

The role of peer review in Norwegian quality assurance: potential consequences for excellence and diversity

Liv Langfeldt · Bjørn Stensaker · Lee Harvey · Jeroen Huisman · Don F. Westerheijden

Published online: 30 June 2009
© Springer Science+Business Media B.V. 2009

Abstract The article analyses the role of peer review within broader external quality assurance schemes. Based on an analytical framework emphasising that modern quality assurance schemes are designed as a balancing act between standardised guidelines and professional judgement, the article uses data from a recent evaluation of NOKUT, the Norwegian Quality Assurance Agency, to investigate whether and how the peer review process has maintained its central role in quality assurance, not least with respect to promoting excellence and diversity. The findings indicate that what is presented as judgements based on peer expertise, turn out to be a rather technical process in which pre-defined rigid criteria and standards are imperative. In the conclusion, the role of peer review is discussed in relation to developments in European higher education.

Keywords Peer review · Quality assurance · European standards and guidelines · Norway · Excellence · Diversity

Introduction

Quality assurance in higher education is a growing activity and attracts much attention. The quality assurance activities may be undertaken for various interrelated purposes, including (1) ensuring that higher education institutions, their procedures or their specific

L. Langfeldt (✉) · B. Stensaker
NIFU STEP, Wergelandsvn. 7, 0167 Oslo, Norway
e-mail: liv.langfeldt@nifustep.no

L. Harvey
CBS Learning Lab, Copenhagen Business School, Kilevej 14 A, 2000 Frederiksberg, Denmark

J. Huisman
International Centre for Higher Education Management, University of Bath, Claverton Down,
BA2 7AY Bath, UK

D. F. Westerheijden
CHEPS, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands

study programmes fulfil required standards, (2) as basis for assigning institutional or programme accreditations, (3) for closing down substandard programmes, and (4) for informing potential students and other stakeholders about the quality of institutions and educations. In general the initiation of quality assurance will be embedded in formal requirements. The overall aim is still most likely to be some kind of improvement; to enhance the quality of the education and of the national higher education system as such. In addition to the ability to enhance quality, there are, however, also other issues attached to national quality assurance systems, such as the concern that the degree of institutional diversity (the variety of institutions within a system, see e.g. Huisman 2000) in higher education may be affected by the characteristics of the quality assurance system. Besides, we look at scholarly diversity, defined in his paper as the variety of theoretical and methodical approaches within a single area of knowledge.

The latter concern is especially relevant with respect to recent understandings of the role higher education should play in society, and the importance of diversity in this process. According to van Vught (2008), diversity in higher education benefits society in a variety of ways. A diverse higher education system enables the sector to cater for the needs of students and society, it stimulates mobility and innovation, enhances the relevance of higher education with respect to the labour market. There is also a hierarchical dimension of diversity: diverse systems allows for the coexistence of elite and mass education. In the current process of globalisation and internationalisation of higher education, it is believed that finding space for elite education and excellence is an important element in attracting the most talented students and staff and, as such, strengthening the competitive advantage of a given country (Castells 2000). However, how diversity and excellence are maintained or supported, can be dependent on how quality assurance systems are designed and how they function. Given the emphasis of the Bologna process on transparency and transferability, one could at least question the degree to which current quality assurance schemes are able to find room for issues of diversity and excellence. Hence, this paper investigates more closely whether current national quality assurance schemes are adequate policy measures for promoting excellence, and how such systems may affect diversity. More specifically, the paper focuses on the role of the peer-review process within such national quality assurance schemes. Peer review is here generally understood as the evaluation of work or performance by experts (in the same field), or inter-collegial, professional, evaluation.

According to the *European standards and guidelines* external quality assurance in higher education should be based on predefined criteria and conducted by groups of experts including student members (ENQA 2009). The empirical scope of the paper is the Norwegian quality assurance system, a system fully in line with the ENQA guidelines (Langfeldt et al. 2008a). The data applied were collected for a recent external evaluation of the Norwegian Agency for Quality Assurance in Education (NOKUT), conducted by the authors of the paper (Langfeldt et al. 2008a, see below for elaboration of the data).

In this article, three main characteristics of the Norwegian quality assurance system are discussed in relation to impact on excellence and diversity. First, which elements of information and what processes are seen as the most important for the reviewers when forming their opinions. Second, analyses of the balance between formal standards and scholarly discretion in the assessment process. Third, exploration of the degree of agreement within review panels and how this is reached. The conclusion discusses the implications of the findings for the Norwegian quality assurance scheme and for the role of peer review in quality assurance in general.

The double logic of quality assurance

Quality assurance schemes have traditionally been perceived as a professional process in which peer review is the key dimension. This emphasis is rooted in the belief that specialised knowledge can only be properly assessed and validated by specialists within that specified knowledge area (Chubin and Hackett 1990). Within higher education, Europe has witnessed a new development in quality assurance in the last decades. Written rules and identification of specific indicators and academic standards have been developed in parallel to the peer-review process. Sometimes these written rules are just meant as guidance to the peer review process, while on other occasions they are meant to ensure equity and provide limits to peer discretion.

The tendency to specify written rules, indicators and guidelines for how knowledge should be assessed, and the establishment of specialised bodies and agencies with responsibilities for these processes could be interpreted as a sign that the whole field of quality assurance is maturing (Stensaker 2007). However, it could also be interpreted as an indicator of how higher education is being transformed along more bureaucratic lines. The weight given to accreditation in quality assurance, the emergence of European standards and guidelines in this area, and the recent establishment of a European Register for quality assurance agencies, are all developments that can be said to relate to a bureaucratic logic (Power 2007). The logic of these processes is to emphasise standards (Brunsson et al. 2000), and as such to reduce the influence of the professional authority (see also Evetts 2002).

Although peer review may be used for checking academic standards and, as such, be related to a bureaucratic logic, studies show that peer review as an assessment process may not be the most appropriate tool in this respect. As Ceci and Peters (1982), Westerheijden (1990, 1991), Stensaker (1998) and Langfeldt (2001) all emphasise, peer review of any academic performance, considered on its own, is a process with a varying degree of reliability, which goes against the logic of bureaucracy with its emphasis on rule-following and consistency ('repeatability'). Peer review may be more geared to identifying excellence and uniqueness than to conformance to minimum standards (see also Prøitz et al. 2004). Following from this, one could argue that peer review is a process that may enhance not only excellence, but also diversity.¹

In this way, modern quality assurance could be conceived as a process that is trying to balance two different logics—bureaucracy and peers' judgements—in a joint procedure. How this balancing act is conducted in practice is an interesting question for analysis, as quality assurance is becoming a regulatory instrument with potentially dramatic consequences for the sector (Evetts 2002).

Methodology and sources

The evaluation of NOKUT proceeded from a research-based approach and collected a broad set of qualitative and quantitative data from a wide variety of sources and stakeholders. NOKUT's practices, experiences and views were studied based on: (1) Self-evaluation reports from NOKUT; (2) site visit to NOKUT and interviews with NOKUT

¹ On the other hand, peer review is sometimes accused of being conservative and stifling innovativeness as assessments are done by well-established experts rejecting ideas differing from their own, i.e. counteracting diversity (as emphasised by the literature referred to and by Langfeldt 2002, p. 39).

Table 1 Number of respondents and informants, by group

Group of respondent/informant	Surveys	Site visits/ interviews institutions	Interviews stakeholders	Site visits/ interviews NOKUT	Total
NOKUT staff and leadership				28	28
Members NOKUT panels (excl students)	335				335
Students	89	10	5	2	106
Staff and leadership at evaluated institutions	326	46			372
Other stakeholders	6		13	3	22
Total	756	56	18	33	863

The table shows number of obtained survey respondents, not the requested sample. Survey response rates: Students, staff and leadership at evaluated higher education institutions: 64% (students only 35%); NOKUT panel members: 80%; Leadership at evaluated vocational schools: 84%

leadership and staff; (3) acts and regulations describing NOKUT's standards, criteria and procedures; (4) studies of NOKUT's evaluation and accreditation reports.

Stakeholders' experiences and views were studied based on: (1) Site visits to institutions subjected to NOKUT evaluations and accreditations; (2) survey of staff, students and leadership at higher education institutions that underwent NOKUT evaluations and accreditations; (3) survey of vocational schools; (4) surveys to all members of NOKUT's accreditation and evaluation panels; (5) interviews with stakeholders (the Norwegian Ministry of Education and Research and national interest organisations for students, academic staff, and business and industry). In total, 107 informants were interviewed, and 756 respondents replied to the surveys (Table 1). The paper is primarily based on analyses of the survey among staff, students and leadership at higher education institutions and the survey among members of accreditation and evaluation panels.

Key features of the quality assurance system in Norwegian higher education

Norwegian higher education is subject to an extensive quality assurance (QA) system, based on: (1) local QA-systems, cyclical external audits of these local QA-systems; (2) institutional self-accreditation of study programmes and external accreditations allocating authority for institutional self-accreditation; (3) external accreditation of study programmes (in cases without institutional self-accreditation authority); and (4) non-regular external reaccreditation of study programmes (unrelated to local accreditation authority).

- (1) *The institutions' internal quality assurance systems* are evaluated by NOKUT at least every sixth year, in so-called *quality audits*. The Ministerial regulations state that internal QA-systems shall cover all processes of importance for the quality of the study programmes, ensure continuous improvements, reveal deficiencies in quality and provide satisfactory documentation of the quality work.² If the QA-system does not pass the audit there is an immediate follow-up procedure for a second audit. The

² Regulations no. 1040 of 8 September 2005 from the Ministry of Education and Research: Regulations governing accreditation, evaluation and approval pursuant to the Norwegian Universities and University Colleges Act.

result of a second failure would be withdrawal of the authority to establish new study programmes. Moreover, an approved quality assurance system is required when applying for institutional accreditation (item (2) below). So far there is no case of an institution not passing the second audit.

- (2) *Institutional accreditations* divide the Norwegian higher education institutions into three different categories, each with a different degree of autonomy in establishing study programmes: accredited university colleges, specialised university institutions and universities. In short, the following is required for all accreditation categories, whereas key differences between the categories concern the number of accredited study programmes at higher levels and the standard of the research activities³:
- Education, research and development and dissemination as primary activity;
 - An organisational model, facilities, infrastructure and services which support its primary activities;
 - Research and development activities;
 - A sufficient body of teaching staff with appropriate qualifications in key subject areas of their study programmes;
 - A satisfactory academic library;
 - Participation in national and international networks.
- (3) *Accreditation of new study programmes*. Universities are fully authorised to establish study programmes at all levels, specialised university institutions are authorised to establish study programmes at all levels within specific fields, accredited university colleges are authorised to establish study programmes at Bachelor level, as well as programmes at Master level in fields where it has obtained an accredited PhD programme, whereas non-accredited institutions are not authorised to establish any programmes on its own. For establishing study programmes which the institution is not authorised to establish on its own, accreditation from NOKUT is needed. The *accreditation criteria* relate to the plan for the programme, the academic staff involved, infrastructure, quality assurance and international cooperation. The specific criteria vary with the level of the study programme, but the same criteria apply regardless of field of study. For example, for Bachelor studies at least 20% of the staff assigned to the programme are required to have senior lecturer or professorial status, whereas for PhD programmes at least 50% are required to hold full professorships and the remaining associate professorships.
- (4) *Reaccreditations*. NOKUT may re-evaluate any previously awarded accreditation. Procedures for the reaccreditations of study programmes are more thorough than for new programmes, including additional data collection to study the academic level and outcome of the programmes, as well as site visits and student representation on the panels. The formal standards and criteria are the same as for accreditation of new programmes, except that the regulations state that emphasis is to be placed on the study programme's academic standards and documented results.⁴

NOKUT was established in 2002 and mandated to establish a quality assurance system as described in the Acts and regulations relating to Norwegian universities and university colleges. The Act states that "NOKUT shall be a professional autonomous state body

³ E.g. for university accreditation, five accredited Master Programmes and PhD programmes in four different subject areas are required.

⁴ In addition to national reaccreditations there are national evaluations of study programmes with no formal link to the accreditation system. Both the reaccreditations and evaluations are performed by NOKUT.

which, by means of accreditation and evaluation, shall monitor the quality of Norwegian institutions that provide higher education” and that the “Accreditation and evaluation activities shall be designed in such a way that the institutions can benefit from them in the course of their quality assurance and development work.”⁵ All evaluations and accreditation are conducted by expert panels appointed and organised by NOKUT, with different rules for the panel composition and different review criteria applying for the specific kind of audit, accreditation or evaluation activity.

Turning to the question of nourishing and maintaining excellence, there is little explicit attention to these matters in the design of the Norwegian QA-system. The system aims at ensuring good quality in all study programmes at all institutions, more than encouraging a concentration on excellence. The criteria focus on a ‘sufficient’ and ‘satisfactory’ quality level, not at ‘excellence’. Moreover, as far as non-universities aim at obtaining university status, the requirement for PhD programmes in four different subject areas may result in a national fragmentation of doctoral education that may undermine excellence. On the other hand, the internal QA-systems are intended to ensure continuous improvements. If they are able to do so, also for the study programmes that already are reasonably good, the QA-system may still promote excellence.

Concerning diversity, inbuilt incentives to obtain university status may threaten *institutional* diversity. University status implies a higher degree of academic autonomy and authority. There is consequently an incentive for the university colleges to prepare their organisation for obtaining university status. Concerning *scholarly* diversity the picture is somewhat more complex. In general, QA-systems that apply the same criteria to all subject fields may endanger diversity. The Norwegian QA-systems still allows for diversity through the general, qualitative criteria that can be interpreted and adjusted to various fields of study. However, there are also some quantitative criteria that do not allow much flexibility and consequently may reduce diversity. A specific concern here is a possible academic drift resulting from subjecting all professional study programmes to the same criteria as academic study programmes. In sum, the system opens for an academic drift: similar criteria applied to the different programmes of different institutions may lead to the blurring of institutions’ distinctive characters.

Standards, excellence and diversity in Norwegian QA

The recent evaluation of NOKUT provides extensive data on the perceived impact of NOKUT’s activity and the Norwegian quality assurance system. Below, these impacts are discussed in relation to excellence and diversity.

The respondents within the higher education institutions identified several positive impacts of the current Norwegian quality assurance. The system helps to put quality issues on the institutional agenda (77%), creates equal framework conditions for public and private institutions (54%), and provides students with information about the quality of education (56%). Hence, the system is assessed as having a quite positive impact on agenda setting, equal framework conditions and information. Respondents are more sceptical about diversity; more than half (53%) answered ‘Don’t know’, 19% answer ‘Positive impact’ and 5% answered that they think the QA-system has negative impact on stimulating diversity or division of labour within the national education system (Table 2).

⁵ http://www.regjeringen.no/upload/kilde/kd/reg/2006/0031/ddd/pdfv/273037-loven_higher_education_act_norway_010405.pdf.

Table 2 Evaluatees' (HEIs) perceptions of impacts of system level, percent

How would you rate the impact of the current national quality assurance system in Norway with respect to:	No impact	Positive impact	Negative impact	Don't know	<i>N</i>
Putting quality issues on the institutional agenda?	7.4	77.2	1.5	13.8	325
Providing students with information about the quality of the education?	18.8	56.0	2.8	22.5	325
Creating equal framework conditions for public and private institutions?	5.2	54.2	1.8	38.8	325
Ensure the public accountability of the education sector?	17.4	49.4	4.3	28.9	322
Facilitate internationalisation of education?	24.8	42.2	1.2	31.7	322
Developing a more solid basis for resource allocation/distribution within the higher educational system in Norway?	22.1	26.2	3.4	48.3	321
Stimulate diversity/division of labour within the national education system	24.1	18.5	4.9	52.5	324

Source: Survey for the evaluation of NOKUT 2007, encompassing leadership, staff and students at higher education institutions exposed to NOKUT accreditation/evaluation

Many stakeholders did not seem to have insight into, or a firm opinion about, the impact of NOKUT's activities and decisions (or more specifically the impacts of the institutional accreditation regime) on the scholarly national landscape.

The Norwegian quality assurance system is expected to maintain system diversity, operationalised by the different institutional categories in the Norwegian higher education. On the other hand, the current regulations invite non-university institutions to try to reach university status, which may imply putting diversity and other important functions of the higher education system under pressure. Current incentives are geared towards being a research-intensive higher education institution, which may undervalue the role of teaching and the qualities of professional/vocational studies. It is notable that three of the most frequently-mentioned negative effects of NOKUT's evaluations relate to internal resource allocation and priorities for research (Table 3). Moreover, a government commission has recently concluded that the current Norwegian accreditation regime leads to fragmentation of Master and PhD education and inhibits specialisation and task division, not least when attempting to acquire the necessary study programmes for obtaining university status.⁶ The current step-by-step design of the institutional accreditation scheme (a number of PhD programmes are accredited separately before a higher education institution is able to apply for university status) may moreover entail randomness in what forms the institutional specialisation and profile. A result may be that, even if the various institutions intend to maintain or develop a unique profile, the diversity in the institutional landscape is reduced.

Whereas a large majority of the higher education respondents find that the Norwegian quality assurance systems helps to put quality issues on the institutional agenda (Table 2), the figures are much lower when it comes to effects on the quality of the education itself. Asked about the effects a specific evaluation or accreditation (the latest of those the respondent had been involved in), 28% think it had a positive effect on the quality of the education, whereas 23% think it had partly positive and partly negative

⁶ NOU 2008:3. Sett under ett. Ny struktur i høyere utdanning. Oslo: Norges offentlige utredninger. ISBN 978-82-583-0953-3. (NOU = *Official Norwegian Reports* accessible at <http://www.regjeringen.no/nb/dok/NOUer>).

Table 3 Evaluatees' (HEIs) opinions about results and follow up, percent

So far, has the evaluation/accreditation had any positive or negative effects at/for your institution concerning:	No effects	Clearly negative	Partly negative/ partly positive	Clearly positive	Don't know	<i>N</i>
a. Effects on the quality of the education/teaching	24.5	1.5	23.2	28.1	22.6	327
b. Effects on the quality assurance of the education/teaching	13.2	.9	27.3	36.8	21.8	326
c. Effects on the scholarly reputation of the institution	15.6	4.9	18.7	31.0	29.8	326
d. Effects on the scholarly discussion on learning and teaching	19.4	2.2	22.8	35.4	20.3	325
e. Effects on staff engagement in learning and teaching questions	21.2	.9	28.8	31.6	17.5	326
f. Effects on student involvement in learning and teaching questions	29.1	.9	19.9	23.5	26.6	327
g. Effects on the development of new courses/programs	29.4	2.5	19.3	25.2	23.6	326
h. Effects on the criteria for the recruitment of teaching staff	30.2	1.2	15.4	21.8	31.4	325
i. Effects on the criteria for the recruitment of research staff	30.2	1.2	14.2	21.6	32.7	324
j. Effects on the criteria for the recruitment of administrative staff	38.2	.6	13.2	9.8	38.2	325
k. Effects on new routines and procedures	14.1	1.5	23.3	40.8	20.2	326
l. Effects on the governance structures of the institution	27.1	1.2	19.7	20.0	32.0	325
m. Effects on the internal resource allocation	29.2	3.4	24.0	13.2	30.2	325
n. Effects on the cooperation between administrative and scholarly staff	29.0	2.5	25.6	18.8	24.1	324
o. Effects on setting priorities for research	27.4	2.8	17.2	25.2	27.4	325
p. Effects on the resources and facilities for research	34.2	3.4	15.7	18.5	28.3	325
q. Effects on the supervision of research students	31.9	.6	15.5	16.7	35.3	323

Source: Survey for the evaluation of NOKUT 2007, encompassing leadership, staff and students at higher education institutions exposed to NOKUT accreditation/evaluation

effects, 25% think it had no effect, 2% think it had negative effect and 23% answer 'Don't know' (Table 3).

Looking at Table 3, the most noticeable positive effect can be found with respect to creating new routines and procedures, an indication that the quality assurance process may emphasise standards rather than diversity and excellence. The fact that those exposed to external quality assurance tend to focus on bureaucracy, establishment of routines and procedures, and emphasising the agenda-setting functions of quality assurance more than perhaps the 'real' effects are very much in line with earlier studies in the field (Brennan and Shah 2000; Newton 2002). Still, there are far fewer studies investigating whether these views are also shared by those performing the evaluation—the reviewers. This is further explored in the next section.

Information and processes emphasised by the reviewers

Although new routines and procedures can be associated with bureaucracy they can, of course, also advance excellence and diversity. One possibility is that excellence and diversity are emphasised during the visits of the peer-review panels, and through the engagement and dialogue that the visit is intended to stimulate. Earlier studies show that site visits by peer-review panels are important for both the institutions and the panels themselves (Stensaker 1998). Such a visit makes it possible for the panels to provide informal comments and advice that are beyond the focus of the formal evaluation process. Furthermore, the site visits are also arenas where the institutions may put their own issues and strategies on the agenda, and where the power of the peer-review panels to control the agenda is challenged by those evaluated.

One way of analysing to what extent such processes have taken place during the site visits, is by reviewing what panel members think are the important information sources and arenas forming their conclusions. As shown in Table 4, the panels tend to think that their own experiences and discussions within the panels are the most important processes forming and influencing the conclusions. Moreover, the panels do seem to emphasise the standards and criteria set by NOKUT (77%) as very important information sources. Surprisingly, as many as 11% of panel members performing site visits considered them unimportant. The importance of the self-evaluation reports seems somewhat lower (62%). In sum, the panel members' replies draw a picture of a process characterised *both* by bureaucracy and standards and by professional discretion. However, since some of the standards and criteria used by NOKUT are very detailed and specific, the actual room for scholarly discretion may be restricted (Langfeldt et al. 2008b).

Balancing standards and scholarly discretion?

Interviews with stakeholders in the Norwegian higher education system show that both the NOKUT criteria and the assessments are disputed by stakeholders. This seems partly dependent on the technicalities of some of the criteria but also on their focus. Since a number of standards and criteria specify minimum demands for teaching resources, infrastructures and the quality assurance system, assessments of learning and teaching outcomes may receive less attention. Moreover, the interviews also exposed different views about the preferable nature of criteria; general, adaptable criteria open to interpretation by the review

Table 4 Panel members' opinions about the importance of information sources, percent

How would you rate the importance of the following information sources for the conclusions of the panel's report?	Unimportant	Somewhat important	Very important	Not applicable	<i>N</i>
NOKUT's standards and criteria	1.2	20.8	76.8	1.2	341
NOKUT's preparing seminars or meetings*	(9) 5.1	(51) 27.8	(39) 21.3	45.8	334
The self assessment report(s)/ documentation/application(s)	0.3	33.3	61.6	4.8	336
Visits to the institution(s) under review*	(11) 5.4	(15) 7.2	(74) 35.4	52.0	333
Your own experiences and competences	0.3	16.2	82.9	0.6	340
The discussions within the panel	1.5	12.6	82.6	3.2	340

Source: Survey for the evaluation of NOKUT 2007 encompassing the members of the NOKUT-panels

*In brackets: Percentages when 'Not applicable' are excluded from the calculations

panel or unambiguous criteria without room for interpretation. Part of the NOKUT criteria are perceived as too vague by the stakeholders, whereas other criteria are criticised for being quantitative and not properly addressing the quality of the education.

An example of a disputed criterion is the quantitative requirement for staff qualifications for the accreditation of study programmes. The required percentage of staff with high formal academic qualifications is higher for Master studies than for Bachelor studies, and even higher for PhD studies. However, the same standards apply regardless of the subject field and regardless of the academic versus professional orientation of the study programme. Many panel members expressed serious concerns regarding this criterion.⁷ Applying the same ‘academic’ teaching staff requirements to all studies, in philosophy and physics as well as in nursing, law and social work, may endanger the individual characters of the various studies, implying less diversity in higher education. Concerning professional studies, the current standards and criteria may lead to academic drift and greater distance from professional practice, due to strong emphasis on formal qualifications and research competence of academic staff. One of the most notable impacts of NOKUT’s reaccreditations so far is increased attention to academic qualifications for Norwegian nursing study staff (Langfeldt et al. 2008b). Giving priority to formal standards and criteria such as these may also have a negative impact on indigenous forms of excellence as good professional teaching and learning environments may be discredited and study programmes even shut down due to failure to comply with such standardised, quantifiable criteria.

On the other hand, bureaucratic standards may be defended on the basis of equality and the necessity for threshold quality levels. Hence, one could expect that procedures emphasising formal standards and criteria would also increase the level of precision with respect to agreeing on minimum standards. Comparing the formal recommendations and decisions made by the panels with their informal assessments indicate that this may not always be the case. Although there are difficulties in studying the mismatch between the formal and informal assessments as we only have data on the informal assessments for the macro-level and not for the specific accreditations and audits, there are signs of discrepancy between what is formally stated and what panel members informally think (see Tables 5 and 6).

The discrepancy is clearest for the accreditation of vocational school study programmes. Here 37% of the panel members answer that the programme was (clearly or just) below the required level (Table 5), whereas only 9% of the programme applications in the period between 2005 and 2007 ended up with a negative conclusion (Table 6). Even when including the 12% of the applications that ended up without a conclusion (withdrawn or dismissed) in the ‘below-required-level’ category, the discrepancy is 16 percentage points. For the higher education programmes the discrepancy is less pronounced (Langfeldt et al. 2008b).

A possible explanation for this discrepancy is that keeping to the NOKUT standards and criteria results in a more positive conclusion than keeping to their intuitive academic assessments of what ought to be the required standard. It is hard to promote quality, and even harder to promote excellence, if some of those not fulfilling perceived scholarly standards are accepted/pass the examination. There are still cases where evaluatees state that they received valuable input for improvements from the review panel. There are, however,

⁷ Moreover, teaching staff may be counted in different ways, and while NOKUT tries to be consistent and gives detailed instructions to the evaluations panels, the institutions struggled to find out whether they fulfilled the criterion. There is no shared understanding about how staff qualifications ought to be assessed, and NOKUT has not yet been able to settle on criteria for assessing staff qualifications that the evaluatees find meaningful and adequate.

Table 5 Panel members' opinions on the quality of the object under review, by type of accreditation percent

What was your opinion about the quality of the institution/programme/ education under review?	Clearly above the required level	Just above the required level	Just below required level	Clearly below the required level	No opinion	<i>N</i>
Recognition of tertiary vocational education	9.4	47.9	19.8	16.7	6.3	96
Accreditation of higher education institutions	25.0	45.0	12.5	17.5		40
Accreditation/revision of accreditation of higher education programmes and courses	15.4	38.2	23.5	18.4	4.4	136
Total	14.7	42.6	20.6	17.6	4.4	272

Source: Survey for the evaluation of NOKUT 2007 encompassing the members of the NOKUT-panels

Table 6 Overview of accreditations of vocational school programmes, by year, percent

Year*	Approved	Not approved	Withdrawn or dismissed	<i>N</i>
2005	26.2	2.4	71.4	42
2006	77.1	13.9	9.0	599
2007**	85.6	3.6	10.8	388
Total	78.6	9.4	12.0	1,048

Source: Table provided by NOKUT

* Year the application was approved/disapproved or the case ended

** As by 7 Sept 07. A large number of applications were under review and are not included in the calculations

indications that such input is more effective for weaker units than for the potentially excellent ones (Langfeldt et al. 2008b). Those who pass the formal requirements have no obligation to follow up the recommendations from the review panel. Moreover, there are often no resources set aside for such follow up.

Reaching agreement within the review panels

A final dimension that can be of importance for promoting excellence and diversity in external quality assurance, is the degree to which disagreements and alternative views are given weight when trying to reach a decision within the review panels. One could argue that reviewer disagreement on preferable methods/procedures, as well as on accomplishments and results, is an indication that (excellence and) diversity, especially scholarly diversity, is prioritised in the process.

The literature on the dynamics of peer-review processes in academia finds limited agreement between the assessments of different reviewers (low inter-reviewer reliability),⁸

⁸ Cole et al. 1981 and Ceci and Peters 1982 are among the classics here, studying inter-reviewer reliability in the review of grant proposals and journal publications. The kind of quasi-experimental design employed in these classic studies is difficult to use when studying expert panel versions of peer review (as practiced by NOKUT), but reviewer disagreement is generally confirmed also in studies of expert panel evaluations (Langfeldt 2002).

and underlines that the outcome of review may depend on the composition and organisation of the panels (Langfeldt 2001). Hence, high level of disagreement in reaching a decision may suggest that the composition of the panel is characterised by various academic perspectives and positions, which at least increases the chances of emphasising diversity.⁹ Analysing the composition of panels through intellectual biographies of members indicating their scholarly diversity would be a more direct measure of potential disagreement; the data did however not allow for such an analysis.

In their input to the evaluation of NOKUT, some higher education institutions reported that they found the evaluation of their institution to be unduly influenced by the composition of the review panel. These cases were dealt with by NOKUT to assure that the final conclusions were not influenced by such bias. Still, the inherent problem with decisions based on expert review and scholarly discretion is to define and detect undue biases, and at the same time give the experts the room needed for discretion to use their expertise. Expert review is needed for adjusting assessments to diverse contexts, and using a broad set of reviewers allows for nuanced conclusions taking the various points of view into consideration. On the other hand, there is a risk that the panel does not represent the relevant and normally-accepted scholarly positions¹⁰ or that the panel dynamics result in conclusions that do not represent such positions. Using review panels, therefore, demands specific competences related to composing and organising panels, and ensuring the panels' role in maintaining diversity and nuanced assessments; and some efforts to obtain the balance of mainstream and diverse opinions needed for the given policy context.

The survey of the NOKUT panel members yielded an interesting finding relating to diversity. In total, 91% answered that reaching agreement in the panel was very easy or relatively easy (Table 7). That the decision making in the panel is easy, may indicate that the criteria give little room for discretion or that the panels are relatively homogeneous with regard to the members' scholarly backgrounds and opinions. Put otherwise, if the evidence provided is perceived to give an unambiguous answer to whether the programme or institution fulfils the required standard, this may be due to common interpretative perspectives of the reviewers or to the required standard being unambiguously defined. Neither homogeneous panels nor unambiguous criteria are likely to promote diversity.

Moreover, as mentioned earlier, the conclusions reached by NOKUT are not always in accordance with the reviewers' opinions. Taken together, this indicates that scholarly assessments and discretion play a relatively small role in NOKUT's quality assurance. In relation to quality assurance, scholarly assessments by review panels have both positive and negative attributes; they provide room for nuanced assessments and maintaining diversity but do not promote consistency and equal treatment. NOKUT seems clearly concerned about the latter attribute, and takes care to delimit the panels' scope for discretion. The panels' role in maintaining (or promoting) diversity consequently seems small.

⁹ On the other hand, the panel may very well consist of members with very different opinions without any evidence of such disagreements in the conclusions. In cases of tacit disagreement—or more explicit but concealed compromises—the conclusions may consist of vague assessments (Langfeldt 2002), and consequently promoting neither excellence nor diversity.

¹⁰ Or have conflicts of interest.

Table 7 Panel members' opinions on the processes of reaching agreement by type of evaluation/accreditation, percent

How would you characterise the expert panel's internal processes of reaching agreement on the conclusions and recommendations of its report?	Very difficult to reach agreement	Somewhat difficult to reach agreement	Relativity easy to reach agreement	Very easy to reach agreement	<i>N</i>
a. Recognition of tertiary vocational education		8.1	62.6	29.3	99
b. Evaluation of internal quality assurance in higher education institutions	2.6	5.3	57.9	34.2	38
c. Accreditation of higher education institutions	7.3	14.6	36.6	41.5	41
d. Accreditation/revision of accreditation of higher education programmes and courses	1.4	2.9	35.3	60.4	139
e. Evaluation of the quality of higher education		23.8	47.6	28.6	21
Total	1.8	7.4	46.7	44.1	338

Source: Survey for the evaluation of NOKUT 2007, encompassing members of NOKUT's accreditation and evaluation panels, 2003–2007

Concluding comments

Scholarly assessment is supposed to be a key characteristic of the current Norwegian accreditation regime. All evaluations and accreditations are carried out by review panels appointed by NOKUT. In general, scholarly assessment by visiting panels could be an adequate method both for promoting excellence and diversity. Expert panels with high competence in the relevant disciplines and in quality enhancement, provide a basis for advice for improving education as well as taking care of institutional and scholarly diversity.

However, this study indicates that the evaluation panels in the Norwegian quality assurance scheme are assigned a technical rather than a scholarly role. The standards and criteria provided by NOKUT are given much weight in the review process and it seems that the reviewers believe there is limited leeway for interpreting the criteria and for adjusting them to the specific study programmes. These restrictions on the panels relate mostly to accreditation as fulfilling minimum requirements, the sanctions if they are not fulfilled, and the resulting need for fairness and equal treatment of the institutions. Hence, the primary task of the panels seems to relate to an analysis of the evidence collected from the institutions and decisions as to whether the requirements defined in government regulations and by NOKUT are fulfilled. This result seems independent of the fact that several standards and criteria defined still demand substantial scholarly discretion, for example, "The academic staff shall engage in active research/artistic development work with proper academic breadth at a high international level".¹¹ The more precise and quantifiable standards and criteria seem to be given much weight compared to those that are more qualitative and that are more open to scholarly discretion. In a system and long-term

¹¹ Regulations Relating to Standards and Criteria for Accreditation of Programme of Study and Criteria for Accreditation of Institutions in Norwegian Higher Education, NOKUT 2006, <http://www.nokut.no/sw482.asp>.

perspective, this may in turn lead to reduced emphasis on excellence and diversity in Norwegian higher education.

The study triggers two more general questions that are also of relevance from an international perspective. First, there are strong indications that the whole peer-review process is under pressure by the emergence of the European standards and guidelines for quality assurance, by the emergence of new national indicators and benchmarks intended to guide the whole review process, and not least by the emergence of qualification frameworks in higher education. The implications of the latter frameworks for quality assurance are not yet seen but there is emerging evidence that the indicators and descriptors used in these frameworks are intended to become central in future quality assurance schemes (Bienefeld et al. 2008). One could argue that giving more space and emphasis to dimensions other than those provided by experts in the field, is a logical development following the increasing interest in higher education from a number of different stakeholders. The paradox is that this might actually reduce diversity, and might have negative consequences for excellence, as the weight given to standards and indicators reduces the space for the unexpected and the innovative.

Second, the findings could be seen as an argument for re-opening one of the classical debates in external quality assurance, namely how such processes should be designed and set-up (cf van Vught and Westerheijden 1994). While one could argue that the role of peer review in quality assurance has, for a long time, been poorly specified, with ample room for scholarly discretion, the data suggest that currently professional autonomy is reduced. If it is further reduced, one could question what role peer review will play in future quality assurance schemes. If the role is further reduced, one could also imagine a new type of “peers” appearing having more generic than discipline-based skills and tasks. In both cases, one can predict that a long-term effect of the diminishing importance of peer review might be that recruitment and engagement in such processes by academics becomes even more difficult (it is already becoming difficult due to the sheer size of the schemes). If peers’ roles become more technical than substantial, one runs the risk of not being able to recruit the best and brightest to these assignments and this could have dramatic impact on furthering excellence, but also on diversity in higher education. On the other hand, if the role of peer review in external evaluations is reduced to technicalities, ‘bad’ peers cannot do much harm—nor could ‘good’ peers do much good. It then comes down to the smartness of standards and criteria with regard to stimulating excellence and diversity. This issue has not been taken up in this study but it may be telling that the question never arose: criteria and standards in the NOKUT exercises are bureaucratic. They concern passing the threshold but do not have inbuilt improvement stimuli, such as requiring quality improvement-oriented quality management in higher education institutions rather than quality control, or requiring proof of excellent scholarship from the staff involved.

References

- Bienefeld, S., Harris, N., Helle, E., Hopbach, A., Maguire, B., Michalk, B., et al. (2008). *Quality assurance and qualification frameworks*. Helsinki: ENQA workshop report 5.
- Brennan, J., & Shah, T. (2000). *Managing quality in higher education. An international perspective on institutional assessment and change*. Buckingham: SRHE/Open University Press.
- Brunsson, N., Jacobsson, B., et al. (2000). *A world of standards*. Oxford: Oxford University Press.
- Castells, M. (2000). Towards a sociology of the network society. *Contemporary Sociology - A Journal of Reviews*, 29(5), 670–693.
- Ceci, S. J., & Peters, D. P. (1982). Peer review: A study of reliability. *Change*, 14(6), 44–48.

- Cole, S., Cole, J. R., & Simon, G. A. (1981). Chance and consensus in peer review. *Science*, 214, 881–886. 20 November.
- Chubin, D. E., & Hackett, E. J. (1990). *Peerless science*. New York: State University of New York Press.
- ENQA. (2009). Standards and guidelines for quality assurance in the European higher education area. ([http://www.enqa.eu/files/ESG_3edition%20\(2\).pdf](http://www.enqa.eu/files/ESG_3edition%20(2).pdf)). Accessed 25 June 2009.
- Evetts, J. (2002). New directions in state and international professional occupations: Discretionary decision-making and acquired regulation. *Work, Employment and Society*, 16(2), 341–353.
- Huisman, J. (2000). Higher education institutions: As different as chalk and cheese? *Higher Education Policy*, 13(1), 41–53.
- Langfeldt, L., Harvey, L., Huisman, J., Westerheijden, D., & Stensaker, B. (2008a). *Evaluation of NOKUT—the norwegian agency for quality assurance in education. Report 1: NOKUT and the European standards and guidelines for external quality assurance agencies*. Oslo: Ministry of Education and Research.
- Langfeldt, L., Harvey, L., Huisman, J., Westerheijden, D., & Stensaker, B. (2008b). *Evaluation of NOKUT—the norwegian agency for quality assurance in education. Report 2: NOKUT's national role*. Oslo: Ministry of Education and Research.
- Langfeldt, L. (2001). The decision-making constraints and processes of grant peer review, and their effects on the review outcome. *Social Studies of Science*, 31(6), 820–841.
- Langfeldt, L. (2002). *Decision-making in expert panels evaluating research. Constraints, processes and bias*. Oslo: Unipub. ISBN 82-570-4806-2.
- Newton, J. (2002). *From policy to reality: enhancing quality is a messy business*. LTSN Generic Centre/The learning and teaching support network, (www.ltsn.ac.uk/genericcentre/projects/qaa/enhancement).
- Power, M. (2007). *Organized uncertainty: Designing a world of risk management*. Oxford: Oxford University Press.
- Prøitz, T. S., Stensaker, B., & Harvey, L. (2004). Accreditation, standards and diversity. An analysis of EQUIS-accreditation reports. *Assessment & Evaluation in Higher Education*, 29(6), 735–750.
- Stensaker, B. (1998). Assessing the assessors: A comparative study. *Quality Assurance in Education*, 6(4), 205–211.
- Stensaker, B. (2007). Quality as fashion. Exploring the translation of a management idea into higher education. In D. F. Westerheijden, B. Stensaker, & M. J. Rosa (Eds.), *Quality assurance in higher education* (pp. 99–118). Dordrecht: Springer.
- Van Vught, F. (2008). Mission diversity and reputation in higher education. *Higher Education Policy*, 21(2), 151–174.
- van Vught, F. A., & Westerheijden, D. F. (1994). Towards a general model of quality assessment in higher education. *Higher Education*, 28(3), 355–371.
- 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 education*. Utrecht: Lemma.
- Westerheijden, D. F. (1991). *Promises, problems and pitfalls of peer review: The use of peer review in external quality assurance in higher education*. In T. W. Banta (Ed.), *Proceedings of the third international conference on assessing quality in higher education* (pp. 130–142). Knoxville, TN: University of Tennessee.