# Evaluation of the publication activity of research teams by means of scientometric indicators*

**Peter Vinkler**

Chemical Research Center, Hungarian Academy of Sciences, Pusztaszeri út 59-67, 1025 Budapest, Hungary

**In the Chemical Research Center of the Hungarian Academy of Sciences, special scientometric indicators have been used for evaluating publication activity of research teams for about 30 years. Modified Garfield impact factors for journals as well as relative citedness of papers are applied as indicators because of differences among subfields in scientometric features of the publications assessed. Our experience has shown that the evaluation of real scientometric systems needs compromises among the parties interested and between the practical applicability and the theoretical requirements of scientometrics.**

## Introduction

Scientific research operates with information as input and produces information as output. Input information is in part already disclosed and in part it originates from the researchers themselves who conduct the respective research work. Output information is (or at least should be), however, novel and/or reorganized knowledge which is disclosed in the form of publications. The primary goal as well as innate need for any individual or team working in scientific research is to contribute to the scientific information production. Consequently, the evaluation of scientific activity should cover the evaluation of publications, both qualitatively and quantitatively. The qualitative assessment may refer, here, to measuring the *international impact* of publications manifested in citations.

The fundamentals of publication evaluation by means of scientometric methods have been laid down by Moravcsik[1], Martin and Irvine[2], Moed *et al.*[3] and Martin[4].

Vinkler reported on the assessment methods applied within an institute for teams working on the same field (chemistry) but on different subfields (organic, physical, bioorganic and polymer chemistry)[5,6], and for institutes working on different fields (chemistry, physics, mathematics, engineering sciences, geo-sciences and computer sciences)[7]. The literature of evaluations by scientometric methods is vast. Instead of recalling all requirements for and conditions of scientometric assessments applied or recommended, I wish to refer here only to one of the most important points, namely to the elaboration and application of appropriate scientometric indicators which correspond to the characteristics of the local system evaluated and meet the goals of the respective assessment and which are based on relevant data.

In real (practical) scientometric systems consisting of several parts to be assessed, the main problem is caused by the difference in the factors as follows:

- Scientometric features (here publication and citation characteristics);
- Size (i.e. research capacity, number of staff);
- Fields and subfields and type of activity. (Type of activity means basic or applied research or development.)

## Applicability of indicators in comparative assessments

In comparative scientometric assessments of various organizations forming a system, three important features should be taken into account: *completeness, thematical homogeneity* and *research capacity.* Some of the most frequently used scientometric indicators for evaluations are given in Table 1. For the classification of the indices, see Vinkler[8].

Scientometric systems may be complete or partial. The former ones may represent the total scientometric system to which belong the organizations studied, e.g. a set of organic chemistry papers abstracted in *Chemical Abstracts* published by all countries in a year. (As an approximation, one could use the set of organic chemistry papers referenced in *Science Citation Index.*) In contrast, the organic chemistry papers published by a country or an institute may be assumed as partial systems.

Scientometric indicators, e.g. publication productivity with a dimension of [number of papers/number of researchers × number of years], or citedness of papers [number of citations/number of papers], etc. may be similar for research teams working on thematically homogeneous basic science fields or subfields (e.g. reaction kinetics and

*Dedicated to Eugene Garfield, one of the founding fathers of scientometrics, on his 75th birthday.
e-mail: pvinkler@cric.chemres.hu

photochemistry). Indicators for thematically different fields, however, (like genetics and mathematics) must be very different. The above statements do not preclude that identical scientometric data and indicators could not be observed for fields which are thematically very far from each other.

An important factor to be considered in evaluating research organizations is their *size*. There are teams working only with 2 to 3 staff members whereas others may have a capacity of, say, 15 to 20 researchers.

In Table 1 there are suggestions on the applicability of scientometric evaluation indicators most frequently used[4]

**Table 1.** Applicability of some scientometric indicators for evaluation purposes. Some frequency data (*M*%) of the indicators applied in practice are given after Martin[4]

| | | | Indicator can be recommended for comparative evaluation of systems which are | | | |
|---|---|---|---|---|---|---|
| | | | Complete | | Partial | |
| | | | Thematically | | | |
| | | | Homogeneous | | Inhomogeneous | |
| Type/name of the indicator | *M*% | Possibility of absolute standard | By size | | | |
| *Gross indicators* | | | Identical | Different | Identical | Different |
| Number of publications | 59.5 | | | | | |
| Number of journal papers | | n.a. | 1 | 4 | 4 | 4 |
| Number of conference proceedings | | n.a. | 1 | 4 | 4 | 4 |
| Number of journal papers in *SCI* | | n.a. | 1 | 4 | 4 | 4 |
| Sum of GFs of journals where the papers studied were published | 7.4 | n.a. | 1 | 4 | 4 | 4 |
| Number of citations obtained | 31.4 | n.a. | 1 | 4 | 4 | 4 |
| Number of outstandingly cited papers | 1.7 | n.a. | 1 | 4 | 4 | 4 |
| *Specific indicators* | | | | | | |
| Number of publications per researcher | | o | 1 | 1 | 1 | 1 |
| Number of papers per researcher | | o | 1 | 1 | 2 | 2 |
| Number of citations per researcher | | o | 1 | 1 | 4 | 4 |
| Sum of GFs per paper | | o | 1 | 1 | 4 | 4 |
| Number of citations per paper | | o | 1 | 1 | 4 | 4 |
| *Distribution indicators* | | | | | | |
| Percentage number of the publications of the organization studied in the total | 11.6 | ~ | 1 | 4 | 4 | 4 |
| Percentage number of the citations of the organization studied in the total | | | 1 | 4 | 4 | 4 |
| Ratio of *SCI* papers in the total | 4.1 | ~ | 1 | 1 | 4 | 4 |
| *Relative indicators* | | | | | | |
| Relative Citation Rate (RCR) | 26.4 | + | 1 | 1 | 1 | 1 |
| Relative Subfield Citedness (RW) | | + | 1 | 1 | 3 | 3 |
| Relative Publication Strategy (RPS) | | + | 1 | 1 | 3 | 3 |
| Relative Reference Strategy (RRS) | | + | 1 | 1 | 3 | 3 |

Absolute standard can be calculated if the data are known for the whole respective system.
+: Absolute standard can be calculated;
~: Absolute standard can be calculated for identical subfields and sizes;
o: There are several data for different subfields using different time windows, no generally accepted standards or methods for calculation exist;
1: Recommended;
2: Application is recommended only with exceptional care;
3: Recommended if standard for the total system is available;
4: Not recommended;
RCR: Number of citations obtained per sum of GFs of the respective papers (journals);
RW: Number of citations obtained per mean citedness of papers (journals) of the respective subfield multiplied by the number of papers published;
RPS: Sum of GFs of papers (journals) published per mean citedness of papers of the respective subfield multiplied by the number of papers published;
RRS: Mean of GFs of papers (journals) referenced per mean citedness of papers (journals) of the respective subfield;
For thematically homogeneous *partial systems* consisting of parts of identical size, the same criteria can be applied as for *complete systems* with the same characteristics.

depending on the dichotomies mentioned. It is obvious that in ideal cases each indicator might be used, i.e. for a complete system consisting of thematically homogeneous teams with identical research capacity. In practice, however, the assessment attempts refer to 'real systems', i.e. comparative evaluations of some research teams (far from the total of the world) working on relatively different fields with different numbers of researchers. Therefore, I completely agree with Hicks[9] that each evaluation process is a research task as well.

In comparative assessments gross indicators cannot be used (Table 1); they should be converted into specific ones which are comparative, as far as, for example, size is concerned (e.g. 'impact factor' for periodicals with different number of papers). If we had absolute standards, like indicators with the dimension [papers/researcher × year] as world averages for chemists and physicists and biologists, etc. the performance of the individual teams working on different fields could be made comparable. For measuring the absolute contribution of an organization (person, team, country, etc.) to world science, however, distribution indicators should be applied. The percentage share in publications and citations may express the relative weight of the organization studied in world science.

Considering the potential of the indicators in Table 1, one may conclude that in real systems only relative indicators could be used, in principle. The use of relative indicators requires, however, absolute standards, which are not always available.

Scientometric assessments can be performed in two ways:

- By comparing performances of the parts of the system studied to one another or;
- By relating each part to a common standard or to individually selected absolute (international) standards.

## Introduction of the Publication Assessment Method

This paper describes a scientometric method applied in the Institute of Chemistry, Chemical Research Center, Hungarian Academy of Sciences, for the evaluation of scientific publications.

The Directory Board of the Institute nominated a Publication Committee in the 70s which consisted of representatives of the research subfields cultivated in the Institute. The Committee had the task of elaborating a comprehensive scientometric assessment method for the publications of the teams and to perform the annual assessment. At the beginning very simple methods were used[5], whereas later more sophisticated indicators have been introduced[6].

During the decades elapsed, the assessment method has become an integral part of the management system of the Institute. About 50 per cent of the grants were distributed annually among the teams on the basis of the score system applied. According to an opinion poll conducted in 1992, the satisfaction rate of the Publication Assessment Method (PAM) was high. The answers of 59 researchers (28.1 per cent of the total staff at that time) were distributed as follows:

- 55 researchers (93 per cent) agreed that *quantitative PAMs* should be applied for distributing grants;
- 83 per cent of the respondents agreed that the *impact factor* of journals should be one of the determining factors in the method applied;
- 86 per cent of the respondents were of the opinion that the modified number of *citations* (used, e.g. in relative indicators) may appropriately reflect the scientific level of the results published.

One of the early results of the introduction of PAM was an increase in the number of papers published in journals indexed by *Science Citation Index* from 40 per cent to 80 per cent between 1975 and 1983. The reasons for introducing PAM were as follows:

- The grant offered to the Center by the Hungarian Academy of Sciences was limited;
- The Board of Directors wanted to increase the grant for teams working at the international level;
- Scientometric methods appeared to be more objective than local peer review assessments.

## Characteristics of the PAM applied

The main goal of the PAM for a short term has been to calculate indices which are appropriate for distributing grants among the research teams in the Institute, whereas its long-term aim has been to contribute to the improvement of the international scientific level of the research activity in the Institute.

The organizations assessed are research teams consisting of 5 to 15 researchers. The authors of the publications assessed are staff members of the teams in the current year and in the preceding two years. PAM takes into account the performance of teams, thus the results attained by persons who were members of the team during the period assessed may be seen as those of the team. Only publications which indicate the Chemical Research Center as affiliation of at least one of the authors are accepted by the Committee.

There are several outputs characteristic of the research performance; nevertheless PAM applies only some of them, viz.

- Eminence of the journals publishing the papers of the team;
- Number of papers published;
- International impact of the results published;
- Length and number of publications other than journal

papers (e.g. books, book-chapters, conference proceedings).

Eminence of the journals is characterized by Garfield (impact) Factors (GF), whereas international impact is approximated by the number of citations related to international standards (see below). All the data are reliable and strictly controlled by the administration staff and the research teams evaluated. (There is a computerized data bank for the publications and citations in question.)

The various output indices of scientific research (e.g. number of papers published, citedness of publications, number of lectures given, number of scientific prizes, number of editorial boards and memberships in different committees, number of Ph D theses) suggested and applied widely in the literature for characterization of scientific performance of individuals or teams are not independent variables. The evaluation process applied in practice should be consistent and easily controllable. All the three parties involved in the process, namely science policy makers, who decide on the goal of the assessment and take measures on the basis of its results, scientometricians and experts in evaluation methods, who elaborate the methods and perform the whole assessment process, and finally persons or teams or institutions to be evaluated, should agree on the methods to be applied. Consequently, the assessment processes used regularly and resulting in consequences may apply only some well-selected basic indicators.

## Assessment of journal papers by the GF of the publishing journals

In the Institute, several thematically different research fields are cultivated (e.g. organic, bioorganic, physical, polymer, analytical and theoretical chemistry) which have different scientometric characteristics. The fields mentioned could be divided, however, further to subfields, e.g. photochemistry, surface chemistry, environmental chemistry, kinetics, corrosion science, electrochemistry, catalysis, etc.

GFs are field and subfield-dependent[10]. Consequently, comparison of publication strategy, mean citedness, etc. referring to research teams across fields or subfields would require reference standards. There are several attempts in the literature for the normalization of GFs of journals. Marshakova-Shaikevich[11] suggests selecting five journals with the greatest GF in each subfield and dividing the sum of their citations by the number of papers published by them. The resulting group impact factor could serve as reference standard for each respective subfield. Sen[12] makes a normalization by giving a value of 10 to the journal of the highest GF on a subfield. Relating the GF of any journal on the field to the maximum value and dividing 10 by this ratio, he obtains *normalized impact factors*.

Both methods mentioned use the journal classification scheme given by *SCI JCR*, where there are 'fields' with a couple of journals (e.g. business, biomethods, remote sensing and sport sciences) and those with several hundreds of them (e.g. pharmacology, biochemistry and molecular biology). The measure of overlaps between the fields is different from field to field. The problem of multidisciplinary journals (e.g. *Journal of the American Chemical Society*) cannot be eliminated by the methods suggested. One of the problems concerning the normalizations mentioned is that there are fields without review journals and others with several reviews and advances which have mostly high GF values. Therefore, it seems to be more appropriate to calculate with data referring to each journal of the respective field. It is obvious that the problem of standardizing GFs by subfields can be traced back to that of the classification of journals. As far as classification of journals is concerned, the application of the concept of Hirst[13] can be preferably recommended. According to this method, one can start with a single journal dedicated to a subfield (e.g. *Journal of Organic Chemistry*) and accept journals preferably referenced by it as a set of primarily important information sources of the respective subfield[14].

Some other, more sophisticated normalization processes have been recommended by Moed *et al.*[15] and Glänzel and Schubert[16].

A selection of a standard set of journals corresponding to the activity of research teams in the Institute would be rather complicated. Some of the reasons for that are:

- The activity of the teams is changing with time;
- The spectrum of the research topics of the teams is relatively wide;
- Publications in multidisciplinary journals represent a relatively great share in the total;
- There are many papers published in cooperation with teams working on different fields.

Nevertheless, we may apply reference standards given in Table 2 ($GF_m$) for some research teams.

The subfields in Table 2 can be classified into three groups by their mean GF values. Biochemistry and molecular biology ($GF_m = 2.897$) represents one of the subfields of life sciences showing high mean GF compared to mathematics and natural sciences. Mean GFs of subfields of chemistry (general, organic, physical, analytical, medicinal) and pharmacology range from 1.478 (chemistry, general) to 1.968 (chemistry, physical). Polymer sciences, applied chemistry and chemical engineering belong to the third group showing a mean GF value of 0.734. The distribution of the GFs of chemistry journals does not correspond to a normal distribution in most cases. Skewness of distribution of GF values ranges from 1.040 (chemistry, applied) to 6.254 (chemistry, analytical).

In order to make comparisons between GFs of journals belonging to different subfields, a simple normalization procedure may be applied as follows.

$$GF_{n,s,i} = \frac{GF_{m,s_{max}}}{GF_{m,s}} GF_{s,i} = SF_s \cdot GF_{s,i}\,, \quad (1)$$

where $GF_{n,s,i}$ is the normalized Garfield Factor of the *i*th journal on subfield s, $GF_{m,s_{max}}$ is the highest value of the mean GF data among the subfields studied, $GF_{m,s}$ is the mean GF of subfield s, and $GF_{s,i}$ is the mean GF of the *i*-th journal on subfield s, $SF_s$ is the *subfield factor* referring to subfield s.

Subfield factors (SF) are multiplicative factors, which might equalize differences in scientometric features among subfields. The precondition for the application of SF values is that the subfields selected should be relatively distinct with few overlaps with others and should show similar distribution by GF values. Organic chemistry, polymer chemistry, analytical chemistry and applied chemistry, may meet this requirement.

According to eq. (1), $GF_{n,s,i}$ is equal to 1.81 and 3.47 for papers which appeared in an organic chemistry and a polymer journal of GF = 1.0, respectively. Normalized Garfield Factors ($GF_{n,s,i}$) might be applied as *scores* for characterizing the value of papers in an assessment process. The unity to 1.92 ratio in favour of polymer papers to those in organic chemistry has been proved, however, unacceptable for the researchers evaluated and even for the Board of Directors of the Institute. Therefore, the concept of normalization has been dropped. Nevertheless, the problem of scoring papers originating from different subfields by the eminence of journals remains to be solved.

There are no other measures characterizing impact of journals which would be available annually and credited by an internationally acknowledged organization, than GFs published by the Institute for Scientific Information, Philadelphia, which was established by one of the founding fathers (E. Garfield) of scientometrics. Consequently, the Committee decided to apply GFs for the assessment of papers. [The Committee is aware of the contradictions and errors in the calculation of impact factors (see, e.g. Moed *et al.*[15]), but their application with some modifications (see later in the article) or in calculating relative indicators makes them acceptable for us.]

**Table 2.** Distribution of Garfield (impact) Factors for chemical subfields by categories, and Subfield Factors and statistical data

| Garfield Factor range | Number and per cent of journals | Chemistry: General | Organic | Physical | Polymer | Analytical | Applied | Engineering | Medical | Biochemistry and molecular biology | Pharmacology |
|---|---|---|---|---|---|---|---|---|---|---|---|
| GF ≤ 0.20 | *n* | 17 | 3 | 1 | 11 | 2 | 7 | 25 | 1 | 10 | 1 |
| | % | 23.94 | 6.67 | 1.10 | 16.42 | 3.08 | 15.91 | 22.73 | 3.33 | 3.44 | 0.58 |
| 0.20 < GF ≤ 0.60 | *n* | 23 | 6 | 14 | 22 | 13 | 13 | 46 | 7 | 26 | 41 |
| | % | 32.39 | 13.33 | 15.38 | 32.84 | 20.00 | 29.55 | 41.82 | 23.33 | 8.93 | 23.56 |
| 0.60 < GF ≤ 1.00 | *n* | 9 | 11 | 18 | 14 | 17 | 7 | 26 | 2 | 51 | 32 |
| | % | 12.68 | 24.44 | 19.78 | 20.90 | 26.15 | 15.91 | 26.64 | 6.67 | 17.53 | 18.39 |
| 1.00 < GF ≤ 2.00 | *n* | 12 | 11 | 33 | 15 | 22 | 15 | 12 | 14 | 82 | 66 |
| | % | 16.90 | 24.44 | 36.26 | 22.39 | 33.85 | 34.10 | 10.91 | 46.67 | 28.18 | 37.93 |
| 2.00 < GF ≤ 3.00 | *n* | 2 | 8 | 14 | 3 | 4 | 2 | 0 | 2 | 51 | 13 |
| | % | 2.82 | 17.78 | 15.38 | 4.48 | 6.15 | 4.55 | 0.00 | 6.67 | 17.53 | 7.47 |
| 3.00 < GF ≤ 4.00 | *n* | 1 | 3 | 3 | 1 | 4 | 0 | 0 | 2 | 17 | 8 |
| | % | 1.41 | 6.67 | 3.30 | 1.49 | 6.15 | 0.00 | 0.00 | 6.67 | 5.84 | 4.60 |
| GF > 4.00 | *n* | 7 | 3 | 8 | 1 | 3 | 0 | 1 | 2 | 54 | 13 |
| | % | 9.86 | 6.67 | 8.79 | 1.49 | 4.62 | 0.00 | 0.91 | 6.67 | 18.56 | 7.47 |
| $GF_m$ | | 1.478 | 1.597 | 1.968 | 0.830 | 1.557 | 0.783 | 0.589 | 1.610 | 2.897 | 1.663 |
| SD | | 3.054 | 1.261 | 2.462 | 0.799 | 2.305 | 0.633 | 0.879 | 1.175 | 4.447 | 2.236 |
| *N* | | 71 | 45 | 91 | 67 | 65 | 44 | 110 | 30 | 291 | 174 |
| Median | | 0.480 | 1.129 | 1.333 | 0.607 | 1.013 | 0.596 | 0.401 | 1.486 | 1.617 | 1.137 |
| Skewness | | 4.403 | 1.188 | 4.027 | 2.265 | 6.254 | 1.040 | 7.331 | 1.172 | 5.069 | 5.363 |
| SF | | 1.96 | 1.81 | 1.47 | 3.47 | 1.86 | 3.70 | 3.30 | 1.80 | 1.00 | 1.74 |

Classification of the subfields corresponds to that in *Science Citation Index*, *Journal Citation Reports*, 1998 except for chemistry, general, which covers chemistry journals publishing papers belonging to several subfields, e.g. *Journal of American Chemical Society*, *Angewandte Chemie Int. Ed.*, *Acta Chim. Hung. – Models in Chemistry.*
*n*: Number of journals in the respective class;
%: Per cent of journals;
*N*: Total number of journals;
$GF_m$: Mean GF;
SD: Standard deviation;
SF: Subfield Factor; SF = $GF_{max}/GF_i$;
$GF_{max}$ = 2.897;
$GF_i$ = GF of the *i*th subfield.

Because of the great discrepancies in the GF values by subfields, the low GF values were suggested to be increased significantly, and at the same time, an upper limit was introduced by the Publication Committee.

Table 2 shows the number and percentage shares of journals grouped into seven ranges by GF values. The percentage shares of journals with GFs lower than 0.20 is relatively low for organic, physical, analytical and medicinal chemistry, and biochemistry and molecular biology and pharmacology. The share of low GF journals is relatively high for subfields which have relatively low mean GF values (polymer science, applied chemistry, engineering). The subfields mentioned show very low shares of journals with GFs higher than 4.0. Only the share of journals with GFs higher than 4.0 attributed to biochemistry and molecular biology is greater than 10 per cent (18.56%). The distribution of journals according to GFs in Table 2 gives support to the scoring system suggested by the Publication Committee as follows.

The Publication Score ($GF_s$) given to a paper published in a journal with $GF_i$ is calculated as:

$$GF_s = f_i(GF_i + 0.2), \qquad GF_{s_{max}} = 4.00, \tag{2}$$

where $f_i$ is the respective Team Credit Share determined by an algorithm (see later in the article).

Table 3 gives some examples for converting GFs to publication scores.

GFs change slowly but significantly with time[17]. Therefore, when evaluating papers published, e.g. in 1999, we apply mean GFs of journals for the years 1996 to 1998.

About 10 per cent of the papers evaluated by the Committee are published in journals which have no GF value. The Committee considers the lifetime, topic, editors, language, form, and type (i.e. whether the journal is devoted to basic or applied science, or for letters or reviews, etc.) of the respective journal, and takes into account analogies when deciding on the scores. Journals written in national languages (e.g. Hungarian, Polish, Czech, etc.) are given generally GF = 0.1 ($GF_s$ = 0.1 + 0.2 = 0.3 scores).

## Distribution of credit among co-authors

### *Team credit share*

Co-authorship is recently a common phenomenon in natural sciences. Most of the papers are published in cooperation. The measure of the score of a paper was suggested

**Table 3.** Scores given by the GF of journals and by the page number (b) of publications (books, book-chapters, conference proceedings)

| GF | $GF_s$ | Increase in per cent | $\frac{100 \cdot GF_s}{4.00}$ | b | $b_s$ | $\frac{100 \cdot b_s}{5.38}$ |
|---|---|---|---|---|---|---|
| 0.1 | 0.3 | 200.0 | 7.5 | 1 | 0.23 | 4.3 |
| 0.2 | 0.4 | 100.0 | 10.0 | 2 | 0.44 | 8.2 |
| 0.3 | 0.5 | 66.7 | 12.5 | 3 | 0.64 | 11.9 |
| 0.4 | 0.6 | 50.0 | 15.0 | 4 | 0.82 | 15.2 |
| 0.5 | 0.7 | 40.0 | 17.5 | 5 | 1.00 | 18.6 |
| 0.6 | 0.8 | 33.3 | 20.0 | 6 | 1.17 | 21.7 |
| 0.7 | 0.9 | 28.6 | 22.5 | 7 | 1.32 | 24.5 |
| 0.8 | 1.0 | 25.0 | 25.0 | 8 | 1.47 | 27.3 |
| 0.9 | 1.1 | 22.2 | 27.5 | 9 | 1.62 | 30.1 |
| 1.0 | 1.2 | 20.0 | 30.0 | 10 | 1.75 | 32.5 |
| 1.2 | 1.4 | 16.7 | 35.0 | 20 | 2.80 | 52.0 |
| 1.4 | 1.6 | 14.3 | 40.0 | 30 | 3.50 | 65.1 |
| 1.6 | 1.8 | 12.5 | 45.0 | 40 | 4.00 | 74.3 |
| 1.8 | 2.0 | 11.1 | 50.0 | 50 | 4.38 | 81.4 |
| 2.0 | 2.2 | 10.0 | 55.0 | 60 | 4.67 | 86.8 |
| 2.3 | 2.5 | 8.7 | 62.5 | 70 | 4.90 | 91.1 |
| 2.6 | 2.8 | 7.7 | 70.0 | 80 | 5.09 | 94.6 |
| 2.9 | 3.1 | 6.9 | 77.5 | 90 | 5.25 | 97.6 |
| 3.0 | 3.2 | 6.7 | 80.0 | 100 | 5.38 | 100.0 |
| 3.4 | 3.6 | 5.9 | 90.0 | 200 | 5.38 | 100.0 |
| 3.8 | 4.0 | 5.3 | 100.0 | | | |
| 4.0 | 4.0 | 0.0 | 100.0 | | | |
| 4.5 | 4.0 | 0.0 | 100.0 | | | |

GF: Garfield Factor;
$GF_s$: Score given by GF;
b: Number of pages;
$b_s$: Score by number of pages.

by the Committee to consist of an impact and a credit part. The impact part consists of a modified GF (see eq. (2)).

The total credit of a paper is unity and it is distributed among the participating teams by the Team Credit Share indices ($f$). This index is the sum of credits attributed to the individual co-authors belonging to the same team. In order to distribute credit of papers among co-authors an algorithm has been developed based on the number and rank of co-authors. Empirical[18] and theoretical[19] investigations give support to the use of the distribution shares applied.

The scoring system in Table 4 rewards co-authors according to their rank number, assuming that the measure of the contribution of co-authors is reflected by their rank.

It is often claimed that the rank of co-authors would reflect the alphabetical order of the names. In order to determine the significance of alphabetical ranking, 5686 chemistry papers were randomly selected from *Current Contents* volumes (1994–1995) and the rank of authors controlled. Data in Table 5 indicate only a slight preference for the alphabetical listing of authors over other rankings. The number of possibilities, namely for alphabetical rankings can be calculated by the permutation formula: $P_n = n!$, where $n$ is the number of authors and $P_n$ is the number of permutations. If $n = 2, 3, 4, 5, 6, 7, 8$, $P_n = 2, 6, 24, 120, 720, 5040, 40320$, respectively. Table 4 shows about 10 per cent excess in favour of the alphabetical listing.

Team credit share indices ($f$) used in eq. (2) refer to team co-operations, which can be understood from the following example. There are four co-authors: a, b, c, d of which a and c are members of the team evaluated ($e_1$), b and d are co-operating partners. The individual credit shares are as follows (see Table 4) – a: 0.35; b: 0.30; c: 0.20; d: 0.15. The sum of the individual shares for the respective team is:

$$f(e_1) = 0.35 + 0.20 = 0.55.$$

**Table 4.** Individual credit shares of co-authors by rank numbers

| Total number of co-authors | Rank number of co-authors 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| 2 | 0.65 | 0.35 | | | | |
| 3 | 0.55 | 0.25 | 0.20 | | | |
| 4 | 0.50 | 0.25 | 0.15 | 0.10 | | |
| 5 | 0.40 | 0.25 | 0.15 | 0.10 | 0.10 | |
| 6 | 0.33 | 0.25 | 0.15 | 0.10 | 0.10 | 0.05 |

## Evaluation of non-journal publications

The evaluation of the scientific impact of books, book-chapters, contributions in conference proceedings, etc. cannot be performed regularly by a couple of peers for a great number of publications or for publications of a thematically wide spectrum. Therefore, the Publication Committee suggested assessing these publications according to their language and length. This is a very crude approximation, namely each non-journal publication would have equal impact. Noting that no abstracts, obituaries, recensions, etc. are accepted. Only contributions which have features similar to scientific papers in journals are taken into account. Conference proceedings are accepted provided that they are published by publishing houses, are edited and have ISSN or ISBN numbers. The Committee has the right to accept or reject any publication after having considered each publication thoroughly. Non-English written publications may get half of the score given by eq. (3).

$$b_s = \frac{7b}{30+b}, \qquad b_{s_{max}} = 5.38, \tag{3}$$

where $b_s$ is the score given for non-journal publications, $b$ is the number of pages.

A paper in a journal of average GF in chemistry (about 1.8) has a score of 2.00, a contribution in a proceeding or

**Table 5.** Percentage shares and probabilities for chemistry papers with or without alphabetical order of the authors as a function of the total number of authors

| Total number of authors | 1 | 2 | | 3 | | 4 | | 5 | | 6 | | 7 | | 8 | | > 8 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Number of papers | 349 | 1596 | | 1594 | | 1050 | | 590 | | 296 | | 115 | | 57 | | 39 | |
| Per cent of papers | 6.14 | 28.07 | | 28.03 | | 18.47 | | 10.37 | | 5.21 | | 2.02 | | 1.00 | | 0.69 | |
| | | abc | n | abc | n | abc | n | abc | n | abc | n | abc | n | abc | n | abc | n |
| Share of papers with alphabetical (abc) or without alphabetical (n) order of authors (per cent) | | 53.57 | 46.43 | 24.91 | 75.01 | 15.71 | 84.29 | 12.20 | 87.80 | 10.14 | 89.96 | 8.70 | 91.30 | 8.770 | 91.230 | 2.26 | 97.44 |
| Theoretical probability (per cent) | | 50.00 | 50.00 | 16.67 | 93.33 | 4.17 | 95.83 | 0.83 | 99.17 | 0.14 | 99.86 | 0.02 | 99.98 | 0.002 | 99.998 | | |

Total number of papers studied: 5686.

a chapter of 12 pages in a book is given the same score. Complete books may cover at least 100 pages, and they are considered to represent greater credit than journal papers. Therefore, the maximum score offered for books is greater than that for journal papers (5.38 > 4.00). It should be stressed again that the assessment method described here has been developed through several compromises, and thus lower and upper limits, e.g. are arbitrary. The reasons for not scoring books higher are as follows. Extremely high scores would distort the whole score system. Books are products of activities performed during longer periods, and the present assessment system refers to annual scoring. Results published in books are generally published in several journal papers before. Books can be seen mostly as products of individual activity. The aim of the evaluation method tackled here is to assess performance of teams and not individuals.

## Total publication score

The Total Publication Score (TPS) of teams can be calculated by summing the scores attained through papers and other publications as follows,

$$\mathrm{TPS} = \sum_{p=1}^{P} f_i(\mathrm{GF}_i + 0.2) + \sum_{b=1}^{B} f_i \frac{7b_i}{30 + b_i}, \tag{4}$$

where $P$ and $B$ are the total number of papers and non-journal publications respectively, $f_i$ is the team credit share of the $i$th paper or publication and $b_i$ is the number of pages of the $i$th non-journal publication.

The TPS index is a growth index, i.e. it is appropriate, for example, for distributing grants among research teams consisting of different number of researchers. TPS consists of an impact and a quantity part. The impact is characterized by the Impact Scores ($\mathrm{GF_s}$) and the number of pages ($b$), whereas quantity by the number of papers ($P$) and non-journal publications ($B$). The dimension of the index is: [scores].

## Evaluation of teams in the Institute by determining relative citedness

The second factor taken into account for evaluating the research performance of teams in the Institute is the *echo* of the results published. Citations are regarded as manifested proofs of scientometric results having indicated impact. The teams assessed are active in various fields, therefore the determination of the relative impact (applying international standards) would be strongly preferred. The measure of the relative impact could be determined through the number of papers published.

Both the citations listed in the *SCI* to journal papers and those found by the researchers themselves in books and proceedings or in journals not figuring on the *SCI* lists both to papers and books and proceedings are taken into consideration. Only independent citations are calculated. 'Independent' means that there is no common author in the citing and cited publications.

The Publication Committee decided to apply an indicator of Relative Citation Rate (RCR) type[20] in the mid-eighties.

Relative Subfield Citedness (RW) index might be *regarded* more appropriate, but we were unable to select pertinent reference standards for the papers of each research team evaluated. The Relative Citedness (RC) recommended by the Committee applies a standard selected by the publishing researchers themselves (see eq. (5)). Therefore, this measure is easily accepted by the teams evaluated. The numerator of the index gives the *total real impact* of the publications of the respective team in terms of citations, whereas the denominator provides a reference standard, i.e. a measure of the mean number of citations obtained by the researchers working worldwide on the same field and having published the same number of papers.

The indicator applied in the publication assessment process is termed as Relative Publication Potential (RPP), and can be obtained as follows.

$$\mathrm{RPP} = \frac{\sum_{p=1}^{P_T} c_i}{\sum_{p=1}^{P} \mathrm{GF}_i} \cdot f_\mathrm{m} \cdot P = \mathrm{RC} \cdot f_\mathrm{m} \cdot P, \tag{5}$$

where $c_i$ is the number of citations obtained by the $i$th publication, $P$ is the number of journal papers in *SCI*, $P_T$ is the total number of publications and $f_\mathrm{m}$ is the mean Team Credit Share (TSC, see eq. (6)).

The impact part of the indicator is represented by RC, whereas the quantitative part is represented by the product ($f_\mathrm{m} \cdot P$).

For practical applications, time windows should be selected. The present assessment method applies a single year as citation time window ($t_\mathrm{c}$), whereas a ten-year period as publication time window ($t_\mathrm{p}$). The researchers evaluated insisted on using a long time period for $t_\mathrm{p}$ because they prefer to follow the international influence of their papers.

The Mean Team Credit Share ($f_\mathrm{m}$) is calculated as follows.

$$f_\mathrm{m} = \frac{\sum_{i=1}^{P} (f_i \cdot \mathrm{GF}_i)}{\sum_{i=1}^{P} \mathrm{GF}_i}, \tag{6}$$

where $P$ is the total number of papers in *SCI*, $\mathrm{GF}_i$ is the GF value of the $i$th paper and $f_i$ is the respective TCS.

According to eq. (6), extensive co-operation resulting in papers in journals of lower GF means less than the same published in high GF journals.

Note that the RPP indicator is a complex growth index, i.e. it is appropriate for distributing grants among research teams with different research potentials. In order to obtain comparable indices, RPP values should be converted into specific indices, e.g. dividing them with the number of researchers working in the individual teams assessed.

## Methodological shortcomings and compromises between evaluators and persons evaluated

For any assessment system to be applied in practice, an agreement between persons who ordered the evaluation and apply its results, those who are assessed and the experts who develop the methods and perform the process is extremely important. Each side should make reasonable compromises. In the present assessment method, there are some inconsistencies concerning scientometric methodology and practical application. Three of them are tackled in some detail here.

(1) In calculating the RPP indicator, not only are the citations found in the *SCI* taken into consideration, although the reference standard contains only these items. The reason is that the researchers evaluated insist on considering citations found in non-*SCI* journals and in books as well. They claim that citations received from monographs or books are even more valuable than those from journal papers. In the opinion of the present author, citations in books indicate that the respective information is closer towards being institutionalized knowledge. Consequently, it would be a fault to disregard such citations in the evaluation. There are no 'impact' scores, however, for books which could be applied as reference standard in eq. (5). The consideration of citations found in books and monographs can be regarded, therefore, as an 'extra award' or 'acknowledgement' given by the evaluation method applied. Furthermore, books review earlier results, therefore readers would not reference the original journal papers but the pertinent books.

(2) GFs refer to $t_c = 1$ year (one year for citations) and $t_p = 2$ years (two years for publications). In eq. (5), the RC indicator applies $t_c = 1$ year and $t_p = 10$ years. Theoretically, time windows for the citations evaluated and those used as reference standards should be the same in length. The mean ratio of impact factors with different length of publication times $(t_p)$[IF(2)/IF(10)], however, was found to be 1.15 for 55 chemistry papers[21], where IF(2) is the GF and IF(10) is the impact factor calculated with $t_c = 1$ and $t_p = 10$ years. The difference of 15 per cent cannot cause significant errors in comparing the research teams by the PAM tackled here. (However, it should be stressed that the IF(2)/IF(10) ratio depends on subfields.) The application of IF(2), i.e. GFs, has practical reasons as well. The researchers evaluated are familiar with the GFs published by Institute for Scientific Information, but would not accept 'ten years' impact factors 'artificially' constructed by the local Publication Committee. GFs are available annually and mean factors referring to the period, e.g. 1988 to 1997 are easy to obtain.

The calculation of RC, eq. (5) refers, of course, to citations obtained in, say, 1996 to papers published between 1986 and 1995 and in the next year to those obtained in 1997 to papers published between 1987 and 1996, etc.

(3) No self-references are taken into account although the reference standard, the denominator of indicator RC (eq. (5)), contains self-references as well. The reason is that the ratio of self-references ranges from as low as 5 per cent up to about 25 per cent, in average, for the teams evaluated. Taking into account self-citations would lead to significant errors in the assessment process.

## Consequences of the evaluation of publications

Evaluations without consequences are useless. Results and experiences of the evaluation can be used by the persons and teams evaluated and by the officials, committees, boards, etc. which ordered the process.

Results of the publication assessment process are summarized in Tables 6 and 7, showing publication and citation data for some representative teams in the Institute.

The management system of the Institute aims at maintaining successful scientific research. Science requires, however, relatively quiet conditions for longer periods. Therefore, the grant system of the Institute is based on

**Table 6.** Publication data for some representative teams of the Institute

| Team | Mean number of researchers ($K$) | Number of journal papers (1999): In Hungarian | Number of journal papers (1999): In foreign languages | Number of publications in books or conference proceedings (1999) | Total number of publications (1999) | Publication scores: 1997 | Publication scores: 1998 | Publication scores: 1999 | Publication scores: Total (TPS) | TPS/$K$ |
|---|---|---|---|---|---|---|---|---|---|---|
| A | 9.52 | – | 8 | – | 8 | 13.111 | 3.369 | 4.665 | 21.145 | 2.221 |
| B | 12.15 | 1 | 13 | 6 | 20 | 18.710 | 22.991 | 22.663 | 64.364 | 5.297 |
| C | 11.25 | 1 | 11 | – | 12 | 9.947 | 10.263 | 10.205 | 30.415 | 2.704 |
| D | 4.50 | – | 11 | 3 | 14 | 10.057 | 10.482 | 13.501 | 34.040 | 7.564 |
| E | 11.25 | – | 24 | 5 | 29 | 36.709 | 18.074 | 18.412 | 73.195 | 6.506 |
| F | 5.60 | 1 | 8 | – | 9 | 4.538 | 8.708 | 7.377 | 20.623 | 3.680 |

several medium and long-term performance factors, which are as follows (the percentage values of the budget depending on the respective factor are given in brackets):

- Total Publication Scores, TPS (27.4%);
- Total Relative Publication Potential, TRPP (11.0%);
- Number of researchers with scientific degree or title (Ph D or Doctor of the Academy, respectively) (28.8%):
- Number of researchers (32.8%).

The percentage ratios given above are results of an agreement between the three parties interested in the evaluation process.

Among the factors mentioned above, the first and the second are competitive of medium term, the third is competitive of long term, whereas the fourth is not competitive and serves as a buffer for eliminating high fluctuations. In addition to the funds mentioned, for each Ph D student and post-doc, the respective team is given a special annual provision.

Table 6 contains the number and score of publications that appeared in 1999 and in 1997 to 1999, respectively for some representative research teams. The sum of the annual Publication Scores (Total Publication Score, TPS) for the teams yields the respective factors for distributing grants (27.4%). In order to make the indices comparable, a specific indicator can be derived by dividing TPS values with the mean number of researchers. The data in Table 6 show a dynamic range of 2.221 to 7.564, which corresponds to a ratio of 1: 3.41.

Similarly, the sum of the RPP values for three years (TRPP) is used for the distribution of 11.0 per cent of the total grant. The dynamic range of the (TRPP/*K*) indices is 0.765 to 3.720, which corresponds to a ratio of 1 : 4.86.

## Further possibilities

There are several possibilities, outside management aspects, offered by the scientometric methods used for obtaining valuable information. Relative Publication Strategy[21], e.g. can be calculated from the data of Tables 2 and 7. This indicator can tell us the relative appropriateness of the information channels used by the research teams compared to the mean of the respective subfield. It relates, namely the sum of GF data of papers (i.e. respective journals) of the team studied to the mean GF of journals dedicated to a field or subfield multiplied by the number of papers published by the respective team. This knowledge would be important for the local science policy makers and can serve as a mirror for the respective researchers themselves.

The Relative Subfield Citedness (RW) indicator[22] applies a standard (mean citedness of papers, i.e. journals) assigned to the subfield which corresponds to the activity of the team assessed. This measure is independent of the publishing author, but strongly depends on the method of calculating the reference standard. The indicator of the respective team with a dimension of [number of citations per paper] can be related to the standard mentioned.

The above indicators can contribute to a deeper insight into the international role of the research teams assessed and can help the scientists to plan an appropriate publication strategy.

One further important indicator could be recommended for assessing research performance of teams, namely the number of highly cited papers related to the total or to the number of researchers. The Committee follows the citedness of individual papers published by the researchers of the Institute for a longer period in order to explore results and topics which attract great international interest. As a criterion for being highly cited in a five-year period after publishing a paper, a citedness of about ten times the GF-value of the respective journal was suggested by Andras Schubert (pers. commun.). The mean number of citations which could be obtained in two, three, four, etc. years after the publication year can be predicted by a simple time-citation function[21].

**Table 7.** Relative Citedness (RC) and Relative Publication Potential (RPP) for some representative teams in the Institute

| Team | Mean number of researchers (*K*) | Citations in 1998: In *SCI*-journals ($C_p$) | In books, etc. | Total (*C*) | Sum of GFs (*H*) 1988–1997 | Number of papers in *SCI*-journals (*P*) | $C_p/P$ | *H*/*P* | RC | $f_m$ | RPP 1996 | RPP 1997 | RPP 1998 | RPP | RPP/K |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| A | 9.52 | 75 | 2 | 77 | 143.238 | 101 | 0.743 | 1.418 | 0.538 | 0.454 | 4.345 | 2.034 | 2.467 | 8.846 | 0.929 |
| B | 12.15 | 113 | 10 | 123 | 108.575 | 63 | 1.794 | 1.723 | 1.133 | 0.778 | 10.880 | 8.117 | 5.553 | 24.550 | 2.021 |
| C | 11.25 | 130 | 4 | 134 | 165.702 | 94 | 1.383 | 1.763 | 0.809 | 0.570 | 6.337 | 5.783 | 4.335 | 16.455 | 1.463 |
| D | 4.50 | 66 | 3 | 69 | 42.419 | 56 | 1.179 | 0.757 | 1.627 | 0.661 | 6.549 | 4.170 | 6.023 | 16.742 | 3.720 |
| E | 11.25 | 190 | 0 | 190 | 265.413 | 146 | 1.301 | 1.818 | 0.716 | 0.659 | 13.459 | 6.077 | 6.889 | 28.995 | 2.577 |
| F | 6.50 | 37 | 0 | 37 | 94.077 | 44 | 0.841 | 2.138 | 0.393 | 0.735 | 1.773 | 0.664 | 1.271 | 4.975 | 0.765 |

*K*: Mean number of researchers in 1999;
H: sum of GFs of papers (i.e. journals) published between 1988 and 1997;
RC = *C*/*H*;
RPP = RC · *P* · $f_m$;
$f_m$ = Mean Team Contribution Share (1988–1997);
*P*: Number of papers in *SCI*-journals, 1988–1997.

## Concluding remarks

Surveying the literature of scientometrics one can conclude that there are no ready recipes for scientometric evaluations; each system has its own characteristics. Scientometric researchers are exploring features of information processes in scientific research but are far from finding the solution to all problems.

Scientific research is a social organization, therefore the possibility for a description by quantitative methods has more limitations than in the field of natural sciences.

In my opinion, we may accept, however, that research teams providing indicators (like RCR, RW, RPS) exceeding or at least approaching the international standards, might have performed well. Other teams with indicators significantly inferior to the international averages may have performed poorly. The reasons for the latter should, however, be investigated thoroughly, to find out whether it is caused by a lack of performance or by some shortcomings in the methods applied.

Scientometric methods should be used by scientometricians because the application requires appropriate knowledge. Performing an evaluation process, however, means personal responsibility as well. Refusing the use of scientometric evaluations whenever appropriate is equally wrong as their incorrect or irresponsible application.

1. Moravcsik, M. J., in *Handbook of Quantitative Studies of Science and Technology* (ed. van Raan, A. F. J.), Elsevier Science Publishers B.V., 1988, pp. 11–30.
2. Martin, B. R. and Irvine, J., *Res. Policy*, 1983, **12,** 61–90.
3. Moed, H. F., de Bruin, R. E. and van Leeuwen, Th. N., *Scientometrics*, 1995, **33**, 381–422.
4. Martin, B. R., *Scientometrics*, 1996, **36,** 343–362.
5. Vinkler, P., *Scientometrics*, 1986, **10**, 157–177.
6. Vinkler, P., in *Evaluating Science and Scientists* (eds Frankel, M. S. and Cave, J.), Central European University Press, Budapest, 1997, pp. 149–159.
7. Vinkler, P., *Scientometrics*, 1998, **41**, 185–200.
8. Vinkler, P., *Scientometrics*, 1988, **13**, 239–259.
9. Hicks, D., *Soc. Stud. Sci.*, 1987, **17,** 295–316.
10. Garfield, E., *Citation Indexing: Its Theory and Application in Science*, John Wiley, New York, 1979.
11. Marshakova-Shaikevich, I., in Proceedings of the Fifth Biennial Conference of the International Society for Scientometrics and Informetrics (eds Koenig, M. E. D. and Bookstein, A.), Learned Information Inc, Medford, NJ, 1995, pp. 323–330.
12. Sen, B. K., *J. Doc.*, 1999, **55,** 325–334.
13. Hirst, G., *J. Am. Soc. Inf. Sci.*, 1978, **29**, 171–174.
14. Vinkler, P., *Scientometrics*, 1988, **14**, 453–474.
15. Moed, H. F., van Leeuwen, Th. N. and Reedijk, J., *Scientometrics*, 1999, **46**, 575–589.
16. Glänzel, W. and Schubert, A., *J. Inf. Sci.*, 1988, **14,** 123–127.
17. Vinkler, P., in *The Web of Knowledge: A Festschrift in Honor of Eugene Garfield* (eds Cronin, B. and Atkins, H. B.), September, 2000.
18. Vinkler, P., *Scientometrics*, 1993, **26**, 213–230.
19. Lukovits, I. and Vinkler, P., *Soc. Indic. Res.*, 1995, **36,** 91–98.
20. Schubert, A. and Braun T., *Scientometrics*, 1986, **9**, 281–291.
21. Vinkler, P., *Scientometrics*, 1999, **46**, 621–633.
22. Vinkler, P., *Scientometrics*, 1997, **40,** 163–169.