Chapter 3

INDICATORS FOR NATIONAL SCIENCE AND TECHNOLOGY POLICY

Their Development, Use, and Possible Misuse

Hariolf Grupp[1] and Mary Ellen Mogee[2]

[1]*Karlsruhe University and Fraunhofer ISI, Karlsruhe, Germany*
E-mail: h.grupp@isi.fraunhofer.de
[2]*Mogee Research & Analysis, Reston, Virginia, USA*

Abstract: The purpose of this chapter is to present a survey of the development of Science and Technology (S&T) indicators and their use in national policy making as well as to provide evidence of the vulnerability of S&T indicators to manipulation. A brief history of the development of S&T indicators begins with the United States followed by their worldwide diffusion, with particular emphasis on Europe. The current status of S&T indicators and newer developments towards composite indicators, benchmarking, and scoreboarding is discussed. To investigate the robustness of innovation scoreboards empirically a sensitivity analysis of one selected case is presented. It is shown that composite scores and rank positions can vary considerably, depending on the selection process. It seems not to be too difficult to argue for a 'country friendly' selection and corresponding weighting of indicators. Thus the use of scoreboards opens space for manipulation in the policymaking system. Further research is needed on alternative methods of calculation to prevent their misuse and abuse.

1. INTRODUCTION

The purpose of this chapter is to present a survey of the development of science and technology (S&T) indicators and their use in national policy making, to provide evidence of the vulnerability of S&T indicators to

H.F. Moed et al. (eds.), Handbook of Quantitative Science and Technology Research, 75-94.

manipulation and to suggest questions for future research.[1] The chapter is organised as follows. In the second section we review briefly the history and development of S&T indicators and the major S&T indicators reports in the United States, Japan, and Europe. The third section presents a sensitivity analysis of a composite indicator. Section 4 presents our conclusions and suggestions for future research.

2. A BRIEF HISTORY OF THE DEVELOPMENT OF S&T INDICATORS

2.1 The Beginning — the United States

Science and technology indicators had their start in the United States. The first *Science Indicators* report was published in 1973. The National Science Board (NSB), the policymaking board of the National Science Foundation (NSF), was mandated by the US Congress to publish the report biennially. (Perhaps not surprisingly, given the NSF's pre-eminent support for basic scientific research, it was not until 1987 that the focus and title of the report were broadened to *Science and Engineering Indicators*). In practice the report has been prepared by the Science and Engineering Indicators Unit (SIU) in the NSF's Science Resources Studies division and reviewed by the NSB, which also prepares a brief discussion piece as part of the report.

The idea of 'science indicators' (SI) was an outgrowth of the move toward 'social indicators', i.e., indicators similar to economic indicators such as Gross National Product (GNP) that would provide measures of the health of society. Based on the model of economic indicators, some observers expected science indicators to be a narrow range of statistics that tells about a larger universe, in the same way that new housing starts tell how the economy in general is doing. To date the science indicators report has not evolved in this direction, but rather has become a compendium of many different statistics to measure the health of US science and technology and to compare the US with other nations.

In the early years there was considerable criticism of the science indicators report including the publication– and patent–based indicators. The NSB member and famous mathematician Saunders MacLane criticised the

[1] An earlier, shorter version of this chapter was presented at the Conference in honour of Keith Pavitt, University of Sussex, November 13th–15th, 2003, and subsequently published in a partly overlapping way in a Special Issue of the journal 'Research Policy'.

publication–based indicators, in particular. Early SI reports counted publications by country instead of by individual's addresses or institution, which resulted in the finding that one third were British. In response to the critical writing the NSF held a world conference on the coverage and validity of the set of journals and the way of counting publications.

Similarly there was considerable scepticism in the beginning about patents as science or technology indicators. The NSF had to do lots of studies to show that the use of patents had methodological backing. However, it was patent indicators which showed the US that it should pay more attention to Japan as a competitive industrial power. In the part devoted to international comparisons increases were found in Japanese R&D funding and in all categories of patenting. This finding — that Japan was a power to contend with — was surprising to many at the time.

In addition to criticism of data and methodologies used in the SI report, some criticism of science indicators was based on resistance in the science community to making government funding decisions based on quantitative indicators. Many scientists believed, and continue to believe, that such decisions should be made on the basis of peer review.

In 1985 the House of Representatives Committee on Science and Technology undertook a Science Policy Study which asked the Office of Technology Assessment to examine "… the extent to which decision making would be improved through the use of quantitative mechanisms associated with the concept of investment". OTA concluded that "… while there are some quantitative techniques that may be of use to Congress in evaluating specific areas of research, basic science is not amenable to the type of economic analysis that might be used for applied research or product development". In his accompanying letter, John H. Gibbons, the OTA Director, stated further:

> "Much of the vitality of the American research system lies in its complex and pluralistic nature. Scientists, citizens, administrators, and Members of Congress all play various roles leading to final decisions on funding. While there may be ways to improve the overall process, reliance on economic quantitative methods is not promising. Expert analysis, openness, experience, and considered judgement are better tools."

Despite resistance, over time wide acceptance of the indicators grew. This has been owed in part to pressure from users and decision makers who wanted to be able to show that investment in S&T had value and impact. In response, the SI unit moved beyond the original indicators to additional indicators of interest to policy makers. There was constant improvement by taking feedback and criticism into account.

It was recognised early on that some indicators represent science and engineering resources or 'inputs' to the process of science and engineering, whilst others represented the process itself ('throughput' or 'flow'), the results or 'outputs' of science and engineering, and the effects or 'impacts' of science engineering. The early indicators tended to be heavy on inputs and throughput and weaker on outputs and impacts, spurring the SIU to look for and develop more indicators of outputs and impacts.

In addition to indicators of the level of scientific and technological activity, indicators were developed which examined international collaboration and intersectoral — e.g., university–industry — collaboration. Funding, citations, and co-authorship were all used as indicators of collaboration.

The SIU also developed indicators of the importance of science and engineering. For instance they looked at publication and patent. They also looked at patents citing literature (see, e.g. Narin and Noma, 1985). Publication citations have been used to indicate the quality of research. Patent citations have been used variously to indicate 'technological significance', 'social economic value' and 'private economic value' (for the many applications of patent analysis, see Breitzmann and Mogee, 2002). Other measures of value or quality from patent indicators include the rate at which patent maintenance fees are paid and the number of patent countries in which patent protection is sought for an invention (patent family size).

Later under the Government Performance and Results Act (GPRA) and the pressure to show use of government funded basic research, one of the key indicators that was helpful to policy makers was the patent citations to basic research. This was used to show that commercial technology was building on the basic research funded by government. Other agencies doing this today are the Office of Naval Research (ONR), the Department of Defense (DOD), the Department of Energy (DoE) and the Department of Veterans Affairs (DVA). Industry also uses patent citations and their references to basic research. It is difficult for funding agencies to show a direct use for the basic research they support. The Office of Management and Budget (OMB) and the Congress both review GPRA reports, and the OMB, in particular, likes numbers, so research funding agencies have used these numbers on patents citing basic research to argue for higher budgets.

The SI unit conducted a series of 4–5 surveys of the usage of the SI report. These were generally biennial, like the Science & Engineering report, and were sent out with the report. At first the questions were not split up by what specific indicators the people were using. Instead they asked about the policy issues the respondents wanted to cover and what they found useful. The users always asked for output indicators. Publications and patents are both output indicators. Therefore they had high priority for SI and they still do. Responses were received from the NSB, scientific community, the Office of Science and Technology Policy (OSTP), OMB, and the OECD, amongst others. They all wanted the indicators to be relevant and useful.

The National Science Board (NSB) and the Director of the National Science Foundation (NSF) are important users of the publication data in the SI report. This is natural because publication indicators pertain largely to academic research, and the NSB and the NSF Director are major players in US science policy. This kind of information is reportedly used in allocating funds between scientific fields. Similarly, when they first found evidence of the high citation rate from patents to basic research, the NSB used it to make the industrial use of basic research the centerpiece of their discussion piece in the SI report.

That the NSB has continued the publication of the SI report, although it make suggestions for changes, suggests that the members find the compilation to be useful. In response to a recent (2003) proposal to cut back the SI report and to include just those statistics that are not published elsewhere, the NSB member Anita Jones stated in a presentation that:

– *Science & Engineering Indicators* (S&EI) is a leading data source for R&D policymakers;
– The data are sound; the definitions and categories change slowly;
– The longitudinal data emphasis is very useful;
– The data are collected in one place and repeatedly updated.

Dr. Jones went on to say that the audience for S&EI includes federal and state political appointees in the R&D area, scientists and engineers on advisory boards such as the NSB, the NSF directorate advisory committee, many levels of agency advisory committees, and National Academies' task forces – all important players in the US science policy system. In Dr. Jones' words, "S&EI is the 'one stop shop' for policy makers who do not study the multiplicity of R&D statistics publications." (July 2003)

The authors are not aware of any instrumental use of publication- or patent–based indicators in national policy, that is, cases in which a decision hinged upon the publication or patent indicators. There is wide recognition, however, that statistics and indicators are often used to justify decisions or to support a particular side in a disputed issue. It is also recognised that there is not necessarily anything wrong with using statistics in this manner, but the limitations and meaning of the data need to be made clear when they *are* used in this way. These types of use of S&T indicators present substantial potential for abuse.

Today the indicators in the S&EI report are in widespread use in the US. A broad range of participants in the policy process cites the statistics. Particularly since the federal government came under pressure from the

GPRA, S&T indicators have been reinvigorated in their use. They tend to be used to show where the US stands with respect to other countries and as a reason for increasing funding to particular areas of science or technology.

Also in the US, the Council on Competitiveness has developed its own set of innovation indicators. It uses patent data together with industry analysis. They are thinking about repeating their innovation summit. It is a lot of work and financially difficult in both publications and patents (particularly patents). It needs to be done in a collaborative way.

2.2 Worldwide Diffusion of S&T Indicators and the OECD

Today many other countries use the US data or have been inspired by the US to develop their own indicators and indicator systems. The US is still encouraging the spread of science indicators by consulting with countries which are establishing systems to track and use science indicators. The indicators are broadly accepted. For example, Latin America has become involved in S&T indicators. SI helped RICYT (La Red Iberoamericana de Indicadores de Ciencia y Tecnología) to start a S&T indicators network. Over the past 5–6 years they have developed comparable data such as the SEI pocketbook data book. It covers all the Americas, including the US and Canada, plus Spain and Portugal. It includes a broad range of indicators such as R&D funding, publications, and patents.

It may be an overstatement to say that other countries have gone further than the US in the use of indicators. However, it seems that European countries may be going in different directions in the use of S&T indicators, using indicators more in benchmarking and in foresight exercises than the US (see section 2.3 below).

The idea of developing a S&T indicator system that includes publications and patents and its use by policy makers is taken further in certain countries and used more rigidly in terms of funding individual researchers than it is in the US. Countries which have a more centralised science policy system tend to use science and technology indicators in a more rigorous way. For example, in France and Mexico quantitative indicators are used in the decision to give individual researchers more funding. In the US S&T indicators are used more widely by management for general policy and awareness of trends than at the level of allocating resources to individual researchers.

Beginning in the late 1970s the Organisation for Economic Co-operation and Development's (OECD) secretariat for science and technology indicators (restructured and renamed several times in the past decades) exerted a very important standardising role within the member states of the

OECD. By inviting researchers, statisticians, and other responsible persons to join workshops, by editing manuals on R&D, patent and innovation measurement, and revising unclear national statistics to OECD standards, the secretariat carried out an important task by making national scoreboards on S&T comparable. The OECD bodies resisted producing only simple aggregated rank tables of countries' innovation performance. Even the most recent STI scoreboard (OECD, 2003), the sixth in a biennial series which started a decade ago, did not produce scalar measures of innovation activities of countries, although it did give particular attention to offering new or improved official measures for international comparisons in the major areas of policy interests. This stands in clear distinction to what is observed at the European level, which will be reported on below.

2.3 S&T Indicators in Japan

In Japan, for reasons of language, major English reports on S&T and the respective indicators began with the establishment of NISTEP (National Institute of Science and Technology Policy) in the year 1988. Certainly, important Japanese sources on R&D expenditures were published before, such as the Report on the Survey of Research and Development (Management and Coordination Agency, various years) issued annually in Japanese with English sub-titles to table and figure captions. Yet the special dedication of the foundation of NISTEP was to bring Japanese S&T information, including indicators, to an international audience. Within the mission of NISTEP to contribute to policy making by taking a sort of task force, major internationally comparable reports on S&T indicators were published. These indicators systematically organised the knowledge about scientific and technological activities of Japan and the corresponding reports were the first to make the overall state of these activities quantitatively comprehensible.

In the context of this chapter, in focussing on aggregation methods of S&T indicators, the approach of NISTEP to establish a Japanese science and technology system must be recognised. The basic method of 'integrating' S&T indicators by the Japanese institution was in using a 'cascade model' and factor analysis (Niva and Tomizawa, 1995; Kodama, 1987). The international comparison of overall strengths in science and technology was processed in such a way that 13 indicators for Japan and other countries were used to illustrate national S&T activities such as inputs in R&D, staff, output, number of scientific paper citations, and so on. That is, the multiplicity of indicators was reduced in a way that looks for similarities in the structure of the data and results in a lower-dimensional array of indicators, which is more than a simple ranking. To the best of our knowledge these activities were not fully continued in the past years.

2.4 S&T Indicators in Europe

In Europe reporting on national science and technology performance has changed markedly in the past ten years. There are two main reasons for this. First, the former communist countries did not keep with OECD conventions, and the little information which was available before around 1990 was often not comparable. It was also widely considered to be systematically overestimated. This started to change around 1990 and since then several national reports from Eastern European countries have been issued, some of them in the English language. Yet in this short chapter it is impossible to give a full account in this respect (see, for instance, Gokhberg et al., 1999; CSRS, 1998).

The second change was the more active European Commission. A landmark in this respect is the first European report on S&T indicators (European Commission, 1994), which consisted of a massive attempt to collect available data of various kinds. The Commission was assisted by a large group of leading European researchers in that area. In 2003 the third such report was issued. This new role of the European Commission triggered numerous competing activities which led some observers to note an ‘oversupply’ of S&T indicator reports (see below).

Before these two trends made themselves felt in the 1990s and in the first years of the 21st century, a variety of non-comparable reporting systems in major Western European countries was in place. Here again we do not attempt to give a full account of the 1970s and the 1980s. Nor is it possible to do justice to every country. Some of these reports were not published regularly but only in exceptional cases and in different formats and many of them are in national languages. These activities were not stopped when the European Union level came up with own products and most of the national series continue today.

To give a few examples let us mention the French report on ‘Science & Technologie – Indicateurs’ which has been published since the inception of the Observatoire des Sciences et de Technologie (OST) in Paris in 1990 (OST, various years). The series of reports is clearly subdivided into the national level, the European level, and the international level. Observers within France are proud of the many data series and the systematic and continuous way the report is published. However, the reports are known to lack analytic and policy sections and assessments to complement the data series.

In the United Kingdom various related publications exist, but periodical reports in a consistent format that provide comparable information over a longer interval were not established. Amongst the newer publications let us just mention the Department of Trade and Industry (DTI) economics paper No. 7, 'Competing in the global economy – the innovation challenge' (2003).

In Germany, since 1965 the 'Bundesbericht Forschung' (Federal Research Report) is published every four years in the German language (BMBF, various years). From time to time more or less abridged English versions are available. The latest such report appeared very recently (May 2004). Because of the federal structure in Germany, this report has a national part, a state (Länder) part, and also some international comparisons. It is more focussed on R&D inputs and R&D infrastructure, and describes large organisations in Germany. This report is assisted by the 'Bericht zur technologischen Leistungsfähigkeit' (report on technological competitiveness), which has been published annually since 1985 (with various editors; for the latest version see Grupp et al., 2003). The latter report is not as complete as the former in terms of compiling R&D data and it has a less official character. It is prepared by research institutes for the German government and is quite analytic and policy oriented. Government officials occasionally were not happy with the assessments and findings. In the case of Germany one can also demonstrate the problems with former communist countries. The very complete R&D statistics of East Germany (former German Democratic Republic) had to be adjusted in a very complicated way in order to be comparable to the Western system. This was done in the first years of unification and now comparable backward information is available. This case may be taken as typical for all the Eastern European countries.

National reports are available from Austria (Pohn-Weidinger et al., 2001; Republik Österreich, 2003), Italy, the Netherlands, the Scandinavian countries, and so forth. Most of them are in the national languages.

Returning to the European Union level, in the past several years, in addition to the three European reports of S&T indicators mentioned previously, a variety of other reporting systems were established. Benchmarking activities were started with the explicit aim of going beyond existing statistics and providing new types of data not available so far (for instance, R&D staff by gender; European Commission, 2002). These activities are co-ordinated by the Directorate General for Research and assisted by a High Level Group of Experts on Benchmarking, Excellence, Co-Ordination of National Policies. The Directorate General for Research also issued the booklet 'Key Figures 2003–2004' (2003b). A preliminary version of an 'Innovation Scoreboard' was published in 2000 by the innovation/Small and Medium Sized Industry programme of the Enterprise Directorate General. Since then the European Innovation Scoreboard has been published regularly. The same directorate also published a Biotechnology Innovation Scoreboard (2003). Both types of reports make use of 'composite indicators', which we discuss in more detail below. Composite indicators are also part of chapter 1 in the most recent European Report on Science and Technology indicators (2003).

It seems that the European Commission is driving S&T indicators in the direction of aggregation of different types of indicators into simpler constructs in order to summarise complex multi-dimensional phenomena.

2.5 Current Status of S&T Indicators

To summarise their development, S&T indicators have evolved over the past 30 years to become a large number of statistics each of which describes a portion of the science and technology system. Because they are by definition partial they must be used in combination with each other and with other kinds of data such as expert opinion to provide a full picture. Every indicator has its own unique strengths and weaknesses, and indicators need to be selected according to the problem or question being addressed (Grupp, 1998).

Progress has been made toward linking particular indicators to particular parts of the S&T system or the innovation process. Researchers have moved from the simple concepts of inputs and outputs to concepts of inputs, throughputs, outputs, and impacts at *various stages of the process*. Publications can be used to indicate the output of basic scientific research, for example, but would be misleading if used (alone) to indicate the output of industrial research and technology development. Patents, on the other hand, are useful as an indicator of applied research and technology development (*loc. cit.*).

It is increasingly recognised that some indicators are appropriate in certain contexts and not in others. For example, if the objective is to understand the development of computer software in the last quarter of the twentieth century, a patent analysis would not be recommended. This is because, although software is increasingly patented, particularly in its early years it was not, so an analysis of software patents would miss much software activity in the early years. However, if one is interested in the extent to which ownership of software patents is concentrated in a few companies, an analysis of patents would make sense.

Progress has also been made in developing indicators of quality, importance, or value, although these concepts themselves have not been defined as well as they should be.

Given the above situation, S&T indicators can be misused or abused, as well as used for positive purposes (Pavitt, 1988). Some of the possible misuses include:

- Reliance on a single indicator;
- Use of an indicator that is inappropriate for the technology, system, or stage of the R&D process;
- Drawing conclusions which are too strong, given the 'indicative' nature of indicators;
- Making inferences that are inappropriate, based on the indicator and its relationship to the phenomenon of interest.

2.6 The Development of Composite Indicators and Related Concepts

To sum up several decades of debate, the measurement of science and technology requires measurements along many dimensions. To date no ideal 'catch all' variable for science or innovation has been developed (Patel & Pavitt, 1995). Therefore in many cases multiple indicators have been used. However, the use of multiple indicators means that conventional methods such as the knowledge production function (Griliches, 1995) and many other concepts of efficiency measurement cannot be supported. Optimal configurations of measurement must be worked out in some other way (such as factor or data envelopment analysis). The recognition of the need to measure multiple dimensions of science and technology has also led to the emerging and pioneering field of composite indicators to enlighten national S&T policies.

A fortiori, the multi-dimensional science and technology (S&T) variables are usually not expressed in monetary terms but rather are measured in other units (such as patent counts, innovation counts, number of citations, etc.) and may not be comparable to each other. Lacking a well defined correspondence between relevant S&T data — for instance a conversion relationship between dollars and patent numbers — the multi-dimensional profiles cannot be aggregated into an overall scalar figure. This situation is fundamentally different from one in which all variables are fully specified in terms of quantities and costs or prices such as such well known economic indicators as the Gross Domestic Product (GDP).

On the micro-level of companies or single innovation projects, decision–oriented measurement practices such as 'benchmarking' or 'scoreboarding' have become well established. Benchmarking is the practice of identifying the organisation (e.g., competitor firm) which is the best at a particular function or activity, such as innovation, and using that organisation's metrics as the goal to be achieved and surpassed. "Benchmarking is the continuous process of measuring products, services, and practices against the toughest competitors or those companies recognised as industry leaders." (Kearnes, 1986). Although this is clearly a quantitative approach, it also has qualitative aspects. Camp, for example, refers to benchmarking as "the search for industry best practises that lead to superior performance" (Camp, 1989, p. 12). Benchmarking in the industrial context is very action oriented, aimed at improving business operations and competitiveness. The development of indicators and their aggregation are the means to an end in industrial benchmarking.

Let us note here, that, before a decision is made in a firm, the available database can be questioned and even partly laid aside. Here the assessment of quality, although it may be difficult, can be solved in some way. The database with all its biases does not automatically determine decisions. This may be different in the context of a national policy (see below).

Another concept related to S&T indicators that has developed in the business world is that of scoreboarding. Like the sports scoreboards which show how many goals, runs, or points have been scored in a competition or match, scoreboards have been developed and used to show companies how they stand with respect to their competitors on aggregate, widely recognized metrics of business performance such as productivity. In the area of industrial research and development (R&D) and innovation, 'R&D Scoreboards' and even 'Patent Scoreboards' have been developed and published.

Recent years have witnessed the increasing application of these methods, relatively uncritically, for national or regional science and technology policy. In particular, the use of composite indicators is being promoted as an emerging and pioneering field (European Commission 2003 and further references given there on p. 433, in footnote 1). At this level innovation scoreboards and the like are not usually used instrumentally to make policy decisions, because decision making in science and technology policy is quite a complicated negotiation procedure between societal interests and interest groups (Edler et al., 2003). Scoreboards of national innovation performance instead function more as 'soccer league tables' telling the public which countries are performing well or second rate, which have caught up or fallen behind.

The problems with this use of benchmarks or scoreboards on a national level lie in the lack of clear theoretical models that tell us which indicators to select, how to weight them and how to handle cross-country differences in the availability of data (Pohn-Weidinger et al., 2001; European Commission, 2003a). To say the least, this use of scoreboards or benchmarking rank tables may be dangerous because the numbers provided are taken at face value with little discussion of their validity. Substantial space exists for manipulation by selection, weighting and aggregating indicators. This chapter attempts to raise this point and to provide empirical examples of the range of interpretation or misinterpretation of national innovation scoreboards.

Successful scoreboard–based analysis should depend on mastering the art of indicator selection and scoreboard design. As a *sine qua non* for reasons of public accountability, scoreboards — as any advanced evaluation method — need a clear and transparent structure and recognised concepts (Tijssen, 2003).

3. CALCULATING COMPOSITE INDICES: ONE EXAMPLE

3.1 Methodology

In this chapter we want to investigate the robustness of innovation scoreboards empirically by sensitivity analysis of one selected case. As we have argued above, this seems to be the European speciality driven to a large part by bodies of the European Commission. In any case, it seems to be a newer development within the long standing tradition of S&T indicators. "By aggregating a number of different variables, composite indicators are able to summarise the big picture in relation to a complex issue with many dimensions." (European Commission, 2003, p. 433). What can we learn from aggregating that goes beyond the detailed information?

The procedure will be as follows: We take the original composite indicator of the European Innovation Scoreboard (2001) and compare the ranking of countries by various ranking methods, namely by

- original, Olympic, average and weighted ranks;[2]
- metric scales (weighted and un-weighted) in distinction to rank positions; and
- selective omission of data in order to 'promote' a countries' position.

We determine the weights by grouping of the various indicators into input, throughput, and output and give each group one third weight, but equal weights within the group. There are many metric scales possible; in our case we use the technometric scale originating from evolutionary

[2] The original ranking is explained below, by 'Olympic' we mean that the ranking is done by descending numbers of 'gold medals' first, then by 'silver medals' (rank position 2), and so on.

economics. It adjusts the interval to 'real market' competitive positions (Grupp, 1998).

The empirical base of our case study is the European Innovation Scoreboard 2001 (see section 2.4). In that base, a combination of 18 indicators is presented, namely S&E graduates, tertiary education, lifelong learning, employment in manufacturing and services, R&D intensity, business expenditures on R&D, European and US patents, SMEs' innovation and co-operation, innovation intensity, venture capital, new capital, new products, internet access, information technology markets, and the high tech value added.

From this scoreboard, a 'tentative summary innovation index' (SSI) is constructed, placing, for instance, Sweden in rank position 1 (score 6.5), the UK in rank position 4 (score 4.4), Germany in average rank position 9 (score 0.6) and Greece in rank position 16 (score -7.9) (see Figure 3.1). The SII is equal to the number of indicators which are 20 per cent above average minus the number of indicators that are 20 per cent below. The index is normalised to the interval [10, –10]. An index of zero represents the EU average.

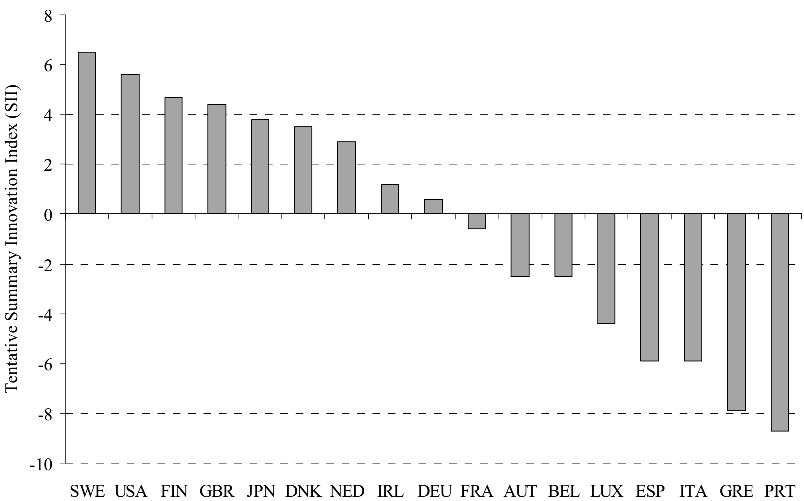

Figure 3.1. Summary innovation index of the European Innovation Scoreboard 2001 (original graph from p. 12)

One may ask several questions, for instance, why this selection of indicators and why is the aggregation done in such a peculiar way? Is it justifiable to give equal weights to 16 out of the 18 indicators but count the two sets of patent data by half each? In addition internal criticism was raised: "While this technique could prove suitable when we have the same number of indicators across countries, its relevance declines sharply when observations are unevenly distributed across countries, as is the case here." (European Trend Chart, 2003, p. 16, footnote 7). In fact, we face the problem that some data for some countries are missing, which raises special problems. Because we consider benchmarking and score boarding designs an art, we do not want to continue with arguments whether or not precisely this procedure is the best solution of all possible alternatives. But we rather want to process the given data in some other ways and compare the sensitivity of the results (the country ranking) to the original method.

3.2 Robustness of Composite Indices — Selected Results of a Sensitivity Analysis

In Figure 3.2 we provide the results of several aggregation procedures for composite indicators other than the original, as suggested in section 5.[3] What we learn from Figure 3.2 is that Sweden is in rank position 1 irrespective of the aggregation procedure. The same is true for Finland in rank position 2. All the other countries vary by one to four rank positions depending on the aggregation procedure, but overall the impression is that the country ranking cannot be completely turned upside down. The countries in top positions are always in a good position and those at the end of the scale are not positioned in the first half of the league irrespective of the aggregation details.

In particular, the average ranking and the weighted average ranking are most similar to the original SSI index, whereas the Olympic scale can change the picture more seriously. Consider the case of France: France is nowhere in best position amongst the 18 variables, thus wins now a 'gold medal' and will be placed behind all other countries with at least one gold medal (number one in one of the 18 variables).

[3] For EU countries only, thus omitting the US and Japan.

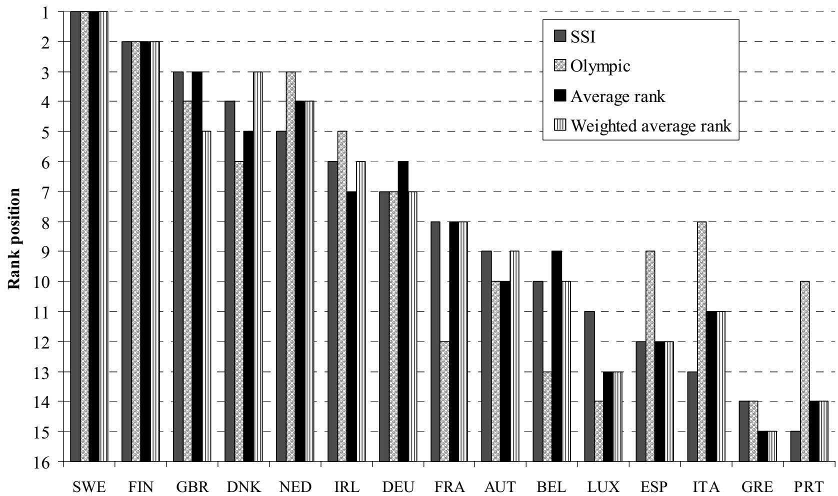

Figure 3.2. Results by selected ranking procedures

Also interesting is the case of the United Kingdom. The UK performs quite strongly in most input indicators but less so in throughput and output variables. As the set of the 18 indicators is very much input biased, equal weighing puts the United Kingdom in a favourable position. When one starts to give all inputs together the same weight as all throughputs and all outputs then this 'natural advantage' in the original SSI ranking vanishes for this country and other countries catch up.

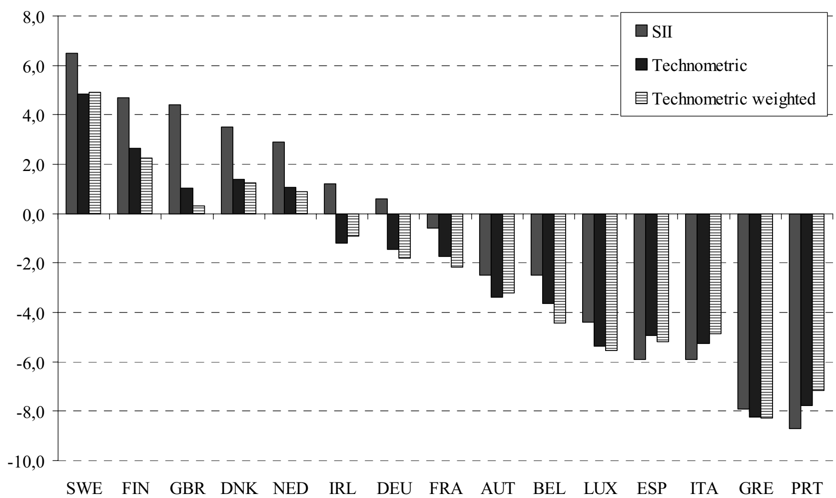

Figure 3.3. Results by selected metric procedures

Figure 3.3 displays the results by comparing the original SII index with metric procedures. Rank tables bear the problem that the distance between any two adjacent positions can be very small in original indicator values or can be large. Metric scales, in distinction, conserve the distances of the original variable values and transform them similarly.

This procedure yields quite different rankings, as is shown in Figure 3.3. Take the example of the United Kingdom again. In the SII scoreboard the country is in third place within the EU countries); however, the distance in most variables to the leading countries is much larger than the rank positions suggests; the United Kingdom seems to be closer to European average if weighted metrics scales are used. Again, the difference is still more pronounced if the weighted metrics are taken, because the 'natural advantage' in input variables vanishes. In the case of Germany the SII index is just above EU average whereas the metric values are clearly below.

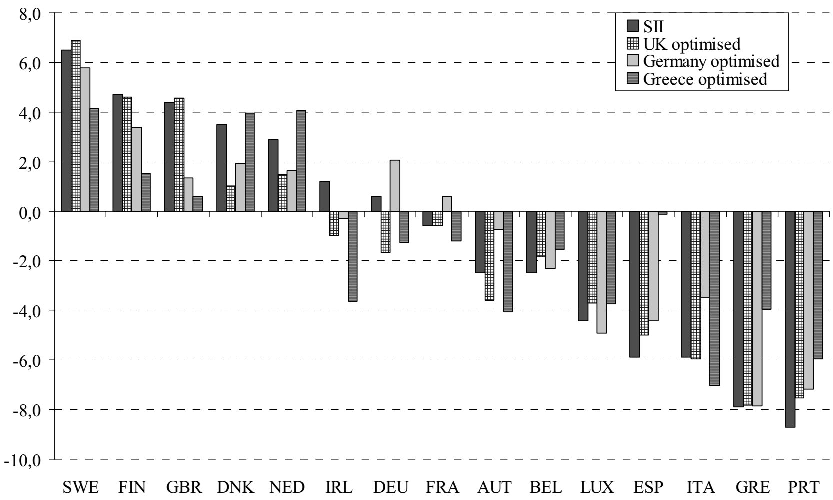

Figure 3.4. Some results by selective omission of variables

A third type of sensitivity exercise consists in selective omission of certain variables. We already have argued that scoreboard design is more an art than a science and it is difficult to argue why 18 indicators are used, not 16 or 20, and why exactly these. If we now tune the calculation by selective omission of those variables in which a certain country does not perform very well, we can try to tune the selection of variables in favour of some countries. In Figure 3.4 we present the results of an optimisation for the UK, for Germany, and for Greece.

The optimisation attempt for the UK is not very successful. The reason is that the set of 18 variables is already optimised in favour of the UK by the original SII (with its many input variables). If we exclude some output and throughput variables in which the UK does not perform very well we cannot really improve the countries' position compared to the original index. This is true for all countries being in lead positions in the original scale; these benefit from the selection by the original scoreboard and cannot really be pushed ahead further.

For countries in middle places and further down the scale one can optimise their position with more success. For instance, Germany can be put in rank position 3 by optimising the selection of indicators, because Germany performs mediocre in life long learning, venture capital, and other variables. If these are taken out of the indicator set the country's position improves considerably. The same is true for Greece, which can be brought upwards by several rank positions if a selection of variables being favourable towards Greece's performance is taken.

How plausible is country tuning? The optimised UK index is achieved by not considering EU patents and giving more weight to US data, not considering business expenditures on R&D but giving more weight to gross expenditures on R&D and to venture capital. These assumptions are not really revolutionary; in particular, in the case of two patent data sets it is questionable why both of them should go into the summary index with half weight (as is the case for the SII). Germany's index, as has been mentioned above, may be improved by not considering venture capital, the opposite assumption as for the UK – but is venture capital really a traditional core S&T indicator? Greece's profile profits from not considering patent indicators at all and leaving out high tech value added.

All these assumptions seem to be soft and can certainly be discussed seriously. Altogether, we think the selection of any one set of indicators does not give equal justice to all countries and thus the selection problem is implicitly a way to tune country positions. This is done by disputable arguments and does not need any heroic assumptions.

4. DISCUSSION AND SUGGESTED RESEARCH QUESTIONS

In this chapter we have given a very brief survey of reports on science and technology indicators in the triad regions. We argue that the European Commission is a latecomer, but is playing a more active role in recent years and is driven in the direction of composite indicators, which do not seem to be of primary concern in most single country S&T indicator efforts.

By applying alternative aggregation procedures to one selected example (introduced by EU bodies) different weights and a different selection of indicators, we have shown that summary scores and rank positions can vary considerably. It seems that ranking is less sensitive to calculating procedures but gives no information on the size of the gaps. Metric scales seem to provide more insights into relative positions of nations in the very sense of benchmarking. Sensitivity analysis further shows that exclusion or inclusion of variables tend to be a bigger problem than a slight variation of indicator scales. It seems not to be too difficult to argue for a 'country friendly' selection and corresponding weighing of indicators. To say the least, this use of scoreboards or benchmarking tables may be dangerous if the summary numbers provided are taken as such with little discussion of their validity.

The space for manipulation of scoreboards by selection, weighing and aggregation is great. Further research should remedy the situation. This chapter attempts to raise this point and to provide empirical examples of the range of interpretation or misinterpretation of national innovation scoreboards. It did not attempt to suggest more viable alternatives.

Triggering the discussion of these problems, an alternative to scoreboarding of national innovation indices is, nevertheless, suggesting, namely, the use of interval based metric scales in order not to hide the size of the gaps. The use of multi-dimensional representations is the minimum requirement, such as 'spider' charts. Maps of similarity between country structures in science and technology may have more explanatory power in particular when combined with non-quantitative methods.

More research is needed on the validity of S&T indicators, their relationship to important S&T policy concepts, their performance in different science and technology domains, their sensitivity to selection, inclusion, and alternative methods of calculation, as well as their use in the policymaking system and means of preventing their misuse and abuse.

REFERENCES

BMBF (Ed.). (various years). Bundesbericht Forschung. Bonn/Berlin.

Breitzman, A.F., Mogee, M.E. (2002). The many applications of patent analysis. Journal of Information Science, 28 (3), 187–205.

Camp, R.C. (1989). Benchmarking: the search for industry best practices that lead to superior performance. Milwaukee: ASQC Quality Press.

CSRS (Ed.). (1998). Technological innovation in Russia. Moscow: Centre for Science Research and Statistics.

DTI (Ed.). (2003). Competing in the global economy.The innovation challenge. London: DTI.

Edler, J., Kuhlmann, S., Behrens, M. (Eds.). (2003). Changing governance of research and technology policy – The European research area. Cheltenham: Edward Elgar Publishing.
European Commission (Ed.). (1994). The European report on science & technology indicators 1994. Brussels: EUR 15897 EN.
European Commission (Ed.). (2001). European innovation scoreboard 2001. Luxemburg: SEC (2001), 1414.
European Commission (Ed.). (2003). Third European report on science & technology indicators 2003. Brussels: EUR 20025 EN.
European Commission (Ed.). (2003a). Biotechnology innovation scoreboard 2003. Brussels: Innovation/SMEs Programme.
European Commission (Ed.). (2002). Benchmarking national research policies. Brussels: EUR 20494 EN.
European Commission (Ed.). (2003b). Towards a European research area science, technology and innovation – key figures 2003–2004. Brussels: EUR 20735 EN.
Gokhberg, L.N. Gorodnikova and J. Wuttke (1999). Volkswirtschaft im Übergang – Ein Vergleich der FuE-Indikatoren in Russland und Deutschland. Moscow and Essen: CSRS and Stifterverband.
Griliches, Z. (1995). R&D and productivity: econometric results and measurement issues. In Stoneman (Ed.), (pp. 52–89).
Grupp H., Legler, H., Gehrke, B., Breitschopf, B. (2003). Zur technologischen Leistungsfähigkeit Deutschlands 2002. Berlin: BMBF.
Grupp, H. (1998). Foundations of the economics of innovation – theory, measurement and practice. Cheltenham: Edward Elgar Publishing.
Kearnes, D.T. (1986). Quality improvement begins at the top, World, 20 (5), 21.
Kodama, F. (1987). A system approach to science indicators. In Grupp (Ed.), Problems of measuring technological change (pp. 65-87). Köln: Verlag TÜV Rheinland.
Management and Coordination Agency (Ed.). (various years). Report on the survey of research and development, Tokyo.
National Science Board (Ed.). Science (& engineering) indicators, Washington, D.C.
Narin, F., Noma, E. (1985). Is technology becoming science? Scientometrics, 7, 369–381.
Niwa, F., Tomizawa, H. (1995). Integrated indicators: international comparisons of overall strengths in Science and Technology. In: NISTEP (Ed.), Science and Technology Indicators 1994 (pp. 345–365). Tokyo: Science and Technology Agency.
OECD (Ed.). (2003). OECD Science, Technology and Industry Scoreboard. Paris: Organisation for Economic Co-operation and Development.
OST (Ed.). (various years). Science & Technologie – Indicateurs. Paris: Economica.
Patel, P., Pavitt, K. (1995). Patterns of technological activity: their measurement and interpretation. In Stoneman (Ed.), (pp. 15–51).
Pavitt, K. (1988). Uses and abuses of patent statistics. In: A.F.J. van Raan, Handbook of quantitative studies of science and technology (pp. 509–536). Amsterdam: North Holland.
Pohn-Weidinger, S., Gassler, H., Gruber, M., Polt, W., Schibany, A., Leo, H. (2001), Innovationsbericht 2001. Wien: Bundesministerium für Wirtschaft und Arbeit.
Republik Österreich (Ed.). (2003). Österreichischer Forschungs- und Technologiebericht 2003. Wien: Bericht der Bundesregierung an den Nationalrat.
Stoneman, P. (Ed.). (1995). Handbook of the Economics of Innovation and Technological Change. Oxford and Cambridge, Mass.: Blackwell.
Tijssen, R.J.W. (2003). Scoreboards of research excellence. Research Evaluation, 12 (2), 91–103.