ISSUE 1 SEPTEMBER 2007

researchtrends

IN THIS ISSUE:

Welcome to the first issue of Research Trends, a bi-monthly newsletter providing objective, up-to-the-minute insights into scientific trends based on bibliometric analysis.

Worldwide, there has been increasing demand for quality research performance measurement and trend-related information by deans, faculty heads, researchers, funding bodies and ranking agencies.

Research Trends has been launched to share interesting insights into developments, approaches and tools in this area.

Research Trends is powered by Scopus, the largest abstract and citation database of peer-reviewed literature and quality Web sources, with smart tools to track, analyze and visualize research.

We hope you enjoy the first issue, and welcome your feedback.

Look out for the next issue in November.

Kind regards,

The Research Trends Editorial Board

Did you know?

How big can a research paper get?

In the era of globalization, science is leading the curve. A microcosm of this trend can be seen in the massive research article entitled 'Precision electroweak measurements on the Z resonance' published in *Physics Reports* in 2006 *(1)*. Spanning 198 pages (in a dedicated double issue of the journal) and listing 405 references, this singular work of scholarship is attributed to 2,535 authors affiliated with 225 institutes in 33 countries, working within 7 collaborative research groups.

(1) Grünewald et al (2006) "Precision electroweak measurements on the Z resonance", *Physics Reports*, Vol. 427, No. 5-6, pp. 257-454.

The value of bibliometric measures

Scientometrics from past to present

Scientometric research, the quantitative mathematical study of science and technology, encompassing both bibliometric and economic analysis *(1)*, is expanding at an enormous pace. This is evidenced in increasing attendance rates at important industry conferences, and the recent launch of the dedicated *Journal of Informetrics*. Indeed, if one were to pick up an issue of any of the leading journals in the field today, one would find research covering article output, citation relationships between disciplines and geographical analysis linked to these. In a two-part article, we explore not only scientometrics' past but also its impact on and relevance in the present.

The origins of bibliometric research can be traced back to the beginning of the 19th century within areas such as law. Shapiro (1999) *(2)* indicates that many aspects of bibliometrics were "practiced in the legal field long before being introduced into scientific literature". Early research in the 1880s was reported by Delmas (1992), who describes documentation in France, but initial studies on qualitative and quantitative analysis of science seem to originate within psychological fields (Godin 2006) *(3)*. Godin cites the work of Buchner in describing the notion of "scientific" psychology as "factual, inductive, measurable and experimental" and in 1920 Boring presented research on subject and geographical analysis of psychologists.

Laying down the Law

Probably the earliest, most definable research within the scientometric field was the work that gave rise to the laws of bibliometrics. The first, which came to be known as Lotka's Law, after Alfred Lotka, can be traced back to 1926 and suggested that within a defined area over a specific period a low number of authors accounted for a large percentage of publications in the area. This was followed in 1935 by the work of George Kingsley Zipf, which describes the frequency of words in a text and became known as Zipf's Law. Zipf's research was refined into two main laws looking at high and low frequency words within a text. In 1948 Samuel Clement Bradford's analysis indicated that within a given area over a specific time a few journals publish a high percent of articles within the area and there are many journals that publish only a few articles each: Bradford's Law. These laws continue to be studied and form the basis of the development of the modern-day scientometric literature.

In 1944, Lehman described the relationship between quantity and quality within scientific writing and this was followed in 1952 by Dennis, who analyzed the effect of scientists' age on these two elements. Again these types of analyses continue to be described in the current literature, and began to direct thinking towards averaged metrics within bibliometrics.

Timeline

Early 19th century
Origins of bibliometric research in areas such as law and psychology.

1926-48
Lotka's Law, Zipf's Law and Bradford's Law developed.

1955
Eugene Garfield first describes the Impact Factor.

1961
Publication of the Science Citation Index.

1978
Launch of first dedicated journal, *Scientometrics*.

1960s-70s
Growth of databases make widespread citation analysis a real possibility.

Journal metrics

One of the most recognized accomplishments in the field of scientometrics is the development of the Impact Factor and the work of Eugene Garfield *(4)*. Garfield first described the Impact Factor in 1955 as a method of selecting journals for inclusion in a genetics citation index he had been developing. This eventually resulted in the publication of the Science Citation Index in 1961 as a means of linking articles together via their references. Since it was first described, journal Impact Factor has developed into a widely used bibliometric indicator.

Around the same time, another key figure, Derek De Solla Price, was working on the study of the exponential growth of science and the citation activity of scientific literature. He published several papers describing the key elements of scientometric analysis, including work on patterns of communication between scientists and the overall history and study of science itself.

There was tremendous growth in the scientometric literature in the 1960s (Herubel 1999) *(5)* and from this point forward the field of scientometrics developed and differentiated into several specializations. These were brought together by the launch of the first journal devoted to the field, *Scientometrics*, founded and edited by Tibor Braun of the Hungarian Academy of Sciences. One of the most notable developments was citation analysis. Once a laborious manual job few scholars would engage in, the emergence of (print) databases allowed citation patterns to be studied with relative speed and ease.

In the next issue, we explore scientometrics' transition into the 21st century: the proliferation of databases, new author-focused indices and the impact of the Web.

References:

(1) Diodato, V. (1994) *Dictionary of Bibliometrics* (1st ed.) New York: Haworth Press.

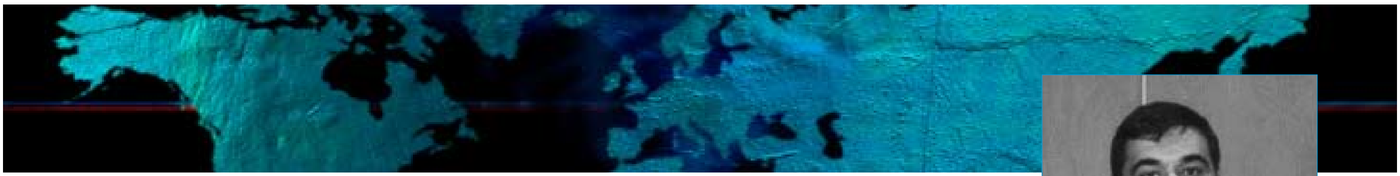
(2) Shapiro, F.R. (1999) "Origins of bibliometrics, citation indexing and citation analysis: The neglected legal literature", *Journal of the American Society for Information Science*, Vol. 43, No. 5, pp. 337-339.

(3) Godin, B. (2006) "On the origin of bibliometrics", *Scientometrics*, Vol. 68, No. 1, pp.109-133.

(4) Garfield, E. (2006) "The history and meaning of the journal impact factor", *Journal of the American Medical Association*, Vol. 295, No. 1, pp. 90-93.

(5) Herubel, V.M.J-P. (1999) "Historical Bibliometrics: Its purpose and significance to the history of disciplines", *Libraries & Culture*, Vol. 34, No. 4, pp.380-388.

Country trends



Focus on Turkey: the influence of policy on research output

With a special contribution from Professor Cem Saraç

When describing research, the American astronomer Dr. Carl Sagan was quoted as saying, "Somewhere, something incredible is waiting to be known." This inspiring quotation reflects the fact that research exists in all parts of the world (and indeed outside of the world, as in the case of Astronomy) and that researchers collaborate to produce incredible breakthroughs in every country. This is the first in a series of articles that reflect the global nature of research. The series covers research trends across countries, and investigates the proliferation of research communication throughout the world.

We are focussing our first analysis on Turkey, a country that has shown strong growth in article output in recent years (see Figure 1 below).

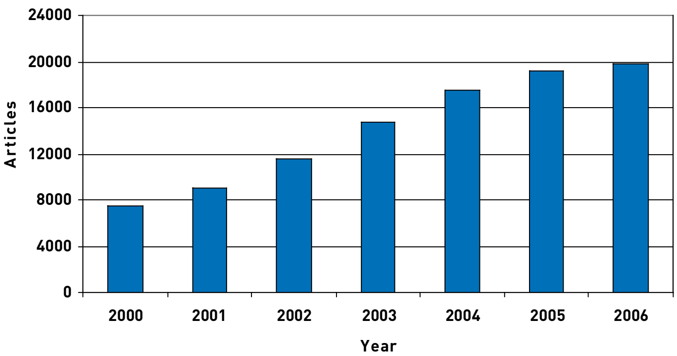


Figure 1 – Output of journal research articles in Turkey 2000–2006
Source: Scopus

The increase in research articles across this period is occurring at an average rate of 17% per annum over the period 2000-2006, as compared with a 3.5% p.a. overall growth in the same period. But how can we explain this increase? Certainly the OECD Main Science and Technology Indicators Vol. 2007 *(1)* identify trends in data, which match this increase in research articles. Figure 2 illustrates the growth in the number of researchers based in Turkey. Comparing the data in the two graphs, we can conclude that the more researchers in a country, the more articles are written and published from institutions within that country.

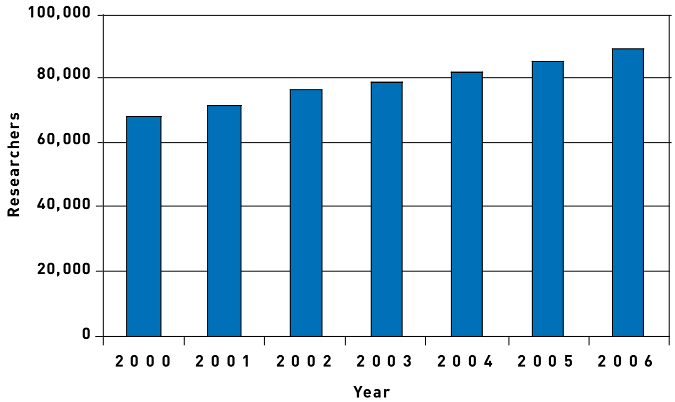


Figure 2 – Researchers active in R&D in Turkey 2000-2004
Source: OSYM 2007 *(2)*

This increase in research articles and number of researchers is also matched by the increase in funding of higher education (HE) within Turkey; Figure 3 illustrates the growth in HE funding across the same period.

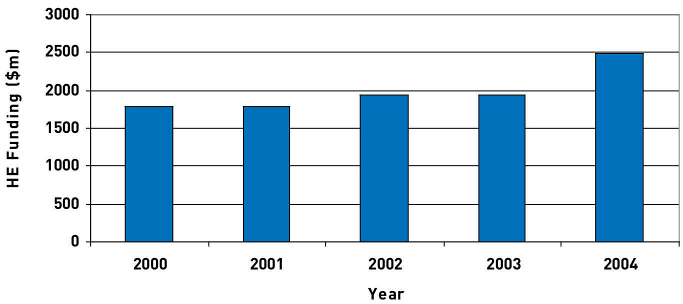


Figure 3 – Higher education funding in Turkey 2000-2004
Source: OECD

While these indicators continue to increase, the difference between subject fields is also evident. Figure 4 illustrates the subject breakdown of Turkish research in 2006 in Scopus and demonstrates that medical and life science research is currently leading the way in terms of published output, but that significant contributions are also being made to the physical and mathematical sciences.

Continued on page 4

Continued from page 3

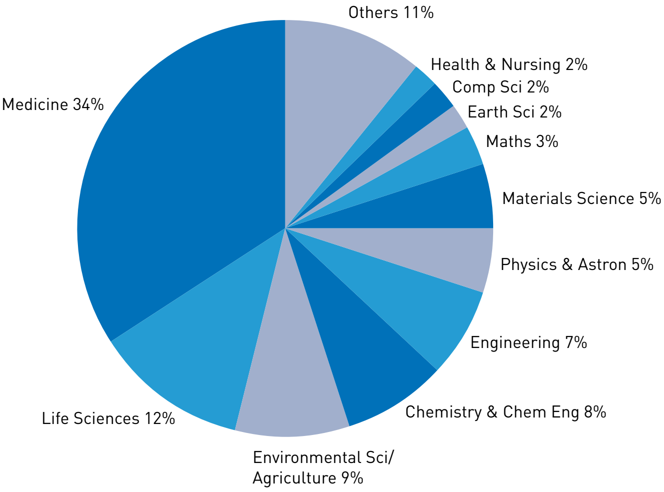


Figure 4 – Subject focus of Turkish research articles 2006
Source: Scopus

A clear relationship exists between research funding, researcher population and article outputs at a national level, and Turkey is no exception. Data like this can inform and guide policymakers at all levels to leverage the infrastructure of the national science system and cultivate a knowledge economy.

As an 'insider' so to speak, Cem Saraç, Professor of Engineering at Hacettepe University, Ankara, says there are two principle reasons that could explain the figures in the tables above *(3)*. Both relate to policy changes. "The first one can be linked to the Turkish Ministry of Health's strategy," he says. Indeed, OECD figures *(4)* show that health spending per capita in Turkey grew, in real terms, by an average of 5.8% per year between 2000 and 2005. This was one of the fastest growth rates in OECD countries and significantly higher than the OECD average of 4.3% per year. In addition, as part of a nationwide performance-based contribution payment system *(5)*, implemented in training and research hospitals in 2004, clinic and deputy chiefs, chief interns and specialists receive additional scores providing they publish a definite number of papers.

"The second reason for the significant growth is the prerequisites, generally initiated after 2000, for applying for university degrees at Lecturer, Assistant Professor, Associate Professor and Full Professor levels," he continues. "My university stipulates that one has to write at least three international papers in order to apply for an Associate Professor Degree and another four international papers for a Full Professor Degree. While each university has its own requirements, prerequisites like these could also have affected article growth."

References:

(1) OECD *Main Science and Technology Indicators*, Vol. 2007

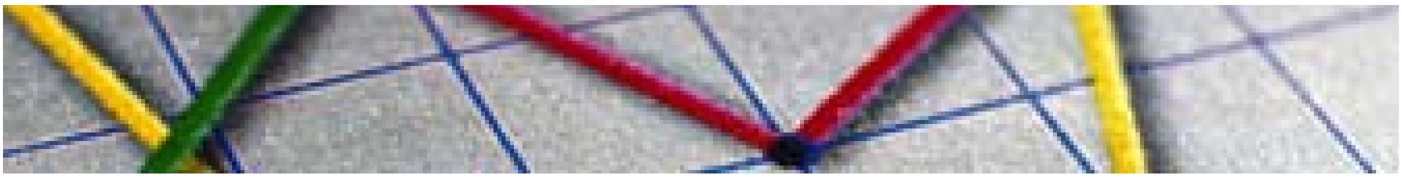
(2) OSYM (2007), Student Selection and Placement Center, Research and Publishing, from the **World Wide Web**

(3) Demirel, I.H., Sarac, C. and Ozgen T. (2007) "Science in Turkey, 1973-2006". *Science Magazine, AAAS.*

(4) OECD (2007) "OECD Health Data 2007, How Does Turkey Compare", Retrieved September 21, 2007 from the **World Wide Web**

(5) The Ministry of Health of Turkey "Performance-based payment system in the Ministry of Health Practices", Retrieved September 21, 2007 from the **World Wide Web**

Research trends



Country analysis: examining the numbers

Research evaluation at country or national level is moving increasingly towards a metric-based system. The most obvious examples of these countries are Australia, with the Research Quality Framework, and the United Kingdom, with the Research Assessment Exercise, where policymakers and administrators are being called upon to submit metrics for national evaluation.

It is interesting to extract two of the indicators researchers, policymakers and administrators focus on, namely article count and citations received at country level. The differences in the number of articles published in each country may not be unexpected but the top 1% and 5% citation thresholds certainly warrant further attention.

Methodology

An analysis was performed in Scopus to extrapolate the top 1% and 5% of cited papers for ten randomly selected countries within 27 subject categories. The results of this analysis can be found in Table 1 (downloadable at **www.researchtrends.com**).

The table denotes the number of papers published in each country for a period of five individual years from 2002-6. These counts are then separated into 27 subject categories (as specified in Scopus.com). For each of these years and for each subject category, the number of papers that forms a part of the top 1% of highly cited papers was derived.

For the purpose of this analysis, it is important to note that the

Continued on page 5

Continued from page 4

cut-off date for the data extrapolation was set at September 18, 2007 resulting in a favoring of 'older' papers in most instances. To illustrate, we would like to take the example of Australia in the year 2002 where the following result was obtained within the subject category Engineering:

- There were 2595 papers published in 2002;
- The top 1% is thus a total of 26 papers (rounded up);
- The citation threshold equals 36 citations (up to September 18, 2007).

For an Australian researcher this means that if s/he has published a paper in Engineering and has obtained a citation count of higher than 36 (considering the citation cut-off date), that researcher belongs to the top 1% of Australian research output in that year. The table also shows these figures for the top 5% citation threshold for all ten countries spread over the 27 subject categories.

This kind of data can also be used to analyze different results observed between countries. As a result, Research Trends will be providing this data for more countries in the future.

There are many interesting directions in which this research can develop and we welcome your **feedback**. This will help us to deliver the content you are most interested in.

A snapshot of the ten subject categories in Australia with the highest number of publications in 2006

Subject Area	Publications	Article #	1% Threshold	Article #	5% Threshold
Agricultural and Biological Sciences	5179	68	9	310	4
Biochemistry, Genetics and Molecular Biology	5411	56	18	297	8
Chemistry	1854	22	14	109	6
Earth and Planetary Sciences	2898	31	12	146	6
Engineering	3814	48	4	349	1
Environmental Science	2702	30	9	167	4
Immunology and Microbiology	1757	19	20	101	9
Medicine	12348	124	18	625	7
Physics and Astronomy	3109	33	10	237	3
Social Sciences	2702	35	4	163	2

A snapshot of publications in Engineering across ten randomly selected countries in 2006

Country	Publications	Article #	1% Threshold	Article #	5% Threshold
Australia	3814	48	4	349	1
Germany	10282	139	5	549	2
Spain	4539	56	5	247	2
France	9132	99	5	457	2
Greece	1791	23	4	149	1
Italy	7080	96	4	698	1
Japan	19254	262	3	1023	1
New Zealand	543	9	5	52	1
Taiwan	6994	96	4	555	1
South Africa	464	5	3	29	1

Expert opinion

From h to g: the evolution of citation indices

Professor Leo Egghe

The h-index has become a familiar term among bibliometricians since its inception in 2005, and is being increasingly adopted by non-bibliometricians. The letter h is often thought to stand for the h in Hirsch, the name of the physicist who developed it, although it is actually short for 'highly cited'. The h-index is therefore the number of papers that receive h or more citations. For example: Professor X has an h-index of 39 if 39 of his 185 papers have at least 39 citations each and the other 146 (185-39) papers have not more than 39 citations each.

Previous indices have tended only to focus on the impact of individual journals, using the average number of times published papers are cited up to two years after publication. This means that one paper in the journal might have been highly cited and another hardly at all but the authors of both are judged equally on the Impact Factor of their journal. While the h-index can measure individual authors, thereby overcoming the shortcomings of journal Impact Factor, it has limitations of its own. "It is insensitive to the tail of infrequently cited papers, which is a good property," says Professor Leo Egghe, Chief Librarian at Hasselt University, Belgium and Editor-in-Chief of the *Journal of Informetrics*, "but it's not sufficiently sensitive to the level of highly cited papers. Once an article belongs to the h top class, the index does not take into account whether that article continues to be cited and, if so, whether it receives 10, 100 or 1000 more citations."

What's in a name?

The g-index is so called for two reasons: Egghe rejected the name 'e-index' on the grounds that it has a different connotation in mathematics. He therefore looked at the two g's in his surname instead. G also falls immediately before h in the alphabet, reinforcing its link to the h-index.

Lotka's Law

This is where the g-index has evolved from its predecessor. It has all the advantages and simplicity of the h-index, but also takes into account the performance of the top articles. It was in direct response to his criticisms of the h-index that Egghe developed the g-index. No newcomer to bibliometrics, Egghe's main area of expertise is Lotka's Law. The premise of this Law is that as the number of articles published increases, the authors producing that many publications decreases. This principle forms the basis of the h- and the g-indices, the formulae for both of which Egghe was the first to prove. The difference between them is that while the top h papers can have many more citations than the h-index would suggest, the g-index is the highest number g of papers that together received g^2 or more citations. This means that the g-index score will be higher than that of the h-index. It also makes the differences between two authors' respective impacts more apparent. "The only disadvantage I've found so far with the g-index is that you need a longer table of numbers to reach your conclusions!" says Egghe.

Access to funds

For many scientists, there is a direct correlation between where they are ranked in their field and the amount of funding they can attract. "Everything is measured these days, which explains the growth of bibliometrics as a whole," says Egghe. "The g-index enables easy analysis of the highest cited papers; but the reality is that as time passes, it's not going to be possible to measure an author's performance using just one tool. A range of indices is needed that together will produce a highly accurate evaluation of an author's impact."