The Sixth *QHE* Seminar

The End of Quality?

Birmingham, 25–26 May, 2001

Running the maze: interpreting review recommendations

Margarita Jeliazkova

CHEPS, University of Twente

Abstract

Major findings and the theoretical insights gained by a study on the impact of external review reports in Dutch higher education are presented. A model of argumentation around review recommendations is presented and illustrated by cases from Dutch higher education institutions. Practical implications are discussed of using the model as a predictive tool for success and failure and as a basis for the development of a practical instrument for optimisation of external quality reviews' effects.

This paper is based on empirical research on the impact of external review reports on education in the Dutch higher education system (Jeliazkova & Westerheijden, 2000)[1]. First, the context and the main issues of the research are presented. Second, a model is introduced for analysing the way higher education institutions respond to review recommendations. The analysis throws light on the conditions under which a system of external quality review can successfully promote quality improvement. Finally, possible applications for enhancing review techniques and quality management are discussed.

When the curtain falls: what happens to review recommendations after the review report is made public?

In the Netherlands, a national system of quality assurance for higher education has been in place since 1987. Very briefly[2], the existing procedure of external quality assessment of education in the Netherlands takes place as follows: a peer review committee,

---

[1] Conducted by CHEPS in 1998–2000 and commissioned by the Dutch Central Court of Audit

[2] We do not consider here the ongoing development towards accreditation, as it does not influence the core of the argument.

appointed especially for the purpose, evaluates clusters of similar programs. The committee bases its judgment on a self-evaluation report and a site-visit. It produces a report for the whole cluster, with specific recommendations for each programme. The report is made public and the programme staff is expected to act in response to the recommendations. The manner of controlling this follow-up, mainly by the Inspectorate for Higher Education, has evolved through the years and varies slightly for higher professional institutions and universities. Ideally, the next visitation completes the cycle by judging the measures taken in response to the previous one. The review system is intended to be self-regulatory. (Vroeijnstein, 1994) The definition of quality is constantly renegotiated, per programme and per visitation committee, thus leaving a lot of space for legitimate questions (and some speculations) about the actual effects of the review system. A growing number of actors inside and outside the academic world utilize the outcomes of these quality reviews for various purposes. Their shared expectation is that any action taken is aimed at quality improvement and will indeed bring about better quality of education.

The Center for Higher Education Policy Studies has followed the developments in the visitation system from the very beginning. (Brennan *et al.*, 1992, van Vught & Westerheijden, 1993, Frederiks *et al.*, 1993, Weusthof, 1994, Westerheijden, 1997, Scheele *et al.*, 1998) The newest opportunity for an in-depth look was provided by the Dutch Central Court of Audit. The chief accountant of the country was interested in the actual impact of external review reports on quality improvement in the reviewed programs. More specifically, the country's Chief Accountant sought the answer to the following question: do the recommendations in external review reports lead to actions aimed at quality improvement?

The question itself was certainly not surprising, who it came from made it exciting. The accountability part of the subject review system could not be taken lightly. It was not enough to claim how difficult it is to pinpoint subtle changes in attitude rather than some clearly depicted results. After all, the 'fuzzy' process of (re)negotiating quality criteria and judging programs did have tangible products: the published external review reports. The seemingly straightforward link between the recommendations in these reports and the action taken by the reviewed programs in response to them became the focus of our research. We had the opportunity to investigate it, hands-on.

Making sense of empirical data: moving beyond description

A multiple-case study was set up, to include four academic and four higher professional institutions in comparable subject fields, and with a 'match' between positive reviews and not-so-favourable ones. We concentrated on review recommendations concerning curriculum improvement, and thereby focused on the programme level, involving the faculty and institutional level where necessary. Next to a document study, sixteen semi-

structured interviews were conducted with persons directly involved in quality assurance of the programs researched.[3]

A model from earlier studies was adopted to classify the types of recommendations (Westerheijden, 1997) and responses (Westerheijden & Maassen, 1998). Four types of responses to recommendation (actions) could be differentiated along two dimensions: time (long-term and short-term), and causality (instrumental and conceptual) (Jeliazkova, 2001) (Figure 1). This scheme proved useful as a descriptive tool: as a solid point of departure towards explaining 'the fate' of external review recommendations in the last stage of the evaluation cycle..

conceptual

• Discussion on report findings and conclusions at various levels • Dissemination of report findings and conclusions • Apparent interest in visitation outcomes • Considering visitation procedure/outcomes important • Use of concepts/argumentation from the report for short term decision making	• Judgment on (purpose of) the visitation procedure/report • Judgment on quality assurance/ quality improvement/ aspects of quality • Judgment on the relative importance of education as a component of performance at individual/disciplinary/ departmental level • Use of concepts/argumentation from report for strategic decision making

short-term / **long-term**

• Actions taken to immediately improve/adjust/legitimize quality assurance policy	• Direct use of recommendations from visitation reports for long term decision making • Use of visitation report's structure/ indicators/criteria for quality testing (op various levels) • Reorganizations of disciplines/sections/ departments

instrumental

Figure 1. Types of action in response to reports

The data revealed that, all other things being equal, (normal organizational and managerial prerequisites for the success of any policy being in place), programs which have undergone an external review process adopt review committee recommendations

---

[3] Including using the Q-sort technique that we will not explain in detail here.

with a different pace and to a different degree. This largely depends on the way and the extent to which these recommendations turn out[4] to fit their views on what good education is all about.

Even if no one ever doubts that actions taken in response to a quality review are aimed at quality improvement, then it is still hard to judge their effectiveness. After all, the point of reference is constantly renegotiated, making quality review reports the outcome of a lengthy process of social interaction. The understanding of the various ways in which actors define quality (Harvey & Green, 1993) and fit their definitions in their larger set of beliefs and ideas, becomes the key element to explain the difference between action and inaction. This task goes beyond the largely descriptive models of research on quality review effects so far.

This is why we turned to a general theoretical framework for evaluating public policy. F. Fischer's framework (Fischer 1985; 1995) is built upon the notion that while ultimate values cannot be subject to scientific verification, this should not mean that values cannot be considered in a reasoned and logical fashion. Rather, they should be discussed in the light and in the context of the existing facts. In order to bring about clarity and consistency in policy deliberations, Fischer places them at four levels: verification, validation, vindication and rational social choice. The discourse at the first two levels takes place within the context of the prevailing value system, and is concerned with questions about effectiveness and validity of goals. At a higher level, a policy deliberation concerns the justification and acceptability of the very value system adopted to judge the policy, sometimes cutting deep to core convictions as preferred social order and way of life.

The framework, after adjustment for our specific purposes, proved useful for discerning the cases in which quality recommendations (fail to) produce results because of basic (dis)agreement on ends (different views on high quality education and on its place in a larger social context) or because of (dis)agreement on means (within the framework of shared understanding of goals, the emerging 'quality culture').

Running the maze: four levels of 'processing' review recommendations

Obviously, when a review report comes out, the recipients of review recommendations must consider acting on it[5], and this action has to fit in the existing practices, beliefs and traditions. The deeper a recommendation cuts into this core, the more difficult, and slow, will it be to rearrange the whole system, the 'vulnerable construction' (in the words of a respondent) in a way that makes room for the new element. When, and only when this process of 'fermentation' and 'maturation' of a 'dialogue' with the authors of the review report runs successfully, a recommendation will be rendered fit to act upon. The way recommendations are finally codified by the recipients determines the manner in which

---

[4] This 'match' may have been the case from the very beginning, but it may be an outcome of a lengthy debate.
[5] The review system has some in-built mechanisms to ensure that *some* action is taken

they will be handled and the type and pace of action they will produce. Thus, at the end, the degree of difficulty and the speed of 'growing towards' a shared viewpoint as a prerequisite for successful problem tackling, depend on the quality of the debate over recommendations (and thus over quality). The usual suspects, the notorious bureaucracy and lack of motivation, are not sufficient explanations for failure to improve quality.

From this perspective, there are two large groups of review recommendations. The first one concerns the programme's specific problems, where the committee and the programme's staff share the same view on the way the programme's general goals are defined, or could be defined. The second group of recommendations places the programme in a larger context of a discipline or a university. The very relevance of the programme's goals (and the criteria for success and failure derived from them) is subject of discussion here, for instance by broadening the programme's context to include demands of the students, the employers and society as a whole. These two groups of recommendations are further divided into four different levels at which the discussion about quality improvement takes place.
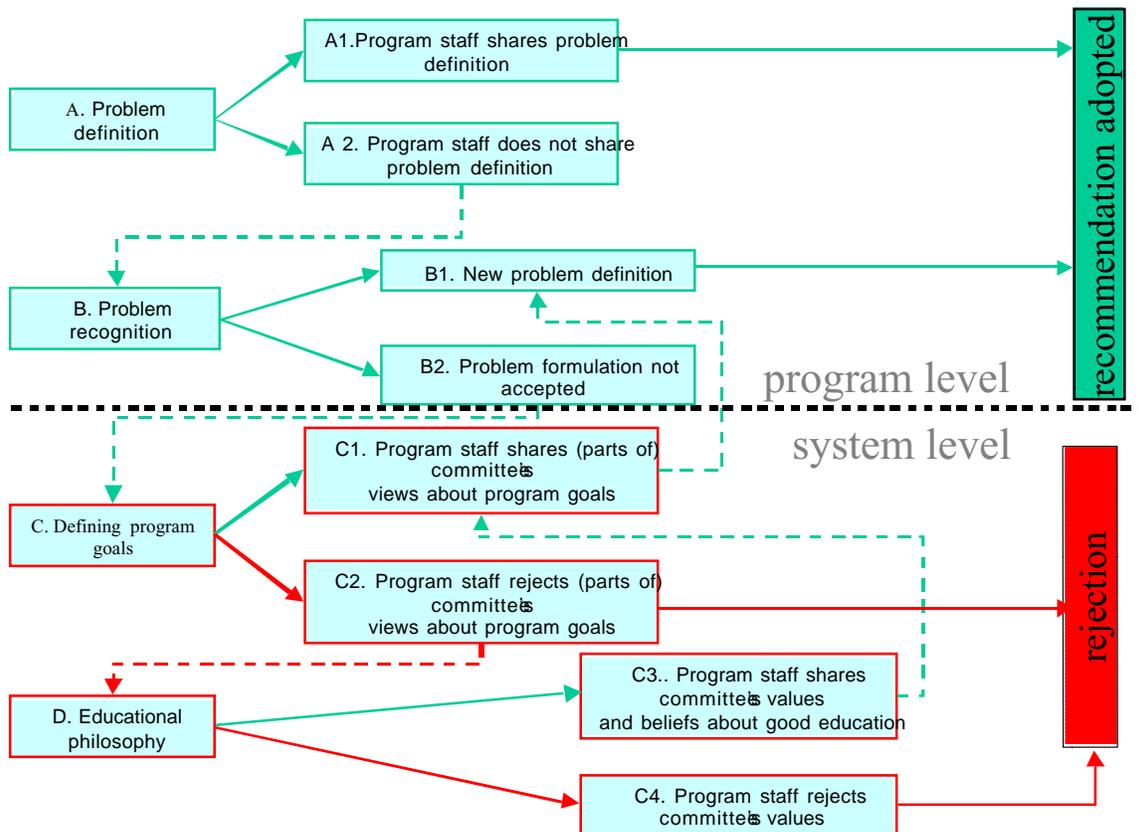

Figure 2: The maze of interpreting review recommendations

Level A: problem definition (problem solving)

The first type of recommendation concerns solutions for specific problems within the programme. At this level, the committee evaluates the programme elements' functioning. The committee recognizes and accepts the way the programme sees and defines these problems and suggests solutions. The question asked is: does the programme function as intended? The suggested changes fit a given structure and the derived quality criteria.

A good example of a level A recommendation is one visitation committee's suggestion to look at another programme's experiences with organizing a writing skill class. Such instances of sharing best practice are generally welcomed by programs. In other cases, the committee takes the role of the outsider who can just pinpoint problems 'hanging in the air'. As one respondent put it, "sometimes you have to name things in order to move further, though we had noticed the symptoms for some time already."

Level B: problem recognition (problem finding)

At the next level, the committee not only offers specific problem solutions; it also puts on the agenda (explicitly or implicitly) new aspects. These may have not been seen as problematic so far but seem to conflict with the programme's goals. Examples of level B recommendations are suggestions for coordination of different programme parts, for adding new elements in the programme, etc. It is clear that, as long as the programme staff shares the committee's basic viewpoints and criteria, it will take the suggestions seriously, and will (re)define them in an acceptable way that assures successful problem solving.

One committee visiting a professional programme at a computer-engineering programme suggested that students lacked communication skills and recommended hiring external experts to teach them. However, the programme staff saw the core of the problem not in the lack of teaching expertise, but in the type of students attracted to such a study. Until that moment, it was considered a given that these students did not have strong communication skills. In the new curriculum, these received a structurally new place. This is how a careful consideration of the problem definition employed (implicitly) by the visitation committee led to redefining the problem and consequently addressing it in a new way. The specific recommendation (the solution) was rejected, but an alternative was adopted which fitted the new definition.

Level C: Defining programme goals

If the programme staff and the review committee do differ in their interpretations of what exactly the problem is, it is time to look at the basis, at the principles around which a programme is organized and which justify its existence. At this level, a programme's goals are reconsidered 'from the outside', in the larger context of an institution, a discipline, or a professional field. The programme is thus judged from the viewpoint of social relevance. Not surprisingly, the discussions on these topics can be quite passionate.

Such processes may take a number of years of discussion, search and compromise seeking before any results become visible. It is at this level where complaints about the committee's bias, insensitivity and infringement on academic autonomy begin to sound seriously.

In one case, only one recommendation actually referred to these core questions. However, all other points were perceived in this context and could not be resolved until the principle discussion was settled. The programme's staff in our example did not accept the way the visitation committee tried to intervene with the programme's profile, which is normally the outcome of national negotiations of all similar programs. Very soon, all other review recommendations came to be seen in this light and perceived as 'biased', thus not to be taken seriously by the programme.

On the positive side, in another example, a very complex merger between two interdisciplinary programs was carried out relatively easy because the committee supported the programme's view on 'interdisciplinarity'. This basic agreement opened the opportunity to implement the recommended merger relatively smoothly. The other recommendations became the necessary supportive means towards this end.

Level D: Educational Philosophy

Occasionally the difference in views between the review committee and the programme touches upon broad social values, such as the place and role of education in society or the ideals of educational institutions. Review committees assume (in most cases rightly so) that this general vision is a given and is shared by all and thus there is no need to discuss it. However, one of our cases demonstrated that under certain circumstances the debate could shift quite surprisingly to this highest level, thereby effectively blocking the tackling of relatively simple specific problems.

This was the case with a programme in environmental studies. The committee rejected the programme's concept of an interdisciplinary study, other actors joined the debate, and very soon, the issue was about the place of environmental education in academic life and modern Western society at large. Even trivial suggestions for curriculum changes did not receive any attention before this discussion was sorted out - four years after the visitation.

In the schematic presentation of the dynamic model, we have shown the cases where 'no action' is expected. We found out that when the programme interprets review recommendations at levels A and B, they always lead to action. On the system level, it is logically possible to reach no agreement and consequently not to act on a recommendation. In our cases however, we found only delays and difficulties in the process, but no cases of 'no action at all.'

Discussion: practical implications

The conclusion is that review recommendations, even when found problematic, have a significant influence on the way programs are run, and eventually contribute to the evolution of staff's deeper convictions and beliefs. Long-term changes on system level probably have more far-reaching effects on the quality of programs than the measures undertaken within a clear and undisputed framework. It is on this system level where change in culture in the long run can be observed.

The levels of argument in the dynamic scheme can be linked to the types of actions that may be characteristic to a particular level. Accepting that every action in response to a review report is based on a (at least implicit) discussion about quality, we may say that 'running the maze' of this discussion from different starting points, at different levels, would produce predominantly different types of actions, as well.

In order to enhance the accountability and efficiency of such processes, it would be useful to arrive empirically at some reasonable period of deliberation for each of the paths in the model. Then, if deviations are detected, two sorts of issues could be addressed. First, delay may be attributable to organizational problems. Second, placing the debate on the wrong level, mostly by ignoring implicit arguments of (and by) some stakeholders and closing the discussion too early, may cause problems.

It could be argued that instrumental action is characteristic to the level of problem solving (A). Because strategic decision-making should not be influenced by suggestions about problems solving, the types of action are predominantly short-term oriented and aimed at immediate remedies.

It is a misunderstanding to expect only superficial, 'cosmetic' changes at this level. One professional programme had serious structural problems of which the staff was aware. The negative review report was not contested. However, internal conflicts prevented the programme from finding a remedy. With some outside help (a new dean and programme director), systematic and specific steps towards programme improvement were set.

At the level of problem finding (B) short-term conceptual type of action may be prevalent, as long as some discussion is taking place regarding accepting the need to address some newly detected problems.

A social sciences programme needed quite some time to realize that the review committee's criticism was not just 'nit-picking'. Rather, the report revealed a serious lack of structure and cohesion in the programme. Real results were achieved only after the programme staff considered and adopted the committee's systematic vision and placed the individual issues in its context. In the words of a respondent: "only then… we realized that we had to work on these issues that did not look important at first…"

At the level of goal definition (C), suggested changes imply enormous stakes for all those involved and have largely a non-incremental character. Roughly, they would fall into the categories of long-term instrumental and long-term conceptual action.

The best way to ensure conformist behaviour and 'window dressing' is to ignore the clash of views and deep beliefs at this level and force change without allowing for natural 'ripening' time. This is what happened in one example when the Inspector for higher education intervened in a most unfortunate effort to speed up action. The result was only more delay, more expensive reorganization and a very cynical attitude of the staff involved.

Finally, at the highest level of educational philosophy (D), crisis occurs and no action follows until some basic agreement can be reached. In an adequately functioning system, this should happen only occasionally. A too rigidly designed quality assurance cycle may not allow it at all.

As stated above, a strength of Dutch quality assurance system is its openness and space for constant renegotiating of the definition of quality. This implies that the procedural rules for creating a shared meaning of quality should be well defined. It is interesting to look also at the way visitation committees formulate and justify their recommendations, and to see whether making their views and beliefs more explicit would bring about different types of responses.

On the other hand, it could be argued, too many incidents of disagreements at the level of social values may indicate a crisis of the system as a whole. Interesting questions from this viewpoint are: how often do discussions about values occur in response to quality reviews? In what way do these discussions contribute to the evolution of thinking about good quality in education? Does the system allow for such an evolution, or does it encourage compliance, if necessary enforced by sanctions? In other words, does the process of quality assurance allow the actors to learn?

Ultimately, if the viability of the quality assurance system depends on building and fostering self-rule, we need the adequate analytical tools to describe this process when it occurs and to extract rules that would help managers and other stakeholders to promote it. Our research is only a first step in this direction.

M.I.Jeliazkova@cheps.utwente.nl

References
Brennan, J., L.C.J.Goedegebuure, D.F. Westerheijden, P.J.M.Weusthof & T.Shah, 1992, *Towards a Methodology for Comparative Quality Assessment in European Higher Education: A pilot study on economics in Germany, the Netherlands and the United Kingdom* (London, CNAA).
Fischer, F., 1985, *Politics, Values, and Public Policy: The problem of methodology* (Boulder, Colorado, Westview Press).
Fischer, F., 1995, E*valuating Public Policy* (Chicago, Nelson-Hall).
Frederiks, M.M.H., Westerheijden, D.F. & Weusthof, P.J.M., 1993, *Externe zorg en interne prikkel* (Den Haag:, Ministerie van OC&W).
Harvey L. & Green, D, 1993, 'Defining qualit*y'*, *Assessment and Evaluation in Higher Education*, 18, pp. 9–34

Jeliazkova, M.I & Westerheijden, D.F., 2000, *Het Zichtbare Eindresultaat* (Den Haag Algemene Rekenkamer)

Jeliazkova, M.. I., 2001, 'Is use a usable concept', in Westerheijden, D.F. & Verkleij, A., (Eds.), *The Next Generation of Quality Evaluation*, in press.

Scheele, J. P., P.A.M. Massen, & D.F. Westerheijden, 1998, *'To be continued'…Follow-up of Quality-Assurance in Higher Education* (Maarssen: Esevier/DeTijdstroom).

Vroeijnstein. T.I., 1994, 'Preparing for the second cycle. external quality assessment in Dutch universities', in Westerheijden, D.F., Brennan, J. & Maassen, P.A.M. (Eds.), 1994, *Changing Contexts of Quality Assessment. Recent trends in West European higher education* (Utrecht:, LEMMA).

Vught F.A. van & Westerheijden, D.F., 1993, *Quality Management and Quality Assurance in European Higher Education: Methods and mechanisms* (Luxembourg: Office for official publications of the EC)

Westerheijden, D.F., Brennan, J. & Maassen, P.A.M. (Eds.), 1994, *Changing Contexts of Quality Assessment. Recent trends in West European higher education* (Utrecht:, LEMMA)

Westerheijden, D.F., 1997, 'A Solid base for decisions: use of the VSNU research assessments in Dutch universities', *Higher Education*, 33: pp. 397–413.

Westerheijden, D.F. & Maassen, P., 1998, 'Following the follow-up', in Scheele, J. P., Maassen, P.A.M. & Westerheijden, D.F. (1998) *'To be continued'…Follow-up of Quality-Assurance in Higher Education*, (Maarssen: Elsevier/DeTijdstroomWeusthof)

Weusthof, P.J.M., 1994, *De interne kwaliteitszorg in het wetenschappelijk onderwijs* (Utrecht, Lemma).