Analysis of the evaluation process of the research performance: An empirical case

Adela García-Aracil, Antonio Gutiérrez Gracia, Marián Pérez-Marín

Institute for Innovation and Knowledge Management, INGENIO (CSIC-UPV),
Technical University of Valencia, SPAIN.

Contact person: Adela García-Aracil, INGENIO (CSIC-UPV), Technical University of Valencia, Camino de Vera, s/n, 46022 Valencia (Spain), Telf: +34 96 387 7048, Fax: +34 96 387 7991, E-mail: agarcia@ingenio.upv.es

Abstract

In this paper we analyze the objectivity of the peer review process of research performance by research groups in the scientific and technological Valencian system, over the period 1998-2002. For that purpose, we use qualitative and quantitative indicators to assess which of them are the most important to determine a research group as excellent one, based on peer review evaluation methodology. The results show that excellence appears to be driven only by publications in SCI/SSCI and the number of *sexenios,* and suggest that the peer review process is not as objective as we expected.

Introduction

Increasingly, evaluation studies of research performance conducted during the past years focus on the identification of research of "highest quality", "top research", or "scientific excellence". Achieving and maintaining excellence has always been crucial for leading researchers and scholars working in the international frontiers of science. The ability to access that level, and to be competitive in the international arena, has also become a strategic goal and an explicit target of research institutes as a whole (Harvey *et al*., 2002; Goldfinch, 2003).

1

Identification of excellence is a matter of *ex ante* assessment or *ex post* evaluation of research performance. Clearly, such a broad and ambiguous concept is not directly measurable in a generally accepted valid manner. To begin with, there are numerous definitions of "scientific prestige", "elite scientists" and "hierarchies of reputation" in the sociological literature, their exact meaning depending on the school of thought, theory or methodological context (Cole and Cole, 1967; Collins, 1982). However, most of these notions can be applied to individual researchers or socio-cognitive collectives, rather than to institutional aggregates.

At the level of entire research groups, departments and institutions, the conceptual and operational problems are further compounded given the diversity of research goals capabilities, resources, facilities and outputs characterizing research organizations and their units (Loch and Tapper, 2002; Ball and Butler, 2004). Managers of research institutions, funding agencies and supra-national governments all face, for different reasons and goals, the same pervasive evaluative question: how can one define, recognize and compare excellence as objectively as possible? (Hauser and Zettelmeyer, 1997; Werner and Souder, 1997).

Regarding research, it is paradoxically difficult to evaluate it in research-oriented universities and research institutes around the world (Johnes, 1990; Korhonen *et al.*, 2001; Tijssen *et al.*, 2002; Coccia, 2004). The problem is universal and several universities have developed their own approaches to this problem. Two extreme approaches are the *process-oriented* and the *results-oriented* approach. In the *process-oriented* approach, the focus is on the research process. This approach is based on the premise that a high quality research-process produces high quality research results. In the *results-oriented* approach various quantitative "yardsticks" (number of publications,

citations, invited talks at conferences, etc.) are used to measure research output/input (Moed *et al.*, 1985; Nagpaul and Roy, 2003).

The principal output of research groups is knowledge and know-how (Hare and Wyatt, 1992; Gulbrandsen and Langfeldt, 2004), which may be presented in various forms: articles and books, innovative prototypes, patents, consultancies or training of personnel. In that respect, publication is, probably, the most critical indicator of research assessment (Bence and Oppenheim, 2004). Publication is central to scholarly activity and recognition (Ramsden, 1994), and widely regarded as: a main source of esteem, a requirement for individual promotion, the evidence of institutional excellence and a way of obtaining competitive research funds. Indeed, in the academic world, it can be argued that research activity only becomes "a work" when it takes on the conventional, physical form of a published paper or its equivalent. Moreover, reaching a top ranking in the global citation impact league within a discipline is, without doubt, a decisive piece of factual information concerning the scientific status and world-wide impact of a research paper and its associated researchers and institutions (Bordons *et al.*, 2002; Moed *et al.*, 2002; Ho *et al.*, 2003). Research publications receiving relatively large quantities of citations indicate significant scientific "impact", a necessary condition for researchers and affiliated institutions to attain broad visibility and success in terms of scientific recognition (Fox, 1983; Egghe and Rousseau, 2002).

Narin (1976) was the first to review the state-of-the-art of citation analysis techniques for evaluating the performance of scientific institutions. His results supported the idea that relative high citation levels correlate with positive peer opinions about scientific papers importance, with peer rankings of research institutions and, more importantly, with other independent indicators of scientific quality of research papers. In the wake of these pioneering studies, citation impact measures have been applied

with increasing frequency in the 1980s and 1990s at institutional level for evaluating and ranking the research performance of university departments and other research organizations (Gulbrandsen, 2000; Frohlich and Resler, 2001).

However, many are the reasons for one research article citing the other, not all of them directly related to the scientific quality of the cited work or the contributing researches and institutions (Weinstock, 1971; Basu and Aggarwal, 2001). Many critics of citation analysis have therefore objected to their use as surrogates for actual scientific communication processes and derivative measures of scientific repute (Edge, 1979; Leydesdorff, 1998). Therefore, the usage of citation measures for quantitative assessments of research visibility and external impact, let alone scientific quality and excellence, is still controversial – especially amongst the advocates of the purely "qualitative" peer review-based evaluation methodologies (Van Raan, 2000).

New developments in the field of quantitative studies of science offer methods to support peer review in order to keep it objective and transparent. The peer review process is one of the most widely accepted by the Scientific Community for selecting and assessing excellence (Gillett, 1989; Roberts, 1999; Kuldell, 2004). Although, not surprisingly, open and fair applications of peer review evaluation may be difficult to achieve (Horrobin, 1990; Moxham and Anderson, 1992).

This paper contributes to these discussions. Our concern is the assessment of research groups' performance in the region of Valencia over the period 1998-2002. In particular, we address questions concerning the processes used in the selection process conducted by the Valencian government peer review to assess research performance and to determine if a research group is excellent.

The paper is organized as follows: Section 2 offers an overview of the scientific and technological policy framework. Section 3 presents the description of our data set and

the methodological approach. The main results of our study are discussed in section 4 and, finally, section 5 provides a summary and concludes.

Scientific and technological policy framework

Excellence in research groups is understood not only as the excellence of each one of its members, but of the whole which is created through their association and their joint programme of activities, at European, national and regional level (CRUE 2001; CEC, 2002, 2003, 2004). One of the objectives of excellent groups is to reinforce scientific and technological excellence through a progressive and lasting integration of existing research capacities. This approach offers a new horizon for scientific and technological activity and for research policy. Essentially, creating favourable conditions can increase the impact of research efforts by strengthening the coherence of research activities and policies conducted in Europe. In particular, it is recognised that world-class excellence exists in practically all areas and disciplines in Europe, but, these competencies, however, are not sufficiently well known across national and regional borders (CEC, 2001).

In Spain, the three key lines of action of the Spanish scientific and technological policy are the National Plan for Scientific Research and Technological Development (National R&D Plan), the actions by the Ministry of Industry, Tourism and Trade (MITYC) and the implementation of the National Agency of Evaluation and Assessment (ANEP). In addition, there are also international actions (joint programmes with European or Latin American countries, etc.), regional actions from the different autonomous communities and even some established by the Central Administration itself (Ballesteros and Rico, 2001; Albert and Plaza, 2004). It should be noted that in Spain there are 17 autonomous regions. As a result of the regionalisation, all agents involved in R&D activities depend on 18 authorities, one central and 17 regional

governments, with different political ideologies and irregular knowledge of what scientific and technological policy is and what it should be.

The National R&D Plan, adopted in 1988, sets the priorities for action, manages the resources available and integrates the R&D actions of the productive sectors, research institutions and universities. The economic efforts of the National Plan are materialised in the provision of the National R&D Fund. They are largely aimed at the enhancement of basic scientific research and the promotion of communication and concerted actions between universities, firms and public research institutions. Thus, while the actions of the National R&D Plan are oriented towards basic research and the precompetitive development of technology, Spanish technological policy is mainly carried out by the MITYC with a view to favouring industrial innovation. The intervention is designed, among other things, to provide an incentive to the efforts in technological development and the incorporation of advanced technologies in firms, and to improve the competitiveness of Spanish industry through an improvement in the quality of its products.

Furthermore, the ANEP was created in 1986 as a mechanism of scientific evaluation, to assess—with maximum rigor and independence—all public scientific-technical research proposals, research groups and entities that request funds to carry out research and/or technological programmes and projects. The ANEP's scientific evaluations are carried out by anonymous experts using peer review and are used by the corresponding institution responsible for the financing. The fact that the evaluation is carried out by a different and independent institution to the one, responsible for the funding and for making the last decision of financing the action or not, is an additional mechanism of guaranteeing the evaluation quality. In this sense, the ANEP has notably contributed to the huge qualitative jump of Spanish science and technology system

during the last years and has become an excellent scientific observatory. The criteria used to achieve its main objective have been described in detail in several reports (Sanz Menéndez, 1995; Modrego, 1995; Fernández de Caleva, 2003; MEC, 2004).

In our particular case, Valencia, a peripheral region of the European Union (OCDE, 1997) with a low absorptive capacity (Cohen and Levinthal, 1989, 1990), is characterized as a small and open economy, based on a traditional micro- and small-sized firm industrial structure, where the owners lack modern business education or research tradition (COTEC, 1999). Moreover, the Valencia level of R&D spending is lower than the already low Spanish level, 0.6 and 0.9 percent of GNP, respectively (INE, 2002). This region's profile has an important influence on the research group's performance.

In this context, the Valencian Government tries to lead and impel a science and technology policy in order to bring the average level of public and private investment up to that of the most advanced regions of the rest of Spain and Europe. Taking as reference the actions planned in the European and National Frameworks, the Valencian Scientific Research, Technological Development and Innovation Plan (PVIDI) was conceived in 1997 (Generalitat Valencia, 2001).

The Valencian Administration proposes through the PVIDI a number of actions to develop the regional potential, mitigate its deficiencies and establish suitable orientations for the future. A "*public call for excellent research groups*" has been included in this plan, the main objectives being related to encouraging creation, consolidation and projection of research groups in the region of Valencia; linking them to priority lines of research and enabling them to compete on the best possible terms with other international groups in obtaining results, projects and financial resources; and fostering interdisciplinary research. In this sense, regional government allocation

criteria of research resources are based on standards of scientific quality. Competitive bidding has become the habitual procedure for the actions, by means of annual public calls that guarantee an objective criterion of excellence in the granting. Thus, the excellence of the group and its ability to reinforce it and spread it beyond its own members is mainly assessed by the regional government using peer review process (group of experts using the ANEP criteria of scientific and technological evaluation).

Relatively little is known about the decision processes used by the regional government peer review to assess the research performance of research groups that apply for the "*public call for excellent research groups*". However, we know that in a scale up to 100 points, the criteria established to define excellence was: for experience a research group could score up to 70 points (including number of papers, patents, participation in projects under the European Union R&D Framework Programme or under the National R&D Plan, collaboration with other organizations and so on); for the scientific-technological newness and importance of the proposal up to 15 points; for the composition, structure and consistency of the group up to 5 points; and, for the adequacy of the R&D activities developed by the research group for the PVIDI framework up to 10 points.

For two reasons, we would expect a less than perfect correlation between the regional government peer review results and our own. First, we use the available quantitative information in a purely objective manner, so as not to allow judgmental factors to influence the outcome of the exercise, whereas the regional government peer reviews allows the impressionistic evidence of qualitative nature to influence the rankings. Secondly, the regional government peer review may impose upon all research groups its own subjective judgement about the relative weight to be attached to each type of publication (e.g. articles published in international conferences are assigned a

relatively low weight, while articles in core journals are considered substantial contributions); such an imposition is not allowed in our exercise.

In this paper, we try to assess excellence features of research groups in the region of Valencia. According to scientific literature, we analyze a range of qualitative and quantitative indicators, each one focusing on different aspects of a group's performance, in order to provide a reasonable and reliable estimate of the contribution to scientific progress made by excellence groups and its influence on the regional scope.

Data and methodology

The data used in this paper was taken from the study *"Análisis de la Especialización Temática del Entorno Científico y Tecnológico de la Comunidad Valenciana"* carried out in 2004 by the Valencian government. The information includes those research groups inside the scientific and technological Valencian system that applied in 2003 for the *"public call for excellent research groups"*, a public aid announcement from the regional government. A total of 227 research groups applied for the public aid (representing an application rate of approximately 20 per cent), of which 185 correspond to university research groups, 22 were R&D public organizations, 12 were part of the R&D foundations and the remaining were technological institutes. It seems probable that researchers with low research activity rates or a low level of publications were under represented in the sample. We cannot forget that the sample is composed by groups that supposedly had high scores in indicators as average experience of the group in research activities (experience in project management, technology transfer and other R&D activities), patents granted, publications, conferences, cooperation with other organizations, number of fellowships, number of PhD steering and defended in the research group, scientific-technological newness and importance of the subject

researched according to the regional, national or European R&D framework program, and so on. It leaves open the possibility of a systematically biased sample.

The information collected was divided into two groups: (i) questions on general characteristics of 2002 research groups, such as size, organization structure, personnel academic status, and so on; (ii) information about the research activity performed by the research group during the period 1998-2002, such as the number of articles published in international refereed journals, number of *sexenios* (for each period of six years, a tenured professor or scientist can present his/her most relevant scientific contribution to a national committee of experts for each discipline in the hope of receiving a positive assessment of his/her individual research activity – the so-called *sexenios*), papers presented in international conferences, funds coming from European, national or regional projects, and funds coming from contracts with different organizations.

Table 1 shows the general characteristics of research groups by type of institutions. Excellence was coded 1 for those research groups that regional government peer review considered as excellent, and 0 for those research groups not evaluated as excellent. Size comprises the number of members. Academic status was coded as 1 if the personnel of the research group got a PhD degree, 2 means personnel with tertiary-type-A-education degree (ISCED 5A), 3 means personnel with tertiary-type-B-education degree (ISCED 5B), and 4 means technical research assistant (OECD, 2004). Labour market status was also measured as code 1 if the personnel had a permanent contract and code 0 if the contract was temporary. Age of group members was coded in years.

Table 1. Research group's characteristics by type of institution, 2002.

Research groups	Excellence		Size	Academic Status*				Labour Status*		Age
	No	Yes		1	2	3	4	Perm.	Temp	
Universities	73	112	18.8	52.5	43.6	1.6	2.2	80.8	19.2	41.2
Tech. Institutes	7	1	20.3	25.5	62.1	2.5	9.8	25.8	74.2	40.4
R&D Foundations	4	8	18.4	42.0	46.1	1.2	10.7	53.5	46.5	42.8
R&D Pub. Organizations	2	20	20.4	45.9	37.9	4.5	11.7	70.7	29.3	41.5
Total	86	141	19.0	50.4	43.8	1.9	3.9	76.5	23.5	41.3

*Note: * rows add 100 percent and differences among groups statistically significant at the 1% level.*

As we can observe in Table 1, 91.0 percent of research groups belonging to R&D public organizations were marked as excellent research groups, followed by the groups in R&D foundations and the groups established in universities, 67.0 and 61.0 percent, respectively. On average, slight differences were found with respect to group's size (19 members). Regarding academic status, we can see that about one half of the personnel holds a PhD degree, except in the case of research groups belonging to technological institutes, where 62.1 percent of the staff holds a tertiary-type-A-education degree. The proportion of tertiary-type-B-education degree personnel was relative high for R&D public organizations (4.5 percent) and for technological institutes (2.5 percent) compared to the overall percentage (1.9 percent). Technical research assistants represented one-tenth of the personnel in R&D public organizations (11.7 percent) and R&D foundations (10.7 percent).

With respect to labour status, Table 1 shows that altogether over 76 percent of the research group members had a permanent contract. The proportion of temporary contracts was by far the highest in technological institutes (74.2 percent). In contrast, temporary contracts in universities represented the smallest percentage (19.2 percent). Finally, we can observe that age was quite similar across all types of institutions.

By subject area, Table 2 presents the same research groups' characteristics as above. The subject areas covered were natural sciences, engineering, medical science, agriculture, social science and humanities. We can see that the proportion of excellent research groups was the highest in natural science (70.1 percent). Humanities came next

(66.7 percent). Regarding group's size, slight differences were found among the groups. Medical science and humanities were the smallest and engineering groups were the biggest. The academic status structure shows us the importance of PhD personnel, specially in humanities (74.1 percent) and social science (59.6 percent). Close to this category was the tertiary-type-A-education staff with high percentages in medical science (48.7 percent), natural science (45.9 percent) and engineering (44.5 percent). In general, personnel with tertiary-type-B-education degrees and research assistants were only representative in engineering and agriculture. Labour status data shows a general tendency to permanent contracts across all areas (76.5 percent). Permanent contracts were clearly lower than average for those research groups in agriculture (69.7 percent), medical science (74.8 percent) and engineering (76.0 percent). On average, age was very similar throughout all scientific areas (around 41 years old), being the oldest those in medical science (43.6) and the youngest those in engineering (40.4).

Table 2. Research groups' characteristics by subject area, 2002.

Research groups	Excellence		Size	Academic Status*				Labour Status*		Age
	No	Yes		1	2	3	4	Perm.	Temp	
Natural Science	32	75	18.7	49.7	45.9	1.0	3.4	77.6	22.4	41.1
Engineering	19	27	21.7	48.3	44.5	2.1	5.1	76.0	24.0	40.4
Medical Science	16	15	15.4	45.9	48.7	1.9	3.6	74.8	25.2	43.6
Agriculture	5	8	19.5	46.5	37.2	5.3	11.0	69.7	30.3	40.9
Social Science	12	12	20.5	59.6	35.0	4.5	0.8	77.0	23.0	40.7
Humanities	2	4	15.5	74.1	25.9	0.0	0.0	81.6	18.4	42.7
Total	86	141	19.0	50.4	43.8	1.9	3.9	76.5	23.5	41.3

*Note: * rows add 100 percent and differences among groups statistically significant at the 1% level.*

Table 3 presents personnel structure of research groups by subject area. We have split each column in two additional ones in order to show differences between the excellent groups and the non-excellent. Contrary to our expectations, research groups did not significantly differ in the proportion of permanent contracts compared to temporary contracts, except in the case of engineering research groups, where the proportion of permanent contracts was significantly higher in excellent groups than in non-excellent groups. On average, the number of pre-doctoral fellowships was higher in

excellent groups than in non-excellent groups across all scientific areas. That difference was statistically significant higher among medical science, engineering and social science research groups. With respect to the number of post-doctoral fellowships, we can observe that in general, research groups hardly had a post-doctoral member in their team. The PhD steering and defended were not significant differed between excellent and non-excellent groups.

Table 3. Personnel structure by subject area, 2002.

Research groups	Permanent Status (%)		Pre-doctoral Fellowships (headcounts)		Post-doctoral Fellowships (headcounts)		PhD Steering and Defended (number)	
	Non-Ex.	Excel.	Non-Ex.	Excel.	Non-Ex.	Excel.	Non-Ex.	Excel.
Natural Science	80.4	76.4	4.2	4.3	0.3	0.9*	4.7	5.5
Engineering	63.3	84.9*	2.6	4.2*	0.2	0.4	4.4	6.1
Medical Science	80.2	69.0	1.8	4.2*	0.3	0.6	6.1	7.3
Agriculture	72.4	68.1	3.9	4.4	0.3	1.5	4.9	4.9
Social Science	78.0	75.9	1.4	2.4*	0.1	0.1	5.2	5.4
Humanities	92.9	76.0	3.7	4.4	0.5	0.1	5.5	7.4
Total	76.1	76.7	3.0	4.1*	0.2	0.7*	5.0	5.8

Note: * differences among groups statistically significant at the 1% level.

For the activity research period from 1998 to 2002, Table 4 shows the total number of *sexenios*, publications in SCI/SSCI and publications related to international congresses for each excellent and non-excellent research group by subject area. On average, we can observe that the number of *sexenios*, publications in SCI/SSCI and publications in international congress were statistically significant higher in excellent groups than in non-excellent groups, except in the case of humanities and agriculture research groups.

Table 4. Quality indicators of publications by subject area, period 1998-2002.

Research groups	Sexenios		Publications in SCI/SSCI		Publications in international congresses	
	Non-Ex.	Excel.	Non-Ex.	Excel.	Non-Ex.	Excel.
Natural Science	8.9	11.3*	26.9	60.9*	25.8	29.3
Engineering	5.9	11.1*	25.9	39.3*	27.2	38.5
Medical Science	9.0	8.0	26.6	47.0*	12.4	42.7*
Agriculture	8.6	8.8	24.4	34.5	15.3	21.3
Social Science	6.5	11.4*	6.3	23.3*	15.6	18.2
Humanities	16.0	10.5	39.8	22.7	27.3	16.8
Total	8.1	10.7*	23.9	49.5*	21.6	30.7*

Note: * differences among groups statistically significant at the 1% level.

Table 5 also presents, for the same period 1998-2002, the funds coming from regional, national, international projects and competitive actions, measured in thousand euros. In general, we can observe that the main financial resources came from national projects, followed by international projects and, lastly, from regional and competitive actions. On average, the amounts were higher in excellent than in non-excellent research groups, although those differences were not statistically significant, except in the case of the high funds coming from national projects for medical science groups, and, on the contrary, the case of high funds coming from international projects for non-excellent social science groups.

Table 5. Funds coming from projects by subject area, period 1998-2002 *(thousand euros)*

Research groups	Regional		National		International		Competitive Actions		Total	
	Non-Ex	Excel.	Non-Ex	Excel.	Non-Ex	Excel.	Non-Ex	Excel.	Non-Ex	Excel.
Natural Science	74.0	85.1	341.8	362.3	118.3	201.3	28.1	44.4	562.2	693.2
Engineering	142.3	75.4	330.8	351.4	170.1	196.4	136.3	81.1	779.5	704.3
Medical Science	29.9	41.1	134.2	312.9*	37.5	134.6	15.1	70.8	216.7	559.4
Agriculture	45.0	83.4	396.2	498.4	101.9	92.0	24.6	26.4	567.6	700.1
Social Science	40.6	80.3	146.5	173.0	233.7	64.2*	17.1	28.3	437.9	345.7
Humanities	75.8	28.9	321.8	471.4	156.5	90.3	50.0	13.2	604.1	603.9
Total	74.6	76.5	276.2	349.7	130.7	172.2	48.4	51.0	529.9	649.4

*Note: * differences among groups statistically significant at the 1% level.*

Table 6 shows, for the same period 1998-2002, the funds coming from R&D contracts, technical support contracts, and other contracts related to consultancies and other similar services provision, measured in thousand euros. We can see that non-excellent engineering groups got their funds mainly from contracts related to service provision and technical support in contrast to their excellent counterparts.

Table 6. Funds coming from contracts by subject area, period 1998-2002 *(thousand euros)*

Research groups	R&D		Technical Support		Service Provision		Total	
	Non-Ex.	Excel.	Non-Ex.	Excel.	Non-Ex.	Excel.	Non-Ex.	Excel.
Natural Science	82.8	161.8	5.6	36.8	18.9	5.3	107.3	204.0
Engineering	351.5	156.8	239.9	35.0	463.7	45.8*	1055.1	237.6
Medical Science	6.0	80.7	5.3	1.5	7.6	3.0	19.0	85.2
Agriculture	103.4	51.3	13.4	25.1	44.1	343.4	160.9	419.8
Social Science	101.8	90.8	45.1	146.8	36.3	42.7	183.3	280.3
Humanities	135.0	33.8	51.0	90.8	67.6	16.9	253.6	141.4
Total	133.0	136.3	64.3	42.9	120.1	35.5*	317.4	214.7

*Note: * differences among groups statistically significant at the 1% level.*

According to this description it seems that the number of publications both in SCI/SSCI and in international congresses is the main difference found between excellent and non-excellent groups in order to settle down by peer reviewers the determination of a group as excellent or non-excellent. In the next section, regression analysis was carried out to estimate which factors were the most relevant determining the excellence of a research group.

Results

To clarify the effect of each explanatory variable on the classification of a research group as excellent, we estimate five different specifications of the probit equation in order to assess the total, the indirect, and the direct effect of publications on excellence (Green, 1997). The first specification uses only general characteristics as regressors, and consequently estimates the indirect effects of publications, via observable composition of the research group, on excellence (Model I). The second specification uses only quality variables about publications, so it estimates the total effects of publication on excellence without looking at actual attributes of the research group structure (Model II). The third and fourth specifications include as explanatory variables the sources' funds coming from projects (Model III) and agreements with firms and administrations (Model IV), respectively. The fifth specification combines all sets of explanatory variables, thus estimating the direct effects of publication on excellence once its indirect effects have been removed (Model V). Additionally, the subject area and type of institution are also added to all models as control variables. The estimation results for all five specifications of the excellence equation are presented in Table 7.

Model I illustrates the effects of diverse observable research-group-specific characteristics on excellence. Since publication production influences the research-group's structure, the estimates here may be understood as a consequence of the indirect

effects of publication production, via actual composition group, on excellence. It is noteworthy to observe that neither size nor personnel labour status influence excellence, however the number of post-doctoral fellowships positively affects excellence (Korhonen *et al.*, 2001; Bozeman and Corley, 2004). When exploring the segmentation of different subject areas, we note that those research groups in medical science tend to be less excellent that those in natural science, the omitted reference category. With respect to the type of institution, those research group in universities and technical institutes tend to be less excellent that those in R&D public organizations, the omitted reference category.

Model II provides information about the total effects of publication-related variables on excellence. Results show, as we expected, that the number of *sexenios* and of publications in SCI/SSCI have an important influence on the determination of a research group as excellent (Tijssen *et al.*, 2002). With respect to the effect of the explanatory variables considered as control variables, research-group's subject area and type of institution, the results are the same as we mentioned in Model I.

Model III and IV present the effects of diverse funding sources on the determination of excellence. Since the financial resources available for a research group affect its research activity, we guess that financial programmes and actions, which stimulate collaboration among other agents, might play a favourable role enhancing research productivity, via raising the number of publications, and thus increasing the likelihood of becoming an excellent research group. Thus, it should be borne in mind that the estimates in both models, III and IV, may be understood as a consequence of the indirect effects of publications, via actual financial resources, on excellence. Contrary as we expected, the total amount of funds coming from projects and from collaborations

with firms and the Administration do not influence the determination of a research group as excellent (Debackere and Glanzel, 2004).

Finally, the direct influence of publications on excellence is analysed by estimating Model V in Table 7. We consider here all explanatory variables together, so estimates reflect both the direct impacts of publication variables on excellence and the indirect effects, via actual research-group characteristics including financial capacity. When all sets of variables are included in the excellence equation, the main results obtained from Model I to Model IV are simultaneously upheld for most of the key variables. Nevertheless some noteworthy differences appear: funds coming from contracts related to service provision have a negative influence on excellence. This finding implies that collaboration activity moves away from excellence. Contrary as the Van Looy et al. (2004) finding, involvement in contract research seems not to stimulate scientific activities rewarded by peer reviewers as particular contributions in the achievement of excellence. This is likely to be the result of the lack of a stimulated incentive system where one of the cornerstones of the researcher concerns is the publication of research results and the opportunity for open discussions between colleagues, whereas firms have a responsibility for and need to protect the value of their investment. These differences in the incentive system of public and private research create challenges with regard to the dissemination of information, the nature of research conducted and the access to research results (Merton, 1968a, b; Hane, 1999). For instance, some forms of publication might be delayed or suppressed, because firms may ask research groups to keep information (temporarily) confidential (Florida and Cohen, 1999, referred to this as the secrecy problem). This might reduce the incentive to publish, and run counter the research group norm of open dissemination of scientific knowledge.

Summarising, excellence appears to be driven by publication to a very high extent. We find evidence of both direct and indirect positive effects of publications on excellence.

Table 7. Probit estimates for excellence of research groups.

	Model 1		Model II		Model III		Model IV		Model V	
	Coef.	z-value	Coef.	z-value	Coef.	z-value	Coef.	z-value	Coef.	z-value
General Characteristics										
Size	-0.005	-0.593							-0.013	-1.251
Temporal staff	0.001	0.217							0.001	0.071
Pre-doc. Fellowship	0.022	0.777							-0.002	-0.063
Post-doc. Fellowship	0.298	2.146							0.189	1.320
Quality of research										
Sexenios			0.059	2.458					0.069	2.645
Pub. in SCI/SSCI			0.014	3.544					0.014	3.047
Pub. Inter. Congress			0.003	0.983					0.004	1.011
Funds coming from projects										
Regional					0.001	0.110			-0.001	-0.094
National					0.002	0.109			0.001	0.181
International					0.001	0.438			-0.001	-0.120
Competitive actions					0.001	0.803			0.001	0.933
Funds coming from collaborations with firms and administrations										
R&D							0.001	0.835	0.001	0.770
Technical support							0.002	1.216	0.002	1.450
Service provision							-0.001	-1.436	-0.001	-1.678
Control variables: subject area (ref. natural science) and type of institution (ref. R&D public organization)										
Engineering	-0.029	-0.116	-0.044	-0.166	-0.136	-0.542	-0.129	-0.514	0.018	0.066
Medical Science	-0.571	-1.988	-0.448	-1.521	-0.605	-2.096	-0.586	-2.045	-0.463	-1.508
Agriculture	-0.501	-1.231	-0.168	-0.412	-0.434	-1.093	-0.357	-0.891	-0.127	-0.297
Social Science	-0.339	-1.132	-0.014	-0.045	-0.489	-1.692	-0.543	-1.861	0.010	0.030
Humanities	0.016	0.029	0.058	0.110	-0.013	-0.024	-0.031	-0.057	0.011	0.021
Universities	-0.904	-2.084	-1.118	-2.506	-1.038	-2.433	-1.203	-2.762	-1.178	-2.375
Tech. Institutes	-2.338	-2.949	-1.839	-2.339	-2.707	-3.627	-2.685	-3.484	-1.975	-2.100
R&D foundations	-0.583	-0.953	-0.447	-0.708	-0.606	-1.026	-0.754	-1.257	-0.449	-0.670
Intercept	1.195	2.484	0.274	0.541	1.427	3.129	1.607	3.636	0.302	0.500
Observations	227		227		227		227		227	
LRchi2(12)	34.03				27.15					
LRchi2(11)			57.16				28.91			
LRchi2(22)									66.42	
Prob>chi2	0.000		0.000		0.000		0.000		0.000	
Log Likelihood	-133.6		-122.0		-137.0		-136.2		-117.4	

Table 8 translates Model V estimates into predicted probability distributions to assess the size of the effect that a marginal change in each explanatory variable has on excellence. The first row gives the probability distribution for a reference research group which is defined as a R&D public organisation in the natural science area with all continuous variables calculated at mean values. The other rows indicate how the probability distribution of this reference group changed when the value of a dummy

variable was altered or if we changed the value of the continuous variable by two standard deviations. The table shows that the net effect of publications on excellence is substantial compared to those corresponding to other determinants.

The results suggest that a raising number of publications in SCI/SSCI and of *sexenios* have a positive effect on the predicted probability distribution of excellence, whereas decreasing funding sources coming from service provision contracts raises the probability of scoring up to 98.3 percent, i.e. an excellent level. In contrast, a research group set up as a technical institute reduces its probability of scoring as excellent group to a mere 35.0 percent, whereas if it is established in a university the prospect is reduced to 65.9 percent. Finally, working in the area of medical science reduces the probability to 87.0 percent.

Table 8. Predicted probabilities of excellence.

	Non excellent	Excellent
Reference research group	5.6%	94.4%
Size	2.8%	97.2%
Temporal staff	5.4%	94.6%
Pre-doc. fellowship	5.5%	94.5%
Post-doc. fellowship	2.1%	97.9%
Sexenios	1.1%	98.9%
Pub. in SCI/SSCI	0.5%	99.5%
Pub. Inter. congress	3.3%	96.7%
Regional	5.3%	94.7%
National	5.1%	94.9%
International	5.3%	94.7%
Competitive actions	3.2%	96.8%
R&D	3.4%	96.6%
Technical support	1.8%	98.2%
Service provision	1.7%	98.3%
Engineering	5.4%	94.6%
Medical Science	13.0%	87.0%
Agriculture	7.2%	92.8%
Social Science	5.5%	94.5%
Humanities	5.5%	94.5%
Universities	34.1%	65.9%
Tech. institutes	65.0%	35.0%
R&D foundations	12.8%	87.2%

In order to determine if the observed state of excellence of a research group differs from the excellence predicted from our model, we calculate the two kinds of errors that can be made in significance testing: (1) a true null hypothesis can be incorrectly rejected

and (2) a false null hypothesis can fail to be rejected. The former error is called a type 1 error and the latter error is called a type 2 error (McFadden, 1984). Assigning the same weight to each research group and taking as the breaking point 50.0 percent, our model correctly classifies 62.8 percent of those who are not excellent and 81.6 percent who are excellent (see Table 9). The type 1 error, or "false negative", consists in predicting that a research group is not excellent when actually it is excellent (11.4 percent). The type 2 error, or "false positive", predicts excellence for a group who actually is not (14.1 percent). We believe that the principal source of incorrect classifications is the omission of objective criteria in the assessment of excellence.

Table 9. Excellence based on the probabilities estimated by the probit model for each research group.

Excellence	Breaking point 50%		
Original	No	Yes	Total
No	54 (23.8%)	32 (14.1%)	86 (37.9%)
Yes	26 (11.4%)	115 (50.7%)	141 (62.1%)
Total	80 (35.2%)	147 (64.8%)	227 (100%)

Conclusions

In this paper we assessed the evaluation process of research group's performance in the region of Valencia over the period 1998-2002. The information comes from those research groups inside the scientific and technological Valencian system that applied in 2003 for the "*public call for excellent research groups*", a public aid announcement from the regional government. Over the last few years, a number of studies focused on research evaluation and evaluation of scientific and technological policies have been published; however, our research proposes two new relevant points of interest. First, it is based on data from a peripheral region with a low-level absorptive capacity and in the context of the European Union, hence we can provide a reasonable estimate of the contribution of research groups activity at the regional scope. The second aspect of our research that should be emphasised is the quality of data regarding research groups' activities. We use a big quantity of performance-related information such as the number

of articles published in international refereed journal, the number of *sexenios*, financial resources coming from European, national or regional projects, as well as funds coming from contracts with different organizations. This comprehensive data set permits us to evaluate the criteria used by peer reviews on the decision of a research group as excellent behind their activity.

Our findings show that publications in SCI/SSCI and the number of *sexenios* are the most determinant indicators of excellence in a research group. Neither size nor personnel labour status influence excellence. On the other hand, research groups belonging to universities and technical institutes tend to be less excellent that those in R&D public organizations. Those belonging to medical science tend to be less excellent that those in natural science. Contrary as we expected, the total amount of funds coming from projects and coming from collaborations with firms and administrations have no influence on the determination of a research group as excellent.

The finding that funding sources do not affect the determination of excellence could be explained by the fact that those papers that disseminate the research results obtained from projects and contracts are interim reports or working papers (Florida and Cohen, 1999; Hance 1999), and not referred journal articles which give research groups a record of high quality research, and thus excellence. Although re-allocation of resources should be one of the ultimate purposes of research evaluations, we do not pursue this aspect formally here. We believe that our research could form part of the basis for such decisions. We also suggest that final resource allocations should depend on aspects that we do not cover.

As other authors have shown, the research activity follows a multi-dimensional strategy (Callon *et al.*, 1994; Laredo, 2003; Nowotny *et al.*, 2003). The data used in this paper confirms that research in a region participates of that strategy due to the

knowledge-production dimension (publications), the interactions within the socio-economical environment (contracts), and the relationships with their counter mates (cooperation) are taken into account by the regional political agents through their R&D policies implemented. Nevertheless, the results obtained in our models show that research groups only are assessed, recognized and rewarded by their scientific dimension (publications). This "practice" may trigger the contrary effect to what should be expected, i.e. to connect the research activity with the problems of the region. Therefore, in a region with low-level absorptive capacity where the scientific and technological policies are mainly focused on the scientific dimension, a science "delocalisation", i.e. distant from the territory, is generated.

To sum up, our results suggest that the peer review process is not so objective as we expected. Moreover, other indicators should be taken into account apart from publications to enhance the impact of the research group activity on regional development. Furthermore, relatively little is found about the "real" formulas and criteria applied to evaluate a research group as excellent.

We encourage other researchers to conduct similar analyses as ours for other regions or countries. More generally, we encourage others to study whether measures of excellence in the supervision of peer reviewers can be usefully incorporated into more general analysis of research group productivity and compensation.

References

Albert, A., Plaza, L.M. (2004), The transfer of knowledge from the Spanish public R&D system to the productive sectors in the field of Biotechnology, *Scientometrics,* 59(1): 3-14.

Ball, D.F., Butler, J. (2004), The implicit use of business concepts in the UK Research Assessment Exercise, *R&D Management*, 34: 87-97.

Ballesteros, J.A., Rico, A.M. (2001), Public financing of cooperative R&D projects in Spain: the concerted project under the National R&D Plan, *Research Policy,* 30: 625-641.

Basu, A., Aggarwal, R. (2001), International collaboration in science in India and its impact on institutional performance, *Scientometrics,* 52: 379-394.

Bence, V., Oppenheim, C. (2004), The role of academic journal publications in the UK research assessment exercise, *Learned Publishing,* 17: 53-68.

Bordons, M., Fernandez, M.T., Gomez, I. (2002), Advantages and limitations in the use of impact factor measures for the assessment of research performance in a peripheral country, *Scientometrics,* 53: 195-206.

Bozeman, B., Corley, E. (2004), Scientists' collaboration strategies: implications for scientific and technical human capital, *Research Policy,* 33: 599-616.

Callon, M., Larédo, P., Mustar, P. (1994), Panorama de la science française, *La Recherche*, 264: 378-383.

CEC, Commission of the European Communities (2004), *Instruments for implementing the Thematic Priorities of the Sixth Framework Programme (2002-2006): Performance indicators in the frame of Networks of Excellence,* Commission Staff Working Paper, Brussels 21/10/2004.

CEC, Commission of the European Communities (2003), *Instruments for implementing the Thematic Priorities of the Sixth Framework Programme (2002-2006): Networks of excellence,* Commission Staff Working Paper, Brussels 12/5/2003.

CEC, Commission of the European Communities (2002), *Instruments for implementing the Thematic Priorities of the Sixth Framework Programme (2002-2006): Presentation of networks of excellence,* Commission Staff Working Paper, Brussels 11/11/2002.

CEC, Commission of the European Communities (2001), *How to map excellence in research and technological development in Europe,* Commission Staff Working Paper, 1.SEC (2001) 434, Brussels 12/3/2004.

Coccia, M. (2004), New models for measuring the R&D performance and identifying the productivity of public research institutes, *R&D Management,* 34: 267-280.

Cohen, W., Levinthal, D. (1990), Absorptive capacity: a new perspective on learning and innovation, *Administrative Science Quarterly,* 35: 128-152.

Cohen, W., Levinthal, D. (1989), Innovation and learning: the two faces of R&D, *The Economic Journal,* 99: 569-596.

Cole, S., Cole, J. (1967), Scientific output and recognition, *American Sociological Review,* 377-390.

Collins, H. M. (1982), Knowledge, norms and rules in sociology of science, *Social Studies of Science,* 12: 299-309.

COTEC (1999), *El sistema español de innovación. Diagnóstico y recomendaciones*, Fundación COTEC, Madrid.

CRUE (2001), *Networks of Excellence*. Directrices para la identificación de potenciales redes de excelencia en el VI PM, *Servicio Europa I+D de la CRUE (Conferencia de Rectores de las Universidades Españolas).* Discussion paper: Seminar on 20 April 2001.Ref: crue-i+d-010-2001.

Debackere, K., Glanzel, W. (2004), Using a bibliometric approach to support research policy making: the case of the Flemish BOF-key, *Scientometrics,* 59: 253-276.

Edge, D. O. (1979), Quantitative measures of communication in science: A critical review, *History of Science,* 17: 102-134.

Egghe, L., Rousseau, R. (2002), A general framework for relative impact indicators, *Canadian Journal of Information and Library Science,* 27: 29-48.

Fernández de Caleva, R. (2003), *Los comienzos de la evaluación científica en España*, report to Institut Municipal d'Investigació Mèdica, IMIM, Barcelona, Spain.

Florida, R., Cohen, W.M. (1999), Engine or infrastructure? The university role in economic development, in Branscomb, L.M., Kodama, F., Florida, R. (Eds.), *Industrializing knowledge: University-Industry linkages in Japan and the United States*, MIT Press, London, pp. 589-610.

Fox, M.F. (1983), Publication productivity among scientists: a critical review, *Social Studies of Science,* 13: 285-305.

Frohlich, C., Resler, L. (2001), Analysis of publications and citations from a geophysics research institute, *Journal of the American Society for Information Science and Technology,* 52: 701-713.

Generalitat Valenciana (2001), *Plan Valenciano de Investigación Científica, Desarrollo Tecnológico e Innovación (PVIDI)*, official document published by Secretaria del PVIDI de la Generalitat Valencia, Valencia, Spain.

Gillett, R. (1989), Research Performance Indicators Based on Peer-Review - A Critical Analysis, *Higher Education Quarterly*, 43(1): 20-38.

Goldfinch, S. (2003), Investing in excellence? The performance-based research fund and its implications for political science departments in New Zealand, *Political Science,* 55: 39-53.

Green, W. H. (1997), *Econometric Analysis,* third edition, Prentice-Hall, New Jersey.

Gulbrandsen, J.M. (2000), Research quality and organizational factors: An investigation of the relationship. Doctoral Dissertation, NTNU, Trondheim.

Gulbrandsen, M., Langfeldt, L. (2004), In search of 'Mode 2': The nature of knowledge production in Norway, *Minerva*, 42(3): 237-250.

Hane, G. (1999), Comparing university-industry linkages in the United States and Japan, in Branscomb , L.M., Kodama, F., Florida, R. (Eds.), *Industrializing knowledge: University-Industry linkages in Japan and the United States*, MIT Press, London, pp. 20-61.

Hare, P.G., Wyatt, G.J.(1992), Economics of academic research and its implications for higher education, *Oxford Review of Economic Policy,* 8(2): 48-66.

Harvey, J., Pettigrew, A., Ferlie, E. (2002), The determinants of research group performance towards mode 2?, *Journal of Management Studies,* 39(6): 747-774.

Hauser, J.R., Zettelmeyer, F. (1997), Metric to evaluate R, D & E, *Research Technology Management,* 40: 32-38.

Ho, Y.S., Chiu, C.H., Tseng, T.M., Chiu, W.T. (2003), Assessing stem cell research productivity, *Scientometrics,* 57: 369-376.

Horrobin, D. (1990), The philosophical basis of peer review and the suppression of on innovation, *Journal of the American Medical Association,* 263: 1438-1441.

INE (2002), Estadística sobre las actividades en investigación científica y desarrollo tecnológico (I+D), Madrid.

Johnes, G. (1990), Measures of research output: university departments of economics in the UK, 1984-88', *Economic Journal,* 100: 556-560.

Korhonen, P., Tainio, R. Wallenius, J. (2001), Value efficiency analysis of academic research, *European Journal of Operational Research,* 130: 121-132.

Kuldell, N. (2004), Scientific Writing: Peer review and scientific journals. *Visionlearning*, SCI-2(2).

Laredo, P. (2003), *Comment caractériser un laboratoire de recherche? La rose des vents de la recherche*, document published in the Laboratoire Techniques Territoires et Sociétés, nº 0/2003.

Leydesdorff, L. (1998), Theories of citation?, *Scientometrics*, 43: 5-25.

Loch, C.H., Tapper, U.A.S. (2002), Implementing a strategy-driven performance measurement system for an applied research group, *Journal of Product Innovation Management,* 19: 185-198.

McFadden, D. (1984), Econometric analysis of qualitative response models, in *Handbook of Econometrics*, vol. II, Griliches, Z., Intriligator, M.D. (Eds.), Elsevier, New York.

MEC, Ministry of Education and Science (2004), http://wwwn.mec.es/ciencia/jsp/plantilla.jsp?area=anep&id=22

Merton, R.K. (1968a), The Matthew effect in science. The reward and communication systems of science considered, *Science*, 159: 56-63.

Merton, R.K. (1968b), The normative structure of science, in Merton, R.K. (Ed.), *The Sociology of Science*, chapter 13, pp. 267-278.

Modrego, A. (1995), *Evaluación de los proyectos concertados del Plan Nacional de I+D*, report to Centre for Industrial Technological Development, CDTI, Madrid, Spain.

Moed, H.F., Luwei, M., Nederhof, A.J. (2002), Towards research performance in the humanities, *Library Trends,* 50: 498-520.

Moed, H.F., Burger, W., Frankfort, J., Van Raan, A. (1985), The use of bibliometric data for the measurement of university performance, *Research Policy,* 14:131-149.

Moxham, H., Anderson, J. (1992), Peer review: a view from the inside, *Science and Technology Policy,* 7-15.

Nagpaul, P.S., Roy, S. (2003), Constructing a multi-objective measure of research performance, *Scientometrics,* 56: 383-402.

Narin (1976), *Evaluative scientometrics: The use of publication and citation analysis in the evaluation of scientific activity*, Report US National Science Foundation.

Nowotny, H., Scott, P., Gibbons, M. (2003), Mode 2 revisited: the new production of knowledge-introduction, *Minerva*, 41(3): 179-194.

OECD (2004), *Education at a glance, OECD Indicators,* Paris, OECD.

OECD (1997), *National Innovation Systems,* Paris, OECD.

Ramsden, P. (1994), Describing and explaining research productivity, *Higher Education,* 28: 207-226.

Roberts, P. (1999), Scholarly Publishing, Peer Review, and the Internet , *Peer-Reviewed Journal on the Internet,* 4(4).

Sanz Menéndez, L. (1995), Research actors and the state: research evaluation and evaluation of science and technology policies in Spain, *Research Evaluation,* 5: 79-88.

Tijssen, R.J.W., Visser, M.S., Van Leeuwen, T.N. (2002), Benchmarking international sicentific excellence: are highly cited research papers an appropriate frame of reference?, *Scientometrics,* 54: 381-397.

Van Looy, B., Ranga, M., Callaert, J., Debackere, K., Zimmermann, E. (2004), Combining entrepreneurial and scientific performance in academia: towards a compounded and reciprocal Matthew-effect?, *Research Policy,* 33: 425-441.

Van Raan, A.F.J. (2000), The pandora's box of citation analysis: measuring scientific excellence – the last evil?, in Cronin, B., Barsky Artkins, H. (Eds.), *The web of Knowledge – A festschrift in honor of Eugene Garfield*, Medford (NJ), ASIS Monograph Series, pp. 301-319.

Weinstock, N. (1971), Citation indexes, in Kent, A. (Ed.), *Encyclopedia of Library and Information Science*, New York, Marcel Dekker, vol. 5, pp. 16-41.

Werner, B.M., Souder, W.E. (1997), Measuring R&D performance – state of the art, *Research Technology Management,* 37: 34-42.