ACADEMIC STAFF EVALUATION AT IST

Pedro Lourtie

Instituto Superior Técnico, Technical University of Lisbon, Portugal

Abstract:

A recent Portuguese law requires public higher education institutions to evaluate their academic staff at least once every three years and establishes the principles according to which the regulations are approved by each institution. Staff member evaluation aims at shaping individual staff member activity and is an instrument to regulate their carrier.

Instituto Superior Técnico (IST), the faculty of engineering of the Technical University of Lisbon, has opted to develop an elaborate process, described in the present paper, base on a multiple criteria approach. The procedure has a significant number of degrees of freedom in order to be aligned with institutional mission and goals and to consider the breath of the activity of the members of the academic staff. On the other hand, its complexity and the sheer number of degrees of freedom may hinder its readability and, as a consequence, its effectiveness.

1. Introduction

The paper addresses the evaluation of the members of the academic staff, as an element of institutional management, including quality assurance, and of the academic carrier. The way evaluation is carried out and the criteria used, including those implicit in the procedures, conditions academic staff's activity. If not immediately, at least in the long run, it will shape the body of faculty and will influence institutional performance and quality.

The evaluation of the academic staff's activities is often conducted by management and is based on curriculum assessment. The guidelines of this assessment must be aligned with the mission and goals of the institution to contribute towards their fulfilment. Although based on guidelines and on the role and performance expected from each academic, it gives the evaluator a significant degree of freedom to take into account differences of profile of individual members of the academic staff and their relevance for institutional performance.

Staff member's evaluation has potentially a dual impact: on the performance of the institution, on the one hand; and on individual staff member activity orientation and performance, given the effect on their carrier and progression, on the other hand.

The Portuguese law, revised in August 2009, requires public universities [ECDU, 2009] and polytechnics [ECPDESP, 2009] to evaluate their academic staff at least once every three years, according to regulations to be approved by each institution based upon a set of principles

established in the law. Most of the regulations approved by higher education institutions take into account four domains of activity – teaching, research, extension and management – with variations in the weight each one has in the final result. Some of the regulations even define different profiles of the members of staff, with different emphasis on each of the four domains, according to which the evaluation takes place.

Instituto Superior Técnico (IST), the faculty of engineering of the Technical University of Lisbon (Universidade Técnica de Lisboa), has opted for an elaborate process, described in the present paper, base on a multiple criteria approach [RADIST, 2010]. This process is presented here as a basis for reflection and discussion of issues related to the evaluation of the academic staff.

2. The academic staff evaluation at IST

In Portugal the members of the academic staff of the public universities and polytechnics are civil servants. Their carrier is strictly regulated by law, including the functions they are supposed to perform, how they are recruited and promoted, their rights and duties and so on. In this law, a couple of articles refer to the evaluation of the members of the academic staff. They establish the principles it must follow and, in general terms, the consequences of positive and negative results of the evaluation.

The last versions of the laws establish that the higher education institutions must define their own evaluation system, following a number of principles. Many institutions already had some form of evaluation of teaching and research units are subject to a national evaluation. The new system, however, is linked to carrier development and focus on the individual academic staff member's performance.

The Technical University of Lisbon approved a general regulation [RADUTL, 2010] that was then detailed at the level of its seven faculties. At IST, an elaborate regulation was established using a multiple criteria approach [RADIST, 2010]. The procedure and the regulation were discussed and approved at the IST's governing bodies and is being applied for the first time this year, with reference to a period ending in 2009.

The process is conducted by a council for the evaluation of academic staff, specifically established for the process, and includes five phases: self-evaluation, evaluation, harmonization, notification of results and final approval of marks. In what follows, we will concentrate on the approach of the evaluation phase.

Diverse profiles of the academic staff

IST's evaluation procedure focuses on the four domains of activity referred above: teaching, research, knowledge transfer – including extension and science diffusion, as well as social and economic use of knowledge – and institutional management. These four domains cover most of the activities of academic staff members, although each one will have a specific profile, with more or less emphasis on each of the domains. This is taken into consideration, by giving each of these four domains a variable weight. These weights have limits that were defined for the first application as: teaching from 20 to 40%; research from 40 to 60%; knowledge transfer

from 5 to 30%; and management with a maximum of 20% but with a minimum that depends on the level of the member of the staff being evaluated, ranging from 0% at the beginning of the carrier to 10% for Full Professors.

The logic behind these variable percentages is to allow for different profiles of academic staff, although it is research that has the highest weight. The percentages to be used for each of the four domains must add up to 100% and are chosen in order to maximize the final result of the evaluation of each individual member of staff.

Some other institutions, where profiles have been defined, have considered that each member of academic staff is allocated to one of a limited number of profiles. Others define the set of goals and targets for each member of the academic staff to achieve.

The relative weight of the domains of activity represents the mix, or latitude of mix, of activities that is expected from each member of the academic staff. To be effective, this mix, or its average, must be aligned with the mission of the higher education institution.

Parameters and criteria considered in each domain

Within each domain, a number of parameters, both qualitative and quantitative, are considered. These are a reference of the attributes to be taken into consideration in evaluating performance in that specific domain. As an example, in the domain of teaching, five parameters are defined:

- Pedagogic contents (pedagogic publications or applications);

- Teaching activity;

- Pedagogic innovation;

- Student supervision;

- Professional non-academic experience.

To translate the parameters into numbers there are criteria that include a qualitative and a quantitative component. Each criterion will contribute to the final result with a number that is the product of a qualitative component and a quantitative component. For instance, in terms of pedagogic contents, the quantitative component will consider the number and type of pedagogic content, such as books, papers, software applications, prototypes, etc., with a correcting factor for the number of authors, each contributing with a specific numeric value. The qualitative component will take into account issues such as originality, depth, maturity, scientific and pedagogic accuracy, diffusion at national and international level, etc.

The qualitative component is translated into numbers by using a scale of five levels, from very positive with a factor of 1.5 to very negative with a factor of 0.5. This means that in the result the contribution of each publication or application may vary from one to three, depending on the evaluator assessment of its quality.

The same applies to all other criteria and makes the result much more dependent on the evaluator opinion than it is apparent at first sight. Each criterion is given a weight, within the domain it belongs to. The criteria and corresponding weights adopted are the following:

- Teaching

  o Pedagogic contents (1/6)

  o Student supervision (2/6)

  o Courses (3/6)

- Research

  o International publications (3/4)

  o Scientific projects (1/4)

- Knowledge transfer

  o Industrial property, legislation, norms and technical publications (1/3)

  o Services, consulting, science and technology projects and diffusion (2/3)

- University management

Comparing the parameters and the criteria for teaching, listed above and at the beginning of this section, it is evident that not all parameters have a correspondent criterion. The reason is that some parameters will only enter as qualitative factors, such as pedagogic innovation.

Targets and ceilings

The way each criterion contributes to the final result of the domain, the weight adopted for that criterion, is obtained by converting the product of the qualitative and the quantitative components into a value through a function that, at least in the present first application of the procedure, is a simple linear function with saturation. This function is parameterised using targets and ceilings. Targets represent what is considered adequate performance in each criterion and correspond to a value of 100 of the function. One may have a higher result than the target, but is limited by the saturation, a value that, although adaptable for future applications, at the moment ranges from 300 to 600, three to six times the reference value, depending on specific criterion.

Targets are a way of normalizing the result of each criterion with reference to the desired performance. While ceilings avoid that one specific criterion dominates the result in one domain and privilege diversification of activities. By the same token, as they are defined by management – the President of IST – they are a way of defining the performance expected from the members of academic staff in each criterion. By changing the targets, the performance required may be more or less demanding in absolute terms or in relative terms among criteria.

Final marks and profiles

Once a result has been obtained for each of the four domains of activity, the final result is composed by weighting the four domains through percentages that vary in given ranges. The specific percentages to be used are those that maximise the final result of each individual staff member.

It is clear that if one allocates the minimum percentages to all four domains, the sum is less than 100% and some percentage is left to distribute. This should be allocated to the domain where de highest result was obtained, up to its maximum percentage, and, if even then there is some left, to the next to highest result. There are 24 possible combinations in ordering the results obtained in the four domains. Considering the ranges of percentages adopted for the four domains, the number of possible combinations of weights is 12 (10 for full professors), because the percentage left to be distributed is spent in, at most, two domains.

This is a more flexible process than that of defining profiles for each member of staff. In fact, in the first discussions of the regulation, the option was to define specific profiles. However, a limited number of profiles do not capture all the richness that results from having a more varied range of academic staff profiles. On the other hand, this flexibility implies that management has less control on the profile of each individual, although it regulates the average (median) profile by choosing targets, ceilings and relative weights.

The final step of the evaluation process, before harmonization and notification, is the conversion of the final result into one of the four levels of the final mark: excellent, very good, good or inadequate.

A summary of the process

The sequence of the procedure is the following:

- The activities, publications, etc. of each member of the academic staff are translated into numbers, according to tables and formulas included in the regulation, producing the quantitative component;

- For each criterion a qualitative component is defined by the evaluator that, multiplied by the quantitative component, gives a value of performance associated with each criterion;

- This performance result is normalised by a target and limited by a ceiling, both defined beforehand at the beginning of each period being evaluated, giving a result for each criterion;

- The results of the criteria are weighted to give the performance in each of the four domains;

- These results are assigned percentages, according to the ranges defined, in order to maximize the final result;

- The final result is converted into a final mark.

The degrees of freedom of the process are the following:

- The value assigned to each item included in a criterion: defined in the regulation;

- The qualitative components: defined by the evaluators;

- The targets and ceilings: defined by the President of IST;

- The weight of each criterion and of the range of percentages of the four domains: defined by the council for evaluation of the academic staff;

- The range of values corresponding to the four levels of the final mark: defined in the regulation, but that may be altered by the council for evaluation of the academic staff.

Such a large number of degrees of freedom allows for the adaptation of the process, although the complexity of the formulas makes it more difficult to visualize the impact of each one of them in general terms. It is, however, probably easier for each individual member of the academic staff to identify the impact on the final mark of his/her options in terms of activities.

The odd ones out

In spite of the completeness and flexibility that was aimed at by this process, there will always be members of the academic staff that have a very atypical activity during a given period of evaluation. To take that into account, the regulation considers an exceptional method of curriculum evaluation that gives the evaluator an even greater freedom to define the result in each domain.

3. First results

The general civil service law, which defined a set of common principles according to which all civil servants are to be evaluated, requires the new processes to be applied since 2004. As the specific academic carriers' legislation was only adapted in 2009, the first application of the procedure described took place in 2010 and covered to two distinct periods: the four years from 2004 to 2007; and the two year period of 2008 and 2009. In the first of these two periods, the members of the academic staff could decide not to be evaluated. Unless one needed points to progress in the carrier, the result obtained was irrelevant.

From the point of view of understanding the procedure and try to explore its functioning, it is the second period, with the evaluation being compulsory, that is relevant, in spite of the procedure having been defined after the period under evaluation, unlike future applications. However, the statistics of the results may give some indication whether it is working properly or not. For instance, as in this period the option for the alternative curriculum evaluation was freely taken by each individual member of the academic staff, a high percentage opting for this alternative would indicate that the multiple criteria approach was viewed with suspicion.

The first results disclosed show that 85% of the 855 academic staff members opted for the multiple criteria approach. This percentage varies according to the department, but it ranges from 76% to 95%.

Other results may give an indication of the way the multiple criteria procedure is working, at least as an alarm indicator if they are very far from what is to be expected. Such indications could come from the percentage of each profile, defined as a pair of the two domains with best results for each member of staff, or that of the domains in which the members of the academic staff have best results. Knowing the academic staff of IST, it is to be expected that the combination of education and research will dominate the profiles, with these two domains clearly prevailing over the other two. If this is not the case, it could mean that the criteria and targets in the different domains are not properly balanced.

Many other conclusions may be drawn as the statistics of the results become publicly available and results outside the expectable range should lead to a reflection on the causes. For instance, the average of the final results will indicate whether the academic staff performance is above, below or at the standard previously defined by the targets adopted. Very poor or very high results should lead to a reflection on whether the targets were adequately chosen.

4. Conclusion

How effective the procedure will be in capturing all aspects of activity of the members of the academic staff and in having a positive effect in shaping that activity, by being aligned with institutional missions, is still not possible to measure. This is a slow process, as is quality improvement as a result of quality assurance processes in general. One may, however, after the present first round of results has been completed, try to predict its impact by questioning those that have been and are to be evaluated on their perception of the process, the implications they foresee for their carrier or recognition at institutional level and, as a consequence, the type of strategies they will adopt in the future. This will give indications on whether the procedure is effectively aligned with mission and goals of the institution.

Such a process aims at being fair, in the sense that those that have better quality results and greater contribution towards the development of the institution have better marks and are more likely to benefit in their carrier. There is no such thing as an absolutely fair evaluation process. Quantitative approaches, although more to the fact, tend to ignore subtleties of quality and qualitative approaches depend more on the evaluator, although they may capture quality. The process described tries to combine quantitative components with qualitative assessment.

The procedure has the potential and the degrees of freedom to be aligned with institutional mission and goals and to consider the richness of the activity of most members of the academic staff. And, whenever it does not, the alternative route of curriculum evaluation is available. Taking this for granted, the issue is whether its complexity and the number of degrees of freedom do hinder its readability and, as a consequence, its effectiveness.

References

Estatuto da Carreira Docente Universitária (ECDU): Decreto-Lei nº 205/2009, *Diário da República*, 1ª série, nº 168, 31 de Agosto de 2009, pp. 5744-5757; Lei nº 8/2010, *Diário da República*, 1ª série, nº 93, 13 de Maio de 2010, pp. 1652-1653.

Estatuto da Carreira do Pessoal Docente do Ensino Superior Politécnico (ECPDESP): Decreto-Lei nº 207/2009, *Diário da República*, 1ª série, nº 168, 31 de Agosto de 2009, pp. 5774-5784; Lei nº 7/2010, *Diário da República*, 1ª série, nº 93, 13 de Maio de 2010, pp. 1648-1651.

Regulamento de avaliação de desempenho dos docentes da UTL (RADUTL): Despacho nº 2809/2010, Universidade Técnica de Lisboa, Reitoria, *Diário da República*, 2ª série, nº 29, 11 de Fevereiro de 2010, pp. 6651-6655.

Regulamento de Avaliação de Desempenho dos Docentes do Instituto Superior Técnico (RADIST): Despacho nº 4576/2010, Universidade Técnica de Lisboa, Reitoria, *Diário da República*, 2ª série, nº 51, 15 de Março de 2010, pp. 12128-12137; Despacho nº 11570/2010, Universidade Técnica de Lisboa, Reitoria, *Diário da República*, 2ª série, nº 136, 15 de Julho de 2010, pp. 38341-38350.

Issues for discussion:

- Are processes more regulated, in which the evaluator opinion has less weight, more objective and/or fairer than those where the evaluator has more to say?

- Can the variety of profiles of the academic staff, relevant for institutional development, be captured by increasing the number of variables? What are the implications for the alignment of the process with the mission/aims of the institution and the activity of those evaluated?

- What is reasonable in terms of complexity on the one hand and readability on the other? May the results of such complex processes of academic staff evaluation justify the effort?