu tun haben. Die Einwände gegen diese Form der Vermessung sind beinahe unübersehbar: sie reichen von der Notwendigkeit der Berücksichtigung der Unterschiede bei Produktion und Publikations-Usancen der Forschungsergebnisse⁶ bis zur Kritik der Verabsolutierung zahlreicher NPM-basierter Indikatoren, die im Widerspruch zu hervorragenden Monografien und zur „armchair research“⁷ stünden.

Schließlich wird noch das Problem des Trittbrettfahrens bei der Anzahl der Zitationen als Gütekriterium thematisiert⁸. Dass Zeitschriften-Rankings und Zitationsindizes ein schmales Band wissenschaftlicher Arbeiten bevorzugen und eher die Orthodoxie statt die Innovation und heterodoxe wissenschaftliche Ansätze fördern, sollte bei der Auswertung dieser Datenbasis nicht vergessen werden.

3. DIE NOTWENDIGE „NEUE“ GOVERNANCE

Die zunehmende Kritik am New Public Management als ungeeignetes Instrumentarium für das Managen komplexer Systeme wie Universitäten hat auch die Diskussion um eine universitätsgerechtere Governance verstärkt. New Public Management wurde immer mehr als „managing networks“⁹ gesehen. Ergänzt werden könnte diese Sicht beispielsweise aus der Systemtheorie, die auf die Komplexität der Gesellschaft rekurriert

1 Vgl. Jansen, Von der Steuerung zur Governance: Wandel der Staatlichkeit? in: Handbuch Wissenschaftspolitik, 2010, 39 – 50

2 Promberger 1998, S.93 f. zit. Nach Eschenbach/Figl/Kraft, Handbuch für Universitätsmanager 2005, S.25

3 Hellmann, Kommunikationsmanagement, 2011, S.24

4 Brüggemeier, Public Management, 2002, S.377 f.

5 Vgl. Burkert (2016), Forschungsevaluation (zum Druck abgeliefert)

6 Kaufmann (2009), Differences, S.10 f.

7 Ebda. S. 19

8 Vgl. Frey/Osterloh (2012) Managing Motivation

9 Rhodes, (1996), The new governance: governing without government. Polit. Stud. XLIV 658: „I use the term network to describe the several interdependent actors involved in delivering services“

und daher meint, dass Governance im besten Fall nur mehr in der Lage sei, Prozesse der „Selbstorganisation, Selbststeuerung und Eigen-Governance“ zu beeinflussen¹⁰. Externe Governance (beispielsweise Ministerien für Wissenschaft und Forschung) könnten in diesem Sinne als „Anleitung zur Selbststeuerung“, „kontrollierte Anregung zur Selbständerung autonomer Systeme“ oder zur Setzung von „Anreizen zur Selbständerung autonomer Systeme“ verstanden werden.

Rektorate österreichischer Universitäten stehen vor der schwierigen Aufgabe, einerseits Leistungen in diversen von der Gesellschaft eingeforderten Bereichen der Forschung – von der „entrepreneurial university“ (als Gesamtvorstellung) bis zu Forschungsinhalten aus der Zivilgesellschaft – zu erbringen, müssen aber gleichzeitig auch auf identifizierte Defizite, vorhandene Potentiale und zukünftig erwartete Chancen und Risiken mit möglichst effektiven und effizienten Maßnahmen zur Verbesserung der Performance reagieren und dies nach innen kommunizieren¹¹.

Es muss daher darauf Bezug genommen werden, was die jeweils zugrundeliegenden Verständnisse von Forschung, Lehre und Qualität sind, damit (interne, externe) Governance auch „epistemisch-treffend“ sein und einen nachhaltigen Beitrag für Qualitäts- und Universitätsentwicklung leisten kann. Der neue Ansatz der „Epistemic Governance“¹² versucht in diesem Verständnis eine Antwort für Governance im Hochschulbereich zu geben. Die Qualitätssicherung der Forschungsleistungen ist dabei ein Mittel, sie kann aber nur über laufende Qualitätsentwicklung erreicht werden. Wie diese Qualitätssicherung auszusehen hat, bedarf einer umfassenden Diskussion.

Die Formen der Abbildung von Forschung in allen ihren Dimensionen sind höchst unterschiedlich, Artikel-Analysen („Bibliometrie“) sind dafür nur eine Möglichkeit¹³. Die „Vermessenheit der Vermessung“ bringt hier ein nicht-lineares dialektisches Verhältnis zum Ausdruck. Alle Vermessungsversuche von wissenschaftlichen Teilbereichen können auch als Versuche interpretiert werden, professionelle Diskurse über den Weg der Qualitätsentwicklung voranzutreiben¹⁴. Da jede Vermessung für sich letztlich nur ein jeweils einmalig geltender Thesenentwurf ist, erscheint er auch nur als Thesenformulierung am sinnvollsten. Damit wird versucht aufzuzeigen, wie sich im Wechselwirkungsbereich von Forschung und Governance interessante und kreative Fragestellungen entwickeln und weiter diskutieren lassen.

4. DIE DIMENSION DER FORSCHUNG: PRÄSENTATION EINER FALLSTUDIE ZU ARTIKELPUBLIKATIONEN IN DEN SOZIAL- UND GEISTESWISSENSCHAFTEN AN ÖSTERREICHS UNIVERSITÄTEN

Artikelpublikationen (in referierten Fachjournalen) werden als Ausdruck der Anwendung anerkannter wissenschaftlicher Standards und Qualität gewertet. Doch häufig fokussieren bibliometrische Artikelpublikationen einseitig auf die Naturwissenschaften. Hier galt jedoch die hier als Grundlage dienende bewusste (und innovative) Überlegung, die disziplinären Bereiche der Sozial- und Geisteswissenschaften in einem längeren Jahrestrend systematischer zu betrachten.¹⁵ In einer Fallstudie zu Österreichs Universitäten wurden Artikelpublikationen in den Sozial- und Geisteswissenschaften für die Jahre 2007, 2010 und 2013 erfasst und verglichen. Ausgangspunkt war eine Analyse des „ISI Web of Science“ und des SSCI (Social Sciences Citation Index) und A&HCI (Arts & Humanities Citation Index).¹⁶ Davon angeleitet wurde eine eigenständige Daten-Basis entwickelt, aufbauend auf dem Prinzip der „Brutto-Institutionen“: Methodisch bedeutet dies, dass in Fällen von Artikelpublikationen mit mehr als einer institutionellen Nennung von AutorInnenschaft, der Artikel jeweils jeder Institution mit „1“ zugeordnet wird. „Brutto-Institutionen“ von Artikelpublikationen repräsentieren und bilden damit eine eigene Forschungslandkarte ab, die Networking und Forschungsk Kooperationen dabei im Besonderen betont. Tabelle 1 und Abbildung 1 projizieren die empirischen Ergebnisse dieser „Vermessung“ (Vermessungs-„Vermessenheit“).

Ausgehend von der Datenlage in Tabelle 1 und Abbildung 1 (und anderen weiterführenden und disaggregierteren Analysen) lassen sich folgende Verlaufstrends zur Diskussion stellen:

10 Willke, (2014), Systemtheorie III: Steuerungstheorie 44 f.

11 Die Innenkommunikation scheint momentan größtenteils noch eine Schwachstelle; vgl. Elias (2014), S. 8

12 Campbell/Carayannis (2013)

13 Hinzuweisen ist hier auf den spannenden Ansatz der Universität Klagenfurt, bei dem erstmals die auf einer breiten internen Diskussion fußenden Indikatorenstellung für eine Forschungsdokumentation auch und gerade für die Geistes-, Sozial- und Kulturwissenschaften erfolgte; vgl. dazu Friederike Wall, Martin Hitz, Erich Schwarz, u.a. Das Projekt „Qualitätsgewichtete Publikationskennzahlen“ der Alpen-Adria-Universität Klagenfurt, 2013

14 Vgl. Burkert (2011), Hochschulplanung, S. 133 ff.

15 Campbell, David F. J. (2015). Forschungs- und Publikationslandkarte Sozialwissenschaften und Geisteswissenschaften Österreich 2013. Studie für das BMWF. Wien.

16 In dieser Fallstudie wurden auch die Zitationen von Artikeln betrachtet, wobei die Zitationszeiträume für die verschiedenen Jahre unterschiedlich lange sind, was das Ergebnis eines gewissen Projektflusses war. Zitationszahlen in Tabelle 1 können daher nicht einfach für eine lineare Outputanalyse verwendet werden, sondern müssten erst in ein analytisches Modell uminterpretiert werden, um Vergleichbarkeiten zu gewährleisten. Für die weiterführende Diskussion im Rahmen dieses Artikels stehen die Zitationszahlen aber auch nicht im Vordergrund.

	Jahr 2007		Jahr 2010		Jahr 2013	
	Artikel („Brutto-Institutionen“) (a)	Zitationen (der 2007-Artikel am 07.12.2010)	Artikel („Brutto-Institutionen“) (a)	Zitationen (der 2010-Artikel am 01.11.2011)	Artikel („Brutto-Institutionen“) (a)	Zitationen (der 2013-Artikel am 11.04.2015)
Akademie der bildenden Künste Wien	2	0	1	1	5	7
Donau-Universität Krems	1	27	5	1	13	30
Institute of Science & Technology Austria (IST Austria)					1	1
Johannes-Kepler-Universität Linz	22	125	44	27	77	111
Karl-Franzens-Universität Graz	30	110	113	129	118	298
Medizinische Universität Graz	28	256	44	61	80	247
Medizinische Universität Innsbruck	46	389	37	90	59	172
Medizinische Universität Wien	90	911	117	253	149	403
Montanuniversität Leoben	2	3	5	6	6	6
Paris-Lodron-Universität Salzburg	31	122	88	135	113	289
Technische Universität Graz	5	48	14	31	26	96
Technische Universität Wien	21	87	32	73	58	201
Universität für angewandte Kunst Wien	2	0	5	0	4	0
Universität für Bodenkultur Wien	8	33	26	15	59	138
Universität für künstlerische und industrielle Gestaltung Linz					1	0
Universität für Musik und darstellende Kunst Graz	1	0	5	0	3	1
Universität für Musik und darstellende Kunst Wien	1	0	6	2	6	8
Universität für Musik und darstellende Kunst, Mozarteum	1	0	2	0	3	0
Universität Innsbruck	62	338	96	84	175	248
Universität Klagenfurt	14	79	42	36	74	145
Universität Wien	156	679	293	307	470	741
Veterinärmedizinische Universität Wien	4	50	7	29	10	17
Wirtschaftsuniversität Wien	38	169	111	105	152	290
Summe (Universitäten öffentlichen Rechts):	565	3426	1093	1385	1662	3449
Österreichische Akademie der Wissenschaften	15	49	51	55	62	63

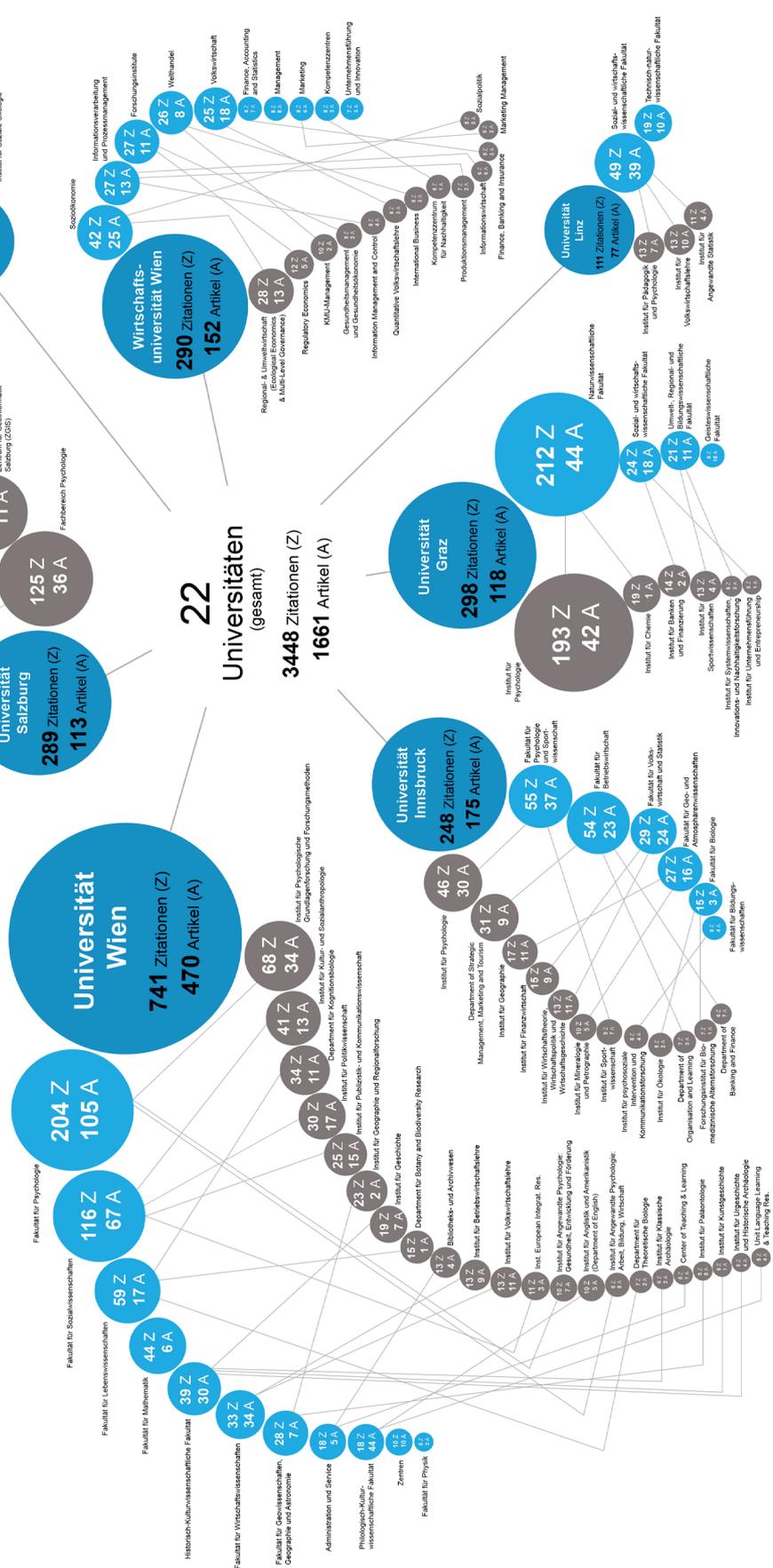
(a): „Brutto-Institutionen“ bedeutet, dass jeder österreichischen Institution der Artikel mit „1“ zugerechnet wird (relevant bei institutionellen Co-AutorInnenchaften innerhalb Österreichs). Dies entspricht im Wesentlichen auch der Logik der Wissensbilanzen.

Quelle: Eigene Berechnung der Autoren.

Tabelle 1: Sozial- und geisteswissenschaftliche Artikelpublikationen an Österreichs Universitäten (2007-2013).

Forschungs- und Publikationslandkarte 'Sozial- und Geisteswissenschaften' in Österreich 2013

Infografik-Beschreibung:
Dargestellt werden sieben ausgewählte Universitäten mit ihren Fakultäten sowie Institutionen ab einer Zitationsgröße von mindestens fünf Zitationen.



Quelle/Grafik: Eigene Berechnung und Darstellung der Autoren

Abbildung 1: Forschungs- und Publikationslandkarte 'Sozial- und Geisteswissenschaften' in Österreich 2013. (Die farbige Version der Abbildung finden Sie in der online Version des Journals unter: www.fteval.at -> Journal -> Archiv)

1. **Wachstum allgemein:** Insgesamt (über den Zeitraum 2007-2013) lässt sich ein deutliches Wachstum der Artikelpublikationen in den Sozial- und Geisteswissenschaften feststellen, das in den Sozialwissenschaften jedoch stärker ausfiel (Artikel und Zitationen). Die Geisteswissenschaften (vor allem bei den Zitationen) platzieren sich dabei hinter den Sozialwissenschaften. Möglicherweise befinden sich die Geisteswissenschaften aber an der Schwelle zu einer deutlicheren Artikelzunahme, etwas zeitversetzt den Sozialwissenschaften nachfolgend.
2. **Wachstum im Detail:** Dieses ist am deutlichsten in der Psychologie und an einigen (aber nicht an allen) Universitäten auch in den Wirtschaftswissenschaften zu erkennen. Die geisteswissenschaftlichen Fächer expandierten vielfach ebenfalls, jedoch tendenziell mehr im Bereich der Artikelpublikationen, und weniger bei den Zitationen. Bei den Sozial- und Geisteswissenschaften hat die Universität Wien die stärkste Position, und dort gab es (bei Artikeln und Zitationen) auch das deutlichste Wachstum in Österreich. Vielfach zeichnet aber jede Universität bei den Sozial- und Geisteswissenschaften ein eigenes Bild. Die Universität Graz zeichnet sich beispielsweise durch ein Wachstum bei den Zitationen aus.
3. **Kein Wachstum:** Die Forschungsleistung an den rechtswissenschaftlichen Fakultäten manifestiert sich – schwerpunktmäßig - nicht (noch nicht) in diesen Artikelpublikationen.

Als systemische Gesamthese lässt sich für das nationale Hochschulsystem Österreichs somit zur Diskussion stellen, dass im Zeitraum 2007-2013 die Artikelpublikationen (aber auch die Zitationen) in den Sozial- und Geisteswissenschaften deutlich zugenommen haben. Mögliche Einzelaussagen zu den konkreten Universitäten müssten jedoch noch erst auch auf der Grundlage der Personalentwicklung (grundfinanziert oder drittmittelfinanziert) entwickelt werden.¹⁷

Neben dieser Output-Diskussion tritt die Input-Diskussion meist in den Hintergrund. Am Beispiel der sieben für die Geistes- und Sozialwissenschaften am meisten relevanten Universitäten wurde bisher die Personalsituation weitestgehend ausgeblendet. Das Personal (globalbudgetiert und drittmittelfinanziert) der Geistes- und Sozialwissenschaften an allen österreichischen Universitäten repräsentiert aber immerhin 30,1% (siehe Abbildung 2).

Die Frage an die Governance der Universitäten sind bei dieser – zugegebenermaßen stark aggregierten – Datensituation, ob die oben dargestellten Output-Relationen für die Weiterentwicklung des Forschungsbereichs der Geistes- und Sozialwissenschaften ausreichend erscheinen oder ob sie aufgrund der gegebenen Personalsituation aus dem Globalbudget Veränderungen für notwendig erachten. Gerade die Drittmittelsituation im naturwissenschaftlichen Bereich (und technischen Bereich) sind hier die Mittel aufgrund der nationalen und internationalen Programme leichter zu lukrieren – mag für Rektorate in deren klassischen Steuerung über Drittmittel als Leistungsindikatoren verlocken, den Ausbau hier voranzutreiben, um für den Leistungsbericht der jeweiligen

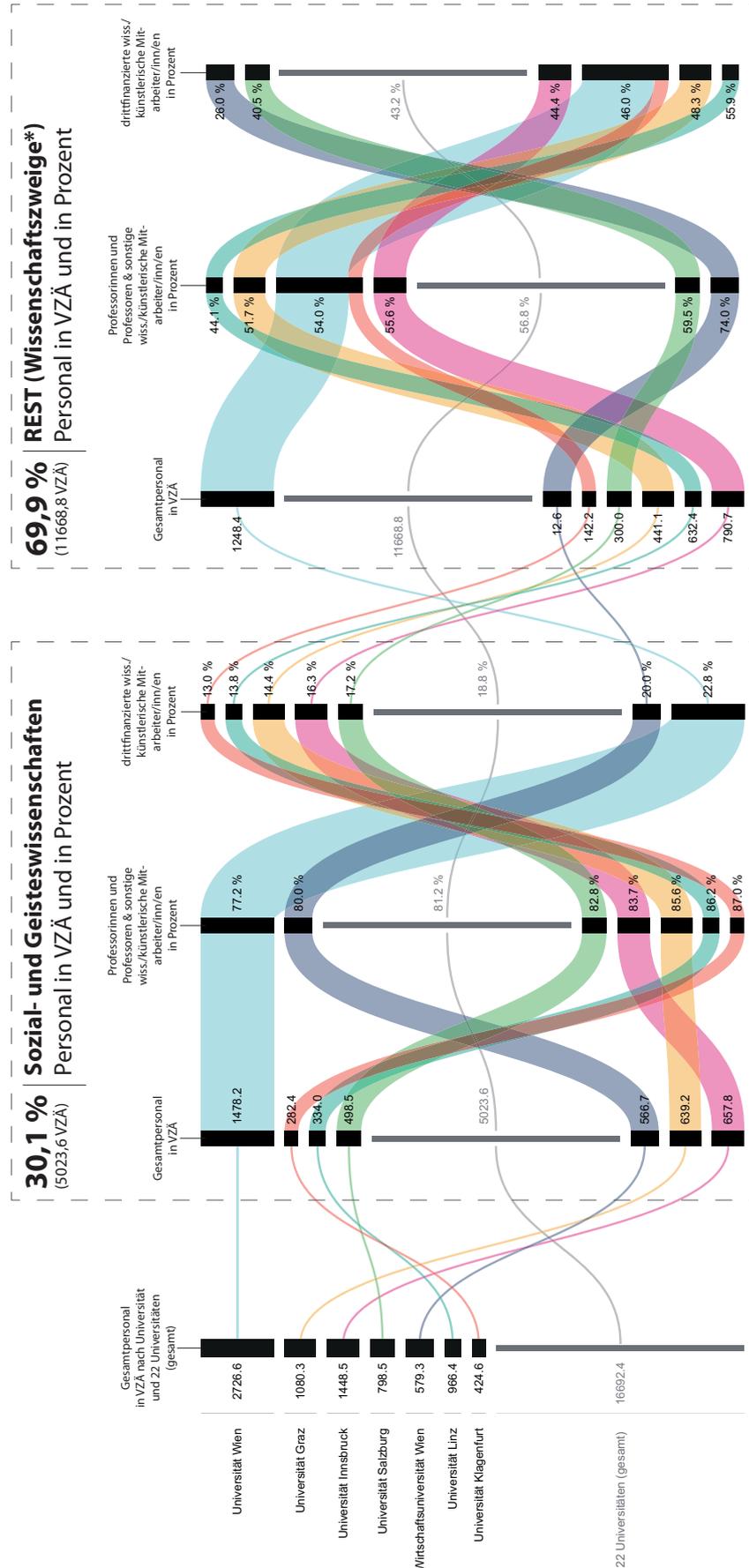
Universität Steigerungen darstellen zu können¹⁸. Im FWF beträgt die Bewilligungssumme für den Bereich Geistes- und Sozialwissenschaften (inklusive Kunst- und Kulturwissenschaften) im Durchschnitt der Jahre 2007-2011 19,8 % (2010-2014 19,0%), während er 2012 18,2% und 2015 20,2% betrug.¹⁹ Während sich daraus keine Hoffnung in Richtung eines gesteigerten Anteils am Gesamtpf herauslesen lässt, sinkt der Anteil der aus dem Gesamtpf des Globalbudgets finanzierten Personals der jeweiligen Universität.

Der Rückschluss, dass das verbliebene globalbudgetierte Personal eine erhöhte Zahl von qualitätsvollen Anträgen verfasste, wird damit kaum zu widerlegen sein.

17 Über Bedeutung und Rolle von Grundfinanzierung und Drittmittelfinanzierung siehe auch die bereits weiter zurück liegende Studie Felderer/Campbell (1994)
 18 Vgl. dazu die Forschungsschwerpunkte an Österreichs Universitäten in der Periode der Leistungsvereinbarungen 2013-2015, die im Anhang des Österreichischen Forschungs- und Technologieberichts 2013, S.181 f. aufgelistet sind
 19 FWF – Der Wissenschaftsfonds (2013). Jahresbericht 2012, S. 22-23; FWF – Der Wissenschaftsfonds (2016). Jahresbericht 2015, S. 30-31.

Explorative Darstellung der Verteilung des Gesamtpersonals in den ‚Sozial- und Geisteswissenschaften‘ und restlichen Wissenschaftszweigen 2013: 7 ausgewählte Universitäten und 22 Universitäten (gesamt) gemäß Wissensbilanz-Kennzahl 2.B.1 in Vollzeitäquivalente (VZÄ) und in Prozent

Wissensbilanz-Kennzahl 2.B.1 - Verwendungsgruppen:
 Vollzeitäquivalente (VZÄ) in den Verwendungen 11, 12, 14, 16, 21, 24 bis 27 und 81 bis 84 gemäß Z 2.6 der Anlage 1 BldokVUni



Quelle: Eigene Berechnung der Autoren; Daten: BMWFW; uni data 2016; Wissensbilanz-Kennzahl 2.B.1, Daten 2013; Design based on Raw Destiny Design 2016
 * NATURWISSENSCHAFTEN, TECHNISCHE WISSENSCHAFTEN, HUMANMEDIZIN, LAND- UND FORSTWIRTSCHAFT, VETERINÄRMEDIZIN

Abbildung 2: Explorative Darstellung der Verteilung des Gesamtpersonals in den ‚Sozial- und Geisteswissenschaften‘ und restliche Wissenschaftszweige 2013. (Die farbige Version der Abbildung finden Sie in der online Version des Journals unter: www.fteval.at -> Journal -> Archiv)

5. ÖSTERREICH ALS KOMPETENZZENTRUM FÜR DIE GEISTES- UND SOZIALWISSENSCHAFTEN

Geht man davon aus, dass einige Bereiche der Geistes- und Sozialwissenschaften weltweite Anerkennung finden und geht man darüber hinaus davon aus, dass an den angeführten Universitäten mehr als 30% des Personals dieser Wissenschaftsbereiche aus dem Grundbudget finanziert wird, könnte die Governance der Universitäten – nämlich jede für sich – beschließen, diese Fachbereiche einfach auslaufen zu lassen. Begründet könnte diese Schließung damit werden, dass die Drittmittel und Publikationszahlen nicht adäquat zur Ausstattung dieser Fächer wären. Damit würden diese Universitäten unter dem Schutz ihrer Autonomie gesamtösterreichische Forschungspolitik machen, für die sie keine Verantwortung tragen können. In extremis könnte das Ergebnis dem deutschen Beispiel ähneln: Dort haben die autonomen Bundesländer keine Notwendigkeit eines Eingriffs in die Schließung der Geografie-Lehrstühle gesehen, die von einzelnen Universitäten durchgeführt wurden bis im gesamten deutschen Bundesgebiet dieses Fach nicht mehr vertreten war. Um derartige „Auswüchse“ autonomer Einzelentscheidungen von Universitäten zu verhindern, bedarf es eines geschärften Gesamtblicks im Verständnis des Hochschulplans²⁰ und des gesamtösterreichischen Universitätsentwicklungsplans²¹.

Die in immer mehr Staaten erfolgte oder erfolgende Schließung von Geistes- und Sozialwissenschaften könnte zu einer USP (unique selling proposition) Österreichs werden, dahingehend, dass sich Österreich als ein internationales Kompetenzzentrum für Geistes- und Sozialwissenschaften erfolgreich profiliert und positioniert: neben zahlreichen anderen weltweit sichtbaren Spitzenforschungen müsste die Zusammenarbeit der betroffenen österreichischen Universitäten im Bereich der Abstimmung, Fokussierung und Ressourcenbündelung gerade dieser Fachbereiche bei gleichzeitig englischsprachigen Studienangeboten zu einem weltweiten Zentrum der geistes- und sozialwissenschaftlichen Disziplinen führen. Die Dimension des Vorhabens erfordert dafür „strategische Partnerschaften“²² der betroffenen Universitäten, womit – zum Unterschied punktueller Kooperationen in einem Einzelprojekt – „großräumige“ und langfristige Ziele verfolgt werden könnten. Neben der verbesserten internationalen Sichtbarkeit würde es damit auch gelingen, ein Solitär in der internationalen Forschungslandschaft für diese Fachbereiche zu werden.

Nach welchen Kriterien die Geistes- und Sozialwissenschaften an den einzelnen Universitäten weiterentwickelt werden, bedarf aber auch einer österreichweiten Abstimmung durch das Ministerium²³. Die Neupositionierung der Geistes- und Sozialwissenschaften an den Universitäten ist deshalb keine alleinige Entscheidung der jeweiligen Rektorate. Sie haben aber die Herausforderung einer speziellen Governance für die Geistes- und Sozialwissenschaften und deren Qualitätsentwicklung. Ansonsten wäre wohl ein grundfinanziertes Personal in einer Größenordnung von mehr als 30% des Gesamtpersonals der betroffenen Universitäten nicht zu rechtfertigen. Positive Beispiele an österreichischen Universitäten sind dafür jedenfalls vielfach vorhanden. Diese aufgreifend könnten die neuen „Innovationsräte“ in ihrer Zusammenschau von Forschung und Lehre im internationalen Vergleich hier bereits neue Akzente setzen²⁴.

20 Österreichischer Hochschulplan. Der Gestaltungsprozess zur Weiterentwicklung des österreichischen Hochschulraums. 2011

21 Pichl, Universitäre Profilbildung, S. 196 verweist auf den zentralen Leistungsvereinbarungsparagrafen - § 13 Abs 2 1 lit a UG 2002 – der festhält, dass die „Universität ihre besonderen Schwerpunkte und Stärken und den daraus abgeleiteten und zur Zielerreichung vorgesehenen Ressourceneinsatz bekannt zu geben“ habe; vgl. auch Koch/Pichl (2014), Governance von Universitäten, S.32

22 Webler, (2013), Strategische Partnerschaften, S.20

23 Die durch das BMWF in Zusammenwirken mit den Betroffenen entwickelte Strategie für die Zukunft der Geistes-, Sozial- und Kulturwissenschaften in Österreich könnte als Grundlage für diese Abstimmung dienen

24 Burkert, Forschungsgovernance durch Forschungssenate? S.110-122

LITERATUR:

Brüggemeier, Martin (2002): Public Management. In: Hanft, Anke (Hrsg.): Grundbegriffe des Hochschulmanagements. Neuwied/Kriftel: Luchterhand. Seiten 377-383.

Burkert, Günther R. (2016): Forschungsevaluation in den Rechtswissenschaften – ein Sonderfall? Herausforderungen universitärer Governance in Österreich, Deutschland und der Schweiz (zum Druck abgeliefert)

Burkert, Günther R. (2016), Forschungsgovernance durch Forschungssenate? Zu einer Neuorganisation inneruniversitärer Verantwortung durch „Innovationsrat“ und Forschungsstärken, in: zfhr 15, S.110-122

Burkert, Günther R. (2011), Hochschulplanung. Vom quantitativen Ausbau zur qualitativen Stabilisierung, in: Kooperation und/oder Wettbewerb? Zum Verhältnis von universitärer und außeruniversitärer Forschung (= Tagungsband 2010 des österreichischen Wissenschaftsrates), S. 133-149

Campbell, David F. J. (2015). Forschungs- und Publikationslandkarte Sozialwissenschaften und Geisteswissenschaften Österreich 2013. Studie für das BMWFV. Wien.

Campbell, David F. J. / Elias G. Carayannis (2013). Epistemic Governance in Higher Education. Quality Enhancement of Universities for Development. (SpringerBriefs in Business.). New York, NY: Springer (<http://www.springer.com/business+%26+management/organization/book/978-1-4614-4417-6>).

Dunleavy, Patrick / Helen Margetts / Simon Bastow / Jane Tinkler (2006). New Public Management Is Dead—Long Live Digital-Era Governance. In: J Public Adm Res Theory (July 2006) 16 (3): 467-494.

Elias, Bernhard (2014): Analyse der Governance- und Management-Strukturen an österreichischen Universitäten – im Lichte der Autonomie der Universitäten seit dem UG 2002

Felderer, Bernhard / David F. J. Campbell (1994). Forschungsfinanzierung in Europa. Trends, Modelle, Empfehlungen für Österreich. Wien. Manz.

Frey, Bruno S./ Osterloh, Margit (2012): Managing Motivation: Wie Sie die neue Motivationsforschung für Ihr Unternehmen nutzen können (Schweizerische Gesellschaft für Organisation und Management)

FWF – Der Wissenschaftsfonds (2013). Jahresbericht 2012, S. 22-23;

FWF – Der Wissenschaftsfonds (2016). Jahresbericht 2015, S. 30-31.

Hellmann, Lisa (2011). „Kommunikationsmanagement an österreichischen Universitäten“ Strategische Kommunikation als Instrument der Profilbildung, Magisterarbeit, Universität Wien.

Jansen, Dorothea (2010). Von der Steuerung zur Governance: Wandel der Staatlichkeit? in: Handbuch Wissenschaftspolitik, 2010, VS Verlag für Sozialwissenschaften, 39 – 50.

Kaufmann, Alexander (2009). Differences in publication practices and the need for a more diversified research evaluation, in: Plattform fteval, Forschungs- und Technologieevaluation 32, 2009.

Eschenbach, Rolf / Figl, Ernst / Kraft, Ingeborg (2005): Handbuch für Universitätsmanager. Controlling, Finanzmanagement, Rechnungswesen, Berichtswesen, Qualitätsmanagement, Evaluierung. Wien: Linde, 2005.

Koch, Sabine/Pichl, Elmar (2014), Governance von Universitäten im Hinblick auf den gesamtösterreichischen Forschungsraum, in: zfhr

Pichl, Elmar (2011): Universitäre Profilbildung im Kontext des österreichischen Hochschulraums und des Universitätsgesetzes 2002, in: zfhr

Rhodes, R. A. W. (1996). The new governance: governing without government. Polit. Stud. XLIV 652-667.

Wall, Friederike, Martin Hitz, Erich Schwarz, Reinhard Stauber, Verena Winiwarter (2013). Das Projekt „Qualitätsgewichtete Publikationskennzahlen“ der Alpen-Adria-Universität Klagenfurt, in: IKN Klagenfurter Beiträge zur Interventionsforschung, Band 12, Dezember 2013 (http://www.uni-klu.ac.at/iff/ogi/downloads/Inhaltsverzeichnis_Band_12.pdf).

Webler, W.D. (2013): Strategische Partnerschaften zwischen wissenschaftlichen Einrichtungen, in: Fo 1+2, S. 20 - 22

Wegrich, Kai (2011). 'Post-New Public Management', in: Blanke, Bernhard; Nullmeier, Frank; Reihard, Christoph; Wewer, Christopher (Hrsg.): Handbuch zur Verwaltungsreform, 4. Auflage (VS Verlag)

Willke, Helmut (2014). Systemtheorie III: Steuerungstheorie: Grundzüge einer Theorie der Steuerung komplexer Sozialsysteme, UTB Verlag.

AUTOREN

GÜNTHER R. BURKERT

Bundesministerium für Wissenschaft, Forschung und Wirtschaft (BMWFV)

E: Guenther.Burkert@bmwfw.gv.at

DAVID F. J. CAMPBELL

Alpen-Adria-Universität Klagenfurt (AAU), Fakultät für interdisziplinäre Forschung und Fortbildung (IFF), Institut für Wissenschaftskommunikation und Hochschulforschung (WIHO)

E: david.campbell@aau.at

THORSTEN D. BARTH

Bundesministerium für Wissenschaft, Forschung und Wirtschaft (BMWFV)

E: Thorsten.Barth@bmwfw.gv.at

EVERYONE'S HAPPY? THE CASE FOR EXPECTATION MANAGEMENT IN COMMISSIONED EVALUATION PROJECTS

ANKE REINHARDT¹ AND ANTON GEYER²

Evaluation studies depend on collaboration. A funding agency commissions an evaluation project to an agent – usually an evaluation company or institute. Despite the agency's and the agent's best efforts in defining the aims and goal of the project, expectations are unlikely to fully converge.

Technopolis' recent evaluation of the International Research Training Groups (IRTG) programme, a funding scheme offered to scientists by the Deutsche Forschungsgemeinschaft (German Research Foundation, DFG), is a case in point. Using this example, the article sketches the motivations, interests and constraints of both parties when starting an evaluation project. It also points to possible difficulties, especially regarding the availability of data and the triangulation of methodologies during the course of the project. After discussing the results of the evaluation, the article describes how the partners dealt with these difficulties to make the evaluation project a joint success.

1. INTRODUCTION

Commissioning an evaluation project externally provides the contracting partner with an independent perspective on the evaluated subject. However, the delegation of the evaluation project generates a principal-agent problem in which principal and agent have partly diverging interests, speak partly different languages, and often fail to precisely communicate their expectations in advance. In addition, not every goal that the commissioning party wishes to achieve can be met, and often both parties fail to adjust expectations given the operational time constraints of the evaluation project. For the commissioning party it would be optimal to receive answers to all evaluation questions. For the agent it would be optimal if all aspects that can be analysed were of interest to the principal. To maximise the middle ground between these rationales, there is a need for the parties to jointly revise the evaluation project as it evolves.

This article discusses some of the challenges an evaluation project poses from the perspective of the commissioning party and the agent. Both parties aim to come up with a sound evaluation study that can be used for programme development (commissioning party) and as reference for other clients (agent). A successful project for the DFG would mean that the study is of high quality so that it is accepted by its stake-

holders and conclusive in regard to the evaluation questions. Ideally, the study gives insights into the impacts of DFG's funding beyond the immediate evaluation questions and would also serve as an example for novel methodologies and good evaluation practices. For the project manager at Technopolis a successful evaluation project is one in which the agent succeeds in putting the information, data and other evidence about the evaluation subject into a new perspective and by so doing, tests the expectations and assumptions the client has about the evaluation subject. The ideal evaluation study does not stop at analysing the programme and its impacts but points the client to those aspects of a programme that are difficult to see from a programme owner's perspective but critical for its understanding.

2. MOTIVATION FOR THE EVALUATION AND PROGRAMME DESIGN

With an annual budget of € 3.0 billion and almost 30 000 funded research projects in all scientific disciplines (Deutsche Forschungsgemeinschaft 2015a), the German Research Foundation (DFG) is the largest funding organisation for basic research in Europe.³ The core of funding is spent on individual grant programmes. However, a substantial share of total funding is allocated through so-called "coordinated programmes", which offer grants for specific science policy goals. For example, with these programmes, the DFG aims to shape the research profiles of research institutions, strengthen research cooperation within and between research institutions, facilitate international research cooperations and unlock the potential of highly qualified early career researchers.

The Research Training Groups (RTG) programme is one of these coordinated programmes. The RTG programme funds collaborative research agendas, carried out by a group of established (principal) researchers and implemented through doctoral projects. RTGs offer a structured qualification framework for doctoral training to provide young researchers with a working environment that facilitates their academic development and careers. Responding to demand voiced by the research community

¹ Deutsche Forschungsgemeinschaft (DFG), Kennedyallee 40, 53175 Bonn, Germany. E-Mail: anke.reinhardt@dfg.de. The opinions expressed in this article are those of the author and do not necessarily reflect the position of the DFG. We are grateful to Thomas Plümer for very valuable comments.

² Affiliation at the time this article was submitted for publication: Technopolis Group Austria, Rudolfsplatz 12/11, 1010 Vienna, Austria. Current affiliation: inspire research, Lange Gasse 65/16, 1080 Vienna, Austria. E-Mail: anton.geyer@inspire-research.at. The opinions expressed in this article are those of the author and do not necessarily reflect the position of Technopolis.

³ More information about the DFG is available on its website, www.dfg.de.

for additional support for international collaboration, in 1997 the DFG added an international collaborative line to the RTG programme, called International Research Training Groups (IRTG).

IRTGs are bilateral doctoral research programmes based at a university in Germany and a partner institution abroad. IRTGs have to meet all the criteria of RTGs to receive DFG funding. In addition, IRTGs must offer obligatory research visits for all doctoral candidates at the partner institution for a period of six to twelve months. The doctoral candidates have two academic supervisors: one at the German university and one at the partner institution abroad. The DFG also expects some degree of symmetry with regard to funding commitments and organisational structures for the IRTG at the partner institution.

IRTGs account for approximately 20 to 25 percent of all RTGs funded by the DFG. Between the start of the programme and June 2014 the DFG approved funding for 109 IRTGs.

It is common practice that the competent bodies at the DFG discuss, review and modify the funding criteria and procedures of its programmes, especially if the research landscape has changed considerably in aspects relating to the programme objectives. In the case of the IRTG programme line no such comprehensive evaluation took place before 2014. In that year DFG decided to commission an evaluation of the IRTG programme to collect information on the participants' experiences and to analyse the effects of the programme with regard to the facilitation and strengthening of international collaboration. In addition, the evaluation was intended to explore the "value added" of international cooperation in doctoral training programmes – generating conclusions which, it was hoped, would also be of interest to other DFG funding instruments that include an international collaboration dimension.

3. THE DFG'S EVALUATION PROCESS

For the DFG, evaluation studies kill two birds with one stone: by analysing the impacts of its funding programmes they provide the "strategic intelligence" needed for programme development and enhancement as well as informing the stakeholders⁴.

Over the years the DFG has implemented a standard evaluation process.⁵ The process invites and involves the use of both subject-specific and methodical expertise. In addition, the DFG has developed clear evaluation standards. Most importantly and to guarantee an independent assessment of the DFG's achievements (or lack thereof), evaluations have to be carried out by independent external providers. The DFG's statutory bodies are responsible for the design of the evaluation project and for the implementation of any actions that are based on the findings and recommendations of the evaluators. The rationale of this approach reflects DFG's nature and conception as a self-governing organisation that autonomously shapes its funding schemes based on the needs and preferences of its constituency, the scientific community in Germany.

4. ORGANISATION OF THE EVALUATION

About one and a half years before the IRTG evaluation started, the DFG set up an internal project team to design and supervise it. For large evaluation projects like this one, the DFG often appoints a steering committee which consists of members of the DFG's decision-making bodies, including academic members and representatives of the DFG's financial supporters, i.e. the German federal government and Länder governments. The inclusion of scientists in the project leadership group also guarantees that academic standards are met and provides methodological input. The involvement of steering committee members throughout the entire evaluation process helps the DFG decision-makers to competently discuss evaluation results and draw conclusions from the recommendations made by the evaluators. This process guarantees that evaluations inform programme development.

In a first step, the DFG project team drafted Terms of Reference for the evaluation study. The steering committee provided a detailed set of evaluation questions as input, thereby not only increasing the legitimacy and the visibility of the evaluation study, but also providing feedback and potentially broadening the goals of the evaluation.

During this planning stage, four objectives were identified:

1. (Theoretical) development of criteria for internationalisation and indicators
2. Description of internationalisation activities and internationalisation effects in IRTGs
3. Evaluation of the programme
4. Assessment of the strengths and weaknesses of the programme.

The steering committee had generated an extensive list of evaluation questions, which were of general interest but also reflected the specific perspective of each member of the committee on the funding scheme. All of these questions were valid and legitimate, but the DFG understood that the evaluation process could not provide answers to all issues raised. Since there was no "chair" of the steering committee, there was no mechanism in place to boil down the evaluation questions to only a few aspects. Additionally, the committee felt unsure which questions were feasible to answer with social science methodology in the timeframe foreseen for the study. For pragmatic reasons, the burden of identifying the most comprehensive set of questions that could be answered was shifted onto the contractor. The public tender procedure showed that this freedom allowed many different approaches to be taken. The DFG awarded the contract to carry out the evaluation study to Technopolis.

4 As well as evaluation studies the DFG uses a range of other instruments to provide strategic intelligence, in particular regular programme data monitoring, e.g. for DFG statutory bodies, review boards and member institutions. The DFG Funding Atlas, the most prominent monitoring instrument, for example, gives a comprehensive overview of third-party research funding provided by the DFG and other German and European funding organisations at universities and non-university research institutions in Germany (Deutsche Forschungsgemeinschaft 2015b).

5 See Reinhardt, Anke: Das Evaluationsvorgehen der Deutschen Forschungsgemeinschaft, *Zeitschrift für Evaluation*, 2/2013, p. 305-314.

5. OBJECTIVES AND KEY DIMENSIONS OF THE EVALUATION STUDY

Technopolis had the challenging task of dealing with the broad and ambitious objectives of the evaluation. While clients of evaluation studies frequently ask for the development of a theory-based analytical framework and guiding working hypotheses as the starting point for further data collection and analysis, the weight given to theory development in the Terms of Reference for the IRTG study was quite unusual for Technopolis. Theory development is mainly an academic endeavour that involves scientific research and builds upon various contributions from and interactions with the scientific community over time to allow for formulating, testing, reviewing and restating the underlying hypotheses of the theoretical model. It posed a challenge for Technopolis to meet the letter and the spirit of this requirement in the offer under the restrictions of a fixed-term, fixed-budget service contract. As academic research is not its core activity, Technopolis collaborated with the Department of Science, Technology, and Policy Studies (STePS) of the University of Twente as sub-contractor to bring in additional academic expertise. Since the time frame of the study (one year) required the parallel execution of tasks and work packages, Technopolis came up with a heuristic model of internationalisation as a starting point both for theory development and for data collection and empirical analysis. The idea was that the research partners at STePS would test and adapt the heuristic model through an (academic) literature review on international, inter-institutional collaboration in doctoral training.

The Terms of Reference also highlighted three key criteria to be covered by the evaluation with regard to the internationalisation effects of IRTGs:

1. The value-added (scientific, professional, personal) for the doctoral researchers in IRTGs
2. The scientific value-added (output), especially on the part of the principal investigators
3. The institutional value-added for the funded universities and research institutions.

These three dimensions were intended to stress the fact that IRTGs are concerned not only with improving the quality of doctoral training through structured doctoral programmes but also with strengthening research collaboration among established researchers and shaping the research profiles of the academic research institutions involved.

6. METHODOLOGICAL APPROACH AND IMPLEMENTATION

In the description of the indicative evaluation questions, the DFG suggested possible methods to be used in the evaluation. The Terms of Reference invited the tenderer to describe the methodological options and limits and explain the reasons for choosing certain approaches. Technopolis proposed a combination of methods for data collection and data analysis to address the evaluation questions as far as possible. The methodological design of the study included work packages on literature review, text analysis of IRTG proposals and reports, analysis of DFG programme funding and programme monitoring data as well as on-site

visits and personal interviews with doctoral researchers and established researchers at ongoing and terminated IRTGs. Telephone interviews with members of the university executives (vice-chancellors or pro-vice-chancellors), telephone interviews with established researchers at the collaborating research institutions abroad and bibliometric analysis of the publication output of doctoral researchers and established researchers in two selected areas of research (chemistry, biology/biomedical sciences) – for both IRTG participants and participants in standard RTGs – complemented the information base.

7. SUMMARY OF THE EVALUATION RESULTS

The analysis conducted by Technopolis supports the view that IRTGs contribute to the internationalisation of early career support and of the academic system as a whole, but differently and in a more subtle manner than the programme's objectives suggest. All three main stakeholder groups – doctoral researchers, established (principal) researchers and universities – profit from IRTGs. In the interviews, the doctoral researchers stressed the benefit of structured supervision and the training framework provided in IRTGs. The informal collaboration and day-to-day contacts with other doctoral researchers significantly contribute to the progress of their research. The feedback from doctoral researchers in IRTGs also indicates that they benefit scientifically and personally from the opportunities offered by international cooperation. According to the interviewees and supported by statistical data, the research visits abroad do not appear to prolong the time to degree.

The study showed that the established researchers and doctoral researchers often take a quite casual approach to the IRTG criterion to spend between six and twelve months for their research at the partner institution. While doctoral researchers in IRTGs do stay longer and more often abroad for research purposes than their classic RTG counterparts, the findings suggest that international collaboration between doctoral researchers has also been growing in standard RTGs over time.

Established researchers view IRTGs primarily as an attractive funding opportunity for their research as it comes with high internal and external reputation. They seek to leverage collaboration in order to access additional scientific resources for their work. The findings of the evaluation suggest that this happens far more often with other established researchers at their home institution than with established researchers at the partner institutions abroad. The bibliometric analysis shows that a fairly small proportion of established IRTG researchers publish with colleagues from their partner institutions, whereas joint publications between established IRTG researchers in Germany are far more commonplace and increase while collaborating in IRTGs. RTGs show a similar pattern with regard to the joint in-house publication activity among established researchers.

The universities' executives see in the IRTG programme an effective framework to increase internal collaboration among established researchers which in turn shapes the research profile of the university and increases its international visibility and scientific reputation. Some smaller universities actively communicated the IRTG "brand" in order to promote their research potential and increase their attractiveness for international research collaborations.

8. DEALING WITH DATA LIMITATIONS AND METHODOLOGICAL AND CONCEPTUAL CHALLENGES

The evaluation raised some practical and conceptual questions that had to be dealt with in the course of the project.

DATA LIMITATIONS

A small subset of the evaluation questions were to be answered by analysing the contents of IRTG proposals and reports and by secondary data analysis of DFG monitoring data on the RTG programme. The DFG introduced the electronic processing of proposals, which makes data and documents easily accessible, in 2000. A number of older IRTG files were missing and could not be included in the analysis.

The evaluation showed that the quality of the contents in the proposal and report documents varied widely, making it very difficult for Technopolis to use the information in a consistent fashion. The reports do not follow a standardised structure, nor do they require a consistent classification for reporting results and outcomes. This is because the DFG wants the reporting researchers to decide which specific (academic) results they deem important to emphasise and which to ignore, and not to burden them with a reporting scheme that is overly bureaucratic. However, without a consistent reporting scheme in place, text analysis of IRTG proposals and reports proved ineffective and provided only limited generalizable results. For future evaluations the DFG will examine the quality of the available documents in advance and assess whether the documents are suitable for qualitative and quantitative text analysis purposes.

The regular collection of information on all scientists of all status groups involved in the DFG RTG/IRTG programme ("DFG monitoring") started in 2005. Evaluation studies of DFG programmes that aim to look back further than 2005 present the DFG with difficulties in data availability. From 2005 till 2012, the DFG monitoring survey produced IRTG "snapshots" at the time of surveying but did not allow a joining of the dots in these snapshots over time to provide a more dynamic picture of the international activity of IRTG researchers. Since 2013 this data in RTGs and IRTGs has been collected using a significantly improved process.⁶ For the IRTG evaluation covering the years 1997 to 2014, the monitoring data was only of limited use.

USE OF BIBLIOMETRIC ANALYSIS

The evaluators employed bibliometric methods to identify and compare the number and share of international publications, including the number and share of co-publications between doctoral researchers and established researchers in IRTGs and RTGs. Bibliometric data was also used to identify co-authorships between IRTG researchers from German universities and their research partners abroad. Gathering bibliometric information often proved challenging, especially for doctoral researchers, who tend to have a limited track record and are difficult to identify in bibliographic databases. The resources required to address a fairly small number of specific explorative evaluation questions using bibliometric

analysis turned out to be substantial. Nevertheless, in the DFG project team's opinion the bibliographic analysis and the explorative approach implemented by Technopolis demonstrated its value. The results were very informative with regard to the general publication behaviour of researchers in RTGs and IRTGs in chemistry and biology/biomedical sciences, even though the findings were inconclusive as to whether there are significant differences between IRTG and RTG researchers with regard to their international publication activities.

THEORY DEVELOPMENT

The contractor made little progress in developing a coherent theoretical approach for studying internationalisation through IRTGs. While the underlying assumptions already spelled out in the offer proved very useful to understand the empirical evidence gathered on IRTGs, the scientific literature review offered little to develop the theoretical framework any further. The DFG project team felt that the analysis of literature remained rather narrow and might not have covered all potentially available literature. The results of the literature review thus fell short of the DFG's expectations. The review could not provide specific answers on what types of added value international collaboration in IRTGs actually generated. It also remained unclear which evaluation results can be generalized to other DFG programmes aiming at international collaborations.

For Technopolis' evaluation team the challenge with the literature review was that, even though there is abundant literature on the mobility of doctoral researchers and on international research cooperation in general, few papers appeared relevant in the context of IRTGs. IRTGs shape the research profiles of universities, pool research resources of established researchers, and deliver high-quality doctoral training programmes. In this respect, internationalisation may contribute to these main aims and outcomes of IRTGs, but internationalisation does not constitute an end in itself.

It turned out that theory development was too ambitious a goal for this evaluation study. The project manager at Technopolis understood the project as a service contract – which it legally was – that specified clear deliverables and hence kept the work very closely to letter of the Terms of Reference and the work plan described in the offer submitted to DFG. Progress in theory development might have profited from a far more open, (academic) research focused approach. This was not an option for Technopolis to follow, though, since (academic) research has unpredictable elements and might have distracted attention from fulfilling the tasks as described in the contract.

In hindsight it looks obvious that relying methodologically mainly on a literature review for theory development was inadequate in the context of this study. The most productive part of the literature review was a joint workshop at the DFG premises in Bonn, where the findings in the literature were discussed more specifically in the context of DFG's programmes and policies. The project might have benefited significantly from applying more of such interactive and explorative work elements in order to put the findings of the literature review into context.

9. USE OF THE EVALUATION AND CONCLUSIONS FOR PROGRAMME DEVELOPMENT

Based on the findings of the evaluation study, the DFG steering committee derived a set of recommendations for the future of the IRTG programme.

The evaluation demonstrated that IRTGs tend to deliver the expected outputs. At first glance, it seems that not many revisions to the funding scheme are required. Nevertheless, the steering committee called for more clearly defined mandatory criteria for IRTG funding focusing on the specific international dimension of the collaboration. Unlike classic RTGs (which allow and invite international cooperation), IRTGs ought to develop a joint research agenda, and a shared research programme that none of the individual research groups is able to deliver without the other institutional partner. Thus, the collaboration between the partner institutions should go well beyond a loose common research theme and frequent research visits by doctoral researchers. In the future, the DFG will put more emphasis on the “sine qua non” condition of a shared collaborative research agenda. Each partner will need to demonstrate its indispensable contribution to the research programme as a whole.

In this respect the DFG steering committee suggested developing a clearer definition of the proposal assessment criterion of “scientific complementarity”. The DFG steering committee also recommended maintaining the bilateral organisational structure of IRTGs with mandatory reciprocal financing on the partner side as a core element of the programme line.⁷

10. CONCLUSION

At the end of the evaluation project both the DFG and Technopolis felt happy with their collaboration, but not necessarily fully satisfied with the results. The evaluation demonstrated that IRTGs are fairly successful instruments and tend to deliver the scientific and policy objectives the DFG and its stakeholders wish to see delivered. In addition, some policy options could be derived from the evaluation study, which promise to alter the IRTG scheme in a way that strengthens the “international” aspect of the funding scheme.

The focus and the emphasis given to certain aspects of the IRTG programme line in the final report is mainly that of Technopolis. In its internal appraisal of the evaluation findings the DFG project team concluded that pinning down and limiting the study to fewer but critical evaluation questions in the Terms of Reference would have given the study more focus. As described before, different stakeholders (steering committee members, DFG project team) had different expectations, which made it impossible to meet them all.

With regard to operational aspects, both parties had the impression that a high degree of re-negotiation and re-planning was necessary during the course of the evaluation study. Clearly, the project would have benefited from a more frequent exchange between the DFG and Tech-

nopolis in order to effectively support the adaptation and convergence of expectations.

Nevertheless, the results of the IRTG evaluation were considered relevant and useful by both sides. The project profited from efficient project management and a high commitment to the project by the DFG and Technopolis alike. Even though data availability and data quality proved challenging, the bibliometric analysis in particular allowed new insights and generated new information on the effectiveness of IRTGs. In hindsight, the extensive use of face-to-face interviews (about 80 in total) was not cost-effective as sufficient theoretical and empirical saturation was reached at an early stage of the interview programme.

In the meantime, the DFG has managed to improve its data base. Future evaluation projects will be able to answer a larger set of questions. The DFG was aware that the abundance of ideas and indicative research questions collected internally in preparation for the tender procedure could not be possibly covered in a one-year service contract study with limited resources. If fully implemented, it would have shifted the character of the evaluation towards a multi-year scientific investigation of internationalisation and early career support in the framework of bilateral doctoral training programmes. The DFG’s strategy of asking the contractor to narrow down research questions raised in the tender of necessarily disappointed some of the stakeholders when they compared the original set of questions with the results of the evaluation. Overall, however, the explorative character of the IRTG evaluation did prove useful. Based on the lessons learnt in the IRTG study, in future evaluation studies and similar projects the DFG will provide room both for conceptual model development to identify the underlying mechanisms of a funding scheme and for methodological exploration and explorative data mining and data analysis. In the future the project manager will pay even more attention to integrating (academic) sub-contractors more effectively in the overall work structure of the evaluation to ensure timely feedback loops on the critical conceptual and operational issues of the evaluation study.

⁷ A written DFG response on the evaluation report (“DFG-Stellungnahme”, Deutsche Forschungsgemeinschaft 2015c) and a summary of the evaluation study (“DFG-Infobrief”, available in German and English, Heidler 2015) are publicly available on the DFG’s website www.dfg.de/dfg_profil/zahlen_fakten/evaluation_studien_monitoring/studien/studie_inter_grk

SOURCES

Deutsche Forschungsgemeinschaft (2015a): Jahresbericht 2014, Bonn. www.dfg.de/dfg_profil/jahresbericht/

Deutsche Forschungsgemeinschaft (2015b): Förderatlas 2015 - Kennzahlen zur öffentlich finanzierten Forschung in Deutschland, Bonn. www.dfg.de/sites/foerderatlas2015/

Deutsche Forschungsgemeinschaft (2015c): Stellungnahme der Deutschen Forschungsgemeinschaft zur Studie „Evaluation der Internationalen Graduiertenkollegs“, Bonn.

Geyer, Anton, Florian Berger, Tobias Dudenbostel und Brigitte Tiefenthaler (2015): Internationale Graduiertenkollegs. Studie im Auftrag der Deutschen Forschungsgemeinschaft (DFG), Bonn. www.dfg.de/dfg_profil/zahlen_fakten/evaluation_studien_monitoring/studien/studie_inter_grk

Heidler, Richard (2015): Schrittmacher der Internationalisierung: Ergebnisse der Evaluation der Internationalen Graduiertenkollegs. DFG-Infobrief 2/2015. www.dfg.de/dfg_profil/zahlen_fakten/evaluation_studien_monitoring/studien/studie_inter_grk

Reinhardt, Anke (2013): Das Evaluationsvorgehen der Deutschen Forschungsgemeinschaft, *Zeitschrift für Evaluation*, 2, pp. 305-314.

AUTHORS

ANKE REINHARDT

Deutsche Forschungsgemeinschaft (DFG)

E: anke.reinhardt@dfg.de

ANTON GEYER

inspire research

E: anton.geyer@inspire-research.at

LAUNCHING CAREERS AND BREAKING NEW GROUND: RESULTS FROM THE EVALUATION OF THE AUSTRIAN SCIENCE FUND'S FWF START & WITTGENSTEIN PROGRAMMES

SARAH SEUS, SUSANNE BÜHRER-TOPCU, NICLAS MEYER, EVA HECKL, MARIO MANDL AND KLAUS ZINÖCKER

This paper presents the results of the evaluation of the FWF's Start Programme and the Wittgenstein Award. It also summarises the methodology employed by the Fraunhofer Institute for Systems and Innovation Research ISI (Karlsruhe, Germany) and the Austrian Institute for SME Research (Vienna, Austria), who were commissioned by the FWF to conduct the evaluation.

The START Programme and Wittgenstein Award were added to the FWF's funding portfolio in 1996. The START Programme¹ is designed to enable highly promising young researchers from all disciplines to plan their research on a long-term basis and with sufficient financial security. By assuming responsibility for the establishment/expansion and management of a research group, the programme allows principal investigators to gain the qualifications necessary for leading positions in the science and research system. The Wittgenstein Award² aims to ensure that top-notch researchers who have reached outstanding achievements in all disciplines are provided with a maximum of freedom and flexibility in carrying out their research work. Through enabling freedom and flexibility, the award also aims at an extraordinary increase in recipients' research output.

Between 1996 and 2014, a total of 114 START projects with an overall funding volume of EUR 132.6 million were awarded. 16% of the projects were awarded to women. The approval rate in the START Programme is approximately 13% (based on 2002–2014 results), and the amount of grants awarded is up to EUR 1.2 million per project.

Between 1996 and 2014, the Wittgenstein Award was conferred upon 30 researchers (including 4 women), who received EUR 42.4 million in overall funding. During that period, a total of 278 researchers were nominated for the prize. The award carries an endowment of up to EUR 1.5 million.

In anticipation of the 20th anniversary of these two programmes, the FWF commissioned a team from the Fraunhofer ISI³ and the Austrian Institute for SME Research⁴ to conduct an evaluation of the START Pro-

gramme and the Wittgenstein Award. The evaluation project was led by Susanne Bühler. The goal of the evaluation was to assess the effects of these programmes and to lay the groundwork for a decision on how to run them in the future.

The evaluation team's assessment of the two programmes is highly positive, and the evaluators recommend that the two programmes be continued in their current forms:

- "A continuation of the START Programme in its present form is recommended without any reservations." (Seus, Heckl & Bühler, 2016)
- "The Wittgenstein Award seems tailor-made to its target group and has an added value to the scientific system in Austria." (Seus, Heckl & Bühler, 2016)

METHODS USED

The evaluation is based on a calibrated mix of multiple sources of data and both qualitative and quantitative methods, including online surveys for three different groups of researchers, a bibliometric analysis, interviews, case studies and a stakeholder workshop.

The core of the evaluation of the START Programme is formed by a comparison of the START grantees with a randomised control group, which was created on the basis of the bibliometric database Scopus. For every START grantee the evaluation team matched up to 3 other researchers in Austria that shared a set of structural characteristics, including gender, publication age, field of research and a similar research output at the point of the award of the START grants. The objective was to select researchers into the control group that had similar starting conditions as the individual START grantees.

In order to measure the discrete impact of the START programme an analysis of the research output of the START grantees and the control group was compared by bibliometric metrics, such as publication out-

1 <https://www.fwf.ac.at/en/research-funding/fwf-programmes/start-programme>

2 <https://www.fwf.ac.at/en/research-funding/fwf-programmes/wittgenstein-award>

3 <http://www.isi.fraunhofer.de/isi-en/index.php>

4 <http://www.kmuforschung.ac.at/index.php/en/>

put, citation rates and co-publication patterns. Those indicators were measured for both groups before, during and after START project was compared, thus combining cross-case comparisons with within-case comparisons. Furthermore, the researchers were invited to participate to an online survey in order to assess effects of the START programme on the career development of the funded researchers. The response rate to the survey for the START grantees was of 83%; for the control group a response rate of 27% was achieved.

In addition to the randomised selected control group, a group of applicants which were not selected for funding but whose project proposals were also deemed to be of an exceptionally high quality by the reviewers), was used as "comparison group". This comparison group also participated at the online survey with a response rate of 51%.

Bibliometric data and methods were also employed to identify and measure the emergence of new lines of research that were expected to be explored as a result of the Wittgenstein Award. First, the evaluation team examined whether the award winners changed their disciplinary focus by analyzing the disciplinary properties of the journals published in, of the co-authors and of the authors citing these publications over time. Starting from the Schumpeterian understanding of innovation as the recombination of existing means of production, or translated to the field of science, the recombination of existing lines of research, it can be expected that new avenues of research are likely to depart from disciplinary traditions and to pierce disciplinary boundaries. Furthermore, the evaluators employed text mining tools to examine whether potential changes in the award holders line of research can be observed in the key words used in the awardees' publications over time.

These methods were complemented by face-to-face interviews with eight START grantees and 7 researcher group members. Institutional stakeholders from the Ministry, the FWF and the START /Wittgenstein Jury have also been interviewed. Interviews were also conducted with seven Wittgenstein Award winners and project group members. Finally, a validation workshop was held with START grantees and Wittgenstein awardees as well as representatives from universities and the FWF.

START – CAREERS AND EXCELLENCE

The evaluators concluded that the START Programme "addresses all the relevant elements needed to promote scientific excellence and help develop the careers of young researchers. This is done for START grantees as well as the group members of a START project" (Seus, Heckl & Bühner, 2016). The funding amount and duration are also deemed appropriate.

START WORKS

From the FWF's perspective, one essential feature of the evaluation was the use of comparison and control groups in order to enable a clear assessment of the programme's effects. In this context, the evaluators come to the following conclusion: "The comparison with the control group shows that START grantees perform better throughout all indicators and also during the different periods analysed" and "The START Programme has contributed to strengthening the grantees' academic performance" (Seus, Heckl & Bühner, 2016).

- The START grantees' expectations of the programme are fulfilled to a high degree.

- The research output of the START grantees rises steadily over time, and it grows faster, more extensively and more visibly than that of the control group.
- START grantees are well connected internationally and occupy more central positions in international publication networks than the researchers of the control group.
- Many START grantees use the experience gained in their projects to establish new cooperation arrangements and to acquire new projects, and a majority of the grantees have also engaged in extensive academic communication activities.

START AS A MECHANISM FOR ENHANCING CAREER PROSPECTS AND DEVELOPMENT

The START Programme does not only have the objective to fund outstanding research but also to enable young researchers to establish themselves in the Austrian science and research system. Young researchers regard the START Programme as the career development mechanism offered by the FWF: "Consequently [START] is seen as the "top" career development instrument. The high prestige of the grant is furthermore conducive to the career development in the Austrian science system" (Seus, Heckl & Bühner, 2016). In consequence, nearly all of the researchers who have received a START grant in the past have also remained in the field of science and research. A majority of START grantees continue their careers at Austrian institutions, while approximately one-third of them move on to work at research institutions abroad. Last but not least, it is worth noting that 80% of START alumnae/i now hold a full professorship. The grantees themselves see the START Programme as an important vehicle for advancing their own careers in science and research, and 80% of the control and comparison groups are of the opinion that START is a suitable means of helping young researchers obtain a permanent position in Austria.

START AS A WAY TO PROMOTE SPECIFIC FOCUS AREAS AND GROUPS

START grants are not distributed evenly throughout the Austrian science and research system. The analysis of the regional and disciplinary distribution of approved START applications shows that researchers from certain disciplines and certain regions are more likely than others to apply and thus to be accepted to the START Programme. On that basis, the evaluators conclude that the Austrian science and research system is characterised by "hotspots of scientific excellence," which make more use of the START Programme than others. In this context, the evaluators specifically mention physics research at the University of Innsbruck or certain fields within the humanities (e.g. archaeology).

WITTGENSTEIN – SERENDIPITY AND FREEDOM

Wittgenstein is the FWF's "high-end" funding programme. It is the only initiative in the FWF's funding portfolio that follows the "fund people, not projects" principle, and its design includes strong "blue sky research" elements. The Wittgenstein Award is used to fund "high risk research [...] in which new fields can be explored and disciplinary borders can be crossed" (Seus, Heckl & Bühner, 2016).

The evaluation demonstrates that the Wittgenstein Award enriches the FWF's funding portfolio with an instrument that is predestined to

promote unconventional and ground breaking research. In this regard, the evaluators identify a number of key contributions: Wittgenstein has contributed to (i) the development of new methods, (ii) the exploration of new interdisciplinary research paths, and (iii) the development of new directions in research: "Not only high-level research has been conducted, but new avenues of research have been explored, scientific performance increased, and the visibility of research in the national and international science community enhanced" (Seus, Heckl & Bührer 2016).

The conditions that a research funding agency such as the FWF need, and does provide for this purpose are described by the evaluation team as "trust" and "flexibility". They emphasise trust because the Wittgenstein Award is not based on a specific research proposal and flexibility because it allows researchers to question the beaten paths in research and to reposition themselves in the science and research system. In this way, researchers can engage in activities, "which would not have been funded in the framework of a 'standard' application processes" (Seus, Heckl & Bührer, 2016). One advantage that should not be underestimated in this context is the fact that Wittgenstein Award winners are, for a certain period of time, relieved of the burden of submitting new grant proposals, meaning that award recipients can focus entirely on their research.

WITTGENSTEIN AND EARLY-STAGE RESEARCHERS

One possible criticism of the Wittgenstein Award is the fact that it may actually exacerbate the "Matthew Effect"; the award is given to Austria's most renowned researchers, whose projects are most likely to receive funding in any event. The evaluation team responds to this criticism by pointing out that the Wittgenstein Award can (also) be regarded as an instrument for supporting junior researchers. Wittgenstein Award recipients largely use their grants to integrate young researchers into their groups. The award winners in particular see the programme as an effective means of promoting young researchers. "*Most of the award is indeed spent on young researchers who get the possibility of working with outstanding researchers*" (Seus, Heckl & Bührer, 2016). For example, over 100 researchers have completed a PhD in the course of Wittgenstein projects, and several have been able to use their work on those projects as a springboard for their research careers.

START AND WITTGENSTEIN IN THE INNOVATION SYSTEM

In the competition for the best and brightest, a functional innovation system should be supported by a well-differentiated funding portfolio. A number of studies (e.g. Janger et al. 2013⁵) have demonstrated the importance of access to suitable third-party funding. The Wittgenstein Award and in particular the START Programme fulfil an important function in this context. For junior researchers faced with a lack of sound career prospects in the Austrian science and research system, the START Programme provides an attractive "starting package": "*The gathered evidence indicates that numerous START grantees would have left Austria without the START Programme in order to take a scientific position abroad*" (Seus, Heckl & Bührer, 2016).

The Wittgenstein Award, for its part, is the only programme in Austria that gives researchers an opportunity to realise "blue sky" ideas in their research. Together with their high visibility and the "critical mass" of funding provided by the FWF, these funding programmes constitute unique characteristics that distinguish the Austrian innovation system.

ERC & START

Since the inception of ERC Starting Grants, an EU-level funding mechanism has been available in parallel to the START Programme. This parallel offering raised a number of questions as to the legitimacy and sensibility of the START Programme. The FWF responded to these questions by requiring applicants to submit their applications to both funding programmes: All applicants who apply for a START grant are also required to submit the same project to the ERC.⁶ This dual submission requirement has been questioned by various institutional stakeholders at Austrian research institutions. However, the results of the survey among grant recipients and the reference groups reveal a more nuanced picture. Many regard the START Programme as an important supplement to the funding offered by the ERC. In fact, approximately 60% of the researchers surveyed would have submitted applications to the ERC and FWF even if it had not been required.

Both the specific and general results of the survey support the legitimacy of the START Programme: (i) The programme is held in high regard by grant recipients and non-recipients alike; it is not regarded as a "second-rate grant" by either of those groups ("*the START Programme is seen as a valuable and equivalent alternative to ERC by the research community in Austria*" (Seus, Heckl & Bührer, 2016)). (ii) START is viewed as a necessary enrichment of the funding portfolio for young post-doctoral researchers; it is considered necessary because funding opportunities at that career stage are already extremely limited due to the low approval rates of the ERC and FWF. (iii) According to the statements made in many interviews, the programme is an important location factor for the Austrian science and research system.

OUTSTANDING MARKS FOR PROGRAMME MANAGEMENT AND EXECUTION

Both the survey results as well as interviews with jury members and grant recipients paint "an overwhelming[ly] positive picture of the interaction with the FWF" (Seus, Heckl & Bührer, 2016). The evaluators confirm the grantees' high degree of satisfaction with the execution and management of the programme. In this respect, both the programme and the FWF are distinguished by their flexibility, which is regarded as a crucial basis for the work of outstanding researchers.

Evaluation of the START Programme and the Wittgenstein Award. Sarah Seus, Eva Heckl, Susanne Bührer; with the participation of Niclas Meyer, Sonia Conchi, Tobias Burst, Christina Schmedes and Barbara Sinnemann. Karlsruhe, 2016. DOI: 10.5281/zenodo.50610

5 Janger, J. et al. (2013): "Academic Careers in a Cross-country Perspective", Vienna.

6 In case that both application are successful, the ERC grant is accepted. The grantee can however further use the title „START grantee“ and gets an additional funding, amounting approx. to 1 year of supplementary funding to the ERC grant.

Annexes - Evaluation of the START Programme and the Wittgenstein Award. Sarah Seus, Eva Heckl, Sonia Gruber. Karlsruhe, 2016. DOI: 10.5281/zenodo.50610

AUTHORS

SARAH SEUS

Fraunhofer ISI

E: sarah.seus@isi.fraunhofer.de

SUSANNE BÜHRER-TOPCU

Fraunhofer ISI

E: Susanne.Buehrer-Topcu@isi.fraunhofer.de

NICLAS MEYER

Fraunhofer ISI

E: niclas.meyer@isi.fraunhofer.de

EVA HECKL

KMFA

E: e.heckl@kmuforschung.ac.at

MARIO MANDL

FWF

E: mario.mandl@fwf.ac.at

KLAUS ZINÖCKER

FWF

E: klaus.zinoecker@fwf.ac.at

WIFAS

EIN MODELL ZUR ABSCHÄTZUNG SOZIALER WIRKUNGEN MISSIONSORIENTIERTER FORSCHUNGSFÖRDERPROGRAMME AM BEISPIEL DER PERSONEN- UND GÜTERMÖBILITÄT¹

PETER KAUFMANN, LAURENZ WOLF, ALEX SCHUBERT UND ALEXANDER NEUMANN

ABSTRACT

Im Auftrag des Bundesministeriums für Verkehr, Innovation und Technologie (bmvit) wurden die Grundlagen zur Bewertung von sozialen Wirkungen von Mobilitätspolitiken erarbeitet. Mit der Unterstützung von Stakeholdern und Akademikern wurden hierzu Wirkungsmodelle für die Personen- und die Gütermobilität entwickelt. Diese konzeptionellen Grundlagen wurden in ein validiertes, empirisches Instrument übertragen, das für eine Umsetzung bereitsteht.

KEYWORDS:

social impact assessment, logic model, indicators, mobility, transport policy

Die Abschätzung der Wirkungen von Forschungsförderungsprogrammen im FTI-Bereich hat in den vergangenen Jahren zunehmend an Bedeutung gewonnen. Zum einen ist ein steigender Legitimationsbedarf in der Verwendung öffentlicher Mittel seitens der Politik zu beobachten, wodurch sich die Wirkungsanalyse (impact assessment) FTI-politischer Maßnahmen als Grundlage für eine evidenzbasierte Politikgestaltung etabliert hat. Zum anderen werden infolgedessen im Rahmen der wirkungsorientierten Haushaltsführung auch immer genauere Aussagen über das tatsächliche Wirkungsspektrum von FTI-politischen Förderprogrammen gefordert. Dies ist auch im Zusammenhang mit der verstärkten Berücksichtigung gesellschaftlicher Herausforderungen zu sehen.

Konzepte für die Wirkungsanalyse eher „klassischer“ FTI-politischer Zielsetzungen (Steigerung der Innovationsfähigkeit und Wettbewerbsfähigkeit, Verbesserung der Kooperationsstrukturen, insb. zwischen Wissenschaft und Wirtschaft etc.) sind dabei in der Vergangenheit bereits umfangreich untersucht und öffentlich diskutiert worden. In den letzten Jahren rückten auch Umweltwirkungen etwas stärker in den Vordergrund. Weniger untersucht wurden bislang die sozialen Wirkungen von FTI-politischen Programmen, wie sie etwa auch im Rahmen der Themenfelder „Personenmobilität“ und „Gütermobilität“ angesprochen werden bzw. als intendierte Effekte oder auch nicht-intendierte Zweiteffekte entstehen.

Vor diesem Hintergrund war das Ziel dieser Studie, ein konzeptionelles Modell zu entwickeln, auf Grundlage dessen sich programminduzierte, soziale Wirkungen von Förderprogrammen im Bereich der Mobilitätsforschung erfassen lassen. Folgende konkrete Fragestellungen waren handlungsleitend:

- Welche sozialen Dimensionen des Wirkungsspektrums sind für die Personen- und Gütermobilität relevant?
- Mit welchen Methoden (und Indikatoren) können diese sinnvoll erfasst werden, um in weiterer Folge zu einer Abschätzung von sozialen Effekten gelangen zu können?

Hierbei war auf alle potenziellen Wirkungen (360-Grad Blick) aufgrund einer möglichen oder tatsächlichen Umsetzung der Forschungsergebnisse und deren Folgewirkungen abzustellen. Dies wurde implementiert, indem in der internationalen Literatur eine Vielzahl an Konzepten für die Analyse von sozialen Wirkungen identifiziert wurde, die zur Einordnung der sozialen Wirkungsanalyse von FTI-Förderprogrammen nützlich sind und in die Entwicklung der konzeptionellen Basis einfließen.

Methodisch erfolgte die Bestimmung der Wirkungspfade und -dimensionen zum einen auf Grund von umfassenden Literaturscreenings über unterschiedliche Disziplinen, die sich mit sozialen Wirkungen (im Mobilitätsbereich) beschäftigten, und zum anderen durch Interviews mit langjährigen MobilitätsexpertInnen für ein „Proof of Concept“ der entwickelten Modelle. Parallel dazu wurde ein systematischer Überblick des Einbezugs von sozialen Effekten in Strategie- und ausgewählten Programmdokumenten FTI-politischer Maßnahmen im Mobilitätsbereich in ausgewählten Ländern erarbeitet. Anschließend wurde anhand von sieben Fallstudien von abgeschlossenen Projekten eine Prüfung auf Konsistenz, Vollständigkeit und Praktikabilität durchgeführt.

Von verschiedenen, vorrangig sozialwissenschaftlichen Forschungsdisziplinen wird der Begriff der ‚Social Impacts‘ (SI), also die Auswirkung von Handlungen (oder auch Unterlassungen) auf bestimmte oder zufällige Individuen und Gruppen, aufgegriffen und beforscht. Dabei werden SI oftmals nur am Rande der disziplinären Kerngebiete bzw. an der Schnittstelle zu angrenzenden Disziplinen untersucht. Die Breite der Begriffsanwendung erschwert eine präzise und interdisziplinäre Konzeptualisierung eines Terminus technicus. In der Psychologie wird der Begriff beispielsweise herangezogen, um den Einfluss von Personen/Gruppen/Massen auf Individuen und deren Verhalten zu erforschen (Latane 1981). Demgegenüber wird etwa in der Sozioökonomie, an der Schnittstelle von Wirtschaft und Soziologie, die wechselseitige Beeinflussung wirtschaftlicher Aktivitäten und gesellschaftlicher Prozesse beobachtet. Eine universelle, Wissenschaftsfelder übergreifende Anwendung und vor allem Konzeptualisierung des Begriffes ist daher nicht existent, vielmehr werden themenfeldspezifische Interpretationen angewendet. Wie die Literaturanalyse verdeutlichte, werden soziale Wirkungen oft entweder

zusammen mit den ökonomischen oder mit den ökologischen Wirkungen bzw. im Sinne eines Drei-Säulen-Modells der nachhaltigen Entwicklung diskutiert. Ökonomische und ökologische Effekte können demnach gleichzeitig auch soziale Effekte aufweisen bzw. können Wirkungen alle drei Dimensionen gleichzeitig tangieren. Dass trotz dieser wechselseitigen Abhängigkeiten die sozialen Wirkungen von Mobilität und Transport gegenüber den ökonomischen und ökologischen Effekten seltener im Fokus der Verkehrsforschung standen (Markovich und Lukas 2011; Jones und Lucas 2012), ist u.a. mit der Betroffenheit über verschiedene Disziplinen hinweg begründet, und mit der Ambiguität der Definition von sozialen Wirkungen. Damit verbunden ist auch eine teilweise weniger eindeutige Möglichkeit zur Messung der Wirkungen; die Quantifizierung von sozialen Wirkungen ist nicht immer möglich.

Wir folgen in dieser Studie einer breiten Definition von sozialen Wirkungen in Bezug auf Mobilität durch Geurs et al. (2009: 71), nämlich als *„...changes in transport sources that (might) positively or negatively influence preferences, well-being, behaviour or perception of individuals, groups, social categories and society in general (in the future).“* Diese breite Definition ist im Kontrast zur bisherigen Literatur mit speziellem Fokus auf Mobilität und Transport zu sehen, die soziale Wirkungen hauptsächlich als Verteilungswirkungen betrachteten (vgl. auch Forkenbrock und Weisbrod 2001, Social Exclusion Unit 2003, Lucas 2006, VCÖ 2009, Atkins 2010, Markovich und Lucas 2011, Lucas et al. 2013). Dieser eingegrenzte Fokus wird in jüngeren Publikationen durchaus kritisch hinterfragt und die soziale Dimension als eigenständiges Konstrukt und gleichwertig mit den ökonomischen und ökologischen Wirkungsdimensionen angesehen (z.B. Jones und Lucas 2012). In welcher Intensität die Wirkungen auftreten, ist in weiterer Folge von soziodemografischen Faktoren der Adressaten, aber auch von räumlichen Gegebenheiten und dem Zeithorizont ab-

hängig, welche die Wirkungsentfaltung beeinflussen. Die Beschreibung der einzelnen Dimensionen an den Schnittstellen ökonomische/soziale, ökologische/soziale aber auch ökonomisch/ökologische (Sub-)Dimensionen wäre somit ident bzw. ähnlich, die konkrete Definition einzelner Indikatoren jedoch unterschiedlich: je nachdem, ob bei Beschäftigungseffekten das abzuleitende Steueraufkommen interessiert (ökonomisch), oder die Wirkung auf die Psyche oder Physis der Individuen bzw. von gesellschaftlichen Gruppen (sozial).

KONZEPTIONELLE GRUNDLAGEN

Im Rahmen der Aufbereitung zum State of the Art konnten wichtige (politische) Initiativen der jüngeren Vergangenheit recherchiert werden, die jedoch meist auf leicht unterschiedliche Anwendungsfälle gemünzt sind, sowie leicht abweichende Begriffsdefinitionen und überlappende, aber dennoch im Detail unterschiedliche Konzeptualisierungen verwenden. Trotz dieses Potenzials für Konfusion kann viel von diesen Ansätzen gelernt werden.

Die Erfassung sozialer Folgewirkungen bzw. die unterschiedlichen Methoden ihrer Erfassung erstrecken sich über zahlreiche wissenschaftliche Disziplinen und politische Ebenen. Dafür steht eine stetig wachsende Anzahl an Richtlinien und Leitfäden zur Verfügung, welche die Konzeptualisierung, Erfassung und Bewertung von sozialen (Folge-)Wirkungen unterstützen sollen. In der nachfolgenden Tabelle wird ein Überblick gegeben, der unterschiedliche Zugänge und Methodenanwendungen auf die verschiedenen politischen Ebenen sowie für verschiedene Anwendungsfälle beinhaltet.

Bei der Entwicklung von Instrumenten zur Bewertung der sozialen Dimen-

	Internationale Organisationen und NGOs	EU und nationale Ebene
(Social) Impact Assessment	Social Impact Assessment: Guidance for assessing and managing the social impacts of projects (IAIA Guidance 2015) Policy Assessment: The State of Art (Adelle und Weiland 2012)	Impact Assessment Guidelines (EC 2009) sowie die Better Regulation Guidelines (EC 2015) Social impact measurement for social enterprises (European Union/OECD 2015) Methodological Guideline for Impact Assessment (Simsa et al. 2014) – FP 7 Projekt IMPACT Handbuch Wirkungsorientierte Folgenabschätzung (Bundeskanzleramt 2013)
Sustainability (Impact) Assessment	Guidance on Sustainability Impact Assessment (OECD 2010) Strategic Environmental Assessment: The State of Art (Tetlow und Hanusch 2012)	Assessing the Social and Environmental Impacts of European Research (EC 2005) Measuring Social Sustainability: Best Practice from Urban Renewal in the EU (Colantonio 2007) und Measuring Socially Sustainable Urban Regeneration in Europe (Colantonio et al. 2009, finanziert durch die Europäische Investitionsbank - EIP) Eine Vielzahl an leicht unterschiedlichen Zugängen auf nationaler Ebene...
Transport-related (social) impact assessment	Social Analysis in Transport Projects: Guidelines for Incorporating Social Dimensions into Bank-Supported Projects (World Bank 2006)	Guidebook for Assessing the Social and Economic Effects of Transportation Projects (Forkenbrock und Weisbrod 2001) USA Guidance for Transport Impact Evaluations - Choosing an evaluation approach to achieve better attribution (Hills und Junge 2010) UK

Legende: Diese stellen nicht nur offizielle Leitfäden diverser Organisationen dar, sondern auch Überblicksartikel aus der akademischen Literatur, tlw. von öffentlichen Organisationen finanziell unterstützt, ohne deren Bestätigung.

Tabelle 1: Richtlinien und Leitfäden zur Erfassung sozialer Wirkungen in unterschiedlichen Kontexten und auf unterschiedlichen politischen Ebenen

sion sowie bei der Ausweitung des Zielkatalogs auf das „Drei-Säulen-Modell“ der nachhaltigen Entwicklung (ökonomisch/ökologisch/sozial) wurde in der akademischen wie auch in der grauen Literatur vermehrt auf Herangehensweisen und Prozesse verwiesen, die eine multikriterielle Bewertung unterstützen bzw. die Resultate von qualitativen und quantitativen Methoden integrieren. Um auf verzerrte Bewertungen weitgehend zu kontrollieren, wird der Prozess vorgegeben und auf Transparenz Wert gelegt. Es wurde vermehrt an der Definition von Prozessschritten anstatt spezifischen Indikatorensets gearbeitet, wobei sich diese beiden Herangehensweisen nicht notwendigerweise ausschließen. Da in der Politik eine gewisse Messbarkeit von Effekten gefordert wird, könnte man an einem Set an Kernindikatoren, und spezifischen Ausprägungen von Indikatoren arbeiten, die jeweils eine Wirkungsdimension abzubilden versuchen. Jedenfalls sollten sowohl qualitative als auch quantitative Indikatoren integriert werden.

Als Rahmen für die Strukturierung des Bewertungsprozesses für Programmwirkungen könnte prinzipiell die Multikriterienanalyse, das Social Impact Assessment (IAIA 2015) oder das Impact Assessment der Europäischen Kommission (EC 2009 und 2015) herangezogen werden. Auch der Input aus dem Literaturkreis zu ‚Sustainable Development‘ über den Einbezug von Stakeholdern in den Bewertungsprozess im Falle von kaum vergleichbaren, multiplen Wirkungsdimensionen stellt eine überlegenswerte Strategie dar.

Aus diesem Literaturkreis folgern wir auch, dass die Berücksichtigung von sozialen Wirkungen bereits in der Projektentwicklung, beginnend mit der Ideenentwicklung erfolgen soll um deren positive Effekte zu maximieren. Damit sind Fördergeber gut beraten, bereits im Antragsstadium dementsprechende Signale an Projektwerber zu senden, damit soziale Wirkungen bereits bei der Projektentwicklung und -umsetzung mitgedacht werden.

Aufbauend auf die bisherige Analyse wurden zwei Wirkungsmodelle bzw. Interventionslogiken entwickelt.² Die Konzeptualisierung der Modelle wird anhand der auf den Folgeseiten dargestellten Modellvarianten zur Personenmobilität und der Gütermobilität veranschaulicht.

Die obere, grau hinterlegte Fläche umfasst dabei alle vorgelagerten Entscheidungen und Maßnahmen zur Anstoßung von verkehrlichen Effekten. Diese werden mit der verkehrspolitischen Zielformulierung initialisiert, und führen über die Ressourcenverteilung und den damit angestoßenen Maßnahmen zur Umsetzung von FTI-Politik. Die verkehrlichen Effekte entstehen über neue bzw. verbesserte Produkte, Dienstleistungen, (systemische) Prozesse, und Geschäftsmodelle in Unternehmen (und Forschungseinheiten). Die verkehrlichen Effekte manifestieren sich in erhöhter Zugänglichkeit oder gesteigerter Nutzung von Mobilitätsangeboten, bzw. verbesserter Verkehrssicherheit. Im unteren Teil der Abbildungen sind hingegen die möglichen sozialen Wirkungen erfasst, die aus der Umsetzung der verkehrlichen Maßnahmen und deren Nutzung bzw. Präsenz resultieren. Die Wirkungspfade sind in drei Hauptdimensionen (soziales Gefüge, Gesundheit, sozio-ökonomische Wirkungen) zusammengefasst und können in weiteren 16 (Personenmobilität) bzw. 14 Wirkungsdimensionen (Gütermobilität) unterteilt werden (weiße Felder). Anhand der orangenen Felder wird eine Kurzdefinition vorgenommen. Die Definition der Wirkungsdimensionen wurde für beide Modelle integriert vorgenommen. In 12 von insgesamt 18 Wirkungsdimensionen stimmen die Personen- und Gütermobilitätsmodelle überein.

Verteilungseffekte sind hier als Querschnittsdimension definiert, da (so-

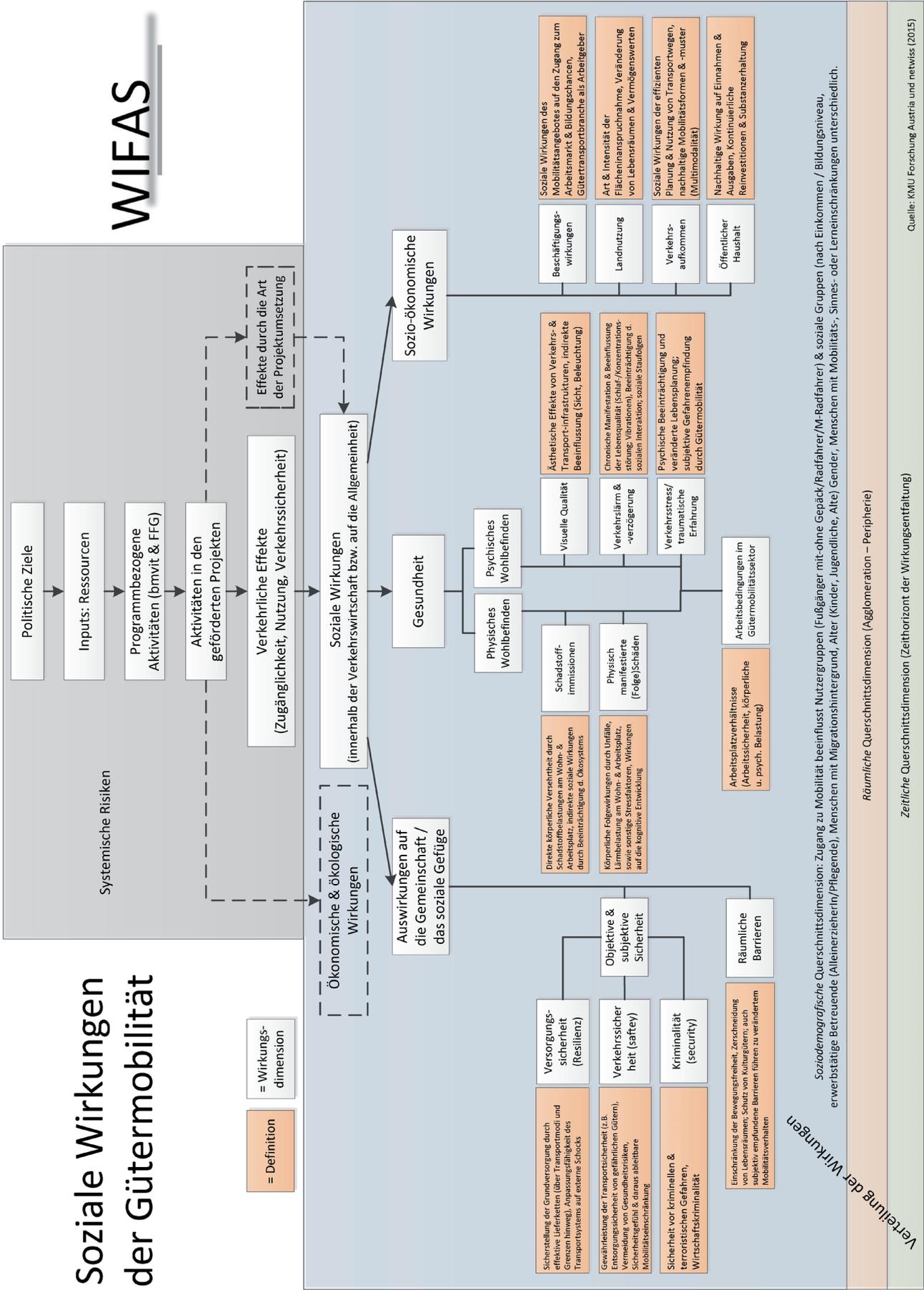
ziale, aber auch ökologische und ökonomische) Wirkungen potenziell immer unterschiedliche Betroffenheit und damit Verteilungseffekte aufweisen können. Wir definieren drei unterschiedliche Querschnittsdimensionen: die sozio-demografische, die räumliche, sowie die zeitliche Querschnittsdimension. Eine Kombination daraus kann zu kumulativen Effekten mit nicht-linearen Steigerungsraten führen.

ÜBERFÜHRUNG DER KONZEPTIONELLEN GRUNDLAGEN IN EIN BEWERTUNGSTRUMENT

Die Europäische Kommission strukturiert im Rahmen ihrer Impact Analysis (EC 2009, 2015) den spezifischen Prozess entlang dreier logischer Schritte. (1) *Identification of impacts*; (2) *Qualitative assessment of the more significant impacts*; (3) *In-depth qualitative and quantitative analysis of the most significant impacts*. Diese relativ einfache Vorgehensweise scheint für den vorliegenden Zweck der Bewertung von Programmwirkungen einen ausreichenden Rahmen darzustellen. In unserem Anwendungsfeld der Abschätzung von Wirkungen von FTI-Programmen mit multiplen Zieldimensionen wird aufgrund der Heterogenität der geförderten Projekte und deren unterschiedlichem Realisierungsstand auf diesen Ansatz aufgebaut, und mittels eines ‚qualitativen Scoping‘, d.h. in unserem Fall anhand eines standardisierten, modellgestützten Fragebogens, in einem ersten Schritt die Breite der Wirkungen und deren Gewichtung zueinander grob erfasst. Aufgrund der qualitativen Bewertung der emergenten Wirkungsmuster könnten in einem nächsten Schritt vertiefende Analysen der wichtigsten bzw. interessantesten Effekte und deren Wechselwirkungen erfolgen. Die Resultate auf Projektebene können über Themenfelder bis zur Programmebene aggregiert werden.

Soziale Wirkungen der Gütermobilität

WIFAS



FAZIT

Die Abschätzung von sozialen Wirkungen von Mobilitätsmaßnahmen soll auf der Projektebene erfolgen, und über die Themenfeldenebene hin bis zum Programm aggregierbar sein. Dieser Bottom-up Ansatz wurde gewählt, da die Heterogenität an Projekten im Förderprogramm sehr unterschiedliche soziale Effekte nach sich ziehen kann. Dies mag für Projekte im Rahmen der Personenmobilität wenig überraschen, aber auch das Modell zur Gütermobilität zeigt, in einem etwas reduzierten Ausmaß, immer noch ein breites Muster an möglichen Wirkungen. Diese Mehrdimensionalität sowie die schwierige Quantifizierung einiger sozialer Wirkungen und ihrer potenziell unterschiedlichen Ausprägungen in verschiedenen Kontexten (abhängig von der konkreten Ausgestaltung der Implementierung, räumlich, nach Adressaten bzw. soziale Gruppen, etc.) unterstützen einen Ansatz, der sowohl auf eine qualitative wie auch quantitative Erfassung von Wirkungen basiert. Zu einem ähnlichen Schluss kamen in den letzten Jahren auch verschiedene internationale Organisationen. Aufgrund von Erfahrungen mit der Quantifizierung, oder Nicht-Berücksichtigung, von weichen Faktoren (wie bspw. soziale Wirkungen im Bereich des gesellschaftlichen Zusammenhalts oder psychischen Wohlbefindens) wird nun eine ausgewogene Herangehensweise empfohlen. Dies reicht von einem „*focus beyond numbers*“ (OECD 2010), über ein qualitatives Scoping zur Identifikation der relevantesten sozialen Wirkungen und der Verwendung eines Mix aus qualitativen und quantitativen Methoden für vertiefende Analysen (EC 2009 und 2015), bis hin zur Stellungnahme einer Expertenkommission der EC und OECD, die empfiehlt, dass „*adopting a measuring process rather than imposing specific metrics or indicators*“ der vielversprechendere Ansatz bei Interventionen mit multiplen Dimensionen darstellt (EC/OECD 2015: 5). Damit wird unterstrichen, dass eine reine Quantifizierung von Effekten, speziell im sozialen Wirkungsspektrum, nicht zielführend ist. Da jedoch in der Realpolitik eine Abschätzung von Effekten gefordert wird, kann man einerseits an einer Definition von Prozessschritten, und andererseits an einem Set von qualitativen und quantitativen Kern-Indikatoren arbeiten, die jeweils eine Wirkungsdimension abzubilden versuchen. Diese beiden Vorgehensweisen schließen sich nicht aus.

Der potenzielle Nutzen der Abschätzung nicht nur von ökonomischen und ökologischen, sondern auch von sozialen Effekten liegt auf der Hand. Einerseits können aufgrund der Analyse vergangener Projekte und deren realpolitischer Umsetzung positive wie auch negative Wirkungen identifiziert, und in den thematischen Feldern aggregiert werden. Hier sollte sich aufgrund von ‚weißen Flecken auf der Landkarte sozialer Wirkungen‘ potenzieller Handlungsbedarf für die Zukunft ableiten lassen. Diese Informationsbasis kann zweitens dazu genutzt werden, um im Zuge von Ausschreibungen die Ziele eines Calls u.a. an sozialen Kriterien festzumachen, und damit bei ProgrammteilnehmerInnen ex-ante eine stärkere Berücksichtigung von sozialen Kriterien bei der Projektentwicklung zu erwirken, bzw. völlig neue Projektideen zu induzieren.

Ein logischer, nun folgender Schritt könnte sein, die hier erarbeiteten Grundlagen für Mobilitätsprogramme umzusetzen. Im Zuge dessen kann ein Survey von vergangenen Projekten deren Wirkungsspektrum flächendeckend erheben. Die Ergebnisse liefern einen Überblick an Wirkungsmustern, die Grundlage für vertiefende Analysen und/oder Informationsgrundlagen für zukünftige Ausschreibungen darstellt.

Ein Vergleich der Berücksichtigung von sozialen Wirkungen in der Mobilitätspolitik anderer Länder brachte zu Tage, dass abgesehen von

Verteilungswirkungen, außer in England, meist nur wenige soziale Effekte in der Mobilitätspolitik mitgedacht werden. Zu diesem Befund kann aus den Interviews mit ExpertInnen im Zuge dieser Studie ergänzt werden, dass die Beschäftigung mit dem Thema ‚soziale Wirkungen‘ auf breite Zustimmung trifft. Der einheitliche Tenor war, dass Österreich hier Themenführerschaft übernehmen sollte. Diese Option ist auch im Zusammenhang mit der Umsetzung der FTI-Strategie des Bundes und dessen mobilitätsrelevanter Handlungsfelder zu sehen: soziale Aspekte wie Selbstbestimmung & Gesundheit, Arbeit, Lernen, Freizeit, Wohnen und Raumentwicklung/Verkehr & Infrastrukturen, sowie Umwelt/Ressourcen (siehe Bauer et al. 2015), betreffen auch die FTI/Technologiepolitik.

AUSZUG AUS DER LITERATUR

Adelle, C., Weiland, S. (2012). Policy assessment: the state of the art. *Impact Assessment and Project Appraisal*, 30(1): 25-33

ASSIST (2013). Assessing the social and economic impacts of past and future sustainable transport policy in Europe. Deliverable D2.1 of the FP 7 project MOVE/FP7/265381/ASSIST, lead contractor for this deliverable: ProgTrans AG, Basel

Bauer, H. et al. (2015). Koordinierte FTI-Strategie und –Maßnahmen für Mobilität und Lebensqualität vor dem Hintergrund des demografischen Wandels. Roadmap der FTI-AG3 „Lebensqualität und demografischer Wandel“ zur Umsetzung der FTI-Strategie der Bundesregierung.

Bundeskanzleramt (2013). Handbuch Wirkungsorientierte Folgenabschätzung, Arbeitsunterlage Version 1.2, Wien

Colantonio, A. Dixon, T. (2009) Measuring Socially Sustainable Urban Regeneration in Europe. Oxford Institute for Sustainable Development (OISD), funded by the European Investment Bank University Research Sponsorship

European Commission (2005), Assessing the Social and Environmental Impacts of European Research (Ricci Report). Directorate-General for Research, Brussels

European Commission (2009). Impact Assessment Guidelines. SEC(2009) 92

European Commission (2015). Better Regulation Guidelines. Commission Staff Working Paper, SWD(2015) 111 final, sowie die dazu gehörende Better Regulation ‘Toolbox’

European Commission (2015). Indicators for promoting and monitoring Responsible Research and Innovation. Report from the Expert Group on Policy Indicators for Responsible Research and Innovation. European Commission, DG Research, Luxembourg

European Commission/OECD (2015). Policy brief on social impact measurement for social enterprises. Luxembourg

Forkenbrock, D.J., Weisbrod, G.E. (2001). Guidebook for Assessing the Social and Economic Effects of Transportation Projects. NCHRP Report 456 for the Transportation Research Board, Washington

Geurs, K. T, Boon, Wouter and Van Wee, Bert (2009). 'Social Impacts of Transport: Literature Review and the State of the Practice of Transport Appraisal in the Netherlands and the United Kingdom Transport Reviews, 29(1): 69–90

IAIA (2015). Social Impact Assessment: Guidance for assessing and managing the social impacts of projects. International Association of Impact Assessment

Jones P., Lucas, K. (2012). The social consequences of transport decision-making: clarifying concepts, synthesising knowledge and assessing implications. *Journal of Transport Geography* 21, 4-16

Lucas, K. (2012). Transport and Social Exclusion: Where Are We Now? *Transport Policy* 20, 105-113

Markovich, J., Lucas, K. (2011). The Social and Distributional Impacts of Transport: A Literature Review, Working Paper No. 1055, Transport Studies Unit, School of Geography and the Environment, University of Oxford

Marsden, G., Kimble, M., Nellthorp, J., Kelly, Ch. (2010). Sustainability Assessment: The Definition Deficit. *International Journal of Sustainable Transportation*, 4(4), 189-211

OECD (2009). Interpreting OECD Social Indicators. In: *Society at a Glance: OECD Social Indicators*, Paris

OECD (2010). *Guidance on Sustainability Impact Assessment*. Paris

Simsa, R., Rauscher, O., Schober, C., Moder, C. (2014). Methodological Guideline for Impact Assessment, FP 7 project IMPACT, working paper No. 01/2014

Vanclay, F. (2002). Conceptualising social impacts. *Environmental Impact Assessment Review* 22: 183-211

World Bank (2006). *Social Analysis in Transport Projects: Guidelines for Incorporating Social Dimensions into Bank-Supported Projects*. Social Analysis Sector Guidance Note Series, Social Development Department of the World Bank

AUTOREN

PETER KAUFMANN

KMU Forschung Austria / Austrian Institute for SME Research

E: p.kaufmann@kmuforschung.ac.at

LAURENZ WOLF

KMU Forschung Austria / Austrian Institute for SME Research

E: l.wolf@kmuforschung.ac.at

ALEX SCHUBERT

Netwiss OG

E: alex.schubert@netwiss.at

ALEXANDER NEUMANN

Netwiss OG

E: alexander.neumann@netwiss.at

BOOK REVIEW: MARK ZACHARY TAYLOR - THE POLITICS OF INNOVATION

STEFAN PHILIPP

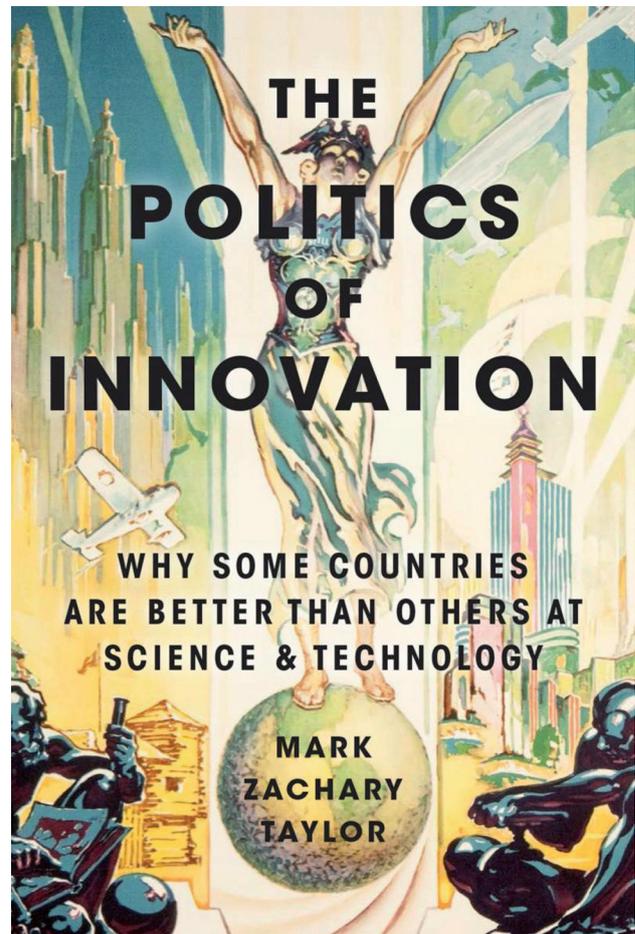
In his book *"The politics of innovation - Why some countries are better than others at science and technology"* (Oxford Press, 2016), Mark Zachary Taylor (Associate Professor of Political Science, Georgia Institute of Technology) investigates what he calls Cardwell's riddle, after the British historian Donald Cardwell who surveyed a thousand years of developments in science and technology (S&T) in the 1970s. Cardwell made a simple observation that has since been called Cardwell's Law. After analysing the innovativeness of the younger human history Cardwell concluded that nationality should not matter because:

"no nation has been very creative for more than a historical short period. Fortunately as each leader has flagged there has always been, up to now, a nation or nations to take over the torch."

While Cardwell was writing about different societies across the grand sweep of history, Taylor is concerned with the time since World War II, a period with increasingly good data availability, allowing him to compare national differences in S&T performance both over time and across different countries within the same time period. He recognises that even in this short period of time something like Cardwell's Law continues to describe national differences in S&T. For example, nowadays a small club of nations dominates the global creation and production of S&T. Among a few others, this elite is comprised by the United States, Japan, Germany, Taiwan, Sweden, Korea and Israel. Some of these leading countries did not innovate much back in 1945 or did not even exist. In the meantime, other countries with long traditions of wealth, democracy, industrialisation and sometimes even S&T production today only show a moderate output if they innovate at all.

For Taylor these national disparities pose a crucial puzzle as national borders and distance should not matter in the twenty-first century when it is easier than ever to copy and transmit knowledge internationally, when human and financial capital as well as equipment move globally at historically low costs. Additionally, a consensus about the advantages of S&T for national economies has been reached and therefore countries today should have similar S&T capabilities. Yet vast differences in S&T outputs are found around the globe. To solve this riddle Taylor not only scrutinises how nations innovate but also why they do so. Separating these two aspects clearly allows him to tackle the problem from different sides.

Many scholars argue that the mix of national policies and institutions, aiming to mitigate the market failures conjoined with S&T, explain these differences and that "good" policies and institutions are the key to succeed as an innovative nation. This is the starting point for Taylor. He



investigates the effects that the five pillars of innovation (property rights, research and development (R&D) subsidies, education, research universities and trade policy) have on national innovation output (measured in patent applications weighted by forward citations). Taylor shows that these five pillars statistically explain up to 90% of the national innovation rates but at the same time the used indicators do not explain well why some countries are high performing or not. There are simply too many countries with "good" institutions and policies that fail to innovate and the other way round.

Therefore, Taylor adds more domestic institutions and policies into his investigation, but takes a broader "marco" approach. In this section he discusses the National Innovation System (NIS) theory and finds that research in this field added insights to the complexity of innovation pro-

cesses but at the same time failed to produce a general theory. In addition, the NIS approach poses a problem for theorising and testing as cases exist where the same institution or policy is attributed with different effects on innovation rates in different countries. Other institutional theories investigated by Taylor include the effects of democracy, political decentralisation and "Varieties of Capitalism". He shows that none of these approaches adequately explains success in S&T across time and space.

In the following chapter Taylor's approach shifts from empirically testing theories towards looking at recent cases of success and failure to identify crucial factors. As success stories the cases of Israel, Taiwan and Ireland are picked while Mexico serves as an example of relative failure. From these examples he learns two things. First, there is no silver bullet. The case studies do not share a single "best" institution or policy design. It is therefore not important which particular institution or policy governments use, but that the market failures get addressed in an appropriate way depending on the country's precondition. Therefore, governments have considerable freedom to customise national strategies for improving their nation's innovation rate. Second, the case studies point towards factors that according to Taylor, are often overlooked in the research of national innovation systems: social networks, industrial clusters and standards. The success stories reveal that informal social networks, both domestic and international, play important roles in determining national S&T achievements and that innovation is not only hampered by market failures but also by network failures. Clusters and standards both help to mitigate the threat of network failures.

This leads Taylor to the next question in this logic chain: if solving market failures and setting up social networks are the keys to national success in S&T, then why do some countries perform these tasks so well while others fail? Taylor explains and successfully tests this phenomenon by the distributive nature of both technological change and the institutions and policies that support S&T since innovation can create winners and losers within a society. Therefore, they can cause new domestic conflicts or excite traditional ones. When these conflicts arise, public support for S&T could wane and political elites may seek to slow it down. At the end, domestic tensions act as a force to slow and obstruct political support for technological change. Taylor further shows that external security concerns can counteract these domestic resistances to innovation. On these finding, he builds his theory of creative insecurity, which he defines as the condition of feeling more threatened by external hazards than by domestic rivals. It is the positive difference between the threats of economic or military competition from abroad and the danger of political-economic rivalries at home. He shows that creative insecurity generally motivates broad and sustained support for S&T. The bottom line of this theory is that countries for which external threats are relatively greater than domestic tensions should have higher national innovation rates than countries for which domestic tensions outweigh external threats. The greater and longer the creative insecurity is the more are people willing to risk their political and economical resources on rapid progress in S&T.

Concluding, the book provides readers from all backgrounds and levels of expertise a comprehensive overview to the debates over national S&T competitiveness. It is directed at readers with different professional backgrounds who want to gain a better understanding of innovation and what governments can do to help or hamper it. It is written in a clear and straightforward style whereat in-depth discussions on different definitions and wordings, the history of innovation measurement and on

measures, data and sources are provided in the annexes. It synthesizes over fifty years of theory and research on national innovation rates, discusses and tests political and economic theories and the latest findings on the politics of S&T. One might be irritated by the approach to find one model to explain differences in national innovation regimes over time and space but if succumbing to this quest, Taylor is taking one on an interesting journey through the world of innovation theories. Did Taylor solve Cardwell's riddle? Of course not. But in *The Politics of Innovation* he delivers a sound examination of the most important pieces of this puzzle and is pointing further research in a certain direction. Even for readers not convinced by Taylor's concept of creative insecurity the book still offers solid evidence of the need for new explanations of national innovation rates.

AUTHOR

STEFAN PHILIPP

*Austrian Platform for Research and Technology Policy Evaluation (fteval)
& Centre for Social Innovation (ZSI)*

E: philipp@fteval.at
philipp@zsi.at

EVALUATION WITH SOUND JUDGEMENT: CONFERENCE ON “EVALUATION OF RESEARCH FUNDING: REVIEW, INSIGHT, OUTLOOK”

ANKE REINHARDT, UTA SASS AND RICHARD HEIDLER

ABSTRACT

In recent decades, the evaluation of research funding has become established as a learning tool for funding organisations in Germany and many other countries. Most funding agencies have set up evaluation units and have acquired extensive experience in conducting evaluations, analysing data, utilising indicators, and discussing and implementing evaluation results. At the conference “Evaluation of Research Funding - Review, Insight, Outlook”¹ on 16 June 2016 in Bonn, participants discussed current trends and new challenges relating to the data basis, the measurement of impacts, the developments in methodology and the professionalisation of research funding evaluation in relation to the German discourse.

1. EVALUATION OF RESEARCH FUNDING: A DEBATE IN GERMANY AND EUROPE

The evaluation of research funding has become firmly established in Europe. The first Research Assessment Exercise conducted in the UK in 1986 was one of the key origins of research evaluation in Europe.

With a view to Germany, the evaluations of the ‘Blue List institutes’ beginning in 1979, the evaluations of the academy institutes in the former GDR in 1990/1991 and the system evaluation of the Max Planck Society (MPG) and the German Research Foundation (DFG) in 1999 (Internationale Kommission 1999) are important early milestones

Since then, the field of evaluation has undergone a notable process of professionalisation as well as a substantial progress of knowledge. This includes, for example, the shift from traditional peer review mechanisms to ‘informed peer review’, where appraisals are based on empirical

evidence (Wissenschaftsrat 2011) and the growing use and refinement of bibliometric indicators (De Rijcke et al. 2016).

The organisational and institutional setting in which evaluation takes place has also been enlarged through the establishment of evaluation units within funding organisations. However, a survey of 20 European funding organisations conducted by the European Science Foundation in 2012 found that most only employ a small number of people in their evaluation units (between one and a maximum of 20 in one case) and that these individuals also fulfil other roles (ESF 2012b). By way of example, this can be demonstrated for German funding agencies. At the DFG, for instance, evaluation is part of Information Management unit; at the VWS it is combined with Internal Auditing and Fund Allocation Auditing; and at the DAAD it is attached to Monitoring and Studies.

Especially for those responsible for evaluation it is extremely beneficial to share information and learn from others through various national and international cooperation arrangements, platforms and networks, and at conferences. For example, northern European countries have formed Nordforsk² to carry out joint and comparative evaluations among other activities. Many other platforms focus on sharing experiences³ or developing common standards or guidelines. The recommendations of the European Foundation Centre, for instance, provide very practical advice on carrying out evaluation studies (EFC 2010).

The Member Forum of the European Science Foundation (ESF)⁴ (2007-2009 and 2010-2012) was another of these multinational dialogue initiatives. Evaluation staff from 33 research performing and research funding organisations discussed issues and challenges of common relevance and published their results (inter alia, ESF 2012a and ESF 2012b)⁵. They identified four developments as current key drivers of research evaluation: 1. Evaluation models of the effects research funding gain in sophistication; 2. Electronic gathering and storage of data creates a much better basis for evaluation; 3. Methodologies advance rapidly and 4. Increasing professionalisation in the field. The ESF members also formulated the

1 German title: „Evaluation der Förderung von Wissenschaft und Forschung – Rückblick, Einblick, Ausblick“

2 <https://www.nordforsk.org/en>

3 For example, the G7 Heads of Research Organisations Working Group on Research Assessment or the European RTD Evaluation Network (EUevalnet) administered by the European Commission. https://ec.europa.eu/research/evaluations/index_en.cfm?pg=network

4 <http://www.esf.org/coordinating-research/mo-fora.html>

5 See also <http://www.esf.org/coordinating-research/mo-fora/evaluation-of-publicly-funded-research.html>

recommendation to develop common evaluation and data standards, to learn from each other, to build networks and to represent common interests (ESF 2012a).

2. BACKGROUND TO THE CONFERENCE

The contribution of the conference “Evaluation of Research Funding - Review, Insight, Outlook” has to be seen against the background of these various topics and the debate taking place both in Germany and in Europe. The conference specifically addressed the challenges outlined by the ESF. The presentations and discussions focused on four topic areas: professionalisation, academisation, new data sources, and the meaning and boundaries of evaluation. The conference was organised by the Alexander von Humboldt Foundation (AvH), the German Academic Exchange Service (DAAD), the Volkswagen Foundation (VWS) and the host organisation, German Research Foundation (DFG).⁶ The idea for the event arose from the informal exchange which had been taking place between these organisations for several years.

The about 70 participants, who mainly came from Germany and other German-speaking countries, included evaluation experts from funding and evaluation organisations, principals of evaluations, and members of the research community. The conference was divided into three parts. After a review of important milestones in the development of research and research funding evaluation in Germany and other countries, the participants turned their attention to current trends and future challenges. In the final part, representatives of the DAAD, the DFG, the VWS and the Swiss National Science Foundation (SNF) provided an insight into the evaluation practice of their organisations.

3. THE CONTRIBUTION OF THE CONFERENCE TO THE DEBATE

The speakers addressed the above-mentioned development trends and future challenges relating to the evaluation of research funding from their individual perspectives. The presentations stimulated sometimes controversial discussion.

In his presentation, Thomas Widmer from the University of Zurich discussed evaluation practice in depth and asked whether evaluation can be considered a profession and/or a discipline. Widmer believes there is a case for both in a broader sense, while making clear that evaluation is more of a transdiscipline than a traditional academic discipline (see also Widmer 2013). He regards the science of evaluation as a driver of evaluation development separate from commercial interests. The same applies to the question of evaluation as a profession: Measured against the five classic criteria defined by Harold Wilensky – full-time work, curriculum and training schools, professional association, official licensing or certification requirements and formal code of ethics – evaluation does not satisfy the requirements for a profession in a narrower sense (see Wilensky 1964). In a broader definition, Widmer does however recognise a serious, high-quality and therefore ‘professional’ approach to evaluati-

on. With regard to evaluation practice, a clear link to research and sound expertise in a given subject area will, in Widmer’s view, contribute to the status of evaluation as a profession and a discipline.

The professionalisation of evaluation at a more ‘practical’ level was also the focus of the presentation given by Katharina Warta, chair of the Austrian Platform for Research & Technology Policy Evaluation (fteval). The networking opportunities, training courses, and discussion forums and publications offered by fteval support the professionalisation of the field in Austria. In her talk she explained the instruments and impacts of the fteval platform, which was established in 1996 as a loose network and now comprises 21 member organisations from politics, research, consulting and research funding. In a self-evaluation, fteval noted that its mission in Austria had evolved from the establishment of an evaluation culture to its perpetuation (Warta and Philipp 2016). One significant challenge for the ongoing development of substance and method and the avoidance of conflicts of interest will be the further internationalisation of the evaluation community in Austria.

Politics and administration are continually seeking meaningful new indicators to assess research and its impact on society. At the same time, more and more data is becoming available thanks to new technical developments. In his presentation, Lutz Bornmann, a sociologist of science and expert in scientometrics at the Administrative Headquarters of the Max Planck Society in Munich, discussed the conflict between the exploratory analysis of information from new data sources and the development of genuinely meaningful and useful indicators. As an example, he quoted a study which investigated the reliability and usefulness of so-called ‘alternative metrics’, such as mentions of research results on social media or networks (e.g. Twitter or Mendeley) and in politically relevant publications (e.g. from UNESCO or the WHO). So far, the results are not encouraging: there is no obvious correlation between citation-based indicators and social media mentions (Barthel et al. 2015), nor is a significant number of scientific studies (taking the example of climate research) mentioned in politically relevant documents (Bornmann et al. 2015). It also remains unclear through which channels societal impact actually takes place and, accordingly, how the citation of a scientific paper in a ‘tweet’ should be assessed (Bornmann 2016). Nevertheless, Bornmann noted that the current development of indicators of the societal impact of research does exhibit parallels with the earlier establishment of other indicators, such as bibliometric measurements. It is therefore possible to learn from this experience.

Stefan Kuhlmann from the University of Twente discussed the ‘fixation on indicators’ manifested in the growing use of quantitative indicators to measure research performance, such as impact factor, Hirsch index or Shanghai ranking. The emergence of new means of gathering more data and assessing it more systematically has actually caused a negative dynamic within the field of evaluation. Kuhlmann added critically that both researchers and decision-makers have a certain self-interest in these ‘simplifications’. Kuhlmann promoted the ideas to involve stakeholders in research evaluations and to recognize evaluation as a space for openness and learning (Kuhlmann et al. 2015).

6 The organising team consisted of Bea Knippenberg, Daniel Lechner (both DAAD), Martina May, Meike Olbrecht (both AvH), Richard Heidler, Anke Reinhardt (both DFG) and Uta Sass (VWS).

In addition to the presentations, the conference offered insights into the evaluation practice of various funding organisations and enabled participants to discuss the organisations' views on the development of research evaluation. Bea Knippenberg (DAAD), Anke Reinhardt (DFG), Uta Sass (VWS) and Katrin Milzow from the Swiss National Science Foundation (SNF) explained how the developments discussed are affecting the activities of their funding organisations. Growing professionalisation can be seen, for example, in the fact that in all four funding organisations, evaluation has been expanded into a field of activity in its own right. After a developmental phase, in some cases a very rapid one, a consolidation phase is now taking place. The institutions have developed evaluation concepts tailored to their own needs (see Sass 2010, Reinhardt 2013) and established evaluation as a practice and 'good governance'. As a result of this learning process evaluation concepts, once established, are adapted to counter undesirable developments and shift emphases. For example, evaluations are not carried out in a routine manner but only in response to specific needs. They are initiated by management bodies and/or programme units or in some cases are agreed on when a programme is first established. In all organisations the observations and recommendations resulting from evaluations are discussed in the organisation's internal bodies and critically reflected in internal statements. They are therefore taken into account in strategy discussions without triggering an automatic implementation process.

4. CONCLUSIONS

The conference addressed the challenges of evaluation identified at European level and offered plenty of starting points for in some cases controversial discussion. Kuhlmann and Widmer diagnosed that in spite of a certain degree of professionalisation, unintended 'pathologies' and undesirable incentives have emerged in evaluation which should be actively countered. Bornmann noted a need for further research in relation to the development of new indicators and alternative metrics for research performance and its societal impacts, as previous attempts have been unconvincing. With regard to new approaches in evaluation, Kuhlmann promoted the involvement of stakeholders. Warta expressed support for further networking of evaluation practitioners and the internationalisation of the field. Widmer also argued for targeted evaluations which are only carried out when the commissioning party actually has the will and the capacity for change processes. To sum up, the speakers stressed that all stakeholders involved in evaluation should apply sound judgement to approach evaluation projects.

Many of the discussions reflected the recommendations of the Member Forum of the European Science Foundation. These have already influenced evaluation practice within the funding organisations. The conference provided participants with a solid knowledge base and an opportunity for a lively exchange of views and networking, which will hopefully contribute to a more intensive dialogue between the organisations in future. Some organisations will continue to discuss the topics mentioned here at European level within the framework of Science Europe⁷, an alliance of research and funding organisations.

REFERENCES

- Barthel, Simon; Tönnies, Sascha; Köhncke, Benjamin; Siehdnel, Patrick; Balke, Wolf-Tilo** (2015). What does Twitter Measure?: Influence of Diverse User Groups in Altmetrics. Paper presented at the meeting of the JC DL, 2015.
- Bornmann, Lutz; Haunschild, Robin; Marx, Werner** (submitted, 2015): Policy documents as sources for measuring societal impact: How is climate Change research perceived in policy documents. <http://arxiv.org/abs/1512.07071>
- Bornmann, Lutz** (2016). What do altmetrics counts mean? A plea for content analyses. *Journal of the Association of Information Science and Technology*, 67(4), pp. 1016-1017.
- De Rijcke, Sarah; Wouters, Paul F.; Rushforth, Alex D; Franssen, Thomas P; Hammarfelt, Björn** (2016): Evaluation practices and effects of indicator use — a literature review. *Research Evaluation*, 25(2): pp. 161-169.
- European Foundation Centre** (2010): Research Forum Evaluation Guidelines 3: How to evaluate. Brussels http://efc.issuelab.org/resource/how_to_evaluate
- European Science Foundation** (2012a): Evaluation in Research and Research Funding Organisations. European Practice. Strasbourg. http://www.esf.org/fileadmin/Public_documents/Publications/mof_evaluation.pdf
- European Science Foundation** (2012b): A guide to Evaluation Activities in Funding Agencies. Strasbourg. <http://www.esf.org/coordinating-research/mo-fora/evaluation-of-publicly-funded-research.html>
- Internationale Kommission** (1999): Forschungsförderung in Deutschland - Bericht der internationalen Kommission zur Systemevaluation der Deutschen Forschungsgemeinschaft und der Max-Planck-Gesellschaft, Hannover. http://www.dfg.de/download/pdf/dfg_im_profil/reden_stellungnahmen/archiv_download/evaluation_forschungsfoerderung_99.pdf
- Kuhlmann, Stefan; Edler, Jakob; Ordóñez-Matamoros, Gonzalo; Randles, Sally; Walhout, Bart; Gough, Clair; Lindner, Ralf** (2015): Responsibility Navigator, Karlsruhe/Germany (Fraunhofer ISI). <http://responsibility-navigator.eu/>
- Reinhardt, Anke** (2013): Das Evaluationsvorgehen der Deutschen Forschungsgemeinschaft, in: *Zeitschrift für Evaluation*, 02/2013, pp. 305-314.
- Sass, Uta** (2010): Evaluationen – Instrumente der Qualitätssicherung der VolkswagenStiftung, in: *Zeitschrift für Evaluation*, 01/2010, pp. 143-152.

Warta, Katharina; Philipp, Stefan (2016): Was bringt die Plattform fteval und zu was hat sie es gebracht? Rückblick und Ergebnisse einer Mitgliederbefragung, in: fteval Journal No. 41, pp. 5-14.

Widmer, Thomas (2012): Evaluation im transdisziplinären Spannungsfeld, in: Soziologische Revue 35 (1), pp. 53-62.

Wilensky, Harold L. (1964): The Professionalization of Everyone?, in: American Journal of Sociology, Vol. 70, No. 2 (Sept., 1964), pp. 137-158.

Wissenschaftsrat (2011): Empfehlungen zur Bewertung und Steuerung von Forschungsleistung, Köln.
www.wissenschaftsrat.de/download/archiv/1656-11.pdf

AUTHORS

ANKE REINHARDT

Deutsche Forschungsgemeinschaft (DFG)

E: anke.reinhardt@dfg.de

UTA SASS

Volkswagen Foundation

E: sass@volkswagenstiftung.de

RICHARD HEIDLER

Deutsche Forschungsgemeinschaft

E: richard.heidler@dfg.de

PREVIOUS ISSUES OF THE FTEVAL JOURNAL FOR RESEARCH AND TECHNOLOGY POLICY EVALUATION:

ISSUE 41

WAS BRINGT DIE PLATTFORM FTEVAL UND ZU WAS HAT SIE ES GEBRACHT? RÜCKBLICK UND ERGEBNISSE EINER MITGLIEDERBEFRAGUNG
Katharina Warta, Stefan Philipp

EVALUATING THE PERFORMANCE OF INNOVATION INTERMEDIARIES: INSIGHTS FROM THE EXPERIENCE OF TUSCANY'S INNOVATION POLES
Margherita Russo, Annalisa Caloffi, Federica Rossi, Valentina Fiordelmondo, Stefano Ghinoi



EVALUATING CLUSTER INITIATIVES: AGENCY, ORGANISATION, FUNCTIONALITY, PERFORMANCE
Michael Rothgang, Bernhard Lageman

AGENCIFICATION REVISITED A CASE FOR THE EVALUATION OF FFG AS AN AGENCY
Rupert Pichler

A REGIONAL FUNDING ORGANIZATION FOR BASIC RESEARCH? UNCONVENTIONAL MISSIONS AND THEIR EVALUATION
Grit Laudel and Michael Strassnig

ZWISCHENEVALUIERUNG DER VOM BMWF BEAUFTRAGTEN REGIONALEN KONTAKTSTELLEN – ODER DIE EVALUIERUNG VON INTERMEDIÄREN STRUKTUREN IN IHREM KONTEXT
Barbara Good

EVALUIERUNG VON „YOUNG SCIENCE“ – EINEM PROJEKT AN DER SCHNITTSTELLE ZWISCHEN SCHULE UND WISSENSCHAFT
Caroline Manahl, Susanne Dobner

WHEN TWO INTERMEDIARIES AND ONE FUNDING MINISTRY MEET FOR A COMMON GOAL: THE CASE OF THE FEDERAL AUSTRIAN INITIATIVE “EVOLVE” FOR THE CREATIVE INDUSTRIES
Alfred Radauer AND Tobias Dudenbostel

AUSWAHLVERFAHREN VON FORSCHUNGSFÖRDERUNGSORGANISATIONEN IM INTERNATIONALEN VERGLEICH
Peter Biegelbauer, Thomas Palfinger

„EVALUATION VS. EDUCATED GUESS“ PROJEKTRÄGER ALS UNTERSCHÄTZE QUELLE FÜR DIE BEWERTUNG VON FÖRDERPROGRAMMEN
Jan Wessels, Christiane Kerlen UND Sonja Kind

ISSUE 40

ZWISCHENEVALUIERUNG DES FÖRDERSCHWERPUNKTS TALENTE
Eva Heckl, Laurenz Wolf

DAS SCHRÖDINGER PROGRAMM DES FWF EVALUIERT. BEMERKUNGEN VON EVALUATORINNEN UND AUFTRAGGEBERINNEN
Susanne Bührer, Niclas Meyer, Barbara Zimmermann, Klaus Zinöcker

DER ORGANISIERTE SERENDIPITÄTSEFFEKT ALS EVALUIERUNGSANSATZ
Fritz Ohler

DIE ROLLE VON BEGLEITENDER EVALUATION UND BEGLEITFORSCHUNG IN DER FTI-POLITIK. EINE BESTANDSAUFNAHME.
Peter Kaufmann, Iris Fischl, Sonja Sheikh

BIBLIOMETRIC CASE STUDIES AS COMPLEMENTARY METHOD FOR S&T PROGRAMME EVALUATIONS
Klaus Schuch, Isabella Wagner, Alexander Degelsegger

WAS KÖNNEN FALLSTUDIEN IM RAHMEN VON EVALUATIONEN LEISTEN? ANSPRUCH UND REALITÄT DER ANWENDUNG.
Iris Fischl, Peter Kaufmann

EVALUIERUNG UND VERÄNDERUNG: LERNEN UM ZU WISSEN, LERNEN UM ZU HANDELN?
Karin Grasenick, Magdalena Pierer, Markus Gruber



EDITORIAL BOARD

Rebecca Allison, **Technopolis UK**; Balázs Borsi, **Eszterházy Károly College**; Elke Dall, **Centre for Social Innovation**; Michael Dinges, **Austrian Institute of Technology**; Leonid Gokhberg, **National Research University Higher School of Economics**; Wolfgang Polt, **Joanneum Research**; Andreas Reinstaller, **WIFO**; Klaus Schuch, **fteval** (chief editor); Michael Stampfer, **WWTF**; Lena Tsipouri, **University of Athens**

AUSTRIAN PLATFORM FOR RESEARCH AND TECHNOLOGY POLICY EVALUATION (fteval)

c/o ZSI – Centre for Social Innovation GmbH

Linke Wienzeile 246, A-1150 Vienna

T +43 1 495 04 42 - 79

F +43 1 495 04 42 - 40

E office@fteval.at

W www.fteval.at

DESIGN

carotte.at

Blindengasse 35/4/97, A-1080 Wien

E caroline@carotte.at

PRINT

REMA-Print-Littura Druck- u. Verlags Ges.m.b.H

Neulerchenfelderstraße 35, A-1160 Vienna

AUSTRIAN PLATFORM FOR RESEARCH AND TECHNOLOGY POLICY EVALUATION

Platform fteval members:

Federal Ministry of Science, Research and Economy (bmwfw), Federal Ministry of Transport, Innovation and Technology (bmvit), Austrian Institute of Technology (AIT), AQ Austria, Austria Wirtschaftsservice (AWS), Christian Doppler Research Association (CDG), convelop cooperative knowledge, Austrian Research Promotion Agency (FFG), Austrian Science Fund (FWF), Institute for Advanced Studies (IHS), Austrian Institute for SME Research (KMFA), Joanneum Research, Ludwig Boltzmann Society (LBG), Austrian Institute of Economic Research (WIFO), Austrian Council for Research and Technology Development, Austrian Academy of Science (ÖAW), Technopolis Austria, Vienna Science and Technology Fund (WWTF), Wirtschaftsagentur Wien, Ein Fonds der Stadt Wien, Centre for Social Innovation (ZSI)



AUSTRIAN PLATFORM
for Research and Technology Policy Evaluation



c/o ZSI – Centre for Social Innovation GmbH
Linke Wienzeile 246, A-1150 Vienna

The fteval Journal for Research and Technology Policy Evaluation is a forum to discuss current evaluation practices in the field of research, technology and innovation policy.

T +43 1 495 04 42 - 79
F +43 1 495 04 42 - 40
E office@fteval.at
W www.fteval.at

ZVR-Zahl: 937261837
ISSN-Nr. 1726-6629
©Vienna 2016