UNIVERSITY AND EVALUATION: A CASE STUDY OF THE ORGANIZATIONAL IMPACT[*]

Eliana Minelli, Gianfranco Rebora, Matteo Turri

Introduction

The evaluative state [Neave, 1988; Pollit, 1993], the regulatory state [Osborne - Gaebler, 1992], the audit society [Power, 1997] have become diffused labels to mean the strengthening trend toward measuring and controlling results of publicly funded activities.

In most European countries, different forms of evaluation or quality assurance have been introduced and gradually reinforced in the field of higher education and specially of universities [Brown, 2004, Turri, 2003]. National agencies have been established in many countries in order to conduct evaluation of academic output –although in rather different forms [Felt, 2004]. These kinds of practices assume the shape of a quest for efficiency from the above, different from "professional development model", more oriented toward quality improvement and effectiveness and managed in a particular institutional and organizational setting [Pollitt, 1987; Barzelay, 1996].

Evaluation imposed and sponsored stated above was quite naturally combined and integrated with performance management and other managerial practices, according to the lessons of new public management movement, mainly in countries like England and Sweden which have been the forerunners of this approach.

This bundle of initiatives, containing evaluative and managerial practices, was often perceived by academics in a not serious way; and a wave of criticism appeared in the last years, with strong voices also in the scientific debate, claiming and advocating for the defence of institutional autonomy, academic freedom, university communitarian life and traditional way to promote knowledge [Henkel - Little, 1999; Brennan – Shah , 2000; Dillard, 2002; Davies – Thomas, 2002; Singh, 2002; Lay, 2004].

Perhaps a part of the reactions moved away from ideological reasons, or merely from academic conservatism, but at least one important component of this current of thought was not adverse to recognise the quest for accountability, for value for money and for high service

quality standard, which European universities must receive in order to legitimate themselves in the present phase of transformation [Hoecht, 2004]. A problematic and risk conscious opinion about evaluation and performance management experiences concerning European universities is shared by representative bodies of the academic system and also by scholars engaged in the process of university reform and also taking part in the evaluative activity [Felt, 2004; Brown, 2004, Brook, 2000].

During 2004, a very relevant number of papers containing this sort of reflective criticism was presented in international conferences concerning evaluation or university reform [EIASM, 2004; EES, 2004; Toulon-Verona Conference, 2004]. It is not the situation of quality improvement to be challenged or to worry about, but the kind of quality coming out from many of the initiatives which are in progress, notably the limits and defaults of current application of new public management practices and techniques in the university context [Hoecht, 2004; Brown, 2004; Krogstrup, 2004].

Our article aims to propose a frame of reference for analyzing evaluation and quality assurance schemes of university core activities, teaching and research. This frame takes into account the ambiguity and the need for interpretation inherent to the evaluation of the core academic activities, research and teaching, and their quality. After having explained it, we use this same frame in order to study the impact of the evaluation carried out in the case of a medium sized Italian public university.

The principal focus consists in the empirical study of the effects of evaluation practices. This subject is rarely treated and discussed in the literature concerning evaluation and also performance management in the public sector [Underwood, 2001; Brennan - Shah, 2000]. In Italy there are not any relevant contributions on this matter.

The systemic structure of evaluation or assessment processes

A formal system of evaluation may take different shapes. In a previous study, we elaborated a prospective framework for the analysis and the approach to planning the evaluation systems, set to perform both at advanced service systems level and of single institutional agent [Rebora, 1999; Minelli – Rebora - Turri, 2002].

The proposed model takes into account the following variables: the idea or concept of assessment, the applied methods, the bodies in charge of evaluation and the way assessment is used. The joint action of these elements, characterized by a more or less high level of coherence, determines the institutional and organizational impact of evaluation.
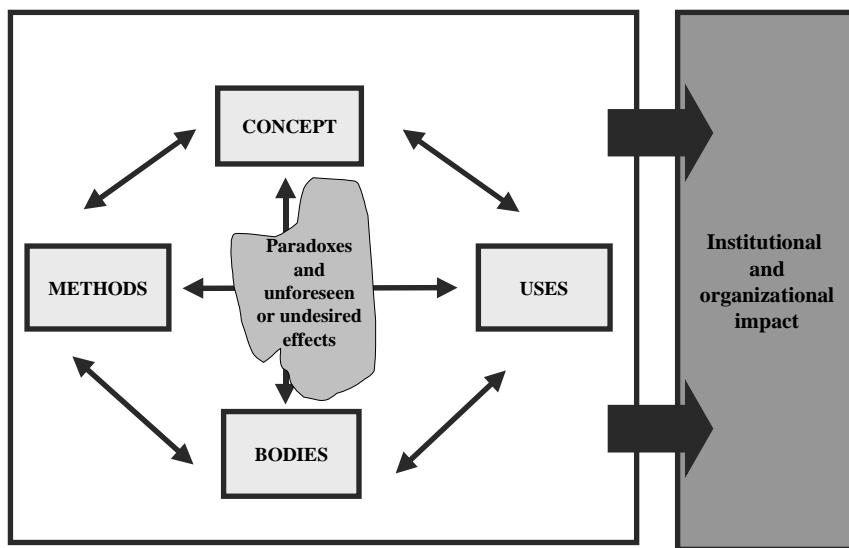
Figure 1 - The systemic Structure of evaluation processes [Rebora, 2003]

The *evaluation idea* corresponds to the official concept coming out from constitution acts, from published documents, programs and reports, or also from formal declarations and speeches expressed by the university leaders, or by the members of government boards and of evaluation bodies and boards. Then, the idea or concept reflects only the ideal and publicly declared purpose, content and process of evaluation, which eventually, and probably, differ from the theory- in-action, the concept which the real practice of evaluation reveals.

The *assessment and evaluative methods* are the specific instruments and techniques which give substance to the activity and process of evaluation. A large variety of applicative approaches and connected instruments is by now available, in consequence of a fast developing trend of disciplines, schools, frameworks of thought which are active in that matter. In the case of university, this kind of instruments not only respond to general methodological rules for proper evaluation, but also have to meet requirements coming from specific activities, namely research, teaching and other services.

The enforcement of whatever evaluation method in the context of complex organizations, like modern universities, asks to be put in charge of evaluations of specific *bodies or agencies*. These kinds of subjects, normally constituted with a professional base, are responsible for a set of tasks used in order to perform evaluation and for the drafting of conclusive reports addressed to governing bodies and eventually to other stakeholders.

The primary value of evaluation concerns the nurturing of dialectic between different subjects, inside and around the involved organization. Then it is necessary to activate bodies which are technically responsible for the entire evaluation process and assume a different role regarding subjects who promote, purchase, utilize the evaluation or have a stake in it.

Evaluation bodies play a fundamental role just in the process of setting this kind of dialectics, which involve their relationship with promoters, purchasers and other utilizers. Specific traits of evaluation body's members assume a strong relevance, mainly technical competence, professional autonomy and independence, organizational status as internal member or, on the contrary, as external practitioner.

As regards to *the way assessment is used*, our model distinguishes three basics modes which evaluation reports or results are utilized for. First: promoting rapid change by sanction, incentive and formal guaranty applied to desired behaviour at the individual, collective and organizational level. Second: convincing and motivating people to work better and then improve the quality. Third: promoting knowledge and eventually obtain some form of learning and influence future decisions.

The first way elaborates information in order to facilitate decisions about programs continuation, expansion, reduction, elimination. This kind of use confers a threatening role to the entire evaluation process, which influences the whole chain of relations. In the second case, when evaluation is used in a logic of improvement, a quite opposite framework appears: different subjects, directly involved in the process, are expected to cooperate in order to improve each other's performance. In the third case, the use is oriented to better knowledge about everything is perceived as important for the organizational activity: practical and immediate implications can be absent, but a concern for individual and collective learning is alive.

The exposed layout of variables confronts the concept of assessment which raises from formal documents or the purposes shown by the leadership with the factors bounded to the real assessment practice in place like the implemented methodologies, the nature and composition of responsible bodies and the ways evaluation outcomes are used. The last three variables are considered the most important features of the evaluation-in-act [Rebora, 1999]. The alignment among the four essential variables which structures the system, namely concept, methods, bodies and uses, is then valued as an important feature which conditions the efficient application of the entire construct.

However, the model also considers the impact of the evaluation practice on institutional and organizational framework of reference. In this way we take into account that the overall effect depends not only from the exposed factors, which are component of the evaluative system according a formally rational view, but also from unmanaged elements which can disarrange the connection of causal relationships formally set, or the expected outcomes: the complexity of institutional and organizational environment and connected service systems often give rise to paradoxes and unexpected effects, frequently nourished even unintentionally by opportunistic behaviours of different players [Stame, 1990; Power, 1997; Vedung 1997]. The alignment

among the four essential variables, which structure the system, namely concept, methods, bodies and uses, represents a crucial condition in order to achieve a positive impact on institutional and organizational design, but doesn't guaranty totally against the emergence of unexpected effects.

On the other side, the lack of coherence among the four kinds of variables ends up undoubtedly in a disarray inside the institutional sphere (that is inside the relationships among different players and stakeholders) and the organizational environment (that is the inside coordination and the proper distribution and implementation of tasks) at the same time. These lacks of alignment not only contrast the efficiency of evaluative system, but opens a greater space for opportunism and unexpected effects.

A case study of an italian university: analyzing the impact of evaluation

Our case study concerns a medium size public university located in northern Italy: the University of Trento, which has a body of 460 professors and researchers, 470 administrative employees and 14,500 students. Six faculties are present and 12 research departments. This university in 2003 had an annual budget of 96 million euros.

The choice to study the experience of this university was motivated by two reasons. First: Trento University from 1998 to 2003 performed the evaluative practices requested by national bodies with full engagement and commitment, publishing annual reports very well elaborated, containing broad information and in depth analysis: a first class practice, compared with other Italian state universities, at least according to the published elements. Second: Trento University in the same period performed also some autonomous practices of evaluation, notably in the research field. Then, the choice of Trento can guaranty the presence of a real and in-depth experience of evaluation during many years, sustained by a serious commitment of the governing bodies.

The data collection process started in November 2003 and ended in the late February 2004. The study was specifically oriented to understand the practice of evaluation in this university in the period of the last five years and the consequences of this activity. Primary sources included semi-structured interviews with people in the key roles of the athenaeum's organization (vice chancellors, deans, research department heads, members of evaluation board, general and administrative managers). On the whole, 26 interviews were released, for a total of nearly 40 hours. All open answers were registered giving rise to more than 200 pages. Secondary sources included documents from internal archives, annual reports and others internal reports, research

reports, minutes of evaluation board meetings. Some direct observations were also carried out during internal meetings and public conferences.

The analysis of assessment activities has been traced to a systemic vision according to the framework shown in the previous paragraph. On the basis of the concepts exposed above, we can summarize the features of the examined assessment practice:

- *The concept of evaluation*: it proves dichotomic. Some elements make it technically advanced and sophisticated, differentiated according to the diverse sorts of activity, namely research and teaching, available to the receiving organizational needs and priorities. On the contrary, other considerations, specifically the lack of uniqueness of concept, lead to make evaluation uncertain and contradictory.

- *Assessment methods*: they are various and well developed. Our analysis collected more than 30 evaluation tools. They are heterogeneous and related to very different basis/principles. As regards to research, the *peer review* by foreign scientists takes a notable interest in order to assess the departments.

- *Bodies in charge of evaluation:* they are two, the Board of Evaluation and the Commission for Scientific Research. A clear task-sharing is still missing. The Board of Evaluation, made up mainly by internal members, shows broad tasks, a disposition to improvement, an effective professional support and a strong autonomy. The Commission for Scientific Research is an autonomous and influential body, with broad tasks, performed with the support of the Athenaeum's administration. Both of the bodies are enjoying a favourable condition (like a state of grace), which allegedly depends on positive circumstances instead of reliable organizational requirements.

- *Uses of evaluation*: they are different with respect to didactics and research. Teaching activities make use of assessment specifically to improve quality and quantity of information about processes. The use of evaluation is soft. This approach may express indifference about evaluation or perhaps has value in order to encourage people to adapt and improve their behaviour without constraints. Research activities assessment allows the Athenaeum's governing bodies and departments to foster accountability. This kind of use supports the decision making process and particularly the allocation of economic resources.

The short description gives the portrayal of a rich evaluation system but not without contradictions: a crossroads of excellent experiences, shown by the extent of implemented methodologies, and of ambiguous situations, like the dichotomic nature of the concept and the lack of coordination between the bodies in charge of evaluation.

The evaluation impact has been analyzed mainly through open questions during the interviews even if, first of all, we show the outcomes of the question asking "the three most important assessment tools implemented in the last years". The analysis of answers allows catching an overview on the main evaluation activities implemented. As presented in the following (Tab.1), this is crucial to the understanding of impacts of the evaluation system.

Table 1 – Evaluation impact. Results from the question "The three most important assessment tools implemented in the last years?"

Tools	Weighted relative frequency
Peer review of research departments	27%
Activities of Evaluation Board (various or not specified)	20%
Activities of Evaluation Board (analysis of students' satisfaction)	12%
Activities of Evaluation Board (analysis of data base of publications about scientific productivity of professors)	6%
Quality system "CampusOne" (promoted by Conferenza dei Rettori delle Università Italiane).	3%
Panel of international experts (evaluation of Ph.D. actuation – research politics)	2%
Activities of Evaluation Board (teaching)	2%
Activities of Evaluation Board (assessment of teaching and of students' career with CATI technique)	2%
Teaching Faculty (various or not specified)	2%
Evaluation of administrative and technical personnel	2%
Teaching Faculty autoevaluation (final reports)	2%
Teaching Faculty autoevaluation (direct interviews on teaching supply)	2%
Activities of Evaluation Board (research)	2%
Data extraction from the information system of athenaeum (SAP)	2%
Management control system	2%
Activities of Evaluation Board (administration)	2%
Review and follow – up of Institution Evaluation Program – European University association.	1%
Activities of Evaluation Board (annual report)	1%
Activities of Evaluation Board (data analysis on students' characteristics and professional opportunities by AlmaLaurea)	1%
Teaching Faculty multiyear development planning	1%
Activities of Evaluation Board (appraisal on finding funds for research by departments)	1%
Workload analysis	1%
Activities of Evaluation Board (full costing analysis of teaching and research)	1%
Activities of Evaluation Board (analysis of graduating students' questionnaire outcomes)	1%
Research department board for publications	1%

The relative weighted frequency of interviewees' answers (which are weighted 1 if they refer to the first evaluation tool pointed out, 0.75 to the second and 0.5 to the third) seems to be focused on four instruments: peer review of departments, assessment activities implemented by the Evaluation Board, analysis of student's satisfaction and analysis of scientific productivity of teachers by looking into the database of publications. Other twenty-one tools are indicated, which show little relative frequency.

Before refining the analysis by gathering the identified tools in meaningful clusters, we can draw some preliminary thoughts:

- the extent of methodologies available to assessment encompasses all areas of academic activity with specific tools, from research to administration;
- the primacy of peer review as an essential evaluation tool, which is implemented in the athenaeum under consideration, points out the centrality of research evaluation, differentiating that athenaeum from other Italian universities, none of which has developed to this extent the research evaluation through peer review, and stresses the importance of professional aspects in assessment of university activities;
- the size of tools and assessment activities implemented by the Board of Evaluation supplies the critical mass to the assessment endeavour of that university (the other three evaluation tools beside the peer review are promoted by the Board of Evaluation).

Subsequently, the evaluation tools, pointed out by the interviewees, are analyzed on the basis of the subjects promoting assessment activities (Tab.2).

Central governing bodies of athenaeum and administration use a set of tools like CRE reports, workload analysis, data extraction from the information system of athenaeum (SAP), evaluation of administrative and technical personnel, management control system.

The Commission for Scientific Research, supervised by the Deputy Chancellor, promotes a panel of international experts, peer review, appraisal on finding funds for research.

The Evaluation Board (whose activities are related to didactics, research, administration) implements students' satisfaction analysis, annual analysis of professors' publications from data base, full costing analysis of teaching and research, analysis of teaching and of students' career by means of CATI technique, analysis of questionnaires filled in by final-year university students, analysis of data from AlmaLaurea database, and makes an annual report.

In addition, there are some self-assessment practices promoted by services in which the assessed processes take place: quality assessment, according the Italian national practice named "*Campusone*", faculty reports, faculty multiyear development planning, direct interviews on teaching supply, faculty self assessment system, department board for publication.

The analysis takes into account, besides the overall values referred to the 26 interviews[1], also the partial outcomes, divided according to the main categories of interviewees, members of the Evaluation Board, directors (deputy chancellor, vice chancellors, general manager, administrative managers), deans and department heads.

Table 2 - Analysis of evaluation tools indicated by interviewees (distribution on the basis of subject promoting the assessment)

Promoter	Implemented tools of evaluation	Frequency per group				Whole sample
		board	directors	deans	head	
Athenaeum central government and administration	Workload analysis Review and follow – up (EUA) Evaluation of administrative and technical personnel Data extraction from SAP Management control system	4%	9%	4%	8%	**8%**
Deputy chancellor-Commission for Scientific Research	Peer review of research departments Appraisal on finding funds for research by departments Panel of international experts	16%	47%	24%	33%	**30%**
Board of Evaluation	Evaluation Board (research) Evaluation Board (teaching) Evaluation Board (assessment with CATI technique) Evaluation Board (various or not specified) Evaluation Board (analysis of students' satisfaction) Evaluation Board (productivity of professors) Evaluation Board (administration) Evaluation Board (annual report) Analysis on students' career by AlmaLaurea Evaluation Board (full costing analysis) Evaluation Board (analysis of graduating students' questionnaire outcomes)	80%	35%	52%	50%	**50%**
Self-assessment	Quality system "CampusOne" Research department board for publications Teaching Faculty multiyear development planning Teaching Faculty (various or not specified) Teaching Faculty final reports Teaching Faculty Direct interviews on teaching supply	0%	9%	20%	8%	**12%**
		100%	100%	100%	100%	**100%**

The interviewees indicate the evaluation tools implemented by the Board of Evaluation at 50%, by the Commission for Scientific Research at 30%, self-assessment at 12% and activities promoted by the central government of athenaeum at the remaining 8%.

The analysis according to the different kinds of interviewees shows that the heads of departments' opinions do not diverge significantly from the overall average; this is true also for the deans, who are nevertheless susceptible to the importance of self-assessment tools, which often are implemented in faculties.

The directors shed light on the evaluation tools promoted by the Commission for Scientific Research (which they are member of), giving them a 47% percentage and containing at 37% the relevance of activities promoted by the Evaluation Board. The members of the Board diverge heavily from the overall average and indicate Evaluation Board's activities as central (80%).

If the concurrent presence of two important players in assessment comes up at a first glence, the Board of Evaluation and the Commission for Scientific Research, a deeper analysis reveals the attitude of promoters to point out their own tools (this is the case of Evaluation Board about didactics, directors about research assessment, or deans about self-assessment).

Generally speaking, all this leads us to believe that:

- the perception of assessment outcomes (thus their likely employment) seems to be suffering from the player implementing the evaluation tools;

- there is a clear orientation to give larger credit to the evaluation efforts whose development players have been involved in actively; thus the issue of participation and sharing of assessment become central;

- besides a greater attention to one's own evaluation tools, there is a tendency to interpret less favourably, and even to oppose to evaluations whose promoters are different.

A deeper analysis is carried out through open questions. The following table (Tab.3) exhibits, in short, the main effects revealed by the interviews, breaking down the answers into classes of interviewees. We prefer to show the outcomes distinguishing didactics from research in order to give account of the differences in the use of the methodologies of assessment. The normal effects are marked with one "+", while the most significant ones are given a "++" mark.

The observed effects are arranged in classes according to the categories developed inside the change management studies [Rebora, 2001]. We can single out effects concerning organizational learning, resources development, power management. Beside these effects, we can pinpoint inertial factors which weaken the change outcomes.

EFFECTS		TEACHING				RESEARCH			
		Administrative managers	Evaluation Board	Deans	Research directors	Administrative managers	Evaluation Board	Deans	Research directors
ON ORGANIZATIONAL LEARNING	Increase of knowledge	++	++	++	++	++	++	++	++
	Transparency of processes		+				+		
	Rise in internal competition							+	+
	Enhanced attention to evaluation	+	+	+		++	+	++	++
RELATED TO RESOURCES DEVELOPMENT	Influence on economic resources allocation (inside the athenaeum)					++	+	++	++
	Standing improvement				+				+
	Reinforcement of university transformation		+	+			+	+	
RELATED TO OWER SYSTEM	Consolidation of bodies in charge of evaluation		+				+		
	Consolidation of assessed bodies								++
	Enhancement in decisional capabilities of athenaeum's governing bodies					++	+	+	+
	Enhancement in decisional capabilities of research departments					+			++
FACTORS OF INERTIA	Increase in paperwork				+				+
	Development of defensive behaviour among academics	+				+			

Table 3 - Effects from evaluation

Effects related to the organizational learning. These effects concerning the learning process are not limited to the development of concepts and knowledge but more widely they include the capability to comply with new conditions in terms of context understanding and behaviour.

As a result of the evaluation exercise, knowledge about teaching and research processes have increased sharply, giving rise to a deeper awareness of what going on in the athenaeum among all organization level. The interviews to the members of the Evaluation Board highlighted how the transparency of the entire spectrum of academic activities as a consequence of the assessment practice has increased, thanks to the enhancement of the knowledge of processes.

Another impact related to organizational learning consists in a rise in internal competition among units and among people who are brought on by evaluation. This influence is still weak and more oriented to research: the researchers perceive a stronger pressure on their own activity, especially toward publications which are the basis of the assessment practices developed in the athenaeum.

The creation of a more competitive internal environment can be considered a factor stimulating learning, by generating positive tension to adapt behaviour, use of resources, objectives and results to new demands in a creative way. In primary instance, direct consequences not are involved in terms of behaviour, but a more generic cultural impact is relevant. It is not a compulsion with direct consequences on behaviour, rather it is a case of cultural effects which makes a department director say: <<…*in research activity, he who twiddles his thumbs will feel like a stranger, a person not involved>>*.

One more effect in the sign of on organizational learning concerns the strengthening of the evaluation itself because the diffusion of the results of evaluation processes stimulates the continuation and development of this kind of practices. A large number of people interviewed argues that the availability of evaluation reports, mainly in the case of research but also referred to teaching activity, generates demands for new information by persons motivated to deepen, to extend, even to deny the results. Evaluation as a source of learning is able to start a process which nourishes itself.

Effects on the development of resources. Evaluation may stimulate the availability and development of human and financial resources within the university and can promote the acquisition of other resources from the external environment.

Evaluation is seen as a mean to influence in a relevant way the resource allocation process concerning research. This kind of effect doesn't appear in the case of teaching. The results of the research activity exercise have been included by the athenaeum's government bodies in the number of principal standards adopted in order to distribute research funds among departments.

The directors of the research departments also recognize an effect of reinforcement to the external reputation as a result coming from the publicity of the university's research activity ranking performed by a national newspaper-- not a big effect, but enough to be noted.

Deans and members of the evaluation board see also the role of evaluation practices as a factor leading the university transformation, particularly teaching structure and process. Evaluation acts simultaneously both as a basic element for the success of university reform, contributing to responsibility assignment and to diffuse information, and as a product or a consequence of the same reform process. In fact, the reform itself requires explanation and data about the efficacy of the implementation process.

Effects on power system. Evaluation can modify power structure within the organization and can influence relations among different internal units. Some members of evaluation board note that the presence of this kind of practices contributes to strengthen the bodies which are in charge of this task.

Confined to the scientific field, the department directors highlight a substantial enhancement of the department as a unit leading research activities. Thanks to evaluation, the departments overcome the traditional vision according which, in a director's words, they are "*unions of scientific independent firms, held together with the parking*" and they gain an organizational role which is crucial to coordinate and direct research activities, getting awareness of their own governing functions.

As far as scientific research is concerned, we see a strong increase in decisional capability as a consequence of the double-level evaluation system:

- The central government of the athenaeum enhances the power to manage and steer the research system; thanks to the evaluation exercises, it acquires a global picture of the existing situations, according to which it can differentiate and even justify its interventions.

- Inside the departments, we can observe the consolidation of the role of the director, who is encouraged by the evaluation exercise to undertake actions to increase the qualitative level of research activities and at the same time is supplied with a lever, which is the result of evaluation, to legitimize his action.

Inertia factors. Besides the described effects, evaluation may generate reactions which weaken its own scope.

The first verified effect, in this sense, is bound to the widespread increase in paperwork which makes the evaluation process run the risk to reduce itself to a routine, failing the true aims for which it has been developed.

The second aspect, though lightly emerging, concerns the presence of defensive approaches to corporative interests among academics, who seem reluctant to express a negative opinion on colleagues or, at least, they seem to be more careful than they are towards administrative activities. The performed analysis puts in light a sharp caesura between didactics and research as far as evaluation effects are concerned. While as regards to didactics, the assessment limits itself essentially to spread knowledge; for research, there are dramatic effects not only related to the organizational learning, but also to the allocation procedures and the strengthening of the roles involved in the evaluation process[2]. A numerical analysis of the 42 effects (of which 25 normal effects and 17 relevant effects) shown in Table 1 supports what was observed:

- effects related to teaching amount to 14, of which just 4 are relevant. The area where assessment got the greatest impact is that of organizational knowledge, specifically in regard to the increase of knowledge.

- The research evaluation shows an overall number of effects (28) which doubles the teaching, of which 13 are relevant. Weighty consequences come up in all examined areas, while the inertial effects turns out to be marginal.

A summary of the relevant finding from the case study

The different impact of evaluation on teaching and on research is the more interesting aspect shown by our analysis. We will take back this argument in the concluding paragraph. But our study allows to identify some other elements which seem advisable to put forward:

1) first of all, a positive vision reveals how the evaluation can contribute to the change process; the inertial factors are pointed out with light influence and the positive effects are considered prevailing.

2) In all interviews, a deep sharing of the incremental logic arises, according to which the effects of assessment remain very light, just for knowledge purposes, not bound to specific organizational interventions, as far as the assessment practice doesn't reach its own critical mass, allowing the consolidation of the evaluation procedures. Only the persisting assessment gives rise to major organizational effects. At this point, we can assume the existence of a breakeven point of evaluation, before which the implementation of assessment activities did not give rise to any effect but costs while overcoming this threshold allows to take advantage of the desired effects.

3) In an organization like a university where the character of hierarchy is dissolved into the collegiality of the deliberative bodies, the evaluation takes on a central role in order to facilitate and legitimize governing actions, which find in assessment outcomes the

possibility to shield from arbitrariness charges. In this regard, *peer review* activities reveal particularly effective because they take advantage of the professional component of the academic activity [Brennan - Shah, 2001]. The loosely coupled organizational structure and the remarkable professional connotation which distinguishes any athenaeum [Reponen, 1999] make the evaluation, when it succeeds to find full expression, become a pivot of the governing process of the athenaeum, acting as an incentive and an external constraint, like a Maastricht standard, to use an interviewee's metaphor, thanks to which the managerial interventions can get a grip on the university.

4) Among the assessment tools introduced in that athenaeum, less than 10 produced direct effects. Among those which had no direct effect, there are the methodologies implemented on the Minister and CNVSU's guidelines. The successful impact of the evaluation tools seems to depend mainly on participation and involvement rather than on external mandatory guidelines.

5) The different players involved in the assessment process, with different roles and in different stages, set out a differentiated perception of the evaluation. Each of them tends to stress those tools, procedures and stages of evaluation which he was more directly involved in, or which he shared more deeply, even setting them against other aspects he had nothing to do with. This is the case of the members of the Evaluation Board's interviews, which show some elements away from other interviews (such as diffusion, transparency, reinforcement of evaluation bodies). Especially when there is a plurality of bodies, the evaluation activity should be thought not as a monolithic process but as the aggregate of different actions, each of them with its own features and purposes.

6) The evaluation exercise implies a plurality of values which the different players can decide to bring into play. The assessment may strengthen the definition of an organizational environment and thus its identity and its institutional importance, may legitimize policies and orientations, first of all facilitating their own formulation and communication, may even foster the awareness of particular issues by giving "objective" substance to aspects otherwise doomed to remain virtual or latent. Generally, the same evaluation outcomes take different features depending on the way they are used by the various organizational players.

7) As a consequence of what set forth, the political games implemented by the different players remain an incumbent and unavoidable element, even more when assessment becomes a powerful activity and is given a considerable attention. The fact that evaluation outcomes are sought and conveyed toward different players' political goals may be considered an evidence of viability and perhaps an essential requisite for

evaluation success, or however some sort of energy to feed its development; of course, the issue of preventing excessive opportunistic and distorted behaviours is still open.

8) The organizational impact of assessment is heavily influenced by players' culture; this aspect may become an obstacle, when some prior judgements or deeply rooted professional values may become the barrier to using the evaluation outcomes. In other circumstances, the assimilation of a shared idea of assessment in specific organizational environment, such as departments and faculties, profiles the likely multiplicative diffusion of evaluation effects. As a matter of fact, assessment may become a mean of cultural change if key players are able to manage the trends such a way.

9) The players are generally very sensitive to the likely feedback of assessment in terms of reputation of the whole university institution or of some of its specific components. The evaluation allows to gain an "external accreditation" and to shed the traditional prejudice (often well grounded) of self-referential approach surrounding the universities. This positive effect is perceived as dependent from its bare existence well before the favourable or unfavourable outcomes, especially if implemented using innovative procedures.

10) The last suggestion from the case under consideration regards the link between evaluation and change inside the university system. They are phenomena that should be analyzed in parallel. In fact, even if evaluation is just one of the elements of the wider process of evolution of the role of the universities, without its implementation the athenaeums will hardly succeed to give substance and permanence to the realized changes. Therefore a biunique relation takes place, which we should consider and study as such.

Outline for future research

Evidence from the exposed case study confirms a weak impact of teaching evaluation which results also in the feelings of several professors and students involved in Italy in this kind of practice. On the contrary, in the case of research, the assessment exercise appears much more involving people interested in it and affecting university's management and decisional processes. These kinds of findings invite to develop some conclusive remark, in order, above all, to trace a frame for future research work.

About the *evaluation impact on teaching*, it is relevant to consider the framework of asymmetric relationships which features this activity in the university context. Professors perform a very complex role in regards to students, which involves some aspects of authority

and which is not expected to receive an instant appreciation by students, like steering toward educational goals or verifying the learning process. On the other hand, the quality of teaching has always been held as the essential resource for the successful educational process and in this sense it is largely brought about by the teachers' quality and professionalism.

The complexity of interpersonal relationships implied in education rejects the only implementation of quality control and certification methods, which are generally applied in industry and services. These tools allow to evaluate static quality but they prove powerless to catch the dynamic quality, which is bound to the uniqueness of teaching performance, in the sense of the contribution of the most brilliant teachers in their best moments [for a philosophical standpoint, see Pirsig, 1981 and 1992].

Scholars who criticise the introduction of quality and performance management in the higher education (and other techniques of NPM) specifically accuse the practice of student evaluation of teaching or student ratings (SRT) to be "clearly designed to position students as consumers in their educational experience" [Singh, 2002]. And thus SRT is seen to legitimise the "commoditisation of students", the "reification" of the student–teacher relationship, and then a very narrow conception of the teacher's professionalism and a very limited vision of the academics' accountability [Lawrence – Sharma, 2002; Singh, 2002].

In the case of Trento this kind of exit doesn't appear to concern the university leading provost. Irrelevance rather than colonisation (see Power, 1997) seems to affect this kinds of practices that are practically imposed by external rules. However, one may put forward the doubt that the very light and soft way to use SRT practices, proper of an early phase, help to avoid the negative effects, in Trento, expected by Sing and others. More research is needed to compare the effects along different stages (early and mature) in the evolution cycle of the use of assessment practices in didactics. In addition, it is also useful to understand which role can play a conscious and careful use of SRT towards improving the teaching quality.

In this order of thinking, we propose a frame of reference supporting the tentative of a soft and intelligent use of SRT practices, based on the differentiated treatment of static and dynamic quality. The static quality identifies what we normally expect from teaching, like the development of a clear, complete and coherent programme and is also brought about by the conformity to an established framework which assures a regular learning process and also by the respect for the rules which represent a normalizing force of educational systems. It also consists in the repetition, in time, of the contents tested through experience.

Nonetheless, it is a dynamic process, also within the normality, which pushes ahead knowledge to a higher level, like a surprise which goes beyond the everyday normality and sparks the emotional involvement which facilitates the learning process. This aspect has been

recognized as dynamic quality and seems to be the outcome of versatility and linking capability of teachers [Pirsig, 1981; Pirsig, 1992].

However, both quality concepts need each other to survive and some tension between the two forces seems essential to stimulate learning and personal growth. If the evaluation process respects these dual and dialectical characteristics of teaching, it will be structured in consistency with them.

At the beginning, it is useful and necessary to assure the static quality with rules and operational systems of control applied to formal aspects of teaching, including that attainment of a threshold of acceptability, in terms of agreement or perceived quality by students.

In this respect, the approval assessment and perceived quality by students represent not just standards to abide by, or targets to maximize but also a mean of communication between students and teachers, anyway leaving to the second ones an active role in understanding and elaborating collected data.

Once the static quality has been assured through rigid procedures, but not excessively interfering, it is certainly easier to foster excellence in teaching, to implement soft tools, which imply agreement, example, involvement and participation. All these concepts, based on the experienced teachers in different ages and in different countries and on philosophical reflections, find an important empirical confirmation in researches on learning achievements, carried out especially in educational institutions in Anglo-Saxon countries. Differences in students' performance are brought about by teachers' quality [Goldhaber, 2002], which seems to have been shaped by intangible elements, often hardly definable and verifiable, like enthusiasm, the willing to pass on knowledge, the communicative and relationship capability.

The *evaluation impact on research* is affected by as much complexity and sensitivity as didactics, even if for different reasons.

In research activity, the complexity of interpersonal relationships is not greatly at stake, apart from the relationships with the customers, if existing. Better yet, the central issues seem to concern the involved competences and the orientation toward innovation and experiment, which gives something of uniqueness to any project, and also the structural tension between acquired knowledge and sources of originality. In light of these considerations, research activity seems even less assessable than didactics using static quality patterns and on the contrary it seems exposed to the sphere of dynamic quality.

The relevance of these characteristics brings the structural risk of control failure and, however, reduces the possibility to resort to quantitative indicators, which show the volume of achieved activity, even if quantitative evaluation is certainly useful, if correctly planned with recourse to multiple and balanced measures. Upon the static quality basis, there is room for

researchers' autonomy and creativity, which can be assessed with the development of methodologies like *peer review*, which imply the involvement of figures competent in the specific field, as stimulus and verifier of self assessment and critical comparison capabilities by researchers.

The examined experience of Trento University confirms this kind of remarks: the extensive recourse at the practice of peer review, performed involving internationally well-known scholars, brings guarantee for fair and serious research assessment. The rooted academic culture is used to deal with peer's opinion about research and more favourable to accept this kind of judgment coming from colleagues, than student's rating about teaching. Moreover, academics in charge of managerial responsibilities can find in research evaluation an effective instrument, which contributes to energize their role in the very participative and loose institutional context of the university, without risk of relevant conflict as happens with SRT in the matter of teaching.

The case of Trento does not give particular evidence to the risk, claimed by some literature in this argument, as rooted inside these systems, when evaluation activities exceed a threshold of intensity and of frequency. Beyond that level, increasing evaluation activity brings about a decrease both in quantitative productivity and, even more, in quality value [Brook, 2000; and also in more general terms Power, 1997 and Brown, 2004]. This effect makes clear that, in this respect, the qualitative value of research work is meant as dynamic quality. The failure of control and assessment methods happens in repeated cycle of assessment, when opportunistic behaviour, even natural and absolutely understandable, gets the success of evaluation procedures to become the real aim of researchers, and to take the place of the orientation toward the substantial success of research [Brook,2000]. These risks of failure may be avoided or contained to a large extent if the approach to evaluation reckons with it appropriately.

The experience of Trento University, also in research assessment, is probably still positioned in an early stage of development and its growth, in term of intensity and frequency, doesn't overcome the threshold which primes a sort of decreasing return by evaluation and opens space for extensive negative effects. However, it offers a vision of a balanced mixture of assessment methods, aligned with the official concept assumed, bodies in charge of evaluation and a fair way to combine the use in the accountability and in the improvement of logic.

For future research, it will be relevant to compare the impact of evaluation practice of research in different stages of development, taking into account the conceptual framework thus outlined.

References

Allégre, C., *University Autonomy, Accademic Accountability and European Integration*, in AA.VV., *Autonomy and Responsibility,* Bologna, Bononia University Press, 2002.

Barzelay, M., Performance Auditing and the New Public Management: Changing Roles and Strategies of Central Audit Institutions, in P.U.M.A., Performance Auditing and the Modernisation of Government, Parigi, OECD (Organisation for Economic Co-operation and Development), 1996a.

Brennan, J. – Shah, T., Quality assessment and istitutional change: experiences from 14 country, «Higher Education», 40 (2000).

Brook, R., The Evaluation of Research: the Benefit, the Risks and Some Simple Rules, Atti del Convegno: La valutazione della ricerca, Roma, CIVR., 2000.

Brown, A., Directions in Quality management in Higher Education in Australia, Toulon-Verona, Quality in Services. Higher Education; Health Care; Local Government, Proceedings of the Septh Conference, Université du Sud , Toulon – Var, Sept. 2-3 2004.

Brown, R., Quality Assurance in Higher Education. The UK Experience Since 1992, London, RoutledgeFalmer, 2004.

Chelimsky, E., A che punto siamo in materia di valutazione: alcune riflessioni, «Problemi di amministrazione pubblica», 1 (1997).

Cohen, S. – Brand, R., Total Quality Management in Government, S. Francisco, Jossey-Bass, 1993.

Conseil Scientifique de l'Evaluation, L'évaluation de l'expertise à la responsabilité, Parigi, La documentation Française, 1991.

Davies, A. – Thomas, R., Managerialism and Accountability in Higher Education: the Gendered Nature of Restructuring and the Cost to Academic Service, «Critical Perspectives on Accounting», 13, 5-6 (2002).

Dillard, J.F., Dialectical Possibilities of Thwarted Responsibilities, «Critical Perspectives on Accounting», 13, 5-6 (2002).

EES (European Evaluation Society), Governance, Democracy and Evaluation, Sixth European Evaluation Society Biennal Conference, Booklet of abstracts, Sept. 30-Oct. 2 2004, Berlin.

EIASM (European Institute for Advanced studies in management), Process of reform of the University across Europe, Workshop of Siena, 24-26 May 2004.

Felt, U., University Autonomy in Europe: a background Study, in Observatory for Fundamental University Values and Rights, Managing University Autonomy. Collective Decision making and Human Resources Policy, Bologna, Bononia University Press, 2002.

Felt, U., University Autonomy in Europe: Shifting Paradigms in University Research, in Observatory for Fundamental University Values and Rights, University Autonomy: Shifting Paradigms in University Research, Bologna, Bononia University Press, 2004.

Ghosal, S. – Bartlett, C.A., Linking Organizational Context and Managerial Action: the Dimensions of Quality of Management, «Strategic Management Journal», 15 (1994).

Goldhaber, D., The Mystery of Good Teaching, Hoover Institution, Education Next, 2002.

Halachmi, A. – Bouckaert, G., Public Productivity through Quality and Strategic Management, Amsterdam, IOS Press, 1995.

Hoecht, A., Accountability, Trust and Professional Autonomy in UK Higher Education Quality Management, EIASM(European Institute for Advanced studies in management), Process of reform of the University across Europe, Workshop of Siena, 24-26 May 2004.

Karol, E.N.H., - Ginsburg, S.G., Managing the Higher Education Enterprise, New York, John Wiley & Sons, 1980.

Krogstrup, H.K., The Old Philosopher and the Teenager who Let the Settee. Evaluation in Time of New Public Management, EIASM (European Institute for Advanced studies in management), Process of reform of the University across Europe, Workshop of Siena, 24-26 May 2004.

Lawrence, S. – Sharma, U., Commodification of Education and Academic Labour, «Critical Perspectives on Accounting», 13, 5-6 (2002).

Lay, S., The interpretation of the Magna Charta Universitatum and its Principles, in Observatory for Fundamental University Values and Rights, The interpretation of the Magna Charta Universitatum and its Principles, Bologna, Bononia University Press, 2004.

Merton, R.K. – Barber, E.G., The Travels and Adventures of Serendipity. A Study in Historical Semantics and the Sociology of Science, Priceton University Press, 2004.

Middlehurst, R., Quality Assurance Implications of New Forms of Higher Education, Helsinki, ENQUA (European Network for Quality Assurance), 2001.

Minelli, E. – Rebora, G. – Turri, M., Il valore dell'università, Milano, Guerini, 2002.

Mintzberg, H., Power In and Around Organizations, Prentice-Hall, Englewood Cliffs, 1983.

Moore, M.H., Creating Public Value, Cambridge Mass, Harvard University Press, 1995.

Neave, G., The Evaluative State Reconsidered, «European Journal of Education», 33, 3 (1998).

Noordegraaf, M. – Abma, T., Management by Measurement? Public Management Practices amidst Ambiguity, «Public Administration», 81, 4, (2003).

Osborne, D. – Gaebler, T., Reinventing Government: How the Entrepreneurial Spirit is Transforming the Public Sector, Addison-Wesley, Reading Mass, 1992.

P.U.M.A., Performance Management in Government, Parigi, OECD (Organisation for Economic Co-operation and Development), 1994.

P.U.M.A., Performance Auditing and the Modernisation of Government, Parigi, OECD (Organisation for Economic Co-operation and Development), 1996a.

P.U.M.A., Integrating People Management into Public Service Reform, Parigi, OECD (Organisation for Economic Co-operation and Development), 1996b.

P.U.M.A., In Search of Results, Performance Management Practices, Parigi, OECD (Organisation for Economic Co-operation and Development), 1997.

Paddock, S.C., Evaluation, in Shafritz J.M., International Encyclopedia of Public Policy and Administration, Boulder Colorado, Westview Press, 1998.

Parker, M. – Jary, D., The McUniversity: Organization, Management and Academic Subjectivity, «Organization», 2, 2, (1995).

Pirsig, R.M., Lo Zen e l'arte della manutenzione della motocicletta, Milano, Adelphi, 1981.

Pirsig, R.M., Lila, Milano, Adelphi, 1992.

Pollitt, C., The Politics of Performance Assess: Lessons for Higher Edutation?, «Studies on Higher Education», 12, 1 (1987).

Power, M., The Audit Society. Rituals of Verification, Oxford, Oxford University Press, 1997.

Rebora, G., La valutazione dei risultati nelle amministrazioni pubbliche, Milano, Guerini, 1999.

Rebora, G., Gli obiettivi della valutazione: miglioramento della qualità, trasparenza delle informazioni, efficienza ed efficacia dei processi, Conference: La valutazione: un indispensabile strumento di garanzia e di governance, Roma 8 ottobre 2003.

Reichert, S. – Tauch, C., Trends 2003 Progress towards the European Higher Education Area, Geneve, European University Association, 2003.

Reponent, T., Is leadership possible at loosely coupled organizations such as university?, «Higher Education Policy», 12 (1999).

Rhodes, R.A.W., The Hollowing Out of the State. The Changing Nature of the Public Service in Britain, «The Political Quarterly», 65, 2 (1994).

Rowan, B. - Correnti R. - Miller R.J., What Large-Scale, Survey Research Tells Us About Teacher Effects on Student Achievement, Consortium for Policy Research in Education, University of Pennsylvania, Graduate School of Education, Research Report Series, November 2002.

Scriven, M., New Frontiers of Evaluation, «Evaluation Practice», 7 (1986).

Shafritz, J.M. - International Encyclopedia of Public Policy and Administration, Boulder Colorado, Westview Press, 1998.

Singh, G., Educational Consumers or Educational Partners: a Critical Theory Analysis, «Critical Perspectives on Accounting», 13, 5-6 (2002).

Stame, N., Valutazione ex post e conseguenze inattese, «Sociologia e ricerca sociale», 31 (1990).

Stame, N., L'esperienza della valutazione, Roma, Studio editoriale AM, 1998.

Stewart, J. – Walsh, K., Change in the management of Public services, «Public Administration», 4 (1992).

Tabatoni, P. – Davies, J. – Barblan, A., Strategic Management and Universities' Istitutional Development, Geneve, European University Association, 2003.

The Danish Evaluation Institute, Quality Procedures in European Higher Education, Helsinki, ENQUA (European Network for Quality Assurance), 2003.

Toulon-Verona, Quality in Services. Higher Education; Health Care; Local Government, Proceedings of the Septh Conference, Université du Sud , Toulon – Var, Sept. 2-3 2004.

Trow, M., Problems in the translation from elite to mass higher education, in OECD, Policies for Higher education. General Report on the Conference on the future structures of post secondary education. Paris: OECD (Organisation for Economic Co-operation and Development), 1974.

Turri, M., Lo stato dell'arte, limiti ed evoluzione nella valutazione a livello di sistema e di singolo ateneo, «Azienda Pubblica», 16, 5-6 (2003).

Unione Europea, Realising the European Higher Education Area, Communiqué of the Conference of Ministers responsible for Higher Education, Berlin on 19 September 2003.

Van Damme, D., Higher Education in the Age of Globalization, in A.A.V.V., Globalization and the market in higher education, Paris, UNESCO, 2002.

Vedung, E., Public Policy and Program Evaluation, New Brunswick, Transaction Publishers, 1997.

Vickers, G., The Art of Judgment. A Study of Policy Making, Londra, Harper & Row, 1965.

Wildawsky, A., Speaking truth to power, Boston, Little Brown, 1979.

Wildawsky, A., The Self Evaluating Organization, in Cook T.D., Evaluation, London, Sage, 1981.

Notes

[*] Previous versions of this paper were presented in the following conferences:
- Toulon-Verona, Quality in Services. Higher Education; Health Care; Local Government, Proceedings of the Septh Conference, Université du Sud , Toulon – Var, Sept. 2-3 2004.
- EIASM (European Institute for Advanced studies in management), *Process of reform of the University across Europe*, Workshop of Siena, 24-26 May 2004.

[1] The whole sample allows for all obtained answers. The sum of partial frequency per group does not correspond to the whole sample because an interviewee may belong to more than one group or none of them (for instance, the interview to the students' representative is not registered in any group, as it is single, but in the total).

[2] The result is not surprising. In the teaching assessment the difficulty to get effects from the evaluation activities is greater. This is shown by the two oldest examples of *assessment* of university activities in Europe, the English one (Drennan, 2001) and the Dutch one (research by Turri, in course of printing).