Research Evaluation, volume 13, number 2, August 2004, pages 87–93, Beech Tree Publishing, 10 Watford Close, Guildford, Surrey GU1 2EP, England

Ranking method

Dealing with diversity in scientific outputs: implications for international research evaluation

M C S Bantilan, S Chandra, P K Mehta and J D H Keatinge

This paper examines the changing role and broadening goals of international agricultural research centers (IARCs), focusing on their evaluation mechanisms and priority setting processes. The case of the International Crops Research Institute for the Semi-Arid Tropics (ICRISAT) is used to identify the relative importance of outputs. It was found that, for IARCs, a wider range of credit items should be used in evaluating the institutional and individual performance. A decentralized process using nested institutional and project logframes would powerfully help to identify milestones for institutional and individual evaluation.

M C S Bantilan is Global Theme Leader; email: c.bantilan@cgiar.org; S Chandra is Principal Scientist (Statistics); email: s.chandra@cgiar.org; P K Mehta is Scientific Officer; email: p.mehta@cgiar.org; and J D H Keatinge is Deputy Director General (Research); email: d.keatinge@cgiar.org; all at SAT Futures, International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru, Andhra Pradesh, India; tel: 91-40-23296161; fax: 91-40-23241239.

THE LAST THREE DECADES have witnessed a substantial broadening in the goals of international institutes engaged in agricultural research. These have developed from food production alone to now include resource management, equity, gender and environmental concerns. Accordingly, there has been an expansion of their work agenda across the full range of the research for development continuum from basic/strategic research to applied and adaptive research with farm- and policy-level applications (Huffman and Evenson, 1993). As a result, creditable products from agricultural research now include a much wider variety of outcomes from new knowledge, ideas, concepts, methods, and techniques relating to strategic and basic research to more downstream developmental outputs including patents, pilot studies, farmer field schools, policy briefs and extension materials. This diversity of agricultural research outputs has resulted in a considerable expansion of the concerns of scientists beyond their traditional academic disciplines and activities.

In the face of these changes, there is a need to review the ways in which the tangible outputs of agricultural research institutes and their individual scientists are assessed. Traditionally, the tools and indicators of research evaluation are based on bibliometric analysis (numbers of publications, citation indices, journal impact factors, etc.) and peer review. But the broadening of the research focus raises questions about the adequacy of these tools to evaluate institutional performance. A single-minded concentration on bibliometric analysis totally ignores a

 0958-2029/04/020087-07 US$08.00

whole range of products that can now be seen to be relevant for research for development and is thus, used alone, an inadequate tool for assessing performance. This is due to a number of factors.

First, published papers are the result of laboratory and field activities; other innovative activities are not traditionally formally published to the same degree. Second, books and journal articles cannot contain all the knowledge produced in research. In addition, these ignore the tacit and material elements of research that cannot be communicated through publication channels (Hicks and Katz, 1997). Moreover, journal impact factors are usually adequate measures of comparison only within a discipline but are highly misleading when used to make comparisons between disciplines (Amin and Mabe, 2000). For example, even within agricultural research the impact factors of most agricultural economics journals are substantially lower than those of soil science or other environmental journals. This does not make the work of agricultural economists either of poorer quality or less intensively reviewed than those of soil scientists. With respect to peer review, while the experiences of the Consultative Group for International Agricultural Research (CGIAR) and the UK Universities Research Assessment Exercise (RAE) show that they can be used at higher levels of aggregation (i.e. institutional, departmental or team level), Phelan (2000) argued that peer reviews are more appropriate when evaluating the work of individual scientists and are much less effective in evaluation of a collection of work such as that produced by an institute.

The conduct of research is a complex activity. Neither a single measure nor use of only a few performance indicators provides adequate assessment of research. Rather it is necessary to use a group of relevant indicators (Butler *et al*, 2002). Additional mechanisms of evaluation that recognize almost all important research outputs are required to evaluate the performance of scientists and institutes in a realistic and objective manner.

Objectives

This paper has two objectives. First, it aims to present additional perspectives in the assessment of research performance by illustrating a mechanism that realistically takes into account the broadening set of objectives, and corresponding tasks and the diversity of outputs reflecting the changing nature of priorities in international agricultural research institutes. Second, it illustrates a method showing how a set of priorities among the whole range of products is processed and computed to measure relative importance.

The broadening of the research continuum

This section expounds on the broadening research continuum, using the case of one of the international agricultural research centers of CGIAR, namely International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), as a basis for illustration.

In the three decades of ICRISAT's history, its scientific and training portfolio has continuously evolved in response to various factors, including the Institute's learning process, changing regional and global priorities and opportunities in the Semi-Arid Tropics (SAT), redefinition of research targets in the light of research findings and the changing research requirements of development investors. In the last 10 years, this evolutionary change has not only been rapid but also has had far-reaching consequences as emphasis was refocused from the principal target of improving crop productivity and food security to that of achieving impact on sustainable livelihoods, poverty eradication and the protection of the environment. These changes have considerably expanded the work agenda of ICRISAT, as it was required to engage itself across the research-for-development continuum according to the capacity of its partners.

For example (as shown in Figure 1), when the Institute was established with a mandate focused on

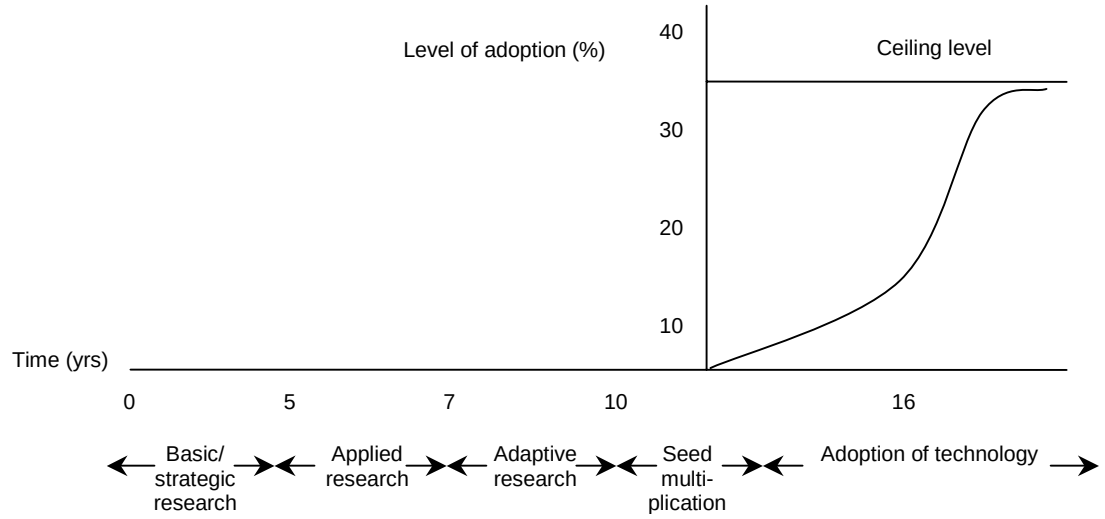

Figure 1. Broadening of research for development continuum
Source: ICRISAT (1992), *ICRISAT Medium Term Plan 1994–1998*

the SAT in the early 1970s, it set priorities oin strategic and basic research in partnership with many institutes including advanced research institutes (ARIs). Figure 1 illustrates the whole research continuum ranging from initial research efforts to impact on the ground. Initial stages involve basic research, such as development of breeding populations and germplasm characterization. Subsequently, scientists are also engaged in both applied (the development of seed-based technology with testing leading to an identifiable product) and adaptive research (the stages of testing leading to release by the national agricultural research systems [NARSs]). Figure 1 depicts development of optimal seed multiplication strategies and adoption of the technology as the final stages to achieve impact. Based on ICRISAT Medium Term Plan data (ICRISAT, 1992), on an average, it takes around five years to undertake basic and strategic research, four or five more years to produce an improved variety, and another five or six years to reach the ceiling level of adoption.

Changing demands encouraged the production of a whole range of research outputs. With an increasing demand for impact and resource mobilization, scientists have been required to be involved in activities including writing project proposals and concept notes, organizing stakeholder workshops, producing public awareness flyers, assisting in farmer field schools and providing policy briefs. The participatory approach has also become an integral part of the working system of the Institute. This means there are many new aspects of research and development excellence, other than books and journal articles, which also need to be recognized when assessing the research performance of the Institute. Thus, the full range of such outputs produced by scientists must be identified and then ranked in importance for the Institute to help in the evaluation of research performance. This ranked output list should also help in getting individual researchers to strike a proper balance of diverse activities along the research for development continuum.

It is important to mention here that our analysis is restricted to the evaluation of the research efforts irrespective of their outcomes. This is mainly because the outcome of the research or its impact is dependent not only on quantifiable variables but also on other variables or factors that are difficult to quantify. It is an exercise to understand the evaluation system irrespective of the outcome of research. Such analysis provides the basis for a retrospective analysis for evaluation.

Statistical methodology

A survey at ICRISAT in December 2002, based on an earlier basic list compiled by scientists at the International Institute of Tropical Agriculture (IITA), produced a comprehensive list of 97 relevant scientific outputs in consultation with scientists situated in all its locations in Africa and Asia. These outputs were placed into four broad categories: Products, Writing, Editing and Training. As listed in the first column of Table 1, Products contains 25, Writing 38, Editing 10 and Training 24 outputs. ICRISAT has identified these as output indicators against which the research performance of each staff member is planned to be regularly monitored. All scientists across the Institute were individually asked to rate each output on a 1 to 10 scale in terms of its *perceived benefit/significance to the Institute* rather than to individual scientists, with 1 being the least important and 10 being the most important. Table 1 provides a comprehensive coverage of all important research outputs, both traditional and non-traditional. While most outputs lend themselves to measurement, it is recognized that not all outputs can be readily quantifiable and there is indeed a need for further work on their measurement or more rigid definition.

The survey was undertaken as part of the Institute's process of developing a new appraisal system and was initiated from the office of the Deputy Director General of Research. The bottom-up participatory group exercise that was conducted enables the development of an evaluation system informed by researcher views. At the time of the survey, there were six global themes (GTs)1 at ICRISAT, which represented its research structure delivery mechanisms. The rating data were analyzed separately for each GT in view of the likely inherent heterogeneity in the way the different GTs might perceive the benefit of an output to the Institute. With a response

Table 1. Relative ranking of outputs in different output categories across global themes at ICRISAT, 2002

Table 1a. Products

Outputs	Weighted median
New techniques for scaling out and up	8.58
New varieties	8.51
Introgression lines for fundamental research	8.31
Biotech products	8.08
New techniques	8.06
Integrated pest management (IPM) strategies	7.95
Integrated natural resource management (INRM) strategies	7.93
Watershed management plans	7.91
Seed system design	7.82
Biotech constructs	7.67
Crop/livestock integration strategies	7.66
Databases/catalogues	7.63
Improved germplasm	7.62
Videos/CDs/audio tapes	7.51
Geographical information system (GIS) maps	7.40
Protocols/tools	7.39
Diagnostic kits/tools	7.38
New food products	7.20
Biosafety protocols	7.08
Pre-breeding derivatives	7.08
Biocontrol agents	7.05
Post harvest machinery and storage design	6.16
Other GIS products	6.06
Computer software	5.88
Chemical products	5.78

Table 1b. Writing

Outputs	Weighted median
Books (in your discipline)	9.31
Journal articles (hard copy)	8.92
Edited books (peer-reviewed conference proceedings, etc.)	8.68
Book chapters (peer-reviewed in your area of specialization)	8.63
Project proposal documents	8.42
Policy briefs for decision-makers	8.03
Concept notes	7.88
Technical bulletins	7.84
Ex post impact reports	7.55
Varietal or chemical product descriptors or germplasm registration notes	7.54
Training manuals	7.50
Electronic papers	7.47
Conference papers	7.44
Monographs	7.29
Extension materials (printed)	7.25
Extension materials (audio visuals)	7.22
Websites or pages	7.11
Patent documents	7.10
Invention disclosures	7.02
Extension posters	7.02
Ex ante impact reports and impact pathway studies	6.99
Consultancy reports	6.98
Network reports	6.81
Public relation (PR) material	6.81
Newsletters	6.56
Conference posters	6.54
Press releases and new items	6.31
Newsletter articles	6.23
Biosafety policy briefs	6.15
Institutional internal policy documents	6.07
Institutional change documents	5.91
Abstracts	5.73
Trademark establishment documents	5.72
Internal reports and research notes	5.70
Institutionally generic power point presentations (PPTs)	5.67
Activity profiles	5.10
Bibliographies	4.83
Engineering and other blueprints	2.77

Table 1c. Editing

Outputs	Weighted median
Books (your discipline)	8.81
Journal special issue editions (your discipline)	8.80
Conference proceedings (your discipline)	8.34
Paper for external journals (assume you are doing five per year)	7.86
Papers peer-reviewed internally for colleagues (assume you are doing five per year)	7.46
Project reports (for donors)	7.30
Institutional public awareness material	7.10
Global theme reports (for ICRISAT)	6.98
Newsletters	6.69
Intellectual property right (IPR) and patent documents	6.25

rate of more than 95%, we have data from nearly the whole population of each GT. Sampling error in our inferences can therefore be considered to be effectively absent. On this basis, as also due to the responding scientists not being a random sample, formal statistical tests of significance were not used in drawing inferences.

Table 1d. Training

Outputs	Weighted median
Partnership building	8.36
Visitors (donors)	8.36
Stake holder workshops	8.28
Higher degree students	7.96
Training workshops	7.81
Young scientist-in-house mentoring	7.69
Field days	7.65
Policy briefings	7.64
Farmer field schools	7.64
Training courses	7.58
Extension/non-government organization (NGO) demonstration days	7.58
Visitors (scientists)	7.56
Inter-center team-building activities	7.48
Seminars	7.38
Visiting scientist mentoring	7.19
Monitoring and evaluation efforts	7.17
Industry dialogues	7.05
Electronic teaching distance learning modules	6.93
Lectures	6.82
Project study tours and retreats	6.72
Non-degree training	6.68
PR collaborations	6.65
Other students	6.46
Visitors (general public for education)	5.94

To get comparable results, we chose two non-parametric measures, the *median* and the *median absolute deviation* (MAD), to describe respectively the *location* (central tendency) and the *scale* (dispersion) characteristics of each GT population. The median and the MAD were computed for responses on each output for each GT separately.

In order to obtain an Institute-wide picture of perceived significance of an output in a given category (Products, Writing, Editing and Training), the average response for the output across the GTs was computed as a weighted median

$$\mu=\sum_j w_j \mu_j,$$

where $w_j=(1/\sigma_j^2)/\sum_j(1/\sigma_j^2)$ is the weight given to the median μ_j, and σ_j is the MAD for j^{th} GT (j=1,…,6).

This weighting scheme duly accounted for the variability in the responses within the GTs in order to obtain an objectively derived Institute-wide average response. These average responses were ranked in each category and across the categories to identify the most important Institute-wide outputs.

In general, the essential steps of the method are:

1. Obtain a list of outputs considered relevant by scientists of the institution.
2. Group the outputs into a few broad categories as relevant to an organization.
3. Let the scientists rate the outputs as perceived by them to be important to the institute, using an appropriate rating scale
4. Apply the statistical tool suggested above to objectively identify important outputs from the whole range.

Results

The four sections of Table 1 depict the Institute-wide average response across the six GTs for the four categories of outputs. Column 2 of the table indicates ranks in a category in their decreasing importance. The ranking of the whole range of outputs across the Institute (broadly classified into Products, Writing, Editing and Training) is obtained by using a weighted median.

The results show a set of outputs belonging to all the four categories that are important from the point of view of the scientists. In the category of Writing, books, journal articles, book chapters, project proposal documents, policy briefs for decision-making and concept notes are viewed as important by the scientists. In the case of Editing, journal special issue editions, conference proceedings and papers for external journals are highly rated. For the Products category, new techniques for scaling up and out, new varieties, introgression lines for fundamental research, biotechnological products and new methods are considered most important. In the category of Training, partnership building, visitors (donors), stakeholder workshops, higher degree students and training workshops are given higher importance.

A discipline-wise (GT) analysis captured the inherent heterogeneity in responses from scientists grouped according to related disciplines in the survey. It was observed that scientists' responses reflected their respective different roles within the Institute and even within a single project. For example, the group of agro-ecosystem scientists have ascribed higher importance to the outputs such as integrated natural resource management (INRM) strategies, new techniques for scaling out and watershed management plans, and lower weights to biosafety protocols and biotech products, whereas the group of breeders and biotechnologists have assigned higher weights to improved germplasm and biotech products. This means that the same suite of evaluation indicators may not be appropriate at all levels of aggregation: individual, group or institution.

A discipline-wise (global theme) analysis captured the inherent heterogeneity in responses from scientists grouped according to related disciplines in the survey

Figure 2 illustrates the relative ranking reflecting high-priority versus low-priority outputs at the Institute-level. It depicts the ranking of all (97) outputs across the four output categories based on the Institute-level average responses. It is clear that products related to Writing activity (books, journal articles, edited books, journal special issue editions, book chapters, etc.) are highly valued by the scientists. High importance is also ascribed to project proposal documents, partnership building and visitors (especially donors), reflecting the increasing amount of time that scientists spend on these activities. They clearly recognize that these have become critical

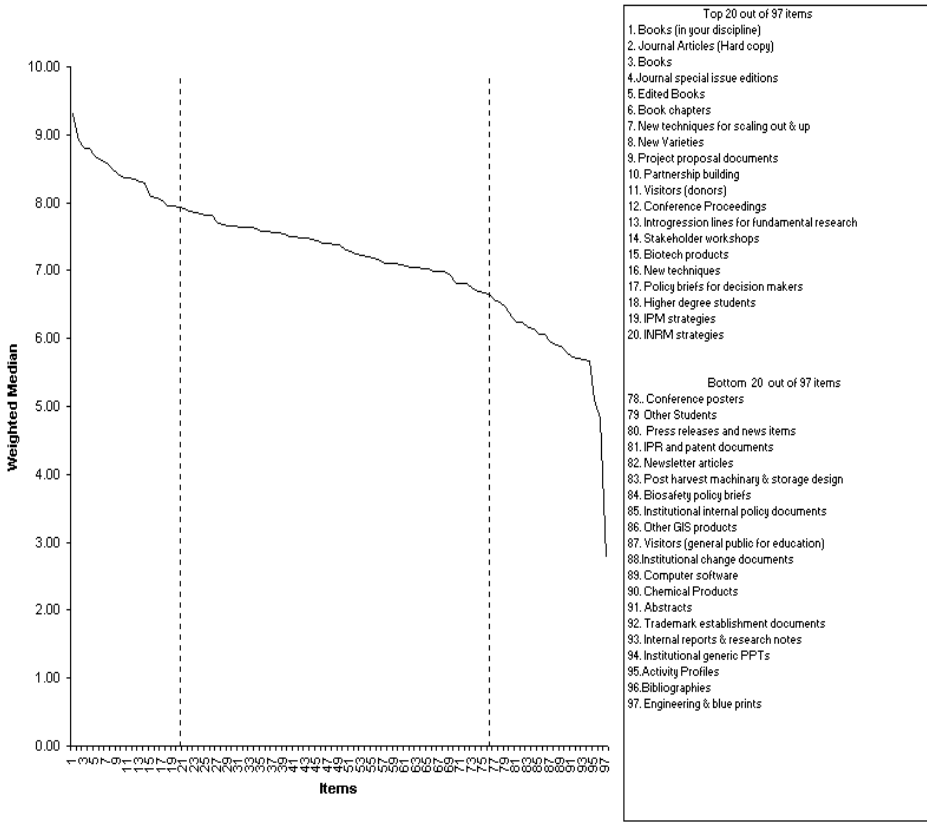

Figure 2. Relative ranking of outputs across output categories

activities for the continuing existence of the Institute and its service to its mandate region. High importance is also placed on the development of an appropriate strategy for the SAT region through its strategic assessments and policy briefs; likewise in INRM and integrated pest management (IPM), which underlies the importance of the quality of international public goods (IPGs). Conference proceedings, stakeholder workshops and interactions with the private sector are ranked important for the Institute in its pursuit of broadened strategic partnerships as required by its external reviews. Traditional partnerships of value, such as higher degree students, remain highly rated.

ICRISAT scientists have thus retained traditional priorities on improved varieties and hybrids and taken on board new priorities of impact generation with new varieties and novel approaches as well as new techniques for scaling out and up. New types of activities, such as policy briefs and biotech products, are also on the priority list. The least preferred items across the Institute include engineering and other blueprints, bibliographies, activity profiles, press releases and news items, along with chemical products, trademark establishment documents, computer software and post harvest machinery. Lower importance is also attributed to institutional internal policy documents, internal reports, research notes and institutional change documents. This indicates that there are institutional outputs that are important to the Institute and remain part of high-priority core activity but are viewed as less important by scientists.

The Institute-level picture drawn above tends to be different, if we further disaggregate and look at the discipline or scientist level. Since each discipline within an institution differs in nature, it is quite obvious that the importance of outputs also varies according to the disciplines. Furthermore, the picture at the scientist level also tends to differ as each scientist within the same discipline has a different suite of work and projects, which results in a different set of research for development outputs. Therefore, the measures that are applied at the aggregate level may not necessarily be the most appropriate at the individual level.

For instance, each scientist is engaged in different kinds of projects and their outputs depend mostly on the objectives and corresponding expected outputs of the project(s). This implies that achievements of the scientists are more closely related to their projects. Since there are several scientists and the duration and outputs of each project differs significantly, a logframe — consisting of the description of the project, expected outputs, progress and achievements — is a more powerful tool to identify suitable criteria for evaluation. In cases such as where the expected output of the project includes journal articles or books, then the evaluation may be based on bibliometric analysis or peer review. But, where the expected output of the project includes production of varieties or improved germplasm, the evaluation criteria must focus on indicators appropriate to those specific categories of output.

Implications

This paper uses the case study of ICRISAT, which is one of the CGIAR centers, and the application of this method could also be appropriate for replication in other sister CGIAR centers and in other institutes, such as ARIs, NARSs, etc. The information given in the paper serves as a firm basis for assessing and prioritizing outputs, according to the differential nature of the specific institute.

The results of this analysis have shown that for an agricultural institute like ICRISAT, evaluation of research performance should include a wide range of outputs across various categories instead of relying only on a few, narrowly based indicators. The main findings of the study demonstrate the need for more comprehensive measures that could be used to evaluate research performance by taking into account the broadened and diverse nature of research objectives today.

The findings of this study have implications for institutional-level policy formulation. First, it brought out the fact that there are research outputs that are important for the institution (e.g. internal strategic documents), but these may be viewed or perceived by scientists as relatively unimportant. This calls for the provision of an appropriate incentive or disincentive system for researchers or substantial team ethos if they are to be fully involved in activities that are critical for the institution rather than for themselves as individuals. Also the importance attributed to project proposals, concept notes, donor visits, partnership building and stakeholder workshops does clearly reflect the increasing amount of time that researchers and scientists have to spend on these activities.

Allocation of time resources for such efforts in annual work planning documents therefore needs to be realistic rather than done in a token fashion, as is often the case at present. With respect to activities related to project proposal development, this result also implies a need for having improved information regarding objectives, priorities and targets of development investors to facilitate the effectiveness of researchers' efforts in project proposal development and to improve success rates.

This analysis can also be used to assess the comparative performance of staff within a discipline/GT, given an appropriate quantitative indicator determined for the discipline/GT. A differential weighting scale may serve as a basis for assessing scientists' performance in a refined evaluation system. Finally, a properly scaled aggregate measure for the institution, as proposed here, may be used as a standard to compare performance across disciplines/GT and evaluation periods.

The data collected through the survey provides important information about the wide range of outputs produced by the Institute. Assigning ranking to the outputs gives a useful picture of the critical or vital outputs for the Institute as a whole. In addition,

the toolbox, which represents all sets of outputs of the Institute, also may be used as a guideline to monitor or evaluate the performance of the scientists and to keep the balance across the whole range of outputs by prioritizing them according to the objectives. For example, each scientist is usually engaged in different set of projects and accordingly their priorities, objectives and hence outputs differ from their colleagues within the same discipline. Thus, the ranking system is useful for the head of the discipline to determine or identify the set of important outputs at the discipline and scientist levels. This information can be used to formulate a strategy to keep the balance of the activities by proper allocation of time and resources. This kind of system also encourages more openness, flexibility and transparency in the system.

Acknowledgements

We would like to thank Dr Jeff Davis (RIRDC, Australia), and all the scientists of ICRISAT and IITA who helped draw up the product list, and especially Drs F Bidinger, H C Sharma, K N Rai, O P Rupela, A Hall, P Singh, S Twomlow, C T Hash, B V S Reddy and C L L Gowda, for their valuable comments.

Note

1. ICRISAT had six global themes (GTs) in 2002–03:
   GT 1 — Harnessing biotechnology for the poor,
   GT 2 — Crop management and utilization for food security and health,
   GT 3 — Water, soil and agro-biodiversity management for ecosystem sustainability,
   GT 4 — Sustainable seed supply systems for productivity,
   GT 5 — Enhancing livestock productivity for wealth creation, and
   GT 6 — SAT futures and development pathways.

References

M Amin and M Mabe (2000), 'Impact factors: use and abuse', *Perspectives in Publishing*, 1, pages 1–6.

L Butler, G Laudel, F Jackson, D Siddle and I Lucas (2002), 'Strategic assessment of research performance indicators', ARC Linkage project, <http://repp.anu.edu.au/Linkage%20grant.htm>

D Hicks and J S Katz (1997), 'The changing shape of British industrial research', STEEP Special Report No. 6, Science Policy Research Unit (SPRU), Brighton.

W E Huffman and R E Evenson (1993), *Science for Agriculture: A Long-term Perspective* (Iowa State University Press).

International Crops Research Institute for the Semi-Arid Tropics (1992), *ICRISAT Medium Term Plan, 1994–1998*.

T J Phelan (2000), 'Evaluation of Scientific Productivity', *Scientist* 14[19], page 39, 2 October.

Royal Academy of Engineering, (2000), 'Measuring excellence in engineering research' (RAE, London) <http://www.raeng.org.uk/news/publications/reports/pdfs/Measuring_Excellence.pdf>