Educational Evaluation around the World

An International Anthology

THE DANISH
EVALUATION INSTITUTE

Educational Evaluation around the World

An International Anthology

2003

THE DANISH
EVALUATION INSTITUTE

Contents

Educational Evaluation around the World

Printed by Vester Kopi

This publication can be ordered from:

The Danish Evaluation Institute
Østbanegade 55.3
DK-2100 Copenhagen Ø

T +45 35 55 01 01
F +45 35 55 10 11

E eva@eva.dk
H www.eva.dk

ISBN 87-7958-132-3

Preface

With this anthology the Danish Evaluation Institute introduces different approaches to educational evaluation from several parts of the world. The purpose of the anthology is to disseminate knowledge about educational evaluation and to provide inspiration for the development and innovation of methods in this field.

All the contributors come from countries with considerable experience within the field of educational evaluation, and they represent different educational sectors. Allow me here to express my sincere gratitude to all the contributors.

Furthermore, I acknowledge the staff members from my institute who have been engaged in developing the concept of the anthology and in the process of editing. Deputy Director Dorte Kristoffersen has been in charge of the project, and the evaluation officers Signe Ploug Hansen, Tommy Hansen, and Rikke Sørup saw the project to the door.

I present this anthology with great pleasure and the expectation that readers will find the new information useful and inspirational.

Christian Thune
Executive Director
The Danish Evaluation Institute

Introduction

With this anthology the Danish Evaluation Institute (EVA) focuses, within the field of educational evaluations, on relations between values, purposes, objects, and methods in a global context. Its contributions come from various parts of the world and represent different educational sectors and evaluation approaches.

The first seeds of the anthology were sowed in 2002. The idea, however, can be dated back to the establishment of EVA in 1999. Besides the main task of implementing systematic evaluations of education and teaching with a view to ensuring and developing the quality of teaching and education in Denmark, EVA is also a national centre of knowledge. This means that EVA is charged with gathering national and international experience within evaluation and quality development. This anthology is one concrete example of this responsibility.

A number of evaluation approaches applied to different educational sectors have been mapped out, embodying various groupings of countries. However, EVA has no knowledge of previous global surveys focussing particularly on relations between values, purposes, objects, and methods for educational evaluations.

At the outset, our expectation was that the contributions to the anthology would make it possible to establish a typology covering combinations of values, objects, purposes and methods. One main problem to be elucidated is the question of what constitutes values and an approach to method. In other words:

Is an establishment of a typology based on, for example, educational sectors or countries feasible? Is it, for example, the case that aspects regarding the school sector are being evaluated based on the same values or a common set of methods across borders? Or is it more likely that each individual country evaluates educational aspects within all educational sectors in the same way? Is it, alternatively, the case that a typology, in fact, has to be generated right across both educational sectors and countries? Or can the picture be so diffuse that it is not feasible to establish a proper typology?

These questions were for operational purposes set out in the following hypotheses:

1. The control and owner relationships of the individual educational sectors determine the values and methods on which evaluations in a given sector are based. To be more explicit, all educational systems under the same control and ownership are evaluated with reference to the same values using the same methods.
2. The culture of a given educational sector determines how it is evaluated, irrespective of the formal control and ownership aspects.
3. The institutional foundation of the evaluating body determines the values underlying its evaluations and choice of methods.
4. Certain evaluation methods develop their own lives and achieve such a degree of status in certain countries that their appropriate application to a given educational sector is not questioned.

In order to be able to analyse the validity of these hypotheses, we asked a number of representatives of organisations responsible for educational evaluations at national and regional levels to contribute to the anthology with a current snapshot of evaluation practice in the first part of the year 2003.

A central focus has been the relationship between, and the combination of, external and internal elements. Hence, it has been an important criterion for the selection of contributors that they collectively represent all possible relevant combinations of internal and external elements. Other important criteria have been a suitable weighting between educational sectors, i.e. broadly speaking between the school sector and higher education, and a geographical variation.

In concrete terms, two of the contributions cover the whole educational scene, five contributions cover higher education solely, and four contributions cover the school sector solely. Six contributions are from Western Europe (Denmark (1), the Netherlands (2), Northern Ireland (1) and France (2)), two are from Eastern Europe (Hungary (2)), two are from North America (Canada (2)), and one contribution is from Oceania (New Zealand (1)).

In etymological terms, the word "anthology" has botanical roots. It means a collection of flowers. With the chosen contributors, we believe that we have composed an attractive bouquet.

Frans L. Leeuw, former Chief Review Officer in the Netherlands Inspectorate of Education has accepted the interesting, but challenging task of analysing the validity of the above hypotheses. His findings are the result of an analysis of all contributions. This analysis constitutes Part One of the anthology.

In order to lighten the task of analysing, the contributors were subjects to a rather detailed frame of reference, including a list of questions categorised into four principal questions:

1. Why do you evaluate? (Concerns values and purposes)
2. What do you evaluate? (Concerns objects)
3. For whom do you evaluate? (Concerns stakeholders)
4. How do you evaluate? (Concerns methods)

The purpose of the framework was to ensure comparability among the contributions. One may claim that this standardisation does not suit the idea of an anthology as a place where different views can bloom side by side, to stick to the botanical metaphor. However, the contributions prove that the frame of reference also allows for individual variation.

Still, a collection of dried flowers put on paper is called a herbarium. Everyone familiar with herbariums knows that flowers, which are dried and pressed, loose their scent, and eventually their colours fade. Likewise, everybody familiar with teaching and evaluation knows that reading articles will never provide you with the feeling and experience you get out in the field.

Hence, Part Two of the anthology, comprising the 11 contributions, is not intended to be read from one end to the other. Rather, we like to think of it as a handbook. For those readers who are particularly interested in evaluation activities in one country or another, the contributors were asked to attach a case story to their contribution, in which they describe values, purposes, objects and methods in the most prominent or interesting form of evaluation practised at present time. These case stories constitute Part Three of the anthology.

Part One - Analysis and Conclusions

Trends, Topics and Theories

Frans L. Leeuw,
Professor, Evaluation Studies, Faculty of Social Sciences, Utrecht University & Director of the Research, Statistics and Information Department, Ministry of Justice, The Netherlands

INTRODUCTION: QUESTIONS, HYPOTHESES AND RATIONALES

The aim of the Danish Evaluation Institute (EVA) for this anthology is to elucidate the relationship between values, purposes, objects and methods for educational evaluation in different educational sectors and in different countries. In more precise terms, the contributors to the anthology address in their papers four questions:

- Why do you evaluate in your country? Here the focus is on values and purposes of the evaluative activities.
- What do you evaluate? Here the object(s) of evaluations are central.
- For whom do you evaluate? This also concerns values and control & agent relationships.
- How do you evaluate? This concerns the methods applied.

The above questions relate to the following hypotheses that will be discussed in this chapter.

Hypothesis 1:
The control and owner relationships of the individual educational sectors determine the values and methods on which evaluations in a given sector are based. Put differently: all educational systems under the same control and ownership are evaluated with reference to the same values using the same methods.

The *rationale* behind this hypothesis can be derived from studies of principal-agent models in the sociology and economics of organizations (Scott, 2002). One of the insights is that principal-agent relations have an impact on the *contents of policies* and on the *selection of policy instruments* in the public sector. Questions asked are the following: under which conditions do agents opt for subsidies and information campaigns to realize their (policy) goals, *or* for accreditation and inspection, or for combinations? And: to what extent does the relationship the agent has with the principal determine these options and decisions? Scott (2001:136-145) adds that there are 'field logics' that dominate an organizational or social 'field'. In line with this, it is reasonable to assume that control-and owner relationships can have an impact on evaluations and the way they are carried out.

Hypothesis 2:
The culture within a given educational sector determines how it is evaluated, irrespective of the formal control and ownership aspects.

Here the *rationale* can be found in sociological studies in a very broad sense. Culture is considered to be a factor of importance with regard to many behaviours and activities, both of persons and of corporate actors. To assume that 'culture' is at least partly determining the way in which evaluations are done therefore is not a strange idea. Anthropologists like Geertz and Douglas, and sociologists such as Berger and Meyer have stressed the importance of 'cultural-cognitive elements of institutions: the shared conceptions that constitute culture, including the frames through which meaning is made' (Scott, 2001:57).

Hypothesis 3:
The institutional foundation of the evaluating body determines the values underlying its evaluations and choice of methods.

This hypothesis can be located within (neo-) institutionalism, one of the more advanced theoretical and empirical traditions within sociology and economics (Scott, 2002; Smelser and Swedberg, 1994). According to this tradition, 'institutions do matter'. They do so because they are transfer agents for routines, artefacts, incentives and approaches. Routines are ´habitualized behaviours that structure activities; they are patterned actions that reflect the tacit knowledge of actors: deeply ingrained habits and procedures' (Scott, 2001: 80 ff). Incentives function as reinforcers or killers of behaviour, while artefacts are technologies embodied in software and hardware, including textbooks, FAQ's, tick & flick lists, Yellow Books on Standards etc. If the importance of this ´equipment´ as a transfer agent is relevant for organizations *in general*, then why would it not be relevant for the evaluation 'industry' *in education.*

Some institutional scholars also use the concept of *'pillars'* to understand *why* institutions matter. Scott distinguishes between the regulative pillar ('rules, regulations, constraints that regularize behaviour'), the normative pillar ('values and norms') and the cultural-cognitive pillar ('shared conceptions, belief systems' etc.). Sociologists have produced knowledge showing how important these pillars are and how they can indeed influence behaviour and activities. If that is true for organizations *in general*, then why would it not be true for the evaluation 'industry' in education?

Hypothesis 4:
Certain evaluation methods develop their own lives and achieve such a degree of status in certain countries that their appropriate application to a given educational sector is not questioned.

The *rationale* here can be found in the work by David (1985). David studied the question of how it can be explained that *QWERTY* is still the most often used keyboard (design) of computers and, in earlier times, typewriters. The central concept is *path-dependency*, indicating that organizations and societies sometimes opt for a route that becomes so immersed in the economy, lines of production, societal belief systems and 'objective knowledge' that hardly anybody questions the (practical) validity of the 'path', even when there are more effective and efficient technologies available. Path dependency is linked to the concept of reification, meaning that certain inventions and 'ways to do' almost develop into 'things-on-their-own'. It is not unreasonable to assume that in the world of (educational) evaluations, path dependence has been taking place. This world is active and alive in many countries, has its own journals, its own (government sponsored) institutions such as Inspectorates, its own official 'international networks' like SICI and other 'transfer agents'. All these conditions make it reasonable to assume that hypothesis # 4 has an empirical content.

ON HYPOTHESIS 1:
LINKING THE CONTROL AND OWNER RELATIONSHIPS OF THE EDUCATIONAL SECTOR WITH THE VALUES AND METHODS ON WHICH EVALUATIONS ARE BASED

In order to find out to what extent hypothesis # 1 corresponds with data from the different country studies, I have looked into three topics.

The first concerns the *role of (basic) values*. In order to find out to what extent educational sectors differ, I have considered the values that, according to the papers, are central in the 7 countries. It appears that these values – overall-- do not differ *largely* between the seven countries. Everywhere, values like the quality of education, the quality of teachers, academic standards, internationalization, autonomy, accountability and independence are high on the agenda. While the system of financing schools – and, therefore, sometimes ownership - differs between some of the countries (e.g. schools in Denmark, the Netherlands and Hungary are largely funded through and by the public sector, while that is somewhat less frequently the case in New Zealand and Canada), basic values do *not* differ very much. Maybe, Bozeman's

concept of the *publicness* of certain organizations and policy fields can explain this finding. Education is such a central element of society that the control and ownership patterns of the educational sector do not determine the values and methods on which evaluations in a given sector are based. Education is conceived by large parts of the population as producing a *collective good*, taken care of by and through the public sector. Even when the production of this 'good' is done by the private sector, basic values remain much the same, including the ones used for their assessment and quality assurance.

Also, from another perspective, the hypothesis that there is a link between the control-owner/principal-agent relationships and basic values can be viewed critically. France, the Netherlands and Denmark strongly adhere to the importance of *dialogues-in-evaluation* that are focused on committing different stakeholders to the evaluations, inspections and quality assurance activities. However, the role of the (central) *state* in these countries differs *strongly*. In France, evaluation procedures are carried out by three different institutions within the school education sector: two are responsible for organising and conducting evaluations (the Inspectorate and the Education Ministry's evaluation division), while the third reviews evaluation findings and methodology (the High Council for Evaluation). All three organizations are part of the government or very strongly related to it. The Danish situation is different, given the role of EVA. In the Netherlands, the Inspectorate has the legal status of an Inspectorate in the 'French' sense of the word, but acts rather similarly to EVA. In these three examples the work done by evaluators is strongly based on dialogue and trust.

The third item to discuss is the question: *to what extent do the methods applied by educational evaluators in the respective countries differ*? Most countries have inspectorates or equivalent organizations working in all fields of education. Sometimes they are called 'inspectorates', sometimes quality assurance and assessments institutes, sometimes evaluation organizations and sometimes accreditation institutes. The methods used are the following:

- District inspections, unannounced inspections, specific inspections in case schools do under perform, systemic inspections, review work, etc.
- Interviews with heads and teachers, based on sampling, is common; interviewing representatives of parents is less so;
- Analysing administrative and other data is also often done; the same is true with regard to evaluating self-evaluation reports, carrying out quality assurance inspections, audits and program evaluations. This is also true for thematic evaluations or 'cross cutting studies'. Often, government documents are also analyzed;
- Other methodologies are also used, such as classroom observation. Information is also gathered for specific scientific studies (e.g. longitudinal studies), and occasionally from employers;
- Process and impact evaluations (also in specific fields, like ICT);
- Meta-evaluations also belong to the tool-kit, though not everywhere;
- Most of the work is criteria based, while the terms of references/norms are usually made public and/or are discussed with the parties representing the object of the inspection or audit. In addition, many countries have implemented education quality indicators programmes;
- Teamwork, site visits (sometimes contingent upon the findings in the self-evaluations and therefore ´proportional´) and 'hearing both sides' are also included in the methodological package.

The differences between the methods applied are *relatively small*. Hungary is rather dependent upon (large scale) surveys, including international ones. France does pay attention to pupil assessments through mass diagnostic tests and summative sample-based assessment, while in other countries that kind of work is carried out outside inspectorates or evaluation agencies. In Canada most provinces use formal, large-scale assessments to obtain systematic information about student achievement. In addition to these, some jurisdictions within Canada have implemented other initiatives to support public accountability and educational system improvement, like a system of school accreditation in which members of the school community and an

external evaluation team reports on the effectiveness (strengths and weaknesses) of various elements associated with the school.

With regard to evaluation approaches applied within *higher education*, the differences are *smaller*. All countries, in one way or another, apply an *approach* consisting of the following steps:

First, a university or polytechnic submits a self-description/self-evaluation, that – next - is peer reviewed by reading and by on-site-visits. Sometimes there is a second level (peer group) review. In the case of the Netherlands, it is the new Accreditation Organization and the (old) Inspectorate that check the validity and reliability of earlier 'visitation reports'. Similar 'modes of operation' can be found, for example, in Canada, New Zealand and Northern Ireland. Next, auditing the quality assurance methods and approaches is also to be found in most of the countries. The final stage is qualified feedback to the evaluand.

Nevertheless, there are differences. One is that in only one country (Hungary), decisions regarding the appointment of professors are made by educational evaluators. A second is that while all countries are engaged in evaluating academic programs, only some are evaluating entire institutions.

To what picture does this overview bring us? Firstly, it is clear that the control-owner relationships do not differ much in practice. Secondly, methods applied in the evaluative work have more similarities than dissimilarities.

ON HYPOTHESIS 2:
LINKING THE CULTURE WITHIN A GIVEN EDUCATIONAL SECTOR WITH THE VALUES AND METHODS ON WHICH EVALUATIONS ARE BASED

We have seen that basic values regarding education do not differ strongly between the countries. Other empirical information relevant to a description of the 'culture' within the educational sector of each country is not provided in the papers.

Nevertheless, in order to test the assumption that an element like 'culture' co-determines the work of educational evaluators, I have considered the goals defined by the educational evaluators of each country. Can it be that the goals of evaluations differ (strongly) between the 7 countries?

What I found is that there is a *high level of agreement* as to *why* evaluations are done. Catchwords are:

- to contribute towards quality improvement, accountability and transparency;
- to provide information;
- to give a public account of quality of education in order to satisfy stakeholders;
- to use valid, reliable and independently produced information about quality;
- to promote the highest possible standards of learning, teaching and achievement;
- to monitor the education system;
- to identify strengths and weaknesses;
- to certify and promote;
- to provide information for curriculum design, instructional methodology and resource allocation.

However, New Zealand frames its goals differently: "The purposes of evaluation in New Zealand higher education, for both quality assurance and quality development, are:

- to protect the interests of learners;
- to ensure learners have access to opportunities for life-long learning;
- to ensure available qualifications are meaningful and credible;

- to assure learners that courses and programmes are well taught;
- to ensure qualifications are obtained in safe environments using appropriate teaching and assessment systems;
- to contribute to the enhancement of quality systems and processes that improve the quality of research, teaching, learning and community service."

Though there is consensus that evaluation is done primarily to safeguard and stimulate (high) quality education and improvement, the focus in most of the countries is on the actual *educational system and its organizations*, whereas in New Zealand the *protection of the interests of learners* is given priority, and *not the institutions* themselves. Putting it somewhat differently: while the papers on most of the European countries describe the relevance of evaluations from a mainly *education-organizational perspective*, New Zealand focuses directly and primarily on *end users and consumers*. It uses the word *'protect'* as the very first goal of its evaluations. In the papers on Canada the authors also sometimes refer to 'consumer protection', but less prominently. It is interesting to see that the youngest democracy of the 7 (Hungary) also refers to 'consumer protection[1]'.

How can this particular focus on ´protection´ in New Zealand be explained? I dare to suggest that it has something to do with the earlier *history and culture of results based management* that was already implemented in the 1980s in New Zealand. What also might play a role is the relative softness with which educational evaluations are carried out in most of the European countries. Dialogue, commitment, reciprocity between subject and inspector is part and parcel in Europe[2]; the instrument of 'unannounced inspections' is only discussed in the paper on Northern Ireland. In the Hungarian situation, it is interesting to see that the goal of 'getting inside the EU' is also directly linked to the attention paid to evaluations.

The fact that consumer protection is formulated so prominently in the New Zealand case, makes it difficult to believe that 'culture' has *nothing* to do with the way in which evaluations are carried out. While most of the other countries ´speak softly´, New Zealanders carry ´a (big) stick´, i.e. focus on consumer protection[3]. Apparently, there is not only something to protect but also something *to protect from*, i.e. educational institutions. It might be argued that those evaluators that stress ´protection from´ have a somewhat different perspective on what (higher) education organizations do and do not do, compared with other evaluation agents.

ON HYPOTHESIS 3:
LINKING THE INSTITUTIONAL FOUNDATION OF THE EVALUATING BODY WITH THE VALUES OF ITS EVALUATIONS AND CHOICE OF METHODS.

According to many sociologists, historians and economists, 'institutions do matter'. Part of the more general concept of 'institutions' is the institutional foundation of organizations, like the ones analysed in this book. The 'institutional foundation' can be seen as a proxy of the way in which evaluations are carried out. To understand the relationship between 'institutional foundation' and 'values of evaluations and choice of methods', I will be looking into the following variables:

- how different or similar is the institutional position of the different countries;
- how different or similar is the anticipated future (next 5 years) with regard to evaluation in practice, as described in the papers?

The institutional positions are described as follows:
It is impossible to be precise about the level of independence, as this is not only dependent upon the legal structure of the institution but varies in accordance with the leadership of the

[1] *In the paper on the Netherlands 1 time 'protection' can be found.*

[2] *Excluding Northern Ireland where still unannounced inspections take place*

[3] *based on a quote used by the American President F.D. Roosevelt.*

board or director(s), the budget freedom, the publication freedom, etc. However, the further down one goes in the list, the smaller the probability that the evaluation actor described would probably be perceived as 'independent' (Standaert, 2003).

Denmark

EVA is an independent institution formed under the auspices of the Ministry of Education. It is required by law to cooperate with the two ministries in charge of education, but it has its own budget and is financially independent of the ministries and the educational institutions. Furthermore, the board of EVA has the right and the obligation to initiate evaluations, and it is mandatory for institutions to participate in evaluations initiated and conducted by EVA.. In the explanatory memorandum to the act it says, "The purpose of carrying out independent evaluations is primarily to contribute to the development and assurance of quality in education, and secondarily to perform an actual control of the goal attainment of the education".

New Zealand

The New Zealand Qualifications Authority *is a statutory body set up by government. It has an overall responsibility for quality assurance in secondary and tertiary education providers other than universities. The* New Zealand Vice-Chancellors Committee *is a statutory body and was given statutory status in 1962. With the Education Act 1989, the Committee was given the responsibility for quality assurance in all universities. To do this, the New Zealand Vice-Chancellors Committee has a standing committee – the* Committee on University Academic Programmes – *that approves new programmes and significant changes to existing programmes and monitors their implementation. In 1993, the New Zealand Vice-Chancellors set up an independent evaluation agency – the* New Zealand Universities Academic Audit Unit – *to audit the universities' systems for monitoring and enhancing quality and to disseminate and commend good practice.*

France: CNE

The National Evaluation Committee for Public Establishments of a Scientific, Cultural and Professional Nature was created by law. The CNE is responsible for examining and evaluating the activities of all universities, engineering schools and institutions under the auspices of the Minister in charge of higher education. The law of 1984 specifies that: "(the CNE) recommends appropriate measures to improve how establishments are run as well as the effectiveness of teaching and research". It is up to each entity to implement the recommendations that concern it.

The Netherlands: the Inspectorate of Education

The Inspectorate is a semi-independent organization, formally part of the Ministry of Education and Science. A specific law describes its work, and its do's and don'ts. An important element is that the Inspectorate publishes its own reports. The inspectorate, by law, is also obliged to assess the effectiveness of the National Accreditation Organization of the Netherlands.

Northern Ireland

The Education and Training Inspectorate (Inspectorate) provides inspection services and information about the quality of education and training in Northern Ireland (NI) to several departments, e.g. the Department of Education (DE) and the Department of Culture Arts and Leisure (DCAL).The organisation is a unitary inspectorate, providing independent advice to all three departments. The legal basis for the Inspectorate's work is set out in the Education Reform.

Canada

There is no one, single organization similar to the Inspectorate in The Netherlands or EVA in Denmark. Most of the organizations are evaluation and accreditation agencies set up by the government. Accreditation, for example, is granted by The Private Colleges Accreditation Board (PCAB), an agency set up by the Government. Ontario and British Columbia have procedures which are quite similar. The department of education reviews all new programs for duplication. In Quebec, new university programs are reviewed from a quality perspective by the «Commission d'évaluation des projets de programmes (CEP)» under the aegis of the «Conférence des

recteurs et des principaux des universités du Québec (CREPUQ)» and their relevance is assessed for financing by the Ministry of Education. New college diploma programs are approved by the Ministry of Education. This is somewhat similar to the situation in the Netherlands (multilayered evaluations).

Hungary
In August 1993, the Higher Education Act (HEA) appeared, and with it the legal framework for PhD training and accreditation was established. The choice between accreditation and audit type evaluation was conscious and well grounded: preliminary accreditation seemed to be the most promising and suitable means to raise strict quality demands and requirements for the new doctoral programmes in particular and for new degree programmes and institutions in general. In the HEA (in force from 1st September, 1993) the Accreditation Committee was, in addition, given the legitimacy to accredit HEIs as institutions and, in general, it was established "for the ongoing supervision of the standard of education and scientific activity in higher education, and for the perfecting of evaluation there". (HEA 1993, Section 80 (1).) Upon the nomination of the HEIs, the HAS, and other organisations in January 1994, members of the Hungarian Accreditation Committee (HAC) received their mandates from the Prime Minister for three years, elected their president, and began work with processing the decisions of the PNAC.

France: the Ministry, the Inspectors and the High Council
Evaluation procedures are carried out by three different institutions within the school education sector: two are responsible for actually organising and conducting evaluations (the Inspectorate and the education ministry's evaluation division), while the third reviews evaluation findings and methodology (the High Council for Evaluation).

At a national level, the General Inspectors, based in the education ministry, participate in the supervision, training and recruitment of some types of teachers. Counselling or assistance for schools is mainly the responsibility of regional inspectors, a group whose main task is to inspect, supervise and assess teaching staff. These inspectors have a regional remit but retain close links with the General Inspectors at the education ministry.

The evaluation division of the education ministry is in charge of the co-ordination of the evaluation and forecasting functions of the ministry. The High Council for Evaluation was set up by the education minister in 2002. Although its members are formally appointed by the education minister, its work is, in practice, independent from ministerial interference.

Netherlands: VSNU
This is the peer review organization of the Dutch universities branch group. It is funded through the university and government (though indirectly). Due to the new Accreditation Organization that has been established in 2003, this part of the VSNU-branch organization is on the verge of disappearing. One of the reasons is its institutional position, vis-à-vis its (perceived) independence.

Perceived level of independence: from high to low(er):

Denmark---------→New Zealand----------→France/CNE----------→ Netherlands/Inspectorate----------→Northern Ireland---------→ Canada-------→ Hungary---------→ France/Ministry -------→ Netherlands/VSNU

With regard to hypothesis # 3, two things can be said. Firstly, that there *are differences in the institutional position and autonomy* of the organizations. Denmark, New Zealand, France-CNE, and the Netherlands-Inspectorate are more autonomous than the French ministerial inspectors, and the organisations of Hungary and the Netherlands (VSNU).

Secondly, again, no clear pattern emerges with regard to the values or methods that these organizations apply. *Largely independent EVA* does not do completely different things to *the Netherlands or France*. The New Zealand agencies approach regarding the quality measurement of universities is not completely different to that of Hungary or the Netherlands (VSNU), with the exception of auditing research and auditing institutions, which is not done in the Netherlands by VSNU. That has very probably something to do with its (perceived) independence.

Let us now consider the question: *to what extent do these institutions differ with regard to the future they foresee in evaluation?* Below, I have indicated the central elements of the views of the future the organizations have presented us with. I have ranked the institutions with regard to the magnitude of the changes they foresee between now and 2007. First and foremost, I have used the impression the authors of the papers have provided. If somebody indicates that only 'gradual changes' are foreseen, or 'that approaches will be the same', then the position is clear. Next, some criteria are applied:

- If the logic of the operations of the organization or the organization itself is expected to change, that is considered a fundamental change;
- If goals or foci of evaluations is expected to change, that is considered a major change;
- If objects of evaluation are to change (more than on an ad hoc basis), that also is seen as a major change;
- If objects are added to the organization's list of current objects, that is considered a minor change;
- If methods and techniques are expected to change, that is also considered a minor change.

Based on major change to minor change, the ranking of the countries can be illustrated in this way:

Major changes----------→ minor changes:

VSNU/Netherlands ------→ New Zealand -----→ Canada -----→ France -----→ Hungary ------→ Northern-Ireland -----→ Netherlands ------→ Denmark

The Netherlands: VSNU

One thing is for sure: five years from now, the Netherlands will have a different system of evaluation, if any system is to survive, i.e. the NAO system. Its tasks will be to:

- *Verify and validate external assessment and grant accreditation for existing programmes;*
- *Assess the quality of new programmes;*
- *Contribute to the introduction of the Bachelor-Master Structure in Dutch Higher Education.*

The introduction of the accreditation system and the establishment of the NAO will have great influence on the current quality assurance system. One of the strong points of the current system is the ownership of the system by Higher Education itself. It did contribute to improvement and enhancement. Will it be replaced by a bureaucratic control system?

New Zealand

The maintenance and enhancement of the quality of core activities will still be the focus of evaluations five years from now. The evaluation of institutions against their own objectives will remain the focus.

The change messages contained in the Ministry of Education's statement of tertiary education priorities include the need for greater alignment with national goals; stronger linkages with business and other external stakeholders; effective partnership arrangements with Maori com-

munities; increased responsiveness to the needs of, and wider access for, learners; more future-focussed strategies; improved global linkages; greater collaboration and rationalisation within the system; increased quality, performance, effectiveness, efficiency and transparency; and a culture of optimism and creativity.

Canada

There will probably be greater focus on the efficiency of its institutions. Evaluations in the future will have to take into account quality indicators as well as efficiency indicators such as graduation rates. Furthermore, it is to be expected that more and more institutions will be judged not only by the quality of their programs, but also by the efficiency of their management and their capacity to establish and meet clearly identified and measurable objectives. The challenge for evaluation commissions will be to stay in support of the institutions and not become pure control agencies.

In describing recent changes in evaluation, it is observed that there has been a general trend, both in classroom and large-scale assessment, toward integrating curriculum, instruction and evaluation to ensure that assessment closely matches learning objectives.

In recent years, there has been a move toward having large-scale assessment results support planning for educational improvements at all levels: province, school board and school. It appears as though this trend will continue into the foreseeable future. The move toward assessment for both accountability and improvement has had some major impacts.

France

As regards pupil assessment, it is likely that the twofold approach which has been followed so far (mass diagnostic tests and summative sample-based assessment) will continue to prevail.

School evaluation will probably try to move beyond the observation of differences to putting forward ideas for interpretation and opportunities for improvement, which may be laid down in a specific procedure that could involve the intervention of a team from outside the school. This school evaluation procedure would necessarily result in a compulsory programme of action with which the school would have to comply.

Based on sound statistical surveys and qualitative evaluations, making forecasts about future trends in the education and training system has been revealed as a major tool for policy development.

The greatest challenge which faces evaluation in the years ahead is undoubtedly to convince stakeholders to make use of it. Finally, it should also be noted that international, and, in France, more specifically European, comparative evaluation will inevitably develop further.

A look at the CNE's past shows consistency in its values and objectives since the very beginning. Secondly, lightening the procedural burden of evaluation is a major challenge that concerns first and foremost the evaluator. A third factor to mention is naturally the entire move towards European convergence and the strengthening of university autonomy.

To conclude, we cannot fail to highlight the fact that European convergence in higher education will be an important factor for change for the CNE and for its working methods in the years to come. The principle of subsidiarity, a guarantee of the respect for national choices and the conditions of mutual trust will impose greater transparency vis-à-vis European partners. The setting up of external evaluation of evaluation agencies and the taking into account of European users in evaluations are two steps the CNE is preparing to face.

Hungary

As to objects and purpose of evaluation, no considerable changes in the next five years are to be expected. Peer-review will certainly remain while academy centeredness is changing only

slightly, at a very slow pace, with the involvement of more experts from the so-called users' sphere in the work of the HAC.

More changes will take place regarding the actual implementations of evaluations. These step-by-step changes to the practical, operational level of methods can slowly give rise to more fundamental changes in the system.

The international involvement of the HAC, and the internationalizing of QA in higher education should be stressed.

Presumably, assessment and evaluation will have a major role in public education. Hungary would continue to be a permanent participant in international surveys.

As an integral part of this development, the field of assessment and evaluation would have to become an essential part of teacher training and in-service training; the lack of these is a shortcoming of the teacher training curriculum. The culture of school self-evaluation also needs further development. The development of key competences and their evaluation will probably be one of the most important fields of public education, but subject-related assessment cannot disappear from practice, either.

Northern Ireland
The framework will, in the near future, consider the inter-play of internal and external evaluation, and highlight the benefits of inspection to organisational improvement. It is anticipated that the Charter and the framework, taken together, will direct and influence the work of the Inspectorate over the next five years and beyond.

In five years from now, it is anticipated that self-evaluation will be further embedded within inspection, and that the Inspectorate's evaluation of an organisation's capability to self-evaluate will become a much more significant part of inspecting and reporting.

Furthermore, the recently piloted strategy, whereby the evaluation of inspection is carried out by a firm of independent consultants, and the findings made public, will become an integral part of the Inspectorate's way of working.

The Netherlands: the Inspectorate of education
There will probably only be gradual changes in our work.
We will be able to focus more on schools that lag behind in their quality development, trying to stimulate them by sharper and more focussed inspections.
And to focus more on inspections in the "better schools", particularly regarding their developmental needs. So, still more flexibility in inspections.

Probably, we will focus more strongly on topics and issues that are seen as relevant for the further development of the system as a whole, and in order to "close the inspection chain" from perspectives of social inclusion and/or integral care for youngsters.

Probably, requests will come for more inspection of subjects in schools; we now do that rather superficially. Not only schools, but probably also parents and students will demand more specific information about how good a certain school is, not only in general terms and quality aspects, but also in terms of expertise, e.g. science, humanities or music teaching.

We also expect a more developed European perspective, in the sense that "international cooperation" will become a more important aspect of quality of schools than it is in our present framework

Denmark
The purpose of the evaluations will continue to be twofold, i.e. they still have to contribute to the quality improvement of the evaluated objects in particular, and the evaluated field in gen-

eral. Furthermore, the evaluations will continue to have a control function, as they inform the stakeholders in a broad sense, both in Denmark and abroad, of the state of quality in the evaluated field.

The majority of the evaluations still take their starting point in the objectives formulated at the national, local and institutional level. However, due to international developments, there will be an interest in the results of education and in creating a higher degree of transparency of quality in education, across borders. There will be a need for quality descriptions which are understandable and acceptable across borders, and it will be necessary to develop other ways of describing quality than in terms of fitness for purpose.

Last but not least, there will be increased focus on competences, i.e. what are the pupils or students capable of when they have completed a particular programme at a certain level of the education system, as another means of making quality judgements comparable.

There will be continued focus on the procedures set up by the institutions themselves to continuously check and improve the quality of their activities and structures. Consequently, there will be a need for external quality assurance to check the effectiveness and sustainability of these internal mechanisms and to undertake measures to provide input to the post-audit improvement activities initiated by the institutions.

It can be concluded that the institutional setting of the educational evaluator is *not* ***the*** *factor* to determine their values and methods of educational evaluation.

ON HYPOTHESIS 4:
CERTAIN EVALUATION METHODS DEVELOP THEIR OWN LIVES AND ACHIEVE SUCH A DEGREE OF STATUS IN CERTAIN COUNTRIES THAT THEIR APPROPRIATE APPLICATION TO A GIVEN EDUCATIONAL SECTOR IS NOT QUESTIONED.

The rationale here can be found in the concept of path-dependency. This means that organizations and societies sometimes opt for a route that becomes so immersed in the economy, lines of production, societal belief systems and 'objective knowledge' that hardly anybody questions the (practical) validity of the 'path', even when there are more effective and efficient technologies available. It can be argued that, within the field of evaluation methods, similar path dependence can be found. Power (1999) and Barzelay (1996) have indicated that this has also happened in the field of auditing.

Let us, therefore, refer to the evidence available in the country-papers. In contrast to the previous three hypotheses where no hard evidence became available, we will hopefully, this time, indeed find some ***corroborative evidence***.

What is the evidence? I will list two elements.

The *first* deals with *the belief (to be found in most of the organizations) that evaluating self-evaluations of institutions and schools is an adequate route to follow*.

The *second* deals with *the belief that carrying out evaluations in which stakeholders play an important role is also an adequate route to follow.* The point I would like to make is that, following *the theory of institutional isomorphism*, educational evaluators, as described in the papers I have used as source material, have started to *become more similar* because of *organizational imitation* (DiMaggio and Powell, 1983).

The evidence, part 1: *evaluating self-evaluations*
All educational evaluators in the 7 countries are becoming more and more involved in *assessing or evaluating school self-evaluation reports* in one way or another. *All* educational evaluators are also getting involved in *auditing* the mechanisms schools and universities use in order to assess the quality, efficiency and effectiveness of their own programs and/or institutions. This is

clearly an example of isomorphism (Power, 1999). Crucial is that the evaluator first and foremost starts with what evaluative knowledge the organization itself has produced. However, as the *criteria* the evaluators work with are *well known* (due to their websites, reports, symposia, feedback to schools, laws, debates etc.), these criteria act as *incentives* to *behave according to these criteria and standards*. There is serious doubt as to whether, by following this line, the goals of the evaluators will indeed be realized. The main reason for this concern lies in the fact that isomorphism not only leads to imitating the good sides of an approach, but also the unintended and even negative side effects (van Thiel and Leeuw, 2002)

The evidence, part 2: *stakeholder involvement*
All evaluation agents believe that involving stakeholders in their work is beneficial to the evaluation process, the outcomes and the utilization of the reports, but not all evaluation agents believe in the same intensity and level of activism with regard to stakeholder involvement. Partly, this is a confirmation of hypothesis # 4. Partly, though, it is not, because the differences in the practice of involving stakeholders are large. Some inform stakeholders and communicate with them, others involve them in practical work, and some make (large) parts of their norms, methods and criteria partly dependent upon the attitudes and responses of stakeholders. One organization even refers to a 'shared evaluation responsibility'. One can, therefore, detect the *following* continuum:

information and communication -----→ involvement -----→ coordination : -----→ agreement with norms, criteria & methods -----→ partnerial evaluations.

In line with this continuum, the different countries can be plotted.

In New Zealand, it is indicated that evaluations, therefore, are directed primarily at the institution's management, academic staff and students, and the evaluation processes are designed to involve those sectors as well as graduates, the professions, business and industry.

In Northern Ireland, Evaluations are also provided for individual organisations via, for example, spoken reports to subject departments, senior management teams, and school governors; and to groups such as the employing authorities, parents and the general public by way of written reports of inspection and follow-up inspections.

The same is true for Canada where the issue of 'stakeholders' is also addressed.

However, in the field of higher education organizations, such as HAC in Hungary and the VSNU in the Netherlands, the intensity of involving stakeholders is increasing. Several of the stakeholders, for example, participate through delegated members in the work of HAC and VSNU.

In France, the role of stakeholders is conceived in terms of 'consultation': 'Whenever a specific type of evaluation is decided at the national level, the object to be assessed is always very clearly identified and defined after consultation with the relevant bodies interested in the results of the work'. And, with regard to the role CNE plays in evaluating higher education programs, again there is a more active approach to stakeholders: 'Evaluation is a concerted approach. When evaluations are carried out, there are many exchanges between the establishments and the National Evaluation Committee – concerted reflection on the evaluation methodology and the questionnaire for internal evaluation; discussion on the themes of expertise chosen for evaluation, on-site visits by members of the French CNE and the general secretariat and the sending of experts. The draft report itself is submitted to those in charge of the institution under review, as they are also responsible for validating the data published in the report. The head of the establishment has the last word; his/her response is published at the end of the evaluation report.

With regard to the situation in the Netherlands, it should be noticed that the many stakeholders play a crucial role in the norms, criteria and methods the evaluators use, based on the frameworks for inspections. These frameworks have to have as much commitment as possible from those who are concerned with the work of the Inspectorate. For this purpose the Inspectorate has to consult with representatives of the educational field and other stakeholders and to take very seriously their opinions. But the Inspectorate remains responsible for the decisions about its own frameworks for inspection. Parliament has created the arrangement that the Senior Chief Inspector has to make decisions regarding the framework, and, following this step, the minister has to approve the proposal and to send it to parliament. This is in order to enable a parliamentary debate between the minister – who is ultimately responsible for the functioning of the inspectorate! – and parliament . This possibility was explicitly opened so that the Inspection could not fix criteria, norms and methods without the social approval of a broad representation of interested parties connected with potential evaluation objects. The first frameworks have been dealt with according to this procedure in November and December 2002 and are now valid for three years. They contain details about the working methods of the school-inspections (see below) and they provide the indicators and criteria for the aspects of quality.

Finally, and closest to the picture of complete stakeholder involvement is EVA, Denmark. *According to the paper: 'Reflecting the fact that EVA's evaluation activities cover the whole public education system, the institute has a very large and diverse group of stakeholders at all levels in society. Amongst others, several ministries are involved, educational councils, labour organizations, employer organizations, local and regional governmental organizations, teacher organizations, etc.* Prior to each evaluation, the institute conducts a preliminary study that typically involves a stakeholder analysis and dialogue. Some stakeholders are, therefore, involved on a case-to-case basis. Other stakeholders, like the Ministry of Education and the education councils, are permanently involved in the activities of the institute through the institute's board, as required by law.

The committee of representatives, which comments on EVA's annual report and the priorities of planned activities, comprises 27 members. The members are appointed by the board on recommendations of organisations from the following sectors: school proprietors, school associations, school boards and employers; rectors' conferences and school managers; management and labour organisations; teacher organisations and student and pupil organisations'.

EVA also points to the phenomenon now known as partnerial evaluations (Pollit, 1999): 'The evaluation model applied by EVA implies a *shared evaluator responsibility* in the sense that the activities of the area under evaluation are partly evaluated through self-evaluation and partly through analysis of the documentation material by the external evaluation group'.

CONCLUSIONS

Educational evaluations are carried out within the *boundaries of societal restrictions and opportunities*.

One such boundary concerns the *control and owner relationships* within the educational sectors that are subjects of evaluation. Based on the material in *this book*, these relationships appear *not* to differ *fundamentally* between the seven countries. Then, it was found that these relationships do not have a large impact on the values within the educational sectors or on the ways in which evaluations are done. Culture was also hypothesized as an important factor, but again: we did *not* find *much evidence* that, based on the material that was presented to us in this bibliography, this factor is crucial in determining values and methods. A third hypothesis stems from neo-institutional theory: does the *institutional foundation* of the educational evaluation institute determine the values and methods? Here we found that there are *indeed* differences between the institutional position and the autonomy of the evaluation agencies in the seven countries. To some extent the organizations can even be ranked in terms of their autonomy.

Additionally, if one compares the different institutions vis-à-vis their respective foundations in society, there is – again — no clear pattern to the values attached to evaluations or to the methods used.

The only hypothesis that seems to be corroborated is number 4: *certain evaluation methods develop their own lives and achieve such a degree of status in certain countries that their appropriate application to a given educational sector is not questioned*. Indeed, there is evidence that a certain ´methodological´ *isomorphism* and *path-dependency* is taking place *within this evaluation community.* Dialogue-driven or stakeholder evaluations are almost *everywhere* to be found, unannounced studies almost *nowhere*; evaluating self-evaluations and auditing quality control mechanisms are *becoming more and more* the ´talk of the day´, and, finally, the *predictions* of where the offices will be ´five years from today´ do *not* present us with a broad spectrum of unexpected or new foci and approaches. It can be argued that educational evaluations, as described by the authors in the country-reports, have become *´institutionalized´*. That, however, does not only have positive effects. Sociologists and economists conscious of neo-institutionalism direct attention to unexpected and even undesirable side-effects of this development, such as: ´predictability´ of the behaviour of counterparts and evaluands; 'teaching to the test'-behaviour; tunnel vision; and – even— the performance paradox (Smith, 1996; van Thiel & Leeuw, 2002). To prevent these phenomena from occurring, educational evaluators should invest in ´reflective practitioners´.

REFERENCES

Barzelay, Michael (1996), PP. 15-57, Performance auditing and the New Public Management: changing roles and strategies of central audit institutions, in: OECD-PUMA, Performance auditing and the modernisation of Government, Paris.

David, Paul. 1985. Clio in the economics of QWERTY, American Economic Review, 75: 332-337

DiMaggio, Paul J. and Walter W. Powell, 1983, The Iron Cage revisited: institutional isomorphism and collective rationality in organizational fields, in: American Sociological Review, 48: 147-160.

Meyer, M.W. & Gupta, V. (1994) The performance paradox. Research in Organizational Behaviour, 16, 309-369.

Power, M. (1999). The audit society. Oxford, Oxford University Press

Shaw,I. Et al, (2003).Do Ofsted inspections of secondary schools make a difference to GCSE results? British Educational Research Journal, (1), 29.

Thiel, Sandra van & Frans L. Leeuw. 2003. The performance paradox in the public sector, in: Public Productivity and Management Review, 25 (3): 267-281

Scott, W.R. (2001). Institutions and organizations. Thousand Oaks, SAGE.

Smelser, N.J. and R.Swedberg (eds.). 1994. The Handbook of economic sociology. Princeton University Press, Princeton, N.J.

Smith, P. (1995). On the unintended consequences of publishing performance data in the public sector. International Journal of Public Administration, 18, 277-310.

Part Two - Country Contributions

Denmark

Dorte Kristoffersen
Deputy Director
The Danish Evaluation Institute (EVA)

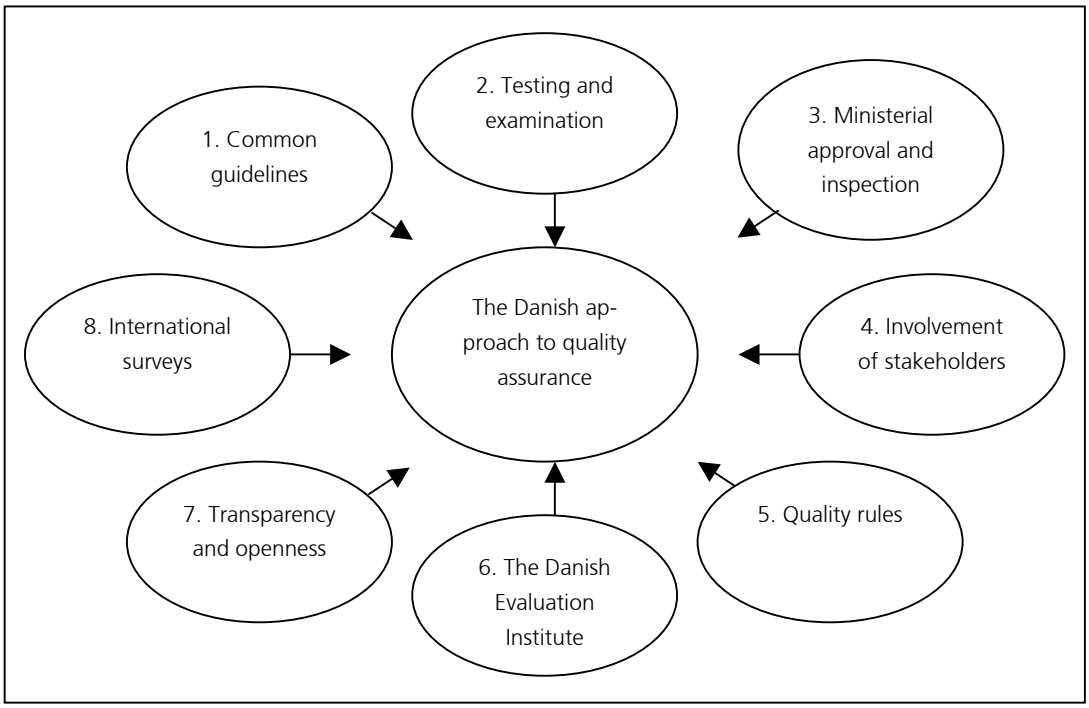
Values and purposes

In Denmark, the Danish Evaluation Institute (EVA) conducts evaluation of education at all levels. Other bodies occasionally conduct educational evaluations, but no other bodies are required by law to conduct evaluations at all educational levels or have educational evaluation as their primary responsibility. This article thus focuses on EVA's evaluations, but first a brief introduction to the Danish approach to quality assurance of education in general.

Systematic quality development in the Danish educational system is based on common principles (see figure 1 below) that are adapted to the various areas of education. This relates, among other things, to the fact that the different levels of the Danish education system are characterised by different principles of government and ownership.

The 12 Danish publicly financed universities are the responsibility of the Ministry of Science, Technology and Innovation, whereas the Ministry of Education regulates almost all other parts of the basic education system including: primary and lower secondary education; upper secondary education; vocational education and training; short- and medium cycle higher education; adult education and continuing vocational training.

Figure 1:
The Danish approach to quality assurance of education



Source: The Danish Ministry of Education

Points 1-4 in figure 1 are traditional and commonly acknowledged quality assurance mechanisms. Testing and examination and ministerial approval and inspection do, however, deserve a few comments.

Both lower and upper secondary education programmes are finalised by examinations. Some tests and written examination questions are produced centrally – hence all pupils answer the same examination questions – and external teachers (i.e. teachers from other schools) take part in the marking of examination papers. In higher education, examination questions and tests are not produced centrally, but for each programme and for each subject a national corps of external examiners is appointed. The corps partly comprises teachers/professors from other institutions, and partly labour market representatives. External examiners take part in a minimum of one third of all final examinations. The role of external examiners, at all levels of the educational system, is to assure that each pupil/student is assessed fairly, and to assure an equivalent national level of assessment across schools and institutions.

Ministerial approval and inspection are other important elements of assurance of national standards. The ministry approves all public institutions. Private institutions may operate without ministerial approval, but if an institution does not meet specified minimum standards, students cannot receive the state student grant. Without grant approval, it is difficult to attract students and, hence, exist as a school. The ministry is, furthermore, responsible for the systematic inspection of all primary schools at institutional level, and all secondary schools at both institutional and subject level. In primary and lower secondary education, local authorities are in charge, whereas in upper secondary education, the ministry has appointed a corps of subject advisors who conduct a form of inspections – however, their advisory function is the more important one.

In the 1990's, the then existing quality assurance mechanisms were supplemented by new initiatives (points 5-8 in the figure above). EVA's predecessor, the Danish Centre for Evaluation of Higher Education, was established in 1992 with the purpose of evaluating all higher education programmes within a 7-year period. EVA was established by act of parliament in 1999 (point 6 in the figure above). The primary mandate of the institute is to evaluate Danish education at all levels and to function as a national centre of knowledge for educational evaluation. The expansion into primary and secondary education was prompted by the results of international surveys (point 8 in the figure above). As Danish pupils in primary and secondary education came out less commendable than expected, the results attracted much public attention.

EVA is an independent institution formed under the auspices of the Ministry of Education. It is required by law to cooperate with the two ministries responsible for education, but it has its own budget and is financially independent of the ministries and the educational institutions. Furthermore, the board of EVA has the right and the obligation to initiate evaluations, and it is mandatory for institutions to participate in evaluations initiated and conducted by EVA.

The explanatory memorandum to the act states, *"The purpose of carrying out independent evaluations is primarily to contribute to the development and assurance of quality in education, and, secondarily, to perform actual control of the goal attainment of education"*. In the act itself, the secondary purpose of control is not mentioned. However, a certain degree of control is understood in connection with the term 'quality assurance', i.e. quality assurance is understood as a short-term purpose, whereas quality development is understood as a long-term process. In practice, this means that EVA has a twofold objective: control and development. Both objectives are prevalent in all the activities that EVA initiates.

Quality is understood as fitness for purpose, with a strong emphasis on the users' perspective. The starting point is partly externally defined in the relevant legislation, and partly internally defined through the objectives formulated for the evaluated activity. EVA examines fitness for purpose through an analysis of the intentions and activities that are supposed to lead to the fulfilment of preset goals. The users involved are typically users closely connected to the evaluation object, i.e. pupils/students, graduates and employers.

Evaluation is understood as a process that leads to quality assurance and quality development. EVA understands evaluation as being a composition of basic, and to a certain degree compulsory, elements, such as self-evaluation, user surveys and assessment by external experts. However, evaluation is also an umbrella-term covering such multifaceted activities as accreditation, benchmarking and audits. EVA has not yet conducted actual audits, but quality assurance mechanisms at the evaluated institutions have increasingly been given attention in the evaluations. This focus may lead to a redefinition of the relationship between evaluation and quality assurance/quality development.

In summary, there is basically consensus in the Danish educational sector on the understanding of external evaluation as presented in this paper. In practice, however, minor variations in the evaluations can be related to the specific educational area. Accordingly, the view of, and expectations to, EVA's role may vary from sector to sector.

Objects

Due to the fact that EVA covers the whole educational system, the fields in which EVA conducts evaluations are numerous and diverse. This implies that it is necessary and appropriate to conduct different forms of evaluations. Despite the diversity among the evaluations that EVA carries out they can generally be grouped under the following seven headings:

- Programme evaluations
- Subject evaluations
- Thematic evaluations
- Evaluations of teaching
- Evaluations of institutions
- System evaluations
- Evaluations of quality assurance mechanisms (audits)

A *programme evaluation* covers a specific programme as a whole, or selected aspects thereof. Most often, the aim is to provide an overall assessment of the programme, and programme evaluations typically encompass all components that influence programme quality. This includes components like the study environment and the organisational framework in which the programme operates.

The main purpose of a *subject evaluation* is to assess the quality of a specific subject within a programme, including the methods of teaching applied in relation to the subject and its context. The focus of subject evaluations is narrower than that of programme evaluations.

Thematic evaluations can in principle be conducted within all programme areas and educational sectors. The aim of this type of evaluation is to assess the quality and practise related to a specific theme that cuts across programmes, institutions and sectors, e.g. the transition between different educational levels.

Evaluations of teaching typically assess the quality of the forms and methods of teaching and learning within one or more programme areas.

Evaluations of institutions aim at assessing the organisational, administrative and managerial framework of an educational institution, e.g. typically covering administration, financing, research, education and quality assurance.

The aim of *system evaluations* is to assess the coherence of central aspects within a system, such as a municipality, and the implications for the quality of the education provided within the system.

Finally, *evaluations of quality assurance mechanisms* aim to assess the quality assurance mechanisms within, typically, a specific system, institution or programme, and how these affect the quality of the educational activities.

Programme evaluations, thematic evaluations and subject evaluations are currently the most common forms of evaluation conducted by EVA. Experience with the other forms is limited, but it is expected that audits in particular will take up a larger share of EVA's activities in the years to come. It should also be mentioned that, despite EVA's limited experience with regular audits, etc., quality assurance mechanisms have always been a prominent focus in most of EVA's evaluations. In other words, the borders between the different types of evaluations listed above are not clear-cut.

When EVA chooses to conduct a specific evaluation, it is selected according to a range of criteria:

- *Relevance*: the evaluation is essential for stakeholders within the educational sector and relevant to the political debate on education;
- *Development perspective*: the evaluation highlights and strengthens quality development within the educational system;
- *Need*: the evaluation reflects an expressed need for quality assessment and development in the educational system;
- *Methodological development*: the evaluation contributes to the development of the methodological portfolio of EVA;
- *Visibility*: the evaluation opens and encourages debate on central issues related to the educational system;
- *Coverage*: the evaluation contributes to ensuring that the evaluation activities of EVA have a broad coverage;
- *Accumulation*: the evaluation provides a possibility to build upon, and add value to, prior evaluations;
- *Theme strategy*: a variety of the different types of evaluations should be represented within the annual action plan.

These criteria reflect the expectations and interests of stakeholders and the ideals of systematic evaluations, quality and independence that EVA has to live up to. The use of these criteria as a frame of reference when selecting the object of an evaluation is expected to ensure quality development of the objects being evaluated. At the same time, the criteria contribute to ensuring quality development of the educational system as a whole. Finally, the criteria also ensure the recognition of EVA as a trustworthy and relevant external cooperating body, in relation to the need for evaluation within education. When selecting the object of an evaluation, the extent to which the above criteria have been taken into account is considered. The criteria are, however, not ranked.

Stakeholders

Reflecting the fact that EVA's evaluation activities cover the whole public education system, the institute has a very large and diverse group of stakeholders at all levels in society. These are listed in figure 2.

Figure 2
Stakeholders

	Government	Organisations and immediate users
International level		ENQA, INQAAHE, etc. Sister organisations around the world
National and local level	Ministry of Education Ministry of Science, Technology and Innovation Educational councils Regional councils Local authority councils	Labour organisations Employers' and management organisations Teachers' organisations Pupil and student organisations Local and regional governments' associations School boards' associations Rectors' conferences
Institutional level	All public educational institutions, from pre-school to universities	Private educational institutions receiving public subsidies
Intra institutional level	Management Teaching staff	Parents Pupils/students

Stakeholders are involved in EVA's activities in various ways. Prior to each evaluation, the institute conducts a preliminary study that typically includes stakeholder analysis and dialogue. Some stakeholders are, therefore, involved on a case-to-case basis. Other stakeholders, like the Ministry of Education and the educational councils, are permanently involved in the activities of the institute through its board, as required by law. The Minister of Education appoints the chairman of the board and, based upon the recommendations of the ministry's educational councils, the board members. The Ministry of Education and the Ministry of Science, Technology and Innovation are also involved in informal, regular discussions of policy, strategy and results. It is a continuous challenge to maintain the balance between independence from and collaboration with the two ministries.

The committee of representatives, which comments on EVA's annual report and the priority of planned activities, comprises 27 members. The members are appointed by the board, based on the recommendations from the following organisation types: school proprietors; school associations; school boards; employers; rector's conferences; school managers; management and labour organisations; teachers' organisations; and student and pupil organisations.

Users of graduates are primarily involved in the evaluations as members of evaluation groups responsible for conclusions and recommendations in final reports. When the institute evaluates higher education, users of graduates are typically employers, and, when evaluating primary and secondary education, users of graduates are generally secondary level respectively higher education institutions. Users of graduates, and graduates themselves, are usually surveyed as part of the evaluation process.

The objects of evaluations obviously have a keen interest in EVA's activities, and are actively involved in the evaluations, first and foremost in the production of self-evaluation reports, but also during site visits and report consultations. Since 2001, EVA has conducted yearly evaluations of its collaboration with the educational institutions under evaluation. For the greater part, these evaluations show that the educational institutions are satisfied with the collaboration, but there is of course room for improvement in some areas. Despite concerted efforts, there are still stakeholders who feel that evaluations are primarily a threat to, and a strain on, their resources.

Student involvement in evaluation is an area under development. Student and pupil organisations have four seats on EVA's committee of representatives, but students are not represented on the board. The institute was recently involved in a working group in the Nordic Network, which looked into the issue of student involvement in quality assessments of higher education

in the Nordic countries. In March 2002, the board also decided to conduct a pilot project with students in evaluation groups to test the concept. So far, experiences are positive.

Evaluation reports have many diverse target groups and are therefore written in a language that allows readers who are unfamiliar with the education system to understand the contents. Most reports are nevertheless lengthy, and, therefore, the institute sometimes publishes so-called "shortcuts" to the reports, i.e. short, concise versions of the reports, in order to get through to a larger group of stakeholders. After any evaluation, the evaluated objects are expected to follow-up on the recommendations given in the report, but there are no direct consequences in terms of funding if they do not. Recommendations are intended as inspiration and suggestions for follow-up, and reports are, therefore, written in a manner that invites dialogue and signals openness. All evaluation reports are made public and are available on the Internet. If the Minister of Education considers that the follow-up plans or activities are insufficient, the minister may intervene.

Methods

Despite the fact that the objects of EVA's evaluations vary and that the evaluations cover various educational levels, they share the following characteristics:

- EVA initiates the evaluation and appoints a *team of evaluation officers* among its permanent staff responsible for the methodological and practical planning and implementation of the evaluation.
- EVA conducts a *preliminary study* for each evaluation. It takes the form of a dialogue with interested parties involved in the subject matter (e.g. a course of education) and encompasses existing material relating to the field of education, e.g. regulations, government circulars, curricula, etc.
- EVA drafts elaborate *terms of reference* for each evaluation, presenting objectives and a framework for the evaluation. The board of the institute approves the terms of reference.
- For each evaluation, an external *evaluation group* is appointed. The members must have either a general or specific expertise in the field concerned.
- The individual educational establishment conducts a *self-evaluation*, presenting and analysing what it perceives as its own strengths and weaknesses with reference to a self-evaluation guide provided by EVA.
- The evaluation group and the team of evaluation officers conduct *site visits* at the educational units under evaluation. The visit is planned in consultation with the individual units.
- In connection with each evaluation, surveys may be conducted. Most often, these are *user surveys* among students, parents, graduates, employers or other groups of stakeholders.
- In its concluding *public report*, the evaluation group presents its analysis, conclusions and recommendations for developing the quality of the area of education in question.

EVA emphasises that the methodology applied in any given evaluation must reflect its specific purpose and focus, and be relevant in the context of the educational unit under evaluation. In each evaluation, the methodological approach can thus vary within the standard framework.

It should also be noted that, although the elements listed above are common to EVA's evaluations and are considered relevant and appropriate, EVA continuously assesses the adequacy of the evaluation model and each of the elements included. Aside from conducting evaluations, a central task of the institute is to continuously ensure methodological development. This is demonstrated by the fact that some evaluations apply a different frame of reference for assessments; generally, EVA applies a fitness for purpose approach, but over recent years, EVA has increasingly experimented with the application of a criteria based approach.

As mentioned above, an external evaluation group is appointed for each evaluation. The evaluation group is responsible for the professional quality of the evaluation. The size and composition of the group is decided in accordance with the purpose and scope of the specific evaluation, and the terms of reference for the evaluation define which competences the group

must possess. The evaluation group is responsible for the conclusions and recommendations provided in the evaluation report. Accordingly, the selection process is crucial. The experts must posses a solid knowledge and understanding of the evaluation object (i.e. the programme, the theme, the subject, etc.) and, at the same time, be independent of the evaluation object. In this connection, it is a well-known small-state problem that it can be very difficult within the narrow confines of a small education system to find experts that may be considered independent and unbiased. The Danish solution is to recruit at least one expert for each evaluation from another Nordic country. The Nordic experts are able to read the documentation in Danish but have the necessary distance to their Danish colleagues.

A team of evaluation officers from EVA supports the evaluation group. The evaluation officers are responsible for the practical and methodological planning of the evaluation and the drafting of the final evaluation report.

Self-evaluations, site visits and user surveys constitute the forms of documentary evidence that are referred to in the analysis and assessment of the evaluation objects. In line with the requirement stated in EVA's legal framework, self-evaluations are included as a mandatory element in all EVA's evaluations. In relation to evaluations within the higher education sector, the self-evaluation is typically conducted by one group of individuals including representatives of different stakeholders in the educational unit under evaluation, such as management, teaching staff, students and administrative staff. For evaluations within the school sector, it is more common that different groups of internal stakeholders conduct the self-evaluation (or parts of it) separately, but one mixed group may also conduct it.

No matter how the self-evaluation process is organised, it is always conducted with reference to a self-evaluation guide provided by EVA. The structure and content of this guide vary from one evaluation to another, but it always includes questions that require the provision of both quantitative and qualitative information. Reflecting the application of the fitness for purpose approach, considerable attention is typically devoted to the values and goals of the evaluation objects and how these are implemented in practice. The self-evaluation guides are always publicly available.

Generally, the self-evaluation is the primary documentation referred to in the succeeding analysis. In fitness for purpose evaluations, the self-evaluation is also the standard against which the evaluation object can judge itself. It provides a framework for establishing a definition of quality; it helps the educational unit decide the extent to which it is achieving its goals, and allows it to form an action plan for development. In a development perspective, the self-evaluation *process* is considered as important as the *outcome* of it. The second, and perhaps even more important, purpose of the self-evaluation is to furnish the evaluation object with valid methods for, and a genuine commitment towards, continued quality assurance.

Quantitative and/or qualitative surveys are conducted in relation to most evaluations. Questionnaires or focus group interviews are the most common forms of user surveys in relation to evaluations within all the educational sectors. These often serve the purpose of collecting the opinions of important groups of stakeholders who are external and, therefore, not represented in the self-evaluation groups, e.g. parents, employers and graduates connected with the evaluated units. In relation to evaluations within the primary school sector, separate sessions with groups of pupils have proven effective in terms of eliciting their perspectives.

The target group and focus of a given survey, as well as the type of survey, are selected with reference to the overall purpose and focus of the evaluation. Accordingly, the focus of the surveys varies, but, generally, surveys among graduates and employers focus on perceptions of the learning outcome, e.g. of the subject or programme under evaluation. Surveys among parents, in relation to evaluations within the primary school sector, typically focus on perceptions of matters relating to the educational processes.

EVA typically engages professional consultancy firms, specialising in qualitative and quantitative surveys, to conduct the user surveys in close contact with the team of evaluation officers and methodological experts from EVA. User surveys are made public as appendices to the evaluation reports.

Site visits conducted by the evaluation group and the team of evaluation officers form the final stage of the data collection process. In relation to programme evaluations, the most common form of evaluation conducted by EVA, a one-day site visit to each of the self-evaluating institutions is the typical form of site visit. In other evaluations the site visits to the self-evaluating units may be shorter or longer depending on the object and focus of the evaluation. Evaluations in which only a selected number of the self-evaluating units are visited do also occur. This is primarily the case when an evaluation includes a large number of self-evaluating units.

During a site visit, the evaluation group and the team of evaluation officers interview the different groups of internal stakeholders separately. Again, the number and length of interviews, and participants, depend on the focus of the evaluation, but always includes interviews with those who were involved in the self-evaluation process. The purpose of the visit is to collect additional documentation for the report and to validate the information provided in the self-evaluation report. Often the site visit concentrates on collecting information about matters not sufficiently covered or elaborated in the self-evaluation report, and gaining a clear picture of how the different stakeholders have been involved in the self-evaluation process, including their individual perspectives.

As the above description of each type of documentation reveals, the combination of sources implies that the evaluation model applied by EVA includes both qualitative and quantitative approaches.

With reference to the organisation of EVA's evaluations, and forms of documentation, the balance between external and internal elements, in relation to the evaluation object, can be summarised as shown in figure 3.

Figure 3
Balance between external and internal elements

	Internal	External
Ownership		X
Evaluator responsibility	X	X

Ownership can be characterised as exclusively external, as EVA is responsible for the initiation, planning and framework of the evaluation. Conversely, EVA's evaluation model implies a shared evaluator responsibility, in that the area under evaluation is partly evaluated through self-evaluation and partly through analysis of the documentation material by the external evaluation group.

As illustrated above, EVA's evaluations are generally characterised by being based on comprehensive documentation material. This emphasis on documentation reflects the view that a solid base of documentation is a prerequisite to the making of any quality judgements, and to the continued development of the evaluated object.

In terms of the purpose of the evaluations, the self-evaluation process most strongly supports the development perspective, and the site visits has the clearest control perspective. The function of the user surveys and the organisation of the evaluations can be placed somewhere in between. The dual purpose of EVA's evaluations thus reveals itself in the selected combination of documentation forms and the organisation of the evaluation. It is also reflected in the characteristics of the reporting of the evaluation, and the requirements for its follow-up.

As mentioned, each evaluation is documented in a report. In the report, the evaluation group presents its conclusions and recommendations for the improvement of the evaluated objects. Furthermore, the report contains a description of the purpose and process of the evaluation, as well as a presentation, assessment and analysis of the different forms of documentation that support it.

Most recommendations are directed towards the evaluated object but may also be directed towards political decision makers. Recommendations must be viable, constructive and realistic. Furthermore, they should be given a clear priority, and preferably it should be evident which recommendations are essential in the short term and in the longer term. Finally, it should be clear who is responsible for follow-up or implementation.

The report is presented to the parties connected with the evaluated object prior to publication, thus providing the immediate stakeholders with the opportunity to comment on the evaluation method and factual errors, if any. The evaluation group may redraft parts of the report in the light of points raised during consultation.

The prime responsibility for follow-up is placed with the evaluated entities. Since 2001, it is now a legal requirement that all self-evaluating entities describe how they intend to follow-up the recommendations of the evaluation report, and they must publish this description on the Internet no later than six months after the publication of the report. This does not apply to Universities.

Past and future

5 years ago, only higher education was subjected to systematic external evaluation. EVA's predecessor, the Danish Centre for Evaluation of Higher Education, was responsible for the systematic evaluation of all higher education. As mentioned earlier, the Centre was set up in 1992 to evaluate all higher education programmes within a 7-year period. When the Centre was established, the idea was that the Centre would start a new round of evaluations after the first 7-year period. This rotation principle would assure the follow-up to the first evaluations.

All evaluations were carried out according to the four-stage model, which includes: self-evaluation, an external expert panel (a steering committee), site visits and a public report. In addition, user surveys were conducted, and the reports of the external examiners were analysed.

The guidelines for self-evaluation contained a number of themes that were relevant for the evaluation of all higher education programmes. The guidelines were, however, adapted to the specific context of a programme when necessary. The definition of quality was fitness for purpose. The evaluations thus took their starting point in the relevant legislation for a programme and the objectives formulated at programme level. Conclusions and recommendations were primarily made in this context.

The purpose of the evaluations was twofold:

Firstly, the evaluations aimed at providing the evaluated programmes with a contribution to their internal quality development. Development was primarily enhanced through a self-evaluation process where the programmes prepared a report on their strengths and weaknesses in relation to the themes presented in the guidelines for self-evaluation. As all programmes within one educational field were evaluated at the same time, the evaluations also sought to contribute to the development of the quality of the programme at national level where relevant. The reports nearly always contained recommendations to the Ministry of Education relating to the need for quality development at the national level.

Secondly, the evaluations also had a control function in as much as they provided the stakeholders with an insight into the quality of the evaluated programmes, through the publication of the evaluation reports. The control function was emphasised through the site visits and the

implementation of user surveys, e.g. among representatives of employers, students or graduates.

The evaluations were thus characterised by having the same focus and the methods were, to a large extent, standardised.

The double purpose of the evaluations has been retained, and the evaluations still build on the same methodological elements, even though EVA's mandate is broader than that of its predecessor.

In the near future the purpose of evaluations will continue to be twofold. They will still have to contribute to the quality improvement of the evaluated units in particular, and the evaluated field in general. Furthermore, the evaluations will continue to have a control function, as they inform stakeholders in the broad sense, both in Denmark and abroad, of the quality-status in the evaluated field.

The majority of evaluations will still use the objectives formulated at the national, local and institutional level as their starting point. However, due to international developments there will be increased interest in the results of education and in creating a higher degree of transparency of education quality across borders. There will be a need for quality definitions that are understandable and acceptable across borders, and it will be necessary to develop other ways of describing quality than in terms of fitness for purpose. One of the means to obtain this higher level of transparency is through pre-defined criteria as the basis for evaluations; another is to focus on output measures, where it is easily identifiable whether expected targets have been met.

Last but not least, there will be an increased focus on competences as another means of making quality judgements comparable, i.e. what are the pupils or students capable of when they have gone through a particular programme at a certain level of the education system. This changes the focus of evaluation from the structures of education to the curriculum and the teaching methods, and the outcomes of teaching and learning.

There will be continued focus on the procedures set up by the institutions themselves to continuously check and improve the quality of their activities and structures. Consequently, there will be a need for external quality assurance to check the effectiveness and sustainability of these internal mechanisms, and to undertake measures that give an input to the improvement activities initiated by the institutions through audit activities. However, that will not be sufficient, due to the European, or even international, demand for comparable assessment of quality. Therefore, there will still be a need to initiate evaluation activities at subject or programme level, but with a transnational dimension.

With the increased international dimension in education, educational systems are becoming more and more complex. Therefore, there will also be a future need for broadness in the foci of evaluations, and in the corresponding methodological elements applied to assess these foci.

Northern Ireland

Stanley Goudie
Assistant Chief Inspector
The Education and Training Inspectorate, Northern Ireland

Values and purposes

The Education and Training Inspectorate (Inspectorate) provides inspection services and information concerning the quality of education and training in Northern Ireland (NI) to the following departments:

- Department of Education (DE);
- Department of Culture Arts and Leisure (DCAL);
- Department for Employment and Learning (DEL).

The organisation is a unitary inspectorate, providing independent advice to all three departments. The legal basis for the work of the Inspectorate is set out in the Education Reform (Northern Ireland) Order 1989 (Article Number 30).

The mission statement of the Inspectorate is, 'Promoting Improvement (in the interest of all learners)'; and the organisation's vision is, 'The Inspectorate will be a highly regarded and influential body, dedicated fully to the education and well being of all learners; its members will treat with respect and consideration, all those with whom they come in contact'.

The purpose of inspection is to promote the highest possible standards of learning, teaching and achievement throughout the education and training sectors, including teacher education within the five higher education institutions, but excluding the education provided by universities per se. In serving this purpose, the Inspectorate:

- provides objective professional evaluation, based on the findings of inspection, of the quality of learning and teaching, including the standards achieved by learners;
- identifies and reports on educational developments and practices, and evaluates their influence on the quality of learning and teaching;
- provides evaluative comment on the influence and outcomes of the policies of the three Departments on education and training services;
- prepares reports on individual organisations, and overview reports on aspects of educational and training provision.

The Inspectorate's mission, vision, purposes, values and service standards are set out in its Charter for Inspection, published in October 2002.

The organisations to be inspected are identified by the Inspectorate. In recent years, however, those being inspected have become increasingly involved in determining the nature, scope, and timing of inspection. These important changes have been facilitated by the development of a range of types of inspection in addition to the more traditional general and focused inspections. The range includes District Inspections, Two-Part Focused Inspections, Unannounced Inspections, Quality Assurance Inspections, and Self-Evaluative Follow-up Inspections.

Evaluation by the Inspectorate is understood to mean the making of informed judgements about the quality of provision and standards in education and training within and across Northern Ireland. Such judgements are based mainly on first hand evidence of observed practice in a range of settings, including classrooms, lecture theatres, youth clubs and workshops, and are

supported, for example, by evidence gleaned from the scrutiny of examination data and documentation provided by the organisations involved.

To help develop the quality of provision in education and training in NI, the Inspectorate has sought increasingly in recent years to foster an ethos of self-evaluation in schools and other organisations. In 1999, for example, the Inspectorate republished its indicators of strong and weak practice in the second edition of the document 'Evaluating Schools'. This was followed by a complementary series of documents 'Evaluating Subjects', as well as 'Evaluating Pastoral Care'; these papers again set out the indicators of strong and weak practice, this time at subject level and in relation to Pastoral Care. Similar documents are available to support the concept of self-evaluation in organisations other than primary and post-primary schools. In further education and in training, for example, there are the 'Improving Quality: Raising Standards' materials, while comparable materials are currently being finalised for use in the pre-school and Youth sectors.

Periodic reports by the Inspectorate on specific issues in education and training also identify aspects of provision which need to be improved, and set out the characteristics of good practice in a subject area, or in an area of provision (such as education of pupils in the sixth form), against which schools and other organisations can evaluate themselves. One example of a series of documents, 'Improving Subjects', was published in 2001, and these leaflets summarise strengths and weaknesses in subject provision in post-primary schools, as well as priorities for action in relation to key areas for improvement. Again, this information is intended to help schools to reflect on and improve their own work.

In late March/early April 2003, the Inspectorate published, through a series of 14 conferences across NI, a set of materials entitled 'Together Towards Improvement'. These materials were prepared in consultation with representatives from the Curriculum Advisory and Support Services (CASS) of the Education and Library Boards (ELBs), the Regional Training Unit (RTU), and the Council for Catholic Maintained Schools (CCMS), and, crucially, with serving principals from primary, post-primary and special schools across Northern Ireland. The materials have been designed to support schools with the process of self-evaluation leading to self-improvement. A key feature of the materials is the inclusion, for the first time in a published document, of a range of the Inspectorate's quality indicators for schools.

In September/October 2003, the Inspectorate intends to publish a series of eight sector-based Digital Versatile Disks (DVDs) as further resources to help support those within the education and training sectors with the process of self-evaluation. Once again the materials have been developed in consultation with serving practitioners and with CASS, and provide guidance on self-evaluation in relation to pre-school, primary, post-primary, special education, youth work, alternative education provision, and vocational education and training.

The developing practices and the materials described above have served to bring about a shared language and understanding across NI of what constitutes good practice, and of the process of self-evaluation leading to self-improvement. Furthermore, they have served to demonstrate the importance of professional dialogue involving organisations, the Inspectorate and the support services in helping to promote lasting improvement in learning and teaching, and in the standards achieved by the learners. The long-term aim is that self-evaluation will become an integral part of all inspection activity, with the Inspectorate taking due cognisance of an organisation's own agenda for improvement in the design and implementation of inspection. In this way, it is hoped that schools and others will increasingly regard inspection as an integral part of the process of continuous improvement, and not as an event to be endured for a period of time.

Objects

The Inspectorate inspects and reports on the quality of provision in individual organisations across the education and training sectors in NI. A wide range of organisations is inspected including:

- pre-school centres;
- nursery, primary, post-primary and special schools;
- further education colleges;
- training organisations;
- the youth and community sector;
- initial teacher education institutions;
- the Curriculum Advisory and Support Services (CASS) of the Education and Library Boards (ELBs);
- the NI Council for the Curriculum, Qualifications and Assessment (CCEA);
- the educational support services, including educational welfare, the educational psychology services, and the peripatetic outreach services.

The materials described in the previous section, along with the structure and content of published reports and pre-inspection briefings by Reporting Inspectors, serve to guide organisations as to the nature, content, scope and purpose of inspection. This guidance has been enhanced significantly by the recent publication of the aforementioned 'Together Towards Improvement' materials, which provide a range of the Inspectorate's quality indicators in relation to Ethos, Learning and Teaching, and Leadership and Management in schools. And, as indicated earlier, there are plans for the publication of similar materials for organisations other than schools, including pre-school centres and youth organisations.

The Inspectorate, in consultation with serving practitioners from across the education and training sectors in NI, is preparing to publish a Common Framework for Inspection in October/November 2003. The generic framework will provide information on the nature, content, scope and purpose of the range of types of inspection available, and include electronic links to further, phase-specific information to complement the generic guidance. The framework will also consider the inter-play of internal and external evaluation, and highlight the benefits of inspection to organisational improvement.

The content and nature of the inspection programme take account of the Citizens' and Parents' Charters, and of initiatives which the Government has introduced, for example, the School Improvement Programme, the New Deal Initiative and the Pre-school Expansion Programme. In addition, the inspection programme is designed or adjusted to provide evidence on issues about which the Departments and their Ministers require particular advice. In such circumstances, surveys, involving visits to a representative sample of organisations, are often the main means of gathering the evidence required, resulting in the publication of a cross-cutting report; in these circumstances, separate reports on the individual organisations involved are not usually provided, although each organisation generally receives a short statement of the Inspectorate's findings. Each year, the scope of inspection activity is outlined in the Inspectorate's business plan.

Stakeholders

As indicated initially, the Inspectorate provides inspection services and information about the quality of education and training in NI for the Department of Education (DE), the Department of Culture, Arts and Leisure (DCAL), and the Department of Employment and Learning (DEL); and the inspection programme is designed or adjusted in order to be able to provide evidence on issues about which the Departments or their Ministers require particular advice.

Evaluations are also provided for individual organisations via, for example, spoken reports to subject departments, senior management teams, and school governors; and to these groups,

the employing authorities, parents and the general public by way of written reports of inspection and of follow-up inspections. In the primary phase, brief spoken reports are provided for each individual teacher at the end of inspection. In the case of cross-cutting reports, derived from survey work, the conclusions to the published reports make it clear that the evaluations are to help the organisations, the Departments and the support services establish a common agenda for improvement. The conclusions also indicate those with the lead responsibility for dealing with particular, identified areas for improvement. Within the last few years, the Inspectorate has taken the lead in communicating the findings of cross-cutting reports directly to those most likely to effect change. Following a recent survey of provision for primary science and technology in a sample of primary schools across NI, for example, the main findings of the published report were relayed directly to the head teachers and science and technology co-ordinators in all primary schools in Northern Ireland. This was done through a series of seminars, organised jointly with specialist officers from CASS.

In June 2003, the now biannual review of education and training in NI by the Chief Inspector is due to be published. Prior to publication, the Chief Inspector will relay the main findings of the report to an invited audience which will include the relevant Ministers, the Permanent Secretaries of DE, DCAL and DEL, and key representatives from the various political parties, the employing authorities and from business and commerce in Northern Ireland.

However, in keeping with its mission statement 'Promoting Improvement (in the interest of all learners)', the Inspectorate views the learners as the key beneficiaries of its evaluations across the education and training sectors.

Methods

As indicated earlier, the organisations to be inspected are identified by the Inspectorate and in recent years, schools and other organisations have become increasingly involved in determining the nature, scope, and timing of inspection. These important changes have been facilitated by the development of a range of types of inspection, in addition to the more traditional general and focused inspections. Furthermore, the increased range of types of inspection has increased the involvement of schools and other organisations in contributing to the assessment of the quality of their own provision. As with general and focused inspections, the findings of District Inspections, Two-part focused inspections, Quality Assurance Inspections, and Self-evaluative Follow-up inspections, are made public.

District Inspections are visits to schools, which are completed by the inspector who has the overview responsibility for a group of schools (the District Inspector). The visits focus on a single theme, e.g. Special Education Needs. A spoken report is given to the school by the inspector, and a short written report on the individual school is also produced. In addition, a report is published on the complete findings (i.e. the overview). The key purposes of the District Inspections are:

- to evaluate and report on a specific aspect of educational provision;
- to further develop links among District Inspectors, head teachers and schools; and
- to encourage and enable schools to effectively monitor and evaluate aspects of their own provision.

As part of the Inspectorate's commitment to its programme of review and continuous improvement, all participating schools have been asked to evaluate the usefulness of the District Inspection. The responses have been positive. In particular, head teachers reported that the District Inspection helped them to evaluate aspects of their work, and they identified action which they had taken afterwards to effect further review and improvement of their work.

In the two-part focused inspection, one aspect of the inspection focuses on whole-school issues, including the school's improvement work, and the second element focuses on a sample of subjects. In the case of the subject inspection, heads of department complete a self-evaluation profile of the work in their department. The quality of the profiles has varied in line with the schools' stages of development in reflecting critically on their own practice, but, re-

gardless of the depth of thinking behind the completed document, the profiles serve as a starting point for discussion between the Inspectorate and the subject department, and can, when completed honestly and in a positive spirit, make a valuable contribution to inspection evidence.

In an Unannounced Inspection, the school is given no prior notice of the Inspectorate's arrival. This type of inspection focuses on pastoral care as evidenced in the school's arrangements and on the work observed. On the first day of the inspection, the school is asked to complete a self-evaluative profile, the content of which will form the basis of the inspection. The completion of the profile is intended to provide the head teacher and staff with the opportunity to contribute to the evaluation process, and to encourage them to be self-reflective practitioners. The profile prompts are arranged under four headings: Ethos, Provision, Management Arrangements, Further Strengths. Under "Ethos", for example, one of the prompts asks the principal to describe and evaluate the ways in which the school takes action to promote good relationships, good standards of behaviour, the self-confidence and self-esteem of the pupils, and clear lines of communication with external support agencies.

The Quality Assurance Inspection is a model of inspection which places the school at the centre of the evaluation process. In quality assurance inspections, the school is more or less in the lead role regarding evaluation; it identifies an aspect of its work, evaluates it and writes a report of its findings. The Inspectorate is then requested, by the school, to evaluate the validity of those findings and the quality of the school's own self-evaluation process, i.e. the Inspectorate has a largely quality assurance role. Following the Quality Assurance Inspection, the Inspectorate publishes a report which states: whether the public can (or cannot) have confidence in the school's self-evaluation; the findings of that evaluation; and on the action proposed to promote improvement.

Of all the forms of inspection, the Quality Assurance Inspection links most closely external evaluation by the Inspectorate to schools' self-evaluation. However, it does take considerable self-confidence on the part of a school to invite in the Inspectorate - and also considerable courage to undertake the standard of rigorous self-evaluation which is needed before the step is taken. To date, this form of inspection has been taking place in post-primary schools only, with one exception in the primary phase. Quality Assurance Inspection has, however, been used by CASS to self-evaluate the service's contribution to the improvement made by schools in the School Support Programme, a programme described in more detail in the case description.

Responsibility to follow-up the findings of inspection evaluations rests with the Inspectorate, and members of the original inspection team return to the organisation within a period of 18 months to two years to undertake a follow-up inspection. The Inspectorate also makes the decision as to the need, or otherwise, to undertake a follow-up inspection.

A growing number of schools are availing themselves of the opportunity to engage in a self-evaluative follow-up inspection. In this type of inspection, the school prepares a written report outlining its own evaluation of the progress it believes it has made in response to the areas for improvement identified by inspection. The Inspectorate then quality assures the schools' findings on the basis of a return visit to the school; the return visit includes classroom observations and the scrutiny of any supporting evidence provided by the school, and the findings are published in a written report.

In Northern Ireland, there is an integrated, partnership-based approach to teacher education. The partnership involves student and newly qualified teachers, principals, CASS of the ELBs, the five higher education institutions (HEIs), teacher-tutors and teachers, all co-ordinating support for student and newly qualified teachers in schools, and school governors. A central place is given to the acquisition of professional competences through the partnerships between the HEIs, the ELBs and the schools, based on common principles. These principles are as follows:

- There should be a common approach to initial teacher education (ITE) across the HEIs.

- As far as practicable, there should be a common profile of competences, used by all HEIs, ELBs and other employing bodies, and all schools – which will underpin the student and newly qualified teachers' professional development.
- The development of links between ITE, Induction and Early Professional Development should enhance the career-long professionalism of teachers and reinforce co-operation among all the partners.
- Underlying the teacher competences are certain "core" qualities of the teacher that enable her/him to employ the individual competences and apply them in a professional context.

It is recognised that the professional development of teachers is a dynamic and complex activity. The professional competences will be acquired progressively during the three stages of early teacher education:

1. Initial Teacher Education – whether a four-year Bachelor of Education (BEd) degree, or a one-year Post-Graduate Certificate of Education (PGCE)
2. Induction – the first year of teaching
3. Early Professional Development – the second and third years of teaching

In order to promote the highest possible standards in early teacher education, the Inspectorate inspects the quality of the partnership arrangements between all of the HEIs, ELBs and schools. At the initial teacher education stage, this involves inspecting the quality of education provided by the HEIs and the efficiency and effectiveness of their partnership arrangements with their partner schools. At the induction stage, the Inspectorate inspects the quality of the support provided to newly qualified teachers by the ELBs and, at the early professional development stage, the quality of the support provided by the schools. In all of this work, the professional competences and the core qualities provide a set of expectations, common to the providers of teacher education and the Inspectorate. To promote the highest quality of continuing professional development for serving teachers, the Inspectorate also inspects the quality of the support provided by CASS. It also inspects the quality of the provision for training future principals in Northern Ireland through the Professional Qualification for Headship (NI).

The procedures and criteria used by the Inspectorate to make its evaluations have been agreed by the members of the organisation and are used and followed closely during inspection. The procedures for inspection have been published and, as indicated earlier, the Inspectorate is in the process of placing its quality indicators in the public domain. A regular programme of staff development and training helps ensure consistency in the use of inspection criteria, and in the outworking of the agreed inspection procedures.

The approaches to evaluation are both quantitative and qualitative. In assessing the various features of an organisation (e.g. Ethos, Learning and Teaching, and Leadership and Management), inspectors relate their evaluations to four grades which may be interpreted as follows:

Grade	Criteria
1	Significant Strengths: good (ranging to outstanding)
2	Strengths Outweigh Weaknesses: satisfactory (ranging to good)
3	Weaknesses Outweigh Strengths: fair (ranging to satisfactory)
4	Significant Weaknesses: poor

The grades, along with supporting text, are recorded subsequently in an electronic database. The grading system is in the public domain but, as yet, the grades awarded to an organisation during inspection are not routinely shared with the organisation, nor do they form part of the published inspection report.

A number of quantitative terms are used in inspection reports to comment on aspects of provision in relation to percentage bands; these terms are in the public domain, and are set out in the table below.

Almost/nearly all	More than 90%
Most	75% to 90%
A Majority	50% to 74%
A significant minority	30% to 49%
A minority	10% to 29%

Part of the Inspectorate's vision statement articulates the organisation's determination to be '… dedicated fully to the education and well-being of all learners; …'. This core value of the organisation is reflected in a concern to evaluate and report on the quality of the process of learning, as well as the appropriateness of the learning outcomes. Equally, the organisation is concerned to evaluate the quality of pastoral care, including the learners' personal, social and health education, as well as their academic achievements. The final phrase of the vision statement – '…its members will treat with respect and consideration, all those with whom they come in contact.' – gives voice to the standard of conduct expected of members of the organisation as they go about their core business, namely the evaluation of learning, teaching and the standards achieved. In addition to direct observation of teaching and learning, evidence about aspects of the life and work of a school is obtained through the use of confidential questionnaires completed by parents, and through meetings with pupils, parents and the governors of the school.

Past and future

Some five years ago, the main types of inspection were the general inspection (GI) and the focused inspection (FI), with general inspections being the dominant form of inspection used. The GI involves an evaluation of the life and work of an organisation as a whole, and is very labour intensive in terms of the number of inspectors involved. The FI involves an evaluation of one or more aspects of the life and work of the organisation, considered against the backcloth of the organisation as a whole.

Over time, the balance between the use of general and focused inspections has shifted significantly, with general inspections now rarely undertaken. In the school sector, this shift in the balance between GIs and FIs was prompted, initially, by the need for the Inspectorate to meet increased Government targets for inspection, while working within the constraints of existing staffing resources. In addition, there was growing evidence from schools that they preferred FIs to GIs, and also a realisation on the part of the Inspectorate of the worth of FIs in giving valuable insights into the life and work of a school, and in a more cost-effective way than GIs. The evaluation of pastoral care was of increasing significance in inspection, irrespective of the type of inspection involved; and the Inspectorate was in the process of developing a broader suite of inspections, a suite which included greater consideration of the contribution of self-evaluation to the inspection process.

The Inspectorate had also begun to further enhance the extent of published information relating to inspection, and also guidance materials to help, in particular, schools and colleges of further education to self-evaluate the quality of their own provision. In addition, the Inspectorate was beginning to increase the involvement of parents and governors in the inspection process, and to include lay members and Associate Assessors (on a pilot basis) in inspection teams. For school inspections, Associate Assessors were usually experienced head teachers or vice-principals, and their participation in inspection allowed them to evaluate practice at first hand in other contexts. Their involvement in inspections in this way also helped them to, subsequently, lead self-evaluation more effectively in their own schools.

The last five years have seen significant developments in inspection. The core values of the Inspectorate remain the same, however, as does the core business of the organisation, namely the evaluation of learning and teaching, and of the standards achieved by learners in whatever

guise they come. The developments over the last number of years have helped consolidate much of the Inspectorate's thinking about its role and function, and provided a sound basis on which to build further developments, some of which are considered below.

The Charter for Inspection, published by the Inspectorate in October 2002, sets out the organisation's mission, vision, purposes, values and service standards. This will be followed in October/November 2003 with the Common Framework for inspection, designed to provide information on the nature, content, scope and purposes of inspection. The new framework will also consider the inter-play of internal and external evaluation, and highlight the benefits of inspection to organisational improvement. It is anticipated that the Charter and the framework, taken together, will direct and influence the work of the Inspectorate over the next five years and beyond.

In five years from now, it is anticipated that self-evaluation will be further embedded within inspection, and that the Inspectorate's evaluation of the capability of an organisation to self-evaluate will become a much more significant part of inspecting and reporting. Such a development makes a great deal of sense, i.e. that the Inspectorate should take due cognisance of an organisation's own priorities for improvement, as set out in its development plan, within the inspection process. It is also anticipated that such a strategy would enable schools and other organisations to see inspection more as process than an event, thereby helping to strengthen the link between inspection and subsequent improvement in learning and teaching, and in the standards achieved.

Within the same time frame, it is envisaged that the full array of the Inspectorate's quality indicators will be in the public domain, and that the grades awarded to the various aspects of an organisation's provision will become part of the published inspection report. This latter development is seen as a logical step from the current position, where the grading system is in the public domain, but the grades are not shared routinely with the organisations inspected. Furthermore, the recently piloted strategy, whereby the evaluation of inspection is carried out by a firm of independent consultants, and the findings made public, will become an integral part of the Inspectorate's way of working.

The Inspectorate will continue to keep inspection under review in order to ensure that the models of inspection used are fit-for-purpose. Whatever changes occur, however, it is envisaged that the core business and core values of the organisation will remain unchanged and, consequently, the promotion of improvement in the interest of all learners will remain central to the work of the Inspectorate.

Canada – The School Sector

Richard M. Jones
Associate Director, Assessment & Reporting
Education Quality and Accountability Office (EQAO), Ontario

Canada is a confederation of ten provinces and three territories. Areas of responsibility are divided between the federal and provincial/territorial governments. The federal level is responsible for portfolios such as national defense, external affairs, fisheries, shipping and railways, telecommunications, energy and the banking and monetary systems. The provinces and territories are responsible for areas such as social services, health, forestry, highways and education.

Hence, there is no federal department of education. This means that each provincial/territorial jurisdiction establishes its own policies related to aspects such as curriculum, teacher certification, school accreditation and reporting of student learning progress. School boards set their local policies within the framework of policies established by the provincial/territorial ministries of education. These policies include aspects such as school operations, curriculum implementation, and the hiring of teachers and support staff. Decisions concerning instruction and classroom practices are made at the school and classroom level. Since there are thirteen provincial/territorial jurisdictions in Canada, this article attempts to summarize education and evaluation systems across all of these jurisdictions.

Public education is provided free to all Canadian citizens and permanent residents until the end of secondary school — usually age eighteen. The ages of compulsory education vary from one jurisdiction to another, but generally, schooling is mandatory from age six or seven to age sixteen.

Canada's public education systems are divided into two levels: elementary and secondary. Most jurisdictions offer a one-year kindergarten program for children prior to grade one. Some school systems offer a two–year program of junior and senior kindergarten. Children aged three to six are in these programs. Elementary education is subdivided into the primary level, including grades one to three, the junior level, including grades four to six and the intermediate level, including grades seven and eight or grades seven to ten. Secondary education is generally subdivided into junior high school, including grades seven to nine and senior high school, including grades ten to twelve (grade eleven in the province of Quebec). In some systems, secondary school begins at grade seven or eight; in others elementary school continues until grade eight.

Educational evaluation – an overview

Student assessment in Canadian schools utilizes a broad range of methods to obtain information about educational achievement for formative and summative purposes. Most provinces/territories have established systems of formal assessments that involve program assessment (providing information about student achievement in key grades and subject areas to education stakeholders for accountability purposes and/or providing information to teachers, students and parents for the purposes of improving student learning and achievement) and/or credentialing examinations required for high school graduation.

At the national level, the provincial/territorial ministers of education have established an organization called the Council of Ministers of Education, Canada (CMEC) to provide a forum for communication on issues of common concern, such as education funding, education programs and assessment. In the early 1990s, the CMEC established a national testing program called the School Achievement Indicators Program (SAIP) in response to the ministers' need for compara-

tive information about student achievement in the areas of reading, writing, mathematics and science. The assessment, developed and coordinated by a consortium of provinces, usually involves the administration of one annual assessment in one of the subjects previously mentioned. The assessment is administered to randomly selected samples of schools and 13 and 16-year-old students. Information on student achievement is reported only at the provincial/territorial and national levels. Because of the sampling framework, no achievement information is available at the individual student, school or school board levels. Further information about the SAIP is available on the CMEC Web site (www.cmec.ca).

From time to time, Canada and/or individual provinces/territories participate in international assessments, such as the Trends in International Mathematics and Science Study (TIMSS) and the Progress in International Reading Literacy Study (PIRLS) conducted under the auspices of the International Association for the Evaluation of Educational Achievement (IEA), as well as the Programme for International Student Assessment (PISA) conducted by the Organisation for Economic Co-operation and Development (OECD). When the country or one or more of its jurisdictions becomes involved in an international assessment, the project is coordinated within Canada by the jurisdictions themselves, through the CMEC, or through another Pan-Canadian organization such as Statistics Canada.

When individual jurisdictions conduct evaluations, or when national testing is administered through the CMEC, the jurisdictions' Education Acts form the legal or statutory basis for this work. Generally, provincial/territorial Education Acts make school participation mandatory.

Values and purposes

In their summary of the purposes of centralized student testing across Canada, Taylor and Tubianosa (2001) found that the main purposes were as follows:

- to monitor the education system;
- to identify strengths and weaknesses;
- to certify and promote;
- to provide information for curriculum design, instructional methodology and resource allocation;
- to gauge change/progress over time;
- to give direction for further research.

To make some determination about the quality or health of the education system, it is necessary to have appropriate indicators/measures of attainment. Some performance indicators such as retention, participation and graduation rates are often available internally to jurisdictions' ministries of education. However, other comparative measures, such as student achievement information, are seen as important additional elements that help educators and the public make decisions about system quality and effectiveness.

Evaluation has the potential to provide information about areas of strength and weakness in student achievement. This type of information can be used to inform future directions for program planning at the jurisdictional and local levels, as well as implementation and/or modification to classroom instructional practices. Program evaluations, based on the intended curriculum, can provide such information in an objective fashion.

Examinations are sometimes used for the purposes of certification/credentialing and promotion. If developed and administered appropriately, they can ensure a level of fairness in the evaluation of students. High stakes tests such as this can be used to place students in programs, classes and schools, based on their achievement, and they can indicate whether a student will be promoted to the next grade or whether a student will graduate from high school.

Information from program evaluations is an important reference for persons such as curriculum developers as they plan for curriculum review and revision. This type of information is also invaluable to those who develop instructional resource materials for classroom use.

In order to determine whether student achievement is improving over time, and whether changes to curriculum and other educational reforms are having the desired effect, it is important to have access to objective information about performance trends. The administration of well-designed, centralized assessments can be the source of such information.

Educational researchers are interested in studying how students learn most effectively and what factors/variables have an impact on, or are associated with, student achievement and attitudes. Secondary analysis of results from centralized program assessments can provide a useful source of data for this type of work.

Areas of consensus and concern

Generally speaking, there is a consensus that educational evaluation, if designed and administered appropriately, has the potential to provide important information related to the foregoing purposes. However, there are those (primarily in the teaching profession, but also some in the broader community) who are concerned about issues related to the impact on the curriculum, test construction, and the misuse of results and other unintended consequences.

Some critics of large-scale student assessment claim that, because of the restrictions imposed by paper-and-pen testing formats, only certain knowledge and skills can be assessed. Another concern is that if teachers teach to the test, then some important curricular elements may be neglected. Others are concerned that since the focus of many testing programs is on so-called basic skill areas, such as reading, writing, mathematics and science, other important subjects that are not usually tested may be considered of lesser importance in the school program.

Many large-scale testing programs, including most provincial/territorial learning assessments, credentialing programs and national and international assessments, are comprised of a variety of test item formats to gather information on a wide range of student knowledge and skills. Some critics, however, contend that assessments that rely heavily on multiple-choice items result in an emphasis on lower-order thinking skills. In addition, there is a concern that if tests are inappropriately constructed, test bias, where items are culture, language, or gender specific, may result in unfairness or advantage to certain groups of students. Other criticisms of large-scale student evaluation relate to the misuse of data, particularly the inappropriate ranking of school results, the unintended consequences of testing, such as stress and anxiety on the part of students and teachers, and the impact of results on grade retention and dropout rates.

Objects

Assessment programs

All of the provinces (except Prince Edward Island) use formal, large-scale assessments to obtain systematic information about student achievement. Most provincial assessment programs involve some form of program/learning assessment, as well as credentialing examinations. However, there are differences in various aspects, such as the subject areas and grade levels tested, the schedule of administration and design elements, including whether all students (a census) or a random sample of students are assessed. The following table, adapted from Taylor and Tubianosa (2001), summarizes the general features of assessment programs among Canadian jurisdictions.

Provincial Assessment Programs

Province	Subject(s)	Grade(s)	Design	Schedule
Alberta	***Provincial Assessments***			
	Math, English Language Arts,	3, 6, & 9	Census	Annual
	French, Language Arts, Science, Social Studies	6 & 9		
	Credentialing Exams			
	11 courses	12	Census	Three times per year
	6 of the courses			
British Columbia	***Provincial Assessments***			
	Reading, Writing, Numeracy	4, 7, & 10	Census	Annual
	Other Subjects	As needed	Sample	As needed
	Credentialing Exams			
	Numerous Subjects	12	Census	Five times per year
Manitoba	***Provincial Assessments***			
	English Program			
	Reading & Numeracy	3	Census	Annual
	Francais Program			
	Lecture, Notions de calcul	3	Census	Annual
	French Immersion Program			
	Reading, Notions de calcul	3	Census	Annual
	Lecture	4	Census	Annual
	Credentialing Exams			
	English Language Arts, Immersion and Anglais Standards Tests	6	Optional	June 5-7, 2001
	Francais langue premiere and Francais langue second, Immersion Standards Tests	6	Optional	May 29-31, 2001
	Math	Secondary 1	Optional	Semester 1, January 24, 2001 Semester 2 June 6, 2001
	English Language Arts Standards Tests	Secondary 4	Sample	May 2001 June 2002
	Immersion Standards Tests			
	Math/mathematiques			
New Brunswick	***Provincial Assessments***			
	Language Arts, Math, Science	3	Census	Annual
	Language Arts, Math, Science, Writing	5	Census	Annual
	French Immersion, Language Arts	6	Census	Annual
	Math and English (Lang. Arts)	8	Census	Annual
	Credentialing Exams			
	Math & English	11	Census	January, June, & August

Newfoundland	***Provincial Assessments***			
	Math, Science, Core French, Writing	3, 6, & 9	Census	Annual
	Canadian Test of Basic Skills			
	Language Arts, Work Study Skills, Science, Math	4, 7, 10, & 12	Census	Four-year cycle, annually
	Credentialing Exams			
	All Subjects	12	Census	Annual, June
Nova Scotia	***Provincial Assessments***			
	Math	5 & 8	Census	Every two years
	Language Arts	6 & 9	Census	Annual
	Credentialing Exams			
	English Language Arts, Science, Math	12	Census	Annual
Ontario	***Provincial Assessments***			
	Reading, Writing, Math	3 & 6	Census	Annual
	Math	9	Census	Twice per year
	Credentialing Exams			
	Literacy (Reading & Writing)	10	Census	Annual
Quebec	***Compulsory Exams***			
	French, Language of Instruction	6, Secondary 3	Census	Annual (June)
	Ministry Exams			
	Language of Instruction, Second language, History of Quebec & Canada, Physical Sciences	Secondary 4, 5	Census	Annual (January, June, & August)
	Complementary Exams			
	French, English Language Arts, Math Secondary level French, Language of Instruction, Math	3, 6	Census	Annual (June)
Saskatchwan	***Provincial Assessments***			
	Math, Language Arts	5, 8, & 11	Sample	Twice/year
	Technological Literacy, Critical/Creative Thinking	5, 8, & 11	Sample	1999, 2001
	Credentialing Exams			
	All Subjects	12	Census	Annual

Other initiatives

In addition to these provincial assessment and credentialing examination programs, some jurisdictions have implemented other initiatives to support public accountability and educational system improvement. For example, for many years the province of British Columbia conducted a system of school accreditation in which members of the school community and an external evaluation team reported on the effectiveness (strengths and weaknesses) of various elements associated with the school. This program operated on a mandatory six-year cycle. However, recently the province abandoned school accreditation in favour of a school district review process.

Recognizing that effective school and school board planning are important components of the improvement process, Ontario mandates that all school boards in the province submit annual improvement plans to the province's Education Quality and Accountability Office (EQAO), the

agency that administers all provincial, national and international assessments within the province. School board improvement plans are reviewed according to specific criteria, and reports are provided back to boards to inform them of the extent to which their improvement plans conform to the required content.

In addition, most Canadian provinces/territories have implemented education quality indicators programs. The programs all vary to some extent, but in general, data and information that have some relevance to education and student achievement are collected from a variety of sources (e.g., surveys, Statistics Canada, Ministry of Education and other government departments) and are reported in a variety of ways. Generally speaking, the data provide contextual information that assists education stakeholders in their interpretation of student achievement results. Some examples of education indicators are socio-economic information; perceptions of school climate/school safety; perceptions of school leadership; and home learning supports, such as access to a computer, quiet study space, and access to books in the home. In some instances, education indicators are reported in their own separate reports. In other cases, education indicators are released concurrently with student achievement data.

Stakeholders

The preceding table indicates that the vast majority of assessment programs and all of the credentialing examinations involve all-student or census testing. This means that in most cases, individual student reports are received/available to students, parents/guardians, and teachers. (In some instances, school administrators and guidance counselors may also see individual student results.)

In the great majority of cases, schools and school boards receive summaries of their own results, and summaries of provincial achievement results are reported by means of print and/or Web-based documents to the public at large, the education community, the Ministry of Education (including curriculum developers) and the media.

Generally, evaluation results are viewed as useful for education planning and improvement purposes, as well as for public accountability. There are always concerns expressed by some about the potential for inappropriate uses of data, such as rank-ordering schools and school boards, and using results to evaluate teachers.

Methods

Policy Design and Implementation

Within the provinces and territories, ministries or departments of education have taken responsibility for developing student assessment policy for both assessment in the classroom and for more formal large-scale provincial student assessments. In some cases, ministries have established detailed assessment policy frameworks, including quite detailed guiding principles. In other cases, ministries and school boards have less formal assessment policies but promote the use of fair testing practices, e.g. through quoting documents such as the *Principles of Fair Student Assessment Practices for Education in Canada* (1993 and 2000).

According to Perry Fagan and Spurrell (1995), in their study of policies and practices of ministries and school boards across the country:

> *"The extent of the jurisdictions' responsibility for student evaluation, policy design and implementation varied depending on the type of evaluation or assessment covered by the policy. For example, there appeared to be a much more specific set of policies for provincial examinations than for other types of evaluation. In five of the provinces with provincial examinations, the department of education is responsible for both design and implementation of policy. In two provinces, the department and the boards are responsible, and in the two territories, the examinations are the responsibility of the department of two other provinces. In other areas of assessment policy, the involvement of others in the system was much more evident. Some provinces indicated that the department and boards were*

equally involved in policy development and design, and others indicated that the department, boards, and schools were all involved. In some cases, where there were no provincial examinations, ministry officials indicated that only boards, or schools, or both, were responsible for assessment policy….All but one of the 12 jurisdictions indicated that they had methods in place for ensuring that policies, practices, and guidelines were implemented." (p.16)

The authors identified nine general approaches that the jurisdictions used to implement policy, and these included:

- creating minister's orders supplemented by policy guidelines;
- publishing documents, handbooks and providing guidelines;
- conducting on-site visits and reviews;
- providing in-service sessions, workshops and seminars;
- providing information and reporting results of provincial examinations and assessments;
- forming specialist implementation teams for new policies;
- assigning coordinators for administering examinations;
- holding centralized marking and training sessions;
- collecting data and reviewing results from school evaluations;
- reviewing school board plans for improvement.

Approach to assessment

Generally, with regard to provincial/territorial assessments, the respective ministries/departments of education are responsible for the evaluations and their methods. Most ministries/departments have student assessment branches whose mandate is to design, administer and report on the assessments. Some ministries rely on external experts to conduct various aspects of the assessments (e.g. item and test development, statistical analysis); however, ultimate responsibility for the assessment remains with the ministry. One jurisdiction (Ontario) has created an office for education quality and accountability to assume responsibility for the design and conduct of provincial assessment initiatives.

Again, generally speaking, provincial assessments are criterion-referenced and are based on the learning objectives and standards outlined in the provincial curriculum. While evaluation approaches are usually more quantitative than qualitative, they are, nevertheless, increasingly becoming both results- and process-oriented.

There is usually a close correspondence between the values (the reasons and purposes of the evaluation) and the method (the way the evaluations are planned and implemented). For example, in Ontario, two guiding principles of the work of the Education Quality and Accountability Office are to provide information for public accountability purposes, as well as for the improvement of student learning. Consequently, it is not enough to simply provide overall scores in various subject areas for individual schools, school boards and the province. Information on strengths and weaknesses is also provided to individual students so that they, their teachers and parents can work together to help students improve.

In all cases, the results of provincial learning assessments are made public through various means such as press releases and public reports. The extent of and approach to follow-up activities related to the evaluations, however, is variable. In some cases, specific recommendations are not directed to stakeholders, and once assessment results are publicly released, there is little formal follow-up to determine what, if any, action has been taken. In other cases, specific recommendations are made to specific stakeholders, and steps are taken to follow up to determine what has been done toward improvement of the system. For example, in Ontario, the annual, provincial reports of the assessments of Grades 3, 6, 9 and 10 provide specific recommendations to stakeholder groups; letters are sent to these organizations, highlighting the recommendations that are specific to them; and, through various means (e.g. discussions at

advisory committee meetings, review of school board improvement plans), indications are obtained about the extent to which recommendations are being attended to.

Past and future

In recent years, there have been some general changes in Canadian large-scale assessment practices. Over the past decade, assessment and evaluation programs have become the focal point of many provincial educational reform agendas. Many Canadian jurisdictions (e.g. Ontario) have either established new assessment programs or expanded or modified existing ones.

In describing recent changes in evaluation across Canada, Perry Fagan and Spurrell (1995) observe that there has been a general trend, both in classroom and large-scale assessment, toward integrating curriculum, instruction and evaluation to ensure that assessment closely matches learning objectives. This is evidenced by most jurisdictions' reliance on criterion- and curriculum-referenced assessment as opposed to norm-referenced testing.

Although most large-scale assessment programs (and also most classroom assessments) continue to rely heavily on paper-and-pen testing approaches, there has been a trend in recent years toward looking for ways to move away from a focus on paper-based assessment. Evidence of this is in practical/performance-based assessment that goes beyond extended written responses to demonstrations of skills acquisition. Current education and assessment theory advocates the use of more authentic, performance-based approaches to assessment, and certainly the use of a variety of assessment item types, even when paper-based assessments are used. Most provincial assessment programs across Canada now contain a blend of item types.

Whereas formerly, large-scale assessment was conducted primarily to generate test results for accountability purposes, in recent years, there has been a move toward having large-scale assessment results support planning for educational improvements at all levels: province, school board and school. It appears as though this trend will continue into the foreseeable future. The move toward assessment for both accountability and improvement has had some major impacts. For instance, it has become increasingly important that the assessments provide an even greater amount of detail about student achievement related to curricular expectations and standards for the purpose of achievement target-setting at all levels. Also, there has been a realization that other information, not just student achievement data, is required if planning for the purpose of improvement is to be conducted, and this has given rise to education quality indicators projects in most Canadian jurisdictions. Consequently, multiple sources of data are needed, and these include sources such as surveys and databases housed in organizations like Statistics Canada and ministries of education and other government departments.

Provincially, nationally and internationally, discussions are being held about how on-line/computer-based assessments can be implemented. Certainly there are numerous technological and other challenges associated with this approach, but this appears to be a direction for future development. The use of technology in assessment, however, is a reality today. Many jurisdictions (such as Ontario) use computer technology as a tool in test item and test development.

Summary

In recent years, all Canadian provincial and territorial ministries of education have taken a more active role in developing student assessment policy for both provincial and classroom assessment purposes. In general, this activity has been associated with provincial education reform initiatives and the perceived need for credible information about student/system performance and curriculum implementation that can support planning, decision-making and policy development at all levels.

References

Author (1996). *Education Quality and Accountability Office: Policy Framework.* Toronto, ON: A policy framework developed by the members of the Education Quality and Accountability Office's Board of Directors.

Earl, L., & Graham, N. (1994). *OECD—School Performance Standards Study.* Toronto, ON: Unpublished paper prepared for the OECD Study on Performance Standards in Education.

Joint Advisory Committee (1993 and 2000). *Principles for Fair Student Assessment Practices for Education in Canada*. Edmonton, AB: University of Alberta, Centre for Research and Applied Measurement and Evaluation.

McLean, L.D. (1985). *The Craft of Student Evaluation in Canada.* Toronto, ON: Canadian Education Association.

Ministry of Education (1991). *OAC Examination Handbook—English: Language and Literature.* Toronto, ON: Queen's Printer for Ontario.

Mullis, I.V.S., Martin, M.O., Kennedy, A.M., & Flaherty, C.C. (2002). *PIRLS 2001 Encyclopedia: A Reference Guide to Reading Education in Countries Participating in IEA's Progress in International Reading Literacy Study (PIRLS).* Boston, MA: International Study Center, Boston College.

Perry Fagan, L., & Spurrell, D. (1995). *Evaluating Achievement of Senior High School Students in Canada.* Toronto, ON: Canadian Education Association.

Robitaille, D.F. (Ed.) (1997). *National Contexts for Mathematics and Science Education: An Encyclopedia of Education Systems participating in TIMSS.* Vancouver, BC: Pacific Educational Press.

Royal Commission on Learning (1994). *For the Love of Learning: Report of the Royal Commission on Learning.* Toronto, ON: Report prepared for the Minister of Education and Training.

Taylor, A.R., & Tubianosa, T.-S. (2001). *Student Assessment in Canada.* Kelowna, BC: Society for the Advancement of Excellence in Education.

Traub, R. (1994). *Standardized Testing in Canada: A Survey of Standardized Achievement Testing by Ministries of Education and School Boards.* Toronto, ON: Canadian Education Association.

France – The School Sector

Gérard Bonnet
Head of European and International cooperation,
Direction de l'évaluation et de la prospective, Ministère de la jeunesse, de l'éducation nationale et de la recherche, France & Visiting Fellow at the Institute of Education, University of London

Using the concepts defined for the Anthology, and leaving aside the question of the informal assessment of pupils by teachers, the question of examinations, as well as the question of individual inspection of teachers, which do not fall within the remit of this survey, evaluation of school education in France can be described as applying to several areas, e.g. schools and pupils, teaching practices. It covers the school education system as a whole, at national, regional and local levels. Evaluation methods and results are also assessed (meta-evaluation). The value and purpose of evaluation are well defined and understood at a conceptual level, but their practical use is an area of concern.

Evaluation procedures are carried out by three different institutions within the school education sector. Two of these are responsible for actually organising and conducting evaluations (the inspectorate and the education ministry's evaluation division), while the third reviews evaluation findings and methodology (the high council for evaluation).

At a national level, the general inspectors, based in the education ministry, participate in the supervision, training and recruitment of some types of teachers. They are also more generally involved in evaluating the school system, either at the national or regional level. For instance, they are called upon to give their assessment of the overall way in which an education region is performing in terms of different focus areas, e.g. its management processes, its pedagogical provisions, the results of its pupils, etc.

Counselling or assistance for schools is mainly the responsibility of regional inspectors, a group whose main task is to inspect, supervise and assess teaching staff. These inspectors have a regional remit but retain close links with the general inspectors at the education ministry. There are 3,000 regional inspectors, a figure which should be compared with the 880,000 teachers in primary and secondary education (an inspector in the primary sector supervises an average of 300 teachers, while in the secondary sector, inspectors are each responsible for about 750 teachers).

The evaluation division of the education ministry (*direction de l'évaluation et de la prospective*) is in charge of the coordination of the evaluation and forecasting functions of the ministry. This division has overall responsibility for the production and analysis of both educational statistics and qualitative and outcome studies. This implies close collaboration with other divisions of the ministry, the *academies* and other actors in the field, including numerous institutions and research teams. Its remit exceeds what is discussed here, as it is in charge of aspects of evaluation for the higher education and research sectors, in particular with respect to the production of certain statistics and indicators.

The high council for evaluation was set up by the education minister in 2002. Although its members are formally appointed by the education minister, its work is, in practice, independent from ministerial interference. Its remit is to assess evaluation procedures and methodologies, not the education system. To do this, it reviews the different assessments of the education system carried out and produces an annual report in which it makes recommendations concerning the state of educational evaluation. It comments on the assessments carried out and published by the education ministry and by public or private bodies. It may also commission specific assessments from public or private organisations in areas where evaluation

specific assessments from public or private organisations in areas where evaluation may be deficient.

Values and purposes

The status of evaluation was clarified in the 1989 Education Reform Act, which states that evaluation is an integral part of the education system. It is mandatory for educational institutions to take part in surveys organised by the ministry with a view to providing statistics, organising national testing, etc. Regional or local evaluation initiatives are also prescribed for the school education sector. However, the Act does not define the term "evaluation" which is seen in the French context as a blanket term covering what might also be described in other countries as quality assurance or quality development.

At first, educational evaluation was mostly quantitative, i.e. derived from basic statistics, such as teacher-pupil ratio, average spending, etc. In the 1960's the education ministry began setting up a statistical information system which became increasingly sophisticated and comprehensive over the years. Nowadays this system encompasses all school education statistics. From a technical point of view, it makes use of up-to-date technology and relies on a unified definition of statistical concepts so that data are comparable throughout the country. One of its characteristics is that it is based on the data collected for management purposes at all levels of the education system and enables the construction of statistics which can be used for general monitoring and steering purposes.

Later, assessment and evaluation surveys were introduced to give some account of the quality of education. The combination of a powerful statistical information system and of quality assessment has created an evaluation system which describes the state of the education system with rigour and accuracy and allows for the construction of indicators. This concept of evaluation is rather broad since it encompasses all the data available to measure and assess the system. It is used as a tool to help with the management of the system but also as a way of making stakeholders aware of the necessary changes in their professional practices.

Evaluation is also a major tool in the current context of greater power and responsibility being transferred from the centre towards the regional and local levels, since this creates the need for a strong monitoring and steering system.

For well over a century, the French education system was organised centrally (nationally). The education system consists of three main levels of responsibility: (i) the national education ministry, (ii) regional education administration (deceptively known as "*académie*"), which also devolves responsibility for primary education to the district level, and (iii) the school. An OECD study conducted a few years ago in France showed that each of these levels made roughly one third of the decisions.

In addition, elected regional, district and communal assemblies have specific responsibilities for lower and upper secondary schools (*collèges* and *lycées*) and primary schools respectively, mostly in the areas of building maintenance and construction, as well as the employment of some ancillary staff.

Central government retains, at each of the three main levels, overall responsibility for teaching staff and all pedagogical aspects of education, even if some of it is devolved to the lower administrative levels. In other words, the crucial, if subtle, difference is between political transfer of responsibility from national government to local elected assemblies, and executive devolution of parts of the power of central government (the education ministry in this case) to national government representatives at regional, district and communal level. It should be apparent from the above description that the real power still remains with central government through its local representatives, with local political bodies enjoying only partial responsibilities.

The question of devolution of power to regional and local levels has been one of the main organisational concerns of administrative reform in education in France over the past couple of

decades. One of the main thrusts has been the transfer of more and more responsibility from the ministry to regional administration (*académie*). Although it is clearly understood that education in France will remain "national", it has also become apparent that more initiative has to be entrusted to the *académies*. The head of each *académie* (there are 30 in all, including overseas territories) is appointed by the education minister and is his representative in each region (this must not be confused with the notion of a locally elected politician responsible for education, which does not exist in France).

In order to facilitate this process, indicators are needed to allow central government to allocate resources equitably throughout the country and to allow the *académie* level to do the same within a given region.

At the regional level, education indicators are also used to make decisions on employment policies (which are partly the responsibility of the regionally elected politicians) based on the output of the education system (number of school graduates, including vocational training, number of university graduates; qualitative information on the qualifications; etc.). The regional level may also establish its own evaluation procedures in order to make decisions and shape policies.

Next to this role in monitoring and steering the education system, evaluation has other purposes. In a country where it is official policy that education should be of comparable standards wherever you happen to be, it is necessary to measure quality, and evaluation fulfils this role in terms of quality assurance.

However, evaluation also has a quality development component, and it is also used as a tool for change. As such, the emphasis is not so much on the summative aspect as on quality development. The French evaluation system includes procedures to encourage educational change and to promote the dissemination of a culture of evaluation among stakeholders throughout the system. This is based on the belief that such stakeholders, among whom teachers are prominent, will improve their professional practices only if they are shown, as in a mirror, the consequences of their actions. For instance, one form of pupil assessment is conducted in such a way as to involve teachers and schools. It is an educational tool used to initiate changes in professional practices. Other tools are also provided to schools to improve their managerial and pedagogical practices. These mechanisms will be fully discussed below.

Evaluation is understood to encompass all these different aspects. The concepts of quality assurance and quality development are closely intertwined within the concept of evaluation. At least this is how it is understood by the vast majority of stakeholders.

The greater part of evaluation initiatives comes from the ministry, where a specific division was set up for the purpose. The lower levels are under an obligation to participate in the processes put in place. They are also encouraged to have their own arrangements. For instance schools are encouraged to have their own assessment tools to monitor their progress and to make use of them to devise their compulsory "school plans".

Objects

As explained above evaluation in France is meant to cover the whole of the school sector, at all levels and in all its components. It embraces both the public and the private sector, at least that part of the private sector which is government funded (i.e. 90% of the private sector).

All types of evaluation can be found: pupils' achievements and attitudes, institutions, programmes, subjects, themes, professional practices, etc. There are specific arrangements in place for meta-evaluation, which is mostly conducted by the high council for evaluation.

Institutions are subject to evaluation in one-way or another. For instance, schools are assessed from the point of view of their administrative and financial policies but also from the point of view of their performance, in particular of their ability to cater for all types of pupils and bring

them to the end of schooling, and of their examinations results in the context of their socio-economic environment. Even whole administrative areas, such as an *académie*, can be subject to evaluation in term of their educational provisions or, sometimes, even their results (percentage of early school leavers, of school graduates, etc.).

Pupils' achievements are assessed nationally through various testing procedures, either to identify their strengths and weaknesses from a teaching perspective or to give a picture of national or regional performance. The results of the development of academic subjects in schools, national policies and innovations are evaluated (for instance the implementation of reforms such as the introduction of foreign language instruction at primary school or of team work for teachers). Themes like literacy, numeracy, ICT policies, provisions and results are looked at. Sensitive issues, such as early school leavers, or violence in schools, are subject to scrutiny. Longitudinal studies are used to understand specific mechanisms or effects of schooling (selection, pupil counselling, dropping out, etc.).

Teachers' professional practices and motivations are investigated through questionnaires or classroom observation of representative samples (this is distinct from individual inspection which also exists).

Whenever a specific type of evaluation is decided upon at the national level, the subject to be assessed is always very clearly identified and defined after consultation with the relevant bodies interested in the results of the work. When evaluation is decided at regional or local level, there is more flexibility so as to tailor the study to specific needs, which may be different from those of the national level. Likewise, the methods used locally may differ from national ones (except for the collection of statistical data, for which specifications are national) and may produce results that cannot be compared outside the area where they were applied.

Overall it is fair to say that the assessment of pupils' achievements and skills is one of the predominant types of evaluation in France. Pupils are consistently targeted through the whole of school education. The reason for this is that evaluation is largely aimed at providing help for teaching and learning. Next to them, schools and educational administration units are the second preferred evaluation targets. This is because is it felt that there is plenty of room for improvement in the way they operate; and their importance in the process of delivering education is also becoming more and more recognised in a country were there is not much in the way of a tradition for school effectiveness.

In most cases the evaluation subject is chosen on the basis of its relevance to policy and to national or local priorities, due to the need to be very selective when deciding what to evaluate. Evaluation is costly in terms of resources and is an added burden on whoever happens to be the subject. In order to be accepted and properly carried out, it needs to be perceived as serving a specific purpose. Hence, besides the need to obtain information on issues under scrutiny, there is an attempt, as far as possible, to target areas which are directly connected to local and national policy imperatives.

Stakeholders

Because evaluation in France covers such a large domain and variety of purposes, it is actually aimed at a whole range of stakeholders, according to the various aspects of education being dealt with.

Evaluation data are available for policy makers (Parliament, ministers, etc.), for educational administrators at local, regional and national level, and for all the stakeholders of the system (parents, teachers, etc.). At the same time, these data are also made available to the press and the general public in order to facilitate discussion about education as part of the normal democratic process. With the creation of the high council for evaluation to be occupied with meta-evaluation, it should be clear that evaluators themselves are concerned with the assessment of their work. The following are examples of the way specific evaluation is aimed at a particular audience and of the impact on that audience.

Diagnostic national testing (which currently exist in three grades) is meant as a tool for teachers to assess the competence of their own pupils in specific subjects. In order to help them place their pupils in relation to others, the results of a representative sample are analysed by the ministry and then fed back to all the teachers, together with pedagogical comments, so that they can compare the ability of their pupils with the national or regional average pupil in the same grade. This is done in printed and electronic form. It is complemented with a bank of assessment items which teachers can use at will for further diagnosis and which is freely available on the Internet.

In the past the diagnostic testing programme has had positive results on teachers. For instance, following the 1989 primary school mathematics tests, the national sample showed a bias in favour of arithmetic as opposed to geometry even though both subjects were given equal weight in the national curriculum. As a result, when teachers realised that this was a general problem they altered the structure of their teaching, and it was observed the following year that competence in geometry had greatly improved without any outside interference.

At the same time, it has also been observed that teachers, particularly at the secondary school level, make insufficient use of the diagnostic testing programme in their teaching. In fact one particular test for upper secondary school had to be discontinued after a few years because not enough teachers used it.

School indicators have been constructed at ministerial level as tools for head teachers to manage their schools. A procedure similar to that described above enables them to compare their processes and results with those of other comparable schools. Three indicators (performance indicators based on national final examination results) are published each year (and made available on the Internet) to give some idea about the value added to upper secondary schools.

A set of regional indicators is published each year by the ministry to show how *académies* compare. Although this publication is national and widely available, it is also intended to help the *recteurs* (the civil servants who represent the minister in the region) and their staff, as well as regional politicians, to formulate relevant policies.

A similar set of national indicators is published annually and is widely used by national policy makers, administrators, parliamentarians and the press to discuss the state of the education system, and even as a basis for drafting national reforms and policies. These national indicators are also one of the elements which the Treasury takes into account in deciding on the share of the national budget that will be allocated to the education ministry.

It is fair to say that the educational world and French society at large have a clear understanding of the aims and audience for each type of evaluation. The resulting data are eagerly awaited and accepted as useful information by those concerned, in particular the press. Consequently, there is no debate about the way existing evaluation is conducted. What is currently being discussed, in particular through the reports and recommendations of the high council for evaluation, is the extent to which evaluation findings are sufficiently used by policy makers, and what new subjects should be targeted, e.g. teachers, in particular.

Methods

Bearing in mind the definitions used in this anthology, it can be argued that most French evaluation is internal, since it is conducted by the ministry of education itself, and not by external independent agencies. Broadly speaking, civil servants from the evaluation division of the ministry conduct the evaluation, publish the results and make recommendations to the political level. However, it could also be described as external to the extent that schools and other educational institutions do not carry out their own evaluation – if one regards the ministry as external to them. Likewise, self-evaluation is not part of the French tradition of evaluation.

External evaluation *stricto sensu* has been attempted by local authorities or *académies*, in particular for the auditing of upper secondary schools, but this has been mostly unsuccessful. The

only true external evaluation of the efficiency of pedagogical management has just been made by the *Cour des Comptes* (National Audit Office). Its official report has been made public but it tends towards making general recommendations.

As explained earlier, there are basically two institutions responsible for internal evaluation of the French school education system, but their roles and methods are very different: the inspectorate (national or regional) and the evaluation division of the education ministry (*direction de l'évaluation et de la prospective*).

There are two contrasting approaches to the brief of the regional inspectors: the first is to reinforce their role as teacher supervisors with more frequent inspections, as promotion for teachers depends on a successful inspection. The principal objective of the second approach is to develop the role of inspectors as counsellors who support teaching staff and thus participate in the pedagogical management of the *académie*. In view of the administrative tasks that must be added to this dual role, it can be said that a satisfactory equilibrium has not yet been found. Despite these problems, the legitimacy of inspections is widely accepted.

There are basically two distinct types of inspection. Firstly, there are some school inspections, limited in scope, whereby aspects of the school are evaluated (accounting procedures, specific problems such as violence, truancy, etc.). They mainly concern the way the school is run and the work of the head teacher. This type of school inspection has no bearing whatsoever on individual teachers' promotion. The second type of inspection is the individual inspection of teachers where inspectors sit in on a lesson. Individual inspections result in the production of confidential reports, discussed with teachers, which may lead to the teacher concerned being awarded faster promotion than would otherwise happen through seniority.

The evaluation division of the ministry contributes to the evaluation, assessment and management of the school system by developing a reliable and fast information system, describing the input, the processes and activities within the system, as well as the outcomes. It is also in charge of forecasts to provide a long-term view of probable evolutions. It makes available to managers of the system at all levels, including teachers, the diagnostic and analytical tools needed. Finally, it reports on results to stakeholders and to society as a whole.

Alongside its statistical activities, the work of the division is important in several other areas: the assessment of the skills and abilities of pupils; the processes, working climate and results of schools; the professional practices of staff involved in the education; and the evaluation of innovations and reforms.

These activities are not restricted to a national level, since the division is also responsible for French participation in international surveys and the production of statistical data and indicators organised by various bodies (European Commission, EUROSTAT, EURYDICE, OECD, UNESCO, IEA, etc.)

While the type of evaluation carried out by the inspectorate (leaving aside inspection of teachers) is more closely related to a specific situation or problem in a given place, and to that extent is more akin to case studies, the division responsible for evaluation in the ministry deals with the more general picture, based on either exhaustive data or on representative samples.

As described above, the high council for evaluation is in charge of reviewing and assessing evaluation work and the methodology followed. It does not conduct evaluations but makes recommendations to the minister on methodological improvements or needs for new evaluation data. In practice these recommendations are aimed at both the evaluation division and the inspectorate.

Most evaluation policies are decided nationally, either by the inspectorate or by the evaluation division, often acting on behalf of specific groups (ministerial advisors, educational administrators in need of particular information). At the regional level, *recteurs* do also commission work,

which may be carried out by their own staff. Head teachers can organise whatever evaluation of their schools they need.

Since the initiative for evaluation is often national, it is expected that the tools and methods will be provided by the ministry that is responsible for devising and developing them, with possible local adaptation for regional evaluation initiatives.

The methodologies utilised are chosen by the relevant people in the ministry. For instance, the inspectorate defines its own inspection methods which normally include school or classroom visits, interviews with head teachers and teachers. Inspectors may decide to put their inspection work into context by using indicators (produced either directly by, or as a result of, projects encouraged by the evaluation division of the ministry) which give suitable background information on specific schools or *académies*.

Evaluation work run directly by the evaluation division is organised by its own staff. Since pupil performance testing is prominent on the list of evaluation tasks, this may involve constructing nationally or regionally representative samples of pupils or classes. It will also imply devising, together with groups of teachers, inspectors and academics, suitable test instruments. Printing these instruments and addressing them to the relevant schools (or to all schools when the tests are for all pupils of an age group) is also the responsibility of the division, as well as organising coding and data processing of the results.

Other methodologies are also used, such as classroom observation by trained teams of researchers, or questionnaires addressed to samples of head teachers, teachers, trainee teachers and inspectors. Information from pupils and parents is also gathered for specific studies (in particular, longitudinal studies of children), as is, occasionally, information from employers.

Where researchers are involved, their role is twofold. Firstly, they act as scientific advisers to the civil servants of the evaluation division at the time of devising methodologies, constructing instruments or conducting specialised psychometric analyses. Secondly, when the data obtained in the context of a specific operation lend themselves to it, researchers may be contracted to carry out further analyses which may be subsequently published. However, when acting in conjunction with the ministry, they are not responsible for the conclusions or recommendations of the work carried out, or for making the results public.

The results of these various types of evaluation are analysed by the teams responsible for them. Inspectors write their own reports. The evaluation division publishes most of the results from its projects.

For a long time, it has been education ministry policy to publish the results of the evaluation studies that have been conducted under its control, even if some inspection reports are not made public, as they may contain recommendations which may not be accepted or refer to particular persons or institutions. Some examples of the specific audiences for which results are provided have been given above.

More generally, it is considered important for the general public – parents and taxpayers – to be fully informed of evaluation and assessment findings. This is achieved through a vigorous publication policy and frequent informing of the press. This point is capital, as the underlying belief is that mentalities – and beyond them professional practices – will evolve only if a sufficiently accurate and well-informed description of the education system is provided on which to base constructive criticism. The editorial policy of the education ministry has thus been expanded to include the work on educational evaluation conducted by the division responsible for evaluation, which contains a variety of authoritative publications that are often quoted and commented on by the press. The better known of these are: *Repères et Références Statistiques sur les enseignements et les formations*, an annual volume containing all official educational statistical data available*; L'état de l'école* and *Géographie de l'école*, published once a year to provide, respectively, national and regional indicators on the education system; numerous four

or six page *Notes d'information* offer updated figures and analyses on a variety of educational matters and, finally, two more academic publications, *Les dossiers d'Education et Formations* and the journal *Education et Formations*, normally feature detailed accounts and findings of recent surveys, in the former, and individual articles on specific subjects, in the latter.

Publishing evaluation results is one thing, making recommendations for action is quite another. Both the inspectorate and the evaluation division make their findings available to ministers and make, where appropriate, relevant suggestions. But it is up to those in charge of actual policies to decide whether or not to act on them. This is one of the main weaknesses of the evaluation of the school sector in France: a lot of the evaluation work is not acted upon. While there is ample data on all aspects of the school education system, which clearly suggest where changes should be implemented, and which describe success or failure of policies and lead to clear recommendations, it is not easy to establish a definite link between this work and policy decisions.

The publication and dissemination of results are of particular interest to managers and decision-makers at the regional and national levels. Taking the example of the national testing programme, the results provide a wealth of fundamental information: there are disparities between groups of pupils at the end of primary school and at entry into lower secondary schools; girls are ahead of boys in French, while, in mathematics, the results of girls and boys have become almost equivalent over the years; disparities in results remain marked over the years; there are geographical disparities at the end of primary school and at entry into lower secondary school, etc.

A few decisions were made based on these findings, for instance, the introduction of a different structure for primary education, the implementation of specific measures for pupils with learning difficulties, and the reorganisation of lower secondary schools. However, it is true that, as a rule, decisions affecting pedagogical organisation and teaching content are not related to the results of national tests. Despite considerable progress, the links are still insufficient between assessment and aspects such as the design of programmes.

This observation can be extended to other aspects of evaluation. Many different elements contribute to the management of the education system. They are, however, disparate and insufficiently connected. It is necessary to seek more co-ordination between the different bodies responsible for evaluation, e.g. to establish links between pupils' assessment results, the design of curricula and the training of teachers. Educational structures and managerial practices should, in turn, take greater account of pupil's outcomes.

Two recent independent reports have outlined the fact that the findings of evaluations are not sufficiently taken into account by policy makers, be they politicians or senior administrators.

Regarding evaluation criteria and procedures, it should be apparent from what has been said that they are very coherent and tend to remain the same, in particular, for the sake of comparability over time where this is possible, but naturally they are updated and adapted when necessary. Although the approach to evaluation is based on reliable statistical data, and is often presented in a quantitative way, more qualitative studies are also conducted, in particular, in the context of case studies of teachers' professional practices, as one example. In all cases where this is possible or meaningful, a qualitative analysis and commentary is provided to help evaluation stakeholders make the best practical use of the findings.

There is a strong bias in evaluation work in favour of results. It is now widely accepted that results are the most relevant yardstick by which to judge any education system. At the same time, results need to be presented and understood in their specific context. Educational processes need, therefore, to be taken into account. Hence the provision of input and process indicators in, for instance, the evaluation of schools, or, as indicated before, the evaluation of teaching processes.

What you evaluate, the reasons why you evaluate it and the way it is done are very closely interwoven. Devising methodological approaches for evaluation is far from being merely a technical question. It is highly political, as the results of what is measured are dependent, in part, on how it is measured. There should be a clear understanding of which different types of methodologies measure best, and the final choice of a particular methodology should be soundly based. One approach is to use national test or examination results to give an absolute measure of the performance of individual schools. Another, that which is used in France, is to provide a relative judgement in the form of a measure of value added, i.e. taking into account the socio-economic origins of the pupils and the various characteristics of the school. The former intends to show parents which school they might choose; the latter tries to show parents, head teachers and staff what is actually achieved to help improve the school. Two different objectives translated into two different methods.

It is, therefore, fundamental to carry out a thorough reflection on the relationship between values and methods before deciding on a particular course to implement an evaluation. A strict link between evaluation values and methods is all the more important, as it is a well-known fact that what is measured tends to become the norm. Education systems are more or less slowly, and more or less directly, changed by evaluation results. To measure the outcomes or processes of education in a market oriented perceptive will inevitably steer the whole system in that direction. France has been careful to define and implement a view of educational evaluation that places the emphasis on improvement through tools for change, rather than on change through judgement.

Past and future

Evaluation has been firmly established in France for many years. It found a new impetus in the second half of the 1980's with the creation of the evaluation division in the ministry, which overtook a mostly statistical division, and with the 1989 Education Act. By the mid 1990's the various evaluation values, methods and procedures that we now know were in place, amid a general political consensus on these issues. In 1997, a socialist education minister was appointed and decided that the evaluation work (statistical and qualitative) carried out in, by and for the ministry was not to be trusted because it was politically biased – which almost everyone else agreed it was not. This coincided with clumsy attempts by some civil servants to try and utilise the results of the diagnostic mass assessment tests to give a summative measure of pupil performance and, furthermore, to do this by comparing the results produced over several years. It is explained in the case description that this is not possible, as these assessments have been devised for other purposes. Inevitably the comparisons showed a sharp decline in performance from one year to the next, since the assessments had been devised precisely to identify and measure pupils' weakest points, and not their strong points. In the ensuing intellectual and political confusion, evaluation was put on the back burner for a couple of years, until a different minister (of the same, socialist, party) was appointed, after which it regained the ground it had lost and even found increased status.

Apart from this politically inspired eclipse, evaluation five years ago was pretty much what it is today, with the exception that the high council for evaluation did not exist then. The main strains of evaluation policy were already defined and most were implemented; they are currently being further developed. Evaluation has not changed very much in the intervening years, there is simply more of it now. This does not mean that new areas for evaluative work are not being defined, and these should start being put into effect in the next few years.

As regards pupil assessment, it is likely that the twofold approach, which has been followed so far (mass diagnostic tests and summative sample-based assessment), will continue to prevail in the future. It has been found that they correspond to two distinct needs and, as such, should be upheld. The diagnostic branch will probably to be called upon to help with the current national literacy and numeracy policies; the diagnostic assessment item bank is poised to play a major part in the process. The summative branch is also destined to expand to provide even more monitoring information. It is likely to develop within the areas of skills in modern foreign languages (to measure the impact of the current policy of introducing foreign language at an

early age at primary school), the assessment of cross-curricular and non-curricular skills (e.g. learning to learn skills, civics), while trying to expose pupil motivation and behaviour (e.g. violence), etc. The impact of motivation and behaviour on school climate and pupil achievement has clearly emerged in recent years and calls for more data to be provided.

School evaluation will probably try to go beyond the observation of differences, to put forward ideas for interpretation and opportunities for improvement, which may be laid down in a specific procedure that could lead to the intervention of a team from outside the school. This procedure could be based, firstly, on the data and indicators that exist at present and, secondly, on supplementary information collected in accordance with an analytical grid, yet to be defined, and to include self-assessment. In this way it should be possible to increase the overall efficiency of the education system, and ensure that it is more equitable. Resources could be made available so as to make the services provided by each school equivalent and as good as possible, i.e. to ensure that all schools strive towards the success of all their pupils whatever their social origin, academic entry level and the courses they are offered. This school evaluation procedure would necessarily result in a compulsory programme of action with which the school would have to comply.

Based on sound statistical surveys and qualitative evaluation, making forecasts about future trends in the education and training system has been revealed as a major tool for policy development. In order to exploit this, it is necessary to conduct very sophisticated regional and national surveys of the demographic and economic factors, in particular the job market, which are relevant to education policy. This in turn has to be analysed in the light of qualitative interpretation. These are no easy tasks to undertake, but they are necessary to provide local and national policy makers with the indicators they need. This is an area where increased co-operation with inspectors could, and probably will be sought.

The greatest challenge facing evaluation in the years ahead is undoubtedly to convince stakeholders to make use of it. It has been argued here that French school sector evaluation is basically devised as a tool to promote and facilitate change, and to monitor and steer the system. However, teachers, administrators and policy makers, although they all recognise the quality of the tool and the correctness of the findings, remain reluctant to draw the necessary conclusions and initiate the necessary evolution. This is not specific to the educational sphere but applies also in other areas of the French administration, mainly for cultural reasons. Nonetheless, it appears more and more unreasonable to spend a lot of time, money and effort to devise and implement sophisticated evaluation procedures if a worrying proportion of the outcomes are not acted upon. Progress has been made, but a lot remains to be done, in particular among teachers and school administrators. In particular, efforts should be made to integrate evaluation and assessment, in the sense applied here, into teacher training courses, in which such notions and skills are curiously absent at the moment.

Finally, it should also be noted that international, and more specifically European, comparative evaluation will inevitably develop further and become more and more relevant for the development of education systems. This is an opportunity that should be welcomed, as it will help steer the school sector towards new directions which can only be beneficial for its future.

References

Bonnet, G. (1996) "Effects of Evaluation Procedures on Educational Policy Decisions in France", in *International Journal of Educational Research*, Vol. 25, n° 3, pp. 249-256, Elsevier Science Ltd (Pergamon), Oxford and New York

Bonnet, G. (1997) "Context and Challenges of Educational Research and Policy in France", in *The Role of Research in Mature Education Systems*, pp. 32-38, National Foundation for Educational Research in England and Wales, Slough.

Loi d'orientation sur l'éducation, 10 juillet 1989 (articles 18, 25, 26, 27 and 28).

Ministère de l'éducation nationale (1996) 3 indicateurs de performance des lycées, in *Les dossiers d'Education et Formations*, n° 66 (Paris, Direction de l'évaluation et de la prospective).

Ministère de l'éducation nationale (2001) Géographie de l'Ecole, n° 7. Direction de la programmation et du développement, Paris.

Ministère de l'éducation nationale (2002) L'état de l'Ecole, n° 12. Direction de la programmation et du développement, Paris.

Ministère de l'éducation nationale (2002) Projection du système éducatif à dix ans, in *Education et Formations*, n° 63. Direction de la programmation et du développement, Paris.

Thélot, C. (1993) L'évaluation du système éducatif. Nathan, Paris.

Hungary – The School Sector

Peter Vari
Director
National Institute of Public Educational Services, Center for Evaluation Studies, Hungary

Values and purposes

Current Hungarian evaluation practice is based on two elements. The first one is a professional tradition that comes from participating in several international surveys. Hungary was the first among the former socialist countries to join the international student achievement surveys, initiated by the IEA in the 1970s. Participating in these surveys gave important experience and knowledge towards establishing the evaluation culture in our country. The Monitor Surveys (the first Hungarian evaluation system) have been developed on this basis. The main features of the surveys are the following: their goal is to measure student achievement in different levels of the educational system (usually in every second year), both in the primary and secondary schools in the fields of reading literacy, mathematics literacy, science, and ICT. This survey system has existed since 1986.

The organizer of these surveys was, and still is, the Center of Evaluation Studies, founded by the Ministry of Education as one of its support institutions at the end of 1960s. The name and structure of this support institute has changed several times over the last three decades, and the number of staff has also changed (from 2 to 12 persons), but the basic tasks and aim of the work of the Centre has remained the same. The Centre now exists as a part of "Árpád Kiss" National Institute of Public Educational Services, with a staff of 12 researchers.

Apart from the Monitor Surveys, there were and are some research projects in the field of evaluation being carried out at the pedagogical departments of certain universities. This research has been developed mainly for diagnostic purposes, but it could not become regular and systematic due to a lack of financial and infrastructural capacities.

International surveys were and are a very important point of reference for judging how successful the Hungarian educational system is in terms of international comparison. Databases and methods stemming from international surveys also provide great opportunities for analyses and evaluations on a national level.

Although Hungary has been continuously participating in surveys organized by the IEA since the early 1970s, and regularly a party to the evaluations organized by the OECD, and has had its national monitoring system for more than 15 years, there was not enough attention paid to questions regarding evaluation in the educational system, until the 1990s. There was only a small group of professionals and researchers who were interested in this field. Publications, results and problems were known by only a very small group of educational specialists.

Regular and well-developed evaluation practices had not been developed, either on different levels of the educational system (e.g. institutional level evaluation) or in different spheres of education (e.g. evaluation of educational instruments).

Before the political changes, at the beginning of the 1990s, a system of inspectors existed in the rather centralised educational system that was meant to ensure the quality of education. Although the system was not able to satisfy this function fully, there were no efforts to develop other elements of quality assurance.

The importance and need for a new system became apparent after the changes, parallel to the abolishment of the inspection system, and with the decentralisation of the educational system. The new documents of educational regulation and legislation formulated this need at the beginning of 1990s.

Given the significantly increased local and institutional independence in the nineties, issues concerning the quality and success of education were raised with more and more emphasis. The external monitoring of schools declined and standardised testing was not introduced until the end of the decade.

Different investigations indicated the fall of test-based school performance, and the growth of the already present discrepancy between location and performance. Several international comparative analyses suggested that one of the weak points of Hungarian education was quality assurance. Concerns about the quality and success of education were only increased following the modification of regulations on content and the resulting transitional conditions.

Though the Public Education Act 1993 prescribed the mandatory evaluation of public educational institutions as part of the education authority's responsibility, there was at the time insufficient professional background for this task. There were no experts, instruments, or evaluation bodies, which could help education authorities, or the schools themselves, with the problem.

The Amendment of the Public Education Act in 1996 detailed the evaluation tasks of the different levels of stakeholders, but the fact was that in everyday practice there were not many real evaluation events or initiatives, except for the national and international surveys mentioned above.

One of the reasons for this phenomena was, probably, that after the political changes, there were numerous reforms, initiatives, changes, and natural problems in other parts of the educational system (for example developing the new National Core Curriculum), which held back development in the field of evaluation of the whole system and everyday practice. The decentralisation of the educational system, the establishment of a new structure for financing and management, the new elements of administrative and control systems, and the development of the new national curriculum presented too many problems to solve in the educational sector.

But, as a result of all these processes, an intense professional and social debate evolved concerning the possible reasons for the inadequacy of quality and the availability of government support in the interests of quality assurance.

The quality assurance issue received special attention in the Amendment to the Public Education Act 1999. At the same time, there began a remarkable process of adaptation of quality assurance and evaluation tools, already in use in other countries. The private sector played an increasing role in the development of a quality assurance system for public education, a role that was consciously backed by government policies. In the quality assurance business, most of the current players are members of the professional sphere of education, with a smaller segment attached to private business rather than the public service sector.

The aims of quality assurance initiatives within education are twofold: the control of quality; the assurance of development. Apart from being familiar with the results of evaluation processes, it has become essential to bring into focus the culture of evaluation, its tools and the possibilities arising from it. Besides the institutions working under the Ministry of Education and professionals who have received different levels of training, a competitive private market in evaluation work has also emerged.

Although the legal framework states the aims of evaluation for all levels of public education, as well as the experience and competence of those executing evaluation processes, by the end of the 1990s one of the primary aims of state educational policy had become the establishing of a quality assurance system at the institutional level. Nationwide assessment at this level coincided with, and already covered a certain section of the desired evaluation work. In order to further the process, the "Comenius 2000" program was established, which, through a competitive tender process and professional administration, secured the creation of a quality assurance system at the institutional level.

At the same time, professionals also responded to the growing culture of quality assurance: pedagogical institutions began transforming their work towards a profile of assessment and evaluation, providing assessment services to schools through the organization of subject surveys, often on the basis of commissions arising from both the institutions and local education authorities. Experts and assessors were also trained more specifically for evaluation work.

According to present interpretations, *control* means the comparison of the actual functioning of public educational institutions with current laws, regulations and the institutions' pedagogical programs. *Evaluation*, on the other hand, is the comparison of the institutions' achievements with the professional aims set up by the very same institutions, their education authorities and the sector's leadership. *Quality assurance* "is the activity whereby the public educational institution continually assures that professional aims and the functioning of the institution come closer to each other as a means of satisfying the needs of teachers, parents, educators, education authorities and the labour market (Paragraph (5), (9) and (23) of § 121, Public Education Act 1993/LXXIX)".

However, having accepted quality assurance as a key term in public education, Hungary has taken a significant step in the formation of views adopted by the public on the success of education. Quality assurance, as a term, is used in two senses both in and outside Hungary. Firstly, it refers to organizational and management techniques at the institutional level, and secondly, it refers to all devices and mechanisms which serve to ensure the successful operation of public education as a whole. In judging quality and success, it is especially important how professional objectives are set, as these are referred to in the evaluation phase.

Having said this, the use and interpretation of terms is still not uniform and well organised. Similarly, there is no real consensus as to the expectations and interpretations of the various actors in the field. Development and organisation will be a challenge for public education over the coming years.

There are two basic standpoints in Hungarian education in terms of aims and values. One of them goes further than the sole evaluation of achievements in traditional curricular subjects and gives an ever-stronger emphasis to cross-curricular competence, literacy and practical knowledge. But there is also an ever-stronger emphasis on the evaluation of achievements compared to earlier achievements, the judging of development or added value, as opposed to assessing performance solely in terms of curricular requirements.

By the turn of the millennia, the basic criteria of a national assessment system had been established: a legal framework, definition of terms (even though their interpretation is not yet fully effective) and the basic professional basis for carrying out evaluation. Of course, assessment work in Hungary is still in its beginnings. The examining of developments in curricular program packages, textbooks, educational tools or other means of public education is still not resolved, not even in terms of initiatives.

But the most urgent tasks are to synchronize assessment work at the different levels and terrains of education; to ensure cooperation between the various agents; and to ensure the coherence of applied tools and their maximum exploitation. Assessment at the institutional level is still limited, but it is being increasingly enhanced by the so-called National Assessment of Basic Competences, described in detail in the "case" chapter.

Reflecting on the hypothesis set up by the editors of the anthology, we can make the following statements about assessment practices in the Hungarian public education system:

- The relationship between education authorities and control agencies is not yet crucial in terms of assessment as education authorities (the local council authorities) request controls in only a few cases.
- The Ministry of Education, as a representative of the aims of, and values connected with, educational policies, and as the body that defines the content regulations of educational institutions, is decisive regarding the aims of the assessments it commissions and the values formulated in these reports. That is to say, in Hungary, management and control are not necessarily in the same hands, and thus neither are values nor the relationships between them.
- Neither does the culture of any given educational sector play a decisive role in evaluation. The "culture" of the entire public educational system is still in the making (exceptions are the national and international assessment studies and the assessments made by university workshops), and so there are no differences between the different sectors.
- Institutional evaluation is also in its early years, but one can expect that the local establishment of values and tools will have a decisive role here. The institution's own pedagogi-

cal program, and the aims formulated in it, can influence the assessment goals and methods.

- In the case of Hungary, one can assume that certain values and methods are of such importance in the educational system that they are simply not questioned. Among them is, at the moment, the issue of key competences, and the basic need that the effectiveness and success of the overall educational system, certain forms of training and institutions, has to be measured and followed-up.

Objects

The object of the evaluation can be defined through various criteria. In accordance with the preceding comments regarding assessment practices in the Hungarian public education system, we can present the following summary:

Evaluation at a national and international level is aimed at the whole of public education and thus makes possible an evaluation of the *system* as a whole. Results may give us a picture of the achievements of schools in different social catchments, or the achievements of different school types. These periodically reoccurring measurements also make the analysis of trends possible.

Schools participating in a survey receive feedback in the form of "school reports" on the performance of the studied pupil group. These reports then serve as references for assessment at the institutional level.

In national surveys, *private schools* are not treated separately but as part of the nationwide sample, in so far as education within the private school/institution is pursued according to the National Core Curriculum. In Hungary, around 90 percent of public educational institutions are managed by the local authorities; only about 10 percent of schools are run by foundations or churches, as "private" schools. On the other hand, if a private school sets its curriculum according the National Core Curriculum, the school becomes eligible for state funding. In this way, education in most, so-called, private schools basically fits into the curriculum of the non-private system. It is only due to their better financial, personnel and infrastructural situation that they can achieve better performance. In brief, private schools do not warrant classification as a separate group in terms of assessment.

In case of the above-mentioned surveys (executed mainly by pedagogical institutions; e.g. Budapest Pedagogical Institute) of curricular subjects, the *object* of the assessment can be clearly determined. It is according to the demands of the schools or school authorities that the knowledge, the performance within the given curricular subjects is measured, mostly with reference to the expectations formulated in the curriculum. Results are, of course, reported back at institutional level. Should the need arise, the conductor of the survey can also provide professional advice to the school.

It is mostly in the field of competence, that measurement and evaluation of certain *topics* can be found. Surveys conducted in the national sample basically fall into this category as they tend not to survey curricular subject areas, but fields of efficiency and competences in a broader, more general sense: reading literacy, mathematics, science literacy.

The subject of the survey is thus basically defined: the framework of international surveys always meticulously outlines it. In the case of national surveys, the first step in all cases is the definition of the research area, the aim and the method. The demands and goals of the current educational policy, the particularities of the area to be surveyed and the possibilities stemming from the size of the sample always play a role in the forming of these cornerstones. Naturally, within this framework, different approaches coexist which will further diversify the definition of the basic elements of evaluation.

Stakeholders

Nationwide surveys are usually commissioned by the political apparatus and are usually prepared considering the aims of national educational policy. However, the results are published to the whole of the educational community (in the form of national reports), to the whole of the nation via the press, and to the participating schools who receive tailor-made school reports. It is a matter for the school to decide how, and in what circle it makes its own results public.

The surveys (methods, tools, processing) and the results, on the other hand, present an important professional challenge to the assessing bodies, research institutes and researchers. To fulfil the professional aims to the highest quality is a basic task for all participants.

Surveys at an institutional level aim at the exposure of a national state of affairs. For such outcomes, they give reliable results at an institutional level, in the form of the school reports mentioned above. There are not surveys of individual student levels, as this is not an aim of the research. It is possible, however, to evaluate individual persons at the local-level and through institutional assessments, and such evaluation is also part of everyday school practice and exams.

Society as a whole and its taxpayers can also be considered as implicated in terms of the evaluations. Society's interest in the educational system (one of the most important areas of society) is traditionally deep. Taxpayers are informed through the daily and specialized media about the results of surveys. In addition, a ranking of certain secondary-school level educational institutes is available each year as a performance indicator. These lists, however, do not rank schools according to a range of criteria but only according to performances in the higher education entry exams.

Results of national and international surveys may be published to parents and/or pupils, according to the individual decision of each school. A number of schools publish the results of surveys via their Web page, school yearbook, or school magazine. Other schools, however, on occasion, do not even publish the outcome amongst their own staff.

As has been mentioned before, the political and administrative sector is a standard participant in the surveys, as the agent who commissions the survey, finances it and will eventually use the outcomes. One of the principal addressees is thus the political elite.

Where the commissioner of an evaluation is an agent from another level of the government system, then it depends on their particular aims and demands as to what circle the outcomes may be published within.

Methods

Surveys are usually initiated by the Ministry of Education, as it determines the test-evaluation tasks of public education agents as set out by the Public Education Act, or through commissioning particular national and/or international assessment works. Commissions for national and international surveys of study achievement are generally handed to the Center for Evaluation Studies of Árpád Kiss National Institute of Public Education Services (KÁOKSZI), the research workshop that has vast experience in this field. The Center is responsible for the forming of methods and tools of evaluation, for the accomplishment of the survey, the publication of the results and their follow-up.

In the case of international surveys, the Center is responsible for maintaining international relations, preparing tools for surveying, organizing and completing the survey, and for the closing of the project (marking process, data entry, cleaning and finalizing database). Following international data processing standards, it is also the Center who processes national data, adapts international results and publishes them.

Regarding national development assessments, the Center formulates the framework, the test-matrix and test-design, prepares and tests items, constructs the tests and chooses the sample(s). Surveys are, in all cases, carried out on representative samples. In the case of primary schools, the basic stratification is in terms of pupil catchment area type, whereas for secondary schools, it is school type. The tests include exercises with closed-end, multiple-choice, open-end, shorter and longer answers. Answers are processed and analyzed after being centrally coded.

The center has the possibility, but also the obligation, to make suggestions in light of the results. On the other hand, due to the size and complexity of national and international surveys, it is necessary to have recommendations and initiatives from other sources, too, regarding future measures (curricular reforms, text book reforms, reform of the structure or basic characteristics of the educational system).

The Ministry, or other participants in the educational field (local authority, school board), may take the initiative concerning other assessment works, in which case the questions of method,

analysis and publication are defined according to arrangements between commissioner and executor.

Results of international and national surveys are, in all cases, published through the previously mentioned channels (i.e. school reports, articles in the daily and specialized media, at symposiums, in reports to the Ministry), meanwhile in the case of other assessments, the method of publication depends on arrangements between the participating agents. For surveys on a district or provincial level, schools participating in the test will obviously get to know the results; if a local authority commissions the survey, it will also receive the results. Apart from this, it is within the remit of the executor of the survey to decide how and in what circle the findings should be published.

The criteria and the process of the evaluation are defined in the investigations of the Center for Evaluation Studies. However, considering the needs of the political apparatus (as commissioner) and the conclusions of earlier assessments, there may be slight changes in any particular surveys.

According to the aims, and due to the methods applied in order to achieve these goals, measuring is quantitative. Qualitative evaluations may be expected primarily in the emerging field of "school self-evaluation".

Nationwide evaluations are generally result-orientated, i.e. pupil performance is investigated and tested. National studies consider processes as changes in the trends of performance. Assessments that place major emphasis on process can be applied to evaluations on a local or institutional level.

There is no strong and theoretically backed relationship in the Hungarian evaluation system between values and methods. On the other hand, it is obvious that international and national representative tests on performance can be accomplished with the use of methods that are being currently applied, and so, besides the aims, the "genre" or evaluation-type defines the method.

Surveys at other levels and in other areas may apply methods in a more flexible manner, and the surveyors may decide which method they find most useful in order to reach the given goal. However, in many cases, it is financial and human capacity, resources and possibilities that define the methods to be used.

We believe that it is not so much the values, but the immediate aims and available possibilities that define the method.

The following is a brief description of the main evaluation methods and procedures to be found in the Hungarian educational system. We shall describe the latest initiative, "National ABC" in the "case" chapter, and, therefore, do not treat it in this section.

The National Monitor Surveys

One of the most important elements of quality assessment in Hungary is the so-called „Monitor Surveys", applied since the 1980s. The main features of the surveys are as follows: they measure student achievement in different grades of the educational system, both in primary and secondary schools, in the field of reading literacy, mathematics literacy, science, and ICT. The Ministry of Education grants the commission to organize and carry out the survey, usually every second year, to a group of experts in the Kiss Árpád National Institute of Public Educational Services. The survey usually involves 150 schools, selected to provide a representative sample. Skilled test administrators run the test under controlled circumstances. There are always background questionnaires for the students, and for the school's principle, too. Finally, test-researchers analyze the outcomes for the participating schools, usually in terms of: national level, school types, pupil catchment area type, gender differences and the differences in family backgrounds. National reports are then prepared and feedback given (school reports). The national report is usually published in form of a book, and short versions appear in educational reviews and on the Internet. The publishing of school reports is the decision of the principle and school staff.

This kind of survey is a good means of highlighting problems and the weaknesses for policy makers, since it produces objective, comparable and "sound" data and results concerning the

status of education. However, as it was mentioned before, there has been insufficient reaction and movement for change.

International Surveys
Hungary has had a long tradition of participating in international surveys, since the early 1970s. The country participated both in surveys run by IEA (Reading Literacy, FIMSS, TIMSS, TIMSS-R, PIRLS, SITES, and TIMSS-R in 2003) and the OECD (SIALS and PISA2000 and PISA2003). The results of these surveys, and the opportunity for international comparisons, provided very important information concerning the strengths and the weaknesses of the Hungarian educational system. But, unfortunately, we are in the same situation regarding the results and the consequences of international outcomes as we are regarding national results: there is little connection with, and impact on, practice, and only few moves towards development.

Past and future

In the Hungarian public education system five years ago, there were no regular assessments with standardized tools and professional assessors (apart from a few, one-off, low volume initiatives, including investigations at the University of Szeged, the Monitor surveys done by the Center for Evaluation Studies and various international surveys).

Although the legal regulatory framework had already acknowledged and stressed the importance and necessity of such evaluations, in practice, basic requirements for the work were missing. Not only society as a whole, but also large segments of the public education sector lacked the view (or at least it was present in a very rudimentary form) that the quality of education and its regular and objective evaluation was important. There were no experts, and the training, the institutional and infrastructural backgrounds were all missing, and, therefore, no step forward could be made in the field. Assessments by the Center for Evaluation Studies reached only a small professional circle. On other levels, and in other spheres and fields, there was generally no evaluation work. What was accomplished was only known amongst the participants.

At the same time, it was also a characteristic that follow-up on the findings of evaluations was not satisfactory. No measures were taken that, taking the results into account, would have made changes in any of the fields of public education.

Presumably, assessment and evaluation will have a major role in public education five years from now. In current educational policy, the concept of the expansion of assessment work (on more levels, in more fields and forms) is strongly emphasised. Hopefully, this intention will also manifest itself in financial and professional investment and, as it has been presented in the latest plans of the Ministry of Education, an independent Public Education Evaluation Center will be established. This would ensure that the Monitor Surveys, launched fifteen years ago, could continue, that Hungary would continue to be a permanent participant in international surveys and that the National Assessment of Basic Competences Survey would become regularly repeated permanent practice. Additional to this, a permanent assessment material development group will have to be set up and a material-bank established.

As an integral part of this development, the field of assessment and evaluation will have to become an integral part of teacher training and in-service training; the lack of it is a shortcoming of the teacher-training curriculum. The culture of school self-evaluation also needs further development.

The development of key competences and there evaluation will probably become one of the most important fields of public education, but discipline-related assessment cannot disappear from practice, either.

The current plans for the transformation of the examination system (multilevel graduation exam in secondary schools; integrating graduation exams in secondary schools with university entrance exams) will present new demands for evaluation.

Another problem that needs to be solved in the future is that, at the moment, there is no comprehensive and sophisticated system of quality assessment for the different - horizontal and vertical - areas of education. The different types of information gathered from different levels and sources are not collected and analyzed centrally. There are several research activities in the back-office institutions of the Ministry of Education, but they rarely have any connection or cooperation with each other. In addition, the results of this fragmented research do not usually

lead to follow-up initiatives or impact on practice, as there is little permeability, and no organization (a department, person or group) to handle these issues.

In the case of curricular programmes or, to use the more recent terminology, pedagogical systems, made available to schools, the problem of quality assurance has not been truly solved. Although generally accepted requirements exist for textbooks, piloting and the publication of the results of piloting to teachers are very rare. The development of material-banks in testing is an interesting initiative in connection with quality assurance. Concerning the success of evaluation, there is a growing need for reliable and quantifiable indicators on which to base an evaluation on. For example, Hungary is an active participant in the Indicators of Education Systems programme of the OECD, and in an indicator-development programme initiated by the European Union, based on the decision made by the Ministers of Education in Prague in 1998 and in Bucharest in 2000.

In order to arrive at a reliable analysis of the success of schools, there is a great need for further indices, e.g. ones which would measure the success of students in finding a job, participating in professional trainings as adults, and integrating into society. The reason is that, for a large proportion of institutions, these indices would be the primary standards for the assessment of success. From this standpoint, it is worth noting that Hungarian schools perform rather well, compared to international standards, when it comes to students' academic knowledge of disciplines. However, their score is fairly low according to indices that measure the school system's success in guiding students towards the world of labour.

The Netherlands – The School Sector

Johan van Bruggen
Staff Inspector for International Co-operation
Central Office Inspectorate of Education, the Netherlands

Values and purposes

As in all other developed countries, we ask the three basic evaluative questions at the different levels of education:

- The level of the individual pupil, student, teacher;
- The level of the individual school;
- The level of the educational system at local, regional or national level.

At the individual level

At the individual level, several questions arise: How good is the student in a particular subject? What improvements have been made over the last few weeks? Is that progress good enough for the student to pass to the next course?

Teachers play a dominant role in these evaluations, sometimes assisted by tests or examinations run by schools, or nationally (see paragraphs below about objects and method). It is clear why these evaluations are important: the decision of pass/fail, and thus about the future of students rest upon them. At this level, the responsibility for evaluation in our country lies in the hands of teachers, who evaluate their students, either alone or as a team.

The inspectorate has no task here, except to assess in an evaluation of a school, whether the function of evaluation of student progress is being fulfilled well enough (i.e. regularly, systematic, honest, complete, with methods that are as objective as those requested by the issue, and other indicators). For some evaluations of decisions about students, the laws concerning schools contain prescriptions, which are often of a general character (e.g. "the school has to have a well-developed policy for assessing the progress of students through the grades"), sometimes more specific (e.g. concerning the organisation and content of school bound examinations for certain subjects at the end of senior secondary schools; and the function of the work of the national institute for assessment and examinations CITO in that process).

It should be clear why the prescriptions for schools, and the checking and evaluating tasks of the inspectorate, are considered important: for the protection of the rights of individual students to fair, professional and comparable decision-making about key stages in their development and about their educational chances; and for guaranteeing society (employers, next schools, authorities) that diplomas, certificates, grades are valid and reliable, and have comparable value across the country and over time. So, here, a general function of the evaluative work of the inspectorate of education is seen as: guaranteeing a bottomline of quality by the inspection of provisions and actual practice in schools and reporting findings to schools, authorities and the public.

In addition, evaluation of the quality of the individual teacher (or principal) is important. Is he/she good enough to receive a new contract? Is he/she good enough to be asked to run this difficult class, or course? Also, here the decision-orientation is clear, guaranteeing adequate quality at this very important input-side.

As mentioned, decisions about these issues are in the hands of the school managers and school boards who evaluate teachers and principals. The inspectorate has to evaluate how these par-

ties fulfil their tasks: an aspect of quality is if the "personnel policy" in a school is well developed; and one of the issues, then, is if the evaluation of teachers and principals is done adequately. The inspectorate has no function in individual decisions about appointments, promotion or dismissal of staff.

So, both "aspects of quality " of a school (a good system of assessment of the progress of students; a good system of evaluation of staff) are evaluated by the inspectorate in its assessment of "the quality of a school".

The institutional level: school, faculty, learning-center, etc.

At this second level of evaluation – the school – more and more people believe that, indeed, 'school matters' (P. Mortimore and others: 'School matters: The Junior years', Open Books, Wells, 1988), and that there are important differences in quality among schools or faculties. I refer to the literature about effective schools, e.g. Reynolds, Scheerens, and many others.

Evaluation has to uncover these differences. This is the background for the rapid development of public self-evaluations carried out by schools themselves, sometimes, as in the Netherlands, Northern Ireland, Scotland and other countries, under a legal obligation. Such an internal evaluation can be done rather informally, or can be only very partial, e.g. by publishing the results of the school in national examinations and comparing these with national averages. The internal evaluation can also be rather formal, with the involvement of a "critical friend", i.e. a form of committee.

The conviction that "school matters" is also the background for the frequent public reports of external evaluators: inspectors of schools, audit-committees for faculties or universities, etc. Also this type of evaluation is done in the Netherlands. Two aspects of quality have already been mentioned that are evaluated in such an external inspection, but, of course, there are more, e.g. the learning results, the pedagogical climate in the classrooms, and other ones. See further when we return to the objects and method of evaluations, which forms the core of the work of the inspectorate. A question immediately arises regarding the evaluation of the learning results: are these assessed in an absolute way, or in a relative way with a leading question and are these learning results good enough taking into account the intake of students and their socio-economic backgrounds? For the "pedagogical climate" important evaluation questions, for example, are: is it safe, warm, challenging, etc.? We shall return to the objects and method of external inspection later, as said, but the examples of questions illustrate the important question. But, firstly, something about values and purposes.

The first 'decision-oriented function" of the evaluative information about the quality of a school is to help the parents (or elder students in senior secondary schools, further and higher education) in their decisions about school-choice and in their involvement in parent committees and other groups involved in schools.

Parents will often ask, "Is this the best school for my child?" To find the answer, they can use public (printed or Internet) reports of the inspectorate about schools, or they can use articles in regional newspapers, where journalists may use the same original reports plus their own sources. Some regional newspapers publish ranking-lists of schools. Students will sometimes make selections of universities using reports of evaluation-committees or ranking-lists published in journals e.g. Focus in Germany; Times in the USA; Le Monde in France; Elsevier or Trouw in the Netherlands.

Of course, there is also a lot of informal "evaluation", done by people at in "over-the-fence" discussions with their neighbours about the differences between two or three primary schools in the neighbourhood; or by young people of 16/17 who ask each other at the beach or a concert about the best faculties in the best universities in the best towns.

An important fact in our country – differing from most other countries – is that there is full freedom of school-choice: parents or elder students are fully free to choose between schools,

e.g. publicly or privately governed, Protestant or Islamic, etc. So, evaluative information about schools can have its place in this selection process. It is known that, in most European countries, school choice is also becoming freer, but not as radically as here.

The second 'decision-oriented function" of the evaluative information – internal self-evaluation or external by the inspectorate about the quality of an individual school, programme or course within a school – is focused on the stimulation of the autonomy of the individual school. This autonomy of a founding board, association or group of parents, to choose a mission, an identity (religious, pedagogical, didactical, organisational), textbooks, teachers, accents in the curriculum, etc., is deeply rooted in the history of Dutch educational policy.

One expression of the desired and promoted autonomy of schools in matters of pedagogy, profile, curriculum (partially free) and organisation is that schools are obliged to make their own school plan and school programme. In these documents, schools have to express their identity (religious, pedagogical, organisational, etc.) and to inform themselves and interested parties what this identity means in terms of practical consequences. The belief – like in most Western countries - is that this obligation to develop a school plan and school programme stimulates the choice – by the schools – of a sharper profile and, thereby, more variety between schools, leading to better quality. In this process, the self-evaluation is also an important – obligatory – step. Put simply, as a school, you cannot develop ideas about new development and improvement if you do not know your own reality, and do not have your own evaluation of that reality. The title of the well-known Scottish framework for self-evaluation and external inspection perfectly expresses this idea: "How good is our school?" External inspection has its place in between, i.e. between the autonomy of a school (and its expression through the responsibility of the school for self-evaluation, and the consequences that the school connects to it) and the desire of society to provide for an external, independent, professional, public evaluation of school quality. This external evaluation, on one side, functions as a guarantee for sufficient "basic quality" but also as a stimulus, a mirror, for the school to develop and to benchmark itself against such an external assessment. Of course, there are tensions here, which we will return to.

We can thus see that the second function of the evaluative information about individual schools (stimulation of autonomy, choice and variety) is connected inseparably with the third function of this same evaluation about the quality of an individual school: to inform authorities (school boards, regional authorities, the minister of education) about the quality of individual schools in order to enable them to take measures, in particular, if this quality is below expected or prescribed standards. We call this the guaranteeing function of the inspectorate.

The system level

At the system level (national, regional, county, provincial, town), the inspectorate also evaluates the quality of education.

Here questions are more general:

- How good is our system compared with other systems, e.g. average learning results for mathematics (TIMSS, PISA)?
- How effective is the use of ICT skills by students of a certain age?
- How good are our schools and our teachers and our pupils in general?
- Or – a very different issue - is the system flexible enough to allow pupils to take their own, best route through its opportunities?
- Does it provide enough innovation-stimuli?
- Is it economic?
- Does it provide value for money?

The inspectorate investigates these types of questions by aggregating the results of the many (ca. 2300 inspections of primary schools, 300 of secondary schools, 15 of the large further education institutions) school inspections per year, and by using results of educational research.

An annual report about the state of education is sent to parliament – through the minister who, by law, has no right to change it but only to give his comments – and this raises much interest and discussion. The inspectorate, and research groups occasionally cooperating with the inspectorate, also deliver thematic reports on certain issues, questions or topics, based on thematic inspections in a sample of schools.

Why is this evaluation at system level important? Because it delivers evaluative information that can contribute to the national public debate about what "we" find important in our schools and about what "we" feel should be improved or changed. So, this evaluative information is helping to set the agenda for policy-making. Of course there is no direct link between information and policy-making, as we know well enough from research about this difficult issue, but it is also known that this information has its place, depending also on its quality, timing, and the standing of the inspectorate. There are many questions and issues here, but I must leave these aside.

Here is a summary of my views about why, in our country, evaluative information at the three levels is found to be important:

a. in order to serve parents and students in their decision making regarding school choices and involvement in school policy and school improvement;
b. in order to help – through providing them with an external, professional, independent evaluation beyond the obligatory self-evaluation – school boards, principals, teachers, consulting and advisory centres to fulfil their responsibilities to better formulate their intentions (mission, aims, content, climate, etc.), and to better work towards there realisation, always within the boundaries of what has been laid down in educational laws regarding aims, content and organisation, etc.
c. in order to guarantee that responsible authorities are able to know whether individual schools and the system as a whole are good enough, i.e. meeting agreed standards; delivering quality comparable with schools in comparable circumstances; serving society's needs well enough; at internationally comparable standards; etc. It is not the inspectorate that has to take action if the evaluation shows that "it" (a school, a theme, the system) is not good enough. The inspectorate is responsible for delivering the evaluative information timely, and so serves as an early warner and as a diagnostician.

Writing about the objects in evaluative work, we will give more details about these three basic perspectives.

Objects

At the student level

The evaluation of achievement and attainment of students, as carried out by teachers, groups of teachers, examination committees, or with the help of tests or observation-instruments produced by the National Institute for Assessment and Examinations (CITO) is, of course, a very important part of daily formative and summative evaluation. It is, however, done in the Netherlands in the same way as in most other European countries. It is worth remarking that increasingly in Dutch schools a "culture of more objective assessment" is developing. Teachers use more and more the tests or instruments that are delivered by textbook-series (formative evaluation) or CITO, or by regional advisory-centres. Groups of teachers even develop their own more objective instruments of assessing the progress of students. More than 85% of primary schools use the well-known CITO-test "End of Primary School" for mathematics, language skills and information skills in order to complete their advice per student for the best track into the secondary school system. The system of combined school-bound and national examinations concluding secondary school, each providing 50% of the final grades, is not seriously in question. Of course, improvement of these aspects of evaluation in training, in-service-training, material delivery, etc. is a continuous task, as is the assurance of the quality of examinations, achieved by inspection.

Examinations with a central, standardised grade is not only a matter of simple paper and pencil tests with multiple choice items – although for reasons of objectivity, cost, speed of correction, that type is frequent in most of our national examinations. 'Free composition' by students on

the basis of centrally developed open questions, dossiers (cases) or films (videos, tapes, CD ROM, etc.) is happening and allows longer answers, drawings, essays, constructions, etc. These are judged by external experts and/or by teachers from other schools who receive these 'free compositions' for grading along with clear instructions about aspects and criteria, without knowing the names of students and schools.

Inspection of the examinations – as set nationally and by schools – is focused on the following issues:

- Do the examinations cover the vital parts of the agreed national/school objectives and national/school content matter?
- Do they have the ´right level` of difficulty?
- Do they have enough integrity, including fraud prevention (no secret questions; no mobile telephones; no 'friend' as a marker; etc.)

Each year the inspectorate examines a sample of schools (further and secondary) and a sample of subject examinations regarding the above aspects.

When we evaluate in a school, if the routine formative and summative assessment of the progress of pupils is done well enough, we look into issues such as: the use of textbook-bound-tests; the administration of, for example, "follow-me-systems"; the learning paths of pupils; pupil dossiers and portfolios. We talk with teachers and students about this work, whether it is fair, comprehensive and well linked with diagnostic and improving actions?

At the school level

Before I can go into the real details of "what is evaluated" by schools that self-evaluate, and inspectors who carry out an external evaluation, it is important to describe the recent legal and political background to this work. This background may be found in the recent Supervision Act.

On 5 June 2001, the Dutch government proposed the 'Wet op het Onderwijstoezicht', Supervision Act (Dutch acronym WOT) by sending it to Parliament. This was the conclusion of a period of three years of "white papers" and rather intensive public and parliamentary debates. In these debates, various aspects of the tensions, mentioned earlier, between the desire to promote more autonomy, profile and variety among schools on one side, and to guarantee minimum quality and to promote outstanding quality though external inspection on the other side were discussed in detail, and brought to workable solutions and "checks and balances" to the systems. In sessions in October and December 2001 and June 2002, the two chambers of parliament proposed amendments, and discussed these with the government. The Law came into force in September 2002, and the new inspection-process, based on this first law on inspection in the 200 years' history of the Dutch Inspectorate of Education, commenced in January 2003.

This Law provides a legal basis for independent, professional and public evaluations of the quality of schools and the system by the Inspectorate of Education. This type of whole – school – evaluation has been developed in the Netherlands beginning with experiments with 180 schools in primary education in 1991-1993; there is much similarity between this type of inspection work and that found in many other countries in Europe, and in Hong Kong, New Zealand etc. Since 1995, the Standing International Conference of Central and General Inspectorates of Education SICI has been the association of, now, 20 European inspectorates of education. Most of these bring into practice this type of external inspection. There are various workshops, reports and other publications of SICI that provide lists of quality indicators, descriptions of frameworks for inspection, guidance on dealing with schools with quality problems, etc. – for more information, see www.sici.org.uk

With the Supervision Act (WOT), the Dutch government intends to 'strengthen the position of the Inspectorate of Education as a provider of independent judgement'. The Supervision Act is important because schools are, on one hand, given more and more autonomy and responsibility for their own quality, with fewer rules and regulations on the input side (curriculum, financial regulations, etc.), but, on the other hand, parents and other stakeholders, including the government, want to hold schools accountable for their quality and want 'somebody' to moni-

tor the schools, take action in the event of problems relating to quality and to stimulate schools to improve continuously, even if their quality is already good, or outstanding.

The Supervision Act states that the 'supervision' is 'assigned' to the Inspectorate. 'Assigned' means that this task is attributed to the Inspectorate. The Senior Chief Inspector performs the task in his/her own name, but the minister of education remains fully responsible for the functioning of the inspectorate. Supervision (of education) means judging if and how far the quality of education (i.e. schools and the education system as a whole) is in accordance with regulations provided in the laws on education. Of course, not all important aspects of quality can be covered by sharp regulations, with which the compliance can be checked. That is the reason that the Law also charges the inspectorate to assess the quality of a number of aspects of the daily reality of each school at regular intervals, including: aspects of performance like teaching and learning; the ethos or pedagogical climate; communication with the outside world of parents, other schools, governing boards, social workers, etc.; the results and outcomes of a school; care for students with learning difficulties; the methods and instruments of evaluating the progress of students; etc. The Law contains a list of these aspects and it is the duty of the inspectorate to provide a public assessment of the quality of all these aspects of a school.

The Supervision Act also states that the Inspectorate, as well as evaluating the school's quality and checking school's compliance with regulations, should try to improve the quality of education by consulting with the management and staff of the school and the local and regional authorities. The Inspectorate's final task is to report on the quality of all schools publicly, and also more generally on the development of the education system. The latter is done in the Annual Report about the State of Education that is published in April and which attracts much attention in the media and in parliament. Like in other countries, including England, Scotland, Flanders, in a different way France, Portugal, the Czech Republic, Slovakia, this report aggregates and analyses the evaluation of individual schools and combines this with thematic evaluations and research into the evaluation of the system as a whole. An important requirement is the existence of a fixed list of quality aspects, broken down into a more detailed list of quality indicators (in total, in our system of inspection has ca. 100 indicators). This list and the corresponding set of definitions must exist for a couple of years in order to enable us to deliver insight into changes in quality over the years. Another important issue is that the aggregated scores deliver reference numbers to individual schools. To clarify this: if over the years we report that 78% of schools, growing in four years from 75 to 79%, score "more strong than weak", or better on a specific indicator such as "enough challenge in the pedagogical climate", in our ca. 2500 primary school evaluations per year, then a school that is evaluated "more weak than strong" against this indicator knows that it belongs to a small minority.

Frameworks for inspection

The way the Inspectorate performs its task of quality inspections has to be laid down in a framework for inspection, the law states. For this framework to be effective, it has to have the full commitment of those who are concerned with the work of the Inspectorate. For this reason, the Inspectorate consults with representatives of the educational field and other stakeholders and takes their opinions very seriously. The Inspectorate remains, however, responsible for the decisions about its own framework for inspection. Parliament has created the procedure that the Senior Chief Inspector has to decide upon the framework and forward it to the minister for approval, who in turn sends it to parliament. This enables, where appropriate, a debate on important matters in parliament. The first frameworks have been dealt with according to this procedure in November and December 2002 and are now valid for three years. They contain details about working methods of the school inspections (see below) and they provide the quality indicators and criteria that are outlined by the Supervision Act (see below).

Proportionality between self-evaluation and external inspection

The principle of 'proportionality ' is another core aspect of the Supervision Act. The better the quality of a school evaluation as carried out by, or on account of, the school itself, the more likely it will be that the Inspectorate will adopt this evaluation as if it were its own external evaluation. So, as it has been boldly stated by members of parliament, the minister and the

senior chief inspector, if a self-evaluation of a school is done well and shows good quality, the school will not be bothered too much by an external inspection. The question of course arises: when is this the case?

The law mentions three indicators for "a good self-evaluation":

a. it has to be complete, i.e. it has to cover all indicators of quality, or most of the important ones, that the inspectorate has elaborated in its framework;
b. it has to be reliable, i.e. with real facts, no window dressing, no "desk-evaluation", but real opinions of: parents, students, next schools where students are taken in, real analysis of learning results, real analysis of the quality of the classroom-teaching by principals or other persons going into classrooms - and not only paper-and-pencil-ratings…etc.
c. it has to show enough ambition in its evaluation of the facts, i.e. are we good enough in this or that aspect, also taking into account our specific situation in terms of student-intake, etc.?

In the frameworks for inspection, the inspectorate has elaborated these three indicators and we have now, following experiments in 2001-2002, begun to practice this "evaluation of the self-evaluation". Of course, we also took part in the project "Effective School Self Evaluation" (ESSE) of the Standing International Conference of Central and General Inspectorates of Education, SICI. On its website, www.sici.org.uk, information about this interesting approach in meta-evaluation may be found; the reports of ESSE have been placed there since April 2003.

The principle of proportionality means that the Inspectorate has to start its work by carrying out only a short investigation in order to check the reliability of the self-evaluation; usually done by a few lesson-visits plus some meetings with staff, pupils and parents, and a check of papers and internal reports.

If the inspectorate does not have any further suspicions concerning failing quality – based upon rumours in newspapers, or complaints of parents, or indications such as an extraordinary high number of teachers' transfers, or a quick and sudden change in examination results, or other early warning indicators – this short check of the reliability of the self-evaluation is enough, and the inspectorate can then adopt the report of the school as its own report. This has not happened until now, because (also in the Netherlands, like in other countries) most schools are only in the beginning of the development of mature and comprehensive systems of self-evaluation that meet the demands of completeness, validity, reliability and ambition. However, much is being done to help schools, through national projects, regional advisory groups and many publications with guides to good practice, and in the last few years we have seen clear progress in this development.

Proportionality also means that the Inspectorate will vary the intensity with which schools are inspected, depending on a quick assessment of their quality. This enables more time to be devoted to the schools that need it. Quick assessments are made through a system of short yearly visits to all schools (one inspector for only half a day for an average primary school of some 200 pupils; two inspectors for one day in a large and complicated secondary school of 2500 pupils) in order to monitor their development (school plan, self evaluation, follow-up after the last periodic inspection from the time before the new law) and to monitor signs of sudden change in quality. Also, desk-research can produce a prompt to inspect a school between the regular planned periodic inspections.

Again: what are the objects?

In order to assess the quality of education of a school, the Inspectorate has to assess the features of quality laid out in the Supervision Act. These features are also known as "key areas of quality" in other inspectorates. Some of these features are also touched by regulations for schools that are to be found in school laws (see what has been written about the Supervision Act), and, of course, the inspectorate has to incorporate these into the lists of indicators. So, it is the responsibility of the inspectorate, in its elaboration of the features of quality into a set of indicators and descriptors of good practice, to combine general interpretations with those parts

of this "definition of quality" that are regulated by specific laws for certain types of school. In other words, the Inspectorate is obliged to use the regulations laid down in the school laws as elements in its evaluation of the quality of the schools and the system, but may also use and take other indicators into account.

The features of quality that the Supervision Act mentions cover most of the internationally known aspects that are relevant to school quality, cf. "How good is our school?" by the Scottish Inspectorate, the frameworks of OFSTED or the frameworks of the Portuguese and Flemish inspectorates, and other colleagues. The Supervision Act mentions the following nine aspects, based on which the inspectorate is obliged to give a public evaluation of the quality of all schools in the periodic, general inspections:

- the learning results and outcomes of pupils;
- the progress in the development of pupils over the years;
- the offer of content (curriculum);
- the arrangement and use of learning time;
- the pedagogical climate in classrooms;
- the general climate in the school;
- the teaching of the teachers;
- the provisions for pupils who need special care;
- the content, level and carrying out of tests, tasks, examinations, assignments and all other forms of assessment of pupils.

Apart from these nine aspects, we, of course, inspect and evaluate the process and products of the self-evaluation of the school, as this, in accordance with the philosophy behind "proportionality", is the first "entrance" for the inspectorate into the school. So, in fact there are ten aspects of quality to be evaluated.

Of course, much could be said about the wording of the features and about the theoretical concepts behind them. I leave that aside, only mentioning that for insiders it will be clear that there is a link with the results of the research into effective schools.

For all aspects of quality, a number of indicators exist in our framework for inspection and, belonging to each indicator, a number of "practice descriptors" has been formulated and agreed. Inspectors have to observe whether these descriptors can be identified in the educational reality of a school in order to conclude that a certain indicator is indeed present, e.g. "the teaching and learning has, where appropriate, an interactive character". This is done with the help of observation and scoring tables. In order to do that correctly, of course training and dual inspections (in pairs) are important in order to guarantee objectivity and comparability. Decision rules guide a team of inspectors through the assessment of a certain aspect of quality, for which a number of indicators may or may not be present, leading to sub-conclusions for a particular aspect, such as "more strong than weak" or "strong', etc. It is not possible here to write about these methods of work in more detail.

It will strike some readers that, in contrast to lists of aspects of quality that are in use with other inspectorates, the important aspect of "management and organisation' is not mentioned. The background to this is that the Supervision Act places emphasis on the "primary process and its results" in a school and, at first glance, disregards important aspects of quality that have a more "conditional" character, such as, "quality assurance" (see above about proportionality in the inspectorate's work), "management and organisation" or "use of staff and money". Of course, the Dutch inspectorate also evaluates these aspects, but only in the so-called "deeper inspection", and only if the evaluation of the primary process leads to suppositions that there are quality problems.

Stakeholders

The answer to the question "for whom do we evaluate" is directly linked to the answer to the question "why we evaluate".

Regarding the evaluation of individual schools, our first clients are the schools themselves, who we try to stimulate in their relative autonomy by: (a) obliging them (in their best interests, as also broadly believed by leaders of teacher unions and other spokespeople) to invest in self-evaluation as an engine for further development; and, (b) providing them with an external, independent assessment of their quality, i.e. a "mirror with authority", in order to stimulate them to improve.

But the second clients are the public in general, and potential users, students and parents in particular. In relation to the traditional free school choice and the desired parent-empowerment, it is important to provide parents, students or journalists with independent, quality information for further dissemination to the general public.

And the third clients are the various forms of education authority, e.g. the school board that may rely on the inspection-reports, or the minister who has to form opinions about proposals of the inspectorate in order to impose sanctions or to offer measures. We are beginning to see local or regional authorities as clients who can use regional analyses of specific quality aspects of educational provisions in formulating their facilitating policy.

At the level of the national system (annual report, thematic reports), it is clear that the general public, as represented by journalists, politicians, non-governmental organisations are also a form of client, or audience.

Methods

How a set of indicators belonging to an aspect of quality is chosen, and how we work with practice descriptors, etc., has been mentioned briefly above. Here, I will merely write something about the flexibility in our system of inspections that is a consequence of the proportionality that we aim to realize.

The baseline is that, from January 2003, all schools will be fully inspected once every three years. Prior to that date, we also did full inspections but somewhat differently, and not on the new legal basis. It is worth noting that all schools have been fully inspected on an experimental basis at least once since 1997/1998. The thoroughness of the general periodic inspection (covering all aspects of quality) will continue to become increasingly proportional to the quality of the self-evaluation of the school and the facts about quality that it reveals. This general periodic inspection will, hopefully, in a growing number of schools, return to a "reliability check", as described.

The baseline also requires that, from January 2003, all schools are inspected annually, in a quick-scan that focuses on the development of the school, its progress with policy development and self-evaluation, and on "early warning" signs. Also, if time allows, there will be focus on a certain topic that the school itself highlights with a view to obtaining an external evaluation of it as support for their own development. Such a yearly quick-scan can lead to the conclusion that an earlier-than-planned full inspection is necessary, as there seem to be quality problems. Thematic inspections in a representative sample of schools can be coupled to the yearly visits (quick-scans) or the periodic (full) inspections, but can also be scheduled separately.

All these inspections are concluded with a provisional oral report in a closing session with the leadership, representatives of the evaluation team and the parents and governors. This is followed up within a few weeks by a draft written report. The board of the school (and the principal) can suggest amendments or react to the facts, conclusions and recommendations presented. If the parties do not agree, the Supervision Act allows the school to have its comments taken into the report in a summarized form and/or to place their comments on the website of the inspectorate, linked to the school report.

If a general inspection – including one that is initiated after a yearly visit – leads to a provisional conclusion that there seem to be quality problems, the inspectorate will carry out a so-called deeper inspection within two or three months, and there will be no public report at that moment. A new team of inspectors will then examine the school more broadly and intensively, looking particularly into all aspects of quality. The investigation will also look for causes of, or backgrounds to the quality problems, which in many cases will lie in the management and governing domain, and also seek interconnections between different aspects of quality.

After the deeper inspection, there is a public report, irrespective of its conclusions. If, indeed, the suppositions of serious problems are confirmed, the school receives – in the closing meeting and in the report – firm recommendations to do certain things that, hopefully, will bring quick improvements, and a notice that, within six or twelve months, the inspectorate will return in order to carry out a so-called "inspection of quality improvement". It is possible that, between the deeper inspection and the inspection of quality improvement, the inspectorate will monitor the school more frequently; not in order to give help, advice or in any way to take away the responsibility from the board and the principal, but in order to keep an eye on the work in progress and to stimulate the improvement by highlighting good practice from other schools, offering links to other schools, or links to the regional or national institutes for help, guidance, in service training or curriculum development.

If the inspection of quality improvement shows that serious problems remain with regard to the quality of education, the Inspectorate, under the new law, has two theoretical possibilities. The first one is for serious problems, clearly caused by the fact that the school does not obey certain important regulations in the school law, e.g. the school does not offer a curriculum that responds adequately to the legal prescriptions in terms of general aims and of attainment targets. In such a case, the inspectorate will have warned the school already after the yearly visit and again after the regular inspection (that, in such a case, will have followed the yearly visit within two months or less) and again after the deeper inspection (that follows the regular inspection, in such a case, within two or three months). If the inspection of quality improvement shows that the school still fails, the inspectorate can advise the minister to withdraw part of the lump sum grant – as a punishment – or to withdraw the permission to take part in national examinations or to give diplomas. In practice, this will mean closure of the school. This "sanction-regime" has always been in place but is now connected with the new Supervision Act. We have yet to see when and which cases will be brought to the minister by the inspectorate in the coming period. The intention is clear: the inspection must have "teeth".

If the serious quality problems that are signalled in the report of the inspection of quality improvement do not have a background in clear disobedience to regulations, but are ambiguous, or have to do with "bad teaching in a majority or large proportion of lessons" (no law for schools says that teachers have to provide good teaching!) or another important aspect of quality, the Supervision Act has introduced the new innovative concept of "the provision". If there are serious or long-term problems of the type just mentioned, the Inspectorate can advise the minister to offer the school some kind of 'provision'. This provision may be financial (e.g. 2 million Euros extra for the coming two years, under the condition that the board gets rid of a bad principal), but may also take the form of extra staff or advice, e.g. a paid contract with an in-service institute under the condition that all staff participate in courses, following an assessment of their working capacities done by a professional institute. Due to the fact that the school has not broken any regulations, provisions cannot be imposed. An intended provision always needs the consent of the school board. The "last thing" that the minister is able to do is to publish the fact that a school refuses an offered "provision" in the hope that this will lead parents and students "to vote with their feet".

These concepts of "sanction" and "provision" are unique for the new Supervision Act in the Netherlands. Of course, at this moment the "provision' is merely a theoretical concept. It will be very important, in the coming years, how the inspectorate applies and develops this concept, which is an important tool.

Some remarks about practical aspects of the inspections

Apart from the remarks above about the flexible system of inspections, and about the use of the set of indicators and practice descriptors, etc., a small number of remaining issues may be of interest:

a. Almost all inspections are notified at least six weeks in advance.
b. Reports of yearly visits take the form of a letter to the school board and are not placed on the Internet, but are passively public.
c. General inspections always take place with a team of at least two inspectors; yearly visits are done alone.
d. Of course there is much investment in training of inspectors and other staff, for the sake of comparability of inspections and, in particular, the use of indicators and criteria. Also external, university-based research is used to assure the important aspect of quality of the inspectorate. The inspectorate is obliged to publish an annual report about its own quality (of inspections) assurance.
e. There is, by law, an independent, external commission that deals with complaints about behaviour of inspectors or about evaluations that are considered unfair.
f. The inspectorate is involved in cooperation with all kinds of institutions and projects in order to stimulate self-evaluation by schools and the use of good sets of indicators and good, comprehensive and reliable procedures. Even though there is no legal requirement or pressure to use the inspectorate's framework of quality indicators from the government or us, the tendency is that more and more schools choose to use that framework, as in Scotland and England.
g. There is, by law, an independent Advisory Commission that advises the Senior Chief Inspector about, in particular, the continued development of the framework for inspections (the set of indicators, criteria and the flexible system of inspections).
h. The inspectorate has a budget of some € 35 million – compared with € 25 billion of state funding for education in general. We employ approximately 260 inspectors plus some 250 other staff, and work in seven regional offices. The so-called support structure for the 8,300 schools and institutions with a total of around 350,000 teachers (university-teachers included) has – for guidance, in service training, curriculum development, testing services, etc. – some 3,000 staff.
i. Of course, one of the most important issues is the breakdown of the quality aspects into indicators and criteria. I have given this some remarks and examples, but a more systematic description of how this works would require too much space. The same applies for the issue of evaluation of the added value of schools

Past and future

Five years ago, we already did this same type of whole school evaluation in primary (from 1992 onwards) and secondary (from 1997 onwards) schools, but were still in more of a developing stage. The principle of proportionality had not been formulated then. We had already done thematic inspections since around 1985 and had developed good methodologies for that type of work; many of the reports have had considerable influence on policy making, teacher training, in service training and textbook development. This work is still continuing, but in the last five years the focus moved more towards whole-school-evaluation instead of thematic evaluations in samples of schools.

Probably, at least we hope, more schools will deliver better self-evaluations in the future. The consequence being that we will able to focus more on schools that lag behind in their quality development, trying to stimulate them by sharper and more focussed inspections. Also, we hope to focus more on inspections in the "better schools" centred on their developmental needs. So, still more flexibility in inspections.

Five years from now we will probably focus more on topics and issues that are seen as relevant for the further development of the system as a whole. For instance, we have just started a project "Synergy in inspections" in which we are trying to develop intensive cooperation with inspectors in some towns regarding health care, youth care, housing authorities and with local

institutions (employment service, churches, libraries, etc.) in order to "close the inspection chain", from perspectives of social inclusion and/or integral care for youngsters.

We anticipate that requests will come for more inspection of individual disciplines in schools; we already do that rather superficially, but probably, in addition to schools, parents and students will demand more specific information about how good a certain school is. This rating will not only be in general terms and quality aspects, but also, for example, specifically relate to science, civics or the music teaching.

We also expect a more developed European perspective. In the sense that "international cooperation" will be a more important aspect of school-quality than it is now, in our present framework. But also in the sense that there will be more cooperation among European inspectorates in the inspection of schools by mixed teams, and/or by sharing benchmarks; more than we already practice with our neighbouring inspectorates in Flanders, England, Lower Saxony and North Rhine Westphalia; and also multilaterally, with our colleagues within SICI.

Now, there is no sign that a fundamental change in our tasks and position is to be expected in coming five years, but, of course, you never know. The Supervision Act will be evaluated in 2007 by some form of external group. As a natural part of our work, we permanently gather evaluative information about what difference the inspectorate makes for the self-confidence of schools, their development and their quality in terms of performance and the achievements of students. So there will probably be gradual changes in our work.

Canada – Higher Education

Jacques L'Écuyer
Chairman
Dany Rondeau
Project coordinator
Commission d'évaluation de l'enseignement collégial du Québec (CEEC)

For the past ten to fifteen years in Canada, evaluation of universities' and colleges' educational programs has been at the center of continuous debates. In various sectors of society, there have been calls for better accountability, more information on the quality and effectiveness of the institutions, performance measurements, quality assurance and monitoring bodies. At the source of these debates, one finds in particular the increasing cost of higher education, its competitiveness in the context of globalization, changes in the needs and behaviour of students and the necessity to increase accessibility without lowering quality.

This context has lead to major changes in the higher education policies and regulation mechanisms in many parts of Canada. Governments have reinforced the need for clear and measurable objectives and effective performance and tracking reporting systems. Alberta and British Columbia legislation on higher education is currently being thoroughly revised. Ontario has recently developed key performance indicators and established a Post-secondary Education Quality Assessment Board. Quebec's universities have had to sign performance contracts with the Ministry of Education.

The present paper attempts to present the situation of the evaluation of higher education in Canada as it is in 2003. It describes the quality assurance mechanisms in the various provinces and analyses their main characteristics. Recent changes have been taken into account.

Overview of Higher Education in Canada

Canada is a federation of ten provinces and three territories. It has two official languages, English and French. Education falls under the jurisdiction of the provinces. Each province and territory has accordingly established its own system of education with its particular structure and regulation mechanisms. Differences between each system can be quite important. In Quebec for instance, the primary and secondary schooling lasts 11 years as compared with 12 in other provinces. But on leaving secondary school, the Quebec student must first receive a two-year college education before entering the university. In other provinces, this obligation does not exist.

The basic structures of the education systems are however quite similar, and the first post-secondary degree, the baccalaureate, is normally obtained after 16 years of schooling. Degrees are usually granted by universities or university-colleges. Colleges of all types, institutes and grant diplomas, generally after two or three years of post-secondary schooling, and certificates or attestations in the case of shorter programs.

Each province has its own quality assurance system and has developed a series of measures to ensure the quality of its post-secondary education system. The legislation has been used extensively to protect the credibility of the degrees granted by institutions and to control the development of the system in a context where the number of would be providers is increasing. In most provinces, the power to grant degree is restricted to institutions formally recognized according to the legislation and regulations in force. Even more importantly, in nearly all provinces, new program proposals and important changes in programs must be approved by the

Ministry of Education or by a mandated intermediate regulatory body. In fact, presently in Canada, nearly all new degree programs are submitted to some kind of review before being offered.

New Program Proposals:
The procedure used to review new program proposals varies according to the province and the type of institution. In Alberta, all new program proposals coming from public institutions, universities or colleges are reviewed by the department of education. The department focuses on how the program reflects the mandate of the institution and balances with other programs in the region. In the case of private institutions, degree program proposals are reviewed in depth and, if found adequate, are accredited for a period of time, normally 6 or 7 years. This accreditation is granted by The Private Colleges Accreditation Board (PCAB), an agency set up by the Government. Ontario and British Columbia have procedures which are quite similar. The department of education reviews all new programs for duplication. When the proposal of a degree program originates from an institution which is not already authorized to offer degree programs, it is reviewed by an advisory body, the Degree Quality Assessment Board (DQAB) in British Columbia and the Post-secondary Education Quality Assessment Board (PEQAB) in Ontario. These agencies have been established by the Government.

In Manitoba, all new university and college programs must be approved by the Council on Post-secondary Education (COPSE). Programs are reviewed from quality, organizational and financial perspectives. In the Maritime Provinces (New Brunswick, Nova Scotia and Prince Edward Island), new university programs are reviewed by the Maritime Provinces Higher Education Commission (MPHEC) to ascertain their suitability given their objectives, structure, resources and relevance. New college programs must be approved by the education department in New Brunswick and Nova Scotia. Both COPSE and MPHEC are government agencies.

In Quebec, new university programs are reviewed from a quality perspective by the « Commission d'évaluation des projets de programmes (CEP) » under the aegis of the « Conférence des recteurs et des principaux des universités du Québec (CREPUQ) » and their relevance is assessed for financing by the Ministry of Education. New college diploma programs are approved by the Ministry of Education.

There are no external reviews of new programs in Saskatchewan, Newfoundland and the Territories.

Provincial Continuous Quality Assurance Practices:
Apart from new program assessments, some provinces have continuous quality assurance mechanisms. Such mechanisms exist in Alberta for the private institutions, as the Private College Accreditation Board must periodically review the operations of the private colleges with accredited programs. A similar provision exists for the British Columbia and Ontario quality assessment boards, DQAB and PEQAB, but it has not yet been tested, as these boards are still very young.

At the university level, Ontario, Quebec and the Maritime Provinces have in operation mechanisms to audit the internal quality processes used by universities for the assessment of their existing undergraduate programs. In Ontario, this is done on a voluntary basis by the Undergraduate Program Review Audit Committee (UPRAC), under the aegis of the Ontario Council of Academic Vice-presidents. In Quebec, each university is required to have a periodic program assessment policy. These periodic assessment processes are reviewed by the « Commission de verification de l'évaluation des programmes (CVEP) » under the aegis of the Conference of Rectors and Principals (CREPUQ). Finally, in the Maritime Provinces, the Maritime Provinces Higher Education Commission (MPHEC) has the mandate to monitor the universities' quality assurance mechanisms.

Ontario is the only Canadian province that has an external quality assessment agency at graduate level. In this province, graduate programs offered or proposed by public universities are appraised on a voluntary basis by the Ontario Council on Graduate Studies (OCGS), under the aegis of the Council of Ontario Universities.

At the college level, Quebec is the only province that has an extensive quality assurance system. In this province, the Commission d'évaluation de l'enseignement collegial (CEEC), a government agency, has the mandate to assess programs as well as public and private institutions.

Professional Accreditation:
At the provincial and national levels, some professional programs are accredited by professional accreditation agencies. In some fields, for instance health sciences, engineering, etc., it is required that applicants for professional registration have graduated from an accredited program. These accreditations are valid for a certain period. Hence the accredited programs are periodically reviewed according to standards and procedures that are frequently modelled on those used by similar organizations in other countries. The Association of Accrediting Agencies of Canada (AAAC) counts 21 such members.

Other Mechanisms:
Other mechanisms are in place to ensure quality in Canadian post-secondary institutions. For instance, the Association of Universities and Colleges of Canada (AUCC) and the Association of Canadian Community Colleges (ACCC) do not accredit or evaluate programs or institutions. However, they require of their members that they deliver high quality programs consistent with the academic standards set out by appropriate jurisdictional authorities. As such they do give some assurance of quality.

Finally, it must be added that a number of ministries of education have established a series of performance indicators. These indicators are published and, in some cases, used for adjusting the level of public funding. Ontario is a leader in this trend. It has launched a Key Performance Indicators Project to account for the performance of the higher education institutions of this province. Employment rates, graduate satisfaction, student satisfaction, employer satisfaction and graduation rates are published on a yearly basis.

The previous overview shows how diverse are the regulatory systems in Canada. Except for a few cases, it is, therefore, not possible to speak of a Canadian higher education quality assurance system. Nevertheless, a few trends or characteristics emerge from this diversity when we try to answer the questions why (values and purposes), what (objects), for whom (stakeholders) and how (method) evaluation is done.

Values and purposes

In Canada, the reasons for doing evaluation vary according to the agencies. Some do it for control reasons, some for quality assurance, some for quality improvement. The distinctions are not always clear between these reasons and one has to look at the objectives and practices of the agencies to determine into which category they fall. For the purpose of this article, we will classify under "control" those evaluations that are done to verify if a program should be offered or if an institution should operate. Quality assurance, on the other hand, means an evaluation that is done to determine if a program or an institution meets predetermined standards. Finally, quality improvement is an evaluation done for the explicit purpose of improving the quality of a program or an institution. This last type of evaluation usually has the following characteristics: it implies that the agency resorts to peers, that it makes recommendations to the institutions themselves and that the evaluation has some kind of follow-up either directly or through periodic evaluations.

Control
In Canada, evaluation is quite often done with the purpose of controlling the development of the higher education system and assuring its quality. As was pointed out, in the major provinces, new degree programs are reviewed in one way or another by government-controlled agencies with a view to:

- avoiding duplication;
- making sure that they are relevant given the needs of the society;
- making sure that they reflect the mandate of the provider;
- making sure that they meet generally agreed standards.

The first three reasons are given in nearly all provinces where evaluation of new programs is done. They are obviously reasons of control coming from the fact that in Canada, the governments heavily subsidize higher education and want to avoid wasting money. They apply mainly, but not exclusively, to public institutions, as even private institutions get government funding in some provinces.

The fourth reason is of a different nature and, in many provinces, concerns only colleges or private institutions that wish to offer a degree granting program but are not recognized as a degree granting institution by a charter or by the provincial legislation. In such cases, permission to offer the program will not be given unless it has been evaluated and shown to be of good quality. Consumer protection and the defense of the credibility of the national and provincial degrees (Baccalaureate, Master, Doctorate) are the main reasons for requiring this evaluation which is, at the same time, a quality assurance and control measure.

Improvement
Evaluation is carried out for the improvement of quality in five of the ten provinces: Ontario, Quebec and the three Maritimes Provinces. In these provinces, an agency has the mandate to periodically audit the internal quality assurance process of each public university. In Ontario, this is done by the Undergraduate Program Review Audit Committee (UPRAC), in Quebec by the « Commission de vérification de l'évaluation des programmes (CVEP) » and in the Maritimes, by the Maritimes Provinces Higher Education Committee (MPHEC). In all three cases, the audit is periodic, formative and the recommendations of the agency are directed to the institutions themselves with a view to providing assistance and advice.

Program evaluation by the Ontario Council on Graduate Study (OCGS) is also done in the perspective of quality improvement. It is periodic, and recommendations are made to the institutions themselves with a view to promoting best practices in graduate education.

The work of the « Commission d'évaluation de l'enseignement collégial (CEEC) » of Quebec will be reviewed in the case chapter. As will be seen, it is done with the explicit purpose of improving the quality of Quebec college education.

Quality Assurance
The evaluations realized by the professional Boards for the purpose of program accreditation usually contain a dimension of improvement superimposed on specific quality assurance objectives. These evaluations are indeed done to verify that programs meet the criteria of the agency and, as such, fall into the category of quality assurance mechanisms. However, these evaluations are done on a periodic basis, and the criteria often contain particular exigencies related to continuous quality improvement. Furthermore, the criteria are themselves periodically reviewed to reflect the state of the art, and, finally, recommendations are addressed to the institutions. All that contributes to the continuous improvement of the accredited programs.

It is interesting to realize that whenever the evaluation is done by a government owned or controlled agency, it is usually done as part of a government policy, and, for this reason, institutions have to submit themselves to the process. In the other cases, Ontario's UPRAC and OCGS and Quebec's CVEP, owned by their respective Conference of Rectors, the process is voluntary, but all members of these Conferences have accepted to comply with it. As for the evaluations done by the professional Boards, they are usually a near obligation, since, otherwise, the graduates from these programs would have difficulties in registering in their profession.

Objects

Generally speaking, Canadian evaluation agencies evaluate programs, not institutions. This is because the authority to grant academic credentials is given by governments through charters or legislation and reserved for institutions with adequate mechanisms for assuring institutional and program quality. These institutions, mostly public, are subsidized and supervised by their respective governments and, therefore, are not subjected to institutional accreditation. There are, however, a few exceptions to this situation. Alberta, British Columbia, Ontario and New Brunswick have opened the field of higher education to private institutions that meet the standards of quality for degree granting institutions. These private institutions are submitted to a periodic institutional evaluation to make sure that these standards are maintained.

As previously mentioned, public universities in Ontario, Quebec and the Maritimes Provinces are audited by an external agency. This audit concerns their quality assurance mechanisms. It is formative and is realized to provide assistance to the universities and, at least in the Quebec case, to testify to the efficiency of their internal quality assurance practices.

At the college level, institutions and programs are usually « registered » (or « licensed ») by governments. This registration is done for consumer protection only. Institution or program quality is not evaluated. There is, however, a major exception to this situation, namely, Quebec where institutional as well as program quality is extensively evaluated by the « Commission d'évaluation de l'enseignement collégial (CEEC) ». This evaluation is done primarily for quality improvement, but the Commission can recommend that a program or an institution be closed if it is of a poor quality.

In all other cases, agencies evaluate programs, not institutions. Agencies like PCAB in Alberta, DQAB in British Columbia, COPSE in Manitoba, PEQAB in Ontario, CEP in Quebec and MPHEC in the Maritimes Provinces evaluate projects of new programs to assess their quality and suitability. Such aspects as the relevance of their objectives and content, the coherence of the curriculum and the adequateness of the resources are assessed.

Professional programs are evaluated periodically by the professional accrediting agencies. This evaluation is realized to ensure that these programs comply with the standards of the agency. For instance, the Canadian Engineering Accreditation Board declares in its documentation that accredited programs « will meet or exceed minimum educational standards acceptable for professional engineering registration in Canada ». These standards cover a wide range of questions: governance, program objectives and content, pedagogical methods, student assessment, student support, student success, resources, etc.

The Ontario Council on Graduate Studies (OCGS) periodically evaluates graduate programs. This evaluation is for improvement of the graduate education. Criteria depend on the type of programs – master, doctorate, professional – and cover aspects such as objectives, content, breadth, research component, analytical skills development, etc.

Stakeholders

Accrediting agencies and evaluation bodies consider that the evaluation may benefit many stakeholders. There are the universities, which receive support during the elaboration of a new program; the colleges, which gain experience in evaluation and improve the quality of their programs; members of accrediting agencies; the students, who benefit from a high quality education; and the public, the consumers, the clients and the employers, who can be assured of the competences of graduates and professionals.

However in any evaluation, there are always stakeholders who benefit more directly than others from the evaluation process. In the case of the evaluation of proposals of new programs, governments appear to be the major beneficiary since the result of the evaluation is used to determine if a program should be offered and funded. It is clear, however, that the consumers, who are protected from receiving an education of poor quality, also benefit from this evaluation. The institutions themselves greatly gain from the comments of the review team.

When the evaluation is done for the purpose of a professional accreditation, the professions themselves are the main beneficiaries. The professional agencies evaluate the programs according to their own criteria, which are certainly in accordance with the needs or the objectives of the profession, but do not necessarily take into account the general context and the resources of the institution. Obviously the society benefits from better-trained professionals, but in the absence of public discussions there is always the danger that the accreditation criteria reflect more the interest of the profession than the true needs of the society. Nevertheless, professional accreditation can be, and usually is, a source of important benefits for the society and the institutions, as well as for the profession.

When quality improvement is the main objective of the agency, as is the case for the Maritimes Provinces Higher Education Commission (MPHEC), the Quebec « Commission d'évaluation de l'enseignement collégial (CEEC) » and the agencies operating under the aegis of the Conference of rectors in Quebec and Ontario, it is clear that the direct beneficiaries are the institutions themselves.

Methods

Evaluation in Canada is closely linked with its purpose. It is, therefore, necessary to distinguish between the various contexts in which it is realized: the evaluation of new programs in public universities; the evaluation of new programs in private institutions; the periodic evaluation of programs as done by professional Boards; and the audits of universities' quality assurance processes.

New programs in public universities

Generally speaking, these evaluations are done mainly to avoid unnecessary or unjustified duplication. The review is not extensive and usually does not involve experts and site visits. The quality assessment of the proposal is done by the universities themselves through their quality assurance mechanism. There are, however, two exceptions: Quebec, for all degree level programs, and Ontario, for graduate level programs. In both cases, proposals for new programs are assessed externally by commissions owned by the universities – CEP in Quebec and OCGS in Ontario. This evaluation is done by experts of the discipline and usually implies a site visit. The framework, content, faculty and resources available are subjected to this evaluation. Relevance, taking into account the needs of the society, the mission of the institution and the presence of similar programs in other institutions, is, of course, a major concern in all proposals for new programs.

New programs in private institutions and other non recognized institutions
In some provinces, private universities, out of province institutions and public colleges can request the « consent » of the Minister of Education to offer degree programs. In such cases, the institution must first undergo an institutional review of the type described in the next paragraph. If the result of this review has been positive, the proposed programs are evaluated extensively by disciplinary experts, with site visits according to criteria similar to those of CEP and OCGS mentioned above, and with special or more emphasis on aspects such as internal quality monitoring, academic freedom and administrative capacity.

Before being allowed to operate, private universities are first reviewed to assess their organizational character, their student protection policies and practices and their capacity to achieve excellence in learning. This review is done by experts in university management and requires a site visit. The institution has to demonstrate its capacity to operate as a degree granting institution. The assessment focuses on the mission of the institution, its administrative and academic capacity, its ethical conduct, its financial capacity, its resources and its student protection mechanisms. In Alberta, PCAB requires first an institutional self-review. In Ontario, PEQAB assesses the institution against a series of standards and benchmarks, which constitute the reference for preparing the application.

Periodic evaluation of programs
Professional programs are regularly evaluated for accreditation by agencies linked with professional boards. Most follow the same pattern, with a self-evaluation followed by a site visit. The visiting committee is composed of academic peers and may include representatives of the professional world. In selecting its evaluation criteria, the agency focuses on what is required to produce a professional of international standard. In this sense, they are outcomes oriented. These criteria can be quite specific. They cover a number of subjects, ranging from program objectives, curriculum, pedagogical methods, resources, student assessment to faculty and governance. This kind of evaluation is seen as an efficient way to ensure that professional programs are in line with the international evolution of the field and meet the highest quality standards. This explains that, even when they are not required to, many institutions will submit their programs to this scrutiny.

The periodic review of graduate programs performed by the Ontario Council for Graduate Studies follows the same pattern, with self-evaluation followed by a site visit by experienced academics. Their criteria are, however, very different. They start with the program objectives and focus on its organization, on the competence of its faculty in view of the fields of knowledge covered and the importance of research, on the resources available and on the outcomes in terms of graduates and other productions, such as publications.

The objectives of the periodic review according to OCGS are to « promote, maintain and improve the quality of graduate education », whereas in the case of professional accreditation, the objective is to identify those programs that meet the criteria of the agency. In the first case, the criteria are more general and tend to consider the program as an instrument to enrich the education of the student and develop his analytical and research skills. In professional accreditation, they are, on the contrary, much more prescriptive in terms of course content, laboratories and faculty composition. In all cases, the periodicity of the evaluation is between 5 to 10 years.

Audits of university quality assurance processes
In Ontario, Quebec and the Maritimes Provinces, the universities are periodically audited with respect to their quality assurance processes. The objective of this audit is to ascertain whether the procedures used by institutions to assess the quality of existing programs are efficient quality control mechanisms. The procedure used in Ontario and Quebec is essentially the same. In these provinces, the universities have agreed on a common framework for the periodic evalua-

tion of their programs. An audit Commission (UPRAC and CVEP) has been set up to verify that the universities comply with this framework and efficiently evaluate their programs. The commission is composed of respected academics. After receiving the necessary information, the Commission selects two or three programs, which have recently been evaluated, and analyses in depth the evaluation files. This examination is completed by a site visit. The Commission can make observations and recommendations to which the university must react. The procedure is essentially similar in the Maritimes, but in the absence of a common framework for periodic program review assurance, the universities are requested to begin the audit by a self-evaluation of their quality assurance processes. Then the Commission proceeds with the examination of the self-assessment, the analysis of the evaluation of a few programs, the site visit and the report to which the institution must react.

Past and Future

When presenting former and future trends in evaluation the CEEC is used as example of the trends that can be observed in the Canadian post-secondary education regulation mechanisms.[1] Five years ago, the CEEC was evaluating programs mainly for quality improvement. Its reports contained observations and recommendations that proved to be efficient means of improving the quality of college education and developing a culture of evaluation in the institutions. The Commission was, however, worried about the efficiency of college education as were other stakeholders.

A year ago, the mandate of the CEEC was widened and now covers the evaluation of the colleges' strategic plans and their plans to increase the success rate of their students. This is a clear indication that the Government of Quebec, as the Governments of other Canadian provinces, is now not only preoccupied by the quality of post-secondary education, but also by the efficiency of its institutions. Evaluations in the future will have to take into account quality indicators as well as efficiency indicators such as graduation rates. Furthermore, it is to be expected that more and more institutions will be judged not only by the quality of their programs, but also by the efficiency of their management and their capacity to establish and meet clearly identified and measurable objectives. The challenge for evaluation commissions like the CEEC will be to stay in support of the institutions and not become pure control agencies.

[1] *Likewise the CEEC is presented in the Canadian case presented in Part three of the anthology.*

France – Higher Education

Bruno Curvale
Project Manager
Comité national d'évaluation (CNE), France

The request from the co-ordinators of this anthology comes at a time when the *Comité national d'évaluation* (CNE) (National Evaluation Committee) is concluding in-depth reflections on evaluation issues and methods. Indeed, since it was created, the CNE has constantly sought the most suitable approaches for the specific context of higher education institutions, at any given moment. The rapid transformations of this context over the last few years have led the Committee to consider a complete overhaul of its approach to the evaluation of higher education institutions. The integration of national higher education into the European higher education area currently being created, the transformation of relationships between the State and institutions, the development of autonomy for universities and the new momentum towards regionalization are all factors, among others, that oblige us to rethink the question of independent external evaluation of institutions and all France's provisions in this area.

As far as possible, in addition to presenting the current transitional situation, the following responses will specify the points on which the CNE intends to change its approach in the future.

Values and purposes

To be fully understood, the CNE's work has to be placed in the wider context of developments in the French higher education system since the 1960's and the general move towards autonomy of universities since that time.

Before 1968, the structures and organization of French universities were still mainly those of the imperial university set up by Napoleon at the end of the 18th century. The university had a territorial basis, the academy. Universities were a set of faculties, institutes or schools organized according to discipline, and with little interlinkage. The Ministry of Public Education's division of higher education automatically had authority over these structures. Around twenty major cities had faculties that were formally grouped into universities that were overseen by the Rector, the representative of the Ministry of Public Education.

The two major transformations in the system comprise the building of pluridisciplinary institutions directed by authorities consisting of elected representatives (1968) and the transformation of the link between the State and institutions (1984 and 1989) through a contractual policy. Each university has to establish a strategy for its various areas of activity – teaching, research, management, etc. – and negotiate (for a four-year period) a contract with the State concerning objectives and means of support.

The main stages in the process of developing university autonomy were as follows:

1968 Higher education bill *(Loi d'orientation sur l'enseignement supérieur)*
Creation of higher education institutions.

1975 Colloquium of the Conference of University Presidents at Villard-de-Lans.
During this colloquium, the definition of evaluation as a condition of autonomy was clearly formalized.

1984 Bill on higher education as a public service and 1985 decree.
Creation of the National Evaluation Committee for Public Institutions of a Scientific, Cultural and Professional Nature (CNE)[1].

1989 Education bill.
The CNE became an independent administrative authority.
Contractual policy set up.

1995 Unique teaching-research contract established.

2001 Transformation of the procedure for "habilitation"
(State authorization to deliver national diploma).

Until then, institutions had to design their courses according to models approved by the Ministry of Public Education in order to be allowed to confer national state-funded diplomas. Institutions now develop their own courses within a framework of wide education sectors that are based on their areas of research excellence, with focus on the professionalization of students. Institutions' proposals are evaluated by a Ministry of Public Education mission (Scientific, Educational and Technical Mission).

Alongside the fundamental move towards developing management autonomy of public higher education institutions, the context of the mass phenomenon of higher education during the 1980s contributed to establishing the legitimacy of evaluation. From the very beginning, those in favour of the National Evaluation Committee clearly assigned evaluation the function of rendering the university system more transparent in the eyes of the public and of political and administrative decision-makers. Indeed, at the time, the image of university institutions was tarnished. There are several reasons behind the doubts that the nation may have had concerning higher university education. Two of them were:

- the fact that the universities, unlike the Grandes Ecoles that were protected by selective entry, bore the brunt of the consequences of increased student numbers on first-degree courses. Other issues, that had to be tackled rapidly, emerged at the same time as this influx of students and put great strain on accommodation capacity, including the need to deal with new populations that were under-prepared for university studies and the question of professional integration of graduates;
- questions concerning the research activities being set up in recently-created universities and their positioning within the national arrangements structured by the major research organizations and universities.

This calling into question of the university as an institution, and doubts as to its capacity to face the challenges being imposed, formed part of the basis for the National Evaluation Committee, whose primary objective is to report to the nation on results and quality of higher education.

Another preoccupation of the CNE's founders was to seek to report on the diversity of the institutions, in order to banish the stereotyped vision of homogeneity of universities and their results throughout France's territory. Reflections on efficient use of public resources and the notion of excellence are linked to this determination to show the diversity of the institutions' circumstances.

The CNE's work primarily concerns institutions and focuses on their policy within the framework of their public service mission. Evaluations of institutions are focused in particular on the quality of the institutions' research and teaching activities, human resource management, the working conditions of students in particular, regional integration and international cooperation. Parallel to the evaluation of institutions, which forms the bulk of the CNE's activities, it evalu-

[1] *Comité national d'évaluation des établissements publics à caractère scientifique, culturel et professionnel (CNE)*

ates sites and disciplines, and carries out theme-based studies on general issues related to higher education.

The CNE's activities are designed to help improve the running of institutions and their results across the board. This assistance is, naturally, in the form of assessment of the quality of the institutions' achievements. The CNE's evaluation work is, therefore, always a two-step process. The first step is analysis, with the aim of assessing the strengths and weaknesses of the institution; the second is a more systemic assessment, aimed at assessing the institution's capacity to govern itself.

The National Evaluation Committee for Public Institutions of a Scientific, Cultural and Professional Nature was created by a law dated 10 January 1984.

The CNE's work is defined by three decrees (dated 21/2/1985, 7/12/1988 and 23/9/1992). The CNE is responsible for examining and evaluating the activities of all universities, engineering schools and other institutions under the auspices of the Minister in charge of higher education. This represents 230 institutions under contract, including 82 universities, i.e. the bulk of France's higher education institutions, which also include private institutions such as business schools, five catholic universities and a few public institutions under the authority of other ministries. In 2002, France had over 2,140,000 students in higher education (CPGE and STS courses included[2]), 82% of whom were in institutions directly under the authority of the Higher Education Directorate of the Ministry of Public Education. Overall, the public higher education sector groups together some 92% of students.

The CNE makes recommendations to the institutions, to the Ministry of Public Education and to the partners of the higher education institutions. The law of 1984 specifies that: "(the CNE) recommends appropriate measures to improve the running of institutions and the effectiveness of teaching and research." It is up to each entity to implement the recommendations that concern it.

The notion of transparency and the search for the foundations of quality, maintaining or improving it, in higher education institutions are elements of the definitions of these three concepts: evaluation, development of quality and quality assurance.

Since quality and improvement of quality cannot be dictated from outside, the evaluator finds him or herself in the situation of a third party who, through their external viewpoint and contribution in terms of expertise, helps those in charge of the institution to improve points as necessary. The image of holding out a mirror is probably the most apt here. Evaluation is both an action, that of assessing the quality of a situation, state, strategy or result, and, through the public nature of its results, an approach to accountability regarding the carrying out of the public service mission entrusted to higher education institutions.

In general, the link between quality assurance and quality development is based on the concept of evaluation, which is defined through its assessment goals, transparency, and practice based on dialogue and trust.

In more concrete terms, the relationship between the concepts of quality assurance, quality development and evaluation is established in a pragmatic manner through the evaluation procedure. Evaluation is independent of both institutions and the Ministry. It aims, through a procedure recognized as legitimate and appropriate, and by establishing a constructive dialogue, to report on the strengths and weaknesses of the institutions. Evaluation results are public and are widely disseminated in the form of reports available in printed or electronic format from the CNE's Web site. This means that the information contained in these reports is available to all

[2] *Preparatory classes for the Grandes Ecoles (Classes préparatoires aux grandes écoles) and Advanced Technical Sections (Sections de techniciens supérieurs). These courses are organized in high schools.*

decision-makers, whoever they may be (e.g. students, political, economic or administrative officials), to assist their judgment and to provide a basis for their decisions. The publication of evaluation results also has a direct impact in encouraging the improvement of situations. The CNE's observations and recommendations are not binding, but the resulting publicity does provide a motivation for those in charge of the institutions. It has often been noted, in this respect, that a significant number of problems highlighted by the evaluators are corrected even before the report is published.

Finally, it should be noted that the CNE's procedures include an internal evaluation stage carried out by the institution itself. This stage is in itself a partial response to the needs for quality assurance and quality development. The results of internal evaluation (an information kit or internal evaluation report) are taken account of in the aims of the external evaluation. The quality of the information it includes and the analyses it contains are evaluated by the CNE.

There is an ongoing debate in the academic community and among its partners concerning the issues of quality, and the aims and uses of evaluation of institutions. We cannot pretend that there is a full consensus concerning the definitions presented above. However, it should be noted that diverging views probably concern less the actual definitions than the uses and results of evaluations. The needs of decision-makers vary depending on whether the issues at stake are clear cut, requiring a "yes" or "no" response, e.g. the decision to fund a given project for an institution, or whether the required decisions concern the appropriate way of envisaging a strategy to solve a specific problem, e.g. improving the results of a given student profile. In this respect, reflection on the use of evaluation should be integrated in the dynamic aspects related to the management of universities. A difficult situation, if it is well understood and well analyzed (evaluated), can, in the long term, provided a suitable policy is implemented, lead to positive changes. Conversely, a favourable situation may deteriorate if it is not monitored, and if the reasons for it are not well known and understood.

Objects

The CNE's main task is to evaluate institutions. This evaluation concerns the institutions' capacity for internal evaluation, and deals mainly with four sectors of their work: governance and management; courses offered; research policy; and the living and working conditions of students. The CNE's work also concerns the relationships institutions have with their environment and their partners.

Since most institutions have already undergone a full evaluation, the CNE's current strategy is to encourage a more selective approach to evaluation themes in order to enable more in-depth work, better suited to the current needs of institutions. This approach also aims to lighten the burden of evaluation procedures by focusing on the most sensitive issues for a given institution.

The CNE does not evaluate higher education institutions on its own initiative, but exclusively under the authority of the ministries responsible for higher education or private institutions. CNE can, however, respond to requests from the management or authorities in charge of such institutions.

The CNE does not carry out evaluations of personnel. This is carried out by the *Conseil national des universités* (National Universities Council) for university teaching staff and by joint committees for administrative staff.

Detailed educational evaluation of curricula is performed by the ministry responsible for higher education. Regarding research, scientific appraisal for the funding of laboratories is carried out by the major research bodies (CNRS[3], INSERM[4], INRIA[5], etc.) and the ministry in charge of research.

[3] *National Centre for Scientific Research (Centre national de la recherché scientifique).*

[4] *National Institute for Health and Medical Research (Institut national de la santé et de la recherche médicale).*

Evaluation is a concerted approach. This is how criteria and indicators were established in 1986, and then redefined in 1994, jointly by the National Evaluation Committee and the Conference of University Presidents (CPU). When evaluations are carried out, there are many exchanges between the institutions and the National Evaluation Committee, e.g. concerted reflection on the evaluation methodology and the questionnaire for internal evaluation; discussion of the themes of expertise chosen for evaluation; on-site visits by members of the CNE, the general secretariat and various experts. The draft report itself is submitted to those in charge of the institution for review, as they are also responsible for validating the data published in the report. The head of the institution has the final word, and his/her response is published at the end of the evaluation report.

The development of evaluation, its acceptance by the academic community and the need to increase its scope, led the CNE, a few years ago, to promote the idea of selectivity in the themes for external evaluation. This selectivity, whose corollary is the abandoning of exhaustive evaluation of institutions, led to the development of the concept of an evaluation contract. This important step in the evaluation procedure, which is performed before the external evaluation procedure carried out by experts, is an agreement between the CNE and the institution concerning the main goals of the evaluation. The CNE seeks to ensure that the external evaluation themes are relevant to the specific problems of the institution in question. These themes are chosen during the preliminary study before evaluation. They can be determined based on the self-evaluation report or other sources. The needs of the State and its partners are taken into account when defining the objectives. For example, current emphasis is often on the setting up of the BMD system (Bachelor-Master-Doctorate), or contractual relationships with local authorities.

Stakeholders

The CNE has multiple tasks. By law, it has to inform the nation about higher education in the country. It evaluates and makes recommendations to encourage improvements in the situations of institutions. Finally, in its reports, it makes particular recommendations on measures that could improve the public service provided by higher education. There is, therefore, a wide range of possible audiences: those in charge of universities and other institutions evaluated, and their staff; the ministries in charge of them; politicians; students; and employers, etc.

The diverse nature of the audiences leads us to look at the issue of drafting the reports. This point is currently being debated at the CNE. Clearly, the current format of the reports does not always facilitate reading by all the various target audiences (families, students, university heads, administrative and political officials) who have different levels of knowledge of the institutions and how they are run. The CNE aims to produce full reports that present both conclusions and recommendations, but also informational and descriptive elements.

The fact that the entire reports are public does not always encourage quality and accuracy of criticisms and recommendations. The difficulty here is linked to the conditions of the exercise, which supposes transparency and honesty between those involved in the evaluation process, while asserting the public nature of the results. For accreditation purposes based on a status at a given moment, this does not pose any real problems. However, if the aim is to assist improvement, the question as to the ideal conditions under which true management can function has to be asked. In concrete terms, the CNE's current evaluation procedure for institutions provides for a meeting with those in charge of the institution evaluated to discuss the content of the draft report.

The difficulties encountered in working towards a consensus indicate that the variety of applications of evaluations and the needs of the various parties are irreducible. Consensus is probably not possible concerning applications. The CNE's approach is to provide elements that are

---

[5] *National Research Institute for Informatics and Automation (Institut national de la recherche en informatique et en automatique).*

useful for reflection and decision-making, and not to substitute itself for the various decision-makers whose motivations and objectives may be different.

Methods

Responsibility

The CNE sets its own agenda, and carries out approximately fifteen institution-evaluations on average per year. In addition to these, there are one or two cross-cutting evaluations (of specific disciplines or themes). At present, the CNE aims to select the institutions to be evaluated from those that are to enter discussions with the ministry concerning objective-oriented contracts.

The CNE is responsible for its evaluation methods. These methods are constantly reviewed by the working groups of the Committee's General Secretariat. There are regular seminars involving Committee members and project managers to review the evaluations carried out. This methodological way of working provides a general framework for the evaluation teams, who have considerable freedom to adapt methods to the purpose of the evaluation. The CNE is not looking to standardize its methods, and gives priority to a flexible approach adapted to each institution.

The aim of the evaluation is a better understanding of each institution and its unique characteristics and to provide elements (analyses, data, etc.) that may support the implementation of improvements. These objectives provide an explanation for the fact that the CNE has not chosen a directive approach to guide the experts' work, as one example. Experts are informed about the institution and the general objectives of the CNE. They have considerable autonomy within the framework of their assignment. In particular, they are responsible for selecting the points to be highlighted in their reports. The exploratory nature of evaluation and mistrust of formalism in the approach are two important characteristics that have an impact on the relevance and the quality of the experts' work. The CNE does, however, seek to reinforce the basic briefing given to the experts. The introduction of a new approach to evaluation based on terms of reference will provide the opportunity to formalize the CNE’s expectations towards the experts.

The CNE is also responsible for the assessments and judgments published in its reports. It uses experts' contributions as a basis for its own assessments for which it then assumes full responsibility. Finally the CNE is responsible for the recommendations expressed in its reports. The evaluation reports are public and they are available in a limited number of hard copies and on the Committee's web site.

Follow-up is carried out by the CNE. The setting up of follow-up procedure is an ongoing project, and a totally satisfactory solution has not yet been found.

Characteristics

The evaluation criteria are adapted to the circumstances. Evaluation of institutions is based on a formal framework that provides for an internal evaluation stage carried out by the institution, an analysis and reflection stage carried out by the CNE, whose purpose is the implementation of an external evaluation carried out by experts mandated by the CNE and adapted to the institution's needs. These methodological documents can also be viewed in their entirety on the CNE's Web site.

Both quantitative and qualitative approaches are used. The search for an objective evaluation rests initially on indicators that have been fine-tuned with experience. The quantitative approach depends and rests on the experts' qualitative assessment.

Evaluation is by nature quantitative and qualitative, with statistical data providing a basis for analysis and allowing the gathering of homogeneous data in a national context. Nevertheless, the National Evaluation Committee feels that the evaluation of higher education is first and foremost qualitative. This is why it does not believe in the exclusive use of performance indica-

tors, and takes into account the context, the specific situation, the evolution and the specific objectives of the institution evaluated.

The evaluations are both process and result-oriented. The CNE is eager to learn about the real situation of an institution (results, professional integration of students, research policy, self-evaluation capacity, etc.) in order to help it improve. But internal quality assurance processes put in place by the institutions are also a focus for external evaluations.

Relationship between values and method

The relationship between the motivations behind evaluation and the actual method used must be as direct as possible. The crucial issue here is that of adapting the methods to the requirements of the evaluators and the organisations evaluated. In this regard, dialogue is a priority, and the CNE seeks, as far as possible, an agreement with the institution heads concerning the methods and objectives.

The question of the relationship between values and method is relevant, and leads us to question the very nature of the evaluation process. The CNE defines its actions in two dimensions – assessment of the quality of the institutions' actions on one hand and, on the other, analysis of their capacity to govern themselves. Nonetheless, alongside this definition of the evaluation through its products in terms of assessments and recommendations in these two areas, is another dimension in which evaluations produce a useful effect on the development of quality in higher education. This is the evaluation report itself as a reasoned description of the situation of an institution or site at a given moment in its history.

This representation of a complex situation that concerns a number of parties and issues is an important product of the evaluation process. The evaluation, if it is successful, leads to a representation of a situation(s) that can be appreciated and utilised by all parties. This representation is, as such, a useful tool for partners and heads of institutions. Because it is shared, i.e. recognized as valid by all, it enables improved choices and decisions, and generally improves decision-making beyond the proposed recommendations.

Here we can see the importance of the method(s) for the subsequent acceptance of the results of the evaluation procedure. The value of the evaluation depends largely on the fact that the parties and partners recognize it as relevant. The choice and the conditions of interpretation of quantitative indicators, for example, must be subject to agreement between the evaluators and those in charge of the institutions evaluated. The same is true regarding the choice of evaluation themes and the way in which evaluations are carried out.

This need for transparency concerning methods and assumptions led the CNE to make its expectations more explicit by drafting them, in partnership with the IGAENR[6], in the form of references proposed to the Conference of University Presidents as shared support material for the evaluation.

Basically, the evaluation is designed in a spirit of partnership and dialogue. The results of evaluations cannot be the subject of negotiations, but upstream in the process there is ample scope for discussion and agreement on the methods and tools in order to ensure the quality of the evaluation and respect for the institutions evaluated.

Past and future

A look at the CNE's past shows consistency in its values and objectives since the very beginning. The objectives are those entrusted by the law. The aim is to help improve institutions and the public service provided by higher education and to inform the nation; the basic values are those of evaluation - dialogue and partnership.

[6] *General Inspection of the Administration of Public Education and Research (Inspection générale de l'administration de l'éducation nationale et de la recherche).*

Having said this, we observe that subjects and methods have, on the contrary, evolved considerably. It would be impossible to try to present the details of the successive points that have changed one by one during a continuous adaptation process. We can, however, go over the main factors behind this adaptation and the trends that underpin the future of the CNE's work.

Regarding the factors that explain the change in preoccupations (the subjects of evaluation) and methods, we can without hesitation say that the main one is the adoption of evaluation by the academic community. Internal evaluation capacities are different from one institution to the next, and not always at the desired level, but, globally, the CNE today deals with institutions that are aware of the importance of evaluation. This evolution becomes quite clear when reading documents produced by the CNE to guide the institutions' work. When it began its work, the CNE used a questionnaire to gather quantitative data and requested global analyses of the strengths and weaknesses from those in charge of institutions; then a greater link was asked for between the data produced by the information systems and self-evaluation capacities through use of a detailed questionnaire; today the CNE asserts legitimate expectations (the references) in collaboration with other partners, and institutions are asked to demonstrate that they satisfy them.

A second factor that encouraged change was the recognition of the task evaluation represents for the institution. It must be recognized that lightening the procedural burden of evaluation is a major challenge that concerns first and foremost the evaluator. In this regard, clarification of the evaluation terms of reference is a decisive step as these references clarify the subject of the evaluation. In addition, the philosophy of demonstration that gives this new approach its meaning, allows the institution a free rein to use the methods it considers appropriate to show that the references are satisfied. It is then up to the evaluator to check the credibility and quality of these demonstrations. By decreasing the formalism of the response and encouraging authentic expression by the institution, we see that it is possible to lighten the burden and to improve the relevance of the responses. In conclusion regarding the reduction in the workload, it should be mentioned that the references prepared together by the CNE and the IGAENR are destined, following improvement (the work is underway), to be recognized by institutions and their partners as an essential basis for internal quality assurance measures.

A third factor to mention is naturally the entire move towards European convergence and strengthening of university autonomy. In this highly demanding context for institutions, it is up to the evaluator to stress the more urgent issues in order to encourage institutions to adapt and succeed.

Concerning methods, the most striking changes are:

- The increasingly wide acceptance of the concept of peers, which today includes experts from any of the stakeholders. This development, which underlines quality and competence as related to a given academic theme, and not just the concept of belonging to the academic milieu, is a good indicator of the legitimacy of evaluation, whose recognition relies above all on demonstrating the relevance of the methods;
- More account taken of the work of other bodies whose subject is higher education (and research), and more coordinated work with them (IGAENR, CNER[7]). Here, there is a direct impact on the work requested of institutions through better coordination of agendas and more reuse of data gathered and results produced;
- The development of group work by experts and extended visits to institutions;
- The reorganization of evaluation agenda. Here the aim is to increase the preliminary time spent before choosing the experts and visiting the institution to encourage better control over the choice of the specific themes for the evaluation.

[7] National Research Evaluation Committee (Comité national d'évaluation de la recherche). The CNER is responsible for assessing the implementation and results of national research and technological development policy defined by the government.

- The finalizing of evaluation terms of reference and development of the philosophy of demonstration. This approach, which respects the autonomy of higher education institutions, leaves institutions free to demonstrate that they do indeed satisfy the evaluators' legitimate expectations.

To conclude, we cannot fail to highlight the fact that European convergence in higher education will be an important factor for change for the CNE and its working methods in the years to come. The principle of subsidiarity, a guarantee of the respect for national choices and the conditions of mutual trust will impose greater transparency vis-à-vis European partners. The setting up of external evaluation of evaluation agencies, and the taking into account of European users in evaluations, are two challenges the CNE is preparing to face.

Hungary – Higher Education

Tibor R. Szanto[1]
Secretary General
Hungarian Accreditation Committee (HAC)

Values and purposes

Quality assurance of higher education (HE) on a national scale was introduced in Hungary in 1992-93. That was the time when universities[2] regained the right to launch doctoral training and award PhD/DLA (Doctor of Liberal Arts) degrees. Until then, following WW2, academic degrees (*Candidatus scientiarum, CSc,* and *Doctor of the Hungarian Academy of Sciences* [HAS], *DSc*) were conferred by the *National Committee of Scientific Qualifications* operating alongside the HAS. Besides these degrees, there was a *doctor of university* (or so-called "small doctor") degree conferred by universities, but this was at a lower level than a western-type PhD, while the *Candidatus* degree was generally more demanding then a normal PhD.

From the 1950's, research and teaching became relatively separate in Hungary. Following the Soviet pattern, a rather large network of research institutes belonging to the Academy of Sciences was established. A considerable proportion of Hungarian scientists worked (and still works, though in decreasing numbers) in these institutes. The academic degree granting mechanism was based – logically enough – on that culture. This was perhaps one of the main reasons for the reallocation of higher degree awarding rights following WW2. The political reform in 1989, however, generated changes in this respect too.

The transformation of Hungarian higher education, approaching the Western European and American patterns and standards, was thought to be necessary for several reasons.[3] Beyond renewing and extending the autonomy of institutions, there was, first of all, a strong desire to become compatible with higher education in the European Community. While, on the other hand, it was hoped that funds could be raised for the restructuring of the Hungarian higher education system.[4]

An important and actual impetus in this respect was that Hungarian universities could not have joined the European associations without having the right to confer PhD degrees. Thus, the task and the direction for change were clear. Then the question arose: should each and every university get the right of launching doctoral training unconditionally? Some experts thought that this was an opportunity for introducing quality standards. The idea was reinforced by the growing international trend to stress quality management and improvement in teaching and learning in higher education. Thus it was agreed, that there was a need for an independent

[1] *The views in this paper are the authors', and HAC or other staff members do not necessarily share them.*
[2] *Hungary has a dual higher education system comprising universities (5-6 years of education) and colleges offering vocationally oriented training of 3-4 years for a degree.*
[3] *Marianne Csáky, Interview with András Róna-Tas.* BUKSZ *1995 Spring, pp. 74-86 (in Hungarian).*
[4] *The latter hopes were realised with the signing in March 1998 of a 5 year loan agreement with the World Bank for USD 150 million (an additional USD 100 million was tied to this amount from the Hungarian state budget) for the structural reform of the system. The result of this was the so-called „integration process", the merging of institutions by January 2000. The loan agreement was, however, terminated by the Hungarian government in May 2001. The former desire, to join the EU, is scheduled to be satisfied in May 2004. On the way to this, on 12 April 2003, 84 % voted in favour of joining the EU (approx. 38% of eligible voters).*

body of highly qualified experts from various disciplines, who would be responsible for supervising the quality of doctoral training and higher education in general in Hungary.

In August 1993, the Higher Education Act (HEA) appeared, and in it the legal framework for PhD training and accreditation was established. The choice between accreditation[5] and "softer", audit type, evaluation was intentional and well justified: preliminary accreditation seemed to be the most promising and suitable means to raise strict quality demands and requirements for the new doctoral programmes in particular, and for new degree programmes and institutions in general.

The *Provisional National Accreditation Committee* (PNAC) was formed in the autumn of 1992, and it immediately invited applications for launching doctoral programmes. Evaluation of applications was done with the promise that the *Hungarian Accreditation Committee*, to be founded, would positively consider the decisions of its provisional forerunner. That was the beginning.

In the HEA (in force from 1 September 1993), the Accreditation Committee was given the additional legitimacy to accredit HEIs as institutions and, in general, was established "*for the ongoing supervision of the standard of education and scientific activity in higher education, and for the perfecting of evaluation there*" (HEA 1993, Section 80 (1)). Upon the nomination of the HEIs, the HAS, and other organisations, in January 1994 members of the Hungarian Accreditation Committee (HAC) received their three year mandates from the Prime Minister, elected their president, and began work on processing the decisions of the PNAC.

Foundation values

Thus, in the background to the formation of the Hungarian accreditation system, six basic values (let us call them "*foundation values*") can be identified:

- *V1 internationalisation*
- *V2 autonomy*
- *V3 commitment to quality and quality development*
- *V4 central/national supervision* (control)
- *V5 accountability* (value for [state] money)[6]
- *V6 independence.*

I think it would be too simple – though alluring – to say that the dominant purpose of creating the system was indirect state control of higher education[7] through a buffer organisation. It certainly cannot be denied that such an intention was present on the part of the state administration. Other actors in the process, however, had different vision and priorities. The initiative for setting up the system actually came from experts of the university sector, and they estab-

[5] "Accreditation" *is understood here as a special kind of evaluation using predetermined quality requirements or standards as threshold criteria for the yes/no accreditation decision. I use the word "*evaluation*" in the wide sense as it is defined in the glossary of this book. Accreditation always involves evaluation, but not every evaluation is accreditation at the same time.*

[6] *As the founding father of the Hungarian accreditation system,* András Róna-Tas*, put it in 1991-92, realising that the sector had to make a deal with government concerning quality, quality development, and allowing quality control for extra resources from state budget: „I had several meetings with* Mihály Kupa*, the minister of finance then, who told me that he was not willing to give an extra penny to the sector until he got guarantees for structural reform and quality improvement." Interview with* András Róna-Tas*, op.cit. p.77.*

[7] *See e.g. V. Tomusk, "When East meets West: decontextualizing the quality of East European higher education"* Quality in Higher Education 6*(2000) No. 3. pp. 175-185.; D.F. Westerheijden, "*Ex oriente lux?*: National and multiple accreditation in Europe after the fall of the Wall and after Bologna"* Quality in Higher Education *7(2001) No. 1. pp. 65-75.*

lished both the framework and the Committee itself.[8] There has not been any representative of the ministry in the HAC. What is more, *civil* servants, i.e. persons working in the state administration, cannot be either HAC or subcommittee members.[9] (HAC staff members are full time employed as *public* servants, i.e. they have the same status as employees of the higher education sector.) The value of independence (V6) thus dominates the value of central/national supervision (V4). I shall return to the discussion of structural relationships of values later.

When, in 1999/2000, the HAC compiled a *Self-evaluation Report* for the purpose of its international external evaluation, it specified the main *objectives* of the accreditation system as follows:

- *"O1 Public protection of the "stakeholders" of HE, students, employers, society at large;*
- *O2 promoting quality improvement;*
- *O3 and to an increasing extent, accountability to the Government and the public about the quality of education."*[10]

Thus, it is important to note that the values and purposes of control and quality development were (and are) both present. The history of the system shows, however, that no matter what the real intentions were, in the actual operation of the HAC, the emphasis in the first eight years (the first cycle of institutional accreditation) was on the accountability and control side. (V5 and V4 periodically dominated V3.) This feature, however, had to do not with the intentions or interventions of the ministry or other state authorities or officials, but rather with the general legal embedment and, more importantly, with the HAC's detailed procedural and operational regulations, and input focussed requirements and evaluations. That is the historical path, the genesis of the system, together with, and reinforced by, internal operational characteristics and peculiarities of applied methods, which influenced or even determined the prevailing values.

However, (Yes, "The times they are a-changin'. ") the focus is now shifting from accountability towards development and improvement. This shift is reflected (externally) in the 2000 amendment of the HEA where, in section 80 on HAC, the word "supervision" was replaced by "validation of quality". Internally, the intention of this shift is clearly articulated in the *Strategic Plan* of the HAC, finalised in February 2002.[11] Implementation is due in the second cycle of institutional accreditation and in the pilot parallel accreditation and evaluation of degree programmes in history and psychology, respectively.

It is clear from the above that the question of purposes and values is a very delicate one, requiring and deserving due consideration and thorough analysis.

I would like to stress that the above foundation values asserted their effect together. They were, and are, intertwined and, moreover, they constitute only a part of the value world of the Hungarian accreditation system. Other values and more structural relationships will be discussed in connection with operation and method later in this paper.

[8] *„The ministry did not take part in the setting up of the Committee, I did not even inform them. (…) I thought that this must be an internal arrangement and agreement by and within the sector."* András Róna-Tas*, op.cit. p.78.*

[9] *Government Decree Nr. 199/2000. (XI.29.) on the operation of the Hungarian Accreditation Committee, Section 23 (1). (In Hungarian)*

[10] The External Evaluation of the Hungarian Accreditation Committee *(ed. T.R.Szanto). Budapest, 2000. p. 15. Accessible also at http:/www.mab.hu/english/index.htm,* Publications *section.*

[11] *See http:/www.mab.hu/english/index.htm,* Regulations *section.*

Concepts and actors

There is no wide consensus on the interpretation of the basic concepts of QA in HE in Hungary, though they were defined in the *Higher Education Act* (2000 amendment) as follows[12]:

- *Quality assurance system*: a system of deliberate and organised activities covering the whole institution which serves the constant approximation of the professional objectives of the institution to the actual operation of it and which is focused on the fulfilling of the needs of the direct and indirect partners, especially of students (including adults participating in further education), employers, those who order researches and the national and international scientific community.
- *Quality control:* examines the compliance of the operation of the higher education institution with the prevailing provisions of law and with the documents of the higher education institution from legal, economic and educational - professional points of view.
- *Quality evaluation:* compares professional objectives with the operation of the institution in respect of result, efficiency and quality on the basis of a system of indicators prepared together with the professional community concerned.
- *Quality validation* (accreditation): the procedure conducted by the Hungarian Accreditation Committee by which – in the framework of institutional and programme accreditation – it examines whether the educational and scientific activity conducted in the higher education institution, and the professional and infrastructural level of the educational programmes, and the personal and organisational conditions of the institution comply with the accreditation requirements prepared and published by the HAC.

"Quality development" is not defined in the Act, though it can be interpreted as part of the quality assurance system.

Concerning the interrelations of concepts, it is interesting to note that there is no explicit link between "evaluation" and "validation" on the conceptual level. In accreditation practice, however, there is a strong connection between the two: validation is based on evaluation, though the emphasis in the latter is not on "objectives" but on the "system of indicators" which are more or less identical with the "accreditation requirements" mentioned in the definition of validation. With the development and strengthening of inner institutional QA systems and activities, and with the HAC's focus-shift from control towards improvement, evaluation of institutional activities and outcomes of teaching and learning against objectives will certainly gain more ground.

As to QA activities, there are three main (groups of) actors in the QA scene in Hungary (this is sometimes called the "quality triangle"):

- higher education institutions,
- the Hungarian Accreditation Committee, and
- the Ministry of Education (ME).

According to a *Guidebook* compiled by the ME, quality policy comprises four main components:

- *Quality control* is the responsibility of the ME.
- *Accreditation* is the responsibility of the HAC.
- *Quality evaluation* is the responsibility of the ME, HAC, the Rectors' Conference, the NUSH (National Union of Students in Hungary), the "market actors", etc. (HEIs themselves do not appear in this list.)
- *Quality assurance* is the responsibility of HEIs.[13]

[12] *Act LXXX of 1993 on Higher Education. Section 124/E. (See http:/www.mab.hu/english/index.htm,* Regulations *section.)*

Among participants, there is more or less consensus on this distribution of responsibilities. However, penetration of the above concepts and intensity of activities differ at various institutions depending on leadership and/or individual interest and the motivation of faculty members.[14] Moreover, there are various interpretations of concepts, even at the "expert level". Special mention is due to profit oriented individual QA experts and firms who are eager to slice out a decent share from the promisingly emerging Hungarian higher education QA cake looming on the horizon.[15]

Objects

In accordance with the detailed legal regulation of the system, there are clear-cut definitions of what has to be evaluated. Beside the respective sections of the HEA there is also a government decree on the tasks and operation of the HAC containing detailed prescriptions on what to accredit and/or evaluate.

Accreditation in Hungary is mandatory for both public and private HEIs, both at the institutional and programme levels in each sector. (The HAC also gives its opinion to the minister before licensing the operation of a foreign HEI in Hungary.) Moreover, the HAC has some other tasks, too, that are not included in the above categories. The most important of these is perhaps the evaluation of applications for professorial appointments, a task that was prescribed for the HAC by the 2000 amendment of the HEA. Here individual persons are the objects of evaluation. (See more on this in section 5.) An additional important activity is the accreditation of national qualification requirements (NQRs)[16].

Thus, according to the *type of object*, there are four main kinds of evaluation performed by the HAC:

- institutional and
- programme accreditation (including doctoral training),
- accreditation of NQRs
- evaluation of individual persons.

According to the *status of the objects* of evaluation, there are:

- "preliminary accreditation" of *new* institutions and faculties to be established and new programmes to be launched, and
- accreditation of *operating* institutions (I), faculties (F), and programmes (P).

Preliminary accreditation and the accreditation of NQRs is based on initiation by organisations and institutions applying for licences of operation for new HEIs, faculties, or programmes. In the decision making process the minister of education asks for the opinion of the HAC as to the quality of the new institution or programme. (The minister licenses new programmes, while new faculties and institutions are licensed – upon the suggestion of the minister – by the government and the Parliament, respectively.[17])

---

[13] Guidebook for the Quality Development of Higher Education Institutions. *Budapest: Ministry of Education, October 2001. (In Hungarian)*

[14] *There are institutions or units where complete QA systems (ISO or EFQM-based) have been introduced while at some other HEIs only meagre elements of conscious QA activities can be found.*

[15] *In 2002 the ME made available HUF 60 million (~ € 250,000) for open competition for HEIs with the aim of supporting the creation and development of institutional QA systems.*

[16] *A new kind of degree programme can be established only after approval and publication of its NQR by the government (undergraduate programmes) or by the minister of education (specialised postgraduate programmes). NQRs are worked out either by the minister or HEIs. They contain the description of training goals, the duration and main fields and subjects of studies, examinations to be accomplished. The HAC gives an opinion to the minister on drafts of NQRs.*

[17] *It must be noted here that the HAC's job in almost each of its fields of activity is to advise the minister, on the basis of assessing quality, as to granting, suggesting to grant, or maintaining a licence of operation. The only*

Accreditation of operating institutions, faculties and programmes is done according to the schedule determined by HAC, and based on relevant regulations (an 8 year cycle). An important characteristic of this category is that in the first cycle accreditation of operating faculties and programmes was performed in the framework of institutional accreditation, i.e. a given institution and its faculties and programmes were evaluated and accredited in the same procedure, at the same time. However, in 2003/04 – in accordance with the *Strategic plan* of the HAC – two pilot projects of parallel accreditation and evaluation of programmes in the fields of history and psychology respectively, will be performed.[18]

Figure 1
Objects of evaluation according to types and status

<table>
<tr><th>Status
Type</th><th>New</th><th>Operating</th></tr>
<tr><td>Institution, faculty</td><td>preliminary accr.</td><td rowspan="2">accreditation
I, F, P together,
8 year cycle</td></tr>
<tr><td>Programme</td><td>preliminary accr.</td><td>planned pilot parallel accr., eval. of programmes</td></tr>
<tr><td>National qualification requirement (NQR)</td><td>accreditation</td><td>———</td></tr>
<tr><td>Individual person</td><td>eval. of applications for professorial appointments
(since 2000/01 only)</td><td>(only when a second [or further] professorship is applied for)</td></tr>
</table>

Stakeholders

The most important stakeholders of the Hungarian accreditation system are the following:

- students (and their parents)
- higher education institutions
- the state administration for (higher) education
- employers
- society at large.

Three of these stakeholder groups participate in the work of the HAC through delegated members:

- students (1 non-voting member [+ 1 representative of PhD students is invited])
- higher education institutions (15 members)
- employers (8 members delegated by the HAS, 7 members delegated by various chambers and professional organisations).

---

exception is the launching and operation of doctoral (PhD/DLA) schools *where the HAC is the "final authority" having the right of licensing, too.*

[18] *See more on this in T.R. Szanto, „Programme accreditation in Hungary: Lessons from the past, plans for the future". Paper presented at the INQAAHE biennial conference in Dublin, Ireland, April 16, 2003. (Forthcoming in* Quality in Higher Education*)*

In order to preserve independence from state administration, it was decided that there should be no representative of the ME in the HAC.[19] Nevertheless, since most of the HAC's resolutions are suggestions to the minister concerning the operation of institutions and programmes, the interest of state administration in the work of the HAC is obvious. Similarly, there is no lay person in the HAC but information on the results of accreditation, on the quality of institutions and programmes is provided through publications, the HAC's website, and answering e-mails and telephone enquiries. As to the direct consequences of accreditation, clearly, HEIs (and students) are affected most.

In spite of the discussed genesis of the system, the above composition of the HAC, and their direct participation, HEIs tend to regard the HAC and the accreditation system more as of a means of state–administrative control than an organisation and arrangement serving their own interests (quality improvement), too. Reasons for this can be found in the legal embedment and detailed regulation, the strong ties to licensing by the ME and Parliament, and the HAC's emphasis on minimum (input) requirements.

Actual interest on the part of students, parents and the public varies. In the first 6-8 years of operation, it was not strong and evolved only slowly. Basically, that has to do with insufficient publicity generated by both the ME and the HAC in the past. The situation is now changing, especially since the HAC renewed and enriched its website in September 2001.

In the light of this brief stakeholder review, two *"stakeholder values"* emerge (one of them was mentioned previously among the foundation values), to which we can add a third one taken from the objectives (O1) discussed earlier:

- (V6) independence
- *V7 harmonisation of viewpoints of stakeholders,*
- *V8 consumer protection* (public protection of the "stakeholders" of HE).

Methods

Figure 2 gives an *overview* of the procedures and methods applied by the HAC according to objects of evaluation.

[19] *Since HE expertise is obviously essential in the work of HAC, independence from HEIs is sought by other means than non-participation. See the later discussion of operational and ethical values.*

Figure 2
Methods of evaluation according to type and status of objects

<table>
<tr><th>Status
Type</th><th>New</th><th>Operating</th></tr>
<tr><td>Institution, faculty</td><td>- submission (self-description)
- peer review (with visit)
- peer group opinion
- second level opinion</td><td rowspan="2">- self-evaluation
- peer-review with visit
- external report
- second level committee opinion
----
same as above with parallel evaluations and comparisons (planned)</td></tr>
<tr><td>Programme</td><td>- submission (self-description)
- 2 or 3 peer judgements
- disciplinary committee opinion
- second level opinion (HAC's respective college)</td></tr>
<tr><td>National qualification requirement (NQR)</td><td>- submission (draft NQR)
- 2 or 3 expert judgements
- disc. committee opinion
- second level opinion (HAC's respective college)</td><td>———</td></tr>
<tr><td>Individual person</td><td colspan="2">- application (CV, list of publications, etc.)
- 2 or 3 peer judgements
- expert group (disciplinary committee) opinion
- Professorial Appointments Committee opinion</td></tr>
</table>

The final result of evaluation in each case is a formal HAC resolution, being the third level of expert group opinion, which is either a decision (PhD/DLA schools, in the "programme" category) or an opinion (recommendation) to the given institution ("individual person" category since 2002/03) or to the minister of education. It must be mentioned, however, that though these "opinions" are not legally binding, the final decisions by the minister are, in most cases, in accordance with them.

There are slight differences in procedures and methods in accordance with the differences in objects (e.g. the evaluation reports on operating institutions/programmes are obviously more detailed than the peer group or disciplinary committee opinions on new ones to be established). Nevertheless, the applied general methodology is basically the same in each category, with peer judgement at the core, and it follows the internationally well-known (and recommended) scheme[20] consisting of:

- self-evaluation
- site visit by external experts
- external report.

[20] *See e.g. D. Kristoffersen, A. Sursock, D. Westerheijden,* Manual of Quality Assurance: Procedures and Practices *(Phare Multi-Country Programme Report). European Training Foundation, November, 1998. p.18.*

Internal and external elements

The relationship between internal (institutional responsibility) and external (HAC responsibility) elements in the evaluation procedure is as follows:

- Evaluations are *initiated* partly *internally* and partly *externally*. Operating institutions and programmes are evaluated in an 8-year cycle (HEA), in the sequence determined externally, by HAC. While institutional initiation is only *quasi internal*, since accreditation (new institutions, faculties, programmes, NQRs) or evaluation (professorial appointments) by the HAC is compulsory in Hungary.
- Evaluation *method* is worked out *externally* by the HAC, though many experts are involved in this process from universities and colleges, too (About two third of the HAC members are full or part time professors themselves.). Considering, also, meetings and discussions with institutions and their representative organisations (Rectors' Conferences), the process could be regarded almost as *quasi external*.
- *Assessments* in the evaluations are, again, *both internal* (self-evaluation) *and external* responsibilities.
- *Recommendations* in the evaluations have been made *externally* by the HAC, but with the focus-shift from control towards improvement this could change a little in the future. Self-evaluation reports are planned to include a SWOT analysis, too. Applying this tool, institutions will have a chance to directly reflect on their capacities for change, and make actual quality improvement proposals internally, for themselves.
- Until now, *the final accreditation decisions have been published* without the detailed evaluations and grades.[21] Since 2002, in accordance with the law on data of public interest, however, the HAC is obliged to make any of these past grades – though, without the detailed evaluations – available upon request from interested parties.
- *Follow-up* of evaluations is the *mutual responsibility* (I+E) of the institutions and the HAC.

Criteria and procedures

Evaluation criteria and procedures are defined and published. It must be mentioned, however, that criteria (requirements) and procedures did, and do, change over time. The institutional accreditation of the first five HEIs, for example, was implemented as part of a pilot project in 1994-95. Criteria were fine-tuned based on the experience of those (and later) actual evaluations. The *Accreditation Guidebook* for institutional (and operating programme) accreditation was revised several times during the first cycle. (It had seven editions.) Supplements dealing with the specialities of church-run institutions and distance education programmes were added as recently as in 1998 and 1999 respectively. While as far as the accreditation of new programmes to be launched is concerned, a comprehensive set of general requirements was published in 1998, followed by a detailed, discipline-specific, set of requirements in 1999.

But these changes, I believe, are more or less normal. The context of higher education, teaching methods, materials and the technological background are all changing, and standards and requirements have to change accordingly. Otherwise, the government decree on the HAC prescribes that accreditation and evaluation criteria and procedures must be reviewed and revised, if necessary, every three years. Corresponding to this, the latest version of accreditation requirements was published in 2002.[22] Also the procedure and criteria for evaluating applications for professorial appointments were revised in 2002 (available on our website in Hungarian). The specific procedure and criteria for the second round of institutional accreditation are to be worked out in 2003. Similarly, the procedure and criteria for the pilot parallel accreditation and evaluation of degree programmes in history and psychology should be completed by July 2003.

[21] *In the first cycle of institutional accreditation, each degree programme was evaluated on a four-value scale, the grades being:* Excellent, Strong, Satisfactory, Not satisfactory. *In the second cycle there will be only three possible evaluation outcomes:* accredited (for 8 years), provisionally or conditionally accredited (for shorter period), and not accredited.

[22] *See http:/www.mab.hu/english/index.htm, Regulations section.*

Evaluation approaches are *both quantitative and qualitative*. The present approach can be briefly considered: the HAC *"measures and considers"*, ("*mér*" and "*mérlegel*"), where the latter word in Hungarian implies both physically balancing on a pair of scales, or weighing machine, and considering or pondering over mentally. (In the HAC's logo a balance is composed of the acronym.) But since external evaluation is a delicate enterprise, and it is not always easy to accept its outcomes for those under the threshold level, the HAC applied (and still applies) quite a lot of quantitative criteria, requiring the institutions to include many data in their accreditation materials.

This characteristic has to do with the context of the establishment of the system, as well. Clear and unequivocal criteria were needed in order to be accepted by the people involved, among them high standing professors, who were to be "measured" by others.

Similarly, this context explains another characteristic of the Hungarian accreditation system: evaluations have been rather input oriented, with few elements of process and outcome assessment. That feature is, to some extent, also connected with the tasks of the HAC, since preliminary accreditation is by nature input-bound. The intention is, however, to shift the emphasis of evaluations from input to process and outcomes (e.g. the renewal of curricula; teaching and learning methods; competencies of graduates).

Values

In the cited *Self-evaluation Report,* compiled in 1999/2000, the HAC identified the most important values it adheres to as follows (values that have already been mentioned are *not* set in italics):

- (V6) independence
- *V9 objectivity*
- *V10 impartiality*
- *V11 transparency*
- *V12 professional rigour*
- (V5) accountability
- (V7) harmonisation of viewpoints of stakeholders
- *V13 collaboration with HEIs*
- *V14 collaboration with the HE Conferences*
- *V15 collaboration with international professional organisations*
- *V16 assistance to HEIs*
- *V17 openness to innovation***.**

These values can be regarded as *operational values*. While in the *Code of Ethics*[23] of the HAC, compiled in November 2000, the following *ethical values* were specified:

- *V18 lawfulness*
- (V6) independence[24]
- (V9) objectivity
- (V10) impartiality
- (V11) transparency
- (V5) accountability *(personal responsibility)*
- (V12) well-preparedness and high professional standards
- (V17) openness to innovation

[23] *http:/www.mab.hu/english/index.htm,* Regulations *section.*

[24] *"Members shall (…) not 'represent' their employer. If a Member is employed by a higher education institution or another organisation, he/she shall fully disassociate himself/herself from such a relationship(s), and shall formulate his/her statements independently of any such commitment."* Code of Ethics*, op.cit. p.1.*

- (V7) harmonisation of the points of view of those involved.

Before proceeding with the analysis it should be mentioned that research into values has uncovered some important features we must bear in mind when discussing the values underpinning the Hungarian accreditation system, and the relationship between values and methods. First of all, values are not necessarily conscious. Therefore, there can be some discrepancies between the stated values of a person or organisation and the values actually (usually subconsciously) influencing or determining that person's or organisation's actions. Moreover, values exist in subject-object reality fields, in concrete realisations; that is, they are always subject bound on the one hand, and situation dependent on the other. (Remember: "the times they are a-changin' …") And finally, as I mentioned earlier, values interact and intertwine; they relate to each other; and through their relationships, they form value structures or value worlds.[25]

From all this, it follows that the above values are not exhaustive, and have varying importance or "weight", exert various influences, stand at different places in the value-world of the HAC.

As it appears from the above, at least four different sets of values can be identified regarding the HAC and the Hungarian accreditation system:

- foundation values,
- stakeholder values,
- operational values, and
- ethical values.

Several values belong to more than one set at the same time. Among these, independence stands out as perhaps the most important stated value in the value world of the HAC, being an element of each set. Accountability and harmonisation of viewpoints of stakeholders both appear in three sets, while objectivity, impartiality, transparency, professional rigour, and openness towards innovation are identified as elements of two sets each.

If we try to trace these values according to their actual influence on the operation of the HAC, the picture is a little bit different from the one above. Independence certainly remains the *dominant* value. On the *second level*, beside accountability (regarding both HEIs and HAC), however, adherence to legal regulations, transparency (meticulous procedures) and central/national supervision (authority-like status and operation) come to the fore. And only on the *third level* comes quality development, objectivity, impartiality, and all the other values, while openness towards innovation, to be honest, is perhaps on a *fourth level*, as it has been more important as a stated value than an effective one.[26]

Of course, one can make something like a chart showing part of these structural relationships (see *Figure 3*), but we must not forget that value structures are dynamic (and more complex than *Figure 3*), changing in time and according to actual situations. There certainly have been HAC decisions – though, fortunately, not too many – where, for example, independence was dominated by other values or external interests. And, as discussed earlier, in the effective (and stated) hierarchy, central/national supervision and quality development seem to change places currently. Similarly, professionalism is coming to the fore.

[25] *The „value world (value structure) is an ordered, though situation dependent, totality of the values (value commitments) of a person" (or group, or organisation). T.R. Szanto, „Value communities in science: The recombinant DNA case". In: Th. Brante et al. (eds),* Controversial Science*. Albany: State University of New York Press, 1993. p.252.*

[26] *See* The External Evaluation of the Hungarian Accreditation Committee*, op.cit., p 108.*

Figure 3
Hierarchies of values of the HAC 1993-2000

Stated	Effective
independence	independence
accountability, harmonisation of viewpoints of stake-holders	accountability, adherence to legal regulations, transparency, central/national supervision
objectivity, impartiality, transparency, professional rigour, openness towards innovation	quality development, objectivity, impartiality, harmonisation of viewpoints of stakeholders etc.
(all other)	openness towards innovation

One more remark to the value structure(s): historically, the four sets, or kinds, of values appeared more or less linearly, beginning (obviously) with foundation values, followed by operational and stakeholder values, while the ethical dimension became apparent and manifest (conscious) only at a later stage of the institutional development of the HAC. In actual operation, however, these sets were completely intertwined and mixed up from the very beginning. Of course, the initial set of (foundation) values greatly influenced the establishment, the genesis and formation of the system. However, there is an opposite direction too: actual activities and operation did and do have an impact on the (intended) value structure, and its manifestations. Not only as far as structure is concerned.

Value structures, value worlds are not only "internally" dynamic. They are not closed, their elements are not given once and for all; they expand, "grow" (and "shrink" at the same time). This can be grasped in the case of the HAC, too. In the outgoing president's report in January 2001 two examples appeared: the quality of life and the short and long term needs of society.[27] Since then these values have gained central (stated) importance as constitutive elements of the mission of the HAC.[28]

Values and method

What can we say now, about the relationship of values and method in the case of the HAC? I must confess that when I was asked to contribute to this anthology and read the proposed framework, hypotheses and questions raised, I thought first that method is much more determined by and dependent upon the object, rather than values of evaluation. My assertion was based on a functional reasoning, similar to the one reflected in this statement: „The functions of systems of quality assurance will determine the layout of the system."[29] Moreover, I stated in my outline for this paper that historical elements and international patterns play important roles in forming the methodological fabric of a national quality assurance system. While, I believe I was right in the latter assumptions, I have, however, on the basis of all the above descriptions and analysis, had second thoughts concerning the relationship of values and method.

Nevertheless, I still think that there is some interdependency between objects and methods of evaluation, but this is, in our case, only at the "second level" of motivations. At the first level there really are values and methodology. The distinction between methodology and methods is, I suggest, very important. By methodology, I mean the general framework of the *How?* of evaluation, while methods present the details, the exact schemes, procedures, and actual ways of assessing a given object.

[27] *A. Róna-Tas, „On the tasks and operation of the HAC". Report to the HAC Plenum, January 26, 2001 (in Hungarian).*

[28] *„The mission of the HAC is to contribute to the improvement of the quality of life of Hungarian society through improving the quality of HE."* The Strategic Plan of the HAC. *(Summary) Budapest, 2002. p. 1. (op.cit.)*

[29] *J.P. Scheele, „Evaluating evaluation in the Netherlands". In:* A Framework for Quality: Evaluating Evaluation. *INQAAHE 5th Biennial Conference, Santiago de Chile, 1999. p.221.*

Thus, values, together with international patterns[30] and historical elements (the "context") do play a role in determining the methodology of evaluation. This is certainly true in the case of evaluation of the higher education sector in Hungary. As it was shown earlier, no matter what the actual object of evaluation is, the general methodological framework is the same: peer-review using self-evaluation or self-description as a starting point. Actual methods applied vary (slightly), according to the object of evaluation, only within the limits of this methodological framework, e.g. site visit vs desk exercise.

Now, exactly which values influenced the basic methodological choice in the formation of the Hungarian system? Some of these values have been presented in the previous sections of this paper. See the foundation values (autonomy plus external supervision plus independence, as basic characteristics of peer-review) and, especially, objectivity, impartiality,[31] and professional rigour from among the operational values. But in the light of the relationship under discussion in this section, a whole bunch of values is still looming for the purpose of analysis: the traditional value set of science, namely, academic values.

Of course, the discussed value sets overlap; they are not at all mutually exclusive. The values enumerated in the previous paragraph can be regarded as part of the academic value set, too. But here, there are more than just these ones. Among others: performance, contribution, (scientific) merits, originality, innovativeness, thoroughness, etc. But for us here, I think, only one small piece is missing for completing the values-methodology puzzle: equality. Yes, the value of equality (think of the "the republic of science" slogan), which is at the core of the concept of "peer", and thus, "peer-review", and which, nevertheless, did not hinder scientists to create delicately stratified communities in the various disciplines. Equality was perhaps the most decisive, guiding value in the background when peer-review was chosen as the methodological framework of the Hungarian higher education accreditation system – even if it has not been manifested explicitly in accreditation documents, strategies and reports. Remember, the original purpose of the establishment of the system was the evaluation of doctoral training or, more exactly (and indirectly), evaluation of professors wishing to launch doctoral training. Who else should evaluate a university professor of high reputation than a "peer", a professor from another university or research institute with corresponding "rank" earned in the regiment of science?

And the other characteristics of methodology and methods also fit into the picture. The weight of external elements, the stress on input indicators, especially the "quality of instructors" as measured by their (formal, academic) degrees and qualifications, and the threshold requirements characteristic of not only foreign accreditation processes, but also the international practice of evaluation of research papers submitted for publication in scientific journals. Editors and referees as "gatekeepers", standing on the *threshold* of acceptance and *licence* for publicity…

Sophisticated multi-dimensional evaluation methods? QA experts? ISO? TQM? EFQM? No thanks. Too complicated; disgracing; not our language (nor our values).

[30] *In the methodological shaping of the Hungarian system the standard-based American accreditation approach, and the internationally wide-spread self-evaluation, site-visit, external report sequence played a central role. While the* Accreditation Guidebook *for the first round of institutional accreditation was written (by* Nóra Halmay*) on the basis of directions by the London-based engineer* Agnes Kaposi*, HAC member in the first term. Peer-review, the methodological core of the Hungarian accreditation system is also an international phenomenon having its roots in the 18th century.*

[31] *"The selection of the papers has been made with as much* impartiality*, and as strict attention to their comparative merits, as could be expected, in decisions of so delicate a nature" (italics added). This sentence can be read in the* Memoirs of the Literary and Philosophical Society of Manchester *and dates back to 1785. Cited in: D.A. Kronick, "Peer Review in 18th-Century Scientific Journalism".* JAMA *263(1990) No. 10. p.1322. (This paper was brought to my attention by* Tobias Lindeberg *[former staff member of EVA], Copenhagen Business School, Denmark.)*

But let us be fair: considering all the circumstances, the traditions, the social-historical context, and existing international patterns, I think peer-review was an adequate methodological choice in Hungary, fitting properly to the academia centred value-world of the founding fathers of the Hungarian accreditation system.[32] That initial choice, however, does not exclude shifts in emphasis and values, gradual changes in the system corresponding to emerging needs, new international trends and "challenges", and the professionalisation of both quality assurance of higher education internationallly[33] and the operation of the Hungarian Accreditation Committee.

Past and future

The approach of this paper has been a dynamic one. Instead of presenting a static picture, I have tried to describe and analyse the Hungarian higher education accreditation system in its genesis and progression, from its establishment through its self-reflection and external evaluation, to the forming of its strategic plan in 2002 and associated follow-up activities. This was something like a "growing-up process", a course of awakening to ever-greater consciousness, at the current stage of which, now, the HAC stands on the threshold of professionalisation. Whether or not it will cross that threshold is yet to be seen.

How was evaluation done 5 years ago? As to *objects*, there were only institutions and programmes, and no evaluation of individuals applying for professorships in 1998. If I had written this paper at that time – I did not then work at the HAC – I would certainly have chosen institutional accreditation for the case description. That was the most important activity of the HAC then. (And, probably, it will be again when the second cycle starts.) But *purpose, methods, and values* were basically the same. Accreditation, i.e. the attestation of quality with control and improvement aspects, the former being dominant; self-evaluation and peer-review, and a science and academe centred value-world in the background.

As I have indicated, slow changes occur with experience gained, lessons learned on the one hand (learning by doing), and the changing context, international interaction and trends (swimming in) on the other. There are some new stated values, shifts in emphasis from control towards improvement, from input to process and outcomes. Saying this, I would not like to imply that the approach of the HAC was "wrong" in the past. No, this is not the case. In the given socio-historical context, I believe the HAC chose the right way. Hungary, thanks to the accreditation system, has no "garage universities" and "diploma mills" of dubious quality, though, of course, the value of the diplomas earned at various institutions are not the same. Never (and nowhere) will they be the same, even if they are regarded as "equivalent". All in all, the system is an organic one, capable of learning and changing in the long run.

If we consider that the Hungarian accreditation system has a history of only 10-11 years, then another 5 year period appears to be quite a long time. While if we think of the relative stability, the "inertia" of well-established systems and organisations, people's thinking, and the strength of the dominant values behind activities and decisions, then this 5 year period seems to be hardly more than a moment. In January 2004, a new HAC term begins, probably about half of the current members will continue while the other half will be changed. That is, elements of both stability and change are present in the system.

[32] *I agree with both* Tomusk *and* Westerheijden *on the major role academics and academic values played in the formation of the accreditation systems in CEE countries (Hungary, in our case). However, I definitely do not share their negative evaluations of this phenomenon. I don't think that academic values are "eroding" (Tomusk, op.cit. p.180.) If this were the case, we would really be in trouble. Think of the cited sentence from 1785: thank God – and the "academic oligarchy" (Westerheijden, op.cit. p.73.) – impartiality as an important value is still with us though, of course, not without occasional cases of less than full adherence to it.*

[33] *D. Woodhouse, „The quality of quality assurance agencies". December, 2002. http://www.inqaahe.nl/public/docs/ThequalityofEQAs.doc*

As to *objects* and *purpose* of evaluation, I do not expect considerable changes in the next five years. Institutions and programmes will probably remain the main targets of evaluation, though new elements of focus may come to the fore. Internal quality assurance of institutions and programmes will certainly be one of these. Another likely change has to do with the necessary transforming of the currently dual Hungarian higher education system to a Bologna-conforming linear one. In line with this process, a reduction in the existing number of study programmes is intended, and "programme groups" are planned to be formed. Students will begin their studies in these wider fields and specialise only in the higher classes. That means that the current system of national qualification requirements will be changed and the HAC will have to give opinion on these new kinds of NQRs.

I do not expect serious changes in the methodological framework applied and the value-world of the HAC, either. Peer-review will certainly remain while academy centredness is changing only slightly, at a very slow pace, with the involvement of more experts from the so-called users' sphere in the work of the HAC.

More changes will take place, however, as far as methods, or the actual implementations of evaluations are concerned. As it looks now, these step-by-step changes on the practical, operational level of methods can slowly give rise to more fundamental changes in the system. That is, it seems that while in the establishment and formation phase of the system values, the given socio-historical context and international patterns and impacts play a constitutive role. At a later, more advanced stage of development, the applied actual methods themselves, together with experience gained during operation, can and will influence the basics, the theoretical-methodological framework and the underlying world of values. But, anyway, in the case of the HAC, for such deep-reaching changes of attitudes and values, I think more time and, perhaps, a new generation of experts are needed.

Of course, beside the "normal", organic evolution ("growing up", learning by doing, stronger self-reflection, professionalisation) there are some – mostly external – factors, which can accelerate changes, exert considerable influence in certain cases on the shape and operation of the system and the HAC:

- evolution (revolution?) of the national system of HE[34]
- state-administration effects (difficult to forecast; almost black box; contingencies, political impacts and embedding)
- internationalising of HE
- international involvement of the HAC
- internationalising of QA of HE (co-operation, networking)
- further effects of the Bologna process.

Of all these factors, I would like to stress here the international involvement of the HAC, and internationalising of QA of higher education. The HAC has had an *International Advisory Board* since the very beginning. Though the major role and membership in the CEE network[35] have not really exerted detectable influence as yet, ENQA and INQAAHE membership certainly did so.[36] The most relevant and effective impact, however, was undoubtedly made by the international panel of experts[37] evaluating the operation of the HAC and the Hungarian accreditation system in 1999/2000. Changes taking place in the last 2-3 years are originating, directly or

[34] *A group of experts commissioned by the minister of education is currently working on the preparation of a completely new HEA. Their proposals will probably touch also the "basics" of the HE system (structure, financing, governance, etc.) which may have major implications for the functions and operation of the HAC too.*

[35] *www.ceenetwork.hu*

[36] *Honorary president* András Róna-Tas *is steering committee member of ENQA, he and HAC staff members regularly take part and present papers on workshops and conferences of ENQA and INQAAHE.*

[37] Alberto Amaral *(chair),* Judith Eaton, Marie-Odile Ottenwaelter, Ulrich Teichler, Christian Thune, Carolyn Campbell, Sami Kanaan.

indirectly, from this exercise. I can only hope that this international involvement and its "returns" will remain with us in the future too.

Conclusions

As to the main theme of this volume, the relationship of values and method, I would like to stress the following, based on the above analysis.

1. It seems to be useful to make a distinction between methodology (principles, framework) and methods (details, procedures).

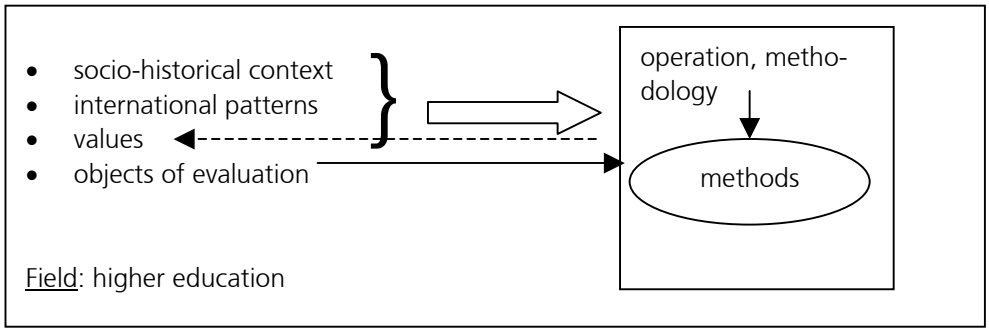
2. Values, together with the socio-historical context and international patterns, greatly influence the establishment, genesis and development, the methodological framework and operation of an evaluation (accreditation) system.

3. However, there is an opposite direction, a "feedback", too: actual activities and operation do have an impact on the state and dynamics of the value-world of the system.

4. Within the given field of investigation namely, higher education, in the case of Hungary, it seems that the objects of evaluation exerted only a "second level" influence on the actual methods applied. (It is to be further investigated whether or not this finding is valid in other countries, in other fields and, especially, between various fields too.)

All this can be illustrated as follows:

Figure 4
Relationship of values, methodology, and methods



The Netherlands – Higher Education

A.I. Vroeijenstijn
Senior Advisor Quality Assurance
Association of the Universities in the Netherlands (VSNU)

Values and purposes

In 1985, the Dutch government changed its policy with regard to Higher Education (HE) from one of detailed regulation to a policy of steering output. The government promised more autonomy to the HE-institutions, based on one condition: quality should be assured. Therefore, evaluation was necessary.

Originally, the idea was that internal evaluation would be carried out by the HE-institutions, and external evaluation would be a task for an outside body, e.g. the Inspectorate. However, the HE-institutions agreed with the minister that both internal and external evaluation was the responsibility of HE. Therefore, the umbrella organisations of the universities (the Association of Universities in the Netherlands (VSNU) and the Association of Universities of Professional Education (the HBO-raad)) started to develop an external quality assessment system. In the beginning, the HBO and VSNU had different approaches. The HBO placed emphasis on the development of an internal quality assurance mechanism in the HE-institutions, planning to add external evaluation to it later. The VSNU started with the development of an external quality assessment system with the idea that it would promote internal quality assurance in the institutions. In later years, the approaches between HBO and VSNU converged, and a system of quality assessment emerged, based on self-assessment, external assessment and public reporting.[1]

Because the HE-institutions took the initiative for external quality assessment and organised their own assessments, an agreement was made between the Minister of Education and the HE-institutions concerning the role of the Inspectorate. The Inspectorate would act as a watchdog, assuring through so-called meta-evaluation that the system was run properly and not becoming an activity of an old boys' network. In a meta-evaluation, the Inspectorate not only checks the quality of the evaluations, but also the quality of the programmes. If the Inspectorate concludes there are serious shortcomings, it gives – in soccer terms – a yellow or red card as warnings that the programme might be closed and the funding stopped if the faculty does not take serious improvement action. The programmes with a yellow or red card come under the special supervision of the Inspectorate.

For the development of the assessment framework and process, the VSNU went shopping around the world to find examples of current best practice. It adapted ideas from the USA, Canada, Australia and the UK to its own national context.

The purposes of the evaluation have been formulated as:

1. a contribution towards quality improvement;
2. ensuring accountability; and
3. providing information.

[1] *The description of external quality assessment in the Netherlands will mainly be based on experiences in the university sector, due to the background of the author. However, most comments also apply to other HE-institutions.*

Because government policy had changed from detailed regulation to output steering, accountability was seen as very important. HE should demonstrate its quality to the outside world: the minister, the parliament, the taxpayer.

HE never disputed the purpose of accountability, because this was seen as the natural consequence of autonomy. However, main function was seen as quality improvement. It became clear that it is not always easy to combine accountability and improvement, as there can be tension between these two aspects[2]. Accountability demands public reporting and the release of information to the outside world, while quality improvement is best served by critical honest discussion behind closed doors, as no institution likes to hang out its dirty laundry in public. During all the years of external quality assessment in the Netherlands, there has been tension.

It is clear that different stakeholders have different ways of looking at external quality assessment. The Minister of Education will stress accountability and use the outcomes of assessment to show parliament that the quality of HE in the Netherlands is still at an acceptable level, even though he is cutting funding every year. HE considers external quality assessment as more of an element within their own internal quality assurance system, offering feedback and recommendations that might lead to improvement.

During the meta-evaluation, the Inspectorate considers how far the public reports of the external committees fulfil the functions of accountability, improvement and providing information. Concerning the aim of providing information, we may wonder whether the Inspectorate always looked at the reports only as a meta-evaluator. Having another task, too (to stay informed about the developments in HE and report on it), the Inspectorate needs information. The transition from detailed regulation towards output steering, changed the information flow between HE and the ministry and to the Inspectorate. Less information became available. It seemed sometimes that the Inspectorate used the meta-evaluation to obtain more information by stressing that the reports should make more information available, although it was not always clear what the connection with quality was.

Concerning the function of "providing information", we see that the outcomes of external assessments are used by others to make league tables and comparisons of programmes in Dutch HE,[3]

Objects

In starting to develop the external quality assessment system, some specific decisions had to be made. The decision about the object of evaluation was not a difficult one. As all public HE-institutions are recognised by law, institutional assessment or institutional audit never was discussed. It was taken for granted that the degree programme should be the target.

For the universities, there has been a short discussion about whether the object of evaluation should include research, too. The decision was made to restrict the evaluation to teaching and learning, based on the following reasons:

- research was already subject to evaluation
- the organisational frameworks of teaching/learning and research were different
- there is no one-to-one relationship between, for example, economic degree programmes and economic research.

There was also a very practical reason: the type of experts needed for the assessment of teaching/learning and those for research are different. Furthermore, there was a general feeling that a combination of research and teaching/learning would again bring the emphasis onto re-

[2] *See A.I. Vroeijenstijn, Improvement and Accountability, navigating between Scylla and Charybdis. Jessica Kingsley publishers 1995*

[3] *Keuzegids Hoger Onderwijs (Consumer guide HE); Elseviers magazine: de beste studies (The best studies)*

search, and education would be again discriminated. In 1993, the VSNU was also placed in charge of the evaluation of research, but this was done via a separate system.[4]

According to the law on HE in force until June 2002, only the degree programmes at the publicly funded HE-institutions had to undertake external evaluation.

The decision about the evaluation object has been made through interaction between HE and government. In Dutch, the wording "program" evaluation is used. It might also be translated as discipline, or subject, evaluation. It is the programme leading to a degree in a certain discipline or subject area that is object of evaluation, e.g. the evaluation of History, Economics, Law, or Physics.

Stakeholders

In the first protocol of external quality evaluation, it is clearly stated that the main client of the external evaluation is the faculty of the university. The external assessment committee reports to the faculty. The committee offers a judgement of the quality, a description of the observed practice and gives recommendations for improvement.

But the evaluations and the public reports also serve additional clients. Because the system is aiming for accountability, the Minister of Education, the parliament and the taxpayers can also be considered as clients. In this sense, the experts do not act as consultants to the faculty, but as accountants, reporting on quality to the outside world.

A special client is the student. The public reports contain a lot of information which students can use in choosing what and where to study. The information is made more accessible to students by the Keuzengids Hoger Onderwijs[5]

In general, there is consensus in the Netherlands concerning which clients are served by, and the different roles of, external quality assessment. The problem arises in the weighting of the different clients. As already stated, there has always been tension between the improvement-orientation (having the faculty as a client) and accountability (where the client is the minister and parliament). In the eyes of the HE-institutions, influencing and enhancing quality is much more important than measuring quality and displaying it to the outside world.

Methods

The method of evaluation in the Netherlands is, according to the generally accepted model:

- self-evaluation by the institution
- external assessment
- public report
- follow-up.

In the law of HE, it only states that an institution has to manage external evaluation of its core activities. There is no legal obligation to adopt a national approach. However, the institutions invited the HBO-raad and the VSNU, respectively, to take the lead and to organise the external evaluations. The invitation was accepted, and the first step was to develop the evaluation framework and to develop a protocol. This was done in close co-operation with all interested parties.

It is either the HBO-raad or VSNU that takes the initiative for an evaluation and decides upon the schedule. The VSNU, for example, has a schedule and evaluation programme for a period of 6 years. For each year, it is known which discipline/subjects will be assessed. It is important to appreciate that the Netherlands has chosen a nation wide assessment approach. This means

[4] *see the VSNU protocol , Assessment of Research Quality 1998 (valid until 2003)*

[5] *see footnote 4*

that one and the same committee will assess all similar programmes at each of the universities (VSNU) or the universities of professional education (HBO-raad), e.g. one committee for history will assess all history programmes at Dutch universities, and one committee for Geography will assess all geography programmes, etc.[6].

Either the VSNU or HBO-raad is responsible for the evaluation method. In the VSNU, the protocol for the external assessment has bee discussed with the universities and finally endorsed by the general board of the VSNU. So far, the department of quality assessment has a steering group, responsible for the process and the outcomes.

Although the universities were the owners of the system, the evaluation method and process were also discussed with the Ministry of Education and the Inspectorate. The latter also has the task of controlling the outcomes of the assessment and seeing that the protocol has been followed. So far, the Inspectorate has refused the outcomes of one assessment of the VSNU.

Responsibility for the assessment and evaluation process is attributed to different actors. The HBO-raad and the VSNU are responsible for the organisation and the process. The external assessment committees are responsible for the judgements and the content of the reports, and are also responsible for the recommendations. The VSNU has no say in the report; it only checks that all topics are treated, but cannot change any part of the judgement. The responsibility of the Inspectorate is to ensure the assessments and the content of the report are both produced independently.

The outcomes of the evaluations are made public. The report is formally sent to the faculty being assessed, but the report is also widely distributed and sent to all stakeholders. The report is also made available on the Website. Beside the public report, the external committee might send a confidential letter to the HE institution management on request.

Follow-up to the external evaluation is seen as the responsibility of the institutions. The HBO-raad and VSNU do not play a role here. According to the latest agreement between the minister, HE and the Inspectorate, it is the Inspectorate that is in charge of the follow up. Two years after the publication of the report, the Inspectorate will visit the institution and discuss what has happened regarding the outcomes of the assessment and the recommendations of the external committee. The Inspectorate publishes their findings.

There is one more check on the follow-up. In the next assessment, the self-evaluation report has to answer the question of what has been done since the recommendations of the external committee, and the new external committee will verify this.

Following the recommendations is not compulsory. Although the recommendations represent the opinions of experts in the field, this does not mean that they are the only truth. As once said by a former chairman of the VSNU: *"You do not need to follow the recommendations, but at least you have to consider them. If you have good reasons for not following the recommendations, then that is acceptable."*

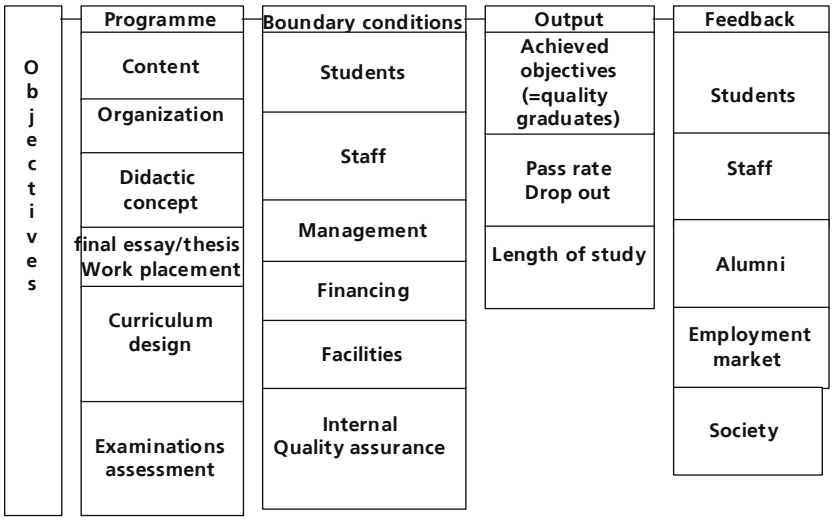
The procedures and the criteria for evaluation are more or less fixed, as all faculties and all programmes have the right to be assessed in an equivalent way. At least for one cycle of assessment, the procedures have to be the same.

Of course there is some variation in the criteria, as the VSNU has always tried to offer tailor-made assessments. At the present time, there is an opportunity for the discipline under evaluation to formulate a specification, for the discipline-adapted evaluation protocol.

[6] *In some cases also programmes from the Flemish universities are included*

Most of the criteria or quality aspects are qualitative, and only a few are quantitative, e.g., the drop-out and pass rates of the students. Generally speaking, external quality assessment is not performance indicator based, but expert based.

The criteria to be applied for the assessment of the quality of a programme, can be illustrated in the following model:



So far, external quality assessment in the Netherlands can be described as process oriented. The philosophy behind it is *fitness for purpose.* The formulated goals and aims of the institution are the starting point for the assessment. The judgement concerns the possibility of achieving the goals and aims. Although the goals and aims are beyond this discussion, questions can be asked concerning input and output quality.

At the moment, with the introduction of accreditation, there is a shift from process assessment towards output assessment. It is no longer only *fitness for purpose*, but also *fitness of purpose*.

Looking at the relation between the purpose of evaluation and the criteria used, one has to make a clear distinction between "how" evaluation is done and "why". Evaluation means the judgement of quality, using a generally accepted method and generally accepted criteria. The outcomes of the evaluation may be used in different ways, perhaps for enhancement, accountability or accreditation, as will be the case in the future. There is no relationship between the value of evaluation and the method used. There are only two types of evaluation: good or bad.

Past and future

The evaluations done 5 years ago, had in broad terms, the same features as described in the case chapter. Keywords were "fitness for purpose" and "improvement". Accreditation was not yet a topic of discussion. The Netherlands had a stable quality assurance system. Of course, the system was adapted to the changing surroundings. Looking at the differences between 5 years ago and nowadays (before the introduction of accreditation), the following points may be highlighted:

- The protocols in use for self-assessment and external evaluation were more elaborate than the protocols that came into use in 1999. The guidelines, nowadays, are less detailed. For the sake of the self-assessment, it offered the faculty more freedom to analyse aspects. However, it made it more difficult to keep all the committees on a correct and similar track.
- The emphasis was on the process (fitness for purpose) and the contents of the programme, although attention was also paid to learning outcomes.
- The assessments were less tailor-made than the 3rd cycle of assessments that started in 1999.

One thing is for sure; five years from now, the Netherlands will have a different system of evaluation, if any system at all survives. While the protocol of 1986 stated that the system of external quality assessment was not aiming at accreditation, the Netherlands has now an accreditation system, based on national legislation.

The Netherlands had an effective system for external quality assurance that also enjoyed international recognition. However, national and international developments made it necessary that the system of external quality assurance be rounded off by the introduction of accreditation, by an accrediting body.

The discussion started in 1998, when the VSNU pleaded for a change to the informal accreditation system of the Inspectorate. Because the criteria for the yellow and red cards were not always clear, the VSNU saw it as important to set up an independent validation council. However, at that moment, it was not yet the right time to establish such a council.

In the HE policy plan of 2000, the Minister of Education introduced the discussion of international accreditation, based on the Bologna declaration. All stakeholders in Dutch HE agreed, and the conclusions were sent to the parliament in the policy document "Keur aan Kwaliteit" ("Accreditation in Dutch HE "):[7]

The reasons for the introduction of accreditation were, as indicated in the above mentioned policy document:

- International recognition for Dutch HE
- Encouraging international benchmarking
- Encouraging transparency of the quality of the programmes on offer
- Reinforcing the independence of quality assessment
- Clarification of the consequences for management in case of lack of quality

In June 2002, the law on HE was accepted, and the Netherlands accreditation organisation (NAO) was established. The tasks of the NAO are:

- Verify and validate external assessment and grant accreditation to existing programmes
- Assess the quality of new programmes
- Contribute to the introduction of the Bachelor-Master Structure in Dutch HE.

Accreditation is defined as, *"Providing a quality label as proof that certain requirements are met."* The quality label will be provided after validation of an external assessment, executed by a Quality Assessment agency on the request of an institution. Accreditation is thus seen as the finishing touch of the quality assurance system. Accreditation builds upon external evaluation. Without external evaluation, no accreditation.

The Netherlands has opted for programme accreditation. Objects of accreditation are all Bachelor and Master programmes at state funded HE-institutions and other institutions of HE that are recognized by the minister.

There is a clear distinction between accreditation of existing programmes and the licensing/testing of new programmes.

The NAO does not itself organise external assessments of existing programmes. This is done by the QA-agencies, such as the departments of External Quality Assessment of the VSNU and the HBO-raad. However, the QA-agencies have to change their organisational structure because of the requirements set by the NAO. The market for external assessment is open to both national and foreign assessors. The expectation is that at least 5 Dutch agencies will act as organisers of

[7] *Ministry of Education, Science and Culture, Accreditation in Dutch HE, July 2000.*

external assessments. The NAO will publish a list of QA-agencies, meeting the requirements set by the NAO. Those requirements concern:

- The independence of the QA-agency
- Composition of the expert teams. The committees should be independent, should have adequate expertise and should command authority.

The QA-agencies have to show how they select committee members, and how they assess authority and expertise. Also, how they assure independence and how they prevent conflicts of interest. If possible, the QA-agency will involve foreign experts, as is already done.

The NAO has formulated 2 frameworks: one for the accreditation of existing programmes and the other for licensing new programmes. The frameworks had to be endorsed by the Parliament. The Parliament endorsed the frameworks in May 2003.

The QA agencies organising the external assessments of existing programmes are expected, as a minimum, to act in accordance to the NAO-framework.

In the future, the process for accreditation will be:

For existing programmes:

- A HE-institution applies to the NAO for accreditation
- The institution looks for a QA-agency to carry out an external assessment
- The institution carries out a self-assessment and produces a self-assessment report
- The QA-agency organises an external expert committee, which visits the institution
- The external committee produces a report; the QA-agency forwards this to the institution
- The institution sends it to the NAO
- The NAO verifies and validates the report.
- If verification and validation are satisfactory, the NAO accepts the final judgements of the committee and provides accreditation (or withholds it)
- The outcomes are made public.

For new programmes:

- An institution applies to the NAO for licensing
- The institution provides an information dossier according to the NAO-rules
- If needed, the NAO organises an external assessment
- Verifying and validating of the report
- If verification and validation are satisfactory, the NAO provides accreditation to the new programme (or withholds it)
- The outcomes are made public.

The external evaluation of existing programmes will de done by a QA-agency and not by the NAO; the external evaluation of new programmes will be done by the NAO.

Accreditation is compulsory for all publicly funded HE and institutions recognised by the Minister of Education, Culture and Science. The benefits of being accredited, for the publicly funded institutions are:

- Funding by the state
- Recognition of diplomas
- Students get student loan.

The accreditation is valid for 6 years.

The introduction of the accreditation system and the establishment of the NAO will have great influence on the current quality assurance system. Although it is said that accreditation is the finishing touch to the quality assurance system, there is some doubts about the real effects. With the strong role the NAO will play, it might become a control mechanism. Will it be possible to combine accreditation and enhancement? Will the HE-institutions only go for getting the quality label and no longer invest in enhancement? Will the NAO become a quality stamp machine?

One of the strong points of the current system is the ownership of the system by HE itself. It did contribute to improvement and enhancement. Will it be replaced by a bureaucratic control system? For the time being, the NAO deserve the benefit of the doubt. A lot will depend on the way the NAO will function in the coming years. All stakeholders agree upon the need for a formal quality label in the Open European HE-area, but nobody is waiting for an extra control mechanism.

Even if the accreditation system indeed develops as the final touch to external quality assurance, the approach of the external assessment has to change. The current quality assessment system is not yet suited for accreditation. Until now, the main purpose has been quality improvement and accountability. The protocols of the HBO-raad and the VSNU will have to be modified to be consistent with the new developments in accreditation. The protocols will have to be matched to the accreditation framework. In any event, the following changes will have to be made:

- *In addition to process orientation, more attention to output orientation.*

Until now, attention has primarily been on the quality of the process, although the product quality also received some attention and the level of the graduates was not entirely neglected. There will, however, have to be more explicit attention paid to the qualifications gained and the standards of education.

- *A greater emphasis on determining the terms of reference.*

At the beginning of the external assessment, the terms of reference for the assessment will have to be clearly and explicitly formulated, more than is presently the case.

- *An overall judgement of the study programme.*

Until now, the external committees have formed opinions of a number of aspects. In order to lead to accreditation, the committee also have to express an overall judgement.

- *A closer monitoring of the independence of the assessments.*

The safeguards for the independence of judgements will have to be made more explicit by the QA-agency than before.

For the organization of external quality assessment by the HBO-raad and VSNU, the introduction of accreditation means the development of a protocol that includes all the conditions for external assessment mentioned in the guidelines for QA-agencies, as published by the NAO. The assessment protocol of the HBO-raad and VSNU should have the assent of the NAO, so that the HBO-raad and VSNU can be registered, and the institutions can have the guarantee that the external assessment reports can be submitted for validation to the NAO.

The years to come will be important. The legislation on accreditation and the NAO-frameworks have been accepted. However, the real work has yet to begin. It is important to build upon the current system, but also to make it as non-bureaucratic as possible. All stakeholders must support it. If the NAO does not have the support of the HE-institutions, and if accreditation is perceived as something "from above", the system is doomed to failure – and the destruction of all the benefits of 14 years of external quality assessment.

New Zealand – Higher Education

John M. Jennings[1]
Director
New Zealand Universities Academic Audit Unit

Values and purposes

The purposes of evaluation within New Zealand higher education, for both quality assurance and quality development, are:

- to protect the interests of learners
- to ensure learners have access to opportunities for life-long learning
- to ensure available qualifications are meaningful and credible
- to assure learners that courses and programmes are well taught
- to ensure qualifications are obtained in safe environments using appropriate teaching and assessment systems
- to contribute to the enhancement of quality systems and processes that improve the quality of research, teaching, learning and community service.

The Education Act 1989 invests authority for accreditation, programme approval and institutional audit in the *New Zealand Qualifications Authority* for all institutions other than universities, and in the *New Zealand Vice-Chancellors Committee* for universities.

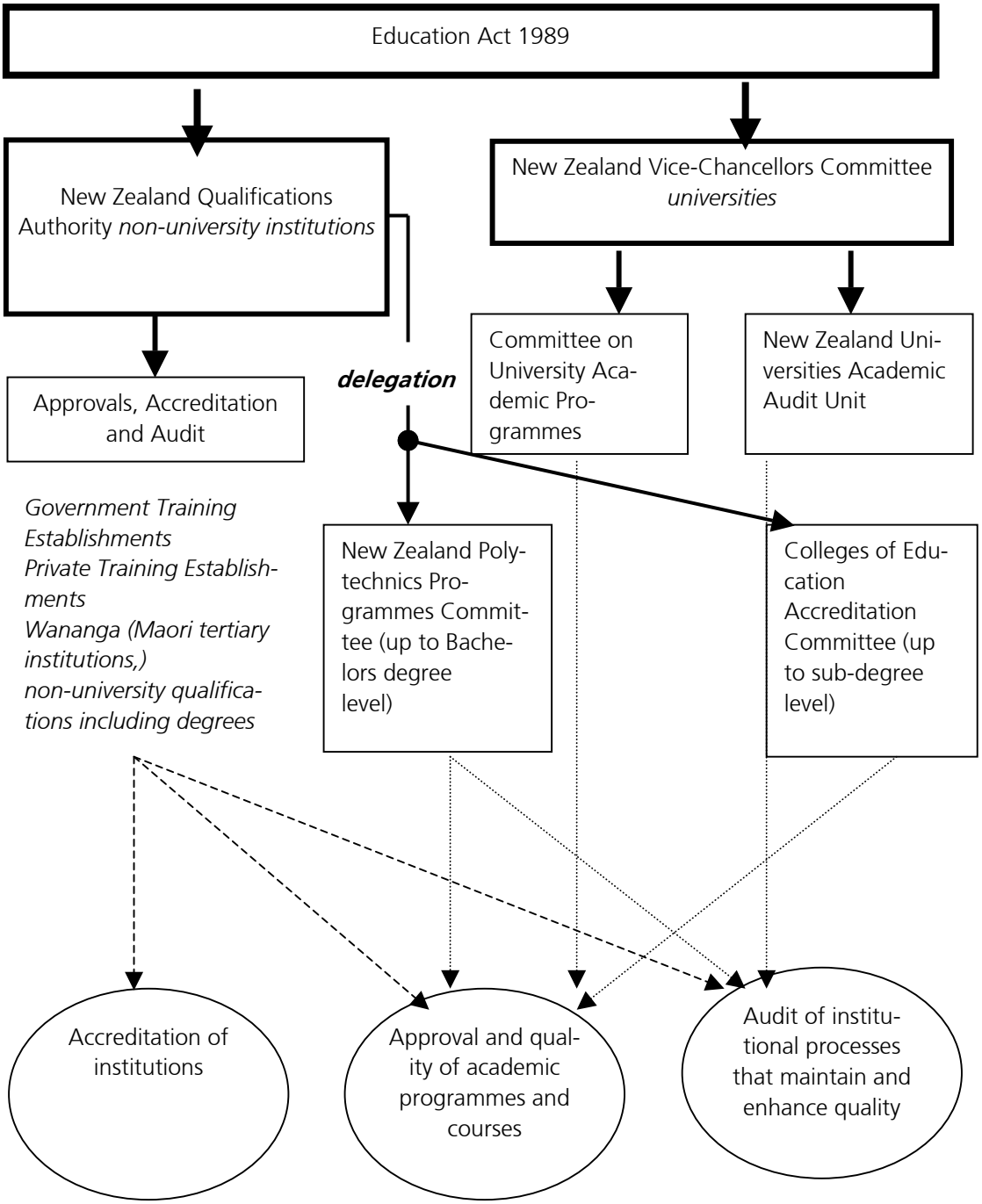
The **New Zealand Qualifications Authority** is a statutory body set up by government as a result of the Education Act 1989. It has an overarching responsibility for quality assurance in secondary and tertiary education providers other than universities. Its *Approvals, Accreditation and Audit* section quality assures courses and qualifications offered by government and private training establishments and wananga (Maori tertiary providers), and degree and related qualifications offered by polytechnics and colleges of education. The Qualifications Authority has delegated quality assurance of sub-degree qualifications to the *New Zealand Polytechnics Programmes Committee* (a committee of the Association of Polytechnics of New Zealand) in the case of polytechnics, and to the Colleges of Education Accreditation Committee (a Committee of the Association of Colleges of Education of New Zealand) in the case of independent teacher training institutions. From July 2003, the delegation to the New Zealand Polytechnics Programmes Committee is to be extended to include Bachelor degrees.

The **New Zealand Vice-Chancellors Committee** is a statutory body which was given statutory status by the government in 1962, at the same time that the federal University of New Zealand was replaced by separate autonomous degree-granting institutions. With the Education Act 1989, the Committee assumed some of the functions of the former New Zealand University Grants Committee at the interface between government and the universities, and was given the responsibility for quality assurance in all universities. To do this, the New Zealand Vice-Chancellors Committee has a standing committee - the *Committee on University Academic Programmes* - which approves new programmes and significant changes to programmes and monitors their implementation. In 1993, the New Zealand Vice-Chancellors set up an independent evaluation agency - the *New Zealand Universities Academic Audit Unit* - to audit uni-

[1] *Acknowledgement is made of the assistance of Michael Steer and Peter Osborn, Approvals, Accreditation and Audit section, New Zealand Qualification Authority*

versities' systems for monitoring and enhancing quality and to disseminate and commend good practice.

The authority for quality assurance in New Zealand higher education



The normal terminology used in New Zealand, and generally accepted in the education sector is accreditation, approval and audit.

- *Evaluation* would be understood to be audit of systems for monitoring academic quality, to assure the quality of core activities (research, teaching, learning, community service) and to assist in the enhancement of those activities.
- *Quality development* would be understood to be the identification, development and implementation of institutional systems and processes to enhance the quality of an institution's core activities.
- *Quality assurance* would be understood to incorporate accreditation, approval and audit, where accreditation is understood to refer to the accreditation and registration of providers to deliver approved programmes and courses of study leading to qualifications, ap-

proval refers to the approval of the programmes and courses leading to qualifications, and audit refers to the evaluation of the effectiveness of institutional systems for monitoring and enhancing academic quality.

Objects

Before government subsidy can be given to providers of higher education who seek to offer approved courses of study leading to qualifications, the providers themselves must be registered, the courses they wish to offer must be approved, and the providers must be accredited to deliver them.

The *New Zealand Qualifications Authority* approves the courses offered by non-university providers and accredits the organisation to deliver them. In 2001, there were 490 non-university institutions delivering government subsidised courses and programmes to 161,803 students:[2]

Institution type	*Number of institutions*	*Student Enrolment*
Tertiary Education Institutions (public):		
polytechnics	21	87,965
colleges of education	4	10,894
wananga (Maori tertiary providers)	3	11,278
Private Training Establishments (private):		
institutions receiving government subsidy	462	51,666

A large number (over 450) of private training establishments offer courses that are not directly funded by the government.

The *Committee on University Academic Programmes* of the New Zealand Vice-Chancellors Committee approves all new programmes of study in universities leading to new qualifications and all major changes to programmes of study in existing qualifications; the *New Zealand Universities Academic Audit Unit* evaluates the quality assurance processes in all universities including the processes associated with the institutional approval of minor changes to programmes of study.

Institution type	*Number of institutions*	*Student Enrolment*
Tertiary Education Institutions (public):		
universities	8	125 668

With respect to the ongoing monitoring of the quality of degree programme delivery and maintenance of standards in non-university institutions, the *New Zealand Qualifications Authority* appoints a Monitor nominated by the approval panel for each new qualification as each has come on stream from 1990. Each Monitor conducts a desktop paper audit and site inspection to ensure the institution implements fairly and adequately the quality management systems and assessment moderation procedures that had been agreed to at the time of the approval of the qualification.

In 2000, after almost ten years of operation in quality assurance, the Qualifications Authority supplemented what it described as a 'front-end check of a provider's policies and their capacity to deliver what they claim they will deliver'[3] with an external quality audit process. This requires self-evaluation by providers, takes account of the advances made by providers in quality assur-

[2] *Statistics taken from Ministry of Education, New Zealand's tertiary education sector: profile and trends 2001, Wellington, Ministry of Education, 2002.*

[3] *OnQ,* [New Zealand Qualifications Authority] *June 2000, p.2*

ance systems, recognises the increasing diversity among providers, and gives providers a greater ownership of the process. With over 900 private training establishments, the audit process was planned to be phased in over a three year period and, at the present time, almost all providers have been audited at least once, and some, where concerns were identified, as many as four times. The ongoing review of over 1100 quality audits has informed the revision of the Quality Assurance Standard for Private Training Establishments, Government Training Establishments and Wananga to enhance its effectiveness.

The most common response of providers to the audit process has been to redevelop their own quality assurance processes so that they are able to more clearly demonstrate that the organisation is meeting the needs of learners and other stakeholders, and is complying with the specific requirements of the Quality Assurance Standard. In general, at this early stage of the implementation of the audit process, organisations have welcomed the opportunity to carry out this revision, and report that they are in a better position to ensure the quality of their operations.

With respect to the ongoing monitoring of the quality of programme delivery and maintenance of standards in universities, the *New Zealand Universities Academic Audit Unit* has completed two cycles of audit. Cycle 1 audits, in 1995-1998, were fully institutional, auditing systems and processes associated with all aspects of an institution's activities. Arising from a review of Cycle 1 audits, it was agreed that Cycle 2 audits would be the first of three theme audits that were to be smaller-scale tightly prescribed audits. The first of the three themes was to be on research matters, the second on staff matters, and the third on academic programmes. Cycle 2 audits - on the national theme of research policy and management, support for research students and the research-teaching nexus, and on a topic nominated by the institution - were carried out within a tighter timeframe of 2000 and 2001. In spite of the intention, the university workload of Cycle 2 was as great as the fully institutional audit of Cycle 1, and the theme concept has been discontinued for the present. Cycle 3 audits, which begin in 2003, will have a focus on teaching quality, the delivery of programmes, and the achievement of learning outcomes, and will be carried out over four years.

Stakeholders

Evaluation is used as a mechanism to assist institutions in their review of quality and to assure stakeholders - particularly students but also government agencies, the professions, business and industry - of the maintenance of quality in the core activities of research, teaching, learning and community service. Evaluation is also designed to contribute to the enhancement of these core activities and to the learning environment in which these activities take place, in order to better equip graduates to add value to society. Evaluations, therefore, are directed primarily at the institution's management, academic staff and students, and the evaluation processes are designed to involve those sectors as well as graduates, the professions, business and industry.

The primary indicator that these reports are written for the institutions themselves is the inclusion - and clear presentation - of requirements and recommendations for improvement. The procedures require follow-up by the institution to the audit agency; while not regarded as public, the Official Information Act 1982 would require an institution to provide a copy of a follow-up report on request from a member of the public.

Audit reports on non-university providers prepared by the New Zealand Qualifications Authority have not been released as public documents, but are available on request to individuals who have valid reasons for requesting them. Individuals requesting a report are referred, in the first instance, to the provider concerned. Where the provider refuses the request, the Qualifications Authority will make a copy available.

Audit reports on universities prepared by the New Zealand Universities Academic Audit Unit are released as public documents and, therefore, they are available to anyone for consideration, and aspects of the reports are often reported by the media. Copies of the reports are sent to a wide range of stakeholders, in particular to government educational and audit agencies.

Methods

The *New Zealand Qualifications Authority's Approvals, Accreditation and Audit* section conducts evaluations:

- at the time an organisation is registered
- each time an organisation applies for course approval and accreditation
- at the time of scheduled quality audits that occur at intervals of between one and three years or more regularly if the Authority is concerned about aspects of the institution's systems and/or quality of teaching and learning environment
- at the time of monitoring visits for degrees and related qualifications.

The *New Zealand Polytechnic Programmes Committee* conducts evaluations:

- each time a polytechnic applies for course approval and accreditation beyond its current scope of accreditation
- at the time of scheduled quality audits that are conducted at intervals of up to four years.

The *Colleges of Education Accreditation Committee* conducts evaluations:

- each time a College of Education applies for course approval and accreditation
- at the time of monitoring visits.

The *New Zealand Universities Academic Audit Unit* evaluations of all universities are conducted according to a timetable agreed to by the universities.

The criteria against which institutions are evaluated for accreditation, and against which courses are evaluated for approval, have been determined by the New Zealand Qualifications Authority in consultation with the New Zealand Vice-Chancellors Committee and have been published by the Government in *The New Zealand Gazette*. The criteria are reviewed from time to time. The quality assurance agencies that undertake the work - the New Zealand Qualifications Authority (and any quality assurance agencies with delegated powers from the Qualifications Authority) and the universities' Committee on University Academic Programmes – have developed their own procedures to operationalise the criteria in ways best suited to the institutions they evaluate. It is the responsibility of quality assurance agencies to ensure proper implementation of the criteria.

The audit evaluations of institutions' quality assurance and quality enhancement systems are carried out by the New Zealand Qualifications Authority and the New Zealand Universities Academic Audit; each has developed its own academic audit processes.

Assessments in the evaluations are undertaken by independent panels appointed by the quality assurance agencies. The panels are responsible for any requirements and/or recommendations that arise from the process. Evaluation panels are of various types depending on purpose.

The Approvals, Accreditation and Audit section of the *New Zealand Qualifications Authority* has responsibility for the evaluation of providers seeking registration, course approval or accreditation.

- Applications for accreditation as a registered provider are evaluated by one of 17 trained and qualified Quality Auditors who assess the application against a clearly defined standard – the Quality Assurance Standard for Private Training Establishments, Government Training Establishments and Wananga.
- Applications for course approval and accreditation are evaluated by a panel consisting of a Quality Auditor and specialists in the subject area(s) of the course. There is no prescribed number of panellists for non-degree courses.

- Applications for degree qualifications are evaluated by a panel of eight members, comprising:
  - two representatives from the universities,
  - two industry representatives,
  - a representative from another tertiary provider offering a similar course,
  - a representative from the applying organisation,
  - a panel chair,
  - a quality systems evaluator.

The New Zealand Qualifications Authority's Quality Auditors conduct audits of providers against the criteria defined in the Quality Assurance Standard for Private Training Establishments, Government Training Establishments and Wananga. Most commonly, only one Quality Auditor is involved, but larger providers may require assisting auditors.

The New Zealand Polytechnic Programmes Committee - under delegation from the New Zealand Qualifications Authority - conducts the quality audits of polytechnics with a panel of four, comprising:

- a New Zealand Polytechnic Programmes Committee Lead Auditor,
- a representative from the polytechnics,
- a representative from the industrial sector,
- a representative from the New Zealand Qualifications Authority (where the polytechnic offers degrees).

The *New Zealand Universities Academic Audit Unit* evaluation panels comprise:

- two or three New Zealand academics,
- one New Zealand-resident non-academic (from business or industry),
- one academic or former academic from outside of New Zealand (usually from Australia),
- the Director of the Unit.

Evaluations of accreditation of providers and approval of new courses are not made public; audit reports, on the other hand, are, or soon will be, public documents in all cases.

The institutions that have been audited are required to present follow-up reports to the quality assurance agency in accordance with an agreed timetable. The New Zealand Qualifications Authority sets down requirements as well as recommendations. Requirements must be satisfied before a new programme is approved, and requirements in an audit must be satisfied by a certain date if accreditation and registration are to continue.

Evaluations for accreditation and approval of courses are undertaken by quality agencies which develop their own procedures to operationalise criteria as published in *The New Zealand Gazette*. The current criteria, gazetted in December 2002, are:

- *Title, aims, learning outcomes and coherence*. The adequacy and appropriateness of the title, aims, stated learning outcomes and coherence of the whole course.
- *Delivery and learning methods*. The adequacy and appropriateness of delivery and learning methods, for all modes of delivery, given the stated learning outcomes.
- *Assessment.* The adequacy of the means of ensuring that assessment procedures are fair, valid, consistent and appropriate, given the stated learning outcomes.
- *Acceptability of the course*. The acceptability of the proposed course to the relevant academic, industrial, professional and other communities, in terms of its stated aims and learning outcomes, nomenclature, content and structure.
- *Regulations.* The adequacy and appropriateness of the regulations that specify requirements for admission, credit for previous study, recognition of prior learning, course length and structure, integration of practical/work-based components, assessment procedures, and normal progression within a course.

- *Resources.* The capacity of the organisation to support sustained delivery of the course, in all delivery modes, with regard to appropriate academic staffing, teaching facilities, physical resources and support services.
- *Evaluation and review.* The adequacy and effectiveness of the provision for evaluation and review of courses for monitoring the ongoing relevance of learning outcomes, course delivery and course standards; for reviewing course regulations and content; for monitoring improvement following evaluation and review; and for determining whether the course shall continue to be offered.
- *Research* (for degrees and postgraduate qualifications). The adequacy of provision of research facilities and support of staff involved in research; the levels of research activity of staff involved in the course and of ways in which the research-teaching links are made in the curriculum.

In the case of audit evaluations, the New Zealand Qualifications Authority evaluates against its own Quality Assurance Standard for Private Training Establishments, Government Training Establishments and Wananga, which requires consideration of the quality of the following:

- Goals and objectives - measurable goals and objectives for education and training,
- Systems to achieve goals and objectives - the implementation of quality management systems to achieve goals and objectives by showing evidence of:
  - adequate and appropriate governance and management,
  - the recruitment, management and development of its people,
  - the provision of adequate and appropriate physical and learning resources,
  - the supply of adequate and appropriate information, entry and support services to learners,
  - the adequate and appropriate design, development, delivery and review of its education and training programmes,
  - the adequate and appropriate systems of assessment and moderation for assessing learners,
  - adequate and appropriate reports on learner achievement,
  - adequate and appropriate means of research where degree programmes are offered
- The achievement of goals and objectives - evidence that it is achieving its goals and objectives and can provide assurance that it will continue to do so.

The Quality Assurance Standard for Private Training Establishments, Government Training Establishments and Wananga was first approved by the New Zealand Qualifications Authority Board in October 2000 and has undergone two revisions since then.

The New Zealand Universities Academic Audit Unit has evolved its own audit manual through three editions between 1995 and 1998, requiring consideration of the quality of the following:

- structure and planning,
- quality management system,
- the Treaty of Waitangi,
- staff matters,
- courses and programmes,
- teaching, learning and assessment,
- research and teaching,
- reviews,
- feedback processes,
- joint, franchised and external programmes,
- support for students,
- facilities and resources.

For the Cycle 2 theme audit on a national theme of research policy and management, support for research students, the research-teaching nexus, and institutional themes agreed to by the Unit, a supplement to the manual was produced detailing detailed expectations with respect to those areas. The December 2002 edition has further refined earlier editions to take account of the improvement to quality systems and infrastructure in universities over the past decade. New Zealand Qualifications Authority and New Zealand Polytechnics Programmes Committee audits provide a qualitative evaluation of how well a provider is meeting the relevant quality standards. Evidence that supports the evaluation may be quantitative or qualitative in nature. Evaluations are conducted on the processes an organisation has in place to support its delivery, and also on the effectiveness of the implementation of those processes.

The New Zealand Universities Academic Audit Unit evaluations are process-oriented but results may well lead to questions about process. For example, cases where qualitative data related to the outputs in areas of the audit suggest a trend (negative or positive). Such investigation would determine either if there are recommendations to make which will enhance processes so as to enhance outputs, or if there are commendations to make in recognition of good practices that are resulting in excellent results. Evaluation panels welcome evidence that demonstrates a qualitative evaluation of outputs as well as an appropriate benchmarking of results against international standards appropriate to the area under review.

The reasons and purposes of audits conducted by the New Zealand Qualifications Authority and the New Zealand Universities Academic Audit Unit are to assure quality and to assist the enhancement of the quality of processes that will lead to the enhancement of the quality of an institution's core activities - research (for degree-granting institutions), teaching, learning and community service. To emphasise this, both agencies expect institutions to take responsibility for the quality of their own education and training programmes. An institutional audit in New Zealand requires each institution to conduct a self-review and produce a portfolio which provides an assessment of the effectiveness of its quality assurance systems relevant to the audit undertaken, and which identifies areas where improvement is needed and which develops action plans for improvement.

The New Zealand Qualifications Authority's Quality Assurance Standard for Private Training Establishments, Government Training Establishments and Wananga is quite specific in its expectations of the self-review. Providers must show that they are reviewing and evaluating all areas of their operation on a regular basis. The New Zealand Universities Academic Audit permitted each university in Cycles 1 and 2 audits to develop the self-review portfolio in a way that the institution considered appropriate to its needs; for Cycle 3, which is just beginning, the self-review portfolio framework is more carefully prescribed with more information indicating the university’s own planned programme of quality improvement arising from the self-review.

For each institution, an independent audit panel appointed by the quality assurance agency makes an on-site visit, analyses documents and conducts interviews with management, academic staff and learners to verify the institution's performance and management processes for achieving quality learning, and its success as an education organisation. The audit panel prepares a report which summarises the panel's findings and sets out its views as to how the institution can improve its quality systems in support of its core activities, and the institution is required to provide a follow-up report on actions taken. At every stage of an evaluation, the purposes of the audit are kept in focus so that the relationship between the values of the evaluations and the method of the evaluations is as strong as possible.

Past and future

Quality evaluation of tertiary institutions in New Zealand five years ago was driven by a requirement to be 'accountable' to stakeholders - the government, students and others who contribute to the funding of the system and to employers who use the services of graduates - in the sense of assuring high quality teaching and learning best suited to the kind of tertiary institution. Nevertheless, audit evaluations have always been undertaken with the purpose of assisting with the enhancement of systems – that is, the improvement of systems in place or

the introduction of new, innovative or more effective systems – which will assist in the improvement of the quality of the core activities of research, teaching and learning, and community service.

Since 1995, the *New Zealand Qualifications Authority* has quality assured non-university degree programmes through processes of annual visits by Authority-appointed monitors, and annual programme reports. Monitors are appointed as part of the initial approval and accreditation process. Their role is to report to the New Zealand Qualifications Authority on the operation of the degree programme and whether it is being implemented as originally planned and approved. The annual programme reports outline the review and evaluation processes that the organisation has applied to the programme to continually enhance the delivery. Prior to the introduction of the quality audit process in 2000, the New Zealand Qualifications Authority quality assured providers at an organisation level through an annual review of their registration. Under this process, providers were required to supply documented evidence to show that they were still meeting the requirements of registration. This evaluation was document based and rarely involved visits to providers.

With respect to the *New Zealand Universities Academic Audit Unit*, evaluations have always been against the university's own objectives. Universities were given freedom in determining the design and implementation of its self-review process, and for the format and content of its self-review portfolio. (Today, the structure of the portfolio is more prescribed, requiring details of the university's own proposed enhancement activities.) While acknowledging the university's own initiatives, audit report recommendations provided advice for enhancement as perceived by the audit panel, but the recommendations did not need to take account of, or build on, the university's own initiatives for quality improvement. (Today, the audit report recommendations will take account of, and build on, the university's own enhancement activities.) Each university was expected to report on follow-up action on the audit report's recommendations only.

The maintenance and enhancement of the quality of core activities will still be the focus of evaluations five years from now. The context in which evaluations are undertaken, however, will have changed, and there will be considerable interest in the nature of the measures of quality of outputs that might be developed by funding agencies.

In 2002, the New Zealand Government released its tertiary education strategy[4], which lays down six strategies:

- strengthen system capability and quality
- contribute to the achievement of Maori development aspirations
- raise foundation skills so that all people can participate in our knowledge society
- develop the skills New Zealanders need for our knowledge society
- educate for Pacific peoples' development and success
- strengthen research, knowledge creation and uptake for our knowledge society.

The change messages contained in the Ministry of Education's statement of tertiary education priorities[5] include the need for greater alignment with national goals; stronger linkages with business and other external stakeholders; effective partnership arrangements with Maori communities; increased responsiveness to the needs of, and wider access for, learners; more future-focussed strategies; improved global linkages; greater collaboration and rationalisation within the system; increased quality, performance, effectiveness, efficiency and transparency; and a culture of optimism and creativity.

[4] Ministry of Education, *Tertiary education strategy 2002/07,* Wellington, The Ministry, 2002.

[5] Ministry of Education, *Interim statement of tertiary education priorities 2002/07,* Wellington, The Ministry, 2002.

The Government now requires the development of improved charters and profiles in which a *charter* describes the characteristics of an institution through statements of mission, role and medium- to long-term plans, and a *profile* sets out the means to reach these ends through operating plans, key policies and proposed activities for the next three years. The charters and profiles will be evaluated by the new Tertiary Education Commission (which came into existence in 2003) against the criteria of excellence (of core activities), relevance (in particular to the Government's tertiary education strategy), access (for all qualified people) and capability (of institutions to deliver programmes that are excellent and relevant).

It might be expected that the Government's demand for institutions to take a more strategic approach to planning will change the character of institutions' outputs and outcomes. If the funding decisions of the Tertiary Education Commission demonstrate a focus on the extent to which institutions comply with Government directives, then the way the Commission measures outputs and outcomes becomes important. Additionally, in this new environment, the way quality evaluation agencies assess the quality of those outputs and outcomes will require proper tests of quality - not just measures such as completion rates, work placements after graduation, attraction of non-governmental research funding – and the best information about quality will be required. The work of quality assurance agencies in their quality enhancement/quality development activities that support the enhancement of the quality of institutions' core activities will increase in importance. The evaluation of institutions against their own objectives will remain the focus, and the Government's promises of steering the tertiary sector in accordance with the Government's strategy will create challenges, particularly for universities with their responsibilities to the world of research and scholarship as well as to national stakeholders.

Part Three - Cases

Denmark

Dorte Kristoffersen
Deputy Director
The Danish Evaluation Institute (EVA)

Programme evaluation within higher education

Evaluation of a higher education programme is one of the most frequent forms of evaluation, and the type of evaluation with which EVA is most experienced. The evaluation model applied in this type of evaluation is typical of the evaluations carried out by EVA, and the evaluation of the ´bachelor of social work` and ´social communicator diploma` programmes from 2000 is, therefore, presented as a typical evaluation.

This evaluation comprised all social work/communicator programmes in Denmark. The first step in the evaluation was the establishment of a team of evaluation officers from EVA, which was responsible for the practical and methodological aspects of the evaluation. The evaluation officers prepared the terms of reference for the evaluation on the basis of a preliminary study, involving an analysis of the previous evaluation of 1994, ministerial orders and other relevant information. The preliminary study also involved meetings with representatives of the institutions running the programmes. The terms of reference were presented to EVA's board for approval in March 2000. Following their approval, the board was not involved in the evaluation process until the final report was presented to them in May 2001.

The 6 institutions running the programme were informed about the evaluation at an early stage. They received the terms of reference, the time schedule and a list of the members of the evaluation group. They were also informed about their role in the evaluation. This information was sent to them in writing, and bilateral meetings were held between EVA and the programmes to inform the latter in more detail about the evaluation. The programmes were consulted about potential inability/incompetence of the members of the evaluation group before it was formally established.

The evaluation group consisted of four members: a director general of a relevant research institute; two representatives of the labour market for social workers/communicators; and a professor from another Nordic country. EVA appointed a chairman for the group. The members of the evaluation group were all paid a fee for their participation in the evaluation. As is always the case, there was a clear division of labour between the evaluation group and the evaluation officers. The evaluation officers functioned as the secretariat for the evaluation group, i.e. prepared meetings, conducted the preliminary analysis of the self-evaluations and other documentation, planned the site visits and drafted the final evaluation report.

When the evaluation group had been established, they met for the first time to be informed about the evaluation and their role, but most importantly to comment on the guidelines for self-evaluation that were prepared by the evaluation officers for this meeting. The focus of the surveys that were going to be carried out was also discussed.

The point of departure for the evaluation was the objectives defined for the programme in the ministerial orders and the objectives formulated by the programmes themselves. Furthermore, the evaluation focused on the quality of the programme as a whole, i.e. input, teaching process and output. Organisational and structural aspects of the programme were also included. The guidelines for self-evaluation consisted of a number of themes, including, study structure, teaching staff composition, curriculum, teaching and assessment methods, internships, etc.

Each theme was covered by a number of specific questions addressing strengths and weaknesses of the programmes.

The programmes then established self-evaluation groups that assumed responsibility for the organisation of the self-evaluation process and the writing of the self-evaluation reports. The self-evaluation groups consisted of representatives of the programme management, teaching staff, students and technical and administrative staff. The duration of the self-evaluation phase was 4½ months, including 2 months' summer vacation.

The self-evaluation reports were of varying quality. Some were voluminous while others were very brief, but more importantly, some lacked an analytical and self-critical dimension. This is a common problem, though EVA tries to prevent it by emphasising the importance of analytical self-evaluation processes at the start-up meetings, and by formulating self-evaluation guidelines designed to encourage reflection.

While the programmes were preparing their self-evaluation reports, a consultancy firm, in co-operation with the evaluation officers, was preparing user surveys. In connection with this evaluation, two surveys were conducted: a qualitative focus group interview survey among graduates, and a quantitative questionnaire-based survey among labour market representatives.

Higher education programmes are required by law to ensure that an external examiner attends a minimum of one third of the final examinations and produce an annual report on the general state of the programme. In relation to this evaluation, a meeting was held between the evaluation group and two external examiners to elaborate on what had been raised in the external examiners' annual report.

When the surveys and the self-evaluation reports were completed, the evaluation group met to discuss the reports and prepare the site visits. Based on discussions of the reports, the evaluation officers prepared a checklist for each site visit, containing the themes that were to be discussed and specific questions that the evaluation group wished to pursue with the different groups that they were going to meet.

The site visits lasted a day per institution. The evaluation group and the evaluation officers conducted separate meetings with the self-evaluation groups, representatives of the permanent and part-time teaching staff, representatives of the students and technical and administrative staff. The visits concluded with a meeting with the management of the institutions. Each meeting lasted between 1½ and 2 hours. The evaluation group stayed together during the visit and all members of the evaluation group conducted all visits together.

Following the site visits, the evaluation officers drafted a report that the evaluation group met to discuss twice before it was sent to the institutions for comments. The institutions corrected a few factual errors and commented on the conclusions and recommendations but these did not lead to major changes to the report.

Prior to publication, EVA's board received and commented on the report and the evaluation report was then sent to the institutions, to the most important stakeholders and to the Ministry of Education. Following publication of the report, EVA arranged a conference for all relevant stakeholders based on the report. The purpose of the conference was to discuss future perspectives for the programmes.

The evaluation of the ´bachelor of social work` and ´social communicator diploma` programmes contained far-reaching recommendations regarding the organisation of the programmes and the internship periods. Many of these recommendations have now been implemented.

Northern Ireland

Stanley Goudie
Assistant Chief Inspector
The Education and Training Inspectorate, Northern Ireland

The School Support Programme

The case study centres around two government-funded, successive improvement programmes, designed to support low-achieving schools. The more recent of the two, the School Support Programme (SSP), replaced the Raising of School Standards Initiative (RSSI) in 1998. Both programmes have at their centre that if schools are to improve, they need a blend of external support on the one hand, and a commitment on the part of the school to self-evaluation and self-improvement on the other hand.

The SSP seeks to:

- improve the quality of school leadership;
- raise morale among pupils and staff;
- raise teacher expectations of pupils; and
- increase pupil self-esteem and self-worth.

The main purpose of the SSP, however, is to raise the standards of attainment of the pupils in the schools involved.

In primary schools, the focus is on improving the pupils' levels of attainment in literacy, numeracy and information and communication technology (ICT).

In post-primary schools, the foci are improving pupil achievement, qualifications, ICT competence, increasing pupils' employability and, where appropriate, improving their standards of behaviour.

For schools entering the SSP, the expectation is that they should be self-sufficient in managing their own improvement, normally within three and, exceptionally, five years of entering the programme.

All of the schools admitted to the SSP were inspected by the Inspectorate, either before or on joining the programme; these baseline inspections were designed to ensure that the staff had a clear view of the areas for improvement they needed to address. The staff of each school was expected to give a commitment to work within a whole-school perspective to bring about improvement in the quality of learning, teaching, school management and leadership, and in setting targets for improvement. The major role and responsibility of CASS, in relation to the SSP, was to manage, monitor and evaluate the effectiveness of the support programme, to assist school review, and to help schools prepare an action plan for improvement.

Schools exit the SSP when they are judged by the ELB Steering Group and the Inspectorate to have reached a level where there is significant improvement in the overall quality of learning and teaching, and in the standards reached by the pupils. In addition, schools exiting the support programme will have developed a culture of self-evaluation and continuous improvement within the normal level of support available internally and from CASS.

Inspection has revealed evidence of improvement in almost all the schools in the SSP. A majority of the schools are increasingly developing a more critical, self-evaluative approach to their

work. Such a development is a crucial factor in promoting improvement and in raising the standards the pupils attain.

Schools which participated in the SSP programme:

- reported positively on the support provided by CASS, and valued the professional relationships which had been developed with the staff;
- agreed with the findings of baseline inspections conducted by the Inspectorate;
- focused on whole-school issues linked to improving the quality and effectiveness of teaching and learning;
- developed strategies which enabled them to increase their involovment in school development planning, action planning and target-setting;
- identified success criteria linked to priorities within the school action plan;
- took increasing responsibility for monitoring, recording and evaluating progress;
- developed strategies and approaches to enable them to take ownership of the improvement process, to evaluate current practice and to plan for improvement;
- enriched the learning experiences of pupils and, to differing degrees, raised the standards they achieved;
- benefited, in almost all instances, from improvement in the effectiveness of the leadership and management of the curriculum; and
- began to share practice within and across year groups and departments and, in some instances, the staff visited other schools to observe practice and to talk with teachers.

On the basis of observed practice, the factors which have enhanced the capacity of the schools to respond positively to inspection findings include:

- follow-up inspection visits when the Inspectorate evaluated progress over the previous year, and the subsequently issued letter acknowledging the progress made and identifying areas where further improvement was still required;
- strong and effective leadership and management of the improvement programme by the principal, senior management and SSP co-ordinator;
- clear action plans which were based on the baseline inspection findings and agreed with the entire staff;
- the collegial working of the staff in the schools;
- the contribution of the designated CASS link officers in providing support and challenge to the schools;
- the allocation of funding which was well matched to school needs and clearly identified success criteria;
- the active support of school governors;
- the direct observation of teaching and learning by, for example, the SSP co-ordinator, head of department, vice-principal or principal;
- the improved resources for teaching and learning such as ICT software and hardware, and library texts.

In conclusion, a recent progress report by the Inspectorate covering the outworking of the SSP over the period 1998 to 2002 makes encouraging reading. In particular, of the 108 schools that have participated in the SSP over this period, 67 schools (or 62%) exited the programme judged as having made sufficient progress to no longer need intensive support.

The report also highlights the crucial part played by baseline inspections and the annual follow-up inspections as key factors – acting as catalysts to improvement and as aids to sustaining improvement respectively. It also highlights the importance of professional dialogue and good working relationships involving the Inspectorate, the schools and CASS in effecting improvement in leadership, management, learning and teaching, and in raising the standards the pupils achieve.

Canada – The School Sector

Richard M. Jones
Associate Director, Assessment & Reporting
Education Quality and Accountability Office (EQAO), Ontario

Assessment in Ontario

According to Earl and Graham (1994), for nearly three decades, there was very little standardized assessment or testing in Ontario. Throughout the 1950s and 1960s there had been standardized exit examinations from Grade 13 (departmental examinations) in all subjects areas, and these formed the sole basis for entry into university. In the mid 1960s, the results from these examinations were combined with teacher marks for university entry. In the late 1960s, the examinations were discontinued, and teacher marks became the basis for university entry. This practice has continued ever since.

In the 1970s and 1980s, some school boards chose to administer norm-referenced, standardized tests at selected grade levels, but the results were used only as additional information to assist teachers in making decisions about individual student placement or program improvement, and the tests were, as a result, "low-stakes." At the provincial level, there was no testing program at this time. Ontario left student assessment in the hands of educators. Teachers were expected to develop evaluation procedures and examinations that measured the achievement of students in courses and programs based on provincial curriculum guidelines.

By the late 1980s and early 1990s, there was growing evidence that the Ontario public was beginning to lose confidence in the education system: a provincial study by the Ministry of Education (1991) showed that there was considerable diversity across the province in OAC (Grade 13) English examinations; the province's results on a series of international assessments were a cause for concern; the universities began asking for a return to provincial examinations; and opinion polls began to show that the majority of the public supported province-wide testing.

As a result of growing public concern, the Government of Ontario conducted a Royal Commission on Learning that consulted widely with educators, parents, students and taxpayers. In its report (1994) the Royal Commission called for the establishment of an independent, arm's-length agency to implement a series of programs to respond to public demands for greater quality and accountability in the education system. Bill 30, the Act to establish the Education Quality and Accountability Office, was given Royal Assent in June, 1996.

Education Quality and Accountability Office (Policy Framework)

The Education Quality and Accountability Office (EQAO) Policy Framework (1996) includes statements about the agency's mission, values, objectives and activities as follows.

Mission:

> *"EQAO will assure greater accountability and contribute to the enhancement of the quality of education in Ontario. This will be done through assessments and reviews based on objective, reliable relevant information, and the timely public release of that information along with recommendations for system improvement."*

Values:

> *"EQAO values:*
> - *the well-being of learners above all other interests,*

- *only that information which has the potential to bring about constructive*
- *change and improvement,*
- *the dedication and expertise of Ontario's educators and will work to*
- *involve them in all of its activities, and*
- *the delivery of its programs and services with equivalent quality in*
- *both English and French."*

Objectives:

"EQAO:

- *provides a broad range of information about quality to support*
- *informed judgement and decision-making,*
- *develops student assessment instruments and practices that contribute to*
- *enhanced teaching and learning,*
- *models positive learning practices within its own organization,*
- *sets its horizons globally, bringing high standards & superior*
- *practices to all its activities, and*
- *seeks markets for its products, services and expertise."*

Activities:

"EQAO:

- *designs and implements a comprehensive program of student assessment*
- *within government-established parameters,*
- *advises the Minister of Education and Training on assessment policy,*
- *develops and implements a provincial performance indicators program,*
- *leads Ontario's participation in national and international assessment and*
- *indicators work,*
- *promotes research in best practices in assessment and accountability,*
- *reports to the Minister, the public and the education community on*
- *system quality issues and recommends improvements, and*
- *conducts quality reviews in cooperation with school boards."*

Education Quality and Accountability Office (Project Summaries)

The activities of EQAO are strategic in that they reflect the need for student achievement and planning information that can be used for improvement, accountability and credentialing purposes.

Grades 3 and 6 Assessments of Reading, Writing and Mathematics:

The Grade 3 and Grade 6 Assessments of Reading, Writing and Mathematics are conducted annually in May and involve all Grade 3 and Grade 6 students in publicly funded schools in Ontario. Students are assessed over a five-day period, approximately two-and-a-half hours each day. The results from these assessments are reported at the individual student, school, school board and provincial levels. The assessments measure how well students have met the provincial learning expectations, as stated in *The Ontario Curriculum, Grades 1-8* in reading, writing and mathematics, that students are expected to have acquired by the end of the grade being assessed. These assessments contain performance-based activities (requiring written responses), short written-response and multiple-choice questions. This combination of question types allows students to demonstrate both the depth and the scope of their learning and provides a variety of ways for students to demonstrate what they know and what they can do.

In the reading component, students demonstrate their knowledge and skills by reading a variety of materials such as factual information, stories and poetry. This part of the assessment measures how well students use various reading strategies and conventions and how effectively

they understand concepts, make inferences and connect ideas. In the writing component, students demonstrate their knowledge and skills by using various forms of writing and by writing for different purposes. Students produce two pieces of written work. One piece models the writing process while the other is a short on-demand piece. In mathematics, students demonstrate their knowledge and skills by solving problems, applying procedures and explaining how they have arrived at their answers.

The Grades 3 and 6 assessments, like all of EQAO's provincial initiatives, involve teachers at every stage. Educators from across the province participate in developing, field-testing and validating assessment materials, as well as the marking of student responses.

Grade 9 Assessment of Mathematics:
All Ontario Grade 9 students in publicly funded schools are expected to participate in this assessment. The assessment measures how well students have met the expectations in *The Ontario Curriculum, Grades 9 and 10: Mathematics,* and is based on the expectations of the Grade 9 mathematics curriculum, both Applied and Academic. Schools have the option of delivering semestered or full-year courses; therefore, EQAO develops and administers separate versions of the assessment annually in January and May/June.

The assessment includes multiple-choice, short-answer and performance-based tasks. There are approximately two-and-a-half hours of testing that take place over two or four days. (There are two optional administrative approaches.) Reports are provided at the individual student, school, school board and provincial level.

Ontario Secondary School Literacy Test (OSSLT):
In 1999, the Ontario Ministry of Education announced that successful completion of the OSSLT would become one of the 32 requirements for an Ontario Secondary School Diploma. The OSSLT is administered annually in October, and is presented in two time blocks of two-and-a-half hours each, one during the morning of the first day and another on the morning of the following day. The first full administration of the OSSLT was in February, 2002. (The administration of the assessment was delayed due to a security breach.)

The test consists of two integrated components: a reading component and a writing component. There are three types of questions: multiple-choice, short-answer and questions requiring an explanation. In reading, students are asked to read and answer questions related to a series of short information, graphic and narrative selections. In writing, students are asked to complete four writing tasks: a summary, a series of paragraphs expressing an opinion, a news report and an information paragraph.

In this assessment, students meet a minimum standard (pass) or do not meet the standard (do not pass). Students may pass/not pass one or both components of the assessment. If they do not pass one or both components, they can retake the assessment at the next administration. On subsequent administrations, students are only required to pass the component(s) in which they were previously unsuccessful. Reports are provided at the individual student, school, school board and provincial levels, with more detailed feedback being provided to unsuccessful students.

Following the second full administration of the OSSLT in October 2002, the Ministry of Education announced that, beginning in the 2003-2004 school year, unsuccessful students who have had two opportunities to write the OSSLT would be eligible to participate in a course named the Ontario Secondary School Literacy Course (following the same curricular design elements as the OSSLT). Successful completion of this new course will serve as the graduation requirement in lieu of passing the OSSLT.

National and International Assessments:
EQAO manages Ontario's participation in the Council of Ministers of Education, Canada's (CMEC) national School Achievement Indicators Program (SAIP) in mathematics, reading and

writing and science. The assessments involve the testing of randomly selected samples of schools and 13- and 16-year-old students. In general, testing occurs in one subject per year on a rotating schedule. Reporting is at the provincial and national levels only. The sampling framework does not allow for reporting at the student, school or school board levels.
At the international level, EQAO manages the province's participation in assessments such as the following: Program for International Student Assessment (PISA), conducted under the auspices of the Organisation for Economic Co-operation and Development (OECD); Trends in International Mathematics and Science Study (TIMSS); Second Information Technology in Education Study (SITES); and Progress in International Reading Literacy Study (PIRLS), conducted under the auspices of the International Association for the Evaluation of Educational Achievement (IEA). As with the national assessments, the international assessments involve randomly selected samples of schools and students at given ages/grades, and reporting is conducted solely at the national and provincial levels.

Education Quality Indicators Program (EQUIP):
The Education Quality Indicators Program (EQUIP) collects and reports information about the many factors inside and outside the school community that affect student achievement across the province. In 2001 – 2002, EQUIP collected data on twenty-one quality indicators at the elementary school level and thirteen quality indicators at the secondary school level.

EQUIP provided demographic and other environmental information to help teachers and administrators in their joint planning for school improvement. EQUIP also provided contextual information for examining and understanding student achievement scores.

Statistical and survey data were used in preparing EQUIP reports. Statistical data were gathered from a variety of sources including Statistics Canada and the Ontario Ministry of Education. EQAO supplemented this quantitative information with survey data collected from individuals such as school administrators, educators, students and parents as part of EQAO's provincial assessment program.

The intent of EQUIP is to present a picture of the entire school environment, not just the achievement results. Test scores are therefore placed in a broader framework of socio-economic data, the teaching and learning environment, school leadership, student mobility, accessibility of educational equipment and resources, as well as other information that can help students, parents, educators and the general public understand the variables that are related to student learning.

Improvement Planning:
EQAO ensures greater accountability in the publicly funded education system in Ontario and helps to improve the quality of education by providing comparable data on student achievement, and by reporting on plans for improvement. Improvement planning is a strategy that brings about educational change by enhancing student outcomes and increasing school boards' and schools' capacity to design and manage change. Improvement plans are the means by which students, teachers, parents, principals and school board staff change conditions for learners to ensure improved achievement. School boards develop and release board-wide plans designed to foster and support improved student achievement. Using the system plan as a foundation, principals and teachers, working with parents, school councils and other representatives of the community, prepare and release school plans for improvement. In support of improvement planning, EQAO publishes various documents, including the *EQAO Guide to School and Board Improvement Planning* and an *Educator Handbook*.

Each year, EQAO mandates that the province's 72 publicly funded school boards submit their improvement plans to EQAO for review. The office reviews the school board improvement plans according to established criteria, and then individual reports are prepared and distributed to the boards to inform them of the extent to which the plans meet the criteria. In addition, EQAO develops a summary, provincial report that outlines the strengths and weaknesses observed in improvement planning.

France – The School Sector

Gérard Bonnet
Head of European and International cooperation
Direction de l'évaluation et de la prospective, Ministère de la jeunesse, de l'éducation nationale et de la recherche, France & Visiting Fellow at the Institute of Education, University of London

Assessment of pupils' achievements & evaluation of schools

The general principles having been described in the article on the evaluation of the school sector in France, it is proposed here to present two concrete examples of evaluation in France: the assessment of pupils' achievements and the evaluation of schools using specific indicators. These case studies should help to understand the principle that evaluation in France is seen first and foremost as providing tools for the practitioners to change their professional practices.

Pupil assessment

The development of specific forms of assessment was established with the twofold objective of, on the one hand, assisting with the monitoring and the steering of the education system in the context of greater regional and local autonomy, and of providing information and tools to teachers to help them evolve better teaching practices, on the other.

Assessment as an observatory of pupil achievement:
This type of assessment systems is intended as a tool to manage the entire education system. Its methodological construction allows comparisons over time. It is based on samples representative of schools, classes and pupils. It is organised at the end of learning stages. In terms of education policy goals, it enables decision makers to see which goals have been attained and which have not. The organisation of teaching, the contexts of learning and the populations thus characterised, it becomes possible to act on the curriculum at national level.

This type of assessment meets a precise need. Based on objective observations, it enables comparison between the results of pedagogical methods in the education system and the goals set, at crucial points in the learning process.

Sample-based assessment was first implemented in France during the 1980's and then somewhat overtaken in importance by diagnostic assessment. In the past couple of years, the former has recovered its place in assessment policies.

The oldest of such sample-based assessment is that which is conducted on a regular basis every few years at the end of lower secondary school – modal age 15 – for both the general and the technological streams. This assessment deals with all the subjects taught in lower secondary schools. In the same survey, non-cognitive attainments are also measured through questionnaires, as are pupils' attitudes and values concerning life in school and society.

Two assessments of this type are currently being organised and will take place during the spring of 2003. The first will assess the skills acquired at the end of lower secondary education, while the second will seek to make similar measurements at the end of primary education. The design of both assessments will be targeting the general skills of pupils and not just those directly linked to curricular objectives.

For the first data collection exercise in 2003, the principal skills assessed will be, in primary schools, general abilities in mastery of spoken and written language, and, in lower secondary schools, the common skills (cutting across the different subjects taught) and skills in modern

languages (German, English and Spanish). In addition, lower secondary school pupils will complete a questionnaire on their motivations and hopes for the future, their experiences in class and in the school, and their relationships with other pupils and adults. The assessments will be supplemented by the collection of information on the environment in which teaching is delivered, based firstly on questionnaires completed by teachers, school heads and inspectors, and secondly on descriptions extracted from existing databases held by the ministry.

To provide the background to all this, longitudinal surveys are conducted with real cohorts of pupils monitored over several years along the various stages of schooling, including higher education and entry into the job market. The methodology involves questionnaires to head-teachers, parents and pupils and is complemented with pupil interviews, and aims to establish links between levels of educational achievements as measured through national testing and extraneous factors such as school organisation, home background, etc. This approach is also used to describe the main trends in school careers: which streams are followed by which types of pupils; relationship between socio-economic background and choice of course; social disparities in repeating years; academic achievements of pupils from ethnic backgrounds at lower secondary school, etc.

The sample-based pupil achievement assessment procedure described above is essentially summative. It provides snapshot information on achievements at a specific time, but, in some cases, it also includes a diachronic dimension since it is possible to trace evolution from one assessment to the next.

The comparison over time, thus measured, is then complemented with a comparison between countries. France is a member of the *International Association for the Evaluation of Educational Achievement (IEA)* and is, or has been, involved with current or past surveys conducted by the *IEA* (*Reading Literacy*; the *Third Mathematics and Science Survey – TIMSS*; the *Progress in International Reading Literacy Study -PIRLS*). France plays also an active part in Network A of the international indicators programme (INES) of the OECD and participates in the *Programme for International Student Assessment - PISA*. Other comparative work is or has been conducted bilaterally or with a number of countries: achievements in mathematics at 8 and 11 in France and Scotland; achievements in English as a foreign language at 15 in 8 European countries (France, Sweden, Germany, Denmark, Norway, Finland, the Netherlands and Spain).

Assessment as a tool for educational change:
The French assessment system includes, alongside the methods used to measure achievements, procedures using standardised tests as educational tools to encourage educational change and to promote a culture of evaluation. This is why this type of pupil assessment is conducted in such a way as to involve teachers and schools and is devised as a diagnostic assessment. Diagnostic assessment has been reinforced over the past few years in the context of new education policy orientations. As the system is defined and organised in France, it is specific to that country.

The first step is to enable teachers to assess the strengths and weaknesses of their pupils through mass national testing. This has now been organised at the beginning of each academic year since 1989 for all pupils aged 8 (grade 3) and 11 (grade 6) in French language and mathematics. A similar survey for pupils aged 16 had been taking place since 1992 in all subjects taught in upper secondary schools but was recently discontinued. Last year it was decided to introduce this assessment for the second year of lower secondary school (grade 7). Around 800,000 pupils for each level, the whole of the age group, are thus tested every year.

Testing is primarily intended for diagnostic use by teachers and parents within the school. However, a representative sample of the test for each level is analysed centrally by the education ministry to obtain a statistically valid national picture of pupil achievement for the skills measured. This complements the information provided by the sample-based assessments described above, but it must be emphasised that the primary objective of mass testing is pedagogical and formative in nature. The tests are different every year as they are intended to re-

flect skills that need to be assessed with reference to changes in the curriculum or policies. They cannot, therefore, be compared (in the psychometric sense) from one year to another; they are not indicators of evolution within the education system.

The tests are voluntarily different in their content and form from those given during lessons by teachers to assess the knowledge of their pupils or train them in a particular skill. Furthermore, a skill is made up of components, and diagnostic assessment enables analysis of these components, as a pupil progresses in its mastery. Diagnostic assessment is not a learning situation, nor a test of progress within learning stages, nor a remedial situation.

The tests are implemented through either multiple choice questions or free responses, according to what is being tested. They are organised into a number of "skills" (e.g. spoken comprehension, written expression, etc.) for which specific items make it possible to assess achievement or otherwise. They are based on what is expected of pupils, given the objectives and standards set by the curriculum. Questionnaires to head teachers, teachers and pupils are also normally used to provide background information. Standardised coding is used, and this is normally carried out by the teachers themselves. The tests are administered by form teachers at the beginning of the school year, since the aim is to provide information on pupil achievement so that work can be organised accordingly. In this way, teachers better accept compulsory testing, since pupil achievement can in no way be regarded as a reflection on their work but merely as a teaching aid.

Tests and assessment instruments are developed under the responsibility of the education ministry. In practice, groups of teachers, academics and inspectors from a number of regions, together with ministry officials, meet throughout the year to devise, develop and test the material. Much emphasis is laid on the principle of "participation", i.e. actively involving all interested parties in designing the tests.

Since the assessment allows measuring the extent to which pupils have attained the objectives for each subject, test outcomes enable teachers to define what shortcomings in pupils' attainments have to be overcome before effective teaching can be delivered. Each school is responsible for conducting the analysis of its own results using specific computer software provided for the purpose and for drawing up a "success chart" for each pupil and each form. It can also compare its performance to the national standard since schools can access the nationally computed results of the national sample of schools through the Internet. Schools receive the main results together with a pedagogical commentary. It is intended that the findings of assessments should be taken into account by inspectors and trainers in the recommendations that are made to teachers and in the course of in-service training programmes.

The policy of mass testing is clearly one which aims to make teachers – who, in France, enjoy a large degree of individual freedom in their teaching and in the conduct of their classes – aware of the usefulness of assessment to gauge pupils' needs. This has begun to happen, as several surveys have indicated. It was shown that the vast majority of a sample of primary and lower secondary school teachers make use of the results of the assessment tests in their teaching, while this was true for only half of upper secondary school teachers, thus reflecting differing traditions in teaching styles and independence. It was on the basis of such findings that testing was discontinued in upper secondary schools.

The education ministry has gone a step further towards this policy by making a bank of standardised assessment items freely available to teachers on the Internet, for use during the year on a self-service basis. It is intended that this assessment item bank will cover subjects at a variety of levels for both primary and secondary school.

Although there is general agreement about the usefulness of diagnostic testing, in particular among primary school teachers, it is also clear that a lot still remains to be done to convince all teachers that it should be viewed as an integral part of teaching and used as such.

Evaluation of schools

One of the questions which educational assessment experts in many countries are currently occupied with is that of school evaluation. For various institutional and cultural reasons, and contrary to popular belief, there has never been a formal and systematic school evaluation procedure in France. One of the reasons is that ideological considerations about equality tended to impose the view that all schools should function in the same way so as to be strictly identical. This situation has evolved over the past twenty or twenty-five years, and the idea that schools can be different (which is entirely legitimate), and that they may not always offer the same type or quality of service (which is much less so) is now accepted.

In addition, there are no inspection procedures on the scale found in some other countries. Despite the fact that there are district inspectors for the primary level, regional inspectors for the secondary level and national inspectors, inspections traditionally do not deal with schools in a global perspective but rather concentrate on individual teachers, not least for promotion purposes. In short, they are not equipped to undertake the kind of overall national assessment of schools that is necessary to monitor the system and provide information to parents.

This explains why, until a few years ago, there was no proper national monitoring of schools. To respond to parent concern, the press had been publishing yearly league tables for upper secondary schools (age 16 to 18+) based on their raw results at the school leaving examination (*baccalauréat*). No information was provided concerning lower secondary schools. As a result, and to make up for such a partial and sketchy description of school performance, the education ministry was prompted to develop an original system for providing information on schools to the public, and to the schools themselves to help them improve their processes and performance. This approach is underpinned by the idea that the evaluation of schools is an educational tool, like the assessment of pupils. This is radically different from the present rationale of school inspection in France, and at the same time departs from the idea that value added can be measured through monitoring individual pupil achievement at various key stages.

From around 1995 onwards, a set of nationally developed standard indicators was made available to secondary schools, together with the necessary computer software, by the evaluation division of the education ministry. Most of the indicators proposed by this system take the form of a feedback of information to the schools. Schools are required to provide data for the national information and management systems, and these data are then automatically returned as "personalised" indicators which are ready for use and accompanied by references (national, regional or district averages), which allows them to place themselves in comparison with others.

These standard indicators constitute the background against which schools can measure themselves. The number of standard indicators is deliberately kept down to a manageable 21. Their objective is to provide an accurate description of the mode of functioning of the schools, and to allow each school to compare its practices with that of similar schools, nationally or in the *académie*. The indicators fall into four categories: input indicators (characteristics of pupils), output indicators (results of the school in examinations, admission of pupils to higher forms or institutions), indicators relating to resources, and indicators for school management and environment.

It is intended that analysis of the information provided by the indicators will lead schools to revise their policy, amongst other things, in terms of pupil admission procedures, selection and opportunities for repeating years. This will naturally take time, as the whole of the school community will have to be made aware and convinced that some changes in their professional practices may be necessary. Next to this, an equally important part of the programme is to encourage schools to devise their own specific indicators, based on local characteristics and needs, to help them develop and assess their own school development plan.

Three indicators among the output indicators (commonly referred to as "performance indicators") are used in combination to estimate the effectiveness of upper secondary schools: the

proportion of pupils in the school passing the examination among those sitting it; the proportion of pupils from the lower forms admitted into the final year; and the proportion of those pupils leaving the school at any time that have passed the examination. This gives a very different account of school performance than simply looking at the single raw indicator of success at the examination, since several aspects of school policy are thus brought to light. By measuring the value calculated for each of the three indicators for a particular school against nationally calculated expected readings for the school – given its specific educational and sociological make-up – against those of comparable schools in the region and in the country, i.e. schools with pupils of the same ages and from the same socio-economic background, the value added for that particular school becomes apparent. Thus a very selective school, which only retains the better pupils through the various stages, will be seen to have no great difficulty in leading the great majority of them to success but it will also be seen to have added less value than one which struggles on with children of mixed abilities but nevertheless manages to bring a significant number to the required standard. This is naturally a first approach that needs to be further refined to avoid unfair treatment of some categories of schools for which external constraints are particularly strong.

It is vital for the education ministry to provide information on the evaluation of secondary schools in order to avoid this being done by the market, which will tend to use criteria chosen to promote certain schools rather than improving them all. Thus, publication of such information contributes to reducing inequalities. For the system as a whole (and for a given school), transparency constitutes a strong incentive for action, and publication ensures that weight will be put behind efforts to improve the situation. The three performance indicators calculated each year for each upper secondary school are available to the general public on the Internet.

An indicator system has been set up for primary education, and its implementation began in 1998. One of its aims is to facilitate a working relationship between primary schools and lower secondary schools in the hope of fostering smoother continuity between the two educational levels. The indicators are currently organised around three thematic domains: local characteristics; functional efficiency; and pupil characteristics. They are applied to three geographical entities: the primary school; the relevant lower secondary school's catchment area; and the educational zone in which the primary school is situated.

Hungary – The School Sector

Peter Vari
Director
National Institute of Public Educational Services, Center for Evaluation Studies, Hungary

National Assessment of Basic Competences Survey

The most current and important momentum in the Hungarian public educational system is the National Assessment of Basic Competences Survey (National ABC), to be organized for the second time in 2003. The survey stems from the previously mentioned precedents, and from the assessment practices and developments within Hungarian public education, i.e. from an ever growing need for a full-scale continuous national evaluation (as opposed to research based on a representative sample) that could fulfil several goals through its structure, methods and tools.

According to plans, the main aims of the survey would be to spread the "culture" of measurement and quality assessment in the everyday practice of schools, and to start systematic, nationwide, compulsory surveying in the country.

The evaluation system that maps the literacy of pupils (from all primary and secondary schools in the country) in reading and mathematics was based on this idea. The tests, that include both open and closed-end exercises were primarily made to test competence and so did not examine the requirements of the school curriculum, only the ability to apply this knowledge.

The main purpose of this survey was:

- To assist schools and the local education authorities in quality development
- To map the general situation in national and regional school-types
- To study the implementation of pedagogical systems
- To develop the assessment and evaluation knowledge of teachers
- To provide means and technology of assessment for all agents in the field of education evaluation

The value that manifests itself in the National Assessment of Basic Competences Survey is, firstly, the spreading of the culture of assessment itself, and, secondly, the highlighting of the areas of competence, with special regard to the Hungarian results of the 2000 OECD-PISA Survey. The aim of the assessment was a nationwide mapping of mathematical and reading competence, applying tools and a system that, at the same time, helped to implant the "culture' of evaluation in everyday practice. The subject of the test was reading-literacy and mathematics; its method comprised: surveying through tests of two chosen grades in public educational institutions (5th and 9th grades in 2001, and 6th and 10th grades in 2003), where the tests consisted of open-ended and multiple choice questions, completed at the same time at all locations and according to a uniform set of rules.

The uniform set of rules that defined how to fill out the tests, and which were obligatory to all schools, were created by the Center for Evaluation Studies. Experts commissioned by the Center controlled that rules were observed.

Tests completed by students were marked by the school's teachers according to a centrally developed marking-guide. Twenty tests, chosen at random in each school, were sent for central correction and analysis. The tests sent to the center were corrected by trained coders, and the analyses were based on the evaluations made by the coders. Analyses were prepared in two forms. Firstly, the so-called "school reports", prepared for the schools themselves, which contained the performance of the given school's pupils in comparison with the national average and the average within the school's own category (population catchment, type of school). Re-

sults were presented with the help of numerical data, charts and graphs. In order to better understand the data provided, schools were also given a handbook that contained the following: a detailed description of the theoretical background for the assessment; background information to the compilation of the tests; a list of statistical terms and their meanings that were imperative in the understanding of the findings; the national average results in relation to the test questions. Schools also received data processing software on a CD, which, by entering all data of the corrected tests, enabled the school to do further detailed analysis of its own results. Besides the school-level analysis and feedback, analysis on the national level was prepared, too. From this, experts could deduce deficiencies and shortcomings of the particular areas and types of exercises that caused problems, and the division of performance according to population catchment-types, regions, school size, and the type of school-program.

Apart from the fact that through the survey we could get a picture of student competence, it was an important aspect that an objective assessment tool would reach the schools and could be used by them. This tool is apt in its content to measure performance in competence fields, makes these kinds of tasks widely known, and practicing teachers can learn the different aspects of correction. But not only its content, also its nature makes it an objective, standardized evaluation tool for teachers, which can be used over the long term, and which was not available to schools before. By these means, the culture of evaluation, local and institutional assessment and school self-evaluation can all be further developed.

As mentioned above, during the first examination in 2001, tests were corrected on two levels. Tests of twenty pupils selected at random in each school were coded, according to a central marking guide, by the school's own teachers. The same marking key was used for central correction. The aim of the double correction was to allow teachers to get to know, and to use in practice, the evaluation of open-end test questions, thus gaining experience of this new assessment form. It was also important to examine the extent to which teachers were capable of consistently evaluating tests with the help of the correction guidelines, which, in turn, gave an indication of the reliability of the results in the national assessment.

Comparative analysis showed that, although teachers mastered and followed the rules of correction to a rather high level, reliability in the statistical sense did not reach acceptable standards.

After the 2001 survey, the Center for Evaluation Studies began a significant development project to prepare question papers for a new round of evaluation.

The National Assessment of Basic Competences Survey will be organized for the second time in May 2003, amongst 6th and 10th grade students. The main features and aims of this year's test are the same as in 2001; the characteristics of the test, the rules of performance and the characteristics of the following analysis are unchanged. In terms of correction, however, a change will be introduced: test papers selected for central correction and analysis will not have to be previously corrected in the given school, although the institutions do have the opportunity to perform this task, as central correction keys will be handed out to schools this year, too. Aspects of the correction will thus become known. This is important, as schools will continue to have the opportunity to process the results of their own corrections with the help of the software. And so (complementing the analysis made by the center), the results of this year's survey may be processed on an institutional level.

This new element of assessment (together with the previous ones) could mark the very important beginning of a new and more effective approach towards quality assessment in Hungary. It appears to be essential to organize the cooperation between the different agents of quality assessment in order to gain the maximum number of advantages possible from carrying out these kinds of surveys, analysing the results, and subsequent further research.

Canada – Higher Education

Jacques L'Écuyer
Chairman
Dany Rondeau
Project coordinator
Commission d'évaluation de l'enseignement collégial du Québec (CEEC)

The « Commission d'évaluation de l'enseignement collégial » of Quebec (CEEC)

The CEEC represents a very special case in Canada. It is the only agency with the mandate to evaluate non-degree granting institutions, both public and private. This mandate covers the evaluation of the institution itself, its program and its academic policies. The CEEC is an autonomous agency created by an Act of the Government of Quebec. It is composed of three commissioners who are assisted by a staff of civil servants. All college level institutions in Quebec are subject to the evaluation of the CEEC.

The CEEC was created in 1993 in the wake of a major reform of college education in Quebec. Before the reform, the college system was highly centralized, with programs completely elaborated by the Ministry of Education. Colleges had little autonomy to adapt these programs to their environment and their internal context. One of the major aims of the reform was to decentralize the system and give more responsibilities to institutions, particularly with respect to the educational programs, in order that they be more responsive to the needs of their community and to the constraints of their environment. The CEEC was created in this context as a mean for the colleges to account for their realizations and for the government to make sure that the decentralization proceeded in an ordered manner.

Values and purposes

At the beginning of its operation, the CEEC clearly indicated that it would fulfil its mandate with the objectives of improving the quality of college education, testifying to the value of the education offered in each institution, and helping the institutions in the development of a culture of evaluation. This last objective came from the fact that few colleges had experience in the field of quality assurance and program evaluation. These objectives clearly define the CEEC as a quality improvement centered agency.

The first objective was pursued in all the evaluations realized by the CEEC, whether they were evaluations of program or later evaluations of institution. The Commission is supportive of the institutions in their evaluations and reports and leaves them space and time for improving their performance. The Commission's reports point out strengths and weaknesses and include observations and recommendations. Recommendations call for appropriate actions to correct the situation. The reports are public documents, posted on the CEEC's website.

In cases where a program or an institution has deteriorated to a point where there is little hope of progress, the Commission recommends that it be closed. In this sense, the CEEC is also a quality assurance agency. This recommendation is made to the Minister of Education. It is seldom used.

The fact that the CEEC's reports are public documents increases considerably their efficiency. On many occasions, institutions have used them to bring about changes that would otherwise

have been difficult to realize. In this sense, the CEEC is seen by many colleges as an efficient agent of change.

As indicated above, at the time of the reform and the creation of the CEEC, colleges had little experience in the field of evaluation. For this reason, the Commission developed detailed guides for program evaluation, gave workshops to help institutions to realize their self-evaluations and selected and trained experts from the college community. In this way, colleges have learned how to realize efficient program evaluations. A recent audit of the quality assurance processes of the colleges has shown the importance of the progress made in the development of a culture of evaluation and the realization of efficient evaluations.

Objects

The initial mandate of the CEEC was to evaluate for each college, public or private:

- the institutional policies on the evaluation of student achievement
- the institutional policies on program evaluation
- the programs of study.

This mandate was recently extended to include the evaluation of the institutions themselves with regard to their academic and organizational management, their strategic plans and their plans to increase their students' success rate.

When evaluating a program, the Commission looks at the following: its relevance with regard to the needs of the society and their students; the coherence of the curriculum taking into account its objectives; the pedagogical methods utilized; the competence of the teachers; the adequacy of the resources; the assessment of the achievement of the students; and finally, at the program management.

When evaluating an institution, the CEEC examines the relevance and clarity of its mission and objectives, the efficiency of its organization, the adequacy of its resources, its planning process, its integrity and its outcomes. The Commission audits separately the quality assurance mechanisms developed in its policy on program evaluation.

The CEEC began its operation by the evaluation of programs and chose programs that were widely offered throughout the province, like computer technology and social science. In this manner, the Commission hoped to diffuse good evaluation practices in the institutions and, at the same time, give a first assessment of the quality of college education. After evaluating a number of programs in each institution, the Commission decided to audit the internal program evaluation practices of each institution. Recently, it began evaluating individual institutions in terms of organization and management. The Commission felt that this last operation would help institutions to better identify the origin of problems that appeared in a recurrent manner in the evaluation of programs. The CEEC also believed that institutional evaluations would permit the measurement of the extent to which the institutions had assumed the new responsibilities they were given by the 1993 reform.

Stakeholders

It should be clear from the above that the institutions are the first and main beneficiaries of the work of the CEEC. Though many stakeholders take advantage of the improvement in the quality of college education – in the first place the students, but also the employers, the universities when they receive college graduates and the society in general –, it is the institutions themselves that directly benefit from the evaluations of the CEEC. It is their activities, which are evaluated, their educational programs, which are strengthened, and their reputation which is enhanced. It is to them that the evaluation reports are addressed and for them that they are prepared, even though they are eventually made public and available on the website of the

CEEC. Finally, the CEEC adopted evaluation procedures and used instruments that were specifically designed to help colleges to develop their own culture of evaluation.

Methods

As indicated before, the CEEC is autonomous. It chooses the institutions and programs to be subjected to a particular evaluation, the evaluation methods and the moment of the evaluation. It is fully responsible for its judgments and its reports.

In all its evaluations, the Commission uses similar procedures. It first develops the necessary instruments, guides, questionnaire samples with the help of an advisory committee comprising academics and professionals from the labor market. In the case of the evaluation of a program, the institutions offering the selected program are then requested to carry out their self-evaluation following the criteria and suggestions of the guide. Upon receiving the self-evaluation reports, the Commission organizes site visits. The visiting committees are composed of one of the commissioners and three or four experts of the field. Each committee is chaired by the commissioner, and a professional of the CEEC acts as secretary. The report of the visiting committee constitutes the basis for the Commission's report. This final report is adopted by the three commissioners during one of the CEEC's formal meetings. It is then first sent to the institution for their reactions. Modifications can be made before a definite report is adopted. In other type of evaluation – audit and institutional review –, the same procedure is followed.

For its program evaluations, the CEEC has developed a general guide in which are presented its criteria. This general guide serves as the basis for the elaboration of more specific guides adapted to the situation of each program. In these specific guides, the CEEC indicates which aspects of the program should be examined more closely depending on the particular context, for instance, actualization and equipment in computer sciences, coherence in social sciences. For each criterion, the specific guides precise what type of information should be collected and make suggestions with regard to the analysis. The CEEC expects that the institution will have identified its weaknesses in its self-analysis and adopted a plan of action for improvement.

Special guides were developed for the audits and institutional evaluations. Of course, the criteria are quite different, but the approach is similar. There is, however, a major difference between these three types of evaluations: the framework of the evaluation.

In the case of a program evaluation, the objectives of the program constitute the reference against which the evaluation must be done. These objectives must be, and generally are, presented as competences that the graduates should have acquired. Therefore, the evaluation looks at the process and resources, as well as at the outcomes. Information related to outcomes, such as employer and graduate surveys, employment rates, must be used as well as internal data on courses, graduation rates, student satisfaction, faculty and resources.

When the CEEC is carrying out an institutional evaluation, it is the mission and objectives of the institution that serve as references. Here the focus is on the organization of the institution under review, to determine if it is optimally adapted to its mission and objectives. The outcomes are obviously looked at, but they are of a more general nature and may depend very much on the type and mission of the institution, e.g. small private institution, urban college, regional institution, etc.

As for the audits, the institutional policy on program evaluation constitutes the reference, and the audit is done to verify that it is in force and constitutes an efficient means of ensuring and improving the quality of the programs. The main data are, therefore, the files of the program evaluations already realized by the institution.

It must be realized that in all these evaluations, the CEEC always pursues its main objective of contributing to the improvement of the quality of college education. This is at the root of the approach it has taken, characterized by support to the institutions through guides, workshops and direct assistance, insistence on efficient self-assessment and use of standards which are for the most part qualitative. Quantitative data are utilized more as a base upon which judgments are made than as absolute norms.

France – Higher Education

Bruno Curvale
Project Manager
Comité national d'évaluation (CNE), France

The three main categories of evaluation

The CNE performs three main categories of evaluation: evaluation of institutions (institutional evaluation), evaluation of sites, and cross-cutting evaluations of disciplines.

Evaluation of institutions (institutional evaluation)

General principles

Evaluation of a higher education institution is carried out in two stages. An internal evaluation is conducted and organized by the institution. It is based on the Evaluation Guide prepared by the National Evaluation Committee and should involve all staff. It can be defined as the view the institution has of itself based on the data it produces. This step is both an important moment in evaluation pedagogy, during which the institution focuses on an approach to self-knowledge, and an essential step in determining the project and themes that will be at the heart of the external evaluation.

Next, the CNE organizes and co-ordinates an external evaluation by calling on experts. This step consists of a "peer" evaluation (academics, technical or administrative managers in higher education, figures from the French or foreign business worlds) in the form of appraisals carried out on site that lead to confidential reports. This step requires a few comments on the role of the experts. Progressively, the CNE is seeking to encourage experts to work together. The basis of their contributions remains their personal contributions as written following their visit to the institution, but today experts are asked to work together on the overall design of the evaluation, including upstream of the visit to the institution and later on with drafting of the final report. This group work is an important opportunity to clarify the implicit terms of reference used by each expert. It is also an important moment to fine-tune the objectives of the task assigned to each expert and to better coordinate their expertise.

How an institution evaluation progresses

The procedures used by the CNE are based on the general principles described above. They have varied over time and will continue to do so in order to adapt to the changes in institutions, and also to take advantage of the experienced acquired through the evaluations conducted. In addition, adaptation of the procedure to its subject is one of the quality criteria desired by the CNE for the evaluation of higher education institutions.

The decision. The decision to evaluate an institution behoves the CNE, which freely chooses its agenda. Evaluation is compulsory and institutions cannot refuse it. The custom is however to set the date after consulting the president or director of the institution so that the procedure does not interfere with other work (inspections by the IGAENR, for example) and that the results will be available at a timely moment for the institution (preparation of the institution's project, start of contract negotiations, etc.)

The CNE takes the decision to launch an evaluation of an institution during its plenary session. Responsibility rests with two (sometimes three) members of the Committee, assisted if necessary by a consultant designated by the Committee's president. Coordination is entrusted to a project manager by the general delegate, who constitute the steering group for the evaluation.

Launching the evaluation. During a first meeting between officials from the Committee, its general secretariat and the president or director of the institution concerned, the principles for the evaluation are agreed and the Evaluation Guide is provided. This meeting generally takes place on the CNE's premises.

A few weeks later, the steering group visits the institution to officially launch the evaluation. This day of meetings with all parties at the institution (teachers, administrative staff, students) provides the opportunity to explain to everyone what an evaluation is and what the CNE's role is. It is an important moment in terms of motivating everyone and, thus, for the quality of the internal evaluation.

The internal evaluation stage. The institution then has two to three months to gather together the elements that comprise the internal evaluation file. The CNE asks the institution to conduct its own evaluation. The CNE offers various tools to help carry out the self-evaluation. Initially, institutions and those in charge of them were asked to supply quantitative information and reports on strengths and weaknesses. Then, the CNE proposed an internal evaluation guide listing the main questions for which institutions should be able to produce answers. Now, the CNE asserts its expectations in the form of references and expects institutions to demonstrate that they satisfy these. The objectives remain the same – to encourage the development of internal evaluation on one hand, and gather the information required by the CNE and its experts to design and satisfactorily carry out the external evaluation step on the other.

In parallel, the general secretariat (project manager, documentalist) gathers together the available information on the institution (statistical data, institution project, contract, etc.).

The external evaluation. This stage has two essential functions. One is to assess the quality of the internal evaluation and the other is to take an outside view of the institution. The main goals of the external evaluation are decided on by the evaluation steering group. Here, again, the CNE has modified its approaches over time. Initially, it sought to cover all the activities of institutions. The current trend is to select specific themes that are directly related to problematic issues for institutions or the problems highlighted by the internal evaluation. After examining the internal evaluation report, those in charge of the evaluation determine the themes for appraisal and choose the experts (10 on average), who are then officially nominated by the president of the National Evaluation Committee. These experts are mainly from academic and research environments, but also from administrative sectors and economic milieus. Comparison of the image of the institution from pre-existing information, what it says about itself and what the experts or steering group will understand from on-site visits is an essential element in the evaluation dynamic.

The role of the experts in preparing the external evaluation. The experts are involved in defining the goals of the evaluation. After accepting to take part in the evaluation, they each receive the institution's internal evaluation report and are invited to participate in a meeting at the CNE's offices (first experts' meeting). The aim of this meeting is to take stock of what each person knows about the institution and enables the external evaluation project to be fine-tuned. Those in charge of the evaluation clarify the expectations of the Committee and define the issues associated with each theme with the experts. Finally, each expert receives a specific "lettre de mission" (detailed description of the appraisal to be performed by the expert) specifying the field of investigation assigned and the general expectations of the steering group. This is a global framework leaving significant room for manoeuvre to the experts, who are asked to remain particularly attentive to anything they may discover or understand on the ground. In some cases, experts may be provided with an extremely detailed plan of the appraisals expected of them.

The evaluation preparation step ends with a meeting of the steering group with the institution's management team to discuss the specific objectives of the evaluation and the methods implemented. This meeting is not always organized, but does generally take place. It is becoming a key moment in the evaluation during which the CNE and the institution conclude their

agreement on the specific objectives of the evaluation process. The move towards selective external evaluation makes this agreement necessary.

Conducting the external evaluation. The central moment of the external evaluation is the visit of the experts and the steering group to the institution. In agreement with those in charge of the evaluation, the experts choose who they wish to meet. The appraisals, organized by the project manager and the institution's administrative departments, generally take place over three days. The experts meet those in charge of the institution, with teachers, researchers, administrative staff, students and external partners. These interviews are usually individual, but can also be group interviews, depending on the experts' requirements. The Committee members also have an agenda of meetings and visits. Their work is usually oriented towards activities linked to the institutions' partners (local authorities, economic milieus, etc.) If necessary, additional on-site visits are organized.

The fact that all those participating in the evaluation are present on the ground enables the team to meet at the end of the day to take stock. These meetings provide an opportunity to compare points of view and to share information gathered with the group as a whole. In some cases, this may lead to meetings that were not initially planned.

The role of the experts after the visit to the institution. Each expert has a month to prepare a report of approximately fifteen pages on the themes entrusted to them. These reports remain confidential within the evaluation team. Each expert receives the reports submitted by his or her colleagues. Based on this exchange of information, the experts meet at the CNE (second experts' meeting) to compare their points of view and achieve consistency in their analyses in order to highlight the main specific characteristics of the institution. This is an opportunity to check points of convergence in the analyses, or to measure the different dynamics identified by the experts. One of the objectives is to determine the outline for the report and the type of report required to express the evaluation conclusions. This aspect of the procedure is particularly important for the personal involvement of the experts and the quality of their work as a group.

The preparation of the preliminary evaluation report. The evaluation steering group prepares the evaluation report using the experts' contributions. In some cases, the experts are asked to participate in the drafting process, individually or in small groups. Once it is finished, this preliminary report is sent to all the experts, who meet for the last time at the CNE (third experts' meeting) to discuss the recommendations that could be made to the institution.

Once finished, the preliminary evaluation report is sent to all members of the CNE to be debated in a plenary meeting. This is when the recommendations are discussed and finalized. This discussion is particularly important, especially since the reports are published under the auspices of the Committee.

The final discussions with the institution. Once it has been debated and approved by the full Committee, the draft report is sent to the president of the institution evaluated without the conclusions and recommendations. During a visit for discussion, the National Evaluation Committee collects the comments from the institution heads, and these comments may be taken into account when writing the final report. The final report, approved by a plenary session of the National Evaluation Committee, is sent with the conclusions and recommendations to the president of the institution, who is invited to draft a response.

The publication of the report. The report is published, with the response from the president or director of the institution. Between 500 and 1,000 copies are printed. Approximately 400 copies are immediately distributed to officials at the Ministry (Cabinet, directors), and to the members of parliament and local authorities concerned. A short document (4 pages), called a *Profile*, presents the institution, the main points of the evaluation and the main recommendations of the National Evaluation Committee. Both can be viewed on the Committee's Web server (http://www.cne-evaluation.fr)

The duration of the evaluation procedure. The entire procedure, from decision to publication generally lasts between 12 and 14 months.

Site evaluations

Site evaluations are evaluations that are particularly meaningful in the context of declining student numbers, a trend which began in the middle of the 90's, and the need to optimize public spending on higher education. These are evaluations that essentially cover the relationships between several institutions located in the same geographical area (usually a city or region), or problems that they have in common.

These evaluations naturally lead to reflection on the autonomy of institutions and their working conditions in the context of increasing competition between institutions. This work is also an ideal moment to tackle questions concerning the relationships institutions have with their environment and the political authorities in particular. The recent re-launch of the decentralization process makes these evaluations even more relevant.

Cross-cutting evaluations

Cross-cutting evaluations of disciplines are characterized by their systematic and isochronous dimension. An entire field is evaluated (e.g. professionalized applied mathematics courses; basic training for lawyers, etc.) on a national scale in one fell swoop. These evaluations provide the opportunity to establish comparative value judgments.

The common characteristics of these evaluations are the implementation of a large number of appraisals and the setting up of a steering group with additional expert consultants to ensure control of the process. The question of method is particularly important when the goal is to compare different entities. The definition of relevant criteria and terms of reference for the assessments is one of the essential stages in the procedure so that the remaining work can be conducted similarly to the institution evaluations.

The question of relations with those in charge of the entities evaluated is a delicate one. In this regard, the CNE feels that further explanations have to be provided to the academic community concerning this field of comparative evaluation. However, it is clear that this work, though long and resource-hungry, is particularly useful for all partners, students in particular.

Hungary – Higher Education

Tibor R. Szanto
Secretary General
Hungarian Accreditation Committee (HAC)

Evaluation of applications for professorial appointments

The evaluation of applications for professorial appointments is a rather new and, as far as I know, an *internationally unique* evaluation activity of the HAC. Therefore, it seems to be promising as far as the elucidation of the relationship between values and method is concerned, since the influences of historical factors and international patterns are not too strong here. Another reason for choosing this activity as subject of a case study has to do with the fact that other prominent activities of the HAC are, from a methodological-methodical point of view, presently in transition. Special ad hoc committees are working on the procedural and methodical details of the second cycle of institutional accreditation and the parallel evaluation of study programs within a discipline.

Object and purpose of evaluation of professorial applications

The number of students entering higher education in Hungary tripled in the 90'ies.[1] That fact obviously involved a need for growth in the number of instructors, too. However, due to other developments implied by the transition of the state driven "socialist" economy to a market oriented capitalist one, the latter growth was subject to some unfortunate constraints, especially in a qualitative sense. Diminishing resources, poor state financing of both higher education and research, together with lucrative incomes and positions offered by the emerging business sector, drew young, talented people towards the latter. The pool of potential university and college professors tightened while more and more new appointments were made.

Realising the quality problems of the professorial appointments, the Ministry of Education initiated changes in procedures. Though the HAC was very much against it, the 2000 amendment of the HEA prescribed that the HAC had to express its opinion on the applications for university and college professorships on an individual basis before the actual appointments were made by the President of the Republic (university professors) and the Prime Minister (college professors).[2] The HAC's standpoint was understandable: though it also realised that there were quality problems in the appointment process, it insisted on its mandate and practice of evaluating institutions, organisational units, or programs only. Behind these official structures, there is always a group of people, and the HAC did not want to take on the burden of evaluating individual persons directly. Since the academic year 2000/01, however, it has had to do so.[3]

In 2000/01 and 2001/02, evaluations of individuals applying for professorial appointments were initiated externally, by the Minister of Education. Each institution sent its proposals for appointment to the Minister who, in turn, asked for the opinion of the HAC. This procedure was, however, successfully appealed against by some individuals in the Constitutional Court. Thus, from 2002/03, institutions themselves ask for the opinion of the HAC on applicants for profes-

[1] The total number of students including those enrolled in distance education was about 102,000 in 1990/91, while it rose to more than 313,000 by 2001/02.

[2] That has been the first and, as yet, the only case where government intentions overruled the arguments and standpoint of the Hungarian Accreditation Committee in connection with its tasks and functions.

[3] As a result, the number of actual appointments decreased considerably. Between 1998 and 2000 there were 230-250 professorial appointments yearly, while this number was reduced to 123 in 2001, and to 157 in 2002.

sorial appointments and then decide on the applications, and send their proposals together with the HAC opinions to the Minister.

Thus, the objects of evaluation here are individual persons applying for university or college professorships. The purpose of evaluation is to give advice to the institutions and the minister on the quality of individuals applying for professorships and, through this process to safeguard the quality of university and college professors in Hungary.

Values and methods of evaluation of professorial applications

Individual persons obviously differ from institutions or programs as objects of evaluation. Nevertheless, the general methodological framework of evaluation is the same here as in the other cases: peer-review. Differences are only in the implementation, the actual procedures, in the details of method. As opposed to institutional accreditation and the accreditation of operating programmes, there is no site visit (or personal interview) here. Interestingly, however, the procedure of evaluating professorial applications and new programmes to be established or launched at operating institutions are almost exactly the same! (See Figure 2 in the previous article.) The only difference is that professorial applications are discussed at the second level, not by the HAC's colleges[4], but by a special standing committee established for this particular purpose.

It is important to note that potential professors – unlike institutions and programmes –are only to be evaluated, there is no formal "accreditation" obligation for the HAC. In spite of this fact, not only the methodological framework but also its cornerstone, the threshold level (accreditation) approach, is the same as in the case of the other three kinds of objects, though actual standards, minimum requirements obviously differ here from those applied in the case of institutions or programmes.

The value-world behind the methodology of the evaluation of professorial appointments is the one that was discussed in the previous article. Here, the importance and the central role of the value of equality can be seen most clearly. Potential professors are evaluated by equals, by their peers. The academic value set, the values of science can also be grasped here, on the standards level.

In the academic year 2002/03 there were five sets of standards applied in the evaluation of applications for university professorships:

- teaching performance
- established scientific/artistic school(s)
- scientific performance and publications
- inventions, patents, or works of art (professional performance)
- activities in professional public life.

These sets were determined on the basis of the prescriptions contained in the Higher Education Act. However, while for university professors the HEA prescribes only a PhD/DLA degree (and habilitation), the HAC raised the standard and, beyond these, in many disciplines expected a Doctor of Science (HAS) degree, or a "scientific performance equivalent with that degree or title".[5] Now, of course, it is obvious that for the evaluation of scientific performance, a scientific standard (degree) should be chosen. In the actual evaluation process, however, this standard became dominant, more important than other requirements. If an applicant possessed this degree, he/she was almost "automatically" on the winning side and could be sure in advance of the HAC's positive opinion. (Though about 30-40 % of the applicants supported by the HAC

[4] On the second (medium) level of decision making the HAC has three Colleges, for Humanities and social sciences, Life sciences (medicine and agriculture), and Natural sciences and technology, respectively. Members of the colleges are HAC members grouped according to their field of expertise.

[5] The Doctor of the HAS qualification can still be earned today as a "title", though it is not a "degree" anymore.

did not possess this qualification.) And while this elevated degree was not required from applicants for college professorships, publications and scientific performance played a decisive role here too. In some actual cases they were regarded as clearly more important than teaching performance and experience, and could "trade-off" weaknesses in the latter.

The Netherlands – Higher Education

A.I. Vroeijenstijn
Senior Advisor Quality Assurance
Association of the Universities in the Netherlands (VSNU)

Quality Assessment of Physics and Astronomy Programmes

To clarify the quality assessment approach in the Netherlands, the case of the quality assessment of Physics and Astronomy programmes will be described, in order to show all the steps in the evaluation process.

The external quality assessment of the Physics and Astronomy discipline was planned for 2001/2002. It is the third time that the discipline has been assessed. It concerned the following programmes:

- Physics
- Astronomy
- Applied Physics.

Nine universities were involved, among them the 3 Technical universities.

The first step was a letter to the board of the universities involved, signalling that the assessment of Physics and Astronomy was to commence in the near future. The letter enquired whether the programmes mentioned would participate, and if there were other programmes that should be included.

The faculty responsible for the programme(s) was invited to carry out a self-assessment, according to the rules of the protocol.[1] The faculty had about 6 months to complete the self-evaluation and to send the self-evaluation report to the VSNU.

At the same time, a letter was sent to the Board of Sciences at the VSNU, inviting the deans of the faculties to propose names for a chairperson. The proposal was checked by the VSNU and, as the proposed chairperson met the requirements, he was appointed by the chairman of the VSNU. Together with the chairman of the Board of Sciences, the chairperson of the external committee looked for potential members of the committee and proposed its composition to the VSNU. The department of quality assurance then checked the proposed members' credentials against the criteria in the protocol. As a minimum, they should be independent, without conflicts of interest, have the necessary expertise and command authority among their colleagues. An external committee normally has 5 to 7 members. There always has to be at least 2 experts from outside the Netherlands. In all cases, one member of the external committee is a student. In this case, the committee comprised 5 *professional* members and one student member. The chairman came from the USA, and another member from Belgium.

The department of External Quality Assessment of the VSNU appointed one of its employees as secretary of the committee. His task was twofold: to support the external committee as much as possible and to monitor the protocol and process.

[1] *VSNU, Protocol for the External Assessment of Educational Programmes 2000-2005. See www.vsnu.nl*

When the members of the external committees were appointed, the composition was made known to the institutions to be assessed and to the Inspectorate. The secretary of the committee started with scheduling the site visits.

In the mean time, the faculties were busy with the self-evaluation. They had to send the self-evaluation report to the VSNU at a certain date. The secretary of the committee checked the self-evaluation reports to see if they were complete and critical enough. In consultation with the members of the external committee, the secretary collected the final essays/master theses from the faculties and sent them to the members for assessment. At the same time the committee members were asked to study the self-evaluation reports and assess their usefulness. The members were also asked to formulate questions for additional information.

Prior to commencing the site visits, the committee members met for 2 days at the VSNU. The main aim of this meeting was to train the committee and to explain what was expected from of them. The time was also used for formulating the terms of reference, e.g. what did the committee expect as learning outcomes of a Master in Physics? What knowledge, skills and attitudes should be achieved after the study? It is important to have these terms of reference ready prior to the site visits, as these contain the domain specific, or subject specific, standards as seen by the external committee.

On the second day, the committee discussed the self-evaluation reports, shared out tasks and prepared questions for use during the site visit.

As mentioned, in the Netherlands, one and the same committee assesses all similar programmes. In this case, the committee for physics had to carry out 9 site visits.

A site visit normally takes two-and-a-half days. It starts with a meeting of the committee to prepare itself for the visit. This is also the time to discuss and endorse the assessment report of the preceding visit. In the evening, there will normally be a dinner with the board of the university and the faculty. The next day there will be interviews with the writers of the self-evaluation report, the board of the faculty, staff, students and supporting staff. The committee has, more or less, the same programme the next day. It also looks at facilities, materials used, minutes of important committee meetings, etc. At the end of the two-and-a-half day visit, the committee will formulate its first findings and give a preliminary report to the faculty.

The secretary will produce a draft report to be discussed at the next visit. Following endorsement by the committee, the report will be sent to the faculty for comment. It is up to the committee to decide what to do with the comments.

When all the visits are concluded, the committee will write the final report. This report not only includes the faculty reports, but also a comparative overview of all programmes being assessed and a statement of current practice of the discipline/subject at the time of evaluation. In the case of Physics, the report also contained some benchmarking with programmes in the USA and Europe.

The report will be sent again to the faculty to be checked for factual inaccuracies. After finishing the round of comments, the report is sent to the QA-department of the VSNU. A steering group will look at the report to see if it fulfils the requirements set out in the protocol. The steering group can ask the committee to be more complete and more precise, can accept or reject the report, but can never change the text of the report.

If the steering group accepts the report, the report will be printed and, on a predetermined day, the chairman of the external committee will hand over the report to the chairman of the VSNU in the presence of, in this case, the deans of the Science faculties.

The report will also be sent to the Inspectorate for the meta-evaluation. The meta-evaluation concerns two aspects:

- Is this report of the committee an adequate and acceptable report? Was the committee independent? Was the necessary expertise present? Did the committee follow the rules of the game? Does the report cover all aspects? Are the judgements clearly supported by evidence?
- The second question has more to do with the quality of the assessed programmes. The Inspectorate considers the report and, where there are serious shortcomings in the quality, will hand over yellow or red cards to the faculty, i.e. a warning that something has to be done. Any judgement concerning shortcomings is confidential.

In fact, the Netherlands already had an accreditation system, although this was informal: if you did not get a yellow or red card, you could consider yourself as accredited. However, there was no formal decision, and the criteria for the yellow and red cards were not always clear.

In the case of Physics, the general pattern of site visits and reporting has been followed. The report was published in 2002. It is waiting for the meta-evaluation.

New Zealand – Higher Education

John M. Jennings
Director
New Zealand Universities Academic Audit Unit

Cycle three audits[1]

The New Zealand Universities Academic Audit Unit is refining its audit processes to take account of the benefit gained from the first two cycles of audit.

The chief function of the Unit is to support New Zealand universities in their continuing achievement of standards of excellence in their academic responsibilities in research and teaching. In fulfilling this function, the Unit must do more in its efforts to enhance the universities' own programmes of continuous improvement. The Unit's objective with respect to Cycle 3 audits (which are to be conducted during 2003-2006) is to have produced audit reports which are acknowledged as being authoritative, rigorous, fair and perceptive and which are acknowledged by the universities as being of assistance to them in improving their own programmes of continuous improvement of quality and added value.

It is clear from a review of audits carried out in Cycle 1 (institutional - 1995-1998) and Cycle 2 (research-related matters including the research-teaching nexus – 2000-2001) that the greatest benefit has been the development and enhancement of the universities' own quality assurance processes and systems. The universities themselves are now much better equipped to maintain their own programmes of quality improvement, and the work of the national academic audit agency in future must take this into account if its audits are to add value.

A weakness in the two cycles of audit in New Zealand has been in the post-audit follow-up in which the focus of the Unit has been on the universities' attention to the audit report recommendations. This has become more of a problem after Cycle 2 arising from the fact that many of the universities identified proposed quality improvement initiatives in the areas of the audit focus in their self-review portfolio. The great majority of audit report recommendations, on the other hand, sit alongside those initiatives and often do not relate to those initiatives. Thus any follow-up report which reports progress on the audit report recommendations only and which ignores progress on the university's own quality improvement initiatives has the potential of divorcing audit from reality. There is the inherent danger of the audit process leading toward a situation where a university will continue to talk to itself about the issues that really interest it - that is, its own initiatives - and will talk to the Unit only to the extent that is necessary to keep the Unit happy.

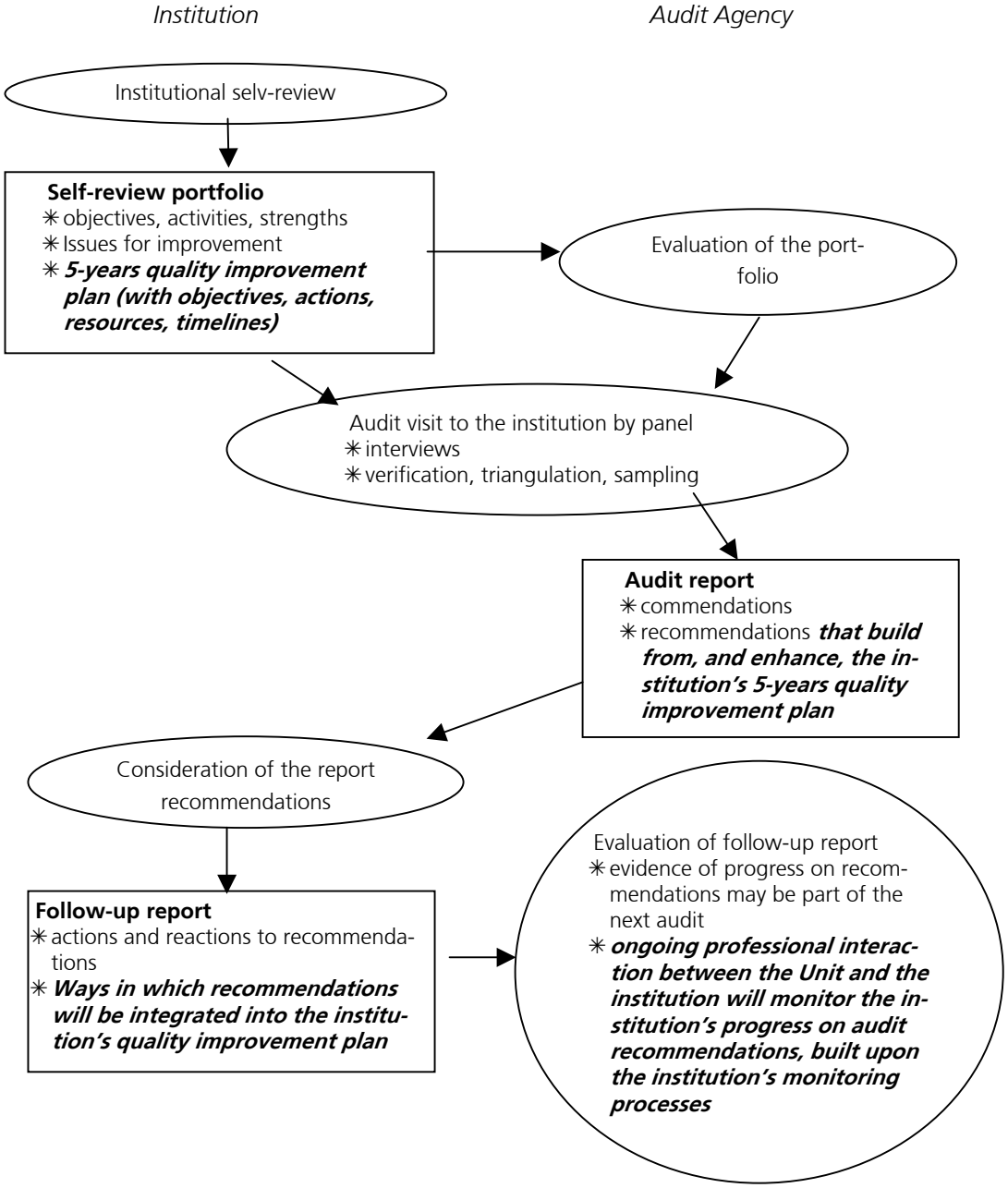
During the first half of 2002, the Unit carried out a planning process, which involved consideration of recommendations made by an independent panel that had reviewed the Unit towards the end of Cycle 2. In addition, the Unit conducted a round of discussions with quality personnel in the universities, auditors on the Unit's Register of Auditors, representatives of governmental educational and audit agencies, and the Board of the Unit. Arising from these discussions, the Unit prepared a plan to respond to the challenge the Unit had set itself – namely, to produce audit reports which achieve the objective of being acknowledged as authoritative, rigorous, fair and perceptive while finding ways to support the universities as they build on

[1] *This case study is based on a paper presented to the Australian Universities Quality Forum in Melbourne June 2003*

improved quality improvement systems and for the audit process to capture, support and enhance the University's own initiatives and energies.

For cycle 3, changes will be made in four areas in particular - the self-review portfolio, the audit report, the follow-up report, and the post audit activities. A diagrammatic representation is given in Figure 2.[2]

Figure 2.
Changes to the evaluation process in New Zealand intended to improve support for universities' own programmes of quality improvement



Note: This figure is a diagrammatic representation of improvements explained in the text. The practice followed in audits conducted in Cycle 1 (1995-1998) and Cycle 2 (2000-2001) is shown in normal font; the intended changes to present practice are shown in ***bold italics****.*

[2] *The Unit will operate Cycle 3 audits in accordance with the procedures set out in John M. Jennings (compiler),* Academic audit manual for use by the New Zealand Universities Academic Audit Unit December 2002, *Wellington, the Unit, 2002.*

The traditional expectations of the *Self-review portfolio* include a statement of goals, objectives, strategies and associated processes in place, an evaluation of progress identifying positive progress towards objectives, strengths and the effectiveness of tools used to measure progress, and areas of weakness requiring enhancement. Future portfolios will be required to include a five-year plan, which identifies the areas and topics, the proposed enhancement activities, the responsibilities for those activities, the expected outputs and outcomes of those activities, the resources to be allocated, the timeline for activities, and the mechanisms to monitor the effectiveness of activities.

The *Audit report* will follow its traditional pattern and include discussion of findings in each area of the audit, with commendations of good practice and recommendations for enhancement. The recommendations in future audit reports, however, will begin from the university's planned enhancement programme and will support the university plan where it can. Where, following its investigations, the audit panel reaches conclusions different to those of the university's planned enhancement programme, the panel will make recommendations that will enhance the planned activities, prioritise them, and/or set out an alternative plan (or part of the plan) preferred by the audit panel. While beginning from the university's planned enhancement programme, an audit panel must also be able to write whatever it believes it must write if the audit process is to achieve its objective.

Besides describing the actions and reactions to the audit report recommendations, the *Follow-up report* by the university will report on the ways the recommendations will be integrated into the institution's quality improvement plan. If the audit report is very critical of the university's plan, the Unit will expect the follow-up report to detail the ways in which the university will refashion its improvement programme to incorporate the recommendations. The aim is to give stronger ownership to the university of the quality enhancement activities following the audit process by endeavouring to have the audit panel's recommendations embedded into the university's own quality improvement programme.

The evaluation of the follow-up reports will take place through *Post-audit activities* that build a professional working relationship between the Unit and the university. As part of its quality assurance processes, the university will want to monitor progress on responding to the recommendations and will want to monitor the effectiveness of changes and improvements made arising from those responses. The Unit will encourage the university to use such periodic reviews of progress as the basis for ongoing reports to, and/or discussions with, the Unit so that the process of post-audit activities is not divorced from the institution's own quality improvement programme. When the Unit receives reports related to post-audit activities, the Unit will assess them and give feedback to the university as part of the ongoing professional interaction between the Unit and the university.

These refinements are being introduced in response to experience in New Zealand. Thune and Harvey, amongst others, have observed that experience suggests that continuous improvement from audit depends on the extent to which the linkage can be made from external audit to the aspects characteristic of internal institution-based quality improvement (Thune 1997:102)[3] and how the outcomes of audit are communicated and linked with the day-to-day research and teaching activities of academic staff (Harvey 1998:242)[4]. Universities cannot do this alone, and audit agencies must heed this advice and refine their audit processes to facilitate a linkage so that quality enhancement of a university's core activities do, in fact, materialise.

[3] *Thune, C. (1997) The balance between accountability and improvement: the Danish experience. In J Brennan, P De Vries & R Williams. Standards and quality in higher education. Higher Education Policy Series 37. pp.87-103. London: Jessica Kingsley.*

[4] *Harvey, L. (1998) An assessment of past and current approaches to quality in higher education, Australian Journal of Education, 42(3), 237-255.*