Title: Lines of Change in the Discourse on Quality Assurance
An Overview of Some Studies into What Impacts Improvement

Paper presented to the

28th ANNUAL EAIR FORUM

30 August to 1 September 2006
Rome, Italy

Name of Author(s)

Don F. Westerheijden
Veerle Hulpiau
Kim Waeytens

Contact Details

CHEPS
University of Twente
+31 53 489 3263
d.f.westerheijden@utwente.nl

Key words

1. Assessment /Evaluation 2. Curriculum design and development 3. Higher education policy / development
4. Quality 5.

Lines of Change in the Discourse on Quality Assurance

An Overview of Some Studies into What Impacts Improvement

Don F. Westerheijden, Veerle Hulpiau & Kim Waeytens

Keywords: Assessment/Evaluation; Curriculum design and development Higher education policy / development; Quality;

Abstract

Attention shifted in recent years from design and implementation to use and usefulness of quality assurance. Scientific studies focus increasingly on quality assurance's impact on curricula and on individual teachers. Which (f)actors influence follow-up activities and what is their relation to improvement of education? One factor that we single out is teachers' experience of quality assurance and how that influences the climate for quality work in higher education institutions, against the backdrop of the social context in which quality assurance was introduced. The paper critically analyses a number of studies on the topic.

1 Introduction

Quality assurance is a theme with an unchallenged position in the discussions around higher education since the 1980s and 1990s, when the first supra-institutional (i.e. national) quality assurance schemes were developed and implemented. The experiences accumulated since then have always given rise to question, discuss and adapt those schemes. The current developments at the European level such as the introduction of three-cycle degree systems and the accreditation that often accompanies this restructuring keep this discourse relevant also today.

How did quality assurance schemes gain a foothold in European countries? How did national quality assurance schemes develop? What decided on their main aims and how did these develop? What has been the impact of quality assurance schemes at the levels of higher education institutions, curricula, and the individual lecturer[2]? Which factors affected these impacts? In this paper we approach these questions by describing some of the main trends, basing ourselves on existing research publications.

[2] We realise—and agree—that staff members in higher education institutions who are active in all kinds of activities besides teaching, may resent being called, reductively, 'lecturers', but for brevity we will do so nevertheless.

2 The Rise of Internal and External Quality assurance Schemes in Europe

In the first decades after World War II, the discourse on higher education was dominated worldwide by the wish to increase the numbers of higher education graduates, i.e. the issue of access and the consequent transformation from 'elite' to 'mass' higher education (Trow, 1974). Moreover, for reasons of social equity, higher education students were to hail from all strata of society (the issue of equality). The quantitative growth has been phenomenal—the social equality issue remains one with which countries are struggling—witnesses for instance the fact that more than half of the higher education institutions in Europe were founded after World War II. A consequence was that the ministries of education reached their 'span of control'. Another was that the claims for higher education on the state's budget increased rapidly, which in combination with the general economic downturns in the 1970s and 1980s gave rise to crisis-like situations. For at least these two reasons, the steering of higher education had to be changed: less expensive per student, and with lower transaction (control) cost for the ministries. Besides, the 1980s saw neo-liberalism become very popular, with its tendency to adopt practices from the business-world in the public sector. In combination with the surge in economic prominence of Japanese companies, whose successes were at least partly ascribed to innovations with regard to quality control, quality assurance became a new a popular policy instrument in higher education.

Until then, evaluation had been in use at the level of individual lecturers as a voluntary instrument for the improvement of their own teaching. But once external authorities began to demand quality assurance schemes, educational evaluation was integrated into the new schemes, which, however, served multiple, and other, goals than the improvement of individual courses.

Such developments were visible all across Western Europe[3] though to different degrees and at different speeds. Quality assurance schemes were introduced as national policy instruments first in France (1984), the United Kingdom (1985) and the Netherlands (1985). If we reduce the policy motives per country to one or two key words—almost like a caricature but not besides truth—the French government was first of all interested in reducing dysfunctional bureaucracy; the government in Great Britain wanted to achieve better linkage of higher education with the labour market and distrusted what it saw as the elitist attitude in universities (van Vught & Westerheijden, 1993). The developments in the Netherlands were a mixture of some of the French and British motives, because there the government tried to re-establish trust between the government and higher education institutions after a budget cut by introducing a new approach to steering. In France a new state agency was introduced, the Comité National d'Evaluation, bypassing the Ministry of National Education. The British University Grants Council was brought under closer gov-

[3] At least until 1990, the development in Central and Eastern Europe was separate from the one in the Western part of the continent, cf. (Schwarz & Westerheijden, 2004b).

ernment scrutiny. In the Netherlands, quality assurance came in the hands of the associations of higher education institutions. The developments in these 'pioneer countries' were disseminated by 'missionaries' who influenced other countries in Europe and worldwide. In 1990, Denmark followed as the number four in Western Europe. In 1991 the universities in Flanders (the Dutch-speaking part of Belgium, which is autonomous with regard to higher education) were made responsible for their quality assurance by the government and a year later the colleges[4] were subjected to a similar regulation. In 1991 quality assurance also came on the agenda of the European Union, first through an ad hoc working group (van Vught & Westerheijden, 1993), then through a pilot project (DG XXII, [1995]; Kern, 1998), which eventually led to the establishment of the European Network of Quality assessment Agencies (ENQA) in 2000. By that time, the majority of Western European countries had introduced formal quality assurance schemes (Schwarz & Westerheijden, 2004a).

Meanwhile, another course of developments had set in in Central and Eastern Europe. Until 1989 the higher education systems in those countries were elite in Trow's sense, which the Stalinist governments in the 1950s and later had fragmented into small, specialised higher education institutions. The fall of communism in 1989-1990 led to swift developments, sometimes (as in Romania) with the mushrooming of many small private higher education institutions. In all these countries the need arose for rapid transformation of study programmes to become ideologically free and adapted to a modern, post-industrial society. The instrument to stimulate that transformation and—where necessary—to control the establishment of private higher education institutions was a centralised accreditation agency, related to the state but staffed by academics. The most archetypal examples are to be found in the Czech Republic and Hungary (Campbell & Rozsnyai, 2002; Westerheijden & Sorensen, 1999).

For most higher education institutions, and maybe even for a number of ministers of higher education, the Bologna Declaration of June 1999 (European Ministers Responsible for Higher Education, 1999) came as a surprise. For quality assurance the ensuing 'Bologna process' meant a new context and a significant acceleration of developments. In the first place, the Declaration meant that quality assurance more than ever was placed in the framework of internationalisation. Second, in relation to internationalisation, other functions of quality assurance schemes were emphasised than before. These tendencies influenced the development of national quality assurance schemes in Europe.

---

[4] We use the term 'colleges' to denote non-university higher education institutions ('hogescholen' in Dutch).

3 Development of 'National' Quality Assurance Schemes

Most quality assurance schemes in higher education systems,[5] are based on the same four general principles (van Vught & Westerheijden, 1994):

1. a coordinating agency for the quality assurance scheme;
2. submission of a self-evaluation report by the unit to be evaluated;
3. site visit by *peers*;
4. (partly) public report on the evaluation results.

Within that very general model, there are systematic variations. To describe those variations, we use the phase model as described in (Jeliazkova & Westerheijden, 2002). The model describes the phases in the development of quality assurance schemes in relation to social issues regarding higher education. The model links the design of quality assurance schemes in each of those phases to the social demands it is to address.

The main assumption of the phase model is that quality assurance schemes operate in a social and policy context in which certain issues are dominant; the issues stand in a more or less hierarchical relation to each other. Once a more basic problem has been 'solved' to a degree that is satisfactory for the policy discourse, the next problem often already emerges. However, as long as a basic problem has not been 'solved', it is often useless to try to address a more esoteric one. For instance, there is little use in addressing issues of efficiency (drop-out, time to degree) as long as there are grave doubts about the level of knowledge achieved in study programmes in higher education. But as soon as a threshold level has been guaranteed, e.g. through accreditation, critics of higher education almost automatically will ask why 'producing' higher education graduates must cost so much tax money—or on the contrary they may wonder if the level can be maintained with continuing budget reductions. These are examples of what they called the 'external dynamics'. Moreover, the 'social dynamics' are important due to changes that may occur in the wider context such as higher education policies in general (introduction of new steering models), the general political climate (more, or less, emphasis on self-regulation), the economy (budget crisis) or demography (reduction of traditional student pool through smaller young cohorts). Finally, there are 'internal dynamics', resulting from the learning effects that results from the actors playing their part in subsequent rounds of quality assurance activities. This may be positive learning, as when lecturers may become interested in making the quality culture in their institution more sustainable once they have reached a certain degree of self-evaluation capacity. But it can also be negative learning, as when lec-

[5] For brevity, we will call supra-institutional quality assurance schemes in a higher education system 'national' schemes, although Flanders may serve as an example for the increasing tendency for higher education to become part of the authority of sub-national state governments (also e.g. the federal states in Germany, the regional governments in Spain, or the four parts of the previously United Kingdom).

turers and quality officers in higher education institutions learn to play the tricks of a quality assurance scheme without letting the 'internal life' of the actual quality of their teaching and research be affected by it ('window dressing').

Table 1 Phases in quality assurance systems

Problems	Role of quality assurance	Information base	Nature of external review
Phase 1: Serious doubts about educational standards.	Identifying sub-standard educational programmes.	Descriptive reports. Performance indicators.	Summative; accreditation, checking standards. Report to state.
Phase 2: Doubts about the efficiency of the higher education system and/or institutions.	a) Public accountability. b) Creating quality awareness in institutions.	Descriptive / strategic reports ('self-selling') covering: a) performance, b) procedures.	Ranking of institutions. One report to state and institutions. Identifying good practices.
Phase 3: Doubt about innovation capacity and quality assurance capacity of institutions.	Stimulate self-regulation capacity of institutions. Public accountability.	Self-evaluation reports about: a) procedures, b) performance.	Audit report to: – the institution – the state
Phase 4: Need to stimulate sustainable quality culture in institutions.	Split between: – improvement based on self-regulation; – public accountability.	Split between: – self-evaluative reports about processes and strategies based on SWOT and benchmarking; – self-reporting about performance indicators.	Split between: – audit report to the institution; – verifying data to be incorporated in public databases.
New challenge: Decreasing transparency across higher education systems.	Market regulation, i.e., informing clients (students, employers).	Performance indicators about 'products' (knowledge and skills of graduates).	

In each of these phases, quality assurance has a different function, which is best reached through different forms. Accreditation, for instance, is eminently able to assure a threshold level of quality, but it is much less suited to stimulate continuing improvement above the threshold level. Still, for instance the combined Flemish-Dutch accreditation organisation NVAO tries to unite these two functions by requiring a continuously functioning internal quality assurance system as one of its accreditation criteria. This example shows on the one hand that the table displays tendencies and emphases rather than mutually exclusive categories, and on the other hand that national quality assurance schemes address several functions simultaneously.

From the outset, both in Flanders and in the Netherlands the official goals of the national quality assurance scheme included both accountability and quality improvement.

The question if and to which extent such multiple goals are possible—or necessary—has long played a major role in the discourse on quality assurance in higher education (Kells, 1995; Vroeijenstijn, 1989; Vroeijenstijn & Acherman, 1990; Westerheijden, 1990) without leading to a theoretically grounded consensus. A practical consensus was reached, nevertheless: the political context required that both goals were included. What the exact balance between the two was—or had to be—was often not explicated, and changed over the years.

The actual developments of quality assurance in Flanders seem to illustrate the internal and external dynamics of the phase model well. In the policy document accompanying the decree on the universities of 1991 (Vlaamse Raad, 1991). Both goals, accountability and quality improvement, were mentioned there as requirements on the universities in exchange for their getting more autonomy. There was not a social problem about the basic level of study programmes in Flanders—just as in most other Western European countries. The situation at the outset rather could be compared to phases 2 or 3. The statements of the Flemish association of universities (VLIR), which coordinated the quality assurance scheme, might be read to put the emphasis (a little) more on quality improvement than on accountability. Whether this was just rhetoric to gain co-operation from universities and lecturers, or whether this was a genuine effort to address phase-3 issues, is not important at this place. It is important, however, hat the actual policies of the VLIR and of the universities focused on phase-3 issues of quality assurance capacities within the universities. When in 1998 an ad hoc committee chaired by Dutch education inspector Mertens was appointed by the minister of education to assess the functioning of the quality assurance scheme (Auditcommissie Kwaliteitszorg in het Academisch Onderwijs in Vlaanderen, 1998), this was by definition an expression of the accountability function as the committee made public statements about how universities had implemented internal quality assurance. Most attention in the committee's report, however, went to the effectiveness of what universities did about the quality of education; the committee searched actively for 'good practices' with the am to advise the Minister on strengthening the quality of academic education. Around the same time, the government's department for education initiated a programme for innovation of education, including quality improvement measures. This situation might be typified as on the border between phases 3 and 4.

A quit different mixture of internal and especially external dynamics prevailed in the Netherlands during the same period. In that country too, the official point of departure of the government was balanced attention to accountability and quality improvement (Ministerie van Onderwijs & Wetenschappen, 1985)—but always in that order. And there too, the coordinators of the quality assurance scheme, the associations of higher education institutions (VSNU and HBO-Council) turned the order of the two around: quality improvement first and accountability second. At least until the second half of the 1990s the internal dynamics as sketched in table 1 seemed to prevail and phase 3 seemed to grow to full stature; by *circa* 1998 the transition to phase 4 seemed to be dawning for many higher

education institutions. The protocol for the third round of evaluations in the universities, called *Quality Assessment Made to Measure* (VSNU, 1999), would confirm that interpretation. Meanwhile, in the college sector the external dynamics took over and made developments go in a different direction: especially employers began to criticise the quality assurance scheme because it did not give transparent information to them as 'end users' about the quality of study programmes. Moreover, colleges launched new study programmes continuously, decreasing the transparency of the offer of programmes and thereby increasing the need for transparent information. The HBO-Council in response started a pilot project for programme accreditation (Goedegebuure, Jeliazkova, Pothof, & Weusthof, 2002). At the same time, voices grew louder in national politics that the quality assurance scheme was too 'soft', too much a matter of mutually back-scratching academics. In brief, the trust in the accountability function waned, more as a result of the social dynamics of changing political insights than as a result of the internal or external dynamics of the phase model.

And that was the moment when Bologna happened. The bottom row of table 1 refers to it: the new problem that quality assurance had to address was transparency across higher education systems—globalisation suddenly took centre stage. To make European higher education a strong competitor on the world market for students, information about it to the main 'customers', students and employers, had to become clear and concise. The first step on that road was to cut drastically the 'jungle of degrees' that existed in Europe until then; restructuring towards the three-cycle bachelor—master—doctorate model was the way. Soon it appeared, though, that relabeling or cutting up existing degrees was not sufficient: the equivalence of study programmes also needed assurance. This was the new challenge for quality assurance. An option that responded to those needs could be the publication of performance indices per programme regarding what graduates knew and could (knowledge and skills), as indicated in the bottom right cell of table 1. However, the desires to protect consumers worldwide against substandard higher education providers—another phenomenon on the rise in the debate about globalisation and cross-border higher education (Knight, 2002; Vlk, 2006; van der Wende, 2001)—and to keep the information as concise as possible, combined to make policy-makers in many European countries opt increasingly for accreditation (Schwarz & Westerheijden, 2004b).[6] Logically, this would require a European consensus on minimum quality standards, which however was

[6] It might be contended that the currently rising popularity of multi-dimensional 'ranking' (or: information for choice of study locations as in the German-Austrian-Swiss CHE ranking [www.che-ranking.de] and its Dutch clone Studiekeuze123 [www.studiekeuze123.nl]) is a belated turn to the more sophisticated type of information provision foreseen in table 1. See e.g. (Van Dyke, 2005).

a political impossibility at the time.[7] This is not the place to expand on this question; let us return to the development of quality assurance schemes.

An aspect that was given attention at the system level only in very general terms was the second goal of quality improvement. As illustrated by the examples of Flanders and the Netherlands, this was seen by politicians in many countries as a secondary motive at best. It could be maintained, however, that also from a political perspective in the long run quality improvement is at least as important as accountability about the level of quality achieved, because a structured process of stimulating improvements and disseminating 'good practices' can ensure that higher education institutions perform better than if one awaits spontaneous dissemination of improvements among lecturers (which was the traditional way to innovate higher education). Besides, there also is a short-term interest in quality improvement, as this is the side of quality assurance that may attract lecturers to co-operate. And that in its turn is one of the prime conditions to make quality assurance schemes successful, because without loyal co-operation at the 'shop-floor' level it is often impossible to collect valid and relevant data. In any case, without lecturers' cooperation it is impossible to make data collection into a true self-*evaluation*, since evaluation requires reflection on the individual and collective work of the lecturers themselves.[8]

The considerations in the previous sentences give rise to the question to which extent quality assurance schemes actually stimulate quality improvement. We will address this question at two levels, viz. to which extent they stimulate quality management in higher education institutions and study programmes, and to which extent lecturers individually are stimulated to improve their practices.

4 Impact and Use of Evaluation Results at Institutional and Programme Levels

There has been not much research into the impact of supra-institutional quality assurance schemes on the quality of study programmes. One of the reasons of the small amount of studies may be that measuring impact is not self-evident. Quality assurance does not take place in a vacuum, but in organisations with complex structures in which developments in many areas take place more or less simultaneously and with varying degrees of cross-impacts—think of Weick's thesis of higher education institutions as 'loosely-coupled' systems (Weick, 1976). This makes it difficult to ascribe effects unambiguously to a single cause, such as quality assurance schemes (Stensaker, 2004). Yet some

---

7 The apparently successful series of *Tuning*-projects seems to indicate a 'workaround' achieving something like a European consensus through cooperation at the academic rather than the political level. See http://tuning.unideusto.org/tuningeu/.

8 A self-report without true evaluation was called a 'self-selling' report by (Frazer, 1997), see also table 1.

studies have tried to shed some light on, if not the impacts caused by supra-institutional quality assurance schemes, then at least on the extent to which higher education institutions and study programme organisations utilised evaluations and site visits.

Some of these studies gave quantitative information on the extent of utilisation. Thus (Weusthof, 1994) found that 87% of faculties used recommendations in external evaluation reports. Hulpiau & Waeijtens found that 49% of problems that had been detected were not followed up (Hulpiau & Waeytens, 2003).[9] Both figures need to be interpreted with some care. On the one hand, Weusthof included passive use (taking cognisance of results) and not only active use (taking actions to implement changes based on recommendations), and Hulpiau & Waeytens included follow-up intentions not just actual activities. On the other hand, while Weusthof conducted a survey, asking for all use of evaluation results in the faculty, Hulpiau & Waeytens performed a document analysis so that follow up that was not reported on paper was not detected by them.

In other studies, the impact of curriculum evaluations on organisational processes was the object of interest. The first round of external evaluations in many higher education institutions were used to set up internal monitoring systems for quality of the educational process, such as automated systems to track student progress through the programme, information systems on research output, or systems to keep curriculum overviews (syllabi) up to date. Generally, the thing most valued about the first round of external evaluations was the communication among the lecturers and researchers about their coexistence and co-operation. After some time and in more abstract terms, the main consequence seems to have been that organisational decision-making was centralised more in the wake of external evaluations (Stensaker, 2004)—bureaucratisation might be the undesired side-effect of this development. Especially after external evaluations the middle-level management (deans, vice-deans, etc.) tend to take (or be given) more responsibilities regarding follow-up of the recommendations made by the reviewers. This articulation of responsibilities at the same time increases the transparency of the organisation; the 'black box' is opened a little.

In Dutch universities it became visible, after a number of years of experience with external evaluations, that the information in the reports resulting from the evaluations was increasingly permeating the decision-making in all kinds of management choices, not just those directly connected to the evaluations, e.g. when deciding about who would get additional (temporary) teaching or research posts, or how to go about reorganisations (Westerheijden, 1997). For Flanders, parallel research results are not available, but the Mertens Committee mentioned above found that although the Flemish universities tended not to have sophisticated internal quality management systems (just like their Dutch counterparts), yet they seemed to incorporate evaluative information about quality

[9] A follow-up percentage around 50% was also reported in (Frederiks, 1996).

fairly well in their existing planning and control systems (Auditcommissie Kwaliteitszorg in het Academisch Onderwijs in Vlaanderen, 1998).

So external quality assurance schemes impact the universities as a whole; how do they affect the study programme level? Just before the turn of the century, the Dutch Audit Chamber (the highest-level controller of legality and effectiveness of governmental behaviour) commissioned a study into the quality improvement activities of study programmes in response to external evaluations. The study concluded that large-scale improvements took place especially in study programmes that were given relatively negative judgements by the external reviewers. For the study programmes that were given positive judgements, there seemed to be no impulse for further improvement (Jeliazkova & Westerheijden, 2000). There were no other significant correlations with organisational variables, except that it was necessary for the leadership of the institution to prioritise quality assurance to achieve effective follow-up of the external evaluation results.

As this study indicated, and as the other studies already suggested, effective follow-up depended on organisational conditions outside the quality management system in a strict sense. Hulpiau & Waeytens (2003) described explanatory factors for follow-up derived from a multi-dimensional model from organisational literature (Bolman & Deal, 1991). They propose four perspectives, with structural, cultural, human resources and political dimensions. From a structural perspective, several explanatory factors can be derived for follow-up of evaluations. There will be less follow-up if the structure does not clearly define who is responsible for achieving certain goals or for undertaking certain activities, if the activities that have to be undertaken are not within the brief of the study programme leaders or if there are no adapted structures within which certain activities can be undertaken. The extent to which education gets official recognition and, for instance, is reflected in the personnel and salary decisions also influences follow-up. This fits in the human resources perspective. The political dimensions stresses a number of other factors, such as the lack of means, or the existence of conflict about the concept of quality as hindrances to follow-up.[10] Looking at lacking follow-up from the cultural dimension, elements playing a role include the lack of a vision on education, academic freedom, the dominance of the research function in the organisational climate, etc.

5 Impact and Use of Evaluation Results at the Level of Individual Lecturers

Not just the higher education institutions or study programmes as a whole, but also the individual lecturers are affected in many ways by initiatives in the domain of quality assurance, and of evaluation of education especially. Parallel to the manner in which quality

[10] Within this perspective, a communication model was elaborated in (Jeliazkova, 2001): only if the debate in the unit evaluated remains, or has returned to, the level of technical solutions, can follow-up occur, but not as long as the conflict extends to values.

assurance schemes have developed over time, the attention in research developed from the impact of such instruments as a whole to their impact on the educational functioning of the lecturers.

The evaluation of educational quality at the level of individual lecturers traces its roots back to the 1920s and saw significant developments with "student feedback questionnaires" and "student ratings" in the 1970s and 1980s (Gosling & D'Andrea, 2001; Marsh, 1987; Wachtel, 1998). Obviously, then, research on evaluation of education at the individual lecturers' level centred on that. The use of "student feedback questionnaires" and "student ratings" became one of the worldwide most popular methods to chart the quality of education (Hendry & Dean, 2002; Kember, Leung, & Kwan, 2002). This student feedback served several goals: quality improvement, providing data for decisions in personnel matters, and informing stakeholders, including students, about quality of education. Some authors explicitly added a fourth function, namely collection of data for educational research (Marsh, 1987), although in fact not much use was made of them for the latter purpose (Nasser & Fresko, 2002). Since the 1970s, educationalist circles gave much attention to the quality of the data collection instruments: reliability and validity, and influence of bias factors on student feedback results (Greenwald, 1997; Marsh & Roche, 1997; van Os, 1999). Although the discussion still continues, the general verdict about student feedback remains positive (Hendry & Dean, 2002; Marsh & Roche, 1997; McKeachie, 1997).

Nevertheless, especially since the middle of the 1990s this almost-monopoly of "student feedback questionnaires" and "student ratings" was called increasingly into question. This tendency seems to parallel the shift in the discourse in the scientific literature from "quality control" and "quality assurance" [11] to terms like "quality development", "quality enhancement" and "quality culture" (Gosling & D'Andrea, 2001). The criticism of the former set of terms is connected largely to their questioning the impact and effectiveness of the quality instruments for quality improvement.

Relatively little is known about the impact of student questionnaires. This may have to do with the difficulty to establish a causal relationship between student questionnaires and quality improvement, because of the many possibly intervening variables (Marsh, 1987).

Studies, both quantitative and qualitative, into the impact of student feedback usually look at quality improvement at the short or middle term. From quantitative studies it appears that relative strengths and weaknesses of individual lecturers are fairly stable (Centra, 1993; Hativa, 1996; Kember, Leung, & Kwan, 2002). Qualitative studies empha-

[11] The fact that in the previous part of this paper we still used "quality assurance" reflects the fact that in much of the *practice* of *external* quality evaluation it remains the prevailing term. Moreover, it was chosen here as a fairly general term, without too many (negative) connotations.

sise that the follow-up of evaluation at this level often remains limited to rather limited measures (Hendry & Dean, 2002).

In the literature several elements are proposed that influence follow-up of educational evaluations by lecturers. Internal and external factors can be distinguished. Internal factors point to elements that have to do with lecturers' attitudes towards quality assurance in general and student feedback questionnaires especially. External factors point to elements in the working context of the lecturers.

5.1 Internal Factors

Lecturers' attitudes towards quality assurance initiatives are an underexposed aspect in research on quality assurance (Nasser & Fresko, 2002; Newton, 2000). It is important, however, because the lecturers are the ones who first and foremost are responsible for the quality of education. Whether educational evaluation has impact, is in their hands, eventually. In line with previous research on how lecturers handle policy and change, (Trowler, 1998), Newton (2000) studied empirically that lecturers react very differently to quality policies. Some will accept it pragmatically, or view it as a meaningless ritual. Others distrust it or feel it as a discouragement. Research into reactions of lecturers to student feedback questionnaires resulted in divergent (Nasser & Fresko, 2002), but sometimes also surprising results, for instance that lecturers in research-oriented universities are more open to summative use of student feedback data than those in more teaching-oriented institutions.

A first internal factor influencing the use of student feedback data by lecturers is the extent to which they accept the instruments, which is mainly dependent on their perception of the reliability and validity of the instruments (Laughton, 2003). If lecturers doubt the reliability of questionnaire results, they will tend to distance themselves from the results and will not engage in follow-up activities. The extent to which they see the questionnaires as valid, also influences lecturers' attitudes to the results. An important factor in this respect is whether lecturers are of the opinion that the questionnaires address what they see as the core of the feedback they need to improve their teaching (or aspects of it). Thus, questionnaires with closed questions are often criticised for their necessarily focusing on what is measurable in numerical data and what is visible for students (Johnson, 2000; Laughton, 2003). Often, the reserves lecturers feel with regard to student feedback questionnaires is associated with bias factors. Lecturers regularly are convinced that that elements unconnected to the quality aspect addressed in the questionnaires influence students' opinions, such as the students' examination results, difficulty of the course, or students' interest in the course. Bias factors like these are object of many studies, yet evidence for a significant and unambiguous influence on evaluation results is not available (Marsh & Roche, 1997; McKeachie, 1997). Nevertheless, lecturers' subjective perceptions may substantially influence their interpretation of student feedback information and in that way may hinder follow-up.

The extent to which lecturers recognise evaluation results is a second major influencing factor. Follow-up correlates positively with lecturers recognizing the results as truthful.[12] Truthfulness has several facets. For one thing, the evaluation must connect to lecturers' opinions about good education. Besides, the evaluation must connect to their teaching practice (Kember, Leung, & Kwan, 2002). Against the backdrop of the need to relate to lecturers' opinions on educational quality, the use of standardised and general questionnaires has been questioned. Moreover, use of such questionnaires may induce the feeling among lecturers that quality of education is a fixed concept about which they themselves have no opinion or control (Johnson, 2000; Kember, Leung, & Kwan, 2002). As a consequence, also the improvement of education may then be seen as outside of the responsibility of the lecturers, rather as something belonging to 'management' or 'quality experts'. Equally problematical is when lecturers do not have an explicit vision on what they want to achieve with their teaching. In that perspective, having an explicit vision on quality of education is an important condition for evaluations to have impact (Hendry & Dean, 2002).

Equally important for serious follow-up by lecturers is the extent to which they can make *prima facie* interpretations of the data. Not only the 'user-friendly' presentation of data facilitates effortless interpretation, but so too do the presence of standards and the abstractness of the questions. Quantitative results usually are presented in the form of averages, possibly with standard deviations or frequency distributions of the answers. In many cases, lecturers do not possess (implicit or explicit) standards for interpretation of data. Several authors point out that lecturers find comments to open questions very useful to aid in interpretation of the data (Cashin, 1990; Hendry & Dean, 2002; Nasser & Fresko, 2002). These authors emphasise that students' comments must fulfil a number of criteria to be taken seriously, such as realism, specificity, and descriptive and non-judgemental language.

Finally, the existence of a discrepancy between how lecturers think they have taught and the evaluation by students hinders follow-up (Centra, 1993).

5.2 External Factors

As far as external factors are concerned, the literature emphasises the support for lecturers in interpreting evaluation results and for follow-up. In general, the organisational climate in which education and evaluation take place are regarded as important.

Changes and improvements take place especially if student feedback goes accompanied by some form of support of lecturers for interpreting the results (Hendry & Dean, 2002; Kember, Leung, & Kwan, 2002; Marsh, 1987; Nasser & Fresko, 2002). Effective sup-

---

[12] This is the same mechanism underlying the 'maze' leading to follow-up at programme level reported in (Jeliazkova, 2001).

port can take several forms and can be given by different persons, from individual coaching by peers or educational specialists to formal trainings.

By far most attention is given in the literature, though, on the general organisational climate in which education and evaluations take place, and in which lecturers are stimulated to work on the quality of education. This brings us to the issue of the appreciation for education shown by the institution's management. The extent to which lecturers have the impression that education is valued and rewarded determines follow-up. However, few have this impression, studies show (Boyer, 1990; Gibbs, 1995; Ramsden & Martin, 1996). A major, though not the only mechanism is obviously the extent to which quality of teaching plays an effective role in appointment, tenure and promotion decisions for lecturers. In this respect it is problematic that many higher education institutions stress that they value quality of education in such decisions, but that the large majority of lecturers do not perceive it that way (Gibbs, 1999). But the problem is not limited to perceptions: the literature increasingly questions how quality of education can be incorporated in personnel decisions. This does not mean that the voice of students is ignored, but ever more there are pleas to integrate it into a well thought-out and balanced mechanism (Magin, 1998; McAlpine & Harris, 2002). This requires a careful definition of quality of education, and that it is not limited, for instance, to that which is easily visible, but rather addresses "the invisible, hidden aspects of the teaching iceberg" (McAlpine & Harris, 2002, p. 9). This also implies that education is not just viewed form the students' perspective, but also form the lecturer's perspective, from peers, alumni, etc.

In this section it has become clear that a complex interplay of factors determines the occurrence of follow-up of evaluations of education. Some of those factors are more readily amenable to manipulation (e.g. organisational factors) than others (e.g. the culture in the institution, or the room for manoeuvre left by government policy). Anyway, quality improvement depends on much more than the limited area of quality assurance.

6 Discussion and Conclusions

In this paper we intended to give an overview of changing insights and research results with regard to quality assurance of the educational function in higher education. We sketched two parallel lines to do so. The first was concerned with changes over time. In fact, there was not a single line over time: the dynamics at different levels moved at different speeds. The internal dynamics of learning effects may take quality assurance in different directions, possibly under the influence of external and social dynamics. Positive learning effects (in other words: continuous improvement) or negative learning effects (window dressing) may come to dominate. Only the external dynamics at the systems level seem to follow a single, continuous line, leading to a hierarchy of problems that will be addressed, and temporarily solved, in a specific order. But that gradual development

may be interrupted by the (as far as we know: unpredictable) emergence of issues in what we call the social dynamics. The differences in the developments of so similar higher education systems as Flanders and the Netherlands illustrated the latter point. An even clearer illustration gives the Bologna process, which in these two higher education systems (and many others) led to a thorough revision of the policy agenda for higher education at all levels. For quality assurance this meant the emergence of accreditation, in conjunction with renewed attention for the accountability goal rather than for the quality improvement goal. From the institutional perspective, this seems to put emphasis on ensuring that accreditation of the study programmes is 'achieved', given the potentially drastic consequences of non-accreditation.

The second line concerned our observation that the attention for quality of education differed systematically across the levels of the higher education system. Briefly, the more one goes down from the system level to the 'chalk face' level of actual teaching, the more the balance seems to go from accountability as the main aim of quality assurance to improvement of educational quality.

Doubtlessly, research on quality assurance has evolved and reflects a changing discourse. While attention at first was devoted mostly to the emergence, design and implementation of quality assurance systems, more recent studies address the sense and usefulness of activities taking place within those systems. Also the impact of evaluations of education have increasingly come centre stage in research. At the same time, attention shifts from the systems level, to the programme level, and to the individual level. Although still modest in sise, the literature on the role of individual lecturers in quality assurance is growing. The fact that it took so long for the attention for their role to be recognised widely is worrying, as eventually lecturers have the most direct influence on the quality of education.

What were the limitations that we encountered in our overview of the literature? Obviously, due to the recent introduction, empirical studies on the impacts of programme accreditation (in contrast to the previous visiting committees scheme) are absent until now. Such studies would be important to substantiate the statement that accreditation has diminished attention to quality improvement at all levels of the higher education system. Almost ironically we may add that such studies must start soon, or programme accreditation will have ended already, as there are signs that there will be only one round of such accreditations in the Netherlands (until 2009) and Flanders (until 2012).

More in general, the question into the actual influence of (changes in) quality assurance and quality management at the level of higher education institutions on changes (improvements?) in education has not been answered definitely. The methodological problem that quality assurance never appears in isolation, but is always bound to a, for many reasons, inherently rapidly changing context. Ascription of changes to one of the many simultaneously occurring potential causes is then a complicated issue, necessitating substantial and sophisticated empirical research.

Methodically less complicated, and perhaps even more relevant than the previous issue, is that the empirical base for statements about how quality assurance impacts on the programme and—even more—the individual levels, is very limited.

Finally, we must observe that the student experience in all of this remains almost *terra incognita*. Students are expected to give feedback to the education they receive, were even allowed to participate in visiting committees in some quality assurance schemes, but how they viewed quality assurance and if they had the opinion that it benefited them, is a question still awaiting an empirical answer.

For the quality of education—and for the quality and effectiveness of our work in higher education either as lecturer, as researcher or as quality officer—it is of great importance to find answers to these and similar questions soon.

References

Auditcommissie Kwaliteitszorg in het Academisch Onderwijs in Vlaanderen. (1998). *Aandacht voor kwaliteit in de Vlaamse universiteiten*. Brussel: Ministerie van de Vlaamse Gemeenschap, Departement Onderwijs.

Bolman, L. G., & Deal, T. E. (1991). Leadership and management effectiveness: A multi-frame, multi-sector analysis. *Human Resource Management, 30*(4), 509-534.

Boyer, E. L. (1990). *Scholarship reconsidered: priorities for the professorate*. Princeton, N.J.: Carnegie Foundation for the advancement of teaching.

Campbell, C., & Rozsnyai, K. (2002). *Quality Assurance and the Development of Course Programmes*. Bucharest: CEPES-UNESCO.

Cashin, W. E. (1990). *Student ratings of teaching: recommendations for use* (No. IDEA Paper No. 22). Manhattan, Ks. : Kansas State University, Center for Faculty Evaluation and Development.

Centra, J. E. (1993). *Determining faculty effectiveness*. San Francisco: Jossey-Bass.

DG XXII. ([1995]). *Information note on the results of the European pilot project for evaluating quality in higher education*. [Brussels]: European Commission, DG XXII 'Education, Training and Youth'.

European Ministers Responsible for Higher Education. (1999). *Joint declaration of the European Ministers of Education Convened in Bologna on the 19th of June 1999*. Bologna.

Frazer, M. (1997). Report on the modalities of external evaluation of higher education in Europe: 1995-1997. *Higher Education in Europe, 12*(3), 349-401.

Frederiks, M. M. H. (1996). *Beslissen over kwaliteit*. Utrecht: De tijdstroom.

Gibbs, G. (1995). The relationship between quality in research and quality in teaching. *Quality in Higher Education, 1*(2), 147-157.

Gibbs, G. (1999). Improving teaching, learning and assessment. *Journal of Geography in Higher Education, 23*(2), 147-155.

Goedegebuure, L. C. J., Jeliazkova, M., Pothof, F., & Weusthof, P. J. M. (2002). *Alle begin is moeilijk : Evaluatie van de proefaccreditering HBO*. Enschede: CHEPS, Universiteit Twente.

Gosling, D., & D'Andrea, V.-M. (2001). Quality Development: a new concept for higher education. *Quality in Higher Education, 7*(1), 7-17.

Greenwald, A. G. (1997). Validity concerns and usefulness of student ratings of instruction. *American Psychologist, 52*(11), 1182-1186.

Hativa, N. (1996). University instructors' ratings profiles: Stability over time, and disciplinary differences. *Research in higher education, 37*(3), 341-365.

Hendry, G. D., & Dean, S. J. (2002). Accountability, evaluation of teaching and expertise in higher education. *International Journal for Academic Development, 7*(1), 75-82.

Hulpiau, V., & Waeytens, K. (2003). Improving quality of education: What makes it actually work? A case study. In C. Prichard & P. R. Trowler (Eds.), *Realizing qualitative research into higher education* (pp. 145-169). Aldershot: Ashgate Publishing.

Jeliazkova, M. (2001). Running the maze: interpreting external review recommendations. *Quality in Higher Education, 8*(1), 89-96.

Jeliazkova, M., & Westerheijden, D. F. (2000). *Het zichtbare eindresultaat*. Den Haag: Algemene Rekenkamer.

Jeliazkova, M., & Westerheijden, D. F. (2002). Systemic adaptation to a changing environment: Towards a next generation of quality assurance models. *Higher Education, 44*(3-4), 433-448.

Johnson, R. (2000). The authority of the student evaluation questionnaire. *Teaching in Higher Education, 5*(4), 419-434.

Kells, H. R. (1995). *Self-Study Processes* (4th ed.). New York: ACE/McMillan.

Kember, D., Leung, D. Y. P., & Kwan, K. P. (2002). Does the Use of Student Feedback Questionnaires Improve the Overall Quality of Teaching? *Assessment & Evaluation in Higher Education, 27*(5), 411-425.

Kern, B. (1998). A European Union Perspective on Follow Up. In J. P. Scheele, P. A. M. Maassen & D. F. Westerheijden (Eds.), *To be continued...: Follow-up of quality assurance in higher education* (pp. 39-63). Maarssen: Elsevier/De Tijdstroom.

Knight, J. (2002). Trade Creep: Implications of GATS for Higher Education Policy. *International Higher Education, 28*(Summer), 5-7.

Laughton, D. (2003). Why was the QAA Approach to teaching quality assessment rejected by academics in UK HE? *Assessment & evaluation in higher education, 28*(3), 309-321.

Magin, D. J. (1998). Rewarding good teaching: a matter of demonstrated proficiency or documented achievement? *International journal for academic development, 3*(2), 124-135.

Marsh, H. W. (1987). Students' evaluations of university teaching: research findings, methodological issues, and directions for future research. *International journal of educational research, 11*, 253-388.

Marsh, H. W., & Roche, L. A. (1997). Making students' evaluations of teaching effectiveness effective. The critical issues of validity, bias and utility. *American Psychologist, 52*(11), 1187-1197.

McAlpine, L., & Harris, R. (2002). Evaluating teaching effectiveness and teaching improvement: a language for institutional policies and academic development practices. *International Journal for Academic Development, 7*(1), 7-19.

McKeachie, W. J. (1997). Student ratings. The validity of use. *American Psychologist, 52*(11), 1218-1225.

Ministerie van Onderwijs & Wetenschappen. (1985). *Hoger Onderwijs: Autonomie en Kwaliteit*: Tweede Kamer, vergaderjaar 1985-1986, 19253, nr. 2.

Nasser, F., & Fresko, B. (2002). Faculty views of student evaluation of college teaching. *Assessment & Evaluation in Higher Education, 27*(2), 188-198.

Newton, J. (2000). Feeding the beast or improving quality? Academics' perceptions of quality assurance and quality monitoring. *Quality in Higher Education, 6*(2), 153-162.

van Os, W. (1999). *Bruikbaarheid en effectiviteit van studentenoordelen over het onderwijs*. Enschede: Universiteit Twente.

Ramsden, P., & Martin, E. (1996). Recognition of good university teaching: Policies from an Australian Study. *Studies in Higher Education, 21*(3), 299-316.

Schwarz, S., & Westerheijden, D. F. (2004a). Accreditation in the Framework of Evaluation Activities: A Comparative Study in the European Higher Education Area. In S. Schwarz & D. F. Westerheijden (Eds.), *Accreditation and Evaluation in the European Higher Education Area* (pp. 1-41). Dordrecht: Kluwer Academic Publishers.

Schwarz, S., & Westerheijden, D. F. (Eds.). (2004b). *Accreditation and Evaluation in the European Higher Education Area*. Dordrecht: Kluwer Academic Publishers.

Stensaker, B. (2004). *The transformation of organisational identities: Interpretations of policies concerning the quality of teaching and learning in Norwegiam higher education*. Enschede: CHEPS, Center for Higher Education and Policy Studies.

Trow, M. (1974). Problems in the Transition from Elite to Mass Higher Education. In OECD (Ed.), *Policies for Higher Education* (pp. 51-101). Paris: OECD.

Trowler, P. (1998). *Academics responding to change. New higher education frameworks and academic cultures*. Buckingham: SRHE & Open University.

Van Dyke, N. (2005). Twenty Years of University Report Cards. *Higher Education in Europe, 30*(2), 103-125.

Vlaamse Raad. (1991). *Memorie van Toelichting bij het decreet betreffende de universiteiten in de Vlaamse Gemeenschap* (No. Stuk 502 (1990-1991) – Nr.1).

Vlk, A. (2006). *Higher Education and GATS: Regulatory consequences and stakeholders' responses*. Enschede: CHEPS, University of Twente.

Vroeijenstijn, A. I. (1989). *Autonomy and assurance of quality: two sides of one coin.* Paper presented at the International Conference on Assessing Quality in HE, Cambridge.

Vroeijenstijn, A. I., & Acherman, J. A. (1990). Control Oriented versus Improvement Oriented Quality Assessment. In L. C. J. Goedegebuure, P. A. M. Maassen & D. F. Westerheijden (Eds.), *Peer Review and Performance Indicators: Quality assessment in British and Dutch higher education*. Utrecht: Lemma.

VSNU. (1999). *Quality Assessment Made to Measure: Protocol for External Assessment of Educational Programmes 2000-2005*. Utrecht: VSNU.

van Vught, F. A., & Westerheijden, D. F. (1993). *Quality Management and Quality Assurance in European Higher Education: Methods and Mechanisms*. Luxembourg: Office for Official Publications of the Commission of the European Communities.

van Vught, F. A., & Westerheijden, D. F. (1994). Towards a general model of quality assessment in higher education. *Higher Education, 28*, 355-371.

Wachtel, H. K. (1998). Student evaluation of college teaching effectiveness: a brief review. *Assessment & Evaluation in Higher Education, 23*(2), 191-211.

Weick, K. F. (1976). Educational Organizations as Loosely Coupled Systems. *Administrative Science Quarterly, 21*.

van der Wende, M. C. (2001). Internationalisation Policies: About New Trends and contrasting Paradigms. *Higher Education Policy, 14*(3), 249-259.

Westerheijden, D. F. (1990). Peers, Performance, and Power: Quality assessment in the Netherlands. In L. C. J. Goedegebuure, P. A. M. Maassen & D. F. Westerheijden (Eds.), *Peer Review and Performance Indicators: Quality assessment in British and Dutch higher education* (pp. 183-207). Utrecht: Lemma.

Westerheijden, D. F. (1997). A solid base for decisions: Use of the VSNU research evaluations in Dutch universities. *Higher Education, 33*(4), 397-413.

Westerheijden, D. F., & Sorensen, K. (1999). People on a Bridge: Central European higher education institutions in a storm of reform. In B. W. A. Jongbloed, P. A. M. Maassen & G. Neave (Eds.), *From the Eye of the storm: Higher education's changing institution* (pp. 13-38). Dordrecht: Kluwer Academic Publishers.

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