Jointly published by Akadémiai Kiadó, Budapest and Kluwer Academic Publishers, Dordrecht

Scientometrics, Vol. 56, No. 3 (2003) 287–299

Evaluating two Austrian university departments: Lessons learned

CHRISTIAN SCHLOEGL, JUAN GORRAIZ, CHRISTOPH BART, MONIKA BARGMANN

University of Graz, Department of Information Science, Graz (Austria)

This paper describes various problems which may occur in quantitative research evaluation. It is shown that problems already arise when trying to define such seemingly simple scientometric elements as "personnel" or "budget". This has major consequences on the construction of indicators. Furthermore, it is demonstrated that different data sources as well as different data and indicators result in different, sometimes even contradicting outcomes.

Introduction

The Austrian university system has been undergoing significant changes for the past few years. In 1993, a new university organization law* was passed, which was followed by the so-called "evaluation regulation"** four years later. With these innovations, evaluations of teaching and research have been established legally in Austria for the first time. In the next step, it is planned to grant autonomy to universities by the end of 2002. This will further increase the importance of quality control at universities.

There is no consistent definition of the term "evaluation". For instance, in the Austrian evaluation regulation, it is described as a review of the effectiveness and efficiency of teaching, research, and other activities at universities (Evaluierungsverordnung 1997, § 1). Generally, there is a distinction between qualitative and quantitative evaluation. While the first one is based on more or less subjective opinions about the evaluated object, the latter relies on hard data only. This paper deals with quantitative research evaluation only.

* Universitaetsorganisationsgesetz 1993 (UOG 1993).
** [Oesterreichische] Evaluierungsverordnung 1997.

Received July 18, 2002

Address for correspondence:
CHRISTIAN SCHLOEGL
University of Graz, Department of Information Science
Universitätsstr. 15/F3, A-8010 Graz, Austria
E-mail:christian.schloegl@uni-graz.at

0138–9130/2003/US $ 20.00
Copyright © 2003 Akadémiai Kiadó, Budapest
All rights reserved

Case study

It was the goal of this study to identify potential problems, which may occur in a quantitative research evaluation. Especially, it will be investigated whether different data sources as well as different data and indicators will produce the same results. For this purpose, two Austrian university departments were compared on a quantitative basis. When selecting the two departments, attention was given to a high degree of similarity. Accordingly, two physics departments working in the same sub-field and having the same name ("department of theoretical physics") were chosen. They will be referred to as department A and department B in the remaining part of this paper.

Because of the research goal, various input and output data were used:

- input:
  - personnel: number of staff;
  - budget:
    - current expenses;
    - capital outlay;
    - funded money;
- output:
  - teaching:
    - curriculum: numbers of hours offered;
    - numbers of master and doctoral theses finished;
  - research:
    - numbers of publications;
    - numbers of citations received (impact).

The departmental reports of activity and bibliographic databases served as the data sources. The reports of activity, which are composed annually, contained data on budget, courses, staff, theses, and publications. The databases were selected to reflect the research topics of the departments. Furthermore, they were intended to enable citation analysis, if possible. According to these requirements, *High Energy Physics* (HEP), *Inspec,* and the *Science Citation Index* (SCI) were chosen.

HEP was established by the SLAC (Stanford Linear Accelerator Center) and the DESY (Deutsches Elektronen-Synchrotron) libraries in 1974. It includes more than 400,000 documents (January 2001), which can be articles in journals and proceedings, preprints, electronic publications, technical reports, or dissertations, in the field of high

energy physics. HEP is free of charge and accessible at various mirror sites.* Besides conventional search, this database offers peculiarities like citation summaries or top-cite ratings.

Inspec, with more than 6.4 million documents, is the most comprehensive bibliographic database on scientific and technical literature. In total, more than 4,000 journals and 2,000 conferences are included. The database covers mainly the areas of physics, electrical engineering, electronics, telecommunication, and computer science.

When using the SCI, it must be taken into account that only 3,500 journals are included in the CDROM version, compared to more than 5,300 journals in the SCI-Expanded. Since a trial search for one year did not show big differences, the CDROM version was applied for reasons of expense.

Problems encountered

In this chapter, all the problems that occurred during the evaluation of the departments will be discussed. They can be classified as

- problems with data sources;
- problems with definitions of the indicators.

Problems with data sources

During the evaluation process, it turned out that each data source had its specific shortcomings. The main insufficiency with the *reports of activity* is that they are not standardized. In one department, reporting is based on the calendar year, while the other department reports at the end of the academic year. Also the structure and the content of the reports are different in a few cases. For example, external funds were only mentioned in the summary of one department. Another problem was the use of different labels for the same notion. For instance, lecturers with PhD qualifications and retired professors were named "assigned staff" in department A. The same designation was applied to guest lecturers and project staff, who are often employed part-time, in department B. At the same time personnel hired for research projects were referred to as "temporary staff" in department A. For these reasons, data classification was difficult and very time consuming.

* E.g. http://www.slac.stanford.edu/spires/hep (November 2001).

The databases appeared to have various shortcomings as well. In SCI, there were mainly homonym and synonym problems with regard to the author names. Since only the initials of the first names are considered, getting the citation frequencies for an author with a common family name may take quite an effort. Another problem with the citations is that only the first author of a paper is included. Because SCI is a universal database, specific topics are better covered by subject databases like HEP.

This database has even more advantages. In contrast to SCI, the references in HEP contain all the authors of a cited paper. HEP also provides some citation statistics like citation summaries or top-cite rankings. The major disadvantage is that citations only refer to those publications that are included in the HEP database as well. If one bears in mind that HEP is free of charge, this database can be an attractive alternative to SCI.

The main deficiency of *Inspec* is that it does not record the reference lists of publications at all. Furthermore, it is only possible to search for the affiliation of the first author.

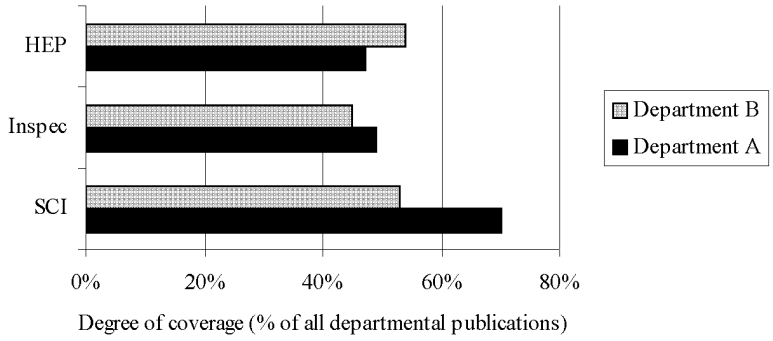

Figure 1. Average coverage of the publications of a department in SCI, *Inspec*, and HEP (1995–1999)

Figure 1 shows the coverage* of all the publications of a department in SCI, *Inspec,* and HEP. As can be seen, the degree of coverage varies among the databases and the

* The number of publications in the databases was determined by searching for the name and/or the address of the departments. Afterwards, the coverage was calculated by dividing the number of the departmental publications found in the database by the number of the publications from the reports of activity.

departments.* For this reason, it is advisable not to rely on only one data source when evaluating departments. The biggest difference in coverage between the departments occurred in SCI. A different research focus could be one reason. So department A may publish in areas that are well covered by SCI. Dissimilarities in publication behavior could be another explanation. The higher inclusion of publications from department B in HEP may indicate that this department does more research in high energy physics. As a consequence, one could conclude that the research scope of department A is broader. One reason for the generally low coverage in the data banks might be that relatively many papers are published only locally.

In order to analyze the cognitive orientation of the two departments, interviews were performed with various staff at the end of the case study. According to the interviewed professors, there are no other more similar departments of theoretical physics in Austria. The only main difference is that department B is more devoted to education in physics. The fact that many of these publications appear in German and in journals which are not included in the SCI can partly explain the difference in coverage between the two departments.

*Problems with definitions of the indicators***

Personnel and budget. As already mentioned, a distinction was made between different kinds of staff in the reports of activity. Apart from naming problems, different measurement units make it difficult to determine the total number of staff. Permanent staff is usually counted in full-time equivalents. This is not possible for project staff who are employed part-time for a certain time period or for a certain project. For this reason, using the total number of researchers in indicator definition could be controversial. This is the same for the total annual budget because capital outlay and external funds can differ considerably from one year to another. Furthermore, it can be difficult to get the data for such external funds.

Table 1 presents the data for staff and budget. It can be recognized that the figures between the departments vary strongly.

* A similar conclusion was drawn by a bibliometric study comparing departments of psychology (*Daniel*, 1988; p. 238).

** It must be remembered that the quality of the indicators depends on the quality of their underlying data sources as well (see previous subsection). The designation "indicator" is used according to *Vinkler* (2001).

Table 1. Mean values for staff and budget per year

		Staff			Budget			
	Depart-ment	Permanent (full time equivalent)	Assigned (no.)	Provi-sional (no.)	Current	Capital outlay	Total	External funds*
Mean value (1995/96 – 1999/2000)	A	16.4	8	34.8	1,156,780	807,970	1,964,750	8,983,772
	B	22.6	18.8	–	418,500	532,500	951,000	–

Curriculum and theses. As illustrated in Table 2, department B is more involved in teaching. This goes hand in hand with the fact that department B has more permanent staff.

Table 2. Mean values for teaching hours and completed theses per academic year

	Department	Courses (no. of hours)	Completed	
			Master theses	Dissertation theses
Mean value (summer semester 1995 – summer semester 1999)	A	269	10	5
	B	353	9	5

When collecting data, we found that not all listed courses had been held in department A. These courses were marked explicitly in the reports of activity. No corresponding remarks were found in the reports of department B. So we cannot automatically assume that all courses took place in department B.

The different types of courses (see Figure 2) are a further reason, which makes it difficult to directly compare the number of courses. While department A offers more laboratory classes, department B contributes more seminars. This could mean that department A has a more practice-oriented approach.

Publications. When determining the number of publications, the following issues are relevant:

- Which sources should be used for retrieving the publication data?
- How should different document types be dealt with?
- How should co-authored publications be treated?

* The values for external funds were only stated in the reports of activity of department A.

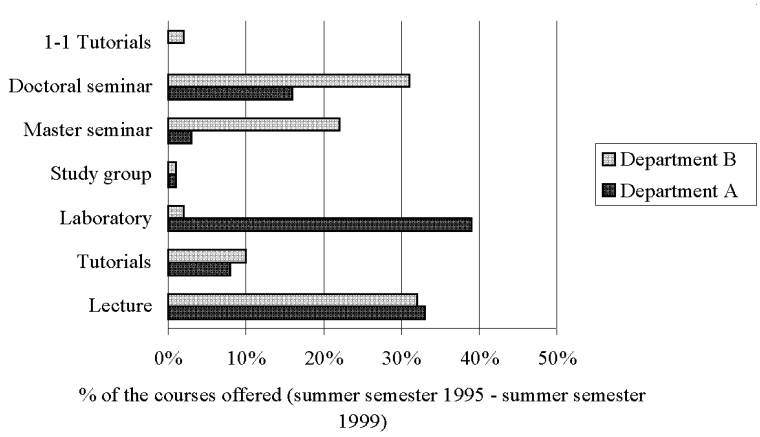

Figure 2. Structure of the courses (summer semester 1995 – summer semester 1999)

The first issue was already discussed in the previous section. Another important point is how to cope with different document types. Since, in physics, most publications are articles,* the weighting of different document types does not seem to have the same importance here as in other disciplines. The only exception were preprints, which were included in the reports of activity. Since they usually result in a publication, they were not considered. Thus, double-counting should be avoided.

Multi-authored papers are usually not such a big issue when evaluating departments as a whole. The share and the order of the authors should only be taken into account if publications are written together with authors from other departments. In order to give an impression of this issue, Table 3 lists how often an external author was ranked first. As can be realized in the case of department A, external authors head publications even more often than permanent staff. At department B, the first-authorship could only be determined for 1999, because only for this year the real order of the authors was given in the reports of activity. As can be seen, at least in this year permanent staff play a much more important role in this department.

* The reports of activity contained hardly any books. As an analysis of SCI showed, nearly all non-book publications were articles (94 %).

Table 3. Number of publications and first-authorship (based on reports of activity)

Year	Depart-ment	Publications (not including preprints)	First authorship		
			Permanent staff	Nonpermanent staff	External
1995	A	86	26	46	14
	B	51	–	–	–
1996	A	112	33	53	26
	B	63	–	–	–
1997	A	116	22	52	42
	B	68	–	–	–
1998	A	115	21	40	54
	B	71	–	–	–
1999	A	75	9	33	33
	B	82	40	26	16
Mean value	A	101	22	45	34
1995–1999	B	67	–	–	–

Impact. The determination of the total number of citations for the two departments turned out to be more difficult for various reasons. The collection of the citations is relatively easy in HEP where the citations can be retrieved by using the departmental name. However, attention has to be paid to the fact that citations are only counted from publications included in the HEP database as well. The gathering of the total departmental citations is much more time-consuming in SCI because of the primary-author problem. Because of the duration of the investigation period (five years) and the size of the departments, it would have gone beyond the scope of this study to collect the citation counts considering each individual publication. This would have been especially true for department B where the real order of a publication's authors was specified in the report of activity for the year 1999 only. For this reason, it was planned to determine the citation frequencies on the basis of the personnel. Because of the high share of non-permanent staff (see Table 1) that changes often, this would still have been excessive. In addition, it would have been difficult to assign citations unambiguously between guest lecturers and departments. Therefore, it was decided to gather citation counts only for permanent staff.*

* Of course, homonym checks were performed for all hits.

Table 4 reveals the citation frequencies of the two departments in SCI and HEP. The divergence between the total citation counts in SCI and HEP can be explained by the different data sources (with their underlying shortcomings) and the collection methods used (see above). Self-citations make up a high proportion of the total citations in which the values in HEP (department A: 42%, department B: 25%) might be more realistic.

Table 4. Number of citations in SCI and HEP

		Citations in SCI (permanent staff)						Citations in HEP (department)			
Year	Depart-ment	No. of citations received			Self-citations		Citations received from other depart-ment	No. of citations received (total)	Self-citations		Citations received from other depart-ment
		Total	for publications		Same author	Same depart-ment			Same author	Same depart-ment	
			<1995	>=1995							
1995	A	286	278	9	75	8	0	351	166	46	8
	B	482	468	14	16	12	7	467	67	12	1
1996	A	317	253	64	88	17	0	469	140	14	6
	B	533	489	44	73	11	8	488	130	13	3
1997	A	499	383	116	110	43	0	358	157	7	0
	B	529	425	104	59	16	7	361	93	12	0
1998	A	339	230	109	40	16	0	323	142	2	0
	B	481	310	171	33	8	5	366	116	8	0
1999	A	393	238	155	66	12	1	130	75	2	0
	B	480	323	157	35	21	4	294	90	5	5
Total	A	1834	1382	453	379 21%	96 5%	1 0%	1631	680 42%	71 4%	14 1%
	B	2505	2015	490	216 9%	68 3%	31 <1%	1976	496 25%	50 3%	9 <1%

In SCI, a distinction was made between citations on articles which were published before 1995 and on those which appeared after and including 1995. Thus, the aging of information in the papers could be accounted for to some extent. As can be seen in Table 4, most citations were obtained on articles which date back further into the past. For instance, from the citations received in 1999, 61% (238) at department A and 67% (323) at department B refer to publications from before 1995.

It was surprising that department B obtained more citations both in SCI and in HEP than department A even though it did not produce so many publications. In order to enable a more objective comparison, three citation indicators (number of citations per researcher* and mean impact factor of the journals used for publication) were computed.

1) Citations per permanent researcher in SCI (CPR^{SCI}):

This impact indicator is calculated as follows:

$$CPR_i^{SCI} = \frac{C_i^{SCI}}{N^p}$$

C_i^{SCI} ...no. of citations received in SCI in the year i by permanent staff

N^p ...mean number of permanent staff (in full time equvalents) between accademic year 1995/96 and 1999/2000

2) Citations per researcher in HEP (CR^{HEP}):

Attention must be paid to the fact that in HEP the citations are counted only from publications included in this database as well. However, since the citations can be retrieved by using the departmental name, they refer to the whole department. For this reason, the citations were related to the total staff.**

$$CR_i^{HEP} = \frac{C_i^{HEP}}{N}$$

C_i^{HEP} ...no. of citations received in HEP in the year i by a (whole) department from publications included in HEP only

N...mean number of total staff between accademic year 1995/96 and 1999/2000

3) Publication strategy index (PSI):

This indicator is computed as follows: sum of impact factors of journals weighted by the number of papers published in them divided by the number of papers. In literature, it is referred to as publications strategy index (Vinkler, 1997).

---

* According to *Vinkler* (1998, p. 188), this impact productivity index is to be given preference over citations per paper. Contrary to the citations per paper index, this indicator is not affected by the publication productivity.

** As already discussed (personal and budget), it could be contentious to use the *total* number of staff in indicators.

$$PSI_i = \frac{\sum_{j=1}^{P_i} IF_{ij}}{P_i}$$

P_i ...number of papers in journals with an IF published in the year i

IF_{ij} ..impact factor of the j^{th} paper (i.e. journal) in the year i

As can be seen in Table 5, the average number of citations per *permanent* researcher (in SCI) is approximately the same for both departments. However, the comparison in HEP shows that department B lists 73% more citations per researcher than department A. The difference between the SCI and HEP based indicators is due to the fact that department B has more permanent staff (22.6 vs. 16.4) whereas department A has more nonpermanent staff (42.8 vs. 18.8). On the assumption that permanent staff usually have a higher impact, department A was at a disadvantage with regard to the "citations per researcher" indicator in HEP. This again shows the difficulty of defining indicators .

Table 5. Comparison among different citation indicators

Year (i)	Department	CPR_i^{SCI}	CR_i^{HEP}	PSI_i
1995	A	17.4	5.9	1.8
	B	21.3	11.3	2.48
1996	A	19.3	7.9	1.66
	B	23.6	11.8	2.54
1997	A	30.4	6.0	2.17
	B	23.4	8.7	3.26
1998	A	20.7	5.5	3.22
	B	21.3	8.8	2.15
1999	A	24.0	2.2	2.07
	B	21.2	7.1	2.27
Mean value	A	22.4	5.5	2.13
	B	22.2	9.5	2.53

Department B has a higher publication strategy index as well what means that it publishes in journals with a higher impact factor. Because department B has more citations on its papers, the impact of its individual articles is also higher.*

Conclusion

It was the goal of this paper to demonstrate problems which may occur in a quantitative research evaluation. This was done by comparing two similar Austrian university departments. During this effort, various problems were encountered. It turned out that it can be even quite difficult to define scientometric elements such as "personnel" or "budget". This can have severe consequences when defining indicators, for instance "number of papers per researcher". Furthermore, it was shown that different data and different indicators** can result in quite different outcomes. As exhibited in Figure 3, department A has more than twice as much nonpermanent staff and budget.

	Dept. A	Dept. B	% by which one dept. is better than the other (worse dept. = 100%)	
Input data:				
Personnel:				
Permanent staff (full time equivalents)	16.4	22.6		**38%**
Nonpermanent staff (no.)	42.8	18.8	**128%**	
Budget:				
Current budget (Euro)	84.066	30.414	**176%**	
Capital outlay (Euro)	58.717	38.698	**52%**	
Output data/indicators:				
Teaching:				
Courses (no of weekly hours)	269	353		**31%**
Master theses finished	10	9	**11%**	
Dissertation theses finished	5	5		
Research:				
No. of publications (reports of activity)	101	67	**51%**	
No. of citations (SCI, only permanent sta	366.8	500.8		**37%**
No. of citations (HEP)	326.2	395.2		**21%**
Citations per permanent researcher (SCI	22.4	22.2	**1%**	
Citations per researcher (HEP)	5.5	9.5		**73%**
Publication Strategy Index	2.13	2.53		**19%**

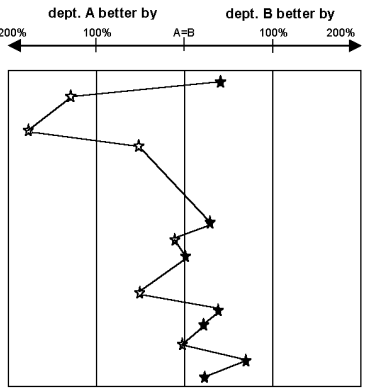

Figure 3. Comparison of different data and indicators

* A higher publication strategy index does not necessarily include a higher impact of the individual articles. This is one source of inaccuracy of this indicator (*Korwitz*, 1995; p. 269 ff.; *Moed* and *Van Leeuwen,* 1995).
** Because of the problem with the definition of the indicators, in most cases absolute data were compared.

However, department B has about 40% more permanent staff which corresponds approximately to the extent to which it is more engaged in teaching. Concerning completed master and dissertation theses, both departments have roughly the same output. The output of publications is 50% higher in department A. However, department B scores better on the total number of citations received, citations per researcher in HEP (73%) and the publication strategy index (19%).

It can be concluded that a quantitative research evaluation should not rely on only one data source and on few data and indicators. Rather, different perspectives are needed to get the full picture and to avoid distorted views (*Lewison*, 1998; p. 15). Even if all these suggestions are taken into account, one should never rely on quantitative data alone. Rather, as some authors suggest, a more comprehensive approach is recommended in which bibliometric methods are complemented by peer review (*King*, 1987; p. 273; *Vinkler*, 1998; p. 199).

*

The authors would like to thank Mr. Gray Chamberlin from the University of Arkansas at Little Rock and Mr. N. G. Berger from the University of Graz for proof reading.

References

ANONYMUS, [Oesterreichische] *Evaluierungsverordnung (EvalVO)* (Verordnung des Bundesministeriums fuer Wissenschaft und Verkehr ueber Grundsaetze fuer die Durchfuehrung von Evaluierungen in Forschung und Lehre der Universitaeten. BGBlNr. II 224/1997), Bundesministerium fuer Wissenschaft und Verkehr, Vienna, 1997, http://www.bmwf.gv.at/3uniwes/03unirecht/evalvo/evalvo.htm (state: November 2001).

DANIEL, H.-D., Methodische Probleme institutsvergleichender Analysen der Forschungsproduktivitaet, In: H.-D. DANIEL, R. FISCH (Eds), *Evaluation von Forschung: Methoden, Ergebnisse, Stellungnahmen,* Universitaetsverlag Konstanz, Konstanz, 1988, 215–241.

KING, J., A review of bibliometric and other science indicators and their role in research evaluation, *Journal of Information Science*, 13 (1987) 261–276.

KORWITZ, U., Welchen "Rang" hat ein Wissenschaftler? *Nachrichten fuer Dokumentation*, 46 (1995) 267–272.

LEWISON, G., New bibliometric techniques for the evaluation of medical schools, *Scientometrics*, 41 (1998) 5–16.

MOED, H. F., TH. N. VAN LEEUWEN, Improving the accuracy of institute for Scientific Information's journal impact factors, *Journal of the American Society for Information Science*, 46 (1995) 461–467.

VINKLER, P., Relations of relative scientometric impact indicators. The Relative Publication Strategy Index, *Scientometrics*, 40 (1997) 163–169.

VINKLER, P., General performance indexes calculated for research institutes of the Hungarian Academy of Sciences based on scientometric indicators, *Scientometrics*, 41 (1998) 185–200.

VINKLER, P., An attempt for defining some basic categories of scientometrics and classifying the indicators of evaluative scientometrics (Opinion Paper), *Scientometrics*, 50 (2001) 539–544.