PLATTFORM fteval
Forschungs- und Technologieevaluierung

Excellence: to pick or to foster?

Nr. 30
June 2007

BUNDESMINISTERIUM FÜR WISSENSCHAFT UND FORSCHUNG,
Minoritenplatz 5, A -1014 Wien
Dr. Reinhard Klang reinhard.klang@bmwf.gv.at
Mag. Irene Danler irene.danler@bmwf.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

Austrian Research Centers GmbH – ARC,
Donau-City-Straße 1, A-1220 Wien
Petra Wagner-Luptacik petra.wagner@arcs.ac.at
Susanne Giesecke susanne.giesecke@arcs.ac.at

FWF-Fonds zur Förderung der wissenschaftlichen Forschung,
Haus der Forschung, Sensengasse 1, A-1090 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
Dr. Rahel Falk rahel.falk@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,
Haus der Forschung, Sensengasse 1, A-1090 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),
Haus der Forschung, Sensengasse 1, A-1090 Wien
Dr. Elvira Welzig
welzig@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
Haus der Forschung, Sensengasse 1, A-1090 Wien
Mag. Klaus Schnitzer
klaus.schnitzer@ffg.at

Rat für Forschung und Technologieentwicklung
Pestalozzigasse 4/DG 1, A-1010 Wien
Mag. Michaela Topolnik
m.topolnik@rat-fte.at

Preface

How can one define, recognize and compare scientific excellence as objectively as possible? During a one day workshop in December 2006 jointly hosted by the Platform Research and Technology Policy Evaluation and the Austrian Council, this prime question was discussed with international experts, Austrian Scientists and policy makers.

Excellence might be seen as the negative buzz word of the scientific year 2006; a critical mind once noted that "scientists put 'excellence' in their tabernacle and never dare to question the appropriateness of a lead concept like this". Anyway, this concept, or, in other words, the search for excellence is the overwhelmingly accepted 'currency' in the scientific world.

Is it possible to recognize excellence and to systematically advance it? We have asked ourselves these questions and discussed it with representatives of relevant organizations. Beate Konze-Thomas, head of the department Coordinated Programmes and Infrastructure in the German Research Foundation (Deutsche Forschungsgemeinschaft DFG), introduced the concept of the German excellence initiative of the federal states. She gave account of emotional discussions, especially in the end of the decision process. The Austrian programmes, e.g. the Excellence programme Science of the bmwf and the FWF, as well as the Competence Center Programme COMET of the FFG, are still awaiting these emotions – they are now in the conceptions' and bidding phase.

Marcel Herbst, owner of the consulting firm 4mation and former head of planning of ETH Zürich, had a critical look at the day-to-day business of the science's sponsors with a view to seven thesis: There is a difference whether such sponsors fall back to existing excellence, or whether they try to develop "Excellence" anew by, for example, creating their general conditions accordingly. He pointed out that the Ivy League Universities in the USA were partly created in the 18th century. According to Herbst, this fact must be taken into consideration in the current discussion.

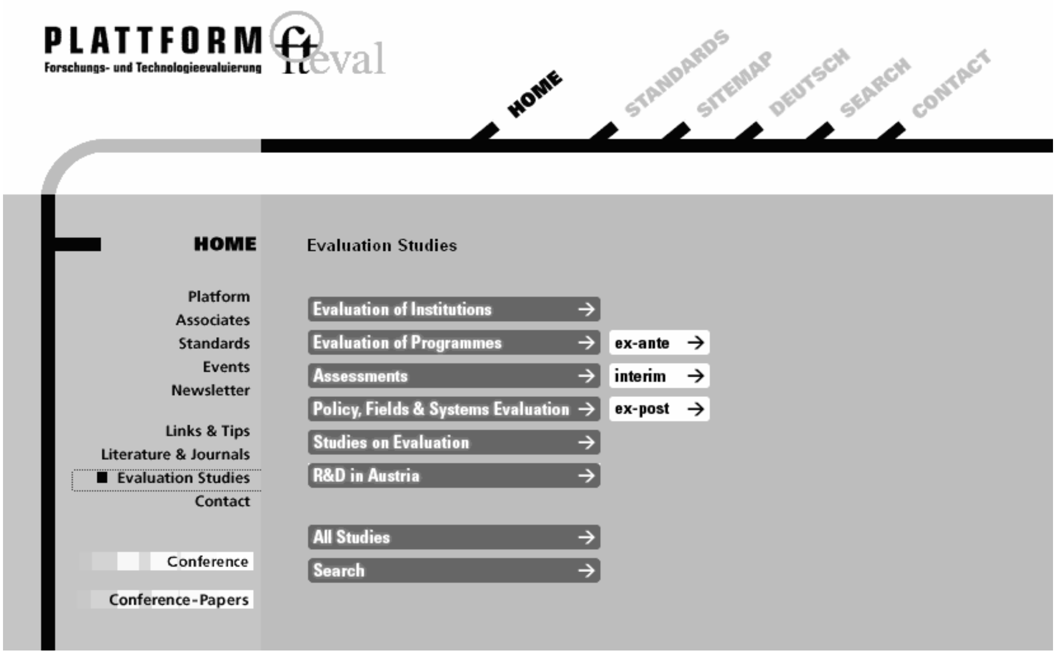
In this Newsletter, you will find the key note address of Dervilla Donnelly from the Austrian Council and Marcel Herbst's paper with a number of observations and theses regarding the state of affairs of European higher education and Humboldtian universities. All the presentations and some pictures of the event "Selecting Excellence" can be downloaded on www.fteval.at.

Beside our main topic, you will find several interesting articles on various aspects of Evaluation. *Thomas M. Pelsoci* describes in his article the use of detailed case studies that build on financial analyses to evaluate programme impacts. He broaches the issue of an advanced new technology which is an important vital sign in emergency medicine.

The differences between action research and empirical social research paradigm are the topic in *Christoph Mandl's* paper. *Stefan Kuhlmann* reviewed "A New Deal for an Effective European Research Policy - The Design and Impacts of the 7th Framework Programme**"** written by Ugur Muldur, Fabienne Corvers, Henri Delanghe, Jim Dratwa, Daniela Heimberger, Brian Sloan, and Sandrijn Vanslembrouck. This book shows elements of a textbook on European societal and economic needs, research performance and research policies.

News from our Homepage: Within the last couple of weeks we made a big effort to improve our homepage www.fteval.at. The domain "Evaluation Studies" with a special search function was introduced, so now all evaluations of the last couple of years are arranged according to the following categories:

- *Evaluation of Institutions*
- *Evaluation of Programmes (ex-ante, interim, ex-post)*
- *Assessments*
- *Policy, Fields & Systems Evaluation*
- *Studies on Evaluation*
- *R&D in Austria*



Furthermore all evaluation studies available on our homepage can be listed at once. (button "All Studies") If you search a specific study use the new search system (see button "Search"). The criteria therefore are: "*Year*", "*Categories*", "*Language*" and the keywords "*Title*", "*Client*", "*Evaluato*r" and "*Content*".

All these improvements ease and accelerate searching for evaluations on our homepage.

Mag. Michaela Topolnik
Austrian Council - Rat für Forschung und Technologieentwicklung
Pestalozzigasse 4, 1010 Wien
m.topolnik@rat-fte.at

Mag. Klaus Zinöcker
WWTF Vienna Science and Technology Fund
Währingerstr. 3/15a, 1090 Wien
Klaus.Zinoecker@wwtf.at

Dervilla Donnelly

Excellence, Scientific Quality and Research Frontiers

Welcome Address given to the joint workshop „Selecting Excellence".

The workshop on "Selecting Excellence" was jointly hosted by the Platform Research & Technology Policy Evaluation and the Austrian Council. The opening speech was given by Prof. Dervilla Donnelly, Member of the Austrian Council (Rat für Forschung und Technologieentwicklung) since 2000.

The first time I considered the subject of excellence in relation to science and science policy was when I was the Irish member of CODEST, the Committee of the European Commission which dealt with funding for fundamental research. Its first Chairman was Sir Peter Swinnerton-Dyer, a truly brilliant mathematician from the United Kingdom. He used to start all meetings with the comment, "We are only going to fund excellence". Something which we indeed did consistently. He laid down the rules to succeed in obtaining a grant: five votes from the members of the Committee were needed. At that time there were twelve countries in the European Union and the larger countries each had three representatives on CODEST, leaving the smaller member countries with just one member each. To negotiate funding for a scientist from a small country required an understanding of simple mathematics and a consummate skill in the art of bargaining. I must admit my country was reasonably happy with its success rate.

We are much more sophisticated today - we have to be, and this is a good thing - but to define excellence in relation to the division of funds is not easy, and so our discussions at the conference and beyond may lead to some helpful guidelines. We all know that scientific quality is a necessary precondition to push back research frontiers and to open up new fields of knowledge. In Austria there is a magnificent history of achievements in science. For example Schroedinger: everybody knows he was a genius and his scientific work was excellent. Achieving and maintaining scientific excellence has always been crucial for leading research workers at the frontiers of science.

In today's world fundamental research is only one of the streams we have to worry about - applied research is another. When we are talking of excellence, it doesn't matter whether research is applied or fundamental. We are looking for excellence in the topic, and so obviously different rules and regulations for assigning excellence will be involved.

We discuss excellence for individual programmes and projects. If we can lay down certain rules, it is much easier to identify excellence but if after five years we can look back and say: "Yes, we made the right decision", that is fine but of little value at this moment in time when the problem is on the table in front of you.

If you are looking for excellence in scientific institutions, that's another and a very different situation than looking for excellence in a scientific business: you have to have success and you have to have success quickly, so for excellence in business you must take a different approach. So when we are defining excellence, we can't have a "one fits all" approach.

Unfortunately today it is a very crowded and competitive marketplace where excellent proposals and the good reputation of those who submit them do not guarantee that they will secure funding or tenure. This is in part due to underfunding, a situation not unknown and well recognized for many years, particularly by scientists. Funding agencies and governments face the same pervasive evaluation question, no matter what country you are in. So the prime question is "How can one define, recognize and compare scientific excellence as objectively as possible?" When thinking of excellence, it may be used as a comparative expression to denote superiority or simply to mean high quality.

Identification of excellence is an ex-ante assessment or an exposed evaluation of research performance. Such a broad concept is not directly measurable in a valid manner. Everybody in the field has had such experiences. Perhaps the older you are and the more experienced you are and when you have been searching for excellence, you may be able to say, "Yes, I recognized the potential."

There are numerous definitions of scientific prestige, elite scientists and hierarchies of reputation in the sociological literature, with their exact meaning depending on the school of thought. Someone can read hundreds of journals and articles on these subjects; however, most of these ideas belong to individual researchers of socio-cognitive collectives rather than institutional aggregates. At the level of entire research groups and institutions the conceptual and operational problems are further complicated given the diversity of research missions, capabilities, resources and facilities characterizing research organizations.

Open and fair applications of peer review evaluation may be difficult to achieve in today's world. However, as every country is talking about transparency, you have to be able to lay down your rules. It is only fair to the individuals or to the businesses. New developments in the field of quantitative studies of science offer methods to support peer review in order to keep it objective and transparent.

I have already touched at this point on excellence in institutions, on the evaluation of excellence in the individual fundamental researcher, excellence in business: they are all different. The question is: "Is it possible to define scientific excellence in a comparative institutional context? Can you measure quality-related features of research institutions?" Are these measurements suitable for cross-discipline benchmarking comparisons at an international level? It has been shown that the average citation impact scores of institutions are not adequate indicators of their performance in terms of producing highly cited research. To capture the intrinsic quality of a world-class scientist, one needs something more than papers in the international scientific literature. The truly accepted measure of scientific quality is the performance of a small and elite group of influential scientists who are considered outstanding and receive prestigious awards and positions on international committees e.g. the Nobel Prize. Citation data by definition reflect the track record and past performance of universities rather than current standing or future scientific potential.

In promoting excellence, Germany has come forward with some very good ideas. Beate Konze-Thomas from the Deutsche Forschungsgemeinschaft (DFG) gave insights into the German excellence initiatives and the selection processes. This is what it is about: we have to keep our eyes open and watch. But Austria isn't lagging behind, it is well to the fore. Austria has the advantage, with such a rich history of excellence. But like every country with a rich history, be it excellence or otherwise, beware of sitting back. Austria is forward-thinking through its funding and policy bodies set up to encourage research. The Austrian Council has published a strategic document – the "Strategy 2010" – where excellence is one of the important points commented upon. It has already been noted by other organizations, for example the FWF (Science Fund), which developed a concept for "Excellence – clusters" as presented by President Kratky. And we have programmes dealing with industrial aspects, the most important one being "COMET". The "Competence Centers for Excellent Technologies" is the new competence centre programme aiming for excellence at the interface between science and industry. The presentation by Michael Binder gave a summary of the concept behind and the selection process applied. Furthermore, the Council is working on an Excellence-strategy, which will offer a broad concept for all areas and sectors.

An international example was given by Alfred Radauer, demonstrating the pro and cons of the PART system applied in the US. Finally, some provocative ideas and thoughts were brought to us from Switzerland by Marcel Herbst on the question "Select or create excellence?"

Scientific quality is a necessary precondition to push back research frontiers and to open up new fields of knowledge Achieving and maintaining scientific excellence has always been crucial for leading researchers and scholars working at the frontiers of science. An ability to achieve excellence at the international level has been and remains the strategic goal of research establishments worldwide.

However, I don't believe that we can give you the answer right now, but I am hopeful that at the end of the day, and after a series of very good papers, we will have some ideas how we evaluate excellence in academic research, excellence in applied research and excellence in institutions. I think those are the three issues we have to keep in mind.

Author:

Prof. Dr. Dervilla Donnelly
c/o Austrian Council - Rat für Forschung und Technologieentwicklung
Pestalozzigasse 4, 1010 Wien
donnelly@stp.dias.ie, donnellydk@eircom.net

Marcel Herbst

Excellence: to pick or to foster?

Abstract

In this note, I shall annotate a number of observations—or theses — regarding the state of affairs of European higher education and Humboldtian universities. These theses were presented at a Forum in Vienna which ran under the title "Exzellenz auswählen" (December 15, 2006). The theses cover a very broad spectrum of concerns, and a full elaboration would require the space of a book. The annotations presented here are very short, but they might induce the reader to reassess higher education and higher education policies.

Excellence has become one of the key foci of higher education, at least in Europe. The concern with excellence is tied to the observation that excellence is rare but needed; it may also be tied to the perception that excellence is to be found elsewhere now—and that it was once an exclusive attribute of European universities.

There is no doubt that most European universities have lost their once dominant position in the world of learning [5]. It is unclear why this has happened, and it is generally thought that the catastrophic events in conjunction with the Nazi Regime and World War II are at least partially responsible for the decline of European science [21, 25, 28] and its slow recovery since. The question remains to be answered, however, why Europe was in a position to recover economically and as an industrial base of strength, even in today's globalized world, while it failed to regain its once noble position in higher education and research. The answer to this riddle may have to found within higher education itself.

Thesis 1: European universities failed to adapt properly to Mass Higher Education

It was Joseph Ben-David who suggested that the European — and in particular the German, Humboldtian — university was unable to adapt properly to a changing environment [3, Chapter 6]. The very forces which propelled the German university in the 19th century to prominence [27] acted in the 20th century as a veritable "strangling noose" for the development of science. Unfortunately, Ben-David's thesis did not fall onto fertile grounds. Researchers do and did see shortcomings of the European — or German — higher education system [10, 14], but Ben-David's challenge — that particular systemic aspects rooted in our higher education cultures ought to be held responsible for the state of affairs now — has generally not been taken up. The failure to embrace Ben-David's analysis may be responsible for many of the ill-fated searches for solutions now.

Thesis 2: US universities adapted far better

The situation in the US can almost be sketched as a complement of the European picture. The US research university came into being in the latter half of the 19th century, and it emulated the German university where many US scientists and professors had received their doctoral training [12, 24]. But the US university retained the college next to the newly formed graduate schools, and it retained the collegiate culture borrowed from Oxford and Cambridge. The US university could not rely on well educated Gymnasiasten (high-school students), and it chose to retain a strong focus on teaching. As a consequence, student-faculty ratios remained fairly stable over the past 100 years, and they did not deteriorate significantly in the face of mass higher education after World War II. The implication of this is that US research-oriented institutions have more faculty than corresponding European institutions, perhaps by a factor of three or more. At US research universities, staff-faculty ratios remained moderate even in engineering or the sciences [19], nurturing a collegiate, less hierarchical setup conducive to cross-disciplinary research. The irony of this is that a system created to foster teaching and learning has now shown to be superior as a research setting [17].

Thesis 3: US research universities are Humboldtian

Wilhelm von Humboldt's dictum of the unity of teaching and research has long been a topos of our higher education debate. While the German university model had a definite impact on other systems [33, 31], the concept of the Humboldtian university gained credence long after the time of Humboldt, and perhaps most prominently at the time when mass higher education became slowly a reality [32]. The longing for the Humboldtian university was accompanied by a gradual deterioration of the unity of teaching and research within the German university. But this unity did not fail everywhere: Gerhard Caspar, a past president of Stanford University and himself a native German, speculated that the US research university had perhaps better preserved Humboldt's ideals.

A number of aspects differentiate US from European universities, and one ought to study these differences in order to find cues for the apparent differences in quality and productivity. First, US higher education is highly diversified [35, 20]: only 96 institutions, roughly 2% of a total of closed to 4,400, are counted (by the Carnegie Foundation) as research universities. Second, US institutions are autonomous to a large degree [6], and many of them embrace what US chief justice Felix Frankfurter called the "fourth academic freedom", namely the freedom to select (within strictly defined rules and regulations) their own student body. Third, power is distributed in different ways: whereas in European universities power is concentrated on the government and the individual faculty member, US universities place power on the executive board and the collegiate culture: US and European institutions differ in their morphological setup [4]. And finally, US universities try to have a broader resource base [7], and teaching and research are funded separately [18]. All these factors together appear to be responsible why US universities were more successful in adapting to mass higher education and are more productive than their European counterparts.

Thesis 4: Excellence is not only a matter of money

It is clear that higher education requires resources, and it is clear that at least the top institutions have to be funded properly. But resources alone, in absolute numbers or as a percentage of GNP, cannot guarantee excellence. Other factors will have strong effects on the quality of institutions, and these factors will have to be studied in conjunction with the levels of resources provided.

I have indicated above that very few US institutions qualify as true research universities. This stands in contrast to European sentiments where almost every university aspires to be a research university. Instead of calling for more funding, or substantially more funding, without a good base of analysis, it might be more fruitful first to look at the distribution of funds provided. Funds that are distributed too thinly are likely not invested effectively. We need discussions on the demand for higher education services, on the mission of the various institutions, on diversity in higher education, and funding ought to reflect the outcome of these discussions. The result might be that fewer universities should assume the (expensive) role of research universities and more institutions should focus on (less expensive) quality teaching.

To institute changes within European higher education, top-down approaches are en vogue. In this mode, changes are decreed: they do not take place because of some adaptation to changing conditions [16]. Because of this predilection, Centres of Excellence are declared instead of discovered, and proponents of a new public management resort to planning techniques which are in opposition to their own political stance and which were thought to be abandoned long ago [18]. Centres of Excellence need not be decreed. The framework within which higher education operates will have to be changed. If we separate the funding for teaching and research, and if the funding of research is meritocratic, centres of excellence will naturally develop as part of an adaptation process.

Separate funding streams for teaching and research are a necessary — but by no means a sufficient — condition for a successful transformation of European higher education. At least three other elements will have to be provided in order to secure a viable development. First, in the case of research universities, the “fourth academic freedom” will have to be implemented. Only thus will the institutions be in a position to properly program their own resource allocations and to gradually improve quality and standing. Second, European institutions will have to focus on a collegiate culture and to completely abandon the concept of a chair system (Lehrstuhl) on which most institutions are still based [19]. Student-faculty ratios will have to be improved significantly, particularly within research universities, with the prospects for more faculty, smaller research teams, less hierarchy, more independent research, and much better career prospects for younger scholars and women. And finally, third, many independent research institutes (such as those of the Max Planck Society) ought to be reintegrated into universities, to stop the bleeding-out process of universities and to invigorate research institutes.

Thesis 5: Excellence is possible on all levels, in all institutions

There is a tendency to associate excellence with research alone, or with output. Excellence is clearly not restricted to research, and it has as much to do with performance or performance improvement as it has with output. Excellence is not confined to elite institutions. We can find it in various endeavors and environments. A range of scholars have pointed out this fact, among them Alexander W. Astin [1, 2]. Astin drew attention to the fact that excellence in education refers foremost to a process, not a static (test) result, where student and school interact. To foster weak learners may be as challenging — and rewarding — as the stimulation of the talented. Societies need educational excellence at all levels of schooling, not only in universities, and not only in so-called 'elite' institutions. Furthermore, modern societies require a balance among their educational strata in order to reap the fruits made available by scientific insights and progress. Otherwise, inventions can be made but not brought to market, and agglomeration economies, so necessary for the well-being of our respective societies, cannot be exploited [26, 30, 29, 13].

Thesis 6: Excellence develops slowly

If we look at the cream of research universities today, we notice that they are generally quite old: they definitely did not set out to become world-class universities, and they developed slowly. There are practically no prominent universities which are less than 100 years old. Newly founded universities — like Stony Brook University (1957) in New York which was to take on the role of the Berkeley of the East, or the University of Waterloo in Canada (1957) — have become good universities, but they are not yet part of the well established set of elite institutions. Israeli institutions — like the Hebrew University (1925), the Weizmann Institute (1934), the Technion (1924)—do astonishingly well, but they are not that young either, and they were founded to serve a vital function. In the foreseeable future some Chinese universities might surprise us, but they are not yet there.

Universities can be destroyed quickly, as we are bound to notice, but it takes decades of continuous and dedicated improvement activities to serve their cause and to bring them to fruition. One cannot count on quick and easy payoffs: concerted and enduring efforts are necessary if one intends to advance the standing and the impact of an institution, or if one intends to found new institutions [9].

Thesis 7: Excellence ought to be fostered, not proclaimed

The current public debate on — or better even: public obsession with — excellence and elite universities is new. Ten years ago, these themes (according to LexisNexis) were barely visible in the (German) press. In the context of today's Initiative for Excellence in Germany, for instance, special research oriented schools (so called Graduiertenkollegs), excellence-clusters or university-wide projects are earmarked to receive special funding [11]. "Visible research beacons" are sought — and proclaimed — that are to shine over the sea of science. For the first two tasks combined Euro 7.5 million per annum are to be distributed on average for each institution which was selected to be funded, and for the third task (benefiting the Universität

Karlsruhe, the Ludwig-Maximilians-Universität München and the Technische Universität München) Euro 21 million per annum are to be distributed.

As I have said above (under Thesis 4), "excellence is not only a matter of money". Will these funds make a lasting difference? Will they help to promote lighthouses which shine over the ocean of science, as the text of the initiative alludes to metaphorically? This is unlikely the case. If the "internationally visible research beacons" are already there, the funding might help them to survive; if they are not yet there, they will not emerge because of the funding. Modern research universities of distinction run on a yearly budget of Euro 0.5 to 1.5 billion or more (without capital outlays). Funding in the amount envisaged today in Germany will not make a difference.

What is required, instead, is a new delineation of the framework within which higher education can operate, a "re-engineering" of the system's boundaries [15], and a reallocation of funds [18]. This can be done gradually, gaining experience and insight along the way. Research universities should not be imposed: they should emerge. They require neighbours and peers; they thrive on competition; they need to be allowed to be entrepreneurial [7, 8]. It will be impossible to find the moneys required to elevate — top-down, so to speak — the many European higher education institutions to the level of research universities, but if we start a sustainable reform along the lines already sketched, true research universities might gradually emerge.

Thesis 8: Excellence might best be promoted bottom-up

As I have observed above, the current national and European Initiatives for Excellence operate mainly in a top-down fashion. Institutions will benefit, not researchers. In this fashion, Initiatives for Excellence will complement institutional base funding as well as competitive— meritocratic— research funding. It is not clear why such an approach, which also forms the foundation of the Research Assessment Exercise in the UK and which has been criticized — among others — by Martin Trow [34], ought to be preferred.

There exists a successful culture of a separation of funding streams: base funding of institutions to sustain teaching, and meritocratic funding of research activities of individual scholars or research teams. Quality assessment exercises — self-assessment and peer review — have long been used to improve teaching and research activities, but these are only effective if they serve the needs — and find the trust — of those assessed [22, 23]. There is enough evidence that professors are generally conservative and reluctant to change: hence the urge to push and shove the faculty. But it appears unwise to forgo good institutional management—or the prospect thereof — and replace that by funding and decision modes which are far removed from the institutions which are to benefit.

Institutions should not be steered from afar, and governance should better recognize its fiduciary role and should not engage in management (and particularly micro-management). Institutions require a proper framework within which they can operate and proser, and if governors, trustees or elected officials are convinced of deficiencies or shortcomings of institutional managers, they should replace these executives, not interfere with their management. At the same time, institutions — and the science enterprise in general—will have to learn to be more introspective and to exploit the insights of institutional research or the sociology and economics of science.

Thesis 9: The reforms themselves require evaluation, and they are possible

A range of features of today's higher education systems are the results of previous reforms, and not all of these reforms appear well advised. Reform measures have to be tested and evaluated locally, and on an individual basis, before they are to be implemented on a grander — national or international—scale. Furthermore, a broad range of reform projects ought to be evaluated.

Among the wide spectrum of possible measures to reform higher education only relative few measures are contemplated and fewer still are tested or implemented. Reforms pertain to two levels: (i) the macro — or inter-institutional — level, and (ii) the micro — or intra-institutional — level. With regard to the macro-level, the following reform measures appear to require closer scrutiny:

1. separation of funding streams for teaching and research;
2. reintegration of dedicated research institutions — such as the Max Planck or Fraunhofer Institutes in Germany or the institutes of the 'Centre nationale de la recherche scientifique' in France — into research universities;
3. installation of the "fourth academic freedom", i.e. the right to select their student bodies, for research universities.

Regarding the micro-level, the following reform measures ought to be contemplated:

4. abolishment of (most) academic staff positions past the post-doctorate that do not enjoy full academic freedom and the eventual abolishment of corresponding academic degrees (e.g., in the German context, no Privatdozentur and no Habilitation);
5. improvements in the faculty-student ratio — and implicitly also the faculty-staff ratio (more faculty, i.e. professorial positions; more and smaller research groups);
6. the promotion of active learning measures and attempts to reduce drop-out rates among students.

Thesis 10: Successful change requires communality, and only common positions are stable positions

All reforms require a perception — and an acknowledgement — of the interests of the various actors or stakeholders of higher education systems. These interests are naturally diverse, not only because individuals have different views and notions, but because views are normally tied to various roles people fulfill in our respective societies. Politicians, regents or higher education executives view higher education and university from other angles than faculty, and they in turn assume other viewpoints than staff or students; parents, industry leaders, employers, citizens and tax payers have yet their own views regarding the role and mission of higher education or technology transfer. In order to move higher education, and in order to change institutions, common positions will have to be sought. A common position, a consensus, cannot be reached without a discourse, an ongoing discussion, and it will also require scholarly immersion as well as experimentation with the unknown — or the unfamiliar.

Author:

Marcel Herbst
4mation
Erligatterweg 65, CH-8038 Zürich, Switzerland
herbst@mat.ch

References

[1] *Alexander W. Astin.* Achieving Educational Excellence: A Critical Assessment of Priorities and Practices in Higher Education. Jossey-Bass Publishers, 1985.

[2] *Alexander W. Astin.* What Matters in College? Four Critical Years Revisited. Jossey-Bass Publishers, 1993.

[3] *Joseph Ben-David.* Scientific Growth: Essays on the Social Organization and Ethos of Science. University of California Press, 1991.

[4] *Pierre Bourdieu.* Homo Academicus. Stanford University Press (Les Editions de Minuit), 1988 (1984).

[5] *CEST.* Universities and Colleges Participating in the 'Champions League': Rankings by four Performance Indicators (1998–2002). Centre d'Etudes de la Science et de la Technologie (CEST), 2004.

[6] *Burton R. Clark.* Diversification, Competitive Autonomy, and Institutional Initiative in Higher Education Systems, chapter 4, pages 37–41. In Herbst et al. [20], 1997.

[7] *Burton R. Clark.* Creating Entrepreneurial Universities: Organizational Pathways of Transformation. Pergamon, 1998.

[8] *Burton R. Clark.* Delineating the Character of the Entrepreneurial University. Higher Education Policy, 17:355–370, 2004.

[9] *Daniel Clery.* Despite a Chilly Reception, the 'European MIT' Advances. Science, 311(5765):1227, March 3 2006.

[10] *Michael Daxner.* Ist die Uni noch zu retten? Zehn Vorschläge und eine Vision. Rowohlt, 1996.

[11] *Deutsche Forschungs-Gesellschaft and Wissenschaftsrat.* Exzellenzinitiative des Bundes und der Länder: Ergebnis der Sitzung des Bewilligungsausschusses am 13. Oktober 2006. www.bmbf.de/pub/Exzellenzinitiative.pdf, 2006.

[12] *Abraham Flexner.* Universities: American, English, German, pages xxxiii–381. In Kerr [24], 1930.

[13] *Masahisa Fujita and Jacques-François Thisse.* Economics of Agglomeration: Cities, Industrial Location, and Regional Growth. Cambridge University Press, 2002.

[14] *Peter Glotz.* Im Kern verrottet? Fünf vor zwölf an Deutschlands Universitäten. Deutsche Verlags-Anstalt, 1996.

[15] *Michael Hammer and James Champy.* Reengineering the Corporation: A Manifesto for Business Revolution. Harper Business, 1993.

[16] *Marcel Herbst.* Change Management: A Classification. Tertiary Education and Management, 5:125–139, 1999.

[17] *Marcel Herbst.* The Production-Morphology Nexus of Research Universities: The Atlantic Split. Higher Education Policy, 17(1):5–21, March 2004.

[18] *Marcel Herbst.* Financing Public Universities: The Case of Performance Funding. Springer Science + Business Media, 2007.

[19] *Marcel Herbst, Urs Hugentobler, and Lydia Snover.* MIT and ETH Zürich: Structures and Cultures Juxtaposed. Centre d'études de la science et de la technologie (CEST 2002/9, www.cest.ch/en/publikationen/cest_reihe.htm), October 2002.

[20] *Marcel Herbst, Günther Latzel, and Leonard Lutz, editors.* Wandel im tertiären Bildungssektor: Zur Position der Schweiz im internationalen Vergleich. Verlag der Fachvereine (vdf), 1997.

[21] *David A. Hollinger.* Science, Jews, and Secular Culture: Studies in Mid-Twentieth-Century American Intellectual History. Princeton University Press, 1996.

[22] *Herbert R. Kells.* Self-Regulation in Higher Education: A Multi-National Perspective on Collaborative Systems of Quality Assurance and Control. Jessica Kingsley, 1992.

[23] *Herbert R. Kells.* Self-Study Processes: A Guide to Self-Evaluation in Higher Education. Series on Higher Education. American Council on Education and Oryx Press, 4th edition, 1995.

[24] *Clark Kerr, editor. Abraham Flexner:* Universities, American, English, German. Transaction Publishers (Oxford University Press), 1994.

[25] *Claus-Dieter Krohn.* Wissenschaft im Exil (Intellectuals in Exile). Campus-Verlag (University of Massachusetts Press), 1987 (1993).

[26] *David Lampe, editor.* The Massachusetts Miracle: High Technology and Economic Revitalization. The MIT Press, 1988.

[27] *Frank R. Lillie, William Trelease, Henry H. Donaldson, William H. Howell, and James Rowland Angell.* Cooperation in Biological Research. Science, XXV:369–386, 1908.

[28] *Jean Medawar and David Pike.* Hitler's Gift: Scientists Who Fled Nazi Germany. Richard Cohen Books and European Jewish Publication Society, 2000.

[29] *Ed Moscovitch, Richard DeKaser, and Paula Fitzgibbon.* MIT: The Impact of Innovation. BankBoston, March 1997.

[30] *Annalee Saxenian.* Regional Advantage: Culture and Competition in Silicon Valley and Route 128. Harvard University Press, 1994.

[31] *Marc Schalenberg.* Humboldt auf Reisen? Die Rezeption des 'deutschen Universitätsmodells' in den französischen und britischen Reformdiskursen (1810-1870). Schwabe & Co, 2002.

[32] *Helmut Schelsky.* Einsamkeit und Freiheit: Idee und Gestalt der deutschen Universität und ihrer Reformen. Rowohlt, 1963.

[33] *Rainer Christoph Schwinges, editor.* Humboldt International: Der Export des deutschen Universiätsmodells im 19. und 20. Jahrhundert, volume 3 of Veröffentlichungen der Gesellschaft für Universitäts- und Wissenschaftsgeschichte. Schwabe & Co., 2001.

[34] *Martin Trow.* Trust, Markets and Accountability in Higher Education: A Comparative Perspective. Higher Education Policy, 9(4):309–324, 1996.

[35] *Martin Trow.* Reflections on Diversity in Higher Education, chapter 3, pages 15–36. In Herbst et al. [20], 1997.

Thomas M. Pelsoci

Measuring Societal Benefits of R&D: Case Study Performance Metrics

This paper describes the use of detailed case studies that build on financial analysis to evaluate program impacts. A benefit-cost analysis traces a government investment from R&D to public benefits. Performance metrics are calculated by multiplying probable cost savings by conservatively projected unit sales over a period of ten years.

An illustrative case study examines the introduction of an advanced new sensor-on-a-chip technology to detect carbon dioxide (CO_2) which is an important vital sign in emergency medicine. Based on comparative medical studies of new and current (defender) technologies for CO_2 detection, medical use of the new sensors is expected to result in more than 112,000 prevented in-ambulance patient deaths over the next ten years and reduced costs of medical care. In addition, another application of the new technology is expected to reduce costs of heating and cooling commercial buildings by 10-20 percent.

Over the next ten years the use of this government funded new technology is projected to result in net cost savings of $143-175 million in present value terms. Expressed in another way, benefits will range from $174 to $212 for every government dollar invested.

The new CO_2 detection technology was funded on a cost-share basis by the Advanced Technology Program (ATP) of the National Institute of Standards and Technology (NIST). The mission of the National Institute of Science and Technology (NIST) dates back to the U.S. Constitution, which assigned oversight for weights and measures to the Federal government. Today, NIST is highly visible in numerous roles, ranging from its precision time signals using radio, Internet and telephone links and for its recent investigations of the collapse of the World Trade Center Towers in New York City and the reliability of electronic voting machines.

NIST carries out its mission in four cooperative programs, including the Advanced Technology Program (ATP), which accelerates the development of innovative technologies for broad national benefit by co-funding R&D partnerships with the private sector.

ATP uses public-private partnerships to improve the quality of life of Americans in the 21st century by investing in early-stage, innovative technologies. Private-sector investments in research and development are focused mainly on incremental improvements to capture faster returns to companies and their shareholders. Few financial institutions, venture capitalists, and angel investors are eager to fund unproven, early-stage technologies. This is where ATP steps in by helping industry invest in longer-term, high-risk research seeking payoffs far beyond private profit.

Drawing upon a portion of a larger research study, this paper presents one example of the experience of NIST and ATP in using detailed case studies that build on a financial analysis foundation to evaluate program impacts and benefits to society using benefit-cost analysis (Pelsoci, 2005).

Performance Evaluation

ATP uses a battery of analytical tools to measure program impact, including surveys and statistical analysis that measure short-term outputs and outcomes, in-depth, benefit-cost case studies that measure longer-term outcomes and impacts, and more. The emphasis is on "What difference did ATP make?" This paper illustrates an example of a NIST economic impact study in the form of a benefit-cost analysis. In addition to sponsoring such studies, ATP seeks to identify methodological steps that can be taken to facilitate consistency and comparability across such studies and aggregation of results of studies performed at different times (Powell, 2006).

Public and Private Benefits

In evaluating the performance of publicly funded science and technology projects, the benefits of innovation can be seen to flow to two classes of beneficiaries.

Economic benefits received by the innovating firms, funded on a cost-share basis with the ATP, are considered **private benefits**. The innovating firm's that these profits from innovation will be realized is a necessary precondition for completing technical development tasks and for undertaking subsequent commercialization.

In contrast, economic and other societal benefits arising from the ATP-funded technologies that are enjoyed by downstream industrial firms and end users of industrial products are considered **public benefits**. Public benefits represent a "spill over" effect, where the degree of "spill over" is that portion of total benefits that the innovating firm is unable to capture for itself. In economic theory and actual practice, public benefits, or the spill over from ATP investment, may be substantial (Mansfield et al., 1977).

A large spill over effect indicates that the innovating firm is unable to appropriate a significant share of technology benefits in the form of profits. As a result, the value of the project to society is higher than its value to innovators and may lead to private sector under-investment and associated loss of benefits to downstream firms, end-user customers, the economy, and society at large (Jaffe, 1996).

ATP compensates for the spill over gap by partially funding the development of high-risk, innovative technologies that private firms are unable to fund due to R&D technical risks and appropriability risks.

In addition to economic benefits, many other public benefits have a "public goods" aspect that society-at-large can enjoy directly, such as:

Reduced harmful environmental emissions

Improved outcomes for medical patients

Conservation of scarce energy resources and

Knowledge diffusion about new technologies.

Retrospective vs. Prospective Benefits

The placement of project benefits within a time frame is another important variable for benefit analysis. ATP uses both retrospective and prospective benefit analysis.

Retrospective analysis: Realized benefits (to date), identified on the basis of well-documented analysis, involve less uncertainty and can lead to higher levels of confidence in the analytical results of impact studies than prospective (future) benefit estimates.

Prospective analysis: Expected future benefits are typically associated with a variety of risks and uncertainties, including market introduction, consumer acceptance and manufacturing ramp-up. While these risks and uncertainties may be mitigated by various means, including interviews with potential customers, prototype sales and market studies, the inherent uncertainties of prospective analysis require the use of expected value constructs, where first-order benefit estimates are scaled back with consensus-based probability estimates.

Some ATP studies, including this case study of a new CO_2 detector technology, are entirely prospective even if conducted years after the ATP project has been successfully completed. Other studies are a combination of prospective and retrospective analysis. Risky, breakthrough technologies, such as ATP seeks to fund, can take more than 20 years to mature.

Attribution of Benefit

Benefit-cost analysis also considers the contributions of other sources of funding to ultimate benefits. As examples, ATP-funded high-risk technology projects may be preceded by basic and applied research projects funded through the NSF or other U.S. agencies. In addition, some projects receive funding from other agencies subsequent to the ATP funding. The fair attribution of project benefits to ATP will depend on specific circumstances of each case, reflecting the relative importance as well as the size and timing of ATP and other investments.

Performance Metrics

Our case study approach to measuring the societal benefits of ATP's projects utilizes performance metrics that capture the value of public benefits attributable specifically to ATP's investment. Benefit cash flow estimates are compared to investment costs to compute three public benefit performance metrics: benefit-to-cost ratios, net present values, and internal rates of return. Metrics are conservatively estimated and represent expected lower bounds.

Benefit-to-cost ratio (B:C): is computed by dividing the present value of "public benefits attributable to ATP" (gains received by U.S. beneficiaries other than the ATP-funded innovator) by the present value of ATP's investment. This measure estimates the benefit to the nation for every dollar of ATP investment.

Net present value (NPV): is calculated by subtracting the present value of ATP's investment from the present value of public benefits attributable to ATP from new technologies. Cash flows are normalized to 2004 dollars and discounted at the 7% rate designated by the U.S. Office of Management and Budget.

Internal rate of return (IRR) is calculated by iterative solution for a rate at which the discounted value of ATP's investment equals the discounted value of benefit cash flows attributable to ATP, thus indicating the rate of return to the nation on ATP's investment.

Case Study

In 1999, Ion Optics, Inc. of Waltham, MA, a start-up technology company, received ATP cost-share funding to develop photonic crystal sensors that can be used to accurately, reliably, and inexpensively measure carbon dioxide (CO_2) levels (the first target gas for which this technology is commercially viable) in the expired breath of emergency room patients and in commercial office buildings. Private funding for direct costs was not available for this early-stage high-risk project to miniaturize the functionalities of non-dispersive infrared (IR) absorption spectroscopy (NDIR) onto an integrated sensor chip, using microelectromechanical systems (MEMS) fabrication technologies. In 2001, the project was successfully completed. Since then, Ion Optics has been pursuing further development and marketing.

There are two initial applications for the new CO_2 detection technology:

Emergency medicine – detecting the faulty placement of breathing tubes in ambulance patients, an error that can result in death or injury of patients

Optimizing the efficiency of commercial buildings' heating, ventilation and cooling systems

Commercial production of the photonic crystal CO_2 sensor, called a sensor-on-a chip, was targeted for 2006 with annual sales expected to ramp up to 400,000 units by 2015 for emergency medicine applications. The chip sensor could prevent the deaths of ambulance patients by detecting a common fatal error, erroneous placement of an emergency breathing tube. The existing "defender" technology used to detect faulty breathing tube placement in ambulances, colorimetric detectors, uses chemically treated paper that changes color when exposed to CO_2. It is inexpensive, has poor reliability, limited shelf life, and does not provide a quantitative measurement of CO_2 levels. A recent medical study reported that "upon arrival in emergency rooms, up to 25% of patients were found to have improperly placed" breathing tubes and faced a 56% fatality rate.

Research and medical studies indicated that the sensor technology had potential for a dramatic reduction in the number tube placement errors and, as a result, a reduction in the number of deaths of emergency patients in ambulances, emergency rooms and other emergency medicine settings. Based on research of deaths resulting from the use of the current defender technology, medical use of photonic crystal CO_2 sensors is expected to result in more than 112,000 prevented in-ambulance deaths of trauma victims and critically ill patients on their way to U.S. emergency rooms over the 2006 to 2015 period, as shown in Table 1.

Table 1. Reduced In-Ambulance Mortality (Prevented Deaths)

	Projected unit sales of CO_2 sensors for emergency medicine capnometry	**Estimated prevention of in-ambulance deaths from timely detection of misintubation**
2006	10,000	950
2007	40,000	3,800
2008	90,000	8,550
2009	130,000	12,350
2010	200,000	19,000
2011	300,000	28,500
2012	400,000	38,000
2013	400,000	38,000
2014	400,000	38,000
2015	400,000	38,000
Total estimated prevented in-transit deaths (2006–2015)		*225,150*
Prevented death attribution to ATP		*112,575*

The 44% who survive an improperly placed breathing tube require additional costly medical treatment and hospital stays.

Example: Estimating Cash Flow Benefits

This is how cash flow benefits from the emergency medicine application of the CO_2 sensor for one year of the study period, 2007, were calculated.

Market analysis and interviews established the following:

Projected unit sales CO_2 sensors in emergency medicine are 40,000 sensors in 2007

The probability of realizing sales projections is 65%

Each sale corresponds to one intubation, the insertion of a medically required oxygen breathing tube into the trachea of a patient in transit to the emergency room

Based on medical journal published statistics, approximately 18% of intubations would fail, with the failure undetected, if colorimetric paper (the current error-detection technology) had been used rather than the CO_2 sensors. In 56% of these cases (9.5 intubations per 100 performed) the patient would have died; in 44% (7.45 intubations per 100 performed), surviving patients would have incurred additional costs for hospital stays and treatment because of the erroneous intubation.

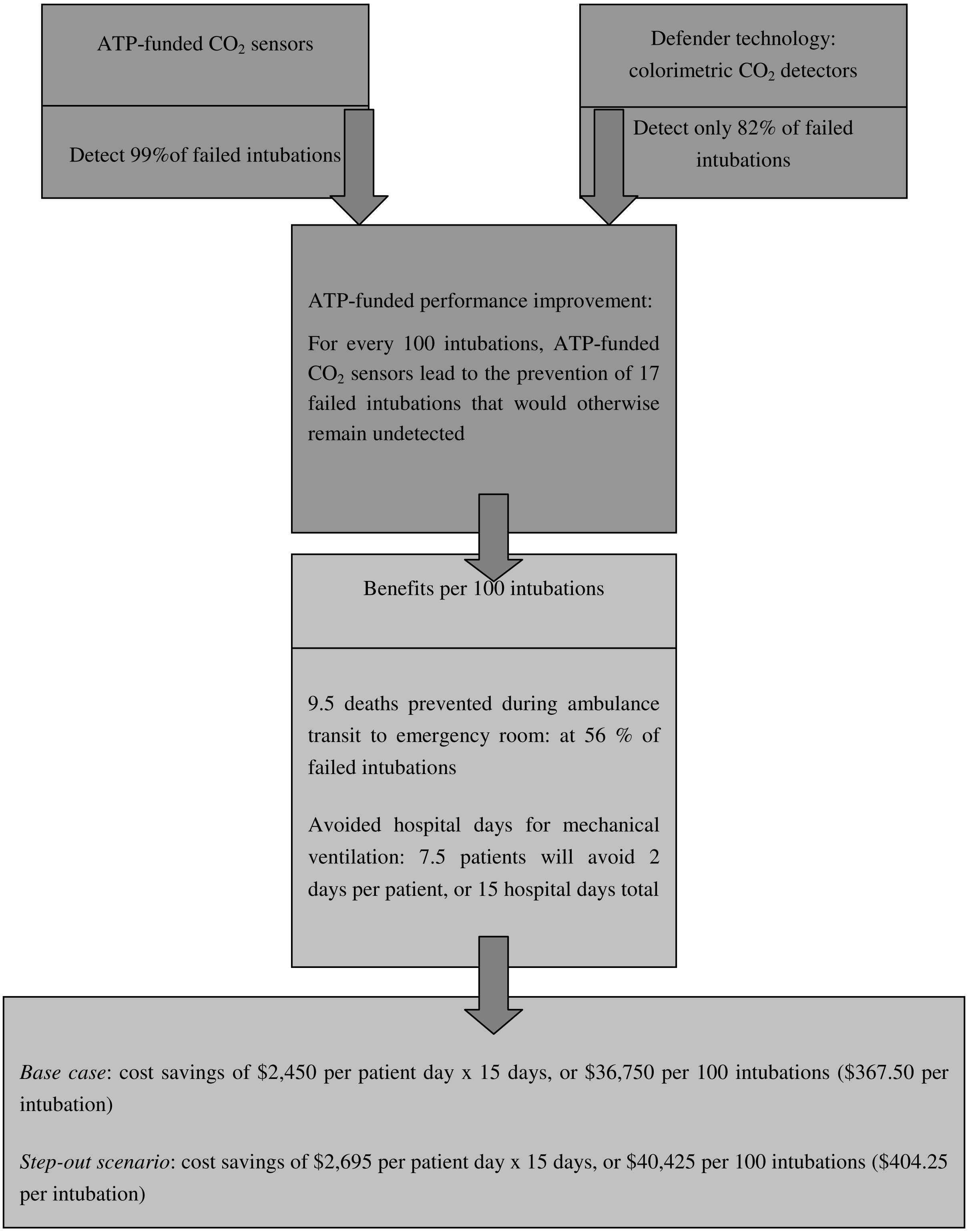
Only a 1% error-detection failure rate is anticipated for the CO_2 sensor, compared to an 18% failure rate for the older technology

Current technology sales losses, the lost sales of the current chemical sensor detectors, are $1.08 per unit (sales revenue of $13 per unit times average U.S. industry profit margin of 8.3%)

Reflecting the relative importance and similar size and timing of funding from ATP and the National Science Foundation, 50% of net benefit cash flows are attributed to NSF and 50% to ATP.

Mortality reductions, expected to number 3,800 in 2007 and over 112,000 cumulatively by 2015, are a key benefit but are not quantified in dollar terms. Avoided costs of hospital stays and treatment, amounting on average to $367.50 for every use of the CO_2 sensor instead of the colorimetric paper detectors, are included in estimated benefit cash flows. The base-case cost savings are $367.50 per intubation and the more optimistic step-out scenario cost savings are $404.25 per intubation, as shown in Figure 1.

Figure 1. Projected Benefits Per 100 Intubations From Improved Detection of Failed Intubations, Using ATP-Funded CO_2 Sensors In Lieu of Colorimetric CO_2 Detectors



The calculation of estimated benefit cash flows from this application for a single year, 2007, is as follows:

Multiplying $367.50 benefit per intubation by 40,000 projected intubations in 2007 will result in total estimated economic benefits of $14,700,000, with an expected value of $9,555,000 after adjusting for the 65% probability.

Subtracting defender technology lost profits of $28,054 yields net benefits of $9,526,946.

Fifty percent, or **$4,763,473** of benefits is attributed to the ATP.

This process is applied for each year and each application.

Following the same methodology as for the emergency medical application, annual cash-flow benefits from another CO_2 sensor application – optimizing commercial building heating, ventilation and air conditioning (HVAC) systems – were similarly calculated for each year. The U.S. Environmental Protection Agency estimates that the use of accurate and cost-effective CO_2 sensors, in combination with variable air flow systems, can reduce office space energy consumption by 10-20 percent. Following the methodology shown for emergency medicine, a similar step-by-step approach was used for the second CO_2 application. Base case cash flows were calculated for both technology applications and inserted into Table 2.

ATP's investments in the photonic crystal technology represent negative cash flows. During the 1999–2001 period, ATP invested $753,000 toward project direct costs. For purposes of cash flow analysis and computation of performance metrics, the ATP investment was normalized to 2004 dollars (as $827,000) and included as one lump sum investment in 2000, the midpoint of the ATP investment period. As its cost share, its industry partner, Ion Optics, invested $626,000 in indirect costs. Several phases of prior and concurrent NSF grants ($850,000) supported prior research and were matched with $500,000 from Ion Optics. The Ion Optics and other agency investments are not included in the performance metrics of this study, which feature the public return on ATP's investment, but they would appropriately be included in a broader analysis of total public and private benefits and costs of the project.

BASE-CASE ANALYSIS: As shown in Table 2, the public return on ATP's investment in the CO_2 sensor technology over the 1999–2015 period can be expressed as a net present value of $143 million. Public benefits attributable to ATP are $174 for every dollar invested and the internal rate of return is estimated at 75%.

Table 2. Base-Case Cash Flows and Performance Metrics for ATP-Funded CO_2 Sensors (2004 Dollars, in Millions)

	Emergency medicine	Building HVAC controls	Total cash flows
	(1)	(2)	(3)
2000			-0.827
2001			
2002			
2003			
2004			
2005			
2006	1.191	0.037	1.228
2007	4.763	1.765	6.529
2008	10.718	3.531	14.249
2009	15.481	5.296	20.778
2010	23.817	7.062	30.879
2011	35.726	8.827	44.553
2012	47.635	9.269	56.903
2013	47.635	9.747	57.381
2014	47.635	10.225	57.860
2015	47.635	10.740	58.374

Net present value	143 million
Benefit-to-cost ratio	174:1
Internal rate of return	75%

Note: A 2000 base year and an OMB-designated 7 % discount rate were used for analysis. Performance metrics were computed from time series assuming ATP investment in 2000 (project midpoint) and prospective cash flow benefits from 2006 to 2015. Positive cash flows represent public benefits attributable to ATP; negative cash flows represent ATP investment costs (assumed to occur at project midpoint).

STEP-OUT ANALYSIS: A step-out scenario analysis investigated the sensitivity of base-case performance metrics to more optimistic assumptions about the projected future benefits of each CO_2 sensor in emergency medicine and office building air quality applications, in combination with a 10 percent increase in projected sales.

As shown in Table 3, the public return on ATP's investment in the CO_2 sensor technology is associated with a net present value of $175 million. Public benefits attributable to ATP are $212 for every dollar invested and the internal rate of return is estimated at 79 percent.

Table 3. Step-Out Scenario Cash Flows and Performance Metrics for ATP-Funded CO_2 Sensors (2004 Dollars, in Millions)

	Emergency medicine	**Building HVAC controls**	**Total cash flows**
	(1)	(2)	(3)
2000			-0.827
2001			
2002			
2003			
2004			
2005			
2006	1.440	0.046	1.487
2007	5.762	2.218	7.979
2008	12.964	4.435	17.399
2009	18.726	6.653	25.379
2010	28.809	8.870	37.679
2011	43.213	11.088	54.301
2012	57.618	11.642	69.260
2013	57.618	12.243	69.861
2014	57.618	12.844	70.461
2015	57.618	13.490	71.108

Net present value	175 million
Benefit-to-cost ratio	212:1
Internal rate of return	79%

Note: A 2000 base year and an OMB-designated 7 % discount rate were used for analysis. Performance metrics were computed from time series assuming ATP investment in 2000 (project midpoint) and prospective cash flow benefits from 2006 to 2015. Positive cash flows represent public benefits attributable to ATP; negative cash flows represent ATP investment costs (assumed to occur at project midpoint).

Summary

Performance metrics presented here demonstrate exceptional results from the original investment in the high-risk project to develop photonics crystals for CO_2 detection. The process systematically researches and collects data, analyzes technical information from various sources and includes commercialization assessments and market analyses. Because this methodology relies on documented scientific sources, along with experts' and manufacturers' input and conservative market projections, the modeling of these future benefits produces compelling and meaningful results.

Businessmen, investors and policymakers can readily follow the logical progression from input to calculation of dramatic performance results – total savings ranging from $174 to $212 for every dollar invested. But perhaps even more striking than cost savings is the result in human deaths prevented in ambulance patients – 112,000 lives over the 2006 to 2015 period.

References and Further Reading

Jaffe, Adam B. 1996. Economic Analysis of Research Spillovers: Implications for the Advanced Technology Program. NIST GCR 97-708. Gaithersburg, MD: National Institute of Standards and Technology http://www.atp.nist.gov/eao/gcr708.htm

Mansfield, Edwin et al. 1977. Social and Private Rates of Return from Industrial Innovation. Quarterly Journal of Economics. 91:221-240.

Pelsoci, Thomas M. 2005. Photonics Technologies: Applications in Petroleum Refining, Building Controls, Emergency Medicine, and Industrial Materials Analysis. NIST GCR 05-879 (Chapter 4), Gaithersburg, MD: National Institute of Standards and Technology http://www.atp.nist.gov/eao/gcr05-879/gcr05-879.pdf

Powell, Jeanne. 2006. Toward a Standard Benefit-Cost Methodology for Publicly Funded Science and Technology Programs. NISTIR 7319. Gaithersburg, MD: National Institute of Standards and Technology http://www.atp.nist.gov/eao/ir-7319/ir-7319final.pdf

About ATP: http://www.atp.nist.gov

NIST's Economic Assessment Office: http://www.atp.nist.gov/eao/eao_main.htm

About the Author

Dr. Thomas Pelsoci is managing director of Delta Research Co. He specializes in the economic assessment of new technologies and manufacturing processes, including prospective economic impact studies during early stage proof of concept and later state demonstration phases. Prior to industrial experience with TRW and management consulting with BearingPoint (KPMG Peat Marwick Management Consultants), Dr. Pelsoci received a degree in Mechanical Engineering from Case Western Reserve University and a Ph.D. in Public Policy and Administration from the University of Minnesota.

Dr. Thomas Pelsoci
Delta Research srl
Via S. Felice 25, 24060 Gorlago, Italy
tpelsoci@deltaresearchco.com

Christoph Mandl

Evaluating Evaluations or The Case for Action Research

In 2004 a panel of 13 people from 13 countries within the European Union was appointed for the Five-Year Assessment of the European Union Research Framework Programmes 1999-2003[1]. I was one of those panel members. This evaluation was in many ways unusually successful not least because the European Commission took its findings and its recommendations very seriously[2]. The Ormala-Report, as it is now being called after the chairman of the panel, was quoted extensively by those dealing with European research policies. It is also one of the prime sources of the European Court of Auditors for its audit of the adequacy and effectiveness of selected FP6 instruments in the achievement of Community RTD objectives.

Having had the privilege of being on quite a number of evaluation panels at regional, at national as well as at European level[3] which have not achieved a similar leverage I wondered what the important ingredients of this specific evaluation were that made it outstanding from the lot? I discussed this question with some of my fellow members of the Five-Year Assessment panel as well as with the head of Unit A.4 - Planning, Programming, Evaluation of Directorate-General for Research who was the prime representative of the European Commission in the Five-Year Assessment. Although these discussions have contributed to my thinking on this question the responsibility for the answers given in this article is solely mine.

What are the differences among evaluations so that some achieve high leverage while others do not? This is the question I am concerned with in this article but not for the sake of understanding. This question has rather practical implications. If one understands these differences one can create evaluations with high leverage deliberately rather than inadvertently as was the case with the Five-Year Assessment. The answer to this question is, however, not straightforward and requires to delve into scientific territories not common to evaluation experts, i.e. social system theory, action research, policy resistance and defensive routines.

At the first meeting of the Five-Year Assessment panel a question emerged immediately. Should members of the European Commission be allowed to participate in all phases of the meetings of the panel or should they be excluded from those phases where the panel members openly share their observations, their findings, their opinions and their conclusions? This led to an interesting discussion revealing the different assumptions people held about evaluation methods, about reaching objectivity in the presence of members of the programmes being evaluated, and about achieving leverage. This discussion was undertaken in the presence of those particular members of the EC we were talking about. I believe that both, the result of this discussion - it was unanimously agreed that members of the EC were invited to join the panel and listen to our discussion - as well as the friendliness of this discussion created a mode of collaboration rather than a mode of examination which I happened to have experienced in other panels. Even though this short discussion and its result may be seen as insignificant from the outskirts there were important undercurrents in it that deserve detailed scrutiny.

What is the purpose of an evaluation?

Evaluators do not share a common understanding of the purpose of an evaluation. With extensive research in the literature and the web I identified three distinct definitions of evaluation. According to the literature an evaluation is[4]

(a) an assessment at a point in time of the value, worth or impact of a project or program.

(b) the assessment of how well a project/activity achieved its objectives.

(c) any effort to use assessment evidence to improve institutional, departmental, divisional or unit effectiveness.

Definition (b) requires some objectives against which an activity is being assessed while (a) does not require them. But the more important difference lies between (c) and the two others because (c) aims at improvement while (a) and (b) aim at assessment.

The great majority of definitions are variations of (a) and (b). However, the present practice of evaluations clearly points towards (c). Nearly all evaluations on the European level include in their published reports recommendations for improvement[5]. It is only in view of definition (c) that the core question of this article bears meaning.

When the panel of the Five-Year Assessment discussed the issue described earlier its members were aware that they were expected to come up with valid recommendations for improvement. In fact, this specific task made it particularly obvious that only a collaborative mode with the European Commission would meet such expectation.

The Paradigm Shift

If one sees evaluation as a process to assess *and* to improve the performance of institutions or programmes then a paradigmatic issue comes up that makes all the difference. This difference can be condensed into one fundamental methodological assumption:

It is possible to assess a social system without profoundly influencing it.

In social sciences there are two methodological schools. One school - the empirical social researchers - hold this assumption to be true while the other school - the action researchers - hold this assumption to be false.

Paradigm shift is the term first used by Thomas Kuhn[6]. It describes a change in basic assumptions within the ruling theory of science. According to Kuhn, a paradigm is what members of a scientific community, and they alone, share.

If one evaluates under the assumption that the assessed system is not profoundly influenced by the assessment then obtaining the needed information from the assessed social system through interviews, hearings and the like can disregard any detrimental effect this might have on the performance of that system.

However, if one evaluates under the assumption that the assessed system might be profoundly influenced by the assessment then the process of obtaining the needed information from the assessed social system becomes a prime concern for the evaluation panel. It then is not sufficient to assume that the verbal or written communication of the recommendations for improvement to the assessed will by itself lead to the improvement of the assessed system. Prior to this communication the interaction with the assessed system might have already destroyed any potential leverage of the recommendations and policy resistance might prevail.

It is for this phenomenon called policy resistance that I find the action research paradigm more convincing than the older empirical social research paradigm. And even though many members of the Five-Year Assessment panel might not have been aware of the action research paradigm - after all many were with a natural science or an engineering background were this paradigm does not apply - they were enacting it out of their intuition.

The Case for Action Research

Since its invention in 1946 by Kurt Lewin[7], one of the pioneers of social psychology, action research has become an established albeit not mainstream field of the social sciences in general and of program evaluation in particular in North America and UK[8] but has not gained much momentum in Central Europe. To clarify its rational and its consequences I find it best to quote one of Lewin's famous scholars, Edgar Schein, professor at the MIT Sloan School of Management. In his notable article[9] about Lewin, Schein lays out the ideas and implications of action research. And even though Schein does not address evaluation as such, his statement can readily be transferred to this domain by replacing the word consultant by evaluator and the word consulting by evaluation.

"The literature is filled with the notion that one first diagnoses a system and then intervenes to change it. I learned early in my own career that this basic model perpetuates a fundamental error in thinking, an error that led Kurt Lewin to the seminal concept of Action Research. The conceptual error is to separate the notion of *diagnosis* from the notion of *intervention*. That distinction comes to us from scientific endeavours where a greater separation exists between the researcher and the researched, particularly from medicine, where the physical processes are assumed to be somewhat independent of the psychological processes (an assumption that is not even holding up in many parts of medicine).

The classical model is that the doctor makes an examination, runs certain tests, decides what is wrong, and writes a prescription which includes recommendations for therapy or, if necessary, for other interventions such as surgery. The consulting industry has perpetuated this model by proposing as a major part of most projects a diagnostic phase in which large numbers of interviews, questionnaires, and observations are made the basis of a set of recommendations given to the client. Consultants differ on whether they feel they should also be accountable for the implementation of the recommendations, but they tend to agree that there is a period in any project that is considered necessary—namely, a diagnosis of the problem—and that the consultant's basic job is done with a set of recommendations for future intervention. If interviews or surveys are done, the attempt is made to be as scientifically objective as possible in gathering the data and to interfere minimally during this phase with the operation of the organization. What is wrong with this picture?

If Lewin was correct that one cannot understand an organization without trying to change it, how is it possible to make an adequate diagnosis without intervening? So either consultants using the classical model are getting an incorrect picture of the organization or they are intervening but are denying it by labelling it just diagnosis. Isn't a better initial model of work with organizations something like the stress test that the cardiologist performs by putting the heart under pressure to see how it will perform, even knowing that there are some risks and that some people have been hurt during the test itself? This risk forces the diagnostician to think about the nature of the diagnostic intervention and to apply clinical criteria for what is safe, rather than purely scientific criteria of what would seemingly give the most definitive answer.

It is my contention that Lewin was correct and that we must all approach our consulting work from a clinical perspective that starts with the assumption that everything we do with a client system is an intervention and that, unless we intervene, we will not learn what some of the essential dynamics of the system really are. If we start from that assumption, we need to develop criteria that balance the amount of information gained from an intervention with the amount of risk to the client from making that intervention. In other words, if the consultant is going to interview all the members of top management, he or she must ask whether the amount of information gained will be worth the risk of perturbing the system by interviewing everybody and, if the answer is yes, must make a further determination of what is to be learned from the reactions of the management to being interviewed. That is, the interview process itself will change the system and the nature of that change will provide some of the most important data about how the system works (i.e., will respondents be paranoid and mistrusting, open and helpful, supportive of each other or hostile in their comments about each other, cooperative or aloof, and so on?) The best information about the dynamics of the organization will be how the organization deals with the consultant, because his or her very presence is de facto an intervention.

Yet the focus in many traditional consultation models is on the objective data obtained in the interview, with nary a reference to how the interviewer felt about the process and what could be inferred from the way he or she was received. The irony in all of this is that Lewin was by training a physicist and knew very well the rules of scientific inquiry and objectivity. For him to have discovered that human systems cannot be treated with that level of objectivity is, therefore, an important insight that is all too often ignored.

Diagnostic activities such as observations, interviews, and questionnaires are already powerful interventions and the processes of learning about a system and changing that system are, in fact, one and the same. This insight has many ramifications, particularly for the ethics of research and consulting. Too many researchers and consultants assume that they can objectively gather data and arrive at a diagnosis without having already changed the system. In fact, the very method of gathering data influences the system and, therefore, must be considered carefully. For example, asking someone in a questionnaire how they feel about their boss gets the respondent thinking about an issue that he or she might not have focused on previously, and it might get them talking to others about the question in a way that would create a common attitude that was not there before."

Overcoming Policy Resistance

Not all evaluations result in such leverage as has been the case with the Five-Year assessment. With other evaluations I have quite different experiences. During such evaluations resistance against some or all findings and even more so against well-intended and well-grounded recommendations based on these findings was being built up by members of the evaluated system resulting in little or no leverage of the evaluation[10].

In 1982 Donella Meadows[11], a pioneering environmental scientist coined this phenomenon policy resistance. It is only within the action research paradigm that such policy resistance can be understood as an adverse effect of the assessment itself. Notably, evaluators' unintended but nevertheless sometimes exam-like interrogations of members of the evaluated system during fact-finding mission inevitably raise resistance against the panel of those being interrogated. People feeling treated unappreciatively by an evaluation panel are hardly the best missionaries for the recommendations coming out of such an experience.

A good action research model of overcoming policy resistance is to start out with the intention of creating an evaluation team consisting of people from outside as well as from inside the evaluated social system. When evaluators and evaluated have joint ownership of the evaluation, both the validity of the assessment and the leverage of subsequent recommendations will be greatly enhanced.

Aiming at leverage evaluators need to become highly attuned to their own insights into what is going on and their own impact on the evaluated system. An evaluation must be viewed from a clinical perspective as a set of interventions that are being guided by their presumed impact on the evaluated system. More emphasis must be put on how different interventions affect a social system than on how to gather scientifically valid information from that very system.

To figure out what needs to change and discover where there is already some motivation to change, evaluators have to find out things about the present state of the system. To gather such information evaluators have to talk to people in the system and ask them questions or conduct surveys. What is especially important to discover is where there is already motivation to change and where there is already survival anxiety.

Initially, evaluators need to focus on what is working well and avoid criticism or problem foci. Interviews should start with what members of the evaluated system are most proud of or what works best. If evaluators focus on success and what works well, they are creating psychological safety that will make it easier for both parties later to discuss problem areas, difficulties, things that need improvement. The prime data that the evaluators should be looking for are where members of the evaluated system see problems or have motivation to change, but the initial assumption has to be that they will not be ready to talk about problems until they feel safe with the evaluators, and they feel safe only if the evaluators display appreciation of what works well.

As interviews or interactions proceed, the evaluators must be constantly alert for changes in mood or feeling on the part of the interviewees, being especially sensitive to issues that may be unappreciative to the interviewees, leading to a shutting down of the flow of information.

For evaluators to become effective as far as their recommendations are concerned they need to create an interaction that provides information to them, that builds trust with the members of the evaluated system, and that gets the members of the evaluated system to think diagnostically and positively about the evaluation such that they welcome another interview or interaction because their curiosity has been aroused. Thus, evaluators need to become facilitators of a learning process that makes sense to the members of the evaluated system.

The Micro-Level of Evaluation

So far, I have drawn attention to the macro-level of an action research guided evaluation by describing such a process. But this does not yet reveal the entire image of a high-leverage evaluation. Even when evaluations are done in a way as to overcome policy resistance I noticed differences. The openness of the conversations can be astoundingly disparate as can be the willingness to cooperate. Not only did I experience this between the panel and the members of the evaluated system but also among the members of the panel itself. In the worst case the evaluation deteriorated to some kind of alibi exercise where everyone was glad when it was over, where nobody cared to think hard about valid recommendations, and where nobody cared to bring these recommendations into being.

What is easily missed in evaluation panels is the mode of communication taking place amongst the members of the panel and also between members of the evaluated system and the panel. This constitutes the micro-level of the evaluation process. It does in fact make a big difference if the panel members treat the interviewees as partners or if they treat them as if they were students to be examined. It also makes a difference if the panel members openly share their views or hold back for whatever reason. Chris Argyris, Professor of Education and Organizational Behaviour at Harvard Business School and another pioneer of action research, provided ample evidenced that highly skilled people can efficiently impede each other in their capacity to learn and to act accordingly[12]. Argyris dubbed this phenomenon defensive routines - a by now well-known term to describe ineffective group behaviour. To get rid of defensive routines it is necessary to adapt ones own communication practice.

	Communication mode supporting teams to become effective[13]
Help and Support	Increase the others' capacity to confront their own ideas, to create a window into their own mind, and to face their non-shallow assumptions, biases, and fears by acting in these ways toward other people.
Respect for Others	Attribute to other people a high capacity for self-reflection and self-examination without becoming so upset that they lose their effectiveness and their sense of self-responsibility and choice. Keep testing this attribution opening.
Strength	Advocate your position and combine it with inquiry and self-reflection. Feeling vulnerable while encouraging inquiry is a sign of strength.
Honesty	Encourage yourself and other people to say what they know yet fear to say. Minimize what would otherwise be subject to distortion and cover-up of the distortion.
Integrity	Advocate your principles, values, and beliefs in a way that invites inquiry into them and encourages other people to do the same.

My own experience is that such a mode of communication requires quite a bit of sometimes frustrating practice at the outset but with surprising occurrences and results as reward.

With hindsight I now believe that the Five-Year Assessment panel succeeded in creating a mode of conversation that surpassed defensive routines. Most of that micro-level adjustment happened when the members of the panel met for the first time in June 2004. It was not an explicit decision but more like a tacit agreement to appreciate who was there and to welcome and honour all experiences and views present particularly when they ran counter to ones own experiences and views. This awareness that the Five-Year Assessment was way too complex to ignore or devalue any person's view and that we therefore needed to listen to each other probably did the trick for us then.

A Personal Remark, Finally

No doubt, evaluations are an indispensable instrument for improving research institutions, research policies, and research programmes. Can we create better evaluations? Or do we have to await an enlightened panel faithfully?

Paul Feyerabend[14], the famous anarchist in the philosophy of science, has argued that any attempt to institutionalize the "best method" is bound to produce inferior results. In this spirit I do not see action research as a "better paradigm" than the empirical social research paradigm. I do believe, however, that in order to achieve progress in the field of evaluation it is important to see that there are paradigmatic options to be chosen from.

Without the experience of the Five-Year Assessment panel I would not have had the courage to write this article because I would not know that action research actually works in evaluations. And without the scholars of action research, policy resistance and defensive routines I would not have been able to make sense of what I might otherwise have interpreted as just a lucky experience with friendly colleagues.

For both, the experience and my making sense of this experience I'm deeply grateful to my fellow panel members, to the members of Unit A.4 of Directorate-General for Research, and to the scholars of action research.

Author:

Univ.Prof. Dr. Christoph Mandl
Mandl, Lüthi & Partner
Sechshauserstraße 75, 1150 Wien
mlp@mlp.co.at

Notes

[1] See Five-Year Assessment of the European Union Research Framework Programmes 1999 2003, http://ec.europa.eu/research/reports/2004/pdf/fya_en.pdf

[2] See Five-Year Assessment of the European Union Research Framework Programmes 1999-2003 - Conclusion and Response from the Commission, http://ec.europa.eu/research/reports/2004/pdf/conclusions_response_commission_en.pdf

[3] E.g. see the reports Monitoring 2004 - Implementation of Activities under the EC and Euratom Framework and Corresponding Specific Programmes, August 2005, http://ec.europa.eu/research/reports/2004/pdf/monitoring_2004_en.pdf and Five-Year Assessment in the field of Promotion of Innovation and encouragement of SME participation, ftp://ftp.cordis.europa.eu/pub/fp5/docs/fp5_panels_final_report_innovation_2000.pdf

[4] See also the description of Program Evaluation on Wikipedia, http://en.wikipedia.org/wiki/Program_evaluation

[5] For example see RTD Framework Programme 5-Year Assessment Reports, http://cordis.europa.eu/fp5/5yr_reports.htm

[6] *Thomas Kuhn*, The Structure of Scientific Revolutions, University of Chicago Press 1962

[7] *Kurt Lewin*, Action Research and Minority Problems, Journal of Social Issues 2, pp 34-46, 1946

[8] See for example the book by *Peter Reason & Hilary Bradbury*, Handbook of Action Research, Sage 2006 and the website by Gene Shackman, Approaches to Evaluation and Politics of Evaluation, http://gsociology.icaap.org/methods/approaches.html

[9] *Edgar H. Schein*, Kurt Lewin's Change Theory in the Field and in the Classroom: Notes Toward a Model of Managed Learning, Systems Practice 9 No. 1, pp 27-47, 1996

[10] E.g. see the report Monitoring 2004 - Implementation of Activities under the EC and Euratom Framework and Corresponding Specific Programmes, August 2005, http://ec.europa.eu/research/reports/2004/pdf/monitoring_2004_en.pdf which mentions that earlier monitoring recommendations have not been put into place

[11] *Donella Meadows*, Whole Earth Models and Systems, Coevolution Quarterly No. 34, Summer 1982

[12] See *Chris Argyris*, Teaching Smart People How to Learn, Harvard Business Review 3, pp 99-109, 1991 and Chris Argyris, Skilled Incompetence, Harvard Business Review 5, pp 74-79, 1986

[13] This table is taken from Chapter 6. Reducing the Organizational Defense Pattern, in Chris Argyris, Overcoming Organizational Defenses, Prentice Hall 1990

[14] See his book Against Method: Outline of an Anarchistic Theory of Knowledge, 1975

Stefan Kuhlmann

Book Review

"A New Deal for an Effective European Research Policy - The Design and Impacts of the 7th Framework Programme", by Ugur Muldur, Fabienne Corvers, Henri Delanghe, Jim Dratwa, Daniela Heimberger, Brian Sloan, and Sandrijn Vanslembrouck; Dordrecht: Springer, 2006, 289 pages (ISBN 978-1-4020-5550-8)

This book suggests a kind of prudential come-back of "policy planning", here in the field of European research and innovation policy.

Policy planning? Once upon a time, this was the cipher and a hopeful instrument for active policymaking, aiming to shape modern society for a better life, not only in socialist planning economies but also in the Keynesian US economy. Since President Roosevelt's "New Deal" programme aiming at economic reform to overcome the Great Depression (1933–37) large-scale political projects have come along with systematic planning efforts and related concepts, such as "Operations Research", "Policy Science and Analysis", "Systems Analysis". In late 1960s the planning hype culminated in the launch of the US Federal Administrations "Programming, Planning, Budgeting System" (PPBS), an ambitious intelligence approach that all-too soon failed. The reasons for failure were nicely addressed in the subtitle of Pressman and Wildavsky's famous book "Implementation" (1973): „How Great Expectations in Washington Are Dashed in Oakland" – the instrumentalist thrust of planning tends to become counterproductive on its own terms because it neglects the complexities and the auto-dynamics of modern political, economic and societal interaction. Too often the underlying assumptions about the working of policies were rather mechanistic, not reflecting the complex interference of diverse "realities" with policy intentions. As a consequence, since the mid-1970, the term "policy planning" was no longer referred to, and even large policy initiatives – such as the European Research Framework Programmes (since the mid-1980s) – were not preceded by ambitious planning efforts. Instead, planning ambitions remained modest, "muddling through" (Lindblom 1959) approaches prevailed.

In our field of research and innovation policy, since the 1990s, intelligence for policymaking became largely based on retrospective analyses of expired initiatives (policy evaluation), only loosely linked with "foresight exercises" aiming at the identification of promising fields for policy action (Martin & Irvine, 1984: "Foresight in Science – Picking the Winners"). Integrated approaches to strategic research and innovation policymaking, based on combined evaluation, foresight, and technology assessment exercises, drawing on distributed strategic intelligence, were called for but hardly implemented (Kuhlmann et al. 1999).

It was only at the beginning of the present decade when in Europe a kind of "policy planning" approach returned, not at least as a response to the European Parliament's request to be informed about the potential impacts of large policy initiatives: Since 2003, every legislative proposal by the European Commission must be accompanied by an in-depth impact assessment report. It replaced existing requirements for a variety of impact assessment (business, gender, environment, trade, etc.) and aimed towards one integrated impact assessment, including trade-offs between different aspects.

The book "A New Deal …" was written by Ugur Muldur and his colleagues, the team in charge of the ex ante Impact Assessment of the 7th European Research Framework Programme (FP). This exercise "… informs decision-makers of the likely consequences of policy choices by answering a common set of questions. It assesses the issues at stake and the objectives to be pursued by the policy proposal. It examines the views of the main stakeholders that will be affected by the policy. It identifies the main policy options for achieving the objectives, and analyses their likely economic, environmental and social impacts. And it outlines the advantages and disadvantages of each option, as well as synergies, trade-offs and risks". (xxiv) Such phrasing signals an ambition and, at the same time, it offers room for discourse and reflexivity ("stakeholders", "options", "disadvantages"). In a way the ambition appears prudent, attempting an "enlightened" policy-planning exercise (as opposed to the 1960s concepts) – this holds at least for the Impact Assessment of the 7th FP, being part of the complex and cumbersome process of launching a huge new programme, bigger in scope, coverage, budget and duration than ever.

The material presented in the book draws extensively on text of this Impact Assessment, but it offers also additional analyses, considerations and statements.

Chapter 1 explores some of the important challenges Europe is facing, the expectations held by the public of science and technology in addressing these challenges. Chapter 2 shows that

Europe needs to tackle a number of key weaknesses in science and technology currently preventing Europe from achieving its full potential. Chapter 3 traces how at the Lisbon European Council of March 2000 and its aftermath a new European research policy context was created, within which it was possible to pursue a FP much more ambitious in scale and scope than any past FP.

The next three chapters of the book look at how the new FP was actually conceived. Chapter 4 takes stock of the experience gained through the implementation of past FPs. The presented evidence endorses the claim that FP 6 has had a clear positive impact on Europe's scientific, technological, and innovative performance, and had also generated wider economic, social, and environmental benefits. Chapter 5 explains how in the development of the Commission's 7th FP proposal, account was taken of the views of outside stakeholders. Chapter 6 offers a model-based (model Némésis) assessment of the likely "European added value", in relation to a number of policy options. These options are analysed with regard to three issues: the maximisation of the FP's macro-economic impacts, overcoming the fragmentation of European basic research, and raising the competitiveness of European industry through research and innovation. The analysis, and hence the Impact Assessment, results in a clear statement "that a much larger Framework Programme would have the largest macroeconomic impacts, and that the establishment of a European Research Council and Joint Technology Initiatives constituted the best means to overcome the fragmentation of European basic research and to raise the competitiveness of European industry through research and innovation." (257)

Chapter 7 describes in detail the cumbersome inter-institutional European decision-making machinery and shows what happened to the ambitious FP 7 proposal in Council and Parliament after it was adopted by the Commission: The funding for the new FP will be 30 per cent less than that initially proposed by the Commission, although representing an increase in average terms vis-à-vis that of the previous FP of approximately 40 per cent. The authors' overall assessment of this process, in one sentence: "The expected impacts of the 7th Framework Programme, though smaller than those of the Commission proposal, are substantially larger than those of a business-as-usual Framework Programme." (258) But on order to come true, a number of conditions would have to be fulfilled (260/2), inter alia a sufficient mid-term adaptability in terms of the FP's scientific content: Neither the Impact Assessment nor the FP (this time running seven years) have addressed all potentially relevant themes in science and technology – just think of the very recent hype of climate change issues.

Finally, the authors afford a view beyond FP 7, making a plea for “a new deal for an effective European research policy”. They admit that seven years after the launch of the European Research Area, some progress has been made in terms of two key dimensions (coordination and excellence). The Open Method of Coordination, the mutual opening up of national research programmes, and Framework Programme instruments such as Networks of Excellence and ERA-Nets had some success. But too much remains to be done: “… the efforts made so far to promote greater coordination and cooperation between S&T players are just a start” (263), a “more dynamic approach is needed towards the realisation of the European Research Area”, meaning “a reallocation of responsibilities (…) going beyond the existing structure of the European research system and redesigning, even reinventing, it.” (264) Here, the prudent and enlightened policy-planners explicitly step out of their function and adopt the role of a policy entrepreneur! After all the evidence-based analysis, this is an acceptable, and in my view laudable, venture.

Thereby, the authors quickly return to their immediate field of competence: A re-invention of the European research system requires relevant and reliable comparative information on the effectiveness of policies at different policy levels. But existing “S&T indicators and datasets cannot provide this kind of information.” A “completely new data effort is needed. The first step in this new generation data effort should consist of the development of new and improved ex-post evaluation methodologies. These should seek to link a particular public research support input with all possible S&T and societal outputs and impacts.” (265)

In sum, an enhanced and enlightened strategic intelligence base is needed, “a pan-European knowledge community, not only in the area of ex-post evaluation, but also in socio-economic and scientific and technical foresight. One cannot hope to restructure the European research system and achieve a true European Research Area if there is no common vision of how major societal and S&T challenges will evolve, how scientific disciplines will develop, and how the research system as such will progress”. (266) Here comes again the policy entrepreneur! And, again, I would agree: We need a European (and beyond!) “knowledge community” around science, innovation, society, policy issues and related strategic intelligence. But this community does not need a “common vision” – competition of ideas will provide for more creativity and robustness.

Should you read this book? Actually, it reads not always easy. The text suffers from some redundancy, which is certainly an effect of multiple authors working on overlapping subject areas. Also the tone of the text changes across chapters, a consequence of several writing styles.

This is not a book you would read in one piece. Rather it displays several characteristics: It shows elements of a textbook on European societal and economic needs, research performance, and research policies. It is a documentation of and gives insight in the ambitious enterprise of developing and applying an enlightened policy-planning process, including a presentation of the methodology, a model-based ex ante policy impact assessment. It is a policy document. And it offers a political plea.

I like this book, not at least because it is such a hybrid. You can browse it and pick pieces for different purposes at various occasions – information, teaching, analysis, enlightened policy planning in the field of science and technology policy. If this is your area of concern you should have the book on your shelf!

Author:

Stefan Kuhlmann
University of Twente, Institute for Governance Studies (IGS)
Capitool 15, P.O: Box 217, 7500 AE Enschede, The Netherlands
s.kuhlmann@utwente.nl

References

Kuhlmann, S./ Boekholt, P./ Georghiou, L./ Guy, K./ Héroud, J.-A./ Laredo, Ph./ Lemola, T./ Loveridge, D./ Luukkonen, T./ Polt, W./ Rip, A./ Sanz-Menendez, L./ Smits, R., Improving Distributed Intelligence in Complex Innovation Systems, Brussels/Luxembourg, 1999 (Office for Official Publications of the European Communities) see also http://www.isi.fhg.de/ti/Final.pdf

Lindblom, C. The Science of Muddling Through, Public Administration Review 19, 1959, pp. 79-88.

Irvine, J./ Martin, B.R.,. Foresight in Science Picking the Winners London, Dover Frances Pinter (Publishers), 1984

Pressman, J.L./ Wildavsky, A., Implementation: how great expectations in Washington are dashed in Oakland; or, Why it's amazing that Federal programs work at all, this being a saga of the Economic Development Administration as told by two sympathetic observers who seek to build morals on a foundation of ruined hopes. Berkeley, University of California Press, 1973

Der Newsletter der Plattform Forschungs- und Technologieevaluierung ist ein unregelmäßig erscheinendes offenes Forum zur Diskussion methodischer und inhaltlicher Evaluierungsfragen in der Forschungs- und Technologiepolitik.
© Wien 2007 ZVR-Zahl: 937261837 ISSN-Nr.: 1726-6629

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

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

PLATTFORM FORSCHUNGS- UND TECHNOLOGIEEVALUIERUNG

Die Plattform Forschungs- und Technologieevaluierung 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.

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