PLATTFORM fteval
Forschungs- und Technologieevaluierung

Special Edition

New Frontiers in Evaluation

Nr. 28
April 2006

BUNDESMINISTERIUM FÜR BILDUNG, WISSENSCHAFT UND KULTUR,
Minoritenplatz 5, A -1014 Wien
Mag. Markus Pasterk markus.pasterk@bmbwk.gv.at
Adir.Reg.R. Ursula Suda ursula.suda@bmbwk.gv.at

BUNDESMINISTERIUM FÜR VERKEHR, INNOVATION UND TECHNOLOGIE,
Renngasse 5, A-1010 Wien
Dr. Rupert Pichler
rupert.pichler@bmvit.gv.at

BUNDESMINISTERIUM FÜR WIRTSCHAFT UND ARBEIT,
Stubenring 1, A-1010 Wien
Mag. Sabine.Pohoryles-Drexel
Sabine.Pohoryles-Drexel@bmwa.gv.at

ARC Systems Research,
Donau-City-Straße 1, A-1220 Wien
Mag. Petra Wagner
petra.wagner@arcs.ac.at

FWF-Fonds zur Förderung der wissenschaftlichen Forschung,
Weyringergasse 35, A-1040 Wien
Dr. Rudolf Nowak
novak@fwf.ac.at

KMU Forschung Austria,
Gusshausstrasse 8, A-1040 Wien
Dr. Sonja Sheikh
s.sheikh@kmuforschung.ac.at

WIFO – ÖSTERREICHISCHES INSTITUT FÜR WIRTSCHAFTSFORSCHUNG,
PF 91, A-1103 Wien
Mag. Hannes Leo
hannes.leo@wifo.ac.at

TECHNOPOLIS,
Prinz Eugen Straße 80/12, A-1040 Wien
Mag. Leonhard Jörg
leonhard.joerg@technopolis-group.com

JOANNEUM RESEARCH, Institut für Technologie- und Regionalpolitik,
Wiedner Hauptstraße 76, A-1040 Wien
Mag. Wolfgang Polt wolfgang.polt@joanneum.at
Mag. Michael Dinges michael.dinges@joanneum.at

WWTF – Wiener Wissenschafts-, Forschungs- und Technologiefonds,
Währinger Straße 3 / 15a, A-1090 Wien
Dr. Michael Stampfer michael.stampfer@wwtf.at
Mag. Klaus Zinöcker klaus.zinoecker@wwtf.at

CHRISTIAN DOPPLER FORSCHUNGSGESELLSCHAFT (CDG),
Weyringergasse 33/3, 1040 Wien
Dr. Laurenz Niel
niel@cdg.ac.at

AQA – Österreichische Qualitätssicherungsagentur
Liechtensteinstraße 22a, 1090 Wien
Mag. Alexander Kohler
alexander.kohler@aqa.ac.at

ZIT – Zentrum für Innovation und Technologie GmbH,
Ebendorferstrasse 4/DG, A-1010 Wien
Robert Mayer-Unterholzner
robert.mayer@zit.co.at

AUSTRIA WIRTSCHAFTSSERVICE GESELLSCHAFT MBH,
Ungargasse 37, 1030 Wien
Dr. Sonja Hammerschmid
s.hammerschmid@awsg.at

FFG-ÖSTERREICHISCHE FORSCHUNGSFÖRDERUNGSGESELLSCHAFT
Canovagasse 7, A-1010 Wien
Dr. Dorothea Sturn dorothea.sturn@ffg.at
Mag. Klaus Schnitzer klaus.schnitzer@ffg.at

Rat für Forschung und Technologieentwicklung
Donau-City-Straße 1, 1220 Wien
Mag. Wolfgang Neurath
w.neurath@rat-fte.at

Rupert Pichler, Michael Dinges, Klaus Zinöcker

Preface

Welcome to this special issue of the Platform – Research and Technology Policy newsletter, which sends you on a journey to some of the key research questions to be discussed at the "New Frontiers in Evaluation" conference hosted by the Platform together with the Ministry of Transport, Innovation and Technology (bmvit) and the Austrian Science Fund (FWF) at the Radisson SAS Palais Hotel, April 24th – 25th 2006, Vienna, Austria.

Policy makers face new challenges at both the European and national levels as the division of labour among European, national, regional and private science financiers increases and has led to a debate about interdependencies among the resulting portfolios of research projects provided by the respective decision makers. With the European Research Council (ERC) a new funding instrument is about to enter the European stage while still little is known about the interaction of different funding instruments, e.g. what a well balanced funding mix actually means with regard to the strategic orientation of research policy and the role of evaluation within this endeavour.

In the meantime – while policy makers are struggling with funding portfolios – institutions such as the evolving ERC need to contribute to the further development of the effectiveness of R&D funding instruments via the implementation of evaluation and selection processes suitable for today's complexities. Peer review already seems to have been pushed beyond its limits in its decision-making capacity, thus leaving room to review peer review and its roles in the science system once again. The newsletter highlights allocation practices of several research councils and discusses the role of evaluation (and review) practices with regard to higher education financing in an article about the Research Assessment Exercise in the UK. The U.S. Programme Assessment Rating Tool (PART) serves as a reference point to compare US and European evaluation approaches.

Not only the design of allocation mechanisms and funding portfolios but also persistent perceptions and strong beliefs on "how things are going" in the field of science and technology policy constitute challenges for evaluation practitioners, funding agencies and policy makers.

The papers in this issue highlight some of the challenges people in the sphere of Science and Technology policy are confronted with and describe practices how the problems are currently dealt with. They all share the same spirit: the belief that evaluation can help to deal with these challenges.

Authors:

Rupert Pichler
Bundesministerium für Verkehr, Innovation und Technologie
Renngasse 5, 1010 Wien, Austria
rupert.pichler@bmvit.gv.at

Michael Dinges
Joanneum Research, Institut für Technologie- und Regionalpolitik - InTeReg
Wiedner Hauptstr. 76, 1040 Wien, Austria
michael.dinges@joanneum.at

Klaus Zinöcker
fteval - Plattform Research and Technology Evaluation
c/o WWTF, Währinger Strasse 3 /15 a, 1090 Vienna, Austria
zin@fteval.at
www.fteval.at

Rupert Pichler

The ERC and Policy Makers' Expectations: Evaluation as a Change Agent

Basic Research Funding at a European Scale: A New Paradigm

Substantial changes are looming at the horizon of Europe's research funding landscape as we are about to witness a fairly significant shift of paradigms within the rationale of EU research funding. The idea of creating a European research council (ERC) rapidly gained momentum from when it was first launched and started to command increasing fascination ever since. This leaves the impression that something must have been missing in Europe's research funding portfolio. At the same time the amount of attention the ERC receives sheds some new light on the validity of justifications of community RTD funding efforts hitherto used.

Until quite recently, Europe's major weakness seemed to have been duly identified: While being good at basic research, Europe was said to be poor at exploiting the results stemming from that research so that, consequently, industry tended to be unwilling to invest adequately in RTD. This was largely the underlying rationale for a series of EU Framework Programmes, made most explicit in preparation of FP 5 (Caracostas and Muldur 1997)[1]. However, such conclusions were not reached on the basis of an analysis of the European situation in its own right. Instead, the pole position of the US as it was perceived by then served as a benchmark for European wishful thinking.

This picture was, of course, nothing new. For decades the gap between the US and Europe had been one of the drivers of RTD policy actions in Europe, leading not only to a whole range of different concepts and approaches at national levels, but also eventually to the introduction of RTD funding as a community competence (Guzzetti 1995). Obviously, the US leadership in most technology based industries still makes Europeans shiver right into the 21st century so that their perception appears to be chiefly guided by the huge scientific and technological opportunities US businesses seem to make use of every day.

This logic appealed to both national and community RTD policy levels while it was assumed that particularly the exploitation of research results required European scale action, as it were the economies of scale that were thought to be scared off by the fragmentation of opportunities

1 The title of this book ("Society – the endless frontier") was obviously chosen in reference to Vannevar Bush's (1945) book, ironically proving not to have had the impact the authors may have wished to achieve.

in Europe. Therefore, FP 5 and FP 6 mainly provided incentives for a concentration of industrial RTD as well as exploitation efforts through bigger and more complex projects.

Up to the late 1990s the division of labour between the Community and the member states remained relatively unchallenged since – according to the great-opportunities-and-economies-of-scale logic – funding of basic research and of innovation oriented applied research (often geared at SMEs) seemed to be more appropriate for the member states' scope of action.

It was not until the re-invention[2] of a European Research Area (ERA) that this division of labour was really challenged and, as a consequence, the analysis of Europe's weaknesses began to be questioned. The reason for that is that by proposing the concept of ERA the original clear-cut division between actions under European community responsibility and actions under national responsibility was judged as a problem. Now, the fragmentation of effort across all RTD policy fields in Europe was being made responsible for lacking efficiency and, therefore, lagging behind obviously well-focussed US efforts.

For two reasons, this meant also addressing the issue of basic research funding. Firstly, the ERA concept aims at motivating the member states to address EU goals within their own policies and strategies. As basic research funding is among the biggest (and most exclusive) national stakes, subjecting national policies to EU goals almost automatically leads right into basic research funding. Nonetheless, the relevant documents largely ignore[3] this focus and concentrate rather on the issue of competitiveness (Commission 2003).

Secondly, after years of benchmarking against the US it began to dawn upon EU policy makers that the dimension of US spending for basic research was a prerequisite for successful science based industries. Such a conclusion is not least supported by statistical evidence that the US lead in RTD intensity[4] may only be explained to 50 % by a higher degree of RTD intensive industries, the rest being explained by spending for military RTD but also basic research (Schibany and Streicher 2005).

This implies two messages for RTD policy making in Europe: basic research funding is among the first and foremost tasks of RTD policy, and, following the American lesson, the positive effects of joining forces are true for basic research funding as for any other field of research funding. As a consequence, the sudden attention at European level for basic research not only linked European issues back to the member states, but created also a new EU agenda taking the form of the ERC project.

It remains interesting though that it took EU politics so long to acknowledge the great tradition of the US institutions for basic research funding. It seems as if Vannevar Bush's (1945)

2 A similar concept – the European Scientific Area – was put forward by Commissioners Spinelli and Dahrendorf in 1973.

3 Like in the "3 % Action Plan" where this issue does not even figure (Commission 2003).

4 In terms of RTD expenditures as share of GDP.

influential ideas[5] have managed to reach Brussels only now and spread there while obviously Donald Stokes' (1997) quadrants[6] encountered lesser difficulties to enter Eurocratic minds.

Knowing History: What to Expect from Evaluation

Having briefly outlined the story of the most relevant changes in policy approaches, what does that mean for evaluation, and, after all, is anybody really interested? The answer to the latter question is, of course, yes. What really matters, however, are those being asked this question. No doubt plenty of institutes specializing in evaluation are ready to get started with whatever evaluation contract they may win addressing such a novel thing like the ERC. But why may we be right to suspect that policy makers will have a particular interest in what can be achieved through evaluation?

Now, that some common sense seems to have been established about how the shift towards basic research came about, it is all but clear what the exact nature of the measures following as a consequence should be. History and international comparison cast some light on how successful funding models should be designed and this may serve as a component of an ex-ante evaluation. Yet these insights can only be part of the story, and this is where those responsible for implementing the ERC start to look around for more helpful devices. Still, a full-scale ex-ante evaluation may not be thought of as an indispensable prerequisite for further action once a set of compelling facts justifying the new approach is at hand. The various documents on the subject that have been published by the Commission as well as independent institutions by now seem to give sufficient testimony for such a tendency (Schregardus and Telkamp 2002; Commission 2005).

There is another, quite simple reason why at this stage evaluation still has to find its place in the process: no money has been spent yet. Once the dimension of spending the taxpayers' money gains preponderance in the decision making process questions will be asked, hoping that the evaluation community may come up with answers.

Commandments from Policy Makers and Temptations to answer for Evaluators: Moving Targets in a Changing Environment

When being in a process of setting up a new research funding institution and, by doing so, putting the existing balance within the funding portfolios of both the member states and the community at stake, evaluation and evaluators are needed. The message, however, may not be heard since policy makers may not be aware of that evaluation is the field they should turn to. While still being in the situation of justifying a new initiative like the ERC, relying on

5 As described in Science: The Endless Frontier, where Bush outlines his proposal for post-war U.S. science and technology policy in a report to president Roosevelt.

6 Stoke re-examined the role of basic science in technological innovation. He challenged the post World War II paradigms emanating from Vannevar Bush's report, which implied an inherent separation of basic and applied research.

traditional policy advice is likely to be the right thing to do. After that, things get different but there are some pitfalls:

Once committing public money to a funding institution or a programme politicians and policy makers want to know what they get. This is the first challenge.

Commandment: Please tell us how we get value for money.

Temptation: Ask us after you have spent the money.

The challenge refers to the difficulties with accountability that all public authorities face. This creates a particular problem where – like in the realm of research – authorities do not have the expertise to evaluate the content of projects that are proposed for public funding, notwithstanding that it remains beyond doubt that funding research is a public core competence. Public authorities therefore need to establish and supervise mechanisms that yield good quality justifying public spending. Thus, the design and implementation of selection procedures is crucial; experience says that whether or not a programme or an institution is regarded a success afterwards is directly linked to the quality of project selection. This aspects gains even more importance the closer we get to basic research. Here, quite naturally, authorities are least probable to have the necessary expertise at disposal. In turn, the pressure upon policy makers to present evidence that they have taken every precaution in order to safeguard the taxpayers' interests is also higher. Therefore, evaluation experts must be prepared to be asked for advice on how to design sound selection (i.e. ex-ante project evaluation) tools and processes delivering robust results. Politics and administration need to establish a certain level of guarantee that their money is well invested. It is essential that evaluators resist their possible inclination to prefer classic and safe ex-post exercises.

Having designed an appropriate set of rules and procedures for the allocation of funding, the question of working out incentives aimed at funding the best research projects arises, while at the same time the requirements of political responsibility and justification need to be reflected.

Commandment: Please spend the money on excellent research and say who gets how much.

Temptation: We can only say what is 'good research' but please do not bother us with dreadful nitty-critty.

Implementing a project selection design that lives up to policy makers' expectations is only part of the story. The flipside of the coin has to do with the responsibility for the consequences of decisions generated by such a design. If we follow the reasoning established above that public authorities must have an interest in appropriate ex-ante project evaluation processes because they themselves lack the relevant expertise, it follows as a consequence that this intention will only be fully achieved if it is not confined to the issue of expressing an expert opinion. The evaluation of a proposal's quality must not be separated from the concrete financial funding decision, and the experts (evaluators and peers) should control for the consequences of their opinion by matching it with the resources available. If evaluation experts are reluctant (or not entrusted) to decide also on the allocation of funds, the weight of their expert opinion will also be undermined, because leaving the financial decision to non-expert authorities or policy makers feeds back directly to the projects concerned and causes a boomerang effect. Therefore, politics is well advised to entrust an ex-ante project evaluation system (as part of the ERC, for instance) with full responsibility for the allocation of the funds available for the projects that it selects. This is clearly the only way not to distort the balance and coherency of decision making within a research funding setting. Those involved as project evaluators or evaluation experts having designed an agency's (or, in the future, the ERC's) evaluation system must therefore resist the temptation to retreat into the safe haven of scientific expertise where they expect not to be troubled by administrative and financial issues. In turn, policy makers must not be lead astray by those reluctant to get the mud of financial responsibility on their hands and must not fall for the misleading concept of entrusting experts only with "strategic" competencies while keeping the decisions on the hard, e.g. financial, facts to themselves.

If an institution has been set up which enjoys a high degree of autonomy, the criteria and procedures according to which projects are selected and funds are allocated will have to be decided.

Commandment: Please develop something new.

Temptation: We already know that there is nothing new.

As every observer of politics knows from experience, a key message in order to trigger decisions is the novelty of the measure proposed. Now, when it comes to the ERC one may object that the whole thing entails plenty of novelty. But when looking at what and how the ERC is expected to do it, it – for what we know – resembles what most national research councils already do. As a consequence, the question what the USP of the ERC really is may be asked in the policy debate. Since it can hardly be a distinction along the lines of certain themes or (this has been excluded from the beginning) the requirement of trans-national co-operation within projects, the major room for manoeuvre left for developing something new is the way projects are selected, including the criteria that are applied. This refers also to the first challenge where the issue of the selection design in itself was dealt with, but not its potential to guiding the form and content of proposals submitted to the ERC. While it seems an inevitable consequence of the ERC's European scope that projects are going to be larger, the trickier aspect are the effects of the peer review process which emerged as a kind of article of faith from all discussions from very early on. Eventually, we have now touched upon this eternal subject about whose efficiency whole libraries have been written. However, the unique opportunity of designing a research council from the scratch at a level where no comparable institution has been active before should also be worth a try in search of evaluation and selection procedures avoiding a tendency towards structural conservatism. It is indeed up to evaluation experts, but also peer reviewers, to advice accordingly.

Provided that everything works out in a more or less ideal manner, one has to acknowledge that any new structure finds itself within an existing system, whose functioning efficiency it inherently affects at the same time because of its novelty.

> *Commandment*: Please think about how you can co-ordinate yourself with the outside world.
>
> *Temptation*: Sorry, this is not our first and foremost concern.

Basic research funding councils or agencies tend to be very particular entities as they act in an area of public intervention where, as has been said above, the expertise of the public administration is very limited. Because of this, most research councils have reached a stage where they have been able to pursue their business fairly independently since early in their history. This does not only mean that the twin principles of autonomy and independence are widely acknowledged, but also that those research councils tend to have been the first research funding institutions established in the history of their countries. Other funding agencies running application oriented schemes are very often much younger and acting in closer connection to the political authorities where the administration has its own technological expertise. As addressing science-industry co-operation is a comparatively new phenomenon these two types of agencies or councils had not much to do with each other. Only when science-industry co-operation and multi-actor projects became a widespread model – not least because of the overwhelming success of the Framework programme stimulating many similar activities also at national level – the pressure towards closer connections between basic and industrial research funding agencies increased. Basic research councils are, however, fearful that, as a consequence, they may lose their autonomy and are reluctant to get themselves into such networks actively. Interestingly, at the EU level we now find the reverse situation: the "classic" Framework programme is the older sibling while the ERC hast just been conceived. Once having put its teething problems behind itself the ERC may change the Community's research funding portfolio substantially. While it remains the ERC's business to determine what this means for researchers, the issue of how it affects the other funding programmes should be addressed in the Framework programme's evaluation and assessment routine.

Towards Increased Responsiveness of Evaluation

It may come as a comforting thought that only four commandments were identified. But even so the issues raised reflect a valid approach towards evaluation matters. Research managers, administrators, civil servants, policy makers and politicians do not see evaluation as goal in itself; to them it is rather a tool in order to get something properly done. Therefore it is, when establishing a new institution (or a new programme), paramount that evaluation expertise is taken on board from the very beginning of such a project.

The evaluation community must, in turn, understand that this includes also touching sensitive ground when evaluators want to be heard. Complying with political agenda-setting and keeping distance at the same time is often difficult, yet necessary if real help during decision-making is asked for from evaluation experts. Better usability in such a process can be achieved when evaluation expertise responds not only to requests like "please evaluate this or that" but also to a broader range of problem-solving demand (like the one put in form of "commandments" above) where it is not primarily directed to the field of evaluation.

Bringing together political decision making with input from evaluation would have two benefits: Firstly, evaluation would be more widely seen as a comprehensive approach assuring quality of public spending from the selection of projects to the ex-post evaluation of the entire institution (or programme), even together with the portfolio it affects. Secondly, it is reasonable to assume that as a consequence the overall quality of the measure in question would be better: once evaluation is built in from the beginning, it is likely that if integrated in a comprehensive evaluation concept also selection criteria and procedures will turn out to have a higher capability to select and to measure good quality projects.

One may object that evaluation may loose on its impact if it gets engaged in political decision-making. Politics and policy are, nonetheless, the gamekeepers. Thus, responding to a concrete political need (during a decision making process) rather than getting involved only afterwards may well be a preferable option for evaluation experts, as compared to letting go direct political agenda setting in research funding. The latter may occur when policy makers have been left alone with their questions and have waited for answers in vain because they did not think of asking evaluators and evaluators did not think they were those being asked.

While from the Austrian experience the involvement of evaluators from early on is only known at programme level, the interaction with political decision-making also in the field of basic research funding has already been the case, rather to a positive effect for the matters discussed (Arnold 2004; Plattform 2005). Whether or not the issues raised in this paper have been addressed satisfactorily in course of the debate on the creation of a European research council remains to be seen. A hint in this direction is an analysis comparing research councils of EU member states undertaken by CREST in preparation of the ERC: A substantial part of the study is devoted to selection and evaluation procedures, obviously taken as prerequisite for successful agency governance and performance (CREST 2006).

References

Arnold, Erik, 2004: Evaluation of the Austrian Industrial Research Promotion Fund (FFF) and the Austrian Science Fund (FWF). Synthesis Report, Brighton: Technopolis.

Bush, Vannevar, 1945: Science: The Endless Frontier, Washington DC: National Science Foundation.

Caracostas, Paraskevas and Muldur, Ugur, 1997: Society, The endless frontier. A European vision of research and innovation policies for the 21st century, Luxembourg: Office for official publications of the European Communities.

Commission of the European Communities, 2003: Communication from the Commission. Investing in research: an action plan for Europe, Brussels: COM(2003)226 final.

Commission of the European Communities, 2005: Frontier Research: The European Challenge. High-Level Expert Group Report, Luxembourg: Office for official publications of the European Communities.

CREST – Scientific and Technical Research Committee, 2006: The funding of basic research through national funding agencies in the European Research Area: sampling the diversity. A comparative analysis for CREST, Brussels: CREST 1201/06.

Guzzetti, Luca, 1995: A Brief History of European Union Research Policy, Luxembourg: Office for official publications of the European Communities.

Plattform Forschungs- und Technologieevaluierung, 2005: How to evaluate funding systems. The example of the FFF/FWF evaluation, Vienna: Plattform Newsletter 25/2005.

Schibany, Andreas and Streicher, Gerhard, 2005: The Way to Lisbon – A Critical Assessment, Vienna/Graz: Joanneum Research InTeReg Research Report 33-2005.

Schregardus, Peter A. and Telkamp, Gerard J., 2002: Towards a European Research Council? An offer that cannot be refused, The Hague: Netherlands Ministry of Education, Culture and Science.

Stokes, Donald, 1997: Pasteur's quadrant: basic science and technological innovation. Washington DC: Brookings Institution Press.

Author:

Rupert Pichler
Bundesministerium für Verkehr, Innovation und Technologie
Renngasse 5, 1010 Wien, Austria
rupert.pichler@bmvit.gv.at

Michael Dinges

Pressure, Deception and Peer Review

Some stories to begin with

Today we are all quite familiar with Gregor Mendel (1822-1884) and his groundbreaking theories of heredity which paved the way to modern genetics. Unfortunately, Gregor Mendel’s peers did not recognize his work at his time. Mendel sent his research report to the eminent Swiss Botanist Karl Nägeli[1](1817-1891), who had developed his own theory of evolution known as orthogenesis. He therefore rejected Gregor Mendel's discovery of how heredity works in pea plants. Mendel wrote to Nägeli, summarizing his results and asking where it would be best to publish them. Nägeli answered that the experiments were worthless and should not be published at all. Charles Darwin also received a letter from Mendel but he did not even read it. In the end Mendel committed his results to a small, privately printed monograph. His results were only re-discovered in the early 1900’s, almost 35 years after he had conducted his research (Botanik Online: Universität Hamburg[2]).

Approximately 150 years later, in December 2005, the “Hwang-affair” traversed the globe. W.S. Hwang et al. had published two papers[33] on presumably ground-breaking results in stem cell research in the renowned peer reviewed magazine *Science*, An investigation of the Seoul National University indicated, however, that a significant amount of the data presented in both papers was fabricated. The editors of *Science* therefore retracted both papers. Only few months later, another scandal shook the scientific community: A paper published in *Lancet* (2005; 366: 1359–1366) that was written by the Norwegian medicine researcher Jon Sudbø from Oslo turned out to be a complete fake. Sudbø, who had already published 38 scientific papers in renowned journals, admitted that there were further manipulations in his publications in the

1 Karl Nägeli (1817-1891) contributed many positive discoveries to botany, such as the life cycle of ferns and the distinction between a nucleus and the rest of the cell (remember that this was before the DNA and the nucleus being the site of the chromosomes were discovered). Nowadays, however, he is mostly remembered as the loser of several very important ideological battles in biology.
(source: http://www.ilmyco.gen.chicago.il.us/Authors/KNgeli593.html)

2 http://www.biologie.uni-hamburg.de/b-online/d08/08b.htm; There is a debate whether this claim holds true: the geneticist, Alexander Weinstein (1977) shows that the belief that Mendel's work was virtually unknown before 1900 dates back to statements made at the turn of the century by the "rediscoverers" of "Mendel's laws", de Vries, Correns and Tschermak. See: A. Weinstein: Journal of the History of Biology 10, 341-364.

3 For titles of papers and reaction of Science see : http://www.sciencemag.org/cgi/content/full/311/5759/335b#ref2

New England Journal of Medicine and in the Journal of Clinical Oncology (Deutsches Ärzteblatt Nr. 103).

Yet another case of scientific fraud was the case of the Jan Hendrik Schön, who, from 1998 to 2001, published on an average of one research paper per every eight days (alone and with co-authors). 17 of those papers had come out in *Nature* and *Science*. Due to allegations of scientific misconduct a committee was set up in 2002 in order to investigate possible scientific fraud. In the final report by the commissions evidence of Schön's misconduct was shown in at least 16 out of 24 allegations (*Nature*, Vol. 419, pp. 419-421; and Vol. 419, pp. 772-776).

Some questions arise…

These stories provide some anecdotic evidence for the fallibility of the human mind and for immoral human behaviour as well. They raise some questions in regard to peer review, or, more generally, about how peer review is related to the scientific system. One could ask:

- What is the role of peer review processes in the present scientific system? On the one hand it is important for the functioning of the whole system, on the other hand it can also fail to acknowledge real scientific breakthroughs and does not (always) prevent or detect deceptive behaviour.
- What are the reasons for scientific misconduct, and how can such swindle reach such proportions?
- Are there any alternatives?

Historical evolution - institutionalisation and different functionalities

Initially, we should take a look at the rationale and the development of peer review in the scientific system. In his essay on peer review in 18th century scientific journalism, Kronick (1990) pointed out that it, „can be said to have existed ever since people began to identify and communicate what they thought was new knowledge... because peer review (whether it occurs before or after publication) is an essential and integral part of consensus building and is inherent and necessary to the growth of knowledge“. Editorial peer review, however, was far from being a well established institution back at Mendel's time.

Rennie (2003) mentions that editorial peer review, or in the stricter sense prepublication review, seems to have begun in the early eighteenth century. Some evidence for this can be found in the first volume of the Royal Society of Edinburgh's „Medical Essays and Observations“, which was published in 1731. Additionally, in the article "Philosophical Transactions“, the Royal Society of London established a 'Committee on Papers' that was supposed to review articles and was empowered to call on „any other members of the society who are knowing and well skilled in that particular branch of Science that shall happen to be the subject of matter...“

Nevertheless, it is obvious that peer review only really became institutionalised after World War II (Rennie (2003), Burnham (1990)). In the course of the post-World war II science boom it became an accepted practice and it reached the height of its power in the US. There the National Science Foundation and the National Institutes of Health were established (Abate 1995) and renowned magazines, like *Science*, improved their peer review practices. Along with the institutionalisation of peer review, its functionalities changed from a discursive consensus building tool, as it had existed ever since, to a decision making tool that affected research funding and publication decisions.

There is no peer review, and yet there is quite a lot of it

But not only has the scope (and consequently the power) of peer review changed. As Rigby (2004) outlines, the forms of expert review have multiplied during the last century. In addition, Fröhlich (2006) has made the very important remark that there is neither „a peer review system" nor „a peer review process" in singular, but only multiple and highly variable practices which have often nothing in common except for the name „peer review".

Nevertheless, without being a standardised practice, peer review is most commonly defined as "the evaluation of scientific research findings or proposals for competence, significance and originality, by qualified experts who research and submit work for publication in the same field (peers)" (Sense about Science, 2004). And despite the heterogeneity of peer review practices, "peer review" seems to constitute a corner stone especially in the dissemination process of scientific research results. It is seen as "a form of scientific quality control" or "an error detection system" that is based on the scientific judgement of other experts who are trying to enhance knowledge in the field as to whether work is competent, significant and original (Science Media Centre, 2003 – Communicating Peer Review in a Soundbite – as quoted in Sense about Science 2004).

For instance, during the process of editorial peer review that affects publication decisions, referees typically comment and make recommendations on various dimensions of a scientific paper (see box 1) in order to fulfil the purpose of quality control. The comments delivered by these referees usually lead to the acceptation (with/without minor amendments or major revisions) or the rejection of a paper.

Table 1: Referees usually comment and make recommendations on some of the following:

1. Significance:	Are the findings original? Is the paper suitable for the subject focus of this journal? Is it sufficiently significant? (Is it a 'me too' paper; is it 'salami slicing?)
2. Presentation:	Is the paper clear, logical and understandable?
3. Scholarship:	Does it take into account relevant current and past research on the topic?
4. Evidence:	Are the methodology, data and analyses sound? Is the statistical design and analysis appropriate? Are there sufficient data to support the conclusions?
5. Reasoning:	Are the logic, arguments, inferences and interpretations sound? Are there counter-arguments or contrary evidence to be taken into account?

6. Theory:	Is the theory sufficiently sound, and supported by the evidence? Is it testable? Is it preferable to competing theories?
7. Length:	Does the article justify its length?
8. Ethics:	In papers describing work on animals or humans, is the work covered by appropriate licensing or ethical approval? (Many biological and medical journals have their own published guidelines for such research.)

Source: Working party "on equipping the public with an understanding of peer review" (see http://www.senseaboutscience.org.uk/peerreview/),

The problems with peer review

Meanwhile the scientific community struggles with the shortcomings and the glories of this non-standardised tool called peer review, which is deemed to be both indispensable and unbearable at the same time. While working parties have been set up in order to make the public understand the necessity of peer review, criticisms and allegations on peer review are quite common as well.

Rennie (2003) summarises these allegations and lists scientific papers that are dealing with them. However, while he has commented on these criticisms in greater detail, I would only like to quote some of the core issues in order to give an idea of the ongoing discussion and to provide some information about the existing literature concerning peer review:

- Peer review is not standardised and therefore idiosyncratic and open to all sorts of bias (Rennie 1993)
- Peer review secrecy leads to irresponsibility because it insulates reviewers from accountability (Harnad S. 1996) and it invites malice. (Rennie 1991, Rennie 1998, Rennie 1994, Altman 1996, Ruderfer 1980, Osmond DH, 1982)
- Peer review stifles innovation, perpetuates the status quo and rewards the prominent (Horrobin 1990, Mahoney 1977). Peer review tends to block work that is either innovative or contrary to the reviewer's perspective. Controversial work is therefore more harshly reviewed (Smith 1997). Horrobin (1990) has cited 18 cases where, in his view, innovation has been blocked by the peer review system.
- Peer review lends a spurious authority to reviewers. The anonymous opinions of the reviewers are set against the documented work of the author and are given exaggerated weight by the editor who appointed the reviewers. (Rennie 1991)
- Peer review must fail because only reviewers that are very familiar with the subject are knowledgeable enough to review. As such, however, they are competitors and therefore disqualified by their conflict of interest.
- Peer review causes unnecessary delay in publication and is very expensive. (Altman 1996)
- Science scarcely benefits because authors usually ignore the reviewers' comments if their manuscript has been rejected. (Armstrong 1997)

On top of all the previous quotations Rennie takes up the accusations that peer review is said to be unreliable, unfair, and fails to validate or authenticate. Several studies have shown that its reliability – which would be indicated by an agreement among the reviewers – is poor (Cicchetti 1997, Armstrong 1997) and that partiality and bias pose a real threat to the peer review process. Ross et al. (2005) found evidence for bias in favour of prestigious institutions within the United States at conference applications for the American Heart Association, which used open peer review from 2000 to 2001 and blind review from 2002 to 2004 for abstracts submitted to its annual scientific sessions. The findings of the study argue for a universal adoption of blind peer review by scientific meetings.

In an Article published in *Nature*, Wenneras and Wold (1997) found evidence for severe gender bias in the peer review of research grant applications to Sweden's Medical Research. Hence, we can conclude that there is plenty of evidence for potential and existing bias in peer review processes. Martin (2000) categorizes 4 different sorts of bias in his papers on research grants – problems and options:

- Success-breeds-success bias: successful applicants are likely to become entrenched, using their grants to produce the outputs necessary to attract further funds, while others never even have the chance to get started. This bias is also known as the "Matthew effect" as introduced by Merton (1998) in an analogy to the Gospel according Mathew (13:12): Whoever has will be given more, and he will have an abundance. Whoever does not have, even what he has will be taken from him.
- Insider bias: decisions are made by cliques of insiders, who think highly of, and award most grants to themselves and a small group of favourites.
- Dominant group bias: there is discrimination against groups such as women, ethnic minorities and lower-status institutions (Peters and Ceci 1982; Wenneras and Wold 1997).
- Conventional approach bias: grants are much more likely to support tried-and-true approaches, while challenging, innovative or unorthodox proposals are seldom funded (Armstrong 1996, 1997; Epstein 1990; Horrobin 1990, 1996).
- Personal bias: administrators or referees obstruct researchers or projects that they do not like (Horrobin 1974).

But these different forms of bias are not the only problems in peer review. Interdisciplinary research can make some of its other limitations visible, when we look at the fact that peers are supposed to be colleagues working in the same field on similar topics. In interdisciplinary research, however, one cannot expect peers to be competent in every aspect of the project (Laudel 2004). Porter and Rossini (1985) found some evidence that interdisciplinary proposals are downgraded in peer review because peer reviewers tend to rate proposals from their own discipline more favourably (as quoted in Rinia E.J. et al 2001). However, in a paper on the influence of interdisciplinary aspects on peer-review and on bibliometric evaluations in physical research in the Netherlands, Rinia et al. (2001) showed that there is no general bias against interdisciplinary projects in peer review assessments and in the scientific impact that bibliometric indicators show. In this case, it seems to matter to which degree and how centrally the research project is interdisciplinary.

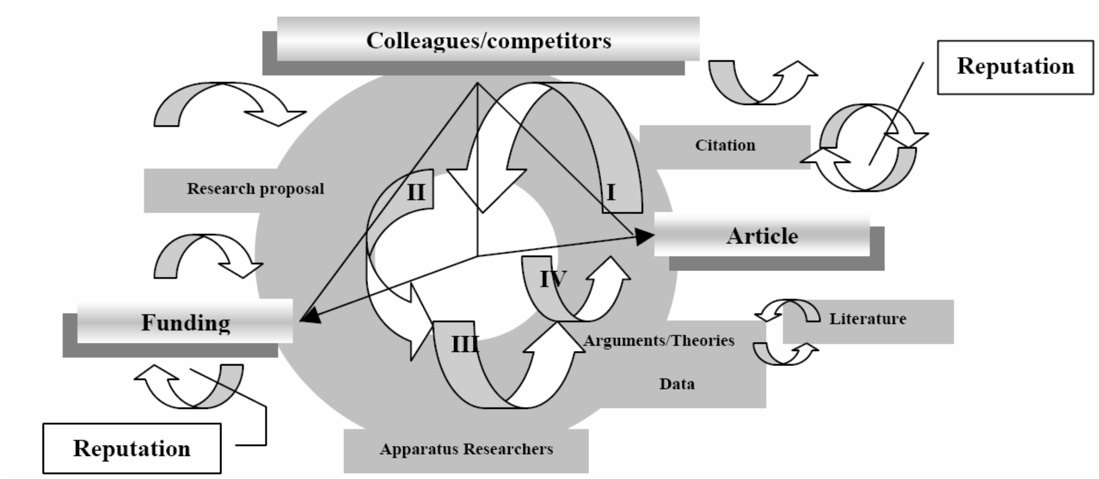
Peer Review, the reputation of scientists and yet more problems

Rennie (2003) is surprised that, given the early history of institutionalised peer review, anyone should be so naive as to imagine that peer review could validate or authenticate scientific work or that it could guarantee its integrity, since, in his view, the peer review process is an effort that is supposed to improve quality but does not guarantee perfection. However, it is exactly this guarantee of perfection that editors, the public, and funding agencies are seeking for. To them, peer review is not primarily a quality improvement tool, but a quality control mechanism, which directly and indirectly affects some important elements of the scientific system, namely funding, publications, and scientific reputation.

Funding decisions for research grants and for scientific publications depend heavily on the results of peer review, and thus peer review also contributes to the scientific reputation of the grant/publication recipients. Latour and Woolgar (1979) were among the first to see research funding as part of a scientist's reputation and credibility cycle, which links production, communication and collective evaluation of research results into a concept of reputation. Whereas the right hand side of Figure 1 shows that the reputation of scientists is created via the production of scientific articles in refereed journals, the left hand side shows that grants distributed via competitive mechanisms constitute yet another dimension of reputation creation.

Analyzing the outcome of competitive funding procedures (grants received) and publications in scientific journals allows us to differentiate research institutions and researchers with regard to their research capabilities, effort, and their competitiveness (Garcia and Sanz Menéndez, 2004).

Figure 1: How everything is connected - the reputation cycle for scientists



Source: García and Sanz Menéndez (2004)

The appeal of such procedures is quite obvious, since we are all eager to search for evidence based decisions on research funding, the "best" researchers in a field and the most prominent research institutions.

Peer review has considerably expanded its role both in direct (via peer review for journal publications and research grants) and indirect ways, as formula based funding schemes entered the scene, where the allocation of resources depends on publication output and input factors in scientific journals. However, these procedures that are ultimately based on a weak peer review tool may provoke unwelcome and unintended consequences:

1. An expertise to the German Ministry of Research and Education, (Gläser et al., 2002) reports that in the Australian funding scheme for university research concerns about formula based funding are being raised and that there are doubts as to whether the currently used formulas should remain in use. Furthermore, studies have documented a significant increase in the country's journal output that was accompanied by a worrying decrease in the relative international impact of these publications as measured by citations.
2. Scientific systems operating under an authoritative "publish or perish" regime contribute to 'salami slicing' tactics in publishing, which divide research results into minimal papers at the threshold of acceptability while including as many co-authors as possible (Fröhlich, 2006). These tendencies definitely present the peer review system with new challenges; as peer review is a self-governing system, we have to be aware of the fact that the attached merit system in science provides incentives to either misuse referee power or to accept salami-slicing so that everyone get's his/ her share in publications.
3. Rennie (2003) makes a clear point stating that the editorial problem has changed from a search to fill empty pages to the selection of the best out of an avalanche of incoming manuscripts. Editors have felt the need to refine their techniques of selection and rejection without good evidence concerning the predictive powers of peer review to distinguish exactly between the best 10 % and the next best 5 %; a matter of great concern to authors submitting to journals with a 90 % rejection rate.

Pressure and scientific misconduct

The linkage of publication, peer review, and career consequences is considerable and has put many burdens on an over-emphasised tool with limited validating power. It has put a lot of pressure on researchers and it may also be a major factor in scientific misconduct.

Martinson et al. (Nature 2005; 435: 737-738[4]) surveyed early- and mid-career scientists, who work in the United States and are funded by the National Institutes of Health (NIH), and asked them to report their own behaviour. Having elaborated 16 forms of possible scientific

[4] http://www.nature.com/nature/journal/v435/n7043/full/435737a.html, for a detailed report written in German see also Deutsches Ärzteblatt Jg. 102, Heft 26, 1. Juli 2005

misbehaviour in expert talks with 51 researchers, a questionnaire was sent to a random sample of 7760 researchers, of which 3427 responded.

Overall, 33 % of the respondents said they had engaged in at least one of the top ten kinds of behaviour during the previous three years. The authors were particularly concerned about the working environment of scientists, about their perceptions of the functioning of resource distribution processes, and about the potentially detrimental effects on the ethical dimensions of scientific work: “These processes are embodied in professional societies, through peer-review systems and other features of the funding and publishing environment, and through markets for research positions, graduate students, journal pages and grants. In ongoing analyses, not yet published, we find significant associations between scientific misbehaviour and perceptions of inequities in the resource distribution processes in science. We believe that acknowledging the existence of such perceptions and recognizing that they may negatively affect scientists' behaviours will help in the search for new ways to promote integrity in science.”

Moving forward

So far, I have tried to outline some of the main issues concerning peer review nowadays. Despite the numerous problems associated with peer review, it is nevertheless spreading and its importance as an allocation mechanism has grown, not only for journals and classical project funding in science funds, but also for R&D programmes. This, however, is reasonable: expert opinions and peer review processes support fair funding decisions, as no one but peers will be able to judge upon the originality or the degree of innovativeness of a research proposal.

As regards the potential biases and pitfalls in peer review, one has to take into account that these biases also affect performance rankings of individuals and research institutions. Hence, incorporating performance based funding schemes may even fortify the distortions resulting from the peer review process.

Peer review should not be seen as an infallible and authoritative tool and the responsibility for funding decisions should always be on the funding agency and the funding committee. The responsibility for publication decisions should finally be on the editor’s side. Furthermore, it is also crucial that the decision making processes are made transparent and that reasons for the decision are given in order to make the process more fair and reliable.

Peer review is often free of cost (for editors and agencies, but not for the peers), since it is used more and more frequently, its limits may not only have been reached due to its inherent obstacles, but the peer-review system could also break down simply because of its frequent use. As a rule of thumb, it is often said that agencies or science funds have to contact 10 scientists to find one peer willing to write a review[5].

[5] The Austrian Science Fund (FWF) seems to be in a somewhat exclusive position, as 70 % of peers accept its invitation to review research grant proposals.

Given the fact that the time of researchers is limited, fears arise that an overstraining of peer review may lower the quality of the reviews, or that there may be conflicts of interest. Since the availability of peers is scarce, funding agencies and journal editors may be tempted to not take an all too close look at those willing to perform the review. Hence, although everyone admits that we have to do all we can to identify different sources of bias and pin them down, the increasing demand for peer review might have an unintended detrimental effect concerning bias and abuse of power.

Wesolowski (2003) states with regard to editorial review processes that authors should routinely be given the benefit of the doubt. Rather than having the review process be the ultimate watchdog, the published "comments and replies" that many journals offer and that are an underutilized method of airing scientific debates, should be encouraged. He admits that publication is an expensive process, but he also thinks that it costs us more as a society, if we squelch controversial observations and hypotheses too arduously. These sometimes prove to have been correct in the first place, and they frequently spark new lines of research and testing in any event.

Smith (2003) writes that at present there do not seem to be serious alternatives to peer review, but he argues that, although peer review has been structurally static since the nineteenth century, advances such as electronic publishing and continuous quality improvement may help to improve the quality of peer review and develop new systems. Abbasi et al (2002) have tried to identify strong forces that may radically change the world of scientific and medical publishing. The most important ones seem to be:

- The appearance and spread of the World Wide Web, facilitating greater communication between authors and their readers while reducing the need for intermediaries.
- Increasing resentment in the academic community about having to pay ever more for information that it effectively produces itself.
- The appearance of new players - such as HighWire Press, BioMed Central, and PubMed Central - who are trying to capture value that currently belongs to publishers.
- The price of information is falling as many organisations such as pharmaceutical companies make information available for free on the internet
- The marginal price of electronic information is effectively zero.
- The real price of long distance telephone calls is close to zero.

Abasi et al. further used scenario planning techniques in order to imagine four different futures of scientific and medical publishing, which they named after the characters from the well-known Simpsons family. The article characterizes the different worlds as follows: The Marge world is the world of academic innovation where academics innovate and publish primarily on the web and not in journals; in it, the publishers must bring out large numbers at low cost to succeed. In the Homer world, the world of the lazy father, there's no great need to change and publishers adapt to the electronic world and continue to publish research as they've done before. Lisa represents a world of global conversations; publishers have largely disappeared, and communication takes place globally and with the help of electronic communication technology. In the Bart world, finally, the big guys have taken over. Publishers have also disappeared and instead large organisations have become the main purveyors of research.

As Smith (2003) has said – despite its clear deficiencies, peer review probably does have a future… The appearance of the internet is likely to transform peer review just as it is likely to transform everything else as we move from the industrial age to the information age. Peer review processes and functionalities have gradually changed in the past and they may even radically change in the future. We are only beginning to see how peer review might work in this electronic age, but it is likely to become more open. And that would be good news.

Literature

Abate, T. What's the verdict on peer review? 1995. The World of Research at Columbia University. http://www.columbia.edu/cu/21stC/issue-1.1/peer.htm

Abbasi, K. et al. Four futures for scientific and medical publishing. BMJ 2002;325;1472-1475.

Altman LK. The Ingelfinger rule, embargoes, and journal peer review – part 1: Lancet 1996;347:1382–6; part 2: 1996;347:1459–63.

Armstrong JS. Peer review for journals: evidence on quality control, fairness, and innovation. Sci Eng Ethics 1997;3:63–84.

Armstrong, J.S., Peer Review for Journals: evidence on quality control, fairness, and innovation. 1997. Science and Engineering Ethics, 3(1),pp 63-84.

Armstrong, J.S., The ombudsman: management folklore and management science – on portfolio planning, escalation bias, and such. 1996. Interfaces, 26(4), pp 25-55.

Burnham JC. The evolution of editorial peer review. JAMA 1990;263:1323–9.

Chubin, Daryl E and Hackett, Edward J 1990, Peerless science: peer review and U.S. science policy, Albany, State University of New York Press.

Epstein, William 1990, 'Confirmational response bias among social work journals', Science, Technology, & Human Values, 15(1), pp 9-38.

Fröhlich, G. Informed Peer Review – Ausgleich der Fehler und Verzerrungen? In HRK, HG. Von der Qualitätssicherung der Lehre zur Qualitätsentwicklung als Prinzip der Hochschulsteuerung. Bonn, 193-204.

Garcia, C. and Sanz Menendez, L., Competition for funding as an indicator of research competitiveness: The Spanish R&D Government Funding, Unidad de Politicas Comparadas 2004, CSIC – Working Paper 04-15.

Gläser, J. et. al., Impact of evaluation-based funding on the production of scientific knowledge: What to worry about, and how to find out (Expertise for the German Ministry for Education and Research), Australian National University & Fraunhofer-Institut für Systemtechnik und Innovationsforschung, 2002, Karlsruhe.

Guston, D.,: The expanding role of peer review processes in the United States. In: Shapira, P.; Kuhlmann, S. (eds.): Learning from Science and Technology Policy Evaluation. Cheltenham: E. Elgar Publishing 2003.

Harnad S. Implementing peer review on the net: scientific quality control in scholarly electronic publications. In: Peek RP, Newby GB, eds. Scholarly publishing. The electronic frontier. Cambridge, MA: MIT Press, 1996:103–18.

Horrobin DF. The philosophical basis of peer review and the suppression of innovation. JAMA 1990;263:1438–41.

Horrobin, D F 1974, 'Referees and research administrators: barriers to scientific advance?', British Medical Journal, (27 April), pp 216-218.

Horrobin, David F 1990, 'The philosophical basis of peer review and the suppression of innovation', Journal of the American Medical Association, 263(10), pp 1438-1441.

Horrobin, David F 1996, 'Peer review of grant applications: a harbinger for mediocrity in clinical research?', Lancet, 347 (9 November), pp 1293-1295.

Kronick DA. Peer review in 18th-century scientific journalism. JAMA 1990;263:1321–2.

Latour, B. and S. Woolgar, Laboratory Life. The construction of scientific facts, 1979Sage: London.

Laudel, Grit,: Conclave in the Tower of Babel: How Peers Review Interdisciplinary Research. In: Proceedings TASA 2004 conference, 2004, Revisioning Institutions: Change in the 21st Century, La Trobe University, Beechworth Campus, 8-11 December, 8 pages.

Mahoney MJ. Publication prejudices: an experimental study of confirmatory bias in the peer review system. Cogn Ther Res 1977;1:161–75.

Martin, B., Research Grants: Problems and Options, Australian Universities' Re-view, 2000. Vol. 43, No. 2, 2000, pp. 17-22.

Merton, Robert K. (1985): Der Matthäus-Effekt in der Wissenschaft. in: ders.: Entwicklung und Wandel von Forschungsinteressen. Aufsätze zur Wissenschaftssoziologie. Suhrkamp, S.147-171.

Martinson, B.C. et al. Scientists behaving badly Nature 435, 737-738 (9 June 2005)

Noah L. Sanctifying scientific peer review: publication as a proxy for regulatory decisionmaking. Univ Pittsburgh Law Rev 1998;59:677–717.

Osmond DH. Malice's wonderland: research funding and peer review. J Neurobiol 1983;14:95–112.

Peters, D.P., Ceci, S.J., Peer-review practices of psychological journals: the fate of published articles, submitted again. 1982. Behavioral and Brain Sciences,5. pp 187-195.

Porter, A.L., Rossini, F. Peer-review of interdisciplinary research proposals. Science, Technology, and Human Values 10, 33-38.

Rennie D. Anonymity of reviewers. Cardiovasc Res 1994;28:1142–3.

Rennie D. Editorial peer review: its development and rationale. In Godlee F, Jefferson T, editors. Peer review in health sciences. Second edition. London: BMJ Books, 2003:1-13.

Rennie D. Freedom and responsibility in medical publication. Setting the balance right. JAMA 1998;280:300–1.

Rennie D. More peering into peer review. JAMA 1993;270:2856–8.

Rennie D. Problems in peer review and fraud: cleave ever to the sunnier side of doubt. In: Balancing act, essays to honor Stephen Lock, editor, BMJ. London: The Keynes Press, 1991:9–19.

Rennie D. The state of medical journals. Lancet 1998;352:SII18–22.

Rigby, J., Making decisions about science and technology – between the devil and the deep blue sea. 2004. Plattform fteval newsletter – www.fteval.at.

Rinia E.J., Influence of interdisciplinarity on peer-review and bibliometric evaluations in physics research. Research Policy 30 (2001) 357-361.

Ruderfer M. The fallacy of peer review – judgement without science and a case history. Speculations Sci Technol 1980;3:533–62.

Sense about Science, Peer Review and the acceptance of new scientific ideas – discussion paper from a working party equipping the public with an understanding of peer review. 2004. www.senseaboutscience.org . London.

Smith R. Peer review: reform or revolution? Time to open the black box of peer review. BMJ 1997;315:759–60.

Smith R. The future of peer review. In Godlee F, Jefferson T, editors. Peer review in health sciences. Second edition. London: BMJ Books, 2003:329-46.

Wennerås, C., and A. Wold. Nepotism and sexism in peer-review. 1997. Nature 387: 341-343.

Wesolowski D.J.: Preserving Anonymity in the Review Process, 2003. EOS, Vol. 84, 583.

Author:

Michael Dinges
Joanneum Research, Institute of Technology and Regional Policy
Wiedner Hauptstr. 76, 1040 Wien, Austria
michael.dinges@joanneum.at
www.intereg.at

Rudolf Novak

Some Common Ways to Distribute Funds – Evidence from International Practice

1. Introduction

Research Councils all over the world have to distribute limited funds and are confronted with the – often increasing – divergence between number and volume of grant applications compared with available funds, increases in which usually lag more or less far behind the requested sums. This challenges research councils to develop adequate procedures that ensure a fair distribution of funds and the selection of the "best" applications while limiting complexity and time of review and decision processes to a reasonable order. In the field of basic research, peer review and decision-making boards represent widely accepted methods for selection of grant applications for funding but considerable differences exist in the application of this method.

In 2004, the Austrian Science Fund (FWF)[1] conducted a fact-finding mission that investigated the selection and decision-making procedures in use in a variety of European funding agencies and research councils as well as in the US NSF. The aim of this mission was to study different models in detail and "on site". Much information about an organization's application and decision-making processes can be collected via brochures and homepages but operational details, which are often crucial for the success of a procedure, usually do not appear there. The results of the fact-finding mission represented the basis for drafting perspectives for modifications and reforms to the FWF's own procedures. This short article compares the advantages and disadvantages of various models from the "practitioner's" point of view. It is to be hoped that some of the comparisons may be of wider interest, especially in the light of the current discussions on the design of the European Research Council (ERC).

2. Rationale for the study

Input for a reorganization of the FWF´s review and decision-making procedures was considered especially necessary for the FWF´s largest funding category, stand-alone projects (which receive about 70 % of FWF funding), for the following reasons:

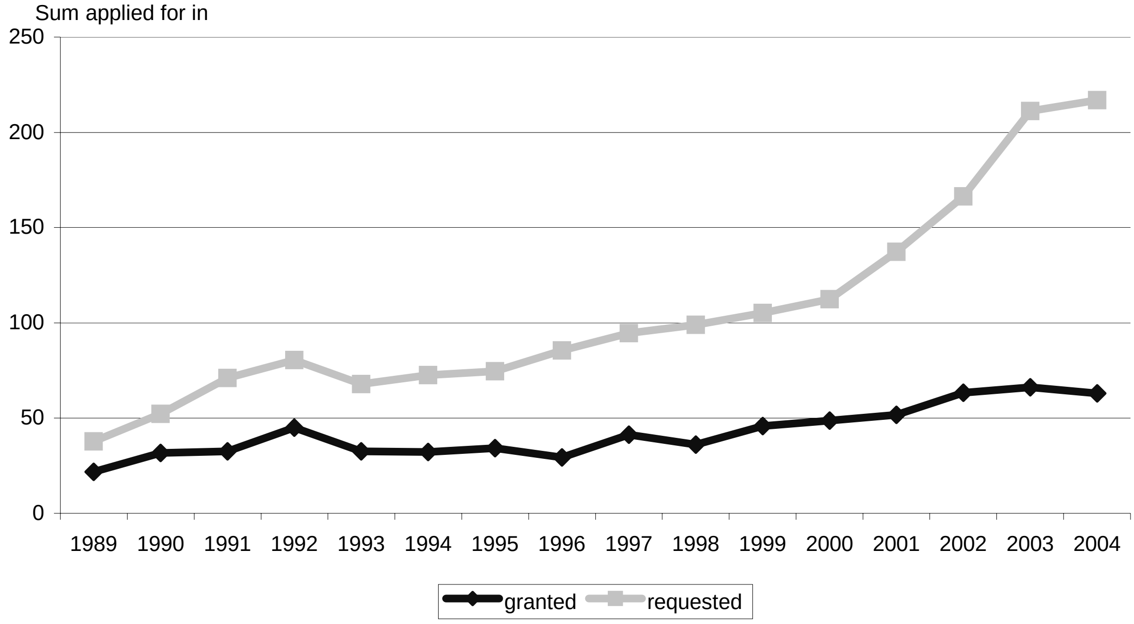
High application numbers and the increasing gap between requested and available funds lie behind the current fall in the acceptance rate of proposals (Figs 1 and 2). The acceptance rate

[1] The FWF is Austria's central body for the promotion of basic research. It is equally committed to all branches of science and in all its activities is guided solely by the standards of the international scientific community. See also www.fwf.at

appears fairly high in international comparison but this point requires discussion in its own right and needs to be seen in the light of the general funding situation in Austria, consideration of which is outside the scope of the present article. The central point in the present context is that the acceptance rate has dropped by 50 % over the last 10 years. As a consequence, an increasing number of applications with very positive reviews cannot be funded.

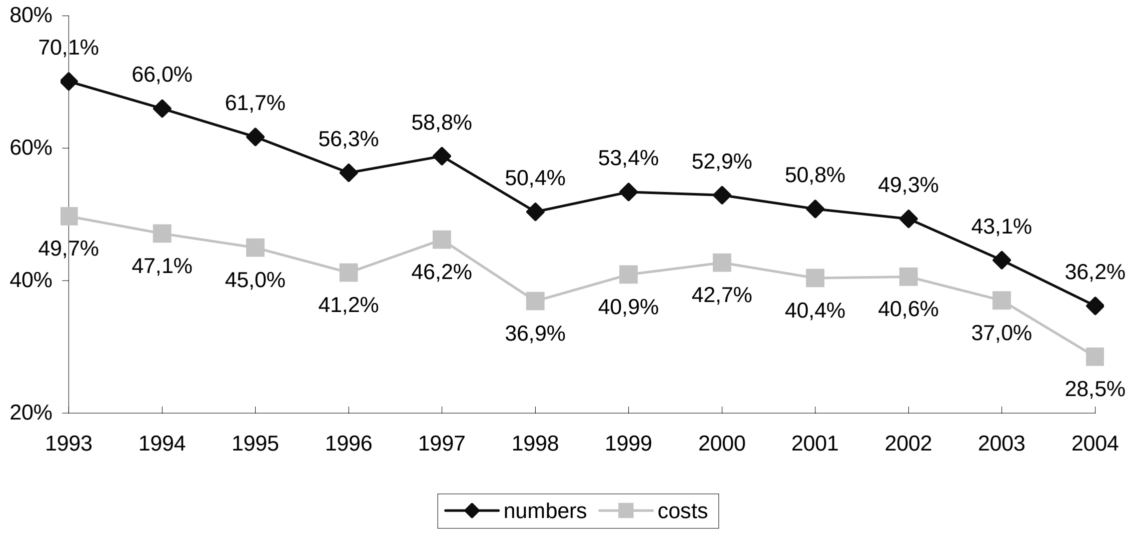
In the FWF's experience, an amazingly high proportion of scientists respond positively to requests for review: more than two thirds of about 5,000 scientists approached annually agree to act as reviewers. However, there is a substantial increase in the need for comparative discussions of applications in the Board meetings, in addition to paper (mail) review of grant applications.

Figure 1: FWF - Applications Stand Alone Projects



Source: FWF

Figure 2: FWF – Acceptance Rates Stand Alone Projects



Source: FWF

The FWF is in the process of re-orientating its position and its activities within the national innovation system. One of the consequences is that more time is needed for strategic discussions by the FWF's Board. Another consequence is that the FWF is administering, or at least involved in, an increasing number of funding programmes for other institutions (Austrian federal ministries, other funding agencies, universities). Decisions relating to these programmes have to be taken by the FWF Board, which still has to deal with its traditional "core business". Time pressure in the Board meetings has increased dramatically in recent years and this problem must be solved by the implementation of more efficient decision-making procedures.

3. Methods

Six research funding organizations in other countries were visited by FWF staff members. The choice was largely based on the sample of organizations that served as benchmarks in the international evaluation of the FWF conducted on behalf of the Austrian Federal Ministry for Technology, Innovation and Transport under the leadership of Technopolis Ltd in 2003/2004. The organizations were: AKA - Academy of Finland, EPSRC – Engineering and Physical Sciences Research Council UK, DFG - German Research Foundation, NWO - Netherlands Organization for Scientific Research, NSF - National Science Foundation USA and SNF - Swiss National Science Foundation.

In addition to an analysing available data (brochures, homepages), FWF staff members visited the organizations and performed standardized interviews with selected colleagues there. In the case of the NSF, the FWF staff member was also invited to participate as a reviewer in an expert panel, which provided valuable additional insights into the NSF's procedures.

An international workshop on peer review procedures organized by the NORFACE ERA-Net (mainly Scandinavian countries) and the FWF's work in various other ERA-Nets provided additional input.

4.1. General trends

Internationality: To an increasing extent, reviewers from abroad are used, i.e.: reviewers do not work in the country of the applicant (funding agency). In smaller countries, this is a central prerequisite for ensuring the impartiality of reviewers. The problem is less striking in large European countries (such as Germany and the UK) and virtually non-existent in the US. Nevertheless, it seems that reviewers from abroad are increasingly used even in some of the larger European countries. There seems to be a widespread consensus that the quality of reviewers and the reviews they provide should be assessed and documented.

Division of labour in review and decision making: European funding agencies show a general trend towards a separation of peer (mail) review, ranking of applications by expert panels (based on written reviews) and decision by council boards. While the final decisions by the board are often purely formal in nature and/or focus mainly on budgetary aspects, scientific discussions take place in the expert panel meetings. The division of labour between review/ expert panels, council boards and council staff members varies considerably. In many funding institutions, staff members are heavily involved in the selection of peers and the preparation of the board decision. Applicants may frequently suggest peers or exclude selected scientists (institutions) from involvement in the review procedure.

The use of study panels/study groups is increasingly good practice. A generally observed advantage is that it facilitates comparative discussions between applications. It is often the case that large review panels or study groups give unconventional (non-mainstream) applications a better chance than small and/or specialized panels. Larger panels often give a competitive advantage to small and very small scientific communities. On the other hand, reviewers are frequently not particularly familiar with the field of some of the applications they are judging. In general, the use of study groups and/or review panels seems to be significantly more expensive that relying entirely on written reviews.

Transparency is one of the "hot issues" of the peer review procedure. The anonymity of peer reviewers, which is a basic rule in many procedures, has been and continues to be heavily criticized. On the one hand, it is considered necessary to "protect" reviewers so as to guarantee the impartiality of their assessments; on the other hand, critics note that reviewers who are "hidden" in the cloak of anonymity have an easier job "killing" applications (papers) from potential or actual competitors. The problem may possibly be resolved, at least in the EU, by a general application of the "freedom of information act", which is presently under discussion. In some countries the names of reviewers are customarily published (e.g. as a general rule in Denmark, or upon request in Finland or only in agreement with the reviewer in the UK). Feedback between applicants and reviewers is a common practice to increase transparency. Applicants are given the chance to respond to reviewers' statements and both the assessments and the applicants' responses are taken into account in the funding decision However, a systematic procedure of reply presupposes a highly specialized review panel that is able to evaluate the reviews and the applicants' replies.

Honorarium for peer reviewers: In the field of basic research, it is still the practice in the majority of European funding agencies that reviewers are not paid for their work. However, there is a widespread consensus that this might not remain so in the future: the number of highly qualified reviewers is limited and reviewers should be financially recompensed for the time they take to produce quality evaluations. Various models are currently under discussion:

(1) Direct payment. This is practised by some Scandinavian funding organizations. The problems are numerous: a) funding organizations enter a competition for reviews that raises prices, b) scientific motivation for acting as a reviewer (reputation) becomes less important; c) weak reviews can hardly be avoided because reimbursement cannot be made conditional on review quality; d) the administrative costs increase significantly in comparison with funding budgets.

(2) Indirect rewarding system. This is a common practice in the UK and the USA. Scientists' reviewing activity is taken into account and rewarded by the home institutions (universities) in the documentation of the performance of individual researchers.

(3) Reviewers may increasingly be involved in the decision-making process via various feedback mechanisms.

It seems obvious that this problem can only be solved at an international, at least a European, level. Discussing payment for reviews in the funding of basic research could and should be a task for, e.g., the EUROHORCS (the association of the Heads of European Research Councils).

Earmarking of budgets: Most funding organizations earmark budgets for different scientific (scholarly) fields and/or programmes. Earmarked budgets have undeniable advantages, of which the most important ones are probably facilitated planning and management. Nevertheless, all discussions underlined a serious drawback: once budgets are fixed, the possibilities for making significant changes are severely restricted. This hinders flexibility and rapid reaction to unforeseen developments and changing demands of the scientific community and may lead to budget shortages and thus high rejections rates in one field (or programme), whereas other fields (programmes) may be over-financed, permitting weaker projects to be funded. The problem is less striking in large scientific communities such as the NSF in the USA, where nearly all money is distributed on an earmarked basis (programme funding). Earmarking and/or division of budgets between scientific (scholarly) disciplines also gives rise to significant problems in the handling of interdisciplinary projects and special programmes must be developed to solve them (e.g. the thematic programmes in the NWO) and to enable cross-border communication of departments or research councils.

Continuous application vs. deadlines: several funding institutions have defined deadlines for submitting grant applications, at least in certain funding schemes. The advantages of “calls” (or application deadlines) seem to be higher visibility of announcements and facilitated planning for the scientists and the funding bodies (both in budgetary and in procedural terms). Some evidence for a possible drawback of deadlines is provided by the observation that the share of proposals with a lower quality seems to increase in the case of fixed calls (due to an increased number of “last minute applications”).

Last but not least: Administration. Restriction of the length of applications, electronic submission of applications and electronic review are increasingly common good practice in research funding organizations. Experience shows that these measures considerably increase the efficiency of administrative procedures.

4.2. Comparison of “typical” Schemes

In the course of the FWF’s Fact Finding Mission (FFM), the structures and procedures of the various funding organizations were studied in more detail. The study referred especially to operational details, such as workload, costs and details of interactions between the various actors (e.g. applicants, reviewers, boards, secretariat etc.), which are of low importance in the context of this paper. However, we found it helpful in our further considerations to group the procedural and structural schemes of the different organizations into three major “forms” and these are discussed briefly below.

4.2.1. The strong council board

In this form, the council board (or council boards) play the predominant part in the review and decision making process. The present structures and procedures of the Swiss National Science Foundation (SNF) and the Austrian Science Fund (FWF) are good examples of this model. In the SNF, reviewers for project grant applications are nominated by Council Board Members, frequently in close cooperation with and with considerable support from staff members of the council secretaries. The “open (bottom-up) funding scheme” of applications submitted by individual researchers accounts for the major part of the budget (about 80 %); the following section refers to this scheme only (funding schemes for targeted research have different rules that are not discussed here). About two thirds of the reviews for the open (bottom-up) funding scheme come from outside Switzerland and reviewers are not paid (although for some time there was a trial in the department for Humanities and Social Sciences of paying about 100 € per review). The review procedure, ranking and decision making is conducted separately in three different departments (Humanities and Social Sciences; Mathematics; and Natural and Engineering Sciences and Biology and Medicine), which have their own budgets. This seems to lead to a tendency of the various departments to develop divergent rules and practices, which is increasingly considered by some leading SNF members as a matter of concern with respect to the organization’s corporate identity. The strengths of the system are undoubtedly its low costs, the low organizational complexity and a relatively high degree of “internationalization” of reviewers and of confidentiality that is ensured by the “closed shop” character of the board sections. The present system is currently being intensively discussed and dramatic changes are under serious consideration, mainly caused by the predicted increase in the number of applications, which is expected to double in the coming years. This raises the concern that the present system may soon reach its limit in terms of the time and effort needed to collect a sufficient number of competent reviews in a reasonable time and with a manageable workload. There are also further concerns with respect to transparency and the treatment of inter-disciplinary projects. In addition, some SNF authorities feel that more space would be needed for the council board to discuss issues of wider strategic and political implications.

4.2.2. Expert Panels and Study Groups

Three of the organizations that were visited in the FWF's FFM implement various types of expert panels and/or study groups, at least in funding schemes that may be termed "open", characterized by a bottom-up approach (no thematic restrictions) and having individual scientists as applicants. A common feature is that reviewing, ranking and decision making are strictly separated. Applications are reviewed by external experts (from the national and/or the international scientific community). Expert Panels, named "Fachkollegien" (DFG), or "Review Advisory Boards" (NWO), then undertake the ranking of applications based on the reviews. In the Academy of Finland, "Review Panels" carry out both reviewing and ranking, occasionally collecting additional external reviews. Members of expert panels are recruited for a limited period of time (a couple of years; honoraria are paid in Finland). All the panels prepare in various ways a suggestion for funding to a council board. This board (or several boards) takes the final funding decision. The major difference from the first form is that in this model a significant number of external (international) experts are committed to cooperate with the council, e.g. in selecting (additional) reviewers and preparing decisions. This is considered to ensure a high level of checks and balances and to facilitate comparative discussions of proposals. A high degree of impartiality is also frequently quoted as an advantage of such schemes. Among the drawbacks are rather complicated logistics, sometimes longer time spans for decisions and higher administrative costs. When such experts panels also act as reviewers, an additional problem may arise with respect to the scientific competence of the panel members: small scientific fields may not be covered adequately and this may lead in turn to rather erratic funding decisions (see the section above on "general trends").

4.2.3. The principle of the Strong manager

The principle of this form may be briefly summarized by the maxim "forget about council boards". Peer review is carried out either by members of large review colleges, where members are assigned for a few years and reviewers are specifically dedicated to individual projects upon demand (about 3,800 review college members in the case of EPSRC), or by specific review panels of on average about 15 members, which are assembled specifically for each call by the programme directors in the funding organization (e.g. the Computer Science Dept of the NSF). The common feature of this decision-making form is that the responsibility for decisions about funding or rejection essentially rests entirely with the programme manager of the funding organization. Various ways of interaction between review panels, programme managers and council CEOs, as well as continuous monitoring by the scientific community and the public, ensure a high degree of transparency and impartiality. In the case of the NSF in the USA, programme managers are usually members of the scientific community and serve in this function for a limited period of time. They usually return to science after their NSF commitment and their work is followed closely by the scientific community. This high level and elaborate system of checks and balances guarantees that programme mangers perform high quality work and minimizes pitfalls such as nepotism. The strength of the model is its efficiency: straight-forward logistics, comprehensive comparative discussions and short decision-making processes. However, a prerequisite seems to be that the scientific community in which the model operates is sufficiently large. In small countries, such as Austria, scientists in various fields of scientific or scholarly research are likely to be too closely interlinked to avoid positive or negative bias in such a procedure.

4.3. Implications for the FWF

The FWF's present system basically follows the scheme described under 4.2.1. and still operates satisfactorily: the average time from proposal submission to decision has decreased over the last few years and now amounts to 4.5 months, which is one of the best for a system relying on written peer review. The response rates of international peers are quite high: about 70 % of requests for reviews are answered positively. Two aspects of the FWF's scheme are worthy of particular mention. 1) The FWF has a "global" budget. FWF funds for various scientific and scholarly fields, as well as for different funding programmes, are "communicating vessels". The FWF strictly applies a bottom-up approach in virtually all of its funding schemes and an increasing demand in a certain field or funding scheme can be met by shifting money from another funding scheme. 2) The FWF is in a unique position among European funding institutions in that it deals with all scientific and scholarly disciplines in a single board. This considerably facilitates the handling of interdisciplinary projects and ensures a high level of checks and balances in the decision-making process. The small size of the Board (fewer than 30 members) ensures sufficient confidentiality of discussions.

Discussions with nearly all sister organizations suggested that the FWF would be well advised to try to maintain the essential features of its present scheme for as long as possible. Thus, the pressure for instant changes or reforms is low. However, numbers and sums of applications and the widening gap between the amount of money requested and the available budget (see section 2) imply that, as is the case in Switzerland, the system may soon run up against its natural limit. The FWF is currently experimenting with various ways of optimizing its present system: recent changes of the legal basis (e.g. a considerable increase in the size of the Board and the available "pool" of internal reporters, which has enabled a marked improvement in internal checks and balances) and organization (electronic processing of applications and internet-based preparation of board decisions) open new opportunities without the immediate need for dramatic changes. Nevertheless, the FWF is closely following developments in other countries and organizations to ensure that it is prepared for forthcoming challenges, such as the establishment of the ERC.

Acknowledgements: I should like to thank the FWF staff members who participated in the Fact Finding Mission in 2004: Sabine Haubenwallner, Laurenz Niel, Anna Prianichnikova, Falk Reckling and Lucas Zinner.

Author:
Rudolf Novak
Head of Department for Development and Strategy – National Programmes
FWF- Der Wissenschaftsfonds (Austrian Science Fund)
Weyringergasse 35, 1040 Wien, Austria
novak@fwf.ac.at

David F.J.Campell

The Research Assessment Exercise 2008 in the United Kingdom

1. Introduction

For the analysis of research evaluation systems (RESs), focusing on academic or university research, Richard Whitley (2005, p. 4) proposes the following typological distinction: "To simplify the analysis, RES are dichotomized into 'strong' and 'weak' forms in order to suggest how stronger variants are likely to have contrasting effects in difficult kinds of funding regimes" (see also Whitley, 2003; furthermore, see Geuna, 1999, and Geuna et al., 2003). From that dichotomized typology it follows that the RES of the United Kingdom (UK) clearly qualifies as a *strong evaluation system*. University research evaluation in the UK is being processed and carried out in the context of the so-called Research Assessment Exercises (RAEs). These RAEs were part of a significant transition and institutional reform process of the whole UK higher education system (Pechar, 2006). In international comparisons of academic governance, also the RAEs themselves already are regarded as a specific ("ideal typical") type of university research evaluation approach (see the discussions in Geuna and Martin, 2001 and 2003). The UK's RAE integrates the attributes of a *comprehensive institutional ex-post evaluation* of university research at national level. In such an understanding, in a European context, the UK (but also countries such as the Netherlands) may be categorized as a "Type A" country. Consequently, "Type B" countries (for example, Germany[1] and Austria) could be defined as not having implemented such comprehensive evaluation schemes (Campbell, 2003, pp. 110, 112; see Figure 1). The growing importance of basic university research for the performance of a knowledge-based society and economy obviously adds to the importance of a proper evaluation system for academic research (on the ramifications of knowledge creation in general, see the chapter contributions in Carayannis and Campbell, 2006; furthermore, see also Campbell, 2006).

[1] Germany, however, is in the process of establishing at the sub-national level, in some provinces (so-called *Länder*), comprehensive university-research evaluation schemes. The *Land* of Lower Saxony serves as a good example for such a recent policy move (Schiene and Schimank, 2005).

The UK already can refer to a sequence of different RAEs. Key criteria for the UK RAEs are: (1) the whole disciplinary spectrum of university research is covered; (2) the national university system experienced several evaluation cycles; (3) methodically, the evaluation procedure is primarily based on peer review, but, secondarily, also refers to indicators; (4) evaluation results are converted into numerical rankings of university departments; and (5) evaluation results directly impact the public basic funding of university research. Our analysis will focus on the concepts and methods of the next RAE, which is scheduled for 2008. Furthermore, this RAE 2008 will be compared with the prior RAE 2001, for the purpose of assessing the evolutionary momentum of university research evaluation in the UK.

Figure 1: A comparative typology of university research evaluation in Europe

A comparative typology of university research evaluation in Europe with regard to the comprehensiveness of institutional "ex post" evaluations.

Type A countries (nations)	
United Kingdom (UK) Netherlands	*Systematic and comprehensive evaluations, at national level and across all disciplines, "systemic and comprehensive approach"; disciplinary-based institutional "ex post" research evaluations.*
Type B countries (nations)	
Germany Austria Finland (1990s?) Switzerland (1990s?)	*Individual and disciplinarily independent evaluations, and without (frequent) references to explicit grading scales: "pluralized and situational approach".*

2. Concepts and methods of the RAE 2008

2.1 Institutional responsibility and supervision of the RAE 2008

The institutional supervision of the RAEs is carried jointly by the four UK higher education funding councils (HEFCs): Higher Education Funding Council for England (HEFCE), the Scottish Higher Education Funding Council (SHEFC), the Higher Education Funding Council for Wales (HEFCW), and the Department for Employment and Learning, Northern Ireland (DEL). These funding councils supply a RAE team, which “manages” the RAE process. Despite the joint responsibility, the HEFCE occupies, by tendency, a “salient” position within this configuration, since the RAE team is locally based at the HEFCE (http://www.hefce.ac.uk), which is in southern England, Bristol.

For the whole RAE process, a website of its own has been set up (http://www.rae.ac.uk). This has the following implications:

- *Transparency:* The RAE 2008 can claim a maximum transparency, crucially supporting arguments in favour of legitimation of the whole evaluation process. All key documents are posted to the RAE website, and are thus exposed to scrutiny and critique.
- *Management quality and communication:* The RAE website defines a crucial reference point for the overall management process of the RAE 2008. All posted key documents can be downloaded for free. Back-and-forth communication between the HEFCs and the UK universities may refer to this website. In-time announcements of the RAE are channelled through the website. In addition, the website serves also as a platform for feedback (and data input).
- *International visibility:* Since the website can be accessed from every location globally, the web-based transparency of the RAE 2008 reinforces also its international visibility. It underscores the UK's lead position of having implemented a comprehensive system of institutional ex-post university-research evaluation. Furthermore, it also sets the stage for an internationally accessible model, which is being reflected by discourses world-wide, and has a potential for replication in other world regions.

2.2 Key documents of the RAE 2008

The first key document, posted to the RAE 2008 website, was: *Initial decisions by the UK funding bodies* (HEFCs, 2004a). Key documents, following in 2004, focused on the process of recruitment of the panel members: *Panel configuration of recruitment* (HEFCs, 2004b) and *Units of assessment and recruitment of panel members* (HEFCs, 2004c). The first key document, released in 2005, offered support to the panel members for developing "assessment criteria" and "working methods" for evaluating the research of UK universities: *Guidance to panels* (HEFCs, 2005a, p. 2). In that line of focus is also placed the second key document of 2005: *Guidance on submissions* (HEFCEs, 2005b, p. 4) concentrates on the "administrative arrangements and data requirements for submissions" to the RAE 2008. This was followed by the draft document *RAE 2008 Consultation on assessment panels' draft criteria and working methods* (HEFCs, 2005c), which gives greater consideration to the "working methods" and "assessment criteria (panel criteria)".[2] These documents are crucial for the management and governance of the RAE 2008. From a discursive viewpoint, they display the unfolding dynamics of the whole process. At the same time, however, they also manifest a high degree of textual redundancy, since a multitude of text blocks are completely identical across these documents. This textual redundancy may be justified by two arguments: taking into account wide-spread phenolmena of a highly selective reading; and offering an opportunity for permanently updating the crucial information. Disadvantages of redundancy are that this might complicate a clear-cut overview and might also increase the efforts to follow a dynamic process.

[2] See: HEFCs, 2005b, p. 4.

2.3 The evolutionary process of the RAE: the recommendations of the Sir Gareth Roberts report

The UK represents a country (national innovation system) that can refer to a well-established and experience-based system of comprehensive institutional ex-post university research evaluations. Five RAEs already were carried out: 1986, 1989, 1992, 1996, and 2001. While the first three evaluation cycles are separated by only three years, these time durations again increased between the RAEs of 1992, 1996 and 2001, extending to four and five years. After completion of the RAE 2001 it was not completely clear, whether the RAE process will (or should) be continued. There were claims that with each new RAE cycle the additionality (or surplus effects), created by earlier evaluations, would decline (see the very interesting debate in Geuna and Martin, 2003; furthermore and more generally, see also Geuna et al., 2003). Earlier arguments for introducing the RAEs, during the period of Conservative governments (Thatcher and Major years), were: (1) applying a market logic to academia; (2) creating a benchmark rationale for "allocating budget cuts" to universities; (3) guaranteeing or even increasing public budgets for the top UK research universities, in an era of general university budget cuts; (4) justifying a continuation of the public basic funding (GUF) for universities, which was (potentially) endangered of becoming replaced by a more expansive public P&P (projects and programs-based) funding (Campbell, 2003, p. 103) through the Research Councils, the main UK institutions for the allocation of public P&P funds.

The political governance shift from the Conservatives to Labour under Tony Blair in 1997, however, did not change the principle direction of the RAE system. The RAE 2001 was implemented and completed, and also the preparations for the new RAE 2008 were put in place. In the UK context we thus can talk about an "evolution of evaluation"[3] or a "co-evolution of research and research evaluation" (Campbell, 2003, pp. 111, 124-125). From a policy decision-making perspective this offers the advantage and opportunity to carry out empirical analyses about the impact of the RAE system on university research in the UK. The HEFCs commissioned and published several reports (for example, see HEFCE, 2005a). The HEFCs (2004a, p. 4) claim that the "RAE is generally agreed to have had a significant positive impact". But also the HEFCs (again 2004a, p. 4) acknowledged the criticism against the RAE system, particularly the following aspects:

- The RAE would favour established disciplines at the cost of interdisciplinarity;
- The RAE would not consider sufficiently application-oriented research;
- The RAE produces an "academic burden";
- The RAE would alter academic "institutional behaviour" in a way to achieve perfect adaptation to the RAE criteria.

[3] Wilhelm Krull coined the phrase of "evolution through evaluation" (cited according to Campbell, 2003, p. 126).

In an effort to encourage learning processes for the whole RAE system, the four UK HEFCs commissioned a comprehensive RAE review to Sir Gareth Roberts (2003). The Roberts report integrated 420 responses and the feedback of 40 meetings. Sir Roberts was supported by a steering group and officers, located at the HEFCE. After completion and out of an interest to encourage further consultation, the Roberts report was posted publicly accessibly on the RAE website. The HEFCs-guided consultation process ended on September 30, 2003. For the HEFCs (2004a, p. 5) the main conclusions were:

- The RAE system claims that there exists a broad consensual support for the REAs: “Overwhelming support for an assessment process built around expert review conducted by disciplinary panels” (HEFCs, 2004a, p. 5);
- Support for a six-year evaluation cycle;
- Support to replace the rating scale by a quality profile;
- Support for a closer cooperation of panels of related disciplines;
- Support for an improved quality recognition in applied, new and transdisciplinary disciplines;
- Support for preventing the development of too complicated review methods by the panels.

2.4 What is the RAE?

In their own words, the HEFCs (2004a, p. 4) offer the following self-definition for the RAE: “From the start, the RAE exercise has been an expert review process in which discipline-based panels of experts – mainly, but not exclusively, people working in research within the higher education sector – assess the quality of research in their own discipline.” Rephrased, an implication is: *The RAE represents a system and a process, in which the quality of all university research, on a comprehensive national level, is evaluated in the context of an institutional ex-post evaluation, on the basis of a disciplinary matrix and by applying the method of peer review.* Broken down in more specific components, additional key features are:

- *Institutional comprehensiveness at the national level:* The RAE addresses all UK universities or, to use a more precise terminology, all UK-based higher education institutions (HEIs). The HEFCs (2005b, p. 3) embodied this in the following wording: “In December 2006 the four UK higher education funding bodies will invite all eligible UK higher education institutions to make submissions in the RAE 2008.” Consequences, however, are: only those UK universities are (continuously) eligible for receiving public basic funding (GUF), which also participate in the RAE. Non-participation is equivalent to being excluded from GUF. Public P&P funding, obviously, is not affected by RAE decisions.

- *Disciplinary-based peer review:* A certain number of disciplines are defined during the preparatory process of the RAE. In technical terminology, these disciplines are called, by the HEFCs, the *units of assessment* (UoAs). Consequently: first, all UK university departments have to decide (self-decide) on their disciplinary assignment. Second, for each discipline (unit of assessment) an expert panel, called *assessment panel*, is being installed. These assessment panels then carry out the disciplinary-based peer review of the RAE.
- *Quality of university research:* The assessment panels, and thus the whole RAE, concentrate on the quality of university research, which represents the (only) core dimension. Consequently, one may postulate that the RAEs focus on *"one" dimension of university research, i.e. quality*. In other words, the RAEs modelled the quality of university research in a one-dimensional understanding.
- *Ex-post profile:* The RAE is ex-post in its profile, because it focuses retrospectively on already conducted and completed research. The RAE 2008 will take the time period January 2001 until December 2007 into account (HEFCs, 2005b, p. 3).

In addition, the RAE 2008 is guided by the following self-proclaimed principles (HEFCs, 2005b, pp. 5-6):

- *Equity:* "All types of research and all forms of research output" across the whole disciple-nary spectrum must be treated on a "fair and equal basis". No segment of the disciplinary spectrum may be discriminated against.
- *Diversity:* The RAE 2008 is interested in reflecting the whole spectrum of diversity of university research across all UK university institutions.
- *Equality:* Equal opportunity measures, and their possible effects on research output, will be recognized.
- *Expert review:* The "discipline-based expert review" represents the core methodic approach. It is understood that panel members themselves have been or are still actively engaged in high-quality research. "Quantitative indicators" only serve as additional information.
- *Consistency:* Concepts and methods of the RAE should be consistent within each cluster of related disciplines (governed by the *main panels*).
- *Continuity:* There is a certain trade-off and challenge for a balance-striking between "continuity" and "development" of the different RAE cycles. In the words of the HEFCs, the "RAE has developed through an evolutionary process, building on learning from previous RAEs".
- *Credibility:* The HEFCs claim a high credibility of the RAE to those who are being assessed. This refers particularly to the "fundamental methodology, format and process employed", based on disciplinary-based peer review. To support the credibility further, the HEFCs reinforces and encourages an embedding of the RAEs in comprehensive

consultation processes – for example with the academic communities, which are exposed to the RAE.

- *Efficiency:* The HEFCs are interested in keeping the costs and the burden of the evaluation cycles for the evaluated UK universities as low as possible. The HEFCs claim that the past RAEs have been "highly cost-effective given the value of public funds distributed through their ratings". For example, the costs of the RAE 1996 were estimated only to amount 0.8 % of the public funds allocated in reference to the results of the RAE 1996.
- *Neutrality:* The RAE wants to measure and improve the quality of university research. The RAE is interested in preventing distortions of university research and, furthermore, is inclined not to function as a misleading benchmark (*environment*) for university research activities.[4]
- *Transparency:* Through transparency of the whole RAE process, and subsequent consultation mechanisms, the credibility of the RAE should be reinforced. Furthermore, decision-making processes, targeting the RAE or referring to RAE results, are openly explained.

2.5 Temporal duration of the evaluation cycles

Five RAEs already were carried out (1986, 1989, 1992, 1996, and 2001). The RAE 2008 represents the sixth RAE. It displays current consensus that RAEs should be carried out in the format of a "six-year cycle" (HEFCs, 2004a, p. 3), implying that the RAE 2008 might be followed by a RAE 2014. Obviously, this pattern of RAE evaluation cycles is subject to potential change in the future. The specific sequence of RAEs supports the impression that the whole RAE system had to learn what a good temporal patterning for institutional ex-post university research evaluations would be. Earlier cycles of RAEs were grouped closer together, whereas later RAEs allowed more extended time intervals in between.

De facto, the RAE 2008 will be carried out 2007-2008 (HEFCs, 2005b, pp. 3, 37). The "closing date" for submissions is November 30, 2007. Submissions reflect publications and/or research output of the "publication period" January 1, 2001, until December 31, 2007. The "census date" for information on active staff is October 31, 2007. Early administrative preparations of the RAE 2008, initiated by the HEFCs, already started in January 2005 with the issuance of guidances for the panels (see HEFCs, 2005a). A year before, 2004, the HEFCs launched the release of official background documents on the whole RAE 2008 procedure (e.g., see HEFCs, 2004a). The submissions, of the UK university departments (due November 2007), will be evaluated by the expert panels during the course of the calendar year of 2008. Publication of the

[4] A successful co-evolution would imply that the RAEs added to a quality improvement of UK university research. It still remains to be tested, whether the HEFCs will use, in future references and documents, the terms of "environment" and "co-evolution" for self-describing and assessing the RAE.

results of the RAE 2008 will be released in December 2008, and will impact the public basic funding formula of UK universities, beginning with the academic year of 2009-2010.

2.6 Number of disciplines

In technical language the disciplines are called units of assessment.[5] It has been decided that the RAE 2008 will refer to 67 disciplines (units of assessment). For each of these units a specific expert panel, a so-called *sub-panel*, is responsible. These 67 units of assessment are grouped together in 15 *main panels*, which represent macro-disciplines or disciplinary clusters (HEFCs, 2005c, pp. 3, 20-21).

In that context two observations are interesting. First, observed over time, there is a tendency that with each RAE cycle the number of disciplines (units of assessment) declines, dropping from 92 (RAE 1989) to 72 (RAE 1992), 69 (RAE 1996), 68 (RAE 2001), and finally 67 for the current RAE 2008. This reflects the circumstance that a too high number of disciplines do not necessarily produce more accuracy, but creates also assignment ambiguities. Second, this disciplinary structure could be regarded as a "conservative approach" (Campbell, 2003, p. 115). This may be one of the consequences of a peer review-based approach, which has to reflect, to a certain degree, the established consensuses of the academic communities, for generating acceptance (among the evaluated) and thus reinforcing the criterion of "credibility" of the whole RAE process (see again Chapter 2.4). Without acknowledging some patterns of disciplinary consensus, an institutional ex-post university research evaluation system would run the risk of being exposed to severe criticism. Disciplines offer further guidelines for "legitimate" procedures and methods of research assessment. *Interdisciplinarity* could be regarded as a quality attribute of disciplines within such a disciplinary framework.

2.7 Smallest institutional unit for research evaluation, content of submission, and membership selection for the expert panels

In principle, the university departments act as the "smallest institutional unit" for the RAE. Every university department must decide on its disciplinary assignment to the 67 units of assessment, organized and assessed by expert sub-panels. The HEFCs (2005b, p. 47) defines a "department" in the following words: "The staff included in a submission to one of the 67 discrete units of assessment recognized by the RAE, and, by extension, their work and the structures which support it. RAE departments are often not identified with a single administrative unit within an HEI". Default is that one department forwards *one submission* to the appropriate and responsible disciplinary-based panel.[6] However, the system is flexible enough for allowing variability (HEFCs, 2005b, pp. 11-12). *Joint submissions* indicate that several (more than one) departments collaborate in putting together a submission for a single

[5] In addition, these disciplines also are classified as specific "subject area[s]" (HEFCs, 2005a, p. 25).

[6] "Normally there should be only one submission per UOA per institution and only exceptionally will this be waived" (HEFCs, 2005b, p. 11).

unit of assessment: in such a situation a virtually expanded university department is being created. Contrary to that, a *multiple submission* implies that a university department is disaggregated into “academically and structurally distinct” units, which are treated independently for the purpose of the RAE. Out of a growing awareness and sensitivity for “inter-disciplinary research”, the RAE 2008 emphasizes two specific tools: (1) different submissions can be partially linked by arrangements of cross-references. (2) Furthermore, *specialist advisers* should be contacted and consulted (HEFCEs, 2005b, pp. 12-13).

The content of the information submission of each university department, forwarded to the units of assessment, must take the following considerations into account:

- *Staff information:*[7] Departments are required to provide comprehensive staff information. Key in that context is the so-called “research active academic staff”, which must be documented individually and also in FTE (full-time equivalents). Research active staff is distinguished according to four different categories (A, B, C, and D). Category A, for example, indicates an active employment status on the census date. Additional staff information refers to: research students; research studentships; individual staff circumstances; and category C staff circumstances (HEFCs, 2005b, pp. 13, 15-19).
- *Research output:* For every research-active staff member (Category A and C) “up to four items … of research output”, which have been “brought into the public domain during the publication period” of January 1, 2001, until December 31, 2007, should be documented.[8] Often, but not necessarily, a research output will be a publication. In that case, a standardized information set must be provided for every named publication output. However, the RAE process also accepts output in a “non-text” format. The crucial benchmark is that the output reached the public domain, implying a public accessibility. The “physical form” of the output should be reported, and also, “where it may be found” (HEFCs, 2005b, pp. 13, 19-21).
- *Additional information* refers to “external research income” and “textual” descriptions. External research income focuses on further sources of financing, placing a special emphasis on different public funds, the “UK-based charities” (foundations), the UK firms, and EU institutions. One purpose of these textual descriptions is to collect more information about the “research environment” and “indicators of esteem” (HEFCs, 2005b, 13, 25-26, 29).

[7] The UK terminology uses this category of *research-active academic university staff*. By contrast, in the U.S. context these British staff categories would be called “faculty”, because the American system associates with “staff” primarily administrative tasks.

[8] This specific criterion underscores that the UK RAE is more oriented toward the quality of key publiccations, and does not one-sidedly favour an aggregation of purely quantitative indicators (Campbell, 2003, p. 117). This fact is sometimes being misperceived outside of the UK.

To support the submission forwarding of the university departments, the RAE team developed a specifically tailored software. All university departments (HEIs) are obliged to use a web-based database application that is hosted at the HEFCE. Furthermore, the RAE 2008 will implement mechanisms of "data verification" of the reported departmental submissions (HEFCs, 2005b, pp. 9-10).

The nomination process for panel membership of the RAE 2008 was crucially governed by the HEFCs. Beginning with July 2004, the HEFCs asked for nominations of members to the panels of the RAE 2008. Specifically, "stakeholder organizations"[9] were asked for their opinions (HEFCs, 2004c). Nominations from individuals or HEIs were not accepted. Most of the organizations, forwarding nomination recommendations, also had already participated in the RAE 2001. However, this list of organizations for the nomination process of the RAE 2008 was extended, after having carried out a consultation process with HEIs (HEFCs, 2004b). All together, the HEFCs received 4948 nominations from 1371 organizations. In a separate tier, the HEFCs advertised in the press for applications for the main panel chairs. From 106 applications, the chief executives of the four HEFCs decided on the 15 main panel chairs (which are from 12 HEIs in the UK) in October 2004. In advice consultation with the main panel chairs, the HEFCs' chief executives selected, furthermore, the chairs and members of the sub-panels. One objective of these panel membership decisions was to represent in a balanced format the diversity of the UK university system. All panel members are documented publicly on the RAE's website (HEFCs, 2006a). On average, the main panels consist of approximately ten members, and the sub-panels of 10-20 members. The actual number of panel members varies across different disciplines (units of assessment). In addition to the panel members, the HEFCs also carried out a nomination procedure for the *specialist advisers*, who should support the sub-panel members in their assessment work, particularly in areas of significant interdisciplinary research (furthermore, see HEFCs, 2004b, pp. 5-9).

Within this *two-tier architecture* of sub-panels and main panels the following functional division of labour operates, creating a multi-level system of enhanced quality checking (HEFCs, 2004a, pp. 8-9; see also HEFCs, 2004c, pp. 8-9):

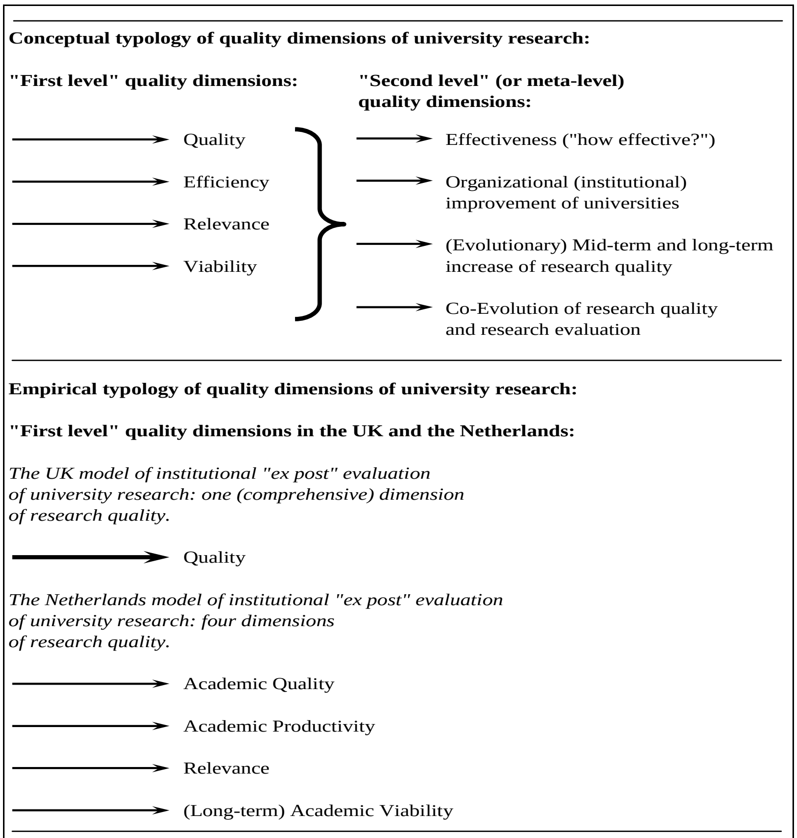
- *Sub-panels:* They are responsible for (1) drafting the "relevant criteria" and "working methods" for each discipline, and for (2) drafting the research assessment results for each university department.
- *Main panels:* Their portfolio is to – (1) finally decide on the "relevant criteria" and "working methods" within the context of each sub-panel (discipline); (2) make the final assessment decision for the research quality of university departments; (3) and to sustain and to progress the communicative interaction with the other main panels.

9 Examples for such stakeholder organizations are: subject associations; professional bodies; practitioner bodies; and commercial organizations.

2.8 Number of dimensions for the quality of university research: the applied grading schemes

Up until currently, the UK RAEs referred only to one quality dimension of university research, understood as *quality*. The one main purpose of the RAEs was exactly to measure the quality of UK university research. Other countries, also applying comprehensive institutional ex-post evaluations of university research, decided to opt for a multi-dimensional modelling of university research. The Netherlands, for example, refer to a four-fold dimensional typology: quality, productivity, relevance, and long-term academic viability. Furthermore, one may also raise the question, whether, from an analytical perspective, a distinction between so-called "first-level" and "second-level" quality dimensions creates additional key information about the evolution of research evaluation systems (Campbell, 2003, pp 109-111, 117; see also Figure 2). The RAE 2001 used the following 7-point rating scale; 1, 2, 3b, 3a, 4, 5, and 5*, with 5* representing the best grading. Since the rating in some disciplines, by tendency, increased with each RAE cycle, it was decided, for the RAE 2001, to allow a broadening of the rating scale. This also was understood as preventing a "ceiling effect" on the upper end of the rating scale: on top of the original 5, therefore, a 5* was added, and 3 became subdivided into 3b and 3a. This, obviously, inspired debates, whether these rating improvements could be regarded as an indication for a quality improvement of UK's university research (Campbell, 2003, p. 118).

Figure 1: Quality dimensions of university research



Source: Campbell (2003, p. 111).

Here, the RAE 2008 marks an important difference, since modifications for the grading ratings were introduced. To allow for a more diversified information about the quality of university research, it was, at least *de facto,* decided to move from a one-dimensional to a multi-dimensional modelling of university research. This also represents a direct consequence and learning effect of the commissioned review of the RAEs by Sir Gareth Roberts (2003). In a follow-up consultation process, the HEFCs (2004a, p. 5) arrived at the conclusion of replacing the earlier "rating scale" with a "quality profile". Already for the RAE 2008, this new *overall quality profile* will be implemented and is being understood as an aggregation of three *separate and distinct dimensions (sub-dimensions).* These *three dimensions*[10] are: (1) research outputs; (2) research environment; (3) and esteem indicators[11]. Research output is defined by the RAE process in a standardized format (HEFCs, 2005a, p. 25; 2005b, pp. 13, 19-21). With regard to the research environment and esteem indicators the main panels have a certain privilege to determine, which component of the departments' submissions fall into the one or other category. As a consequence, the drawing of the (conceptual) boundaries between research environment and esteem indicators may differ across clusters of disciplines (HEFCs, 2005a, p. 25). Also the aggregation of these three dimensions to the overall quality profile may vary, and again is being finally decided by the main panels. The HEFCs (2005a, pp. 25-26) only define the *minimum* "percentage weightings" for the dimensions: 50% for research outputs, 5% for research environment, and 5% for the esteem indicators.

The quality profile of each dimension is expressed in terms of "quality levels", leveraging a 4-point rating scale: 1*, 2*, 3*, and 4*, where again 4* qualifies as the highest score. One could argue that the category "unclassified" may be regarded as a virtual fifth rating point, placing on the bottom. Basically, "unclassified" implies that research quality falls below a minimum or acceptable quality threshold or, alternatively, represents research work, which is not compatible with the research definitions of the RAE. New for the RAE 2008 is also that for each dimension, and their aggregation to the *overall quality profile,* not an average score value is being reported, but the overall distribution, ranging from "unclassified" to 4*. The score distribution of the aggregated *overall quality profile* represents the final *quality level* (HEFCs, 2005a, pp. 24-25; 2005c, pp. 16-19). *Score distributions* (in opposite to *score averages*) express the advantage of displaying a more diversified picture of research quality performance of university departments. To exemplify the content or meaning of these different grades, we quote the RAE's definitions for 1* and 4* (HEFCs, 2005c, p. 16):

- *1*:* "Quality that is recognized nationally in terms of originality, significance and rigour".
- *4*:* "Quality that is world-leading in terms of originality, significance and rigour".

[10]) The RAE system labels these *dimensions* as "different components" (HEFCs, 2005a, p. 25).

[11]) Put in a different wording, these "esteem indicators" may be interpreted as indicators that express, what the reputation of a university department is.

3. Conclusion: The Long-Term Effects of the RAEs

British evaluation results have a direct and formalized impact on the public basic funding of universities. In the UK, these evaluation scores are tied into a funding formula, which determines the amount of public funds, which the universities receive. For the section of the formula, referring to the financial support of university research, operates the following mechanism: *amount = quality x volume* (Campbell, 2003, pp. 119-120). The different components mean: (1) *amount* is the amount of public funds; (2) *quality* are the resulting (numeric, quantitative) quality scores of the RAEs; and (3) *volume* indicates the research active academic staff.[12] For example, every year the HEFCE (Higher Education Funding Council for England) publishes a report with detailed insights into how it allocates its funds to HEIs: of the total public basic funding for universities in England, in the academic year 2005-06, about 19.8 % were devoted to research. This research money is calculated on the basis of a "quality-related (QR) funding", mainly acknowledging the criteria of "quality" and "volume". HEIs, which, in the context of the RAE 2001, only received evaluation scores in a range from 1 to 3a, are excluded from receiving public basic funding for university research (HEFCE, 2005b, pp. 6, 18-21).

Since the RAEs already operate over a longer period of time in the UK, this offers an opportunity for engaging in analyses on the long-term effects of the RAEs on the UK university base and on UK university research. Systems of university research and university research evaluation are tied together in processes of co-evolution (see again Chapter 2.3). The RAEs were exposed to systematic assessments from outside (e.g., Roberts, 2003; House of Commons, 2004). The RAE interest focuses on creating synergies between *excellence* and *selectivity* (Adams et al., 2000). The already mentioned quality-related (QR) funding formula displays a highly selective orientation, which is also explicitly underscored by the HEFCE, formulated in the following words: "As a result, our funding of research is highly selective" (HEFCE, 2005b, p. 20). Possible ramifications of the RAEs on the behaviour of UK academics and their publication strategies for scholarly journals have been investigated in greater detail (e.g., see Walford, 1999). In September 2005 the HEFCE released a commissioned study that assessed the effects of the quality-related funding formula on research behaviour and research performance. Key findings, claimed by that specific study, are (HEFCE, 2005b, pp. 35-42):

- *Performance of the system:* Research performance of the UK's universities improved since the mid-1980s (the first RAE was completed in 1986). Crucial in that context is academic research performance, measured by citations of scholarly work.
- *Performance of institutions:* The UK's university departments (and faculties, schools) are currently more inclined, or in a better position, to expand their research portfolio that is based on research grants.

[12]) This formula design implies that an underreporting of research active academic staff will result in decreases of public basic financing.

- *Behaviour of institutions:* There may be a growing gap between the better and weaker performing HEI's institutions in the UK. The higher performing attract more public resources, thus leading to speculations that we should be prepared to expect increasing differences across UK's domestic university base in the future. University management reacts more sensitively to evaluation outcomes.

- *Performance and behaviour of individuals:* Pressures for academic individuals, to increase their research output, have increased. The incentives of strong research evaluation systems (RESs) may foster the more established mainstream work, whereas interdisciplinary research could be perceived as being coupled with greater risks, since it involves uncertain outcomes.

Despite such in-depth documentations of the effects of the RAEs on the research of UK's universities, *it is still not easy to find data that demonstrate systematic shifts of allocations of public funds.* Studies frequently claim that funds increasingly concentrate on the top performers, whereas the lower performers fall back: "Resources have become more concentrated among the highest performers, so there is proportionately less for others. It is a reasonable hypothesis that this may have led to a situation in which the good get better and the less good decline further, but whether this is actually what happened remains unclear" (HEFCE, 2005a, pp. 37-38). An empirical test would have to juxtapose – perhaps in the context of scenarios –, what the differences would be of (1) a funding formula, primarily focusing on "volume" (representing an old "input-oriented" system, where past financial transfers determine future financial transfers), and (2) the UK funding formula that is currently in place, combining "volume" with "quality" (evaluation outcomes support an "output- oriented" design). Perhaps such studies already were conducted. But then they are not (easily) accessible through the internet. *This marks crucially a weak point of the current debates about the effects of the RAEs on UK's university research*.

References:

Adams, Jonathan / Nicholas Cook / Graham Law / Stuart Marshall / David Mount / David Smith/ David Wilkinson / Nancy Bayers / David Pendlebury / Henry Small / Jessie Stephenson (2000). The Role of Selectivity and the Characteristics of Excellence. Report to the Higher Education Funding Council for England: A consultancy study within the Fundamental Review of Research Policy and Funding. Leeds: Evidence Ltd. [http://www.evidenceuk.com/downloads/selectivity-report.pdf].

Campbell, David F. J. (2003). The Evaluation of University Research in the United Kingdom and the Netherlands, Germany and Austria, 98-131, in: Philip Shapira / Stefan Kuhlmann (eds.): Learning from Science and Technology Policy Evaluation: Experiences from the United States and Europe. Camberley: Edward Elgar.

Campbell, David F. J. (2006). Nationale Forschungssysteme im Vergleich. Strukturen, Herausforderungen und Entwicklungsoptionen. *Österreichische Zeitschrift für Politikwissenschaft* 35 (*forthcoming*) [http://www.oezp.at].

Carayannis, Elias G. / David F. J. Campbell (eds.) (2006). Knowledge Creation, Diffusion and Use in Innovation Networks and Knowledge Clusters: A Comparative Systems Approach Across the United States, Europa and Asia. Westport, Connecticut: Praeger [Book summary available via: http://www.greenwood.com/books/bookdetail.asp?sku=Q486].

Geuna, Aldo (1999). The Economics of Knowledge Production. Funding and the Structure of University Research. Cheltenham: Edward Elgar.

Geuna, Aldo / Ben R. Martin (2001). University Research Evaluation and Funding: An International Comparison. Brighton: Electronic Working Paper Series. Paper No. 71 (University of Sussex) [http://www.sussex.ac.uk/spru/publications/imprint/sewps/sewp71/sewp71.pdf].

Geuna, Aldo / Ben R. Martin (2003). University Research Evaluation and Funding: An International Comparison. *Minerva* 41, 277-304.

Geuna, Aldo / Ammon J. Salter / W. Edward Steinmueller (2003). Science and Innovation. Rethinking the Rationales for Funding and Governance. Cheltenham: Edward Elgar.

HEFCE (eds.) (2005a). Impact of selective funding of research in England, and the specific outcomes of HEFCE research funding. A desk-based review for HEFCE and the Department for Education and Skills by Evidence Ltd. Summary and Report. HEFCE: Bristol [http://www.hefce.ac.uk/pubs/rdreports/2005/rd21_05/].

HEFCE (2005b). Funding higher education in England. How HEFCE allocates its funds (July 2005/34). HEFCE: Bristol [http://www.hefce.ac.uk/pubs/hefce/2005/05_34/].

HEFCs (2004a). Initital deisions by the UK funding bodies (Ref RAE 01/2004). HEFCE: Bristol [http://www.rae.ac.uk/pubs/2004/01/].

HEFCs (2004b). Panel configuration and recruitment (Ref RAE 02/2004). HEFCE: Bristol [http://www.rae.ac.uk/pubs/2004/02/].

HEFCs (2004c). Units of assessment and recruitment of panel members (Ref RAE 03/2004). HEFCE: Bristol [http://www.rae.ac.uk/pubs/2004/03/].

HEFCs (2005a). Guidance to panels (Ref RAE 01/2005). HEFCE: Bristol [http://www.rae.ac.uk/pubs/2005/01/].

HEFCs (2005b). Guidance on submissions (Ref RAE 03/2005). HEFCE: Bristol [http://www.rae.ac.uk/pubs/2005/03/].

HEFCs (2005c). RAE 2008 Consultation on assessment panels' draft criteria and working methods (Ref RAE 04/2005). HEFCE: Bristol [http://www.rae.ac.uk/pubs/2005/04/].

HEFCs (2006a). Panel members for the 2008 RAE. HEFCE: Bristol [http://www.rae.ac.uk/panels/members/members.xls].

House of Commons (2004). Research Assessment Exercise: a re-assessment. Eleventh Report of Session 2003-04. Report, together with formal minutes, oral and written evidence. Ordered by the House of Commons to be printed 15 September 2004. London: The Stationary Office Limited [http://www.publications.parliament.uk/pa/cm200304/cmselect/cmsctech/cmsctech.htm].

Pechar, Hans (2006). Vom Vertrauensvorschuss zur Rechenschaftspflicht. Der Paradigmenwechsel in der bitischen Hochschul- und Forschungspolitik seit 1980. *Österreichische Zeitschrift für Politikwissenschaft* 35 (*forthcoming*) [http://www.oezp.at].

Roberts, Gareth (2003). Review of research assessment. Report by Sir Gareth Roberts to the UK funding bodies. Issued for consultation May 2003. HEFCE: Bristol [http://www.ra-review.ac.uk/reports/roberts.asp].

Schiene, Christof / Uwe Schimank (2005). Research Evaluations as Organizational Development. Some Experiences from the Work of the Academic Advisory Council in Lower Saxony. Draft Paper for the Sociology of the Sciences Yearbook Conference on “Changing Knowledge Production through Evaluation” (June 9-11, 2005). Bielefeld: University of Bielefeld.

Walford, Leo (1999). The Research Assessment Exercise: its effect on scholarly journal publishing. *Learned Publishing* 13, 49-52 [http://www.ingentaconnect.com/content/alpsp/lp/2000/00000013/00000001/art00007].

Whitley, Richard (2003). Competition and Pluralism in the Public Sciences: The Impact of Institutional Frameworks on the Organization of Academic Science. *Research Policy* 32, 1015-1029.

Whitley, Richard (2005). Path Dependency and Institutional Change in the Public Sciences: The Consequences of Research Evaluation Systems for Knowledge Production in Different Research Systems and Scientific Fields. Draft Paper for the Sociology of the Sciences Yearbook Conference on “Changing Knowledge Production through Evaluation” (June 9-11, 2005). Bielefeld: University of Bielefeld.

Author:

David F.J. Campbell
IFF (Faculty for Interdisciplinary Studies, University of Klagenfurt)
Hochschulforschung | Higher Education Research
Schottenfeldgasse 29, 1070 Vienna, Austria
david.campbell@uni-klu.ac.at

Alfred Radauer, Klaus Zinöcker

The Usage of PART in the European Context - Possibilities and Caveats

1. Introduction

New institutions, programmes and initiatives usually come into being in a rather complicated way. The players involved consists of political representatives, civil servants, experts and employees of funding institutions, as well as researchers and representatives of private enterprises who are also integrated accordingly in the planning, consulting and discussion processes. All are part of the R&D stakeholder community.

Relevant knowledge and information is indispensable for all those involved. The more knowledge is openly available and reflected upon by stakeholders together, and the more stakeholders hold joint beliefs, the more successful discussion and work processes can be and the more things will go according to plan. Evaluation has been an important pillar (however, not the only one) in this context, especially programme evaluations. Despite different methodological issues, programme evaluations try to answer the question “Does the programme work?” and, further to that, provide information to policy makers with regard to possible respective policy actions.

Apart from that basic question, which is founded in the demand for greater accountability for the way taxpayer’s money is spent, one can observe that different countries have also developed different evaluation cultures. In the U.S., peer-reviewed evaluations dominate and there is a strong focus also on quantitative outcome measurement. By contrast, the European evaluation system does not emphasise either qualitative or quantitative methods and uses evaluation studies according to the development stages of a programme in a loop-like manner by applying a so-called “policy-cycle model”. Both approaches have their proponents and opponents, and it is interesting to see whether one side can learn and benefit from the achievements of the other side.

One such case in point is PART, the Programme Assessment Rating Tool, developed by the Office of Management and Budget (OMB), and its application to measure and compare outputs and outcomes of federally funded support programmes. This paper analyses the possibilities and caveats of using such a tool in the European evaluation system, as similar instruments do not yet exist in Europe. We will start off by briefly describing and comparing the characteristics of the U.S. and the European evaluation systems; continue with laying down the foundations of PART and analysing the experiences American researchers have had with its usage, describing advantages and drawbacks with the European way of doing evaluation (also by taking examples from Austria); and, eventually, gauge the possibilities of implementing PART in Europe. The latter could and should be seen as a starting point for further discussion.

1 The Evolution of Programme Evaluation Practices

1.1 The Evaluation System in the U.S.A

One of the main features of the evaluation system for R&D programmes in the U.S. (if compared to Europe) seems to be its interest in (and now mandatory application) of quantitative output and outcome measurement with the aim to increase accountability to political authorities (Roessner 2002 for the U.S. system, in comparison to Luukkonen 2002 who describes the European system). The first attempt to implement government-wide formal management and output measurement techniques in the U.S., the Planning-Programming-Budgeting System (PPBS), dates back to the 1960s. It was followed by the Management by Objectives Technique (MBO) during the Nixon administration, by Zero-Based Budgeting (ZBB, Carter Administration), and by Total Quality Management TQM (Roessner 2002). However, methods such as Cost-Benefit Analysis have been used as a decision tool as early as the 1930s (Moore 1995).

Of the more modern approaches, it is especially the GPRA (Government Performance and Results Act) -- coming from the legislative branch of government during the Clinton Administration -- and PART (Programme Rating Assessment Tool) – coming from the executive branch of government as a component of President Bush’s Management Agenda -- which are noteworthy and have stirred up controversial debates.[1] The GPRA, enacted in 1993, calls for the development, application and yearly update of outcome/performance measurements for federally-funded support programmes. Under GPRA, federal agencies are required to develop long-term strategic plans, and to publish annual reports which describe, on one hand, actual performance vs. expected performance goals and, on the other hand, ways to achieve the long-term results aimed for in the strategy plan (Mercer 2005). PART is, by contrast, a specific instrument “*to evaluate programme performance, determine the causes for strong or weak performance and take action to remedy deficiencies and achieve better results*”.(PART Update, cit. in Ruegg 2004) It is been said that PART gave “teeth” to the GPRA (e.g., Riggle 2005). The way PART functions will be described in detail below.

Roessner states that there is some belief that GPRA and PART will eventually lead to the creation of performance management systems embedded in strategic planning that will make programme evaluation obsolete (Roessner 2002). This may be regarded as a provocative statement, overemphasizing an ideal (maybe by some sought-for) case where the PART questionnaire can replace all “proprietary” evaluation techniques. In fact, it seems that PART may need to rely heavily on program evaluations to convince OMB examiners of the well-functioning of the programmes. Programme evaluation studies have been playing an important role in reviewing and analysing S&T programmes, both before and during the GPRA/PART era. It has to be noted, too, that agencies are free to choose when and how to conduct evaluation exercises in the context of GPRA and PART, as long as they provide the required evidence.

[1] For a more detailed description of GPRA and PART, see Mercer 2005.

Programme evaluations carried out in the U.S. can be classified according to the research questions they addressed (Roessner 2002):

- Studies that investigate whether a programme should be continued or terminated (summative evaluation)
- Studies that analyse the economic payoffs (ROI, cost/benefit analysis)
- Studies that scrutinize how outcomes and impacts from the programme are realized (process or formative evaluation)
- Studies that investigate whether the programme generates results in line with accepted standards of scientific quality (merit review, peer review).
- Studies that analyse the measurable outputs from a programme (monitoring, performance management).

According to Roessner, evaluations studies using peer reviews have been used most in the U.S. and are still prevalent today. Peer reviews can be thus considered the second outstanding feature of the American evaluation system. They are the backdrop against which all other types of research evaluations appear and the standard against which other methods have been judged. The significance of peer reviews (not only for the ex ante assessment of research proposals, but also for the ex-post analysis of R&D programmes in basic and applied research) is underlined by several statements of the National Academy of Sciences. For example, in 1982 the Academy stated that *"...any additional evaluation procedures [beyond peer review] should be introduced only if they clearly enhance rather than constrict the environment in which the research proceeds, and that formal techniques cannot usefully replace informed technical judgement."* Similarly, the Academy assessed in 1999 that *"… the most effective means of evaluating federally funded research programmes is expert review."* With the influence of the National Academy of Sciences in mind, one can infer that – despite GPRA and PART – peer reviews will continue to be the method of choice for the ex-post analysis of R&D programmes. It has to be noted, though, that peer reviewers increasingly revert to quantitative methods for their analysis (e.g., by commissioning respective studies), which can be regarded as a departure from the "pure" peer review process (e.g., Wessner 2000).

Evaluation studies using other methods were first conducted more in the 1960s and became even more popular in the 1970s. These studies (and also a large part of the evaluation studies afterwards) applied mainly quantitative methods, such as econometric analysis or formal cost/benefit methods. Roessner suggests that the popularity of theses studies stems from fact that quantitative methods appear very structured and rigorous and at the same time deliver operational figures (i.e., "hard facts") to policy makers.

For the future, it is assumed that improvements in outcome measurement will be linked to advances in innovation theory. There is considerable criticism in the U.S. on the way performance measures are currently used, given that innovation processes in R&D are far from being fully understood (Roessner 2002, but also Feller 2002).[2] Opponents of the current system point to three aspects which should be taken into account more in outcome measurement and where they see a need for further research:

1. There is evidence that networking plays a much more crucial role for the creation and application of knowledge than commonly anticipated. Networking effects are, however, traditionally hardly considered in performance review systems. This shortcoming is addressed by more recent efforts (Valdez 2005).
2. It has been also noted that the "value" of research outputs (however measured) is highly skewed (i.e., a very small number of supported projects in a R&D programme usually accounts for the lion's share of the programme outcome). Again, current outcome measurements schemes are hardly able to reflect this type of distribution.
3. In order to grasp the full scale of effects of how research activities yield socially desirable outcomes, it is suggested to emphasise formative evaluation methods, especially case studies.

1.2 The Evaluation System in Europe

In analysing the European evaluation system, one has to keep, first of all, the heterogeneity of the European countries in mind: Different European countries have different legislative systems which influence the way evaluations are carried out (Luukkonen 2002). The U.K and France[3], for example, have very centralised frameworks as opposed to Germany or the Netherlands, where evaluation was well-established but uncoordinated between ministries and agencies. Most of the Southern European countries feature rigid legislative frameworks which constrain the development of an evaluative culture. The Nordic countries, on the other hand, have a long-standing tradition in evaluation involving heavy use of overseas/foreign panellists. Laredo (Laredo 2005) describes the fast changing policy context within Europe which directly influences evaluation practices: new processes for priority settings, new overarching objectives, new policy instruments to implement EU interventions. Taken all together, Luukkonen states that "*there is no single way of doing research evaluation in Europe*". However, there are certain trends or styles that clearly distinguish European evaluations from U.S. evaluations.

One difference between the European and the U.S. system can be seen in the more pronounced usage of the term "evaluation" in Europe as opposed to the terms "outcome measurement" and/or "performance review" which are more popular in the U.S. (Luukkonen 2002). It seems

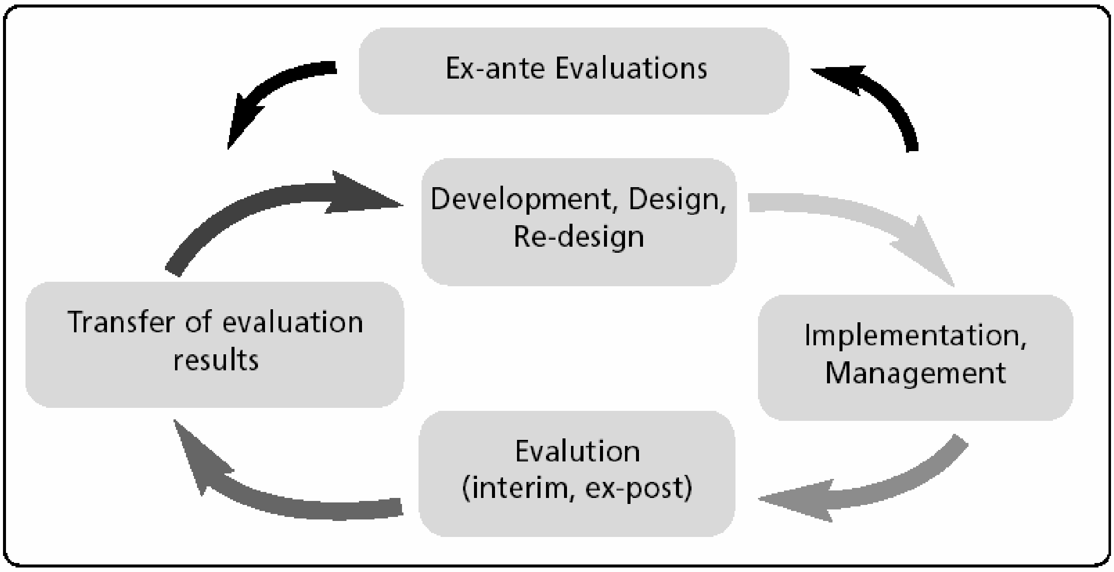
[2] However, similar discussion take also place in Europe with regard to the European evaluation system (see Mollas-Gallart/Davies 2006)

[3] The French evaluation system is in upheaval, as the legal framework is on reform.

that in the U.S. evaluation is used more as a synonym for outcome measurement, while the European understanding of evaluation implies a more neutral stance with respect to the methods employed (covering a wider variety of both qualitative and quantitative methods, whereby no preference is given to either of these) and the objectives sought for.

Another fundamental difference can be found in the fact that evaluations in Europe are (ideally) carried out in a cyclical pattern referred to as the "policy cycle". Fig 1 illustrates this concept by showing an ideal type of policy cycle for research and technology programmes. A programme is conceived and developed in order to address a pre-defined problem. The ex-ante evaluation assesses the suitability of the preliminary design with respect to this problem. Recommendations regarding any necessary modifications spelled out in the ex-ante evaluations are incorporated into the design of the programme, after which the programme is implemented - i.e. in general, given to an institution which usually operates support programmes. An interim evaluation makes a provisional appraisal of the programme: It looks at the status of the projects as well as the programme management and makes recommendations on how to continue. The evaluation results in one stage - those of the interim evaluation as well as those of the final evaluation after the running time of the programme - have consequences for further stages of programme development and implementation: whether the programme is to be continued or not, modified or not, expanded or downsized, etc.

Figure 1: Evaluation in the policy cycle



Source: Platform Research and Technology Evaluation 2003-2005

The diagram also shows that there are always several levels to evaluations, namely a strategic level responsible for development and design, a management level which carries out the evaluation, and the target group of the programme, the clients and the recipients. Clear dividing lines need to be drawn between the work which is done at these different levels, but the lines of communication between them need to be just as clear. Attention is drawn here to only one such line of communication: The results of evaluations must be brought to the attention of policy makers and should, as a consequence, contribute to learning and decision-making activities; in other words, they are to go beyond those directly responsible for the programmes. If this policy cycle is not completed, there is a risk of such things as rogue projects, unwanted continuity, or the lack of it, and acting against one's better knowledge (or what ought to be known)[4]. Again, it has to be stressed that Fig. 1[5] represents an ideal picture; deviations from it are, of course, not uncommon – we do not deny that there might be "pork programmes" in Europe.

The policy cycle model allows evaluations to take place at various stages of programme design and implementation, be it ex-ante or ex-post. This sets evaluation in the European context further aside from outcome measurement used in the U.S., which mainly addresses the ex-post effects of support programmes (Luukkonen 2002). The policy cycle also allows for the evaluation studies to be tailored to the specifics of the programmes, especially with respect to the methodology chosen. As a result, the programme can be analysed more thoroughly. On the other hand, the application of the policy cycle model makes comparisons between different programmes difficult because there is no set of standardised indicators against which the programmes can be assessed. It is thus for a policy maker considerably trickier to get an overview over the "performance" of all programmes (however one-sided that may be).

Some more recent trends in European evaluation practices address, among others, this issue:

- Resulting from the increased demand for evaluations, there is a growing number of (semi-) professional evaluators (Luukkonen 2002). The evaluation scene in Europe is (at least at its top) comprised of researchers who have a background in different science and technology policy fields and use the methods of these research areas in evaluations of related support programmes. The knowledge gathered is afterwards disseminated to the whole evaluation community. This has led, among others, to the creation of evaluation standards, such as the *Evaluation Standards in Research and Technology Policy* of Platform fteval. While standards as these do not go as far as developing standardised performance indicators, they provide some type of code of conduct for evaluators and a list of requirements a "state-of-the-art" evaluation has to fulfil.

[4] This reasoning reflects basically the criteria laid down in the Austrian Standards Platform fteval 2003-2005.

[5] This same diagram is often used in the US, but more often within the context of individual programs that choose to design and redesign themselves in response to evaluation, e.g. in the context of logic modelling.

- On the methodological side, a departure from the pure panel review technique can be observed as modified peer reviews and mixed panels are becoming more common (Rigby 2004; Luukkonen 2002). These panels may include, besides experts from the respective technology/research area (just like in a traditional panel review), experts from other fields (such as evaluation methods) but also representatives of stakeholder groups (i.e., people who have important interests in the research policy area). The aim of the mixed panels is, according to Luukkonen, *"…to increase credibility of the recommendations and the likelihood of their being accepted"*.

- There is also an increasing practise in national evaluation exercises to put evaluation studies out to tender (Lukkonen 2002). Together with the increased "professionalism" of the evaluation scene, this has led to competitions between groups of evaluation experts in order to obtain the evaluation contracts. Further to that, some researchers specialised almost entirely on evaluations, others organised themselves in consulting companies. A new trend within this trend is that some tenders require a transnational consortium to carry out the review. This has advantages (such as increased transparency), but also disadvantages (additional communication and travel costs, which, especially in the face of tight time schedules, may pose a considerable problem).

Critics of the European evaluation system point to the following issues:

- <u>Fragmented Patterns:</u> The methods employed, data availability and the organisations involved in evaluation exercises differ a lot across Europe, but also within the individual nations. This prompts concerns about the lack of comparability and the inability to aggregate individual evaluation results (Arnold 2005). In addition, evidence from the past suggests a rather fragmented pattern of evaluation, with some experimental elements and a strong reliance on approaches incapable of addressing policy questions on a higher level.

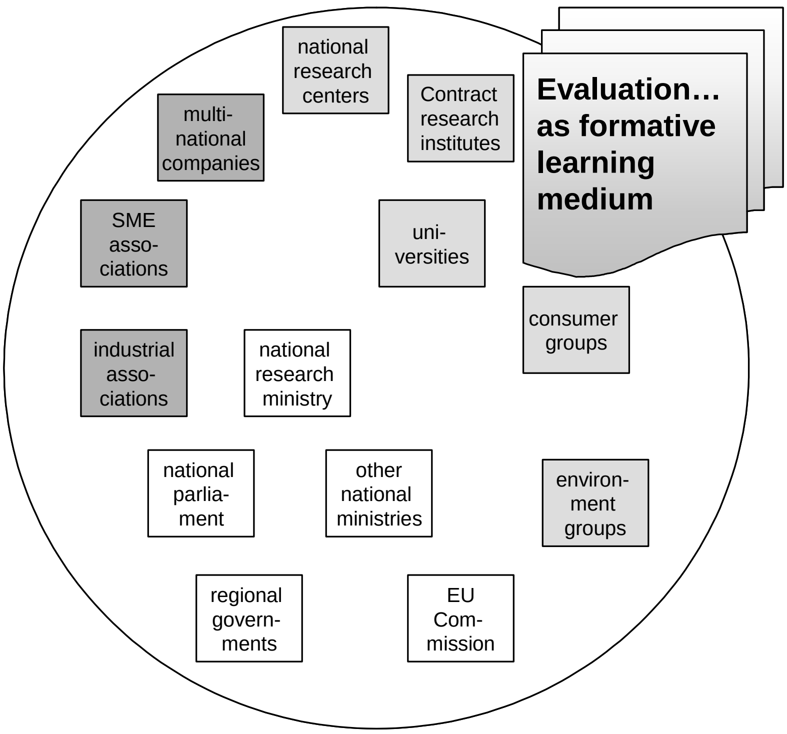
- <u>Evaluations lag innovation theory:</u> In a most recent article, Molas-Gallart and Davies (Molas-Gallart/Davies 2006) argue that the European practice of policy evaluation does not incorporate the latest advances in innovation theory. In the authors' point of view, the evaluation practice in the EU continues to favour the usage of methods which implicitly assume outdated linear views of the innovation processes.

- <u>Lack of long-term impact analysis:</u> Arnold further concludes that *"…as elsewhere, there is lots of mid-term and in-process evaluation [on the EU-level]. There is little evaluation done sufficiently far after the event [in order to be able] to understand [long-term] impacts in a clear way."* (Arnold 2005)

Of course, evaluations are not the only - and often not the most important - instruments of information and orientation. They can neither replace political responsibility nor the development of concepts; and most definitely not the processes of considering, jointly developing and discussing new instruments; they can only be seen as an important part of the whole process.

Kuhlmann et. al. have therefore developed a somewhat more elaborate concept of how ‘strategic policy intelligence’ can influence the policy cycle (see Fig. 2). They have introduced ideas like the “strategic policy arena” which should transform policymaking from a “mechanistic” to a “reflexive” process. Kuhlmann stresses the need for multi-perspective, “reflexive” evaluation and learning, “strategic intelligence” and defines a “research and innovation system and stakeholder arena”, where evaluation comes in as strategic intelligence, but does not replace policy decisions. Kuhlmann’s policy arena is a step forward: Strategic policy development should be more than pure performance measurement. It should be reflexive policy learning with different stakeholders.

Figure 2 Evaluation in the policy arena



Source: Kuhlmann, S/Edler, J. 2004

These trends in the evaluation community in Europe point, taken together, in two directions: First of all, the scope of evaluations becomes broader up to the point where they also tackle, at least partially, and embedded in the policy arena, policy debates. Second of all, through the setting of standards (the overall idea would go as far as to establish professional bodies that issue certificates for professional evaluators or provide quality control functions) it is attempted to increase transparency and credibility of the evaluations. Pursuing the first direction seems only possible if the road towards high professionalism and sensible quality standards is also – successfully (!) – taken. If not, and if strategic stands taken prevail over evidence-based judgements, evaluation may lose credibility.

2 PART in Action

2.1 Design and Application

There is consensus among European evaluators that evaluation studies ought to take many different types of evidence into account when carrying out their analysis, in order to highlight all different aspects and secure a certain quality standard. In this context, it can be interesting to see whether a tool such as PART can be used in the European evaluation landscape for the purpose of providing evidence on outputs and outcomes. It is also of interest to see if and how the tool may help agencies with the accounting of their activities and to what extent programme performances can become comparable.

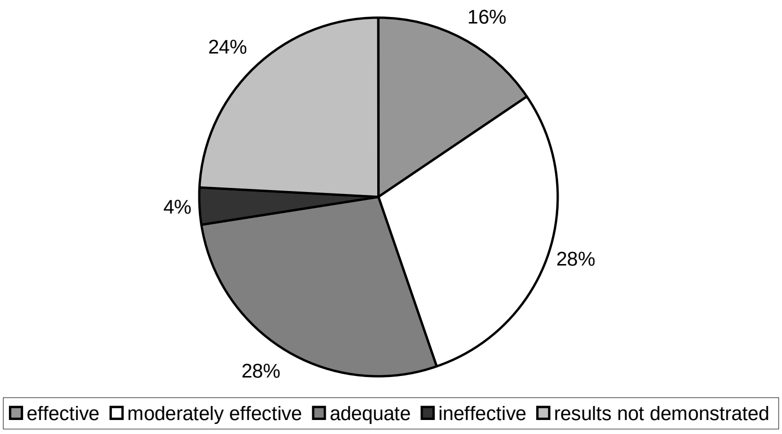
The Programme Assessment Rating Tool is basically a questionnaire. It addresses four critical areas of assessment: purpose and design of a programme, strategic planning, results and accountability (OMB 2004). Questions developed to rate purpose and design aim at analysing whether these two aspects are clearly defined and defensible. The second set of questions refers to the targets and goals of the programme and tries to capture whether the funding agency has set valid annual and long-term goals. Management issues are to be analysed with a third set of questions, whereby emphasis has been placed on financial oversight and programme improvement actions. The last set of questions gauges whether the programmes is demonstrating results with accuracy and consistency. For each question, there is a short answer and a detailed explanation with supporting evidence.

In total, there are around 30 questions to be answered, depending on the type of programme analysed. The questions are standardised (in a "yes" or "no" fashion), and the answers are (based on the evidence provided) graded in each of the four sections on a scale from 0 to 100, with 100 being the best score. The four scores are then combined together in order to rate the programme on a single four-tiered ordinal scale, ranging from "effective", "moderately effecttive", "adequate" to "ineffective". In case programmes do not have acceptable performance measures or have not yet collected performance data they receive the label "results not demonstrated". The full list of questions is given in Annex 1.

PART has been used for the first time in the year 2002 for the analysis of 234 federally funded programmes. The programmes chosen varied, among others, by type (such as regulatory, grants, or direct assistance programmes). About half of the reviewed programmes did not demonstrate results. 5 % were deemed ineffective, 15 % adequately effective and 24 % moderately effective. The highest grade of "effective" was given to only 6 % of the programmes in question. By 2006, 793 programmes were rated. The more recent results showed that 16 % of the programmes were effective, 28 % moderately effective and 28 % adequately effective. 4 % were considered ineffective and 24 % could not demonstrate results (see Graph 1).

The U.S. government plans to conduct PART assessments on about one fifth of all federal programmes (which amount to around 1,000) every year. Consequently, every federal programme should have been rated with PART by 2008.

Figure 3 PART Ratings of Federal U.S. Programmes, 2006



Source: Office of Management and Budget 2006

2.2 Benefits and Challenges of PART

The application of PART seems to have stirred a great deal of controversy among U.S. agencies, researchers and beneficiaries of support programmes. Both positive and negative aspects have been noted, though the number of issues that are of concern greatly outnumber the identified benefits. The identified benefits include (Ruegg 2004, p. 3.):

- The revitalisation of evaluation within an agency
- Self Assessment and subsequent planning for enhancements
- Development of new assessment tools and performance measures
- Attention on evaluation as a useful tool for management
- Interest in a standardized review of agency programmes

- Interest in aligning PART closely to agency programmes

On the negative side, the following shortcomings and possible areas for improvements have been described:

- *Development of adequate performance measures:* One of the key elements needed to have PART working in its intended way is the availability of clearly defined and operationally viable performance measures. In order to create such measures, a number of problems have to be tackled. Some measures are impractical in the sense that data collection is not possible on an annual basis. "Right" performance measures might not even exist for some programmes at all (OMB 2004).
- *Usage of a 100-point scale:* With the problem of finding good performance indicators in mind, it quickly becomes clear that using a 100-tiered scale might misleadingly indicate a very high level of accuracy. Despite the fact that OMB cautions not to use single numerical ratings it can still happen that managers invest significant time in trying to justify a difference of several points. On the other side, different examiners will give a different score even in the case that they would essentially agree on the level of performance observed. As a consequence, it has been suggested to use a 5-point or 10-point scale instead (Schurr 2003).
- *Subjectivity:* The fact that some level of subjectivity on the part of the examiners will be present (see point above) is recognised by the OMB (OMB 2004).
- *Overall context:* As PART addresses only individual programmes, redundancies among different programmes are not captured. Similarly, it is also difficult to assess whether some programmes complement each other in a positive way. The OMB itself calls for suggestions on how to improve PART in this respect (OMB 2004).
- *Programme improvements:* Because of being mainly results-oriented, PART can only hardly reflect programme improvements. Only after the improvements lead to observable results (usually with a delay of one to several years) is it that improvement actions are honoured (OMB 2004, Schurr 2003).
- *Administrative costs:* In order to provide a complete picture and in order to monitor programme improvements, it would be necessary to scrutinize the programmes using PART regularly. Ideally, an assessment should take place every year. Due to budgetary constraints this does not seem to be possible, however (OMB 2004, Schurr 2003).
- *Connection between GPRA and PART:* As it seems, GPRA and PART are used in parallel. This might lead to a redundancy, for example in terms of the number of performance plans to be developed (Schurr 2003). In fact, one report written up by the Government Accountability Office concluded that PART and GPRA were competing approaches and that *"by using the PART to review and sometimes replace GPRA goals and measures, OMB is substituting its judgement for a wide range of stakeholder interests."* (GAO 2005, cit. in OMB Watch 2005a*)*

- *Comparability:* The overall question remains whether programmes from completely different areas are really comparable using PART scores and results. Not only can it be that PART might not capture the specific thematic context of a particular programme, it could already be misused in the actual scoring process for political reasons (see, for example, OMB Watch 2005b).

These are only some of the issues described in the literature. At a WREN (Washington Research Evaluation Network) workshop in 2003, even more were discussed. The table below gives an overview over the problem areas encountered, whereby a distinction has been made between structure issues and process issues.

Table 1 Identified challenges of PART

Number	Process Issue (PI)	Structure Issue (SI)	Description of Challenge
1	PI		OMB lacks a clear definition of "programme"
2	PI		OMB's roll-up of multiple agency programmes into a single programme for PART assessment results in meaningless results
3	PI		Examiners often lack knowledge of evaluation, leading to failure to use data/evidence provided as intended
4	PI		Examiners are inconsistent in applying PART
5	PI		Multiple requirements for centralised reporting are confusing (GPRA, PART, etc.)
6	PI		The use of PART as a political tool devalues its use as a fair assessment tool
7	PI		The link is unclear between a programme's PART score and its budget success in face of budgetary requirements
8		SI	Requiring binary (yes/no) choices is too restrictive
9		SI	Using a rating system based on 100 points implies a level of accuracy that is unwarranted
10	PI		Low scores lead the public to believe programmes are mismanaged when mark-down may actually reflect something beyond management's control
11	PI		Scores have uncertain meaning
12	PI		Programme administrators may lack funding to develop the data/evidence required for PART
13		SI	PART's emphasis on annual measures may not fit programmes (like forestry or basic science research) whose yield does not map to an annual cycle
14	PI		Better models are needed for evaluating research programmes and providing data/evidence under PART

15		SI	PART questions on spending efficiency are by nature difficult for basic research
16	PI		Application of PART is expected to be sporadic and not meaningfully coordinated with programme improvements
17	PI		PART is implemented w/o regard to agency internal reorganisation issues
18	PI		Norms are lacking for comparisons of programmes of differing size and type
19	PI		Performance indicators are used without consideration of context; and OMB examiners and agency staff hold conflicting views regarding appropriate performance indicators
20		SI	PART assumes a direct, linear relationship between research and outcomes which is not necessarily accurate, and, in any case, is difficult to show
21	PI		Both OMB and agencies lack analysis support
22	PI		OMB appeals process is unclear
23	PI		All the evidential burden is on the agency; none on the examiners
24	PI		Congress appears to have little or no interest in PART

Source: Ruegg 2004. p. 4f.

3 General Remarks on the Design of PART from a European (Austrian) Perspective

To further fuel the discussion about PART usage in the European context, we took a look at the Austrian evaluation system.

In the second half of the Nineties, Austrian research and technology policy has brought into being as many as fifty, maybe more, initiatives, programmes, bodies, funding campaigns, etc., at federal as well as at the regional level. This goes hand in hand with a general rise of public attention on R&D Policy and the willingness to spend more money for such initiatives, not only in Austria, but all over Europe. While some of these programmes reflect, at least to a certain extent, Kuhlmann's and the Plattform's thoughts, others don't. The Austrian protec programme[6] to foster technology transfer activities can be considered a positive example. Austrian Evaluators (Technopolis Austria and Joanneum) evaluated the predecessor of protec, the ITF Techtransfer programme, in 2000. Its results were used for the redesign of the programme. A monitoring effort was subsequently commissioned by the ministry. Now, 5 years later, a new evaluation exercise was carried out by another Austrian evaluator, the Austrian Institute for SME Research. Another best practice example is the competence center assessment (Edler et al

[6] www.bmwa.gv.at/protec

2004) which was explicitly used in the policy making process. Yet another one would be the programme FH Plus. There is a long list of programmes, or more generally spoken, of initiatives (e.g. GEN-AU, the Austrian genome research programme or the evaluation of the FIT IT programme), where evaluation reports were used during a policy cycle as described above for redesign and fine-tuning purposes (Zinöcker et al. 2005a, 2005b). The evaluation of the Austrian Science Fund FWF and the Austrian Industrial Research Promotion Fund FFF (Arnold 2004), an endeavour to evaluate the Austrian funding system, were carried out in reflective loop processes. In these evaluation exercises, the evaluators provided the stakeholders with a broad range of suggestions for improvement; a majority of which were then also adopted by the funds. In this context, Pichler describes the evaluation process as an *"intense integration of the evaluators' expertise into the political process as they [the Evaluators] became involved in substantial discussion with reform stake holders".* (Pichler 2005)

One could continue this list and add a considerable number of other mentionable examples, not only in Austria, but also in other European countries. But it would be euphemistic to claim that this is the status quo. The development of policy interventions in the field of R&D can be often also described as a black box: Why programmes are launched, priorities are set or budgets cut down remains hidden behind non-transparent and rarely traceable decisions, and, at least at first sight, the role of evaluations in this process is often ambiguous. An example that turned out positive in the end (again the FFF/FWF evaluation) in this context describes the nature of the ambiguity: *"The reform [a merger of three different federal agencies] has come after a long lasting and intense debate about "bottom up" vs. programme funding and coupling vs. uncoupling policy design and funding mechanisms. It was triggered by some sort of Austrian "parallel action": On the one hand a policy debate – reaching back to the late 1990s – culminated finally in a law-making process in 2003 / 2004. On the other hand, the Federal Government commissioned a big international evaluation of the Funds. The evaluators had to hurry to finalise their work before the new law came into being. A benevolent Hegelian 'Weltgeist' provided, nevertheless, a mainstreaming of ideas in both strands of activities, so in the end a new era of the Austrian RTDI policy could begin, with an integrated agency, more strategic capacities, new governance principles for the still autonomous FWF and a few tasks for the ministries still to install."* (Stampfer 2005)

A negative example described by Grießler (Grießler 2003) is the disillusioning story of how a policy initiative was put on the chopping block in Austria in the early 1990s: ATMOS, the Austrian Technology Monitoring System was launched in spring 1990. After six month of testing, the initiative was evaluated by an international expert team. Before this study was concluded, the programme had already been abandoned: Cloudy competence distributions that led to conflicting decision making at the policy level, old boys networks, and an unclear communication structure resulted in the premature termination of a R&D programme. This setup depicts policymaking as a game rather than an evidence based policy arena. Again: There is no doubt that the positive (but also the negative) list could be longer and enriched by examples from other countries in Europe.

Could PART help increase transparency?

At first sight, GPRA and PART are interesting starting points to structure the decision making processes within a framework as it is described here: Programme managers start a structured discussion with policy makers in order to come to a decision regarding the budget and/or to a stop-or-go decision for the next periods. This process has a clear advantage as it is mandatory and at the same time seems transparent. In this context, PART is in particular worth a second look: It claims to make different policy strategies (programmes) comparable and could be an important tool for policy making in the fields of science and technology.

However, the enormous (and, for the most part, sensible) number of "challenges" makes us hesitant and lets us ask the question whether the introduction of PART would not create more problems than it would help solve. In 2003, we had the possibility to attend a WREN workshop "Meeting the OMB PART Challenge." Among the challenges that were described (see also above) were issues that would apply only to the U.S. (such as the relationship between GPRA and PART), but others would clearly be significant problem areas in Europe, too.

Among these were:

- The perceived rigidity of the system: PART gives the impression to have been implemented top-down and not in a reflexive process as suggested and pursued in Europe. Some ideas on how to deal with the "challenges" were perplexing and, probably as a result of the top-down approach, very down to Earth: One idea to meet better OMB's expectations was to *"use Microsoft Word tools to highlight the central passages in the paperwork, so that one can deal better with the provided text";* or, generally spoken, to use OMB's attention for a programme quite economically. Is the obvious lack of time in making decisions an advantage?
- Administrative load and opportunity costs: Given the amount of work that goes into performing a PART analysis, we would have to ask ourselves whether the PART system is actually compatible with the policy cycle/policy arena model used in Europe. Is there still a meaningful place for detailed evaluation studies? If yes, for whom? What is foregone if PART is the only tool being applied, due to limited resources? Especially, if evaluations are needed as evidence for PART and thus both have to be performed?
- "Safety by numbers?": Until now, researchers in Europe (as well as policy makers) seem to have been hesitant to apply rather simple 100-point scales (or any other scales for that matter, despite of their appeal) for the purpose of analysing and judging complicated settings. Evaluators have been discussing intensely how long it takes for R&D programmes to realize economic and social returns, how tricky it is to trace impacts and, eventually, cater for accountability; on the other hand PART/OMB includes questions like "did the programme maximize net benefits?" which are to be asked, at least in the ideal case, on a yearly basis. This approach would certainly not reflect the discussion process in Europe.

- ”One Size fits all?”: Anecdotal evidence suggests that a high share of evaluators and, especially, evaluated bodies (research institutions, etc.) do not feel comfortable with a standardised “one-size-fits-all” approach for reviewing performance. To this end, there have been frequently outcries with regard to context variables which have not been (supposedly and/or exhaustingly) taken into consideration. This seems to be very similar to the situation in the U.S., at least with regard to the evaluated bodies (see, for example, OMB Watch 2005b). However, while it looks like policy makers in Europe take a more neutral stance on this issue, they do seem at the same time also more hesitant to weigh, say, defence programmes with mobility programmes for young researchers directly against each other. The compulsion on the policy side to implement instruments such as PART is evidently in Europe much less pronounced than in the U.S.
- Correlation between Budgeting Decisions and PART performance: One of the goals of PART is to put budgeting decision on a firmer level. However, there is evidence that PART ratings do not correlate entirely with budgeting decisions - i.e., programmes with low PART scores are continued/expanded, while others with high scores are cancelled/reduced (OMB Watch 2005a).[7] This can be seen positively (the administration makes use also of other information/data available) or negatively (while one of the most positive aspects of PART should be the transparency aspect, it does not, in reality, contribute to transparency at all; PART cannot be used for budgeting decisions; PART is ignored by decision makers; etc.). Once again, the costs of PART have to be weighed against its benefits, according to U.S. researchers.

Finally, systems like PART cannot replace evaluations. Evaluations fulfil various functions: a legitimating function (e.g., justifying the use of public funds); an information function (providing the public with information on how public funds are being used and to what effect); an information-oriented learning function for those funding and/or implementing programmes (decision-makers in the field of technology policy, scientists, etc.); a steering function for establishing policy objectives and/or planning measures for the future; and, eventually, a controlling function, as in private enterprises. PART alone seems to be able to fulfil these functions only in a very limited way, as implied by the “challenges” described above. If, however, PART is used in conjunction with evaluations (the latter as detailed evidence in section 4 of the PART questionnaire), PART may fulfil the described functions.

The shortcomings described make it difficult to believe that implementing PART 1:1 in, say, Austria would significantly boost accountability and performance of publicly funded support programmes, however interesting the idea might be. On the other hand, by implementing subsets of PART and integrating them wisely, as complementary tools, it might indeed contribute to higher transparency. Corresponding suggestions and ideas are presented below in the conclusions.

[7] A similar argument holds also for GPRA (Feller 2002)

4 Conclusion – Compatibility with other Evaluation Methods and Applicability in Europe

There are aspects of PART that make it at least partially appealing for application in the European context. First of all, it has to be acknowledged that there is a legitimate desire on the part of policy makers to get a structured overview on support programmes offered and operated. The number of programmes has multiplied over the past two decades, and so has also the diversity of support instruments employed. Yet, at certain points, qualified decisions have to be made on what programmes to axe and what programmes to continue or fund even more (based also on policy considerations). While it is important and desirable to look at each programme individually, a minimum degree of comparability seems to be helpful if policy wants to stay in control over the programme. That desire seems similar to that of analysts in the business sector who want to obtain, at least for a quick review, key data and figures on businesses before they reach investment decisions.

In principle, elements of the current PART can fulfil such "the programme at a glance" features:

- PART emphasizes the fact that all support programmes have features in common: All programmes have goals that are to be stated and can be broken down into intermediate goals that are to be achieved. All programmes have outputs and outcomes that have to be described or defined, however different they may be. By using PART, awareness is created with respect to presenting operation and key features of a programme in a very clear way. In fact, it has been suggested in the U.S. to use logic models as one of the means to meet the PART challenges (Ruegg 2004) – a tool widely used in evaluations exactly for the purpose of displaying the operation of programme with a single chart.
- While PART seems to have serious issues concerning the measurements of outcome with its "one size fits all" approach, it seems to have a clear connection to the management of the programme and thus to its efficiency. Again, management issues are issues that are very similar across different programmes.
- Finally, PART can be considered a systematic and transparent approach to analysing a R&D programme, because of its standardized nature.[8] It should be also noted that PART takes many aspects of a programme design into account and, in principal, seems to pose the right questions for many cases. A significant amount of difficulties seem to appear when the answers to this questions are valued by personnel who are not trained in R&D (or S&T) policy and/or evaluation methods and who are not familiar with the context of the programme.

A scaled down version of PART could be, in our opinion, an interesting additional tool and add value to the European evaluation system, if its usage would meet certain provisions:

[8] It has to be noted though that the evidence itself is not standardised which stresses the importance of training measures for OMB/PART examiners. This would in our opinion favour a decentralisation approach where external evaluation experts would work with PART (see also further down the text).

- Inclusion of Context Studies: Notwithstanding a limited number of customized questions in the PART questionnaire, the „One size fits all"-approach seems to have, as described, its limits when comparing programmes from completely different areas. As in business, where analysts do not only look at the balance sheet of companies alone but also regularly scrutinize the industry the companies are operating in as a whole, we believe that there is a strong need for context studies when applying PART (i.e., studies that look into certain programme types or areas programmes are operating in as whole). This would help policy makers in putting quantitative output/outcome data into perspective when looking at different programmes and might help to isolate efficiency issues from output/outcome issues. Such context studies could make use of a variety of methods (e.g., case studies). Benchmarking studies, as they are carried out by and for the European Commission, could be a starting point but have to be modified to take account of potential PART measures. The same applies to meta evaluations.

- Notwithstanding context studies, we still claim that a tool such as PART will never be able to live up to the variety of support programmes present in terms of the problems they address, the areas they are operating in (e.g., research, defence, labour market, etc.), the way they actually operate, the different instruments they employ and, consequently, the different outcome measures that have to be derived for each individual programme in order to allow for meaningful analysis. Thus, it should not replace programme evaluations.[9] What would be intriguing, however, is the thought of PART being a mandatory element of programme evaluations (i.e., something like a standardised summary fact sheet or "balance sheet" to be used as front/cover page for evaluation reports). This would imply a merger of programme evaluations with PART in the sense that PART becomes a cover for evaluation reports, for example on outcomes. The integration would have several advantages: It would decentralize PART work and put it into the hands of professional evaluators, familiar with the context of the programme, rather than have an overworked staff of a central authority dig into a multitude of programmes. By putting the evaluations out to tender and thus inducing competition, costs may be reduced even further (of course, this could also have negative effects if the contract were to be awarded only to the lowest bidders). A central board could be established to define and monitor the quality standards of the evaluations and the PART requirements (i.e., fulfil the function of evaluating the evaluators). This authority could also collect the completed PARTs of the evaluations and compile them into a central registry policy makers can revert to. Using PART in such a way would also integrate it completely into the policy cycle model. The tasks and authority power assigned to such a central body would have to be thoroughly discussed and carefully chosen, however. This decentralized approach may also reduce the imposition of political views over objective evaluation results.

[9] Interestingly, OMB itself made this point, too (see OMB 2004).

Trying to resume

Yes, the "European system", presented here as a "strategic policy arena" or as an evaluation policy circle has serious weaknesses: It highly depends on the devotion of policy makers to strategic policy intelligence and their willingness to accept less degrees of freedom in their decisions. On the other hand, it provides the policy maker at least with the possibility of a reflexive process. In this context we do not see the superiority of the PART/OMB system, as long it is hastily implemented, concentrating only on pure performance indicators and trying to compare apples and oranges. If it were to be used as a complementary tool to allow policy makers to get a better overview, taking into account the provisions laid out above, it might indeed prove to be a valuable addition to the evaluation system and might also constitute a prerequisite for system evaluations. The main question that remains, and which is rather independent of the evaluation system in use, is whether policy will actually consider evaluation results in its decision making.

Literature

Arnold, E. et al (2004): Evaluation of the Austrian Industrial Research Promotion Fund (FFF) and the Austrian Science Fund (FWF). Synthesis Report, Brighton – Wien

Arnold, E. (2005): What the Evaluation Record tells us about Framework Programme Performance.

European Commission (2004): Performance Assessment of Public Research, Technology and Development Programmes – Workshop Report of workshop held in Brussels, June 17 to June 18

Edler, J./Bührer, S./Lo, V./Sheikh, S. (2004): Assessment „Zukunft der Kompetenzzentrenprogramme (K plus und K ind/net) und Zukunft der Kompetenzzentren“

Fahrenkrog, G./Polt, W./Rojo, J./Tübke, K./Zinöcker, K. (eds.) (2002): EPUB RTD Evaluation Toolbox: Assessing the Socio-Economic Impact of RTD Policies, STRATA Project HPV 1 CT 1999-00005, IPTS Technical Report Series, EUR 20382 EN, Seville: IPTS

Feller, I. (2002): The good, the indifferent, and the less attractive: II, in: Research Evaluation, August 2002

GAO, Performance Budgeting; PART Focues Attention on Program Performance, but More Can be Done to Engage Congress, GAO-06-28, Oct. 2005.

Georghiou, L. / Rigby J. / Cameron H (eds), Assessing the Socio-economic Impacts of the Framework Programme (ASIF), University of Manchester: 2002

Grießler, E. (2003): Innovation und Politikgestaltung: Administrative Kulturen in der Technologiepolitik. Ein Vergleich zwischen Österreich und den Niederlanden, *in:* Pichler, R. (Hg.) Innovationsmuster in der österreichischen Wirtschaftsgeschichte

Hicks, D. et. al. (2002): Quantitative Methods of Research Evaluation Used by the U.S. Federal Government, NISTEP Study Material, No. 86, May 2002

Kuhlmann, S./Edler, J. (2004): Tailor-made evaluation concepts for innovation policy learning. Presentation held at the Research and the Knowledge Based Society – Measuring the Link conference, 24th May 2004, NUI Galway, Ireland

Laredo, P. (2005): Research Evaluation for Policy Making : American, European and World-wide Perspectives. Meeting of the American Association for the Advancement of Science, Feb 18th 2005

Luukkonen, T. (2002): Outcomes and impacts – Research evaluation in Europe: state of the art, in: Research Evaluation, August 2002, p. 81–84.

Mercer, J. (2006): GPRA and Performance Management – Homepage, http://www.john-mercer.com/

Molas-Gallart, J. / Davies A. (2006): Toward Theory-Led Evaluation The Experience of European Science, Technology, and Innovation Policy. American Journal of Evaluation, Vol. 27, No. 1, 64-82

Moore, J.L. (1995): Cost-Benefit Analysis: Issues in Its use in Regulation, CRS Report for Congress

OMB (2003*):* Budget FY 2004.

OMB (2006): FAQ on expectmore.gov, http://www.whitehouse.gov/omb/expectmore/faq.html#016, requested on March, 21 2006

OMB Watch (2005a): How Do You Measure Programme Results?, published on Feb 22, 2005, http://www.ombwatch.org/article/articleview/2680/1/2

OMB Watch (2005b): House Considers CDBG But Avoids Attacking PART, published on 05/31/2005, http://www.ombwatch.org/article/articleview/2865/1/85

Pichler, R. (2005): Preface. In Plattform Newsletter 25: How to evaluate funding systems. The example of the FFF/ FWF evaluation

Plattform Research and Technology Policy Evaluation (2003-2005): Evaluation Standards in Research and Technology Policy, *www.fteval.at*

Riggle, D. (2005): A Role for Program Assessment Rating Tool in the Future Federal Budget Crisis, SRA International eNews, July 2005

Roessner, D.J. (2002): Programme perfomance – Outcome measurement in the USA: state of the art, in: Research Evaluation, August 2002, p. 85–93.

Ruegg, R. (2004): WREN Workshop Executive Summary Report – Planning for Performance and Evaluating Results of Public R&D Programmes: Meeting the OMB PART Challenge

Ruegg, R. / Feller, I. (2003): A Toolkit for Evaluating Public R&D Investments: Models, Methods, and Findings from ATP's First Decade. NIST GCR 03-857

Stampfer, M. (2005): How These Things Came About: A Short Note on the Early Years of FFF and FWF. In Plattform Newsletter 25: How to evaluate funding systems. The example of the FFF/ FWF evaluation

Valdez, B. (2005): Evaluation Research Policy Development, DOE – Office of Science

Wessner, C. (2000): The Small Business Innovation Research Program SBIR, An Assessment of the Department of Defense Fast Track Initiative, National Academy Press, Washington D.C.

Zinöcker, K. / Schindler, J. / Gude, S. / Schmidmayer J. / Stampfer M. (*2005a):* Interims-evaluierung FIT – IT. Konzepte, Rahmenbedingungen, Design, Prozesse. Snapshots auf Wirkung und Additionalität. Studie im Auftrag des bmvit. Wien

Zinöcker, K. /Tempelmaier, B./Radauer, A. /Fischl, I./Steiner, R. /Ruegg, R. (2005b): Austrian Genome Research Programme GEN-AU: Mid Term Programme Management Evaluation, Final Report

Appendix

PART Questions

I. Programme Purpose & Design (Weighted 20%)

1.1. Is the programme purpose clear?

1.2. Does the programme address a specific and existing problem, interest or need?

1.3. Is the programme designed so that it is not redundant or duplicative of any other Federal, state, local or private effort?

1.4. Is the programme design free of major flaws that would limit the programme's effectiveness or efficiency?

1.5. Is the programme effectively targeted, so that resources will reach intended beneficiaries and/or otherwise address the programme's purpose directly?

II. Strategic Planning (Weighted 10 %)

2.1. Does the programme have a limited number of specific long-term performance measures that focus on outcomes and meaningfully reflect the purpose of the programme?

2.2. Does the programme have ambitious targets and timeframes for its long-term measures?

2.3. Does the programme have a limited number of specific annual performance measures that can demonstrate progress toward achieving the programme's long-term goals?

2.4. Does the programme have baselines and ambitious targets for its annual measures?

2.5. Do all partners (including grantees, sub-grantees, contractors, cost-sharing partners, and other government partners) commit to and work toward the annual and/or long-term goals of the programme?

2.6. Are independent evaluations of sufficient scope and quality conducted on a regular basis or as needed to support programme improvements and evaluate effectiveness and relevance to the problem, interest, or need?

2.7. Are budget requests explicitly tied to accomplishment of the annual and long-term performance goals, and are the resource needs presented in a complete and transparent manner in the programme's budget?

2.8. Has the programme taken meaningful steps to correct its strategic planning deficiencies?

Specific Strategic Planning Questions by Programme Type

Regulatory Based Programmes

2.RG1. Are all regulations issued by the programme/agency necessary to meet the stated goals of the programme, and do all regulations clearly indicate how the rules contribute to achievement of the goals?

Capital Assets and Service Acquisition Programmes

2.CA1. Has the agency/programme conducted a recent, meaningful, credible analysis of alternatives that includes trade-offs between cost, schedule, risk, and performance goals and used the results to guide the resulting activity?

Research and Development Programmes

R&D programmes addressing technology development or the construction or operation of a facility should answer the Capital Assets and Service Acquisition question (2.CA1).

2.RD1. If applicable, does the programme assess and compare the potential benefits of efforts within the programme and (if relevant) to other efforts in other programmes that have similar goals?

2.RD2. Does the programme use a prioritization process to guide budget requests and funding decisions?

III. Programme Management (Weighted 20 %)

3.1. Does the agency regularly collect timely and credible performance information, including information from key programme partners, and use it to manage the programme and improve performance?

3.2. Are Federal managers and programme partners (including grantees, sub-grantees, contractors, cost-sharing partners, and other government partners) held accountable for cost, schedule and performance results?

3.3. Are funds (Federal and partners') obligated in a timely manner and spent for the intended purpose?

3.4. Does the programme have procedures (e.g. competitive sourcing/cost comparisons, IT improvements, appropriate incentives) to measure and achieve efficiencies and cost effectiveness in programme execution?

3.5. Does the programme collaborate and coordinate effectively with related programmes?

3.6. Does the programme use strong financial management practices?

3.7. Has the programme taken meaningful steps to address its management deficiencies?

Specific Programme Management Questions by Programme Type

Competitive Grant Programmes

3.CO1. Are grants awarded based on a clear competitive process that includes a qualified assessment of merit?

3.CO2. Does the programme have oversight practices that provide sufficient knowledge of grantee activities?

3.CO3. Does the programme collect grantee performance data on an annual basis and make it available to the public in a transparent and meaningful manner?

Block/Formula Grant Programme

3.BF1. Does the programme have oversight practices that provide sufficient knowledge of grantee activities?

3.BF2. Does the programme collect grantee performance data on an annual basis and make it available to the public in a transparent and meaningful manner?

Regulatory Based Programmes

3.RG1. Did the programme seek and take into account the views of all affected parties (e.g., consumers; large and small businesses; State, local and tribal governments; beneficiaries; and the general public) when developing significant regulations?

3.RG2. Did the programme prepare adequate regulatory impact analyses if required by Executive Order 12866, regulatory flexibility analyses if required by the Regulatory Flexibility Act and SBREFA, and cost-benefit analyses if required under the Unfunded Mandates Reform Act; and did those analyses comply with OMB regulations?

3.RG3. Does the programme systematically review its current regulations to ensure consistency among all regulations in accomplishing programme goals?

3.RG4. Are the regulations designed to achieve programme goals, to the extent practicable, by maximizing the net benefits of its regulatory activity?

Capital Assets and Service Acquisition Programmes

3.CA1. Is the programme managed by maintaining clearly defined deliverables, capability/ performance characteristics, and appropriate, credible cost and schedule goals?

Credit Programmes

3.CR1. Is the programme managed on an ongoing basis to assure credit quality remains sound, collections and disbursements are timely, and reporting requirements are fulfilled?

3.CR2. Do the programme's credit models adequately provide reliable, consistent, accurate and transparent estimates of costs and the risk to the Government?

Research and Development Programmes

R&D programmes addressing technology development or the construction of a facility should answer the Capital Assets and Service Acquisition question (3.CA1).

R&D programmes that use competitive grants should answer the Competitive Grants questions (3.CO1, CO2, CO3).

3.RD1. For R&D programmes other than competitive grants programmes, does the programme allocate funds and use management processes that maintain programme quality?

IV. Programme Results/Accountability (Weighted 50%)
4.1. Has the programme demonstrated adequate progress in achieving its long-term performance goals?
4.2. Does the programme (including programme partners) achieve its annual performance goals
4.3. Does the programme demonstrate improved efficiencies or cost effectiveness in achieving programme goals each year?
4.4. Does the performance of this programme compare favourably to other programmes, including government, private, etc., with similar purpose and goals?
4.5. Do independent evaluations of sufficient scope and quality indicate that the programme is effective and achieving results?
Specific Results Questions by Programme Type
Regulatory Based Programmes
4.RG1. Were programme goals (and benefits) achieved at the least incremental societal cost and did the programme maximize net benefits?
Capital Assets and Service Acquisition Programmes
4.CA1. Were programme goals achieved within budgeted costs and established schedules?
Research and Development Programmes
R&D programmes addressing technology development or the construction or operation of a facility should answer the Capital Assets and Service Acquisition question (4.CA1).

Source: Mercer (2005).

Authors:

Alfred Radauer
Austrian Institute for SME Research
Gusshausstrasse 8, 1040 Vienna
a.radauer@kmuforschung.ac.at

Klaus Zinöcker
Platform Research and Technology Policy Evaluation www.fteval.at &
WWTF Vienna Science and Technology Fund
Währingerstraße 3 /15a, 1090 Vienna
klaus.zinoecker@wwtf.at

Klaus Zinöcker, Michaela Glanz,
Brigitte Tempelmaier, Michael Dinges

Five Myths about Funding Scientific Research (in Austria)

"A myth is a narrative
which discloses a sacred world"
(Lawrence J. Hatab in: Myth and Philosophy (1990))

Myths have to be understood as simplified explanations for rather complex issues and relate to generally accepted beliefs that are unsubstantiated by fact. They are strongly related to expectations formulated by one's environment, which often draw on myths themselves. Myths thus function as stabilizing elements for the society and its actors. Institutions, for example, anticipate in their structures and their behaviour the expectations (myths) formulated in their relevant environments. Thus, they create certain kinds of expectations towards others, which in turn…

Why are we dealing with myths? Being part of the 'Austrian RTI (that is, Research, Technology & Innovation) system' during the past few years, the authors have been confronted with strong assumptions, persistent pictures and beliefs about "how things are going" in the field of science and technology policy. We have learned about these assumptions and beliefs in the context of numerous interviews, but also when simply chatting with scientists, policy makers or agency people.

Eventually, we started to become curious: is it actually true that the Humanities are being starved of resources? Is it true that networks are fretworks? Should the "funding gap" perception still be the guiding principle for designing new policy interventions in Austria? Should there really be more emphasis on engineering when it comes to the funding of basic research? And finally, what can we realistically expect from the intended impacts of an intervention in RTI policy? Can policy makers and programme managers enhance Austria's competitiveness' with every policy intervention that they plan? Can we support these myths or can we 'deconstruct' them?

This paper has a clear mission: it should contribute to a more structured and evidence based discourse about issues that clearly influence policy design in the field of RTI policy. Moreover, we will try to show how evaluation can play a role in this process.

Myth 1: “The Social Sciences and Humanities (in Austria) are being starved“

“In times of scarce resources, isn’t it obvious, that the Humanities are getting a raw deal?” (A. Freitag in: Der Standard, 31 May 2005)

“Research and teaching (in the Humanities) are being starved” (Austrian Green Party, Press release, 13 May 2005)

“There is a special precaution seen for the Humanities being in threat of marginalisation” (M. Nießen, DFG)

Do the Social Sciences and Humanities not receive enough attention in Austria? Are they provided with too few funds (for research)? Statements like the ones above are nothing new to those who are engaged in research policy, research funding or research activities themselves. However, the underlying problem seems to have received increasing attention over the last few years, not at least due to the more diversified funding landscape for (academic) research and teaching - diversified in terms of the actors involved and the funding criteria applied - that bring along a call for increased accountability.

An important question is: how do the Social Sciences and Humanities (SSH) “sell” themselves? The answer seems to be quite clear: not as good as they would like to. If we take the societal debate about SSH as a starting point, we are confronted with narratives about the poor (academic) SSH that receive too little money and have too little academic staff, which in turn must teach too many students and therefore doesn’t have enough time for conducting (proper) research. And if there was enough time for it there would not be sufficient funding since the funding priorities are set elsewhere. This is one story. Another one tells us about the desperately fragmented SSH, where researchers avoid both cooperation and competition by clinging to their research niches while arguing heatedly that there are no reasonable chances to get a research project funded as nobody is interested in what they are doing.

However, the basic myth behind these two – out of a few more – stories seems to be quite prominent: “The SSH (in Austria) are being starved.” Now the question for the authors was, whether the empirical evidence we could find would support this myth or deconstruct it. Being far away from concluding which of the stories told is the “true” one, we just want to illustrate whether or not the SSH are indeed facing an overwhelming teaching load, if they have little research active personnel, if their R&D (Research & Development) expenditures are scarce and if they are being discriminated against in research funding in particular.

But before we get into the data, we would like to point out the following: In Austria neither the SSH as a whole, nor one of its disciplines / subfields have been subject to evaluation up to now. Although there have been several efforts to map the Austrian SSH landscape (initiated, for example, by the Austrian Council for Research and Technology Development) and its research activities, the investigations mainly focus on specific sub-fields so that there is still lack of comprehensive data. Our analysis therefore relies on diverse publications and sources of data that do not exclusively focus on the SSH, but provide hints for the framework conditions and research activities within the SSH. Still, we have the problem that we hardly know anything about the non-university SSH research scene.

Some structural features about SSH in Austria

The Austrian Research and Technology Report (bm:bwk et al. 2005), which, for the first time, differentiates between disciplinary fields in its statistical annex, teaches us that the SSH landscape is quite fragmented indeed, and that it includes a high number of research actors: for 2002 the authors have identified more than 600 SSH research units in Austria; most of them (namely 411) belong to the higher education sector, about one quarter (namely 160) belongs to the governmental sector and a few units are public non-profit institutions (namely 37). Differentiating between the Social Sciences and the Humanities, the number of higher education research units is almost equal, while in the other two sectors Social Science units are more frequent (see Table 1).

Table 1: Structural features of the Social Sciences and Humanities (2002)

	number of research units	research active personnel		RTI expenditures in 1000 EUR
		headcount	FTE	
Higher Education	411	6,757	4,993.8	301,813
>> Social Sciences	208	3,775	2,718.4	165,755
>> Humanities	203	2,982	2,275.4	136,058
Governmental	160	2,239	1,750.1	100,691
>> Social Sciences	92	913	658.3	37,561
>> Humanities	68	1,326	1,091.8	63,130
Public non-profit	37	316	201.1	6,463
>> Social Sciences	28	292	189.5	6,155
>> Humanities	9	24	11.7	308

Source: bm:bwk et al. (2005)

The image of such fragmented SSH is reinforced when compared to other disciplinary fields within the higher education sector, which is most important for the SSH, and especially when looking at the size of the research units measured in FTE (full-time equivalents). With an average of 12.2 FTE per research unit (13.1 for the Social Sciences and 11.2. for the Humanities) the SSH have the smallest ones. Leaving Medicine aside, since this figure also includes clinics, the average size in the other disciplinary fields ranges from 15.6 to 24.7 FTE per research unit (see Table 2). This may also hint at the assumption that the SSH landscape is quite fragmented as regards sub-fields and content of research.

Table 2: Higher Education Sector (2002)

	Number of research units	research active personnel			RTI expenditures in 1000 EUR
		Headcount	FTE	FTE per unit	
Natural Sciences	197	6,469	4,865.2	24.7	387,193
Technical Sciences	173	3,502	2,690.6	15.6	173,493
Medicine (incl. clinics)	144	7,284	6,025.6	41.8	333,516
Agriculture, Forestry & Veterinary Medicine	44	1,060	847.5	19.3	70,089
Social Sciences & Humanities	411	6,757	4,993.8	12.2	301,813
>> Social Sciences	208	3,775	2,718.4	13.1	165,755
>> Humanities	203	2,982	2,275.4	11.2	136,058

Source: bm:bwk et al. (2005)

The figures also confirm one of the urban legends illustrated at the beginning: there are rather heterogeneous average work loads between the different disciplines (looking at higher education institutions only). While in the SSH almost half of the working time (45 %) is spent on teaching and training, this amount is significantly smaller in every other disciplinary field. It is therefore hardly astonishing that the SSH rank last in terms of time devoted to research activities (taking Medicine without clinics) (see Table 3).

Table 3: Higher Education Sector: working loads of scientific personnel in % (2002)

	teaching & training	research	other tasks
Natural sciences	29.5	64.4	6.1
Technical sciences	31.3	61.5	7.2
Medicine (incl. clinics)	16.8	36.7	46.5
>> without clinics	24.7	65.8	9.5
>> clinics	14.0	26.3	59.7
Agriculture, Forestry & Veterinary Medicine	25.6	57.0	17.4
Social Sciences & Humanities	45.0	47.4	7.6
>> Social Sciences	43.8	48.5	7.7
>> Humanities	46.5	46.1	7.4

Source: bm:bwk et al. (2005)

… and where does the money come from?

The most important financier for RTI in any disciplinary field is still the public sector. The figures for 2002 show that the SSH can draw on almost the same amount of public research funds as Medicine (EUR 293.9 million vs. EUR 309.0 million). Taking the Social Sciences and the Humanities separately, their available funding is comparable to that of the Technical Sciences, although there are differences concerning the origin of the money (see Table 4.). With regard to the relevance of European level funding, the participation of SSH researchers in respective programmes (like Citizens or Science and Society) is quite low compared to other disciplinary fields (bm:bwk et al 2005).

Table 4: Higher Education Sector: Where does the RTI money come from? (2002) (in 1000 EUR)

	public	private, non-profit	foreign (excl. EU)	EU
Natural Sciences	348,201	987	4,877	19,738
Technical Sciences	142,708	924	3,751	8,785
Medicine (incl. clinics)	309,012	3,937	1,908	4,519
Agriculture, Forestry & Veterinary Medicine	63,123	918	696	1,666
Social Sciences & Humanities	293,905	1,422	564	3,136
>> Social Sciences	160,501	639	153	2,526
>> Humanities	133,404	783	411	610

Source: bm:bwk et al. (2005)

If we look at state funding as distributed by the federal ministries in 2003 (bm:bwk et al 2005), we see that the SSH are the main beneficiaries, since 38.4 % (EUR 21.7 million) of state RTI funding goes into this field. The SSH are mainly funded by the bm:bwk (Federal Ministry for Education, Science and Culture), while the bmvit (Federal Ministry of Transport, Innovation and Technology) and the BMWA (Federal Ministry of Economics and Labour) as well as other ministries play a minor role (see Table 5).

Table 5: Higher Education Sector: where does the state money for RTI come from? (2003)

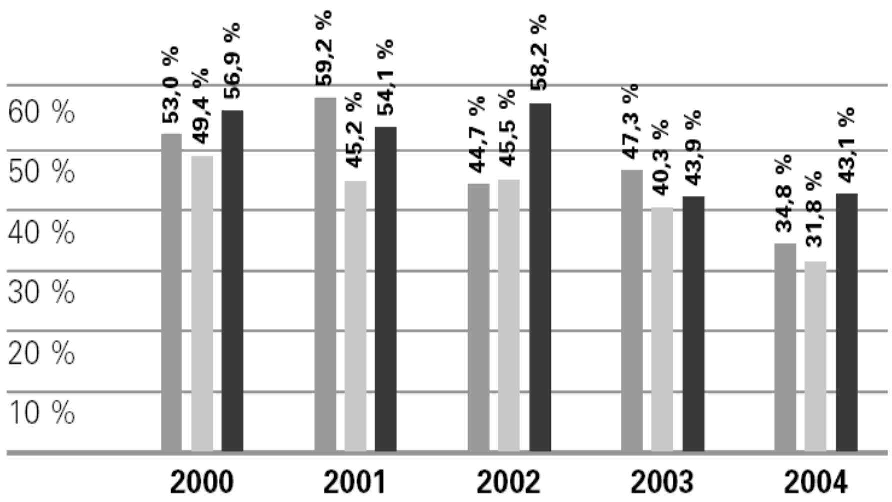
	total sum	in %	bm:bwk	bmvit	BMWA
Natural Sciences	11,099,561	19.6	8,794,489	617,055	187,097
Technical Sciences	7,472,237	13.2	1,686,947	5,214,199	132,180
Medicine (incl. clinics)	13,264,064	23.5	12,848,845	89,969	-
Agriculture, Forestry and Veterinary Medicine	2,997,521	5.3	464,558	-	10,433
Social Sciences & Humanities	21,698,122	38.4	18,189,699	1,129,160	950,817
>> Social Sciences	14,735,356	26.1	11,226,933	1,129,160	950,817
>> Humanities	6,962,766	12.3	6,962,766	-	-

Source: bm:bwk et al. (2005)

… and how successful is competitive funding?

Beside state funding the Austrian Science Fund (FWF) is the main funding body for basic research, also in the field of SSH. The guiding principle of the FWF is to try to fund basic research projects based on quality, without taking the discipline into account. However, the fund is frequently accused of having a disciplinary bias in funding. While we cannot give answers to this issue on a project per project basis (since it is not our task here to analyse single funding decisions) we do have some hints that there is no disciplinary discrimination[1] with respect to acceptance rates and to amounts of money granted. Analysing the acceptance rates concerning FWF project funding over time (FWF 2004) we find that the SSH do not face significantly lower acceptance rates in comparison to Biology/Human Medicine and Natural Sciences/ Technical Sciences. The fact that the rate decreased from 53.0 % in 2000 to 34.8 % in 2004 is to be ascribed to changing numbers of applications for funding and also holds for the other disciplinary fields (see Figure 1).

Figure 1: FWF: Approval rates – stand alone projects



Source: FWF (2004)

1 Assuming that the number of low quality research proposals handed in is comparable in any of the disciplines, the following arguments also draw conclusions on these questions. However, as already illustrated in the two introductory narratives, this is not a shared perception. Thus, we will handle the empirical data that do not give any information about quality aspects with care and will be cautious in drawing our conclusions.

In order to stick to the question of application numbers: In the period from 1998 until 2003 a total of 4,242 applications for funding of stand-alone projects were handed in at the FWF. The SSH account for 25.1 % of the proposals, while 45.5 % of all applications originate in the Natural Sciences. With respect to the number of accepted proposals the share of the SSH remains almost unchanged, namely 24.8 % of successful proposals belong to this scientific field. However, the picture changes a little bit when referring to the share of granted budget, as only 19.6 % of the FWF-money goes into the SSH, compared to 54.6 % for the Natural Sciences. Interestingly, SSH account for roughly the same size of grants as Human Medicine (see Table 6). What we can further learn from these figures is that Human Medicine and the Natural Sciences plan the biggest projects (around EUR 210,000 per application), while the SSH only apply for about EUR 150,000 for an average project (Streicher et al. 2004).

Table 6: Applications and grants by main field of science, 1998 - 2003

	shares [%] in:			
scientific fields	**# proposals**	**# accepted proposals**	**solicited amount**	**granted amount**
no classification	0.1	0.1	0.4	0.8
Natural Sciences	45.5	51.2	49.5	54.6
Technical Sciences	5.8	5.1	5.2	4,8
Human Medicine	22.1	17.8	24.3	19.2
Agriculture, Forestry & Veterinary Medicine	1.5	1.0	1.4	1.0
Social Sciences & Humanities	25.1	24.8	19.2	19.6
>> Social Sciences	7.5	5.0	5.8	3.8
>> Humanities	17.6	19.8	13.4	15.8

Source: Streicher, G. et al. (2004)

On the basis of the presented empirical evidence, we can not confirm the myth that the SSH are being systematically financially starved. Although we should be aware of the fact that, due to the heterogeneity and the multiplicity of actors, some might indeed be starved, the picture doesn't completely hold up under scrutiny for either the Social Sciences nor for the Humanities. In order to grasp the likely underlying (structural) problems in the SSH that this myth is based on, further and particularly focused research is needed and strongly recommended.

Further interesting results are presented in an evaluation study on the FWF, drawing on a probability-analysis with respect to approval rates (Streicher et al 2004). The "benchmark project" that the authors identified on the basis of FWF application data was a Natural Sciences project submitted by a male non-professor coordinator in his forties asking for EUR 150,000 to

EUR 250,000.[2] Starting from this "benchmark project", which has a typical approval rate of 52.4 %, different scientific disciplines face the following differences in probabilities of approval (see Table 7). We can learn from these figures that projects within different scientific fields have indeed significantly different chances of getting funded, if all other factors are being held constant. Just like most of the other scientific fields the Social Sciences are rejected far more frequently than the "benchmark project". However, we see that the Humanities are more successful. Thus, again, we cannot seriously conclude a bias against the SSH (alone) on the basis of this model.

Table 7: A "Benchmarking Exercise"

Benchmark: Natural Sciences, male coordinator, non-professor, 40-50 years, 150,000 - 250,000 € funding requested, approval rate: 52.4 %

changing Variable	% difference in approval rate
Technical Sciences	- 8.5
Human Medicine	-15.1
Agriculture, Forestry and Veterinary Medicine	-18.1
Social Sciences	-19.2
Humanities	+ 4.5

Source: Streicher, G. et al. (2004)

Cautionary remark: Do not interpret the coefficients causally.

We have already pointed out that the field of SSH is a fairly heterogeneous one. If we go deeper into the individual disciplines, we can conclude the following: talking about SSH as one group is not only misleading because of differences between the Social Sciences and the Humanities (e.g. as seen before), but also because of quite astonishing differences within these two groups. In Table 8 we can see that funding in the Humanities considerably differs between individual disciplines: the Historical Sciences are far ahead with respect to grants awarded by the FWF. A similar picture goes for FWF-funding for the Social Sciences (see Table 9), where Economics clearly takes the lead.

[2] Further characteristics were included but do not make a difference for our argumentation. For details see Streicher, G. et al. (2004): 19 et seqq.

Table 8: FWF: Funding in the Humanities, 2002 – 2004

	2002		2003		2004	
	in m EUR	in % of total FWF fundings	in m EUR	in % of total FWF fundings	in m EUR	in % of total FWF fundings
Philosophy	1.20	1.31	0.70	0.70	1.05	0.98
Theology	0.14	0.15	0.34	0.34	0.55	0.52
Historical Sciences	6.20	6.77	5.86	5.89	5.29	4.96
Linguistics and Literature	2.05	2.24	2.37	2.38	3.15	2.95
Other Philological and Cultural Studies	1.30	1.42	1.40	1.41	1.71	1.60
Aesthetics and Art History	2.05	2.24	1.57	1.58	1.47	1.38
Other	0.92	1.01	0.58	0.58	0.32	0.30
Total	13.86	15.14	12.82	12.88	13.54	12.69

Source: FWF (2004)

Table 9: FWF: Funding in the Social Sciences, 2002 – 2004

	2002		2003		2004	
	in m EUR	in % of total FWF fundings	in m EUR	in % of total FWF fundings	in m EUR	in % of total FWF fundings
Political Science	0.26	0.28	0.22	0.22	0.16	0.15
Legal Science	0.54	0.59	1.44	1.44	0.47	0.44
Economics	0.43	0.47	1.87	1.88	3.18	2.98
Sociology	0.35	0.38	1.43	1.44	0.51	0.48
Psychology	0.58	0.63	0.72	0.72	0.70	0.68
Regional Planning	0.03	0.03	0.04	0.04	0.01	0.01
Applied Statistics	0.07	0.08	0.18	0.18	0.75	0.70
Pedagogy, Educational Science	0.03	0.03	0.07	0.07	0.09	0.08
Other	0.62	0.68	1.09	1.10	1.19	1.12
Total	2.91	3.17	7.06	7.09	7.06	6.64

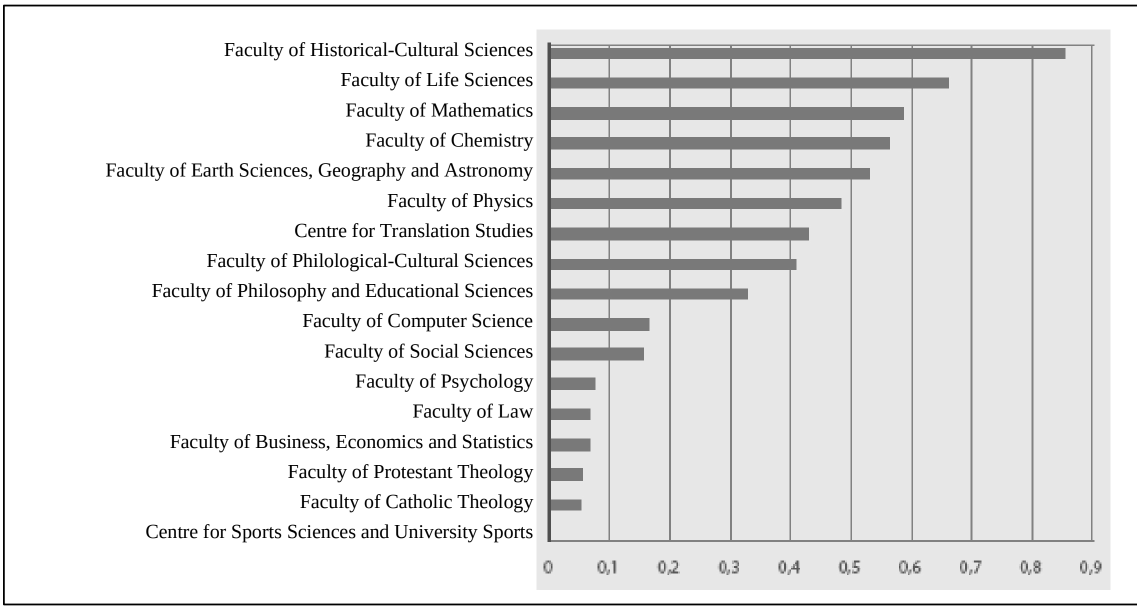
Source: FWF (2004)

Close up: the University of Vienna

Taking the University of Vienna, the largest and most comprehensive Austrian university, as an example, we can see that the average number of FWF funding per scientist (in full time equivalents, FTE) in the period from 2001 to 2003 is highest at the Faculty of Historical-Cultural Sciences (more than 0.8 funding approvals per FTE scientist), followed by the Faculty of Life Sciences, the Faculty of Mathematics and the Faculty of Chemistry. In the last quarter of the ranking (with less than 0.1 approvals per FTE scientist) we can find the two theological Faculties as well as the Faculty of Business, Economics and Statistics, the Faculty of Law and the Faculty of Psychology. Other SSH Faculties rank in between (see Figure 2). So we can conclude that, on the basis of the number of FWF funded projects per scientist in FTE at the University of Vienna, a Faculty within the Humanities astonishingly takes the unchallenged lead (University of Vienna 2005).

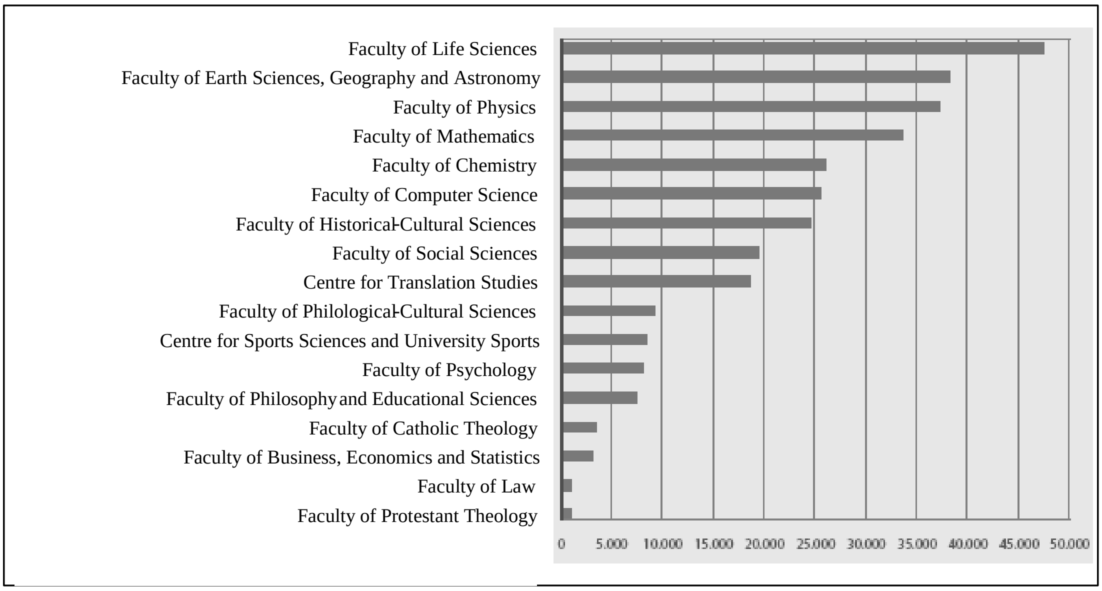
Taking the average budget obtained via third party funding as the basis for comparison the picture changes. As one can imagine, typical research projects in the Natural Sciences are more cost-intensive than projects in the SSH because of the expensive equipment and instruments needed for laboratory work. With respect to average third party funds (including FWF funding) acquired per scientist (in FTE) in the years 2001 to 2003, the Faculty of Life Sciences reigns supreme, while the Faculties within the SSH rest in the middle of the ranking (see Figure 2.). Again, the myth of starving the SSH cannot be confirmed on the basis of the figures available.

Figure 2: the University of Vienna: average number of FWF-funding per scientist (FTE), 2001-2003



Source: University of Vienna (2005)

Figure 3: University of Vienna: average third party funds acquired per scientist (FTE), 2001-2003



Source: Universität Wien (2005)

Conclusions

We have tried to show in figures that the general myth of the starving of the SSH (in Austria) can not clearly be upheld on the basis of the empirical data presented. Taking into consideration that the kind of argumentation we are using is sooner an empirically based discussion of the issue than a comprehensive analysis, we want to emphasise that no discriminatory behaviour against the Austrian SSH can be found with respect to research funding.

However, we do see some hints that there might be some structural disadvantages that the SSH are facing and that lie beyond too little money available for SSH. Therefore, we strongly recommend further empirical research, which should focus on the peculiarities of the SSH in order to not only grasp the origin and the survival of the myth but also to understand whether the SSH are given equal chances but do not make proper use of them (i.e. do not provide high-quality products) or whether they make most efficient use of their potentially unequal starting positions. We would thereby recommend taking issues like the aging of research active staff, fragmentation of research units, lack of career perspectives for young researchers etc. into consideration. And, like in any other scientific field, quality assessment and the use of adequate indicators should be addressed with most effort.

MYTH 2: THE FUNDING GAP

"It is highly visible that a historic research promotion gap between basic research which does not directly refer to applications and applied research with immediate potential for application exists in Austria."
(Forschung in Österreich. Eine forschungspolitische Stellungnahme der Österreichischen Forschungsgesellschaft, Vienna, February 2003)

"The Austrian Industrial Research Promotion Fund (FFF) should readjust its research promotion towards long-term pre-competitive research projects; the share of projects gearing towards radical innovations has to increase. A particular priority is to strengthen the co-operation with the Austrian Science Fund (FWF) in order to bridge the gap between basic and applied research promotion."
(Austrian Council for Research and Technology Development, 3 December 2002)

"The research promotion gap on the interface between basic and applied research, which was criticised for a long time, is to be closed."
(Austria Innovativ, 27 January 2005)

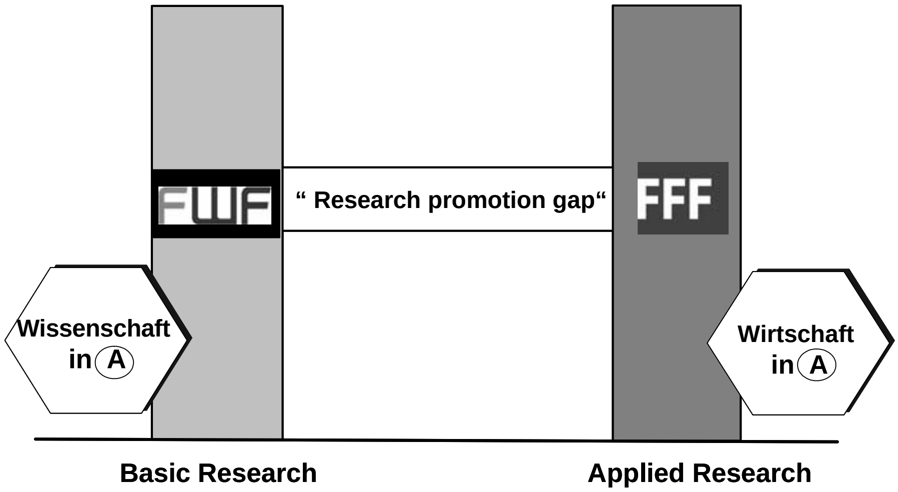
These statements provide evidence for a long lasting myth about research promotion in Austria: the existence of a research promotion gap between basic and applied research. We seek to dismantle this myth by identifying its evolvement and by looking at some data on research funding in Austria.

The evolvement of the myth: FFF & FWF

The perceived gap in research promotion funding is closely linked with the basic principles of the RTI funding system in Austria, which was laid down in the late sixties of the 20th century. The Research Promotion Act constituted Austria's two main funding agencies, the Austrian Industrial Research Promotion Fund (FFF) and the Austrian Science Fund (FWF). Both agencies act upon the basic principal of bottom-up project funding, and whereas the FWF was restricted to fund basic research within the academic sector, the FFF focused on financing industrial innovation. (Stampfer 2005)

In 2004, the FFF was integrated into the Austrian Research Promotion Agency (FFG), which was established via the merger of the FFF and three other research promotion and funding agencies – namely the Technologieimpulse Gesellschaft (TiG), the Bureau for International Technology Co-operations (BIT) and the Austrian Space Agency (ASA). Since then, FFF-funding is operating just like before under the Division "General Programmes", which constitutes Austria's most important source of finance for research and development projects carried out by industry.

Figure 4: Common perception of the funding gap



Based upon: Kratky 2004

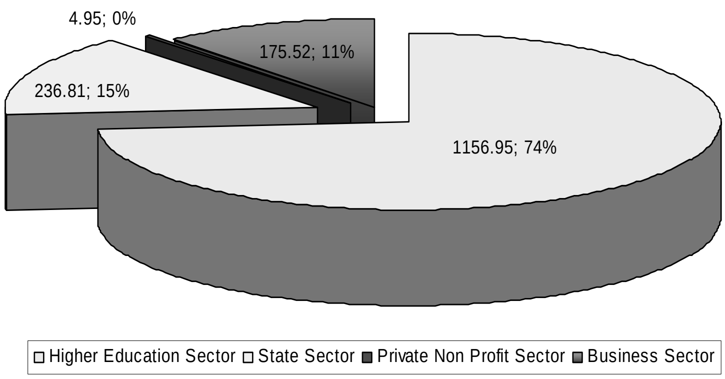
As Jörg (2004) stresses, the foundation of FFF and FWF as two autonomous funding bodies with a fairly narrow mandate on bottom-up project funding has set out important conditions for the evolvement of the governance structure of RTI-policy: the short and medium term steering power at the policy level has remained limited. Big shares of available financial resources have been tied up with general university funds and basic funding of the major research performers in the public realm. Bottom-up project funding has remained the dominant funding instrument throughout the last three decades. Priority settings along selected strategic fields as well as the introduction of new funding instruments have remained difficult.

The underlying rationale of this funding system in which the FWF exclusively funds basic research and the FFF exclusively funds applied research, was the idea of a linear innovation system in which basic science leads to applied science, which in turn leads to engineering, innovation, jobs and ultimately wealth (see Arnold (2004)). This oversimplifying linear model of science, which is dating back to the influential paper of Vannevar Bush, has nowadays gained momentum again, for instance with the creation of the ERC (European Research Council), gearing towards a separation of basic and applied research at the EU policy level (see Pichler's article on the expectation of policy makers).

Public research funding in Austria

The largest part of public R&D funding still goes to the higher education organizations (in 2002: EUR 1,160 million; 75 %) and is mainly distributed via the General University Funds (GUF). Funds to the private sector only account for 11 % of total public funding and the state sector has a share of 15 %.

Figure 5: Public research funding by sectors in Austria, 2002 (in m EUR)



Remark: Data exclude contributions to international Organisations
Source: bm:bwk et al. 2005 (Statistical Annex – Table 18)

As reported in the Austrian Research and Technology Report 2005, public R&D funding in Austria in 2002 amounted to EUR 1,574 million[3]. EUR 1,183 million stem from the federal government, additional EUR 212.2 million were distributed via the two research promotion funds (FFF and FWF EUR 210.4 million) and the Innovation and Technology Fund (ITF). The federal provinces provide EUR 171 million and municipalities provided EUR 7.7 million.

Hence, the FFF and the FWF contributed to 13.4 % of the public research funding in 2002. This is quite a bit, but we may ask: where do the other funds go to, how are these resources distributed, and is there really a funding gap between basic and applied research?

FFF, FWF and “the others”

General university funds definitely make up the largest share of public R&D funding. Of the EUR 1,156 million of funding for the Higher Education sector in 2002 approximately EUR 920 million are characterised as general university funds and EUR 92 million stem from the FWF.

For the business sector government financing is rather limited: business R&D expenditures financed by the public sector amounted to EUR 176 million in 2002, which is only 5.6 % of total business R&D expenditures (EUR 3,131 million). FFF funding, which is mainly devoted to the industry sector, amounted EUR 118 million in 2002.

Despite the importance of funding provided by FWF and FFF, these two funds are not the only actors on the research promotion scene: several other programmes and institutions have emerged within the last few decades, which focus exactly on collaborative R&D financing at the interface between academic institutions and the application oriented industry R&D base. The central aim of these initiatives is to foster know-how and knowledge transfer in both directions that can lead, e.g., to novel products or improved service and production processes.

For a differentiation of promotion schemes that fit into the concept of the funding gap, one can distinguish between institutional and programme oriented research promotion schemes.

For the institutional level we would like to mention the following programmes:

- The Competence Centre Programmes Kplus, Kind, Knet (Edler et al. 2003)
- The Christian Doppler Society (Schibany et al. 2005).

At the programme level one may mention the programmes:

- Aeronautics (TAKE OFF)
- Information & Communication Technologies (FIT-IT) (Zinöcker 2005a)
- Intelligent Transport Systems & Services (IV2S)
- Nanotechnology Initiative (NANO)
- Sustainable Development -Nachhaltig Wirtschaften (Paula 2003)
- The Austrian Space Programme ASAP & ARTIST

All the above mentioned programmes support explicitly selected national thematic priorities to foster R&D projects involving both Austrian companies and research organisations. Funding from these thematic programmes and institutional funding schemes is considerable, but difficult to trace as there is no monitoring system that would allow us to follow funding flows easily.

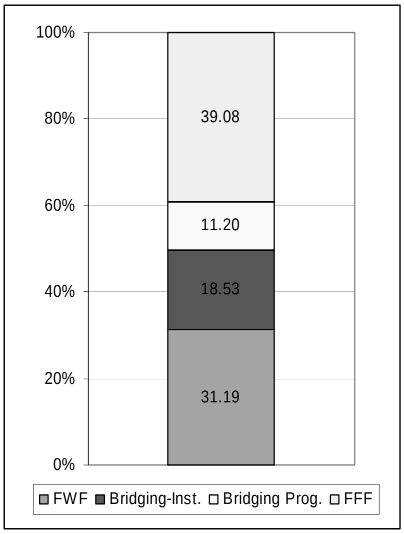
[3] The official R&D survey was conducted in 2002.

In addition to these programmes a non-thematic Translational Research Programme was set up by the FWF in 2004 in order to bridge the gap between basic and applied research. Academic research with applications potential is funded, no business partners are intended.

Taking funding from FFG-BP and FWF as the two poles between innovation oriented research and basic research promotion, the following image emerges for the year 2004 (see Figure 6): FWF funding, excluding the Translational Research Programme, amounts to EUR 101 million (31 % of the total R&D expenditures) whereas the FFG-BP has provided about EUR 127 million (39 %) for applied research. Institutional bridging initiatives (Competence Centres, CDG) amount to EUR 60.3 million (19 %) of the public funding, and bridging programmes amount to EUR 36.5 million (11 %).

Figure 6: A funding Gap?

Funding Instrument	*Million Euros*
FWF – Basic Research:	***101.51***
FFF – Applied Research:	***127.17***
Kplus	30.2
Kind/Knet	24.4
CDG	5.7
Institutional Bridging Initiatives:	***60.3***
TAKE OFF	3.97
FIT-IT	5.49
IV2S	9.48
NANO	3.92
Sustainable Development	6.47
ASAP & ARTIST	2.02
FWF - Translational Research	5.11
Bridging Programmes:	36,46
Total:	***325.44***



Sources: Austrian Research and Technology Report 2005, Evaluierung der CDG 2005

Conclusions

A simple look at the data reveals that a funding gap between basic and applied research does not exist (anymore). The actors of the Austrian research and technology policy area have developed a variety of instruments in order to bridge the funding gap that has been detected more than a decade ago. Whereas institutional programmes seek to establish long-term science and industry relations, thematic research programmes focus on both research and innovation in specific fields that are considered to be of national importance.

Overall, the funding gap per se is not a guiding theme with regard to the development of novel funding programmes. However, we should also make a cautionary remark at the end: while currently there is enough budget for “gap–programmes”, we have not discussed whether or not it would be necessary to redesign them.

Myth 3: Engineering ignored

"Austrian Science Fund FWF: 'Engineering= Application?' Yuk! " [4]
(Ernst Bonek, farewell lecture 2004)

Currently, the field of engineering has a marginal position in the FWF budget. Only 6 % of all applications (5 % of funded projects) between 1998 and 2003 were in the field of technical sciences. Only 5 % of the FWF's budget for research in the same time span went to engineering or construction projects. This is a considerable decrease since the early 1990's, when about 9 % of the FWF budget went to the 'technical sciences'.[5]

A remarkably low position

The FFF/FWF evaluation (especially Van der Meulen 2004) expressed some concerns about these data: "The low position for engineering is remarkable as the Engineering Sciences have made considerable progress in the past decades and have grown to the basic sciences. Internationally, one can see increasing budgets for the Engineering Sciences […]."[6]

An evaluation of the Swiss National Science Fund (SNF) came to a comparable conclusion: "It is estimated that engineering and computer sciences together account for […] a very small share of the overall budget of the SNF", whereas at the same time the industries that depend critically on those fields are considered "the pride of Switzerland" (SNF (2001): 12 et seq.).

Fair treatment documented

There could be several explanations for this situation: one of them could be that, for what ever reason, the Science Funds might treat engineering in an unfair way. The FWF evaluation clearly rejects this innuendo by stating that Vienna's University of Technology, the most important university in the field of engineering in Austria, has the second highest approval rate at the FWF of all Austrian universities. The overall approval rate for Technical Sciences is 47 %, which is better than for Human Medicine or Social Sciences, but worse than for the Natural Sciences or the Humanities (Streicher 2004).

[4] Translation of the original: "FWF: Ingenieurswissenschaften = Anwendung? Pfui!"

[5] We do not know if attribution problems in the classification of projects might contribute to this decline.

[6] Van der Meulen 2004. Interestingly, Van der Meulen does not provide the reader with any evidence for this statement.

Looking beyond the borders of the FWF: a vivid programme scene for technical sciences

In this respect one should note that the Austrian technical universities (which clearly focus on engineering, Computer Sciences and/or construction) do not depend on FWF funding. Engineering faculties only draw on the FWF for 10-15% of their external funding, whereas the FWF's share is more than half of the total funding for the faculties of Natural Sciences or Humanities (Streicher (2004)). One explanation for this is that the technical sector is supported by a vivid programme funding scene that is independent from the Austrian Science Fund and attracts scientists from the field of engineering, computer sciences and construction. There is a budget of more than EUR 30 million p.a. for applied research programmes, e.g. in the fields of micro- and nano- technologies, IT, or space. This amount would be higher by far, if we also considered the competence centres.

Blurring, but still valuable categories of research

Can those special programmes (often endowed with a clear linkage to industries) replace classical project funding for scientific research? Are “scientific research” and “applied research” still valuable terms or do they refer to blurring categories of research? The evaluations of special programmes in Austria (Zinöcker et al 2004, 2005a, 2005b) do not provide too much evidence in this respect. Two hypotheses: the interest of engineers in FWF project funding is not very high; their interests and needs are easily covered by the special programmes and the applied research mentioned. Or: scientists in the field of engineering in Austria lapse into a vicious circle: due to their strong belief that the Science Funds discriminate them, they concentrate on special programmes and on applied research; therefore they disregard ‘pure’ scientific research. We have not found any evidence that would falsify or verify either of these hypotheses.

Different framework conditions

Within the programme logic of the FWF, engineering projects are treated in the same way as projects in the Humanities or Natural Sciences in every respect. For reasons of fairness, this is more than appropriate. On the other hand, when funding basic research projects, is equal treatment always fair? E. Bonek (2004), among others, stresses the fact that, according to the rules of the FWF, young researchers who are working in FWF-funded projects earn half of what they could earn in the private sector. Organisations comparable to the FWF, like the German Research Foundation (DFG), take these framework conditions into account and pay higher wages: for example, PhD students in engineering projects earn up to EUR 58,600 p.a., which is considerably above the FWF-rates (about EUR 28,000).

Special programmes are successful

As a result of these different framework conditions (and as a reaction to a possibly low level of visibility for the field of engineering) some funds have launched special programmes that particularly address the field of engineering. The National Science Foundation (NSF) in the United States, for example, has launched its own engineering directorate in the mid 1980s, which today accounts for over 25 % of NSF's annual budget. The Swedish Foundation for Strategic Research, as another example, intends to fund research to carry out creative projects of the highest scientific quality in the fields of Natural Sciences, engineering and Medicine. More than 40 % of its grants, partly earmarked, go to projects in the fields of engineering.

Conclusions

The evaluation of the FWF has shown that 'engineering' is not being treated in an unfair way within the programme logic of the Austrian Science Fund. The low number of engineering projects may result from the vivid programme scene for engineers and technicians outside the realm of the FWF, and/or the special framework conditions (e.g. the job market that is not taken into account by the FWF), more complex research approaches in engineering (e.g. involving cooperation with diverse business sectors) or even from classification problems.

Although there are international examples of special programmes (or even divisions) for engineering which are successful, this paper is not intended to advocate for the establishment of such programmes in Austria. First, a system (evaluation) approach is needed to create a sound basis for such a decision and to identify reasons for the creation of such programmes. Questions such as the following should be answered:

What is, from a programme portfolio perspective, the situation of engineering in Austria? Are there, besides the Science Fund, enough possibilities (e.g. in thematic programmes) for Engineering Sciences?

Besides the job market, what are the special framework conditions for Engineering Sciences?

Are there any factors which define the success of scientific projects in the field of engineering that make the classical Science Fund project funding unattractive (e.g. the necessity of networking, or collaborations with industry)?

Myth 4: Networks, fretworks7

"It is widely assumed that collaboration in research is 'a good thing' and that it should be encouraged." […]
"Implicit in the enthusiasm for research collaboration and the policies aimed at fostering it are a number of assumptions: one of them is that "more collaboration is actually better" and that we can measure the effects of collaboration and hence determine whether or not it is changing as a result of a particular policy"
(J. Sylvan Katz and Ben R. Martin (1997): What is research collaboration? Research policy 26, 1-18)

"The Austrian NANO Initiative has the […] objective to strengthen and network the Austrian NANO players in science and industry" (www.nanoinitiative.at)

"There is nothing inherently wrong with networks", but… . Networks are seen as "powerful instruments of scientific innovation" and "collaborations and networks are part of scientist's daily life" but they should be "allowed to form spontaneously and not be enforced. […] Scientists want to cast nets, not to be caught in them" (all quotes from Schatz (2004))

Were policymakers on the wrong track when they designed network programmes during the last years? Would all necessary scientific networks have been formed anyway, without any kind of public intervention? Or is there a need to nudge scientists to cooperate and to think in other dimensions than those within the walls of their laboratories and cubicles?

Today, networks seem to be 'en vogue' in all parts of our society. This also holds true for research and development activities. Here, networks are known to form spontaneously and/or are encouraged (and sometimes enforced) by RTI policy. Over the last decades, programme managers have developed special support measures and public funding instruments to foster formal as well as informal collaboration. On the one hand, for example, there is funding for joint research projects based on sophisticated consortium agreements or for mobility programmes. These possibilities parallel the classical stand-alone research projects of science funds targeted at academia or industry. On the other hand there is also room for establishing and cultivating personal relationships at regular informal meetings organized by single researchers, associations and institutions, but also in the context of funding programmes.

A broad range of initiatives currently supports regional, national and international partnering involving academia, industry or both. The public support of collaboration has a lot of different and sometimes also contradictory facets and rationales. There are multiple programmes and the expectations of policy makers regarding their impact are high.

[7] The quote "Networks, Fretworks" was used in Gottfried Schatz (2003): Jeff's View: Networks, fretworks. FEBS Letters 27577, 1-2

Some Austrian examples for funding instruments aiming at joint RTI projects:

(1) Academia only. There are instruments aiming at interconnecting excellent academic groups on a regional level (Special Research Programme, SFB) or nation-wide (National Research Networks, NFN). These programmes allow for mid-term to long-term research work on complex research topics.

(2) Academia only / academia and industry / industry only. The Austrian Genome Research Programme (GEN-AU, www.genau.at) has been set up to strengthen, focus and integrate the respective national research capacities in academia and industry. Any partner constellation is possible.

(3) Compulsory academia – industry partnership. In order to create knowledge that might be of economic use and to encourage the application of existing know-how for example Christian Doppler Laboratories (Schibany et al. 2005), Competence Centre Initiatives (K*plus*, K*ind*/K*net*, Edler et al. 2003) and special thematic programmes (e.g. FIT-IT www.fit-it.at/ Zinöcker et al. 2005a), NANO Initiative, www.nanoinitaitve.at) have been established. These initiatives also aim at making Austria's enterprises more innovative.

Funding of academic research by the Austrian Science Fund (FWF):
Harmony of stand-alone projects and networks

As described in the FWF research network evaluation report (PREST / ISI Fraunhofer 2004), multi-site and bottom-up collaborative research (Joint Research Programmes, FSP) has been funded since the early 1970s as well. Thus, Austria has a more than 30-year-long tradition of collaborative research. In 1993 a second network funding instrument has been developed (Special Research Programmes, SFB), which aims at single-site, bottom-up, long-term and interdisciplinary work on complex research topics. In 2004 both funding programmes accounted for about 22 % of FWF's budget compared to 61 % for stand-alone projects. (See Table 10)

Looking back into the 1980s and 1990s, funding for stand-alone projects ranged between roughly 60 % and more than 80 %, with multiple ups and downs. The funding for network programmes shows a complementary trend and was in the range of 10 to 30 %. In the 1990s, a shift from joint research programmes to special research programmes can be observed. (See Fig. 7)

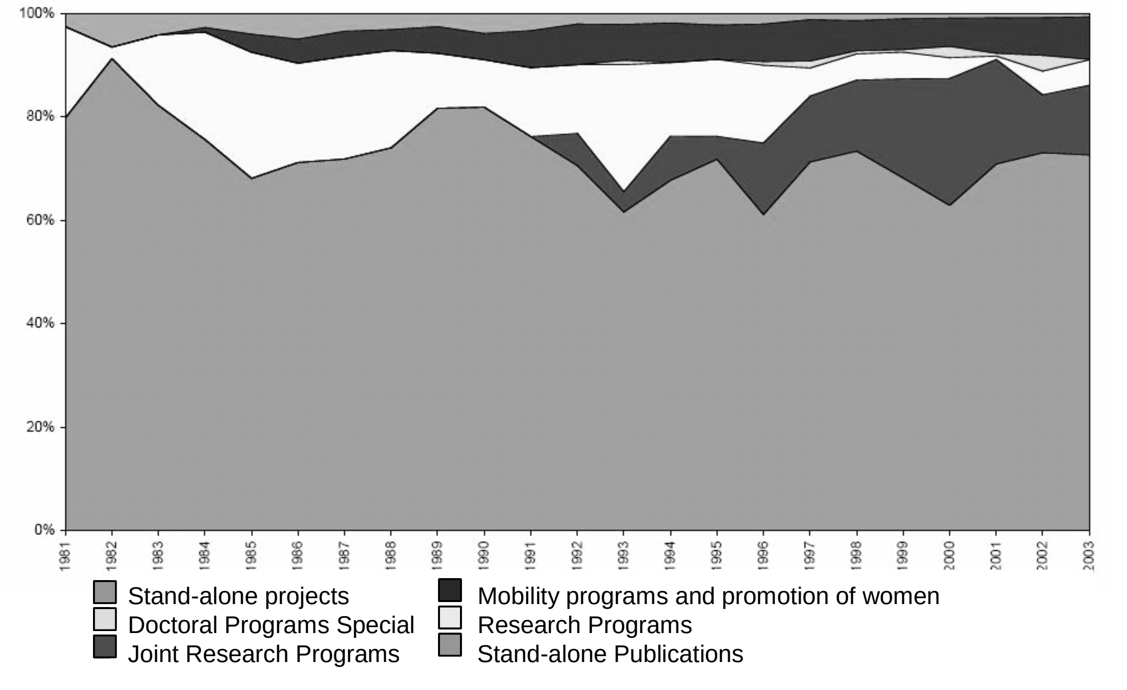
Table 10: Stand-alone projects versus collaborative projects

Funding programmes	FWF Stand-alone Projects	Joint Research Programs (FSP)[8]	Special Research Programs (SFB)
Established in	1967	1972	1993
Thematic direction	Bottom-up	Bottom-up	Bottom-up
Duration	3 years	6 years	10 years
Location	Single-site	Multi-site	Single-site
Average number of subprojects	1	~ 8	~ 12
Objectives	Funding of individual research projects not oriented at financial profit	Promotion of the establishment of "priority" research areas, generally by building up nation-wide research networks for the multidisciplinary, distributed and medium-term work on large-scale research projects	Establishment of research networks based on intern. standards through autonomous research concentration at a single university location. Building up of extremely productive, tightly interconnected research establishments for long-term and interdisciplinary work on complex research topics
Budget share (2004)	~ 61%	~ 6%	~ 16%

Sources: Jakob Edler und John Rigby (2004): Research Network Programmes Evaluation for the Austrian Science Fund (FWF) and www.fwf.ac.at

[8] Renamed „NFN (National Research Networks)" beginning of 2005

Figure 7: FWS's funding portfolio over time



Source: Erik Arnold et al (2004): Evaluation of the Austrian Industrial Research Promotion Fund (FFF) and the Austrian Science Fund (FWF) - Synthesis Report

FWF networks are developed in a bottom-up manner and there are no thematic restrictions. The number of partners and subprojects is highly variable. On average there are 12 subprojects for SFBs (Special Research Programmes) and 8 subprojects for FSPs (Joint Research Programmes). Because of the high variation in network size these numbers do not point to something like 'the ideal network project'.

According to the evaluation of the FWF's network programmes, existing collaborations with network partners are beneficial for the success of an application. With regard to the implementation a strong leading team, high commitment and a pronounced sense of responsibility of the partners are described to be crucial.

The Austrian Genome Research Programme (GEN-AU): Multiple facets of networking.

In 2001 the Austrian Genome Research Programme (GEN-AU) has been launched in order to strengthen, focus and integrate genome research in Austria and to foster networking among all relevant stakeholders in academia and industry. Each year about EUR 10.7 million are spent on the programme, making it Austria's largest thematic programme. Large cooperative projects and networks, pilot projects and accompanying research projects addressing questions related to modern life sciences from a social science point of view are carried out. GEN-AU invites academic and/or industrial partners to suggest RTI projects. In phase 1 of the programme a trend towards rather basic research projects has been observed. In addition, the project consortia turned out to be smaller than expected. Predominantly national joint RTI projects are carried out. 72 % of the budget spent on RTI projects has been invested in four cooperation projects and two networks.

Table 11: Projects in phase I of GEN-AU

Project type	Number of projects	Minimum number of partners	Average number of partnering institutions	Share of project budget
Cooperative projects	4	4	6	59 %
Networks	2	*)	9	13 %
Pilot projects	6	*)	3	16 %
Associated projects	5	*)	2	7 %
ELSA projects	6	-	3	5 %

*Source: Klaus Zinöcker et al (2005): Austrian Genome Research Program GEN-AU': Mid Term Program Management Evaluation; *) funding instruments have only been developed during the first round of applications in phase I of the program*

The Social Network Analysis (SNA) conducted within the GEN-AU evaluation exercise lead to the image of a complex pattern of intensive communication within this cast of actors. The evaluators got the impression of a complex functional network that involves a remarkably high degree of project- and cluster-spanning cooperation. Moreover, cooperation across different projects and project types could be identified. The SNA detected a multiplex structure of relationships that is based on a complex system of communication as well as an intensive exchange of research related information, material and personnel between the projects.

To reflect programmes like GEN-AU in a more contextual manner, a portfolio point of view is necessary. Numerous scientists interviewed in the context of the GEN-AU programme management evaluation took such a broad view and immediately expressed the fear that programmes like GEN-AU could decrease or substitute the funding of classical stand-alone projects. These have been described to form the core of Austria's research system and would either need to be left untouched or, even better, be expanded. The interviewees expressed very clearly that they did not call for another network programme if the total funds for RTI stayed as they were, but neither did they complain about the "network forces" within the programme.

Costs and benefits: Taking the good with the bad

The evaluation of the FWF's network programmes and authors cited therein (PREST / ISI Fraunhofer 2004) clearly point out that network benefits come at a considerable cost, especially in case of joint RTI activities. Lots of time and effort need to be invested in the set-up of networks and in keeping the collaborations alive. This is especially true when it comes to joint applications for research projects. These, however, often allow obtaining substantial financial input for larger research projects that probably cannot be conducted on the basis of only piecing together multiple smaller grants. On the other hand, larger funds also go together with increasing costs for administration and management. And larger projects also mean increasing bureaucracy. Furthermore, different IPR strategies, different accounting standards and contradictory management cultures might make collaboration cumbersome.

Networking at the same time allows for combining skills, division of work, efficient use of abilities and access to (tacit) knowledge. Due to simultaneous memberships in multiple networks a single partnering decision can often also give access to the additional networks of the respective partner (whether this is something positive or not, is another question). Critical masses and fruitful learning environments can be created through personal interaction that also provides social support.

Nevertheless, there might also be compatibility problems in case of joint projects. These could range from disputes regarding the overall focus of the project, concerning research questions and methods or the handling of results and the consequences of the research activities. Different worlds of thought, however, are not only a source of conflict but might also be powerful sources of creativity for inter- and transdisciplinary research. Networking also allows for the rapid dissemination of results and contributes to increased visibility. Increased meeting hours and travel costs need to be accepted in return.

Would these networks have been formed without the substantial amounts of money that the science fund has spent? Would these positive effects have occurred without the FWF funding scheme, motivating (and forcing) scientists to form networks? This should be doubted, indeed.

Conclusions

Networking is still an important issue in designing RTI initiatives. In any case, whenever we think about making collaboration obligatory and formal, we should never forget that network benefits come at a cost.

Networks are part of the daily routines of scientists and business people. They do not always require rigid guidelines and also form up beyond any coordinated funding programme. In addition, teams are in a flux and collaboration changes constantly.

Evaluating network funding is quite a challenge and should neither be underestimated nor be tied up with exaggerated expectations. Collaboration is a multi facetted, often long-term social endeavour. It is hard to grasp with numbers (e.g. joint publications: co-author analyses – for example Katz and Hicks 1997 - are not sufficient but rather misleading) but calls for novel qualitative approaches. We know little about how the different kinds of network funding change day-to-day RTI. We know little about the budget portfolios of those carrying out joint research projects. No evaluation approach has tried to construct the counterfactual to networking.

Existing evaluations mainly concentrate on the output side (comparatively easily available data on co-publications) instead of taking the different stages of research collaboration into account (informal discussions, joint applications, collaborative day-to-day research work, co-publications, initiation of future projects, etcetera, just to mention a few). Thus, qualitative research is needed in order to provide a more detailed understanding of (collaborative) research as an improved basis for the design of RTI programmes.

Are networks fretworks? The answer that we find is uncomfortable for policy makers, because there is no unambiguous answer. When planning new initiatives, policy makers should never forget about the bottom-up character of collaboration. Some forms of interaction can hardly be initiated the top-down way. On the other hand, we could clearly observe additional effects of network programmes: things that would have never happened without public intervention. Networks have no intrinsic value, but initiatives that foster networking activities, if carefully planned, are clearly legitimate and useful.

Myth 5: Impact, Impact, Impact!

„The European Council agreed to set the goal of increasing Europe's research investment to three per cent of GDP by 2010"
(www.cordis.lu)

"Evaluators conduct systematic, data-based inquiries-"
(American Evaluation Association. Guiding Principles for Evaluators, www.eval.org)

"Carefully formulated objectives include both strategic and operational or if possible, even quantified targets, with the relation between the different objectives clearly defined so that they form a transparent system of objectives."
(Evaluation Standards, Platform Research and Technology Policy Evaluation, www.fteval.at)

"Enhancing the quality of life, increasing Austria's competitiveness, maintenance of Austria as a business location, enhancing R&D quality in Austra, giving Austria's society a vision."
(Initiative "Technologies for a Sustainable Development", www.nachhaltigwirtschaften.at)

Evaluators usually regard transparent and – if possible – quantifiable tasks as a prerequisite for assessing RTI programmes and they expect interventions (and their initiators) to set themselves such tasks in that field. Programme and policy makers should have a clear vision of 'what they want to achieve' using a special intervention: this is an essential requirement, which is expected from "good governance" within RTI policy.

A clear set of objectives does not only add colour to RTI programmes but is central in defining the overall orientation of programmes, their territories and borders, the instruments applied etcetera. At the same time, a clear set of objectives eases and actually allows the work of evaluators. Intangible agendas of possible impacts of a programme and an unclear definition of the intended outcomes cause negative effects on the quality of evaluation reports. Needless to say: at the end of their work, evaluators should give more substantial and meaningful recommendations than, e.g., “focus and specify your system of objectives”. Such recommendations are clearly of a limited value for policy makers and programme managers.

The clearer and the more quantifiable, the more transparent a system of objectives is; the greater also are the evaluators’ possibilities and the higher is the value of the evaluations for programme managers and policy makers. Through such a transparent system of objectives the spectrum of possible methods is widened and the need for a data collection throughout an RTI programme becomes clearer. The work with (quantitative) data[9] is one of the prerequisites for a good evaluation.

At the same time it is one of the greatest challenges in the planning of RTI programmes to set out a clear and achievable mission / vision that is connected to a set of objectives. To “increase the competitiveness within a national economy” is a highly reputable mission or goal, although, within a programme evaluation, impossible to be evaluated. The „creation of 15 new science-industry linkages“ may be seen as much more unattractive, although it might be more realistic, at least at first sight.

Towards a „homeopathic“ RTI policy?

In the planning of a system of objectives, budgetary restrictions of policy interventions need to be taken into consideration: within Austria yearly budgets of programmes range between EUR 2 and 10 million. These budgets are usually split in completely different programme lines and are categorized under various sub-goals. It must be questioned, how it is possible to, e.g., “increase the quality of life” with such small funds:

A nation’s GDP could be seen as an important yardstick on how to increase a nation’s quality of life. The Austrian GDP in 2004 was EUR 237 billion. The initiative “Technologies for Sustainable Development” (www.nachhaltigwirtschaften.at), for example, is an Austrian programme that initiates and supports “trend setting research and development projects” contributing to a “general research orientation towards sustainability”: (in good years) it has a budget of about EUR 7 million (Paula 2004), which represents 0,003 % of Austria’s GDP. The programme’s goals are “enhancing the quality of life” and ‘enhancing Austria’s competitiveness’. Quite optimistic goals, indeed.

Cumbersome ways from idea to market

An often-mentioned goal of RTI programmes is the development of start-ups, spin-offs, or high-risk innovations. An example of how hard it can be for a start-up to introduce risky ideas to a global market is the biotech company Intercell www.intercell.com.

[9] Which was hopefully collected during the lifespan of the programme / the project / the policy or can be collected during the evaluation exercise.

Intercell is one of Austria's most noted companies in the field of Life Sciences: "Intercell develops vaccines for the prevention and treatment of infectious diseases" (Intercell's mission statement). The company was founded in 1998 as an academic spin-off; it now has 130 employees and it can look back on a successful financing history. Since its foundation, Intercell has managed to raise more than EUR 150 million including over EUR 50 million out of its Initial Public Offering (IPO) in 2005. EUR 9 million (out of EUR 150 million) are public funds - including funds from Austrian regional and federal agencies like WWFF, ZIT, FFG and aws contributing approx. 6 % of the required funds (Kandera 2005). There is no doubt that, especially in the beginning, these 6 % have been crucial, but they are only one link in the chain.

Today, eight years after its foundation, Intercell has a diversified research portfolio with strategic partnerships (e.g. with major multinational pharmaceutical companies but also recently founded Intercell spin-offs), a range of products in the pipeline (e.g. vaccines to prevent travellers diarrhoea and tuberculosis) and two advanced clinical product candidates (a prophylactic Japanese Encephalitis vaccine and a therapeutic vaccine against Hepatitis C). The market launch of the vaccine against Japanese Encephalitis is planned for 2007. Despite good evidence in clinical trials, it is still not clear whether the vaccines will be accepted for the market launch by the responsible regulatory offices and whether Intercell's products will be successful on the vaccines market. So, there is a long (7 years) and expensive (more than EUR 150 million) way from idea to IPO, while the market has not been reached yet.

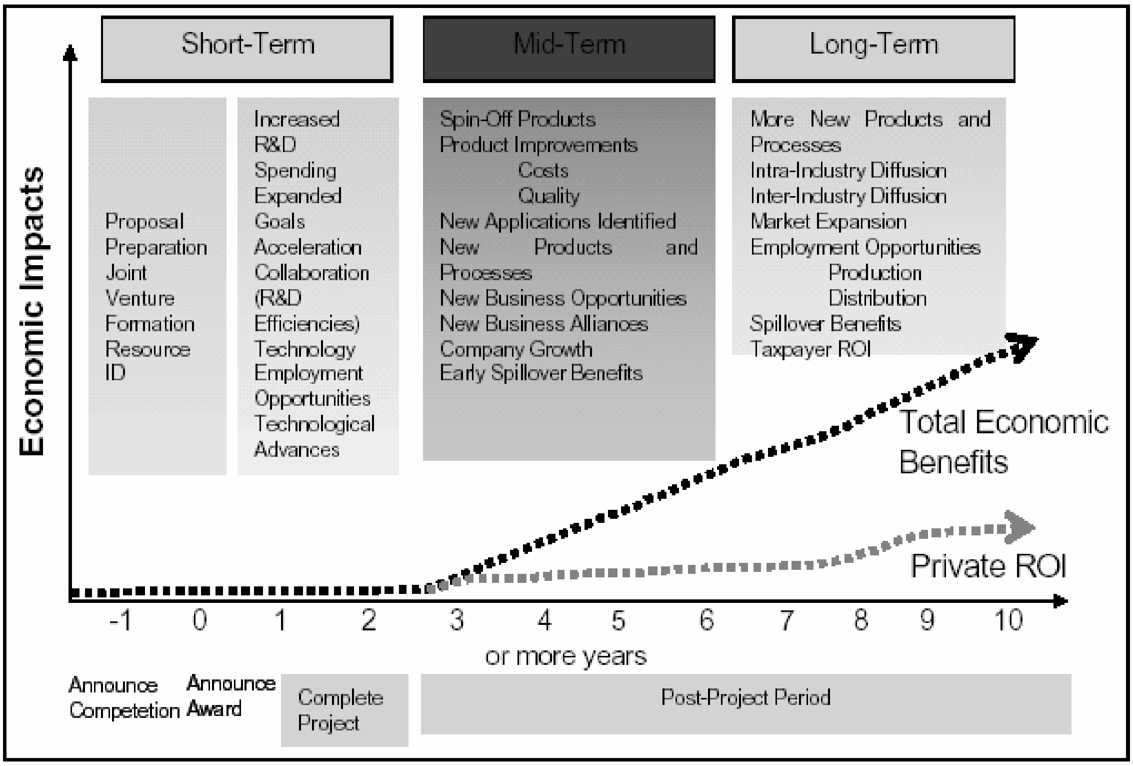
Impulses

Let's compare Intercell (EUR 9 million public funds, a more than EUR 150 million budget, no product on the market) with the most important federal programme to fund scientific research in the field of Life Sciences: the 'Austrian Genome Research Programme GEN-AU'. As described above, the programme is planned for a period of nine years and has a budget of about EUR 100 million. In the first three years several research projects with about 90 partners have been funded. Beside its main goal, which can be described as the promotion of high quality research, it also stipulates a number of other objectives: create knowledge in order to enhance health, create new jobs, promote women, create patents, create start-ups, promote the location and increase the public acceptance of life sciences. Whereas the "promote science"-goal of the programme is ambitious but realistic, the GEN-AU evaluators have criticized the goal overload of the programme: of course, GEN-AU can contribute to, e.g., 'promoting the location', but GEN-AU is only one (tiny) issue among a broad range of activities addressing this goal. The same holds true for the other goals of GEN-AU. Therefore, even the most ambitious (and well-financed) programmes can only give impulses rather than be the one and only factor for reaching programme goals.

Impatience

With interventions in research and development, policy makers intend to reach a variety of more or less "classical" outcomes and impacts. Some of them have already been mentioned, i.e. "quality of life", "competitiveness", "sustainable new science industry linkages", "foundation of new companies", "more and better jobs" etc. All these impacts have one thing in common: they need time to bear fruit. (And therefore "output in numbers" is not available for the next press conference on the schedule.) While, e.g. 'collaboration' can be built up fairly quickly, it takes years to see how sustainable such collaborations are. Especially social benefits (e.g. positive environmental effects and other public goods) can only be expected within a five-year period or probably even at a later stage. Ruegg and Feller (2003) are very confident in their description of these factors for economic impacts and indicate that they would only become visible after 10 years.

Figure 8: Impact of technology oriented RTI programmes



Quelle: Ruegg & Feller 2003

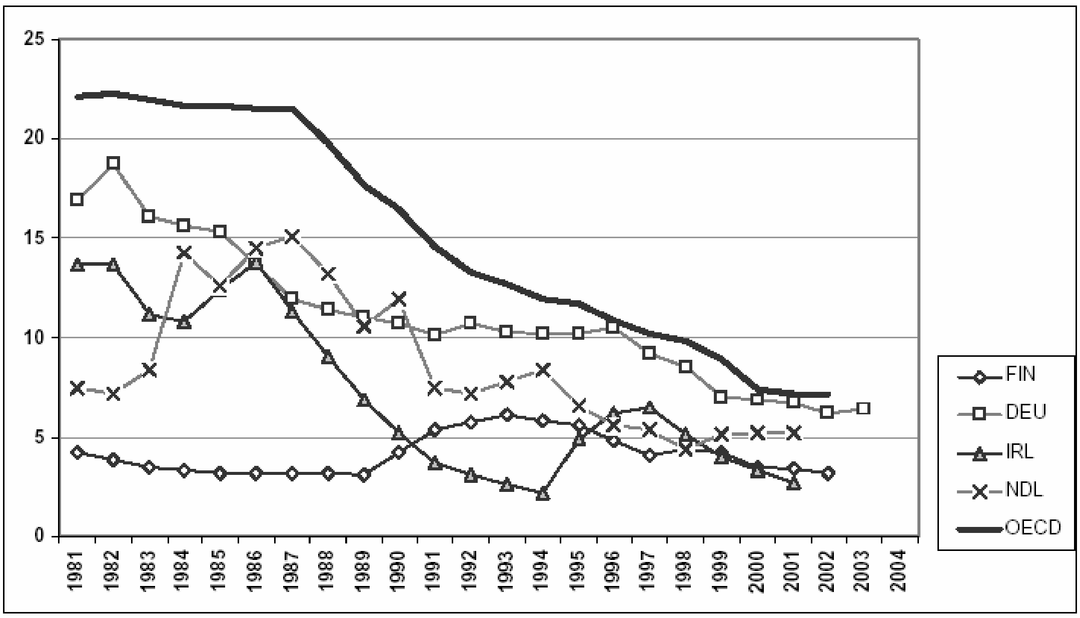
However, evaluators are only in a few, rare cases given the opportunity to analyse such long term effects extensively. Policy makers and programme managers are up to inform and legitimize their activities here and now. If possible, they adjust their activities accordingly while an initiative is running, since after its end there are no more opportunities for improving the programme. Therefore it is evident that those who are commissioning programme evaluations rarely have the motivation to initiate long-term ex post studies. Moreover, ten to fifteen years after the start of an initiative there are different stakeholders in charge for RTI programmes, the knowledge accumulated in the initiatives is partly lost, and the priorities and interests on the policy agenda may have changed.

In any case the introduction of such long-term studies would be appropriate. Without such studies a whole stream of information is lost, which could, however, be very relevant for RTI policy: What could legitimize a new RTD programme better than providing an overview of all the contributions to social and economic benefits of a previous programme? What could give hints in regard to what might be realistic expectations? How can a realistic system of objectives be introduced? What else can tell what is possible through intervention in the field of RTI – and what is not?

Modesty

Schibany und Jörg (2005) show that the share of the state in financing R&D in enterprises has continually decreased. Whereas in 1980 in the OECD area there was still an average of 23 % of state financed RTI in enterprises, today this share is clearly below 6 %. In Austria the rate is around 3.6 % (without ever having been considerably higher). (See Figure 9)

Figure 9: Share of the direct state funding of industry internal R&D (in % of the complete internal RTI expenditures of industry).



Source: Schibany & Jörg 2005

Policy makers expect a lot from investing in RTI: some of the intended outcomes and impacts have been specified above. Looking specifically at the Austrian share of 3.6% of industrial RTI expenditures, one should be even more careful with goals such as "raising the living standards" or "increasing the chances in the competition for location": Explaining positive effects only with public investments in RTI might be audacious. The actual endeavour of the evaluators is to show the contribution of such investments to the emergence of positive effects. Accountability represents only a part of the whole picture.

Conclusions and Consequences

Moderate attitude of policy makers: there are large sums of public funds that are invested in RTI. Nevertheless it would be immodest to state that each public intervention in RTI will have important economic effects. A more modest and more realistic planning of goals would be appropriate.

Moderate attitude of evaluators: There is not a great deal of tradition in analysing long term social and economic impacts; accountability is still an endeavour; a lot of it is new territory and hard to grasp methodically. Evaluators should not be tempted to underestimate the methodological challenges (and promise too much).

Conditioning expectations through long-term ex post evaluations: Policy makers should feel encouraged to call for evaluation studies, which are not directly relevant for their day-to-day business but increase the general understanding of the impact of their interventions and thereby allow a more realistic planning of RTI programmes.

Wrap-up

We have discussed five myths: strong beliefs, assumptions, and persistent pictures on "how things are going" in the (Austrian) RTI landscape and we have tried to discuss them in a more evidence-based manner. What have we learned?

- We have presented relevant data on Social Sciences and the Humanities, out of various sources: research information systems, agency data and data from evaluation reports. On the basis of these data, we can not hold up the myth that the SSH are systematically financially starved. Also, a bias or discriminatory behaviour against SSH cannot seriously be concluded.
- There is no funding gap (anymore).
- We have shown that complaints about "too little attention" given to engineering are premature. Anyhow, special framework conditions (e.g. the job market) or, possibly, more complex research approaches in engineering have to be taken into account.
- Networks are not always fretworks – Networks in RTI have clear advantages, but whether they have an intrinsic value is debatable.

- We have shown that stakeholders might be too optimistic when it comes to the intended impacts of their interventions.

We have also shown some starting points for further research:

- We strongly recommend an evaluation effort on the SSH in Austria.
- While there is enough budget for ‘gap-programmes’ in Austria, their design should be scrutinized. Maybe, there is enough money, but in the wrong pipelines. Maybe FFG-BP and FWF funding have to change, moving from a system seeking to fill a gap to something completely new.
- Determinates of success in the field of engineering may be different from other disciplines. What are the consequences of differences in the behaviour of ‘academic tribes ‘for agencies?
- New ideas for grasping the consequences of network funding would be of use.
- There is a need for long term impact evaluations of public interventions in the field of RTI.

A “fast and furious” debate of RTI policy issues is misleading. Maybe the first and foremost function of evaluation in an innovation system is to bring the debates back to evidence.

Literature

Arnold, Erik et al (2004): Evaluation of the Austrian Industrial Research Promotion Fund (FFF) and the Austrian Science Fund (FWF). Synthesis Report, Brighton - Wien

bm:bwk, bmvit, BMWA (2005): Forschungs- und Technologiebericht 2005 (Austrian Research and Technology Report 2005.

Bonek, E (2004): farewell elcture, Institute of Communications and Radio-Frequency Engenieering, Vienna University of Technology, Vienna 2004

Edler, J. / Bührer, S. / Lo V. / Sheikh, S. (2003): Assessment „Zukunft der Kompetenzzentren-programme und Zukunft der Kompetenzzentren“ (Assessment „Future of the Competence Centres Programmes and Future of the Competence Centres“), Evaluation Report, Fraunhofer Institute for Systems and Innovation Research, Karlsruhe / Austrian Institute for SME Research, Vienna

FWF (2004): FWF Statistik 2004, Vienna 2004

Jörg, L (2004): Policy profile Austria - Input paper for the OECD NIS MONIT Network, www.tip.ac.at

Jörg, L. / Falk, R. (2004): Evaluation of the Austrian Industrial Research Promotion Fund (FFF) and the Austrian Science Fund (FWF). Background report 3.1.2., FFF: Internal functioning and customer satisfaction. Wien

Kandera, R (2005): From Idea to IPO. Presentation during the award ceremony of the business plan competition “BOB - Best of Biotech”, Vienna 14. 6. 2005.

Katz J.S. and Hicks D. (1997): How much is a collaboration worth? A calibrated bibliometric model, Proceeding s on the Sixth Conference of the International Society for Scientometric and Informetric, Jerusalem, Israel, June 16-19

Katz J.S. and Martin B. R. (1997) What is research collaboration? Research policy 26, 1-18

Kratky, G (2004): Vortrag - Aktuelles Angebot und geplante Programme des FWF, FWF. October 7th 2004

Paula M. (2004): Impulsprogramm Nachhaltig Wirtschaften. Zwischenbilanz 2004, bm:vit

PREST / ISI Fraunhofer (2004): Research Network Programmes Evaluation Report for the Austrian Science Fund.

Rigby J and Edler J (2005): Peering inside research networks: Some observations on the effect of the intensity of collaboration on the variability of research quality, Research policy 34, 784-794

Ruegg, Rosalie / Feller, Irwin (2003): A Toolkit for Evaluating Public R&D Investments: Models, Methods, and Findings from ATP's First Decade. NIST GCR 03-857

Schatz, G. (2003): Jeff's View – Networks, Fretworks. FEBS letters 27577, p. 1-2

Schibany, A. / Nones, B. / Schmidmayer J. / Jörg,L. / Warta K. / Sheikh S. / Edler J (2005): Evaluierung der Christian Doppler Gesellschaft CDG. Joanneum Research, KMU Forschung Austria, Frauenhofer – ISI

Schibany, A., Jörg, L (2005).: Instruments of the Technology development and their mix. Intereg Research Report Series Nr 37-2005.

SNF (2001): The Swiss National Science Foundation (SNF). Achievements, Performance, Perspectives., Report of the external evaluation group, Page 12f, Site Visit 10 – 12 September 2001

Stampfer, M (2005): How These Things Came About: A Short Note on the Early Years of FFF and FWF. In: How to Evaluate Funding Systems. The example of the FFF/FWF Evaluation. Platform Newsletter 25, www.fteval.at

Streicher, G. / Schibany A. / Dinges M. (2004): Evaluation FWF - Impact Analysis. Background Report 4.2. InTeReg Research Report no. 23-2004, Vienna.

University of Vienna (2005): Universität Wien 2010. Entwicklungsplan der Universität Wien. Wien.

Van der Meulen, Barend (2004): Evaluation of the Austrian Industrial Research Promotion Fund (FFF) and the Austrian Science Fund (FWF). FWF Governance and Processes. Background report 4.1. Enschede

Zinöcker, Klaus / Dinges, Michael (2004): Evaluation of the Austrian Industrial Research Promotion Fund (FFF) and the Austrian Science Fund (FWF). Positionierung des FFF und FWF vis-a-vis anderen Finanzierungsinstrumenten. Wien

Zinöcker, K. / Dinges, M. / Schibany, A. (2004): Mid Term Evaluation Microtechnics, Joanneum Research 2004

Zinöcker, K. / Schindler, J. / Gude, S. / Schmidmayer J. / Stampfer M. (*2005a):* Interims-evaluierung FIT – IT. Konzepte, Rahmenbedingungen, Design, Prozesse. Snapshots auf Wirkung und Additionalität. Studie im Auftrag des bmvit. Wien

Zinöcker, K. / B.Tempelmaier / A. Radauer / I. Fischl / R. Steiner / R. Ruegg (2005b): Austrian Genome Research Programme GEN-AU: Mid Term Programme Management Evaluation, Final Report

Authors:

Klaus Zinöcker
Platform Research and Technology Policy Evaluation www.fteval.at &
WWTF Vienna Science and Technology Fund
Währingerstraße 3 /15a, 1090 Vienna,
klaus.zinoecker@wwtf.at

Michaela Glanz
WWTF Vienna Science and Technology Fund
Währingerstraße 3 /15a, 1090 Vienna
michaela.glanz@wwtf.at

Brigitte Tempelmaier
WWTF Vienna Science and Technology Fund
Währingerstraße 3 /15a, 1090 Vienna
brigitte.tempelmaier@wwtf.at

Michael Dinges
Joannem Research INTEREG
Wiedner Hauptstraße 76, 1040 Vienna
michael.dinges@joanneum.at

Der Newsletter der Plattform Forschungs- und Technologieevaluierung GesbR
ist ein unregelmäßig erscheinendes offenes Forum zur Diskussion methodischer
und inhaltlicher Evaluierungsfragen in der Forschungs- und Technologiepolitik.

ISSN: 1726-6629

Herausgabe und Versand:
Dr. Rupert Pichler, Dr. Sonja Sheikh, Mag. Klaus Zinöcker
Plattform Forschungs- und Technologieevaluierung GesbR,
A-1090 Wien, Währingerstraße 3/15a, office@fteval.at

Für den Inhalt dieser Ausgabe verantwortlich:
Plattform Forschungs- und Technologieevaluierung GesbR
Mag. Klaus Zinöcker

PLATTFORM fteval
Forschungs- und Technologieevaluierung

PLATTFORM FORSCHUNGS- UND TECHNOLOGIEEVALUIERUNG

Die Plattform Forschungs- und Technologieevaluierung GesbR ist eine Initiative der folgenden Organisationen: Österreichisches Bundesministerium für Bildung, Wissenschaft und Kultur (bmbwk), Bundesministerium für Verkehr, Innovation und Technologie (bm:vit), Bundesministerium für Wirtschaft und Arbeit (bmwa), Österreichische Forschungsförderungsgesellschaft mbH (FFG), Fonds zur Förderung der wissenschaftlichen Forschung (FWF), Joanneum Research, KMU Forschung Austria, ARC Systems Research, Technopolis Austria GmbH, Österreichisches Institut für Wirtschaftsforschung (WIFO), Wiener Wissenschafts-, Forschungs- und Technologiefonds (WWTF) und dem Zentrum für Innovation und Technologie GmbH (ZIT), Rat für Forschung und Technologieentwicklung, AQA – Österreichische Qualitätssicherungsagentur, Christian Doppler Gesellschaft (CDG), Austria Wirtschaftsservice (awsg). Im Rahmen der Plattform werden Themenstellungen zur Forschungs- und Technologieevaluierung erarbeitet und – z.T. unter Einbeziehung namhafter ExpertInnen – in einem Fachkreis diskutiert. Der Newsletter beinhaltet Fachbeiträge zu Fragen der forschungs- und technologiepolitischen Evaluierung. Die Herausgabe erfolgt in zeitlicher als auch inhaltlicher Abstimmung mit Plattform-Veranstaltungen, um die Synergiewirkungen eines breiten Austauschforums zu nutzen.

"new frontiers in evaluation"

Conference

April 24th – 25th 2006
Radisson SAS Palais Hotel

Vienna, Austria

Plattform Forschungs- und Technologieevaluierung GesbR.
A-1090 Wien, Währingerstraße 3/15a, Austria
www.fteval.at