



Barend
van der Meulen

A letter to German colleagues on Dutch experiences

Im November 1999 fand in der BBAW eine Tagung zum Thema »Qualität« statt. Die Niederlande gelten als Vorreiter bei der sukzessiven Umstellung von Universitäten und Forschung auf externe Qualitätsbewertung. Barend van der Meulen, der an dieser Tagung teilnahm, fühlte sich in vieler Hinsicht an Diskussionen erinnert, die in den Niederlanden vor 16 Jahren stattgefunden haben. Den folgenden Brief schrieb er nach den Gesprächen in Berlin. Barend van der Meulen betonte jedoch, dass er damit nicht sagen möchte, sein Land sei Deutschland voraus oder habe die Probleme schon hinter sich: »There is no reason to assume that evaluation practices in Germany will develop similarly to those in the Netherlands.«

Dutch universities faced the first formal evaluations of academic research in 1982, when the Government decided to make institutional funding conditional on positive ex ante assessments of research programmes. Universities had to organise at least 80% of their research in five years research programmes of considerable size – a task accomplished rather easily in the natural sciences, but not in the social sciences and humanities. The latter felt to be forced into an uncomfortable straitjacket, but in the end

The new regime broke paths at the beginning, and certainly functioned in identifying stagnant research at the universities. Researchers soon learned how to cope, how to formulate programmes, and about required performances.

succeeded in organising programmes – at least on paper. The academic community underwent a phase of uncertainty about quality criteria. The assessment was organised by the »Ministry of Education and Sciences« and carried out by disciplinary peer committees, who were to merit

each programme regarding scientific quality and social relevance. Results of the evaluations showed that the peer committees had evaluated using different standards. Some, more strict, tried to select only the best of the programmes. Others, in contrast, were generous and defended the discipline (including their own) against the new pressures. Only a few of the committees were able to cope with the social relevance criteria. The new regime broke paths at the beginning, and certainly functioned in identifying stagnant research at the universities. Researchers soon learned how to cope, how to formulate programmes, and about required performances.

Another evaluation practice also developed in the eighties. These evaluations were linked to Government's ambitions to improve strategic planning of university research and to implement budget cuts. From 1978 to 1989, about 14 ad hoc disciplinary committees were initiated by the »Minister of Education and Sciences« in fields ranging from physics to theology and from chemistry to political sciences. Each of these committees was asked to assess research performances, to prognosticate the discipline's development, to advise about priorities, posteriors (negative priorities) and to come up with recommendations for structural reforms. Some of the committees were also asked to advise on closing of departments. Considering the budget restrictions the Dutch universities faced in the eighties, it goes without saying that most of these committees operated in a high stakes context and their findings were often disputed.

With hindsight, probably the largest impact made by these committees was the acceptance of bibliometric indicators for evaluative purposes. Each committee tried to find the best set of indicators, some relying more on scientific publications, others emphasising citation scores; some limiting themselves to a single indicator, others combining several indicators. The immediate impact made by these indicators was often a debate about these methods. In an evaluation of these committees around 1990, we showed

that opinions on the bibliometric methods often correlated with opinions on the recommendations. The then »Advisory Council for Science Policy« (now the »Advisory Council for Science and Technology Policy« – AWT) deviated from this pattern, and developed a consistent view on bibliometric tools, which, over time has more or less become the Communist opinion of the Dutch academe. University managers and researchers have learned to interpret and use simple indicators – such as number of publications and citations – as modest tools for evaluation and decision-making.

By now, almost twenty years later, the evaluations are still important for effectuating accountability of universities towards their main sponsor, the Government. However, as with publishers in the well-known peer review of manuscripts, the existence and good organisation of the evaluation ensures the sponsor that research is of good quality. The Government is no longer involved in the organisation, left alone in the formulation of criteria or determining the kind of conclusions and recommendations. The evaluations now function mainly to inform the university management (at department, faculty and university level). The direct linkage with institutional funding has disappeared, although excellent evaluations are usually rewarded with a bonus. The »Association of the Dutch Universities« took over the responsibility for the evaluations in 1992. Evaluations are now organised discipline-wise in a four-year rolling scheme. I won't go here into the details of the evaluation procedure here, but instead mention the three pillars on which it is based.

The first pillar is the self-evaluation report, written by the department and faculty. The basic function of the report is, of course, to inform the evaluation committee on the mission, performances, productivity and future plans of the research group. But the act of self-evaluation also helps the research management understand the strengths, weaknesses and commitment to the evaluation and its outcomes. The second pillar is the evaluation committee, with a membership of peer researchers working





abroad and chaired by an eminent Dutch researcher with knowledge of the organisation of Dutch universities. In principle, the committees will visit the departments. The third pillar is the protocol that defines the evaluation procedure and the evaluation criteria. The protocol leaves some room for variety, but basically every evaluation is carried out in a similar way. Researchers know what the evaluation criteria are and how they will be evaluated, evaluation committees don't have to discuss extensively the evaluation procedure.

Every research programme is evaluated on its quality, its productivity, its viability and its relevance. The four criteria are not independent of each other and it is easy to show that in practice, the scores correlate significantly. Nevertheless, the distinction between the four has mitigated discussions on the nature of scientific quality and the validity of bibliometric indicators. More importantly that the distinction enables evaluation of young groups or groups that have recently changed its research programmes radically without direct budget consequences for such groups. Such groups may not have had the time to develop and show quality and productivity. Based on the future plans, however, the committee can assess viability and relevance, and do so often.

of my PhD committee, a well known Dutch philosopher, declared at the public defence of the thesis that in the eighties he had strongly opposed the interventions of the government, but looking back he could not but confess that they had been necessary and were of benefit.

Present evaluations do not have such major impact and, (as a consequence?) are widely accepted within the academic community. Outcomes are not mechanically translated into allocation of resources, but used intelligently in all sorts of research management decisions. With the ease by which current evaluations are done, one would almost forget the turbulence induced by the first evaluations. There are still some scars, as silent memories: full evaluations are not made public, but the scores on the four dimensions plus a short explanation are published. And if citation analyses are made, the report publishes the results anonymously. The message is then ambiguous. If, as a university researcher, one could choose between the evaluation practices as implemented in the eighties and those operating now, certainly the latter seems to be much more attractive. But it is clear that the normality of these practices is based upon the changes enforced by the evaluations in the eighties. In that respect, it is not a choice between two evaluation practices. The real choice is whether the evaluations should induce radical changes or are meant to optimise current research strategies. For that choice, a discussion of evaluation methods might be a useful detraction of the actual problems, but is certainly not helpful in solving them.

The implementation of evaluations ... contributed to a major change in the research climate at universities. Dutch university researchers publish more, and more internationally also.

The implementation of evaluations in the eighties contributed to a major change in the research climate at universities. Dutch university researchers publish more, and more internationally also. Peers from abroad are often impressed by the concentration of high quality research in the Netherlands. Good research performances within the universities are recognised and rewarded. In 1992, my own dissertation contained a chapter on the effects of the intervention on the philosophy departments. A member

Dieter Simon

Aspekte der Qualität

Alle sagen es – wenn auch mit verschiedenen Worten. Neidhardt sagt es, Mittelstraß sagt es, und ich sage es auch: In Deutschland herrscht Evaluitis. Die Evaluitis ist eine fiebrige Erkrankung, die unversehens den Körper der Wissenschaft in seiner Gesamtheit erfasst hat. Und nicht nur ihn. Alle Stätten der Wissensproduktion werden einer generellen Revision unterzogen. Es gibt keine nichtbefallenen Teile mehr. »Optimierung der Qualität« ist zum ubiquitären Bedürfnis geworden. Alles und alle sind betroffen. Universitäten und außeruniversitäre Einrichtungen. Fachhochschulen, Gesamtschulen und Fortbildungsinstitute. Der einzelne Forscher als Wissenschaftler, als akademischer Lehrer und als Mensch. Die Gesamtheit der Gelehrten, sei es mit ihrer Einrichtung, sei es ohne sie. Die Totalität der wissenschaftlichen Anstalten als selbständiges System. Die Akademien und die Förderungsinstitutionen, die Krankenhäuser, das Dienstrecht, die Parkplätze und die Bibliotheken. Die Preisverleihungen. Der große Nobelpreis und die kleinen Nobelpreise. Überall die gleiche bange Frage: Wie können wir die Gutachter zu der Überzeugung bringen, dass wir wirklich »Qualität« liefern?

1974 publizierte Robert M. Pirsig einen autobiografischen Text (»Zen and the Art of Motorcycle Maintenance«), der sich auch in Deutschland schnell zu einem Bestseller und Kultbuch (»Zen und die Kunst ein Motorrad zu warten«) entwickelte und heute (mit der inzwischen 25. Auflage) immer noch an Lesern gewinnt. »What the hell is quality? What is it?« ist die zentrale Frage des Buches. Eine Frage, die den Helden am Ende seiner wissenschaftlichen Erkenntnisfahrt in den Wahnsinn treibt. Eine Frage, die zurückführt in die mythische und relative Welt, in eine Welt, die noch ohne das sokratisch-platonische Ideal der menschenunabhängigen, »absoluten« Wahrheit lebte. Die Schicksalsfrage aller Gutachter, Bewerter und Evaluatoren: Was zum Teufel ist Qualität? Eine Frage, an der heute,

obwohl wir nicht in den Mythos zurückgekehrt sind, gleichwohl keiner mehr zerbricht. Diesen Zustand zu erreichen war nicht schwer. Es musste lediglich eine bekannte, beunruhigende Beobachtung verdrängt und eine andere in ein Vertrauen erweckendes Phänomen umgedeutet werden.

Die Beunruhigung: Es gibt die leidvolle Erfahrung, dass es sehr schwer, ja nahezu unmöglich ist, eine handfeste Definition dafür zu geben, was Qualität »an sich« eigentlich ausmacht. Wieso hat ein Bild, ein Buch, ein Theaterstück, ein Gebäude, ein wissenschaftliches Resultat Qualität und ein anderes nicht? Wie soll man messen, was man nicht definieren kann? Denn schließlich kann man ohne Maß nicht vergleichen und ohne Vergleich nicht rational entscheiden. Man sieht doch täglich, wie sich die Experten winden, wenn sie ein Urteil detailliert, aber ohne Bezug auf andere Wissende begründen müssen.

Die Umdeutung: Der Beunruhigung steht die zumindest irritierende Feststellung gegenüber, dass innerhalb und außerhalb der Disziplinen offenkundig ein weit verbreitetes Einverständnis darüber herrscht, ob und wann Qualität vorliegt. Man »weiß« es einfach. Ein Umstand, der für den Helden Pirsigs, der er unter dem Namen Phaidros selber ist, zum Ausgangspunkt weit greifender Reflexionen wird: Wieso kennt man etwas, das man ohne den Gegenstand, dem es anhaftet, nicht exakt beschreiben kann, so dass man es eigentlich nicht kennt? Was Phaidros/Pirsig als Ausgangspunkt nehmen, kann man allerdings auch zum Endpunkt allen Nachdenkens machen. Die selbstbewusste, abendländische Frage des Sokrates an den zeitgenössischen Dialogpartner und Sophisten Phaidros: »Was aber gut ist, Phaidros, und was nicht – müssen wir danach erst andere fragen?« kann ohne Umstände als Antwort ausgegeben werden. Man braucht andere nicht zu fragen, weil man das Wissen in sich trägt. Man weiß zwar nicht, warum, aber man weiß, dass. Das genügt.