This report does not necessarily reflect the ideas of the service providers of the five individual reports, but is based on the views and opinions of its authors.
REVIEW OF THE NRF RATING SYSTEM:
SYNTHESIS REPORT

By

Thomas Auf der Heyde & Johann Mouton

2 November 2007
EXECUTIVE SUMMARY.......................................................................................................... 1
  Background....................................................................................................................... 1
  Findings ............................................................................................................................ 1
1. BACKGROUND TO THE REVIEW AND THE FIVE STUDIES............................... 7
2. METHODOLOGICAL APPROACH OF THIS REVIEW.................................................... 8
3. ANALYSIS OF INDIVIDUAL REPORTS.................................................................. 10
  3.1 Summary of and comments on Report 2.1 (An historical review and analysis of
      the NRF rating system: 1983-2005; S. Krige)........................................................ 10
  3.1.1 The brief and questions addressed .................................................................... 10
  3.1.2 Approach and methodology ............................................................................. 11
  3.1.3 Key findings and conclusions of report............................................................. 12
  3.2 Summary of and comments on Report 2.2 (Mapping the formal and informal use
      of the rating system over time by various institutions; C. Lombard)................. 32
  3.2.1 The brief and questions addressed .................................................................... 32
  3.2.2 Approach and methodology ............................................................................. 33
  3.2.3 Key findings and conclusions of report............................................................. 35
  3.3 Summary of and comments on Report 2.3 (Impact of the NRF evaluation and
      rating system: A review; H. Marais).................................................................. 46
  3.3.1 The brief and questions addressed .................................................................... 46
  3.3.2 Approach and methodology ............................................................................. 47
  3.3.3 Key findings and conclusions of report............................................................. 49
  3.4 Summary of and comments on Report 2.4 (The NRF evaluation and rating
      system in the world context; A. Pouris)................................................................. 58
  3.4.1 The brief and questions addressed .................................................................... 58
  3.4.2 Approach and methodology ............................................................................. 58
  3.4.3 Key findings and conclusions of report............................................................. 60
  3.5 Summary of and comments on Report 2.5 (Review of processes used to map the
      rating of individual researchers; M. Madikizela).................................................... 71
  3.5.1 The brief and questions addressed .................................................................... 71
  3.5.2 Approach and methodology ............................................................................. 71
  3.5.3 Key findings and conclusions of report............................................................. 75
APPENDICES....................................................................................................................... 79
  A) List of abbreviations and acronyms ........................................................................ 79
  B) Terms of reference of review.................................................................................. 81

LIST OF TABLES
Table 1: Categories identified by De Wet........................................................................... 16
Table 2: Rating classification system................................................................................ 18
Table 3: Map of uses made of rating system by universities........................................... 38
Table 4: Uses made of the rating system by museums....................................................... 40
Table 5: Lapsed ratings of researchers............................................................................. 50
Table 6: Average per capita article count: rated/non-rated researchers, 1990-2005........ 51
Table 7: Relative publication performance of rated researchers..................................... 52
Table 8: Evaluation systems that involve "rating"*............................................................ 61
Table 9: 2007 PBRF Indicative Funding.......................................................................... 66
Table 10: Members of the SNI by category and level, 1992–2005..................................... 70
Table 11: Strengths and weaknesses in key procedural steps......................................... 75
LIST OF FIGURES

Figure 1: Funding sources NRF-rated and unrated researchers .................................................. 56
Figure 2: Relationship: Number of rated researchers and NRF funding leveraged .............. 57
Figure 3: Members of the SNI, 1984–2005, Mexico ................................................................. 69
Executive summary

Background

This report constitutes a synthesis report of five separate reports commissioned by Higher Education South Africa (HESA) and the National Research Foundation (NRF) as part of a joint evaluation being undertaken by them of the NRF Rating and Evaluation System. The five reports are titled as follows:

- Mapping the formal and informal use of the rating system over time by various institutions, by C. Lombard
- Impact of the NRF evaluation and rating system: A review, by H. Marais
- The NRF evaluation and rating system in the world context, by A. Pouris
- Review of processes used to map the rating of individual researchers, by M. Madikizela

In the subsequent Executive Summary the five reports are referred to by the surname of their respective author.

Findings

1. The NRF rating system was initially designed and implemented as a recognition and reward system

Two major and related forces led to the establishment of the NRF rating system in 1984. On the one hand, South African science at the time was focusing increasingly on strategic modes of research with the bulk of R&D expenditure being spent on military and energy projects. This required closer co-operation between the CSIR (the pre-eminent research council at the time), the universities and industry. The result was a growing perception that basic and self-directed research at the universities was suffering and in fact a study at the time showed a decline in research output in the SET fields. On the other hand, the growing isolation of South African science caused increasing numbers of top scientists to leave the country and seek employment overseas.

The rationale behind the proposed rating system found its justification in a philosophy that prioritised the individual scientist and scholar. The aim would be to recognise and reward the individual scientist and in doing this, strengthen the basic science base of the country and (hopefully) reverse the increasing brain drain of top scientists. Recognition of scientific excellence would be based on an assessment of the past performance of the individual scientist. Based on
this assessment, which would be quantified in a rating score, a monetary reward commensurate with the rating would then be disbursed to the individual.

The main “design features” of the original model as proposed by De Wet in 1984 and implemented soon afterwards were the following:

- **Recognition of excellence**: The focus on the individual scientist and his/her past research performance
- **Simple categorisation of the rating outcome**: The initial system involved three categorisations only: A, B and C.
- **Appreciation of the value of fundamental, self-initiated research**: The specific research interests of the individual scientists were explicitly not taken into account.
- **Direct and automatic linking of funding**: The system involved an allocation of funding commensurate with the rating score (A-rated scientists receiving more money than B-rated scientists)

With the delinking of the financial reward component around 2000, the system effectively became one that recognises and honours scientific excellence only, with no direct link to funding.

**2. The rating system has become increasingly and unnecessarily complex**

The rating system has over the years become increasingly complex in its administration. The initial concept of the rating system worked with a relatively simple categorisation of scientists into three categories. Twenty years later the system is overly complex and unnecessarily elaborate. Various factors have contributed to this situation: the addition of social science and humanities scholars in 2003 which increased the workload of the system administration; the increase in rating categories from an initial 3 to the current 11, and more elaborate internal screening; assessment and appeals procedures to address concerns about bias and subjectivity expressed in various reviews of the system.

**3. Key substantive concerns about the system remain unresolved**

Various substantive concerns about the system were raised from the very first external review in 1991 onwards. These included questions about the meaning of the rating descriptors, too much emphasis on quantity rather than quality of output, the low funding for C-rated scientists, the appropriateness of the system for engineers, social scientists and those in the performing arts, and the delinking of automatic funding around 2000. Unfortunately none of these substantive issues have been addressed and they remain, in most cases, unresolved. As illustrated in the Krige
Report there is still no more clarity on the meaning of key terminology (e.g. “international recognition”, “leader in a field”, “proven track record”). In fact, with the inclusion of the social sciences and humanities in 2003 these terminological concerns resurfaced and became even more pronounced. The fact that a significant portion of the scientific community, i.e. medical and health scientists, do not participate in this system, seriously questions any claim as to the inclusiveness of the system.

4. The rating system faces growing scepticism and disillusionment amongst key sectors of academia

The Krige Report records a growing scepticism and even disillusionment by more recent reviews of the system and most notably the 2005 Institutional Review which concluded that the whole rationale for the system needs to be rethought. Unfortunately, as the report also shows, the NRF has over the years neglected to engage with the more substantive concerns and criticisms of the various reviews undertaken, in most cases only responding to technical and procedural issues.

Some of the reasons for the growing disillusionment with the system are covered in the Lombard Report, which identifies the perceived lack of incentives (since funding is no longer linked to rating) as a major negative factor. The effective delinking of research grants from the rating system as a consequence of the growth in funding requests considerably exceeding available funds has had, according to the report, an extremely negative effect on everyone’s perception of the rating system. Researchers are subject to competition in a dual process now – rating and project evaluation – but only the latter actually brings with it measurable rewards.

In addition there continues to exist a strong divide between the Science, Engineering and Technology (SET) and Social Sciences, Humanities, and Law (SSHL) disciplines over the acceptability of the rating system (although in a few universities this is not the case). Many in the SSHL domain remain convinced that there is a fundamental discontinuity between the rating system and the prevalent system of science in this area because much research in SSHL is (a) multidisciplinary whereas the rating system seeks depth rather than breadth of scope, (b) locally applied, and therefore unsuitable for international peer reviewed academic journals, a key benchmark of the rating system, and (c) many of the disciplines in this domain have little or no counterparts outside South Africa, thus making meaningless the rating system notion of “international standing”. Broadly-speaking, researchers in SSHL believe the rating system to have been crafted around the scientific practices of SET, and its applicability to SSHL is therefore in question. In the health sciences a specific hurdle is the conflict between most academics’ dual appointment as a clinician (focused on patient care) and an academic (expected to do research).

There is also some evidence of growing apathy towards the rating system. The Lombard study canvassed the opinion of 1700 rated scientists: only 11% took the trouble to complete the
questionnaire. This might be indicative of the more general phenomenon of survey fatigue in the higher education sector, but might also signal some degree of disinterest amongst (rated) scientists in the system. Certainly another piece of evidence (presented in the Marais Report) points to increasing apathy: for four consecutive years of the last five there has been a steady and significant increase in the number of researchers whose rating has lapsed.

5. The rating system has had a discernible influence at university strategic, research and performance management level

The Lombard Report reveals a significant formal and informal incorporation of the rating system into university strategic, research and performance management, but a much less developed approach in other research-performing organisations. Largely, the rating system is used more broadly in the domains of SET than in SSHL, though in some universities the rating system is also used strongly within the latter disciplines. No other research-funding organisations make explicit use of the rating system, not even other state funding agencies, although large corporate research sponsors indicate that the rating of a researcher does influence their funding decisions.

The Lombard Report also makes an interesting observation in that there is no clear link between the number of rated researchers and extent of utilisation of the rating system in an institution: at the extremes, UCT with the largest number makes little apparent use of the system, whereas TUT, for instance, with a small number of rated researchers makes extensive use of the rating system.

6. The rating system has only been minimally implemented by other research-performing and research-supporting organisations outside the universities and NRF environment

The Lombard Report also demonstrates the very limited implementation and/or adoption of the rating system by other organisations within the National System of Innovation, especially public ones such as the Medical Research Council, Water Research Commission, Council for Scientific and Industrial Research, and so on. This means that economies of scale associated with the costs of the rating system have not been maximised across the National System of Innovation.

7. Most researchers experience little or no benefit from the rating system

The Lombard Report shows that although some rated researchers derive financial benefits through relevant institutional policies, except for the most highly-rated ones researchers largely experience little or no benefit from the rating system, which improves neither their research profile nor networking abilities, nor their ability to attract funding. Although the applicability of Marais’s report to this finding is very limited because of methodological constraints, he also recorded “cynical rejection” on behalf of respondents (though he also reported some support for the rating system).
Marais also reported the increasing number of rated researchers who let their ratings lapse, which may also signal increasing apathy about the system on the part of researchers.

8. **The internal administration of the system is well-managed and perceived to be efficient but no insight was gained on the cost or efficiency of processes external to the Evaluation Centre (EC)**

The Madikizela Report concludes that the processes supporting the rating system are mostly efficient and effective, although a few modifications are recommended in some cases. There certainly is much anecdotal evidence that the internal administrative processes in the EC are efficiently managed, this being confirmed by our own experience of the EC’s operations. But this conclusion provides little or no insight into those parts of the rating system that operate outside – and, importantly, beyond the control of – the EC: the brief for this report clearly intended those to also be reviewed. Inasmuch as the sustainability of the rating system is critically dependent on the support and buy-in of the country’s researchers and their ability to sustain it, their almost complete absence in the Madikizela study constitutes a significant constraint to the usefulness of its findings.

The Madikizela investigation therefore suffers from two major shortcomings – the first is that processes outside the EC were not probed at much depth, and the second is that there are no obvious benchmarks against which to evaluate the performance of those processes, especially those internal to the EC. The latter constraint is intrinsic to the task set the author of the report, and although the absence of external benchmarks provides a significant conundrum, not much evidence of attempts to address it can be found in the report. For this reason too, therefore, it is difficult to assess the performance of the entire scope of processes. In addition the assessment of the rating system’s cost effectiveness has been insufficiently elaborated.

9. **The impact of the rating system on research productivity is not discernable**

The Marais Report demonstrates that rated researchers are significantly more productive on a per capita basis in terms of journal articles than unrated researchers, but it was not able to establish that this is true also for other forms of research output like books and book chapters, conference proceedings, patents, artefacts, and students. Analysis also revealed rated researchers to constitute 82% of the 100 most productive researchers in natural sciences, 76% in humanities, and 43% in social sciences. Unfortunately, the applicability of these findings to the evaluation of the potential impact of the rating system on research and knowledge production is limited by an internal philosophic constraint: for at least 7-10 years since the delinking of funds from rating, the rating system has been an instrument for an after-the-fact classification of the most productive researchers – Marais’s findings therefore constitute an external validation of the rating system’s internal efficiency, but do not causally link it to increased productivity. Another methodological constraint lies in the absence often of separate quantitative data for rated and unrated researchers.
as well as historic trends; as a consequence, the discussion of the relationship between productivity and the rating system is largely descriptive and time-bound, rather than analytic and historical, and unequivocal conclusions are not readily drawn.

10. The NRF rating system is unusual if not unique when compared to other research assessment systems

Is the NRF Rating system unique in the world of research evaluation? The answer is YES. With the exception of the Mexican system, which comes closest to the NRF, no other national funding council in any of the major science systems of the world has as its primary and final goal to rate (score) scientists. What is specifically unique to the South African system is that the NRF employs two different assessment systems: the standard project review process which is comparable to what most research councils worldwide do and the individual rating system (which is not).

It is important to point out that the NRF system is unusual in that it rates individual scientists as a goal in itself. Some systems rate individual scientists but as one phase in the calculation of a higher-level aggregate score of a subject area of university (the New Zealand PBRF System). Some other systems rate university departments again as a step in the calculation of a university rating (the UK RAE). But it is important to emphasise that in both these cases (New Zealand and UK systems), these systems constitute the respective government’s approaches to allocating block funding in higher education on a fair and just basis. In both systems, other bodies (the research councils in the UK) administer the more standard project-based assessment system.

What is thus truly unique about the NRF system is that it is the only funding agency that combines a dual system within itself: an individual rating system (that is currently delinked from the funding allocation) and a more standard project-based evaluation system (currently linked to the Focus Areas and Institutional Capacity Development Programme Frameworks).

1 The Pouris report, unfortunately, was of limited use to the compilers of this synthesis report. Instead of comparing the NRF Rating system with other rating systems, the report devoted most of its space to a discussion of peer review processes followed by agencies such as the NSF, Australian Research Council and British research councils. Only two systems that also utilise individual ratings – the New Zealand PBRF and the Mexican ISN – were discussed very briefly. This led Pouris to draw the rather strange and misleading conclusion that the NRF system is not unique because it also involves peer review! It also meant that the authors of this report had to consult and gather additional materials to draw conclusions on the international comparability of the NRF rating system.

2 The Mexican system (SNI) was established in 1984 evidently to reverse the brain drain of Mexican scientists due mainly to very poor academic salaries. Scientists who participate in the system are rated into one of three categories. On the basis of their rating, the individual scientist then receives a funding amount to augment their academic salary.
1. Background to the review and the five studies

Higher Education South Africa (HESA) and the National Research Foundation (NRF) have co-convened an in-depth review of the NRF evaluation and rating system of individual researchers in response to one of the recommendations of the NRF Institutional Review in 2004.

The Institutional Review, covering the period 1999 to 2004, was conducted at the request of the Department of Science and Technology (DST) by a review panel comprising experts from the United States of America, New Zealand and South Africa. The purpose of the review was, inter alia, to provide a retrospective view on the performance of the NRF during the first five years of its existence, an assessment of the outcomes and impact of its activities as well as recommendations regarding the strategic direction and operational execution of the NRF’s missions.

The review report includes various recommendations, one of which concerns the evaluation and rating of individual researchers, i.e., “[to review] the rating system, in terms of its fundamental purpose and utility…”. Researchers at South African higher education institutions, museums, science councils and approved institutions can apply to the NRF for evaluation and rating of their recent research activities and outputs.

The Review Steering Committee, responsible for planning, overseeing and guiding the process of this in-depth review of the rating system, called for proposals and on the basis of the submissions, commissioned the following studies:

- A document-based historical review and analysis of the rating system for individual researchers at South African higher education institutions and museums since its inception in 1984. This report was commissioned from Ms S Krige, and was numbered Report 2.1 in the original call for proposals – this is the nomenclature we will also utilise in this report, and we also refer to this as the “Krige Report”.

- A mapping of the formal and informal use of the rating system over time by various institutions, both research-performing and -funding. This report was commissioned from Ms C Lombard, and was numbered Report 2.2 in the original call for proposals – this is the nomenclature we will also utilise in this report, and we also refer to this as the “Lombard Report”.

- A report on the impact of the rating system within specific disciplines or fields of scholarship on the scholarly productivity of South African academics. This report was commissioned from Dr B Marais, and was numbered Report 2.3 in the original call for proposals – this is the nomenclature we will also utilise in this report, and we also refer to this as the “Marais Report”.


- A comparative international report on the rating of individuals for funding purposes. This report was commissioned from Prof A Pouris, and was numbered Report 2.4 in the original call for proposals – this is the nomenclature we will also utilise in this report, and we also refer to this as the “Pouris Report”.

- A review of the processes used to manage the rating system in the past five years. This report was commissioned from Dr M Madikizela, and was numbered Report 2.5 in the original call for proposals – this is the nomenclature we will also utilise in this report, and we also refer to this as the “Madikizela Report”.

In order to facilitate consideration of these reports by the Steering Committee, HESA and the NRF jointly commissioned this synthesis report, whose purpose it is to review the five separate commissioned reports, summarise and assess their findings, and present them in an accessible form to members of the Steering Committee.

2. **Methodological approach of this review**

Essentially this summary or synthesis report constitutes a desktop study of the five reviews mentioned above, incorporating for each review a methodological analysis, a summary of the key findings, and a critical evaluation of its content and recommendations. For all five studies, we report on and assess the following:

- the review’s brief and the questions it addressed;

- its approach and methodology; and

- the review’s key findings.

The individual reports outline and list numerous inputs, findings, and observations, on occasion also making recommendations. It seemed to us that simply editing the content of the reports would have defeated the aim of drafting a summary or synthesis report, and it was therefore necessary to make choices about what to include in this synthesis from the original reports. Naturally, those choices are subjective, but we endeavoured to lift out those aspects we considered most directly relevant to the task of the Review Steering Committee.

Lastly, where the reports have suggested possible oversights or extensions of the current work or specific areas requiring attention, we have merely listed and briefly outlined them, seeking to avoid recommendations on our behalf in order not to dilute the role and function of the Steering Committee for this review of the rating system.
In compiling this report, we assume a certain level of operational familiarity on behalf of readers with the NRF rating and evaluation system, and for this reason do not include generally-known aspects of the system even where they are addressed in the individual reports. For instance, we do not capture the full outline of the specific processes involved within the Evaluation Centre (EC; the subject of report 2.5) because we assume readers to know about the fundamental process involving assessors, review panels and appeals, as well as the operational processes whereby applications are submitted.

In citing specific parts of the five reports we list the page for the reference, followed by a capital “R” and the number of the corresponding report as listed in the original NRF/HESA specifications for this overall review of the rating system: hence “p. 6, R2.4” refers to page number 6 in the report number 2.4 (“A comparative analysis of the use of ratings systems” by A Pouris). In instances where we cite directly original sources (rather than the author of a specific report), the cited text is italicised and the reference, naturally, to the original source.

There are several instances where the original reports are very concisely and clearly formulated around specific critical points that we felt needed to be highlighted in this summary; in such cases, instead of seeking to paraphrase what is already very elegantly put, we have allowed ourselves to simply lift the sentences or paragraph from the relevant report, always citing where in the report it came from, but not always enclosing the entire excerpt in quotes – the latter custom was followed for shorter excerpts.

Although the formal brief provided for this synthesis report encouraged interaction between its authors and those of the five reports reviewed here, time constraints were such that we were unable to pursue this option. Consequently, there may be a few references in this document to conclusions or ambiguities arising from apparent omissions of data or information, which could have been avoided given the opportunity for brief communications between us and the respective authors.

Lastly, the five reports under review here differ from each other in several substantive ways; e.g., Reports 2.1 and 2.5, respectively, dealing with the history of and the processes underpinning the rating system differ considerably in their scope, their subject material, and their methodologies – leaving aside stylistic differences that might also exist between the respective authors. These differences constrain the extent to which our review of the five reports can be stylistically uniform.
3. **Analysis of individual reports**

3.1 **Summary of and comments on Report 2.1 (An historical review and analysis of the NRF rating system: 1983-2005; S. Krige)**

3.1.1 **The brief and questions addressed**

THE BRIEF: To undertake a document-based historical review and analysis of the rating system for individual researchers at South African higher education institutions and museums since its inception in 1984. The analysis should review its introduction and implementation, reflecting on issues related to its fundamental purpose and utility over the years. The review should include an analysis of the findings of the previous reviews on the rating system and reflect on the implementation of the recommendations made in the reviews.

The report is divided into two parts. The first part of this report reviews the genesis of the rating system between 1980 and 1987 and addresses the following questions:

- What were the underlying principles, form and purpose of the system?
- What were the international, national and institutional contexts and forces which informed the development of the system?

The second part of the review continues to contextualise the implementation of the system, and addresses the following questions:

- What did the reviews/evaluations of the system say about the implementation and “utility” of the system?
- Did their evaluations focus on the implementation of, rather than the principles underlying, the system itself?
- What changes and continuities can be charted in the rating system?
- How responsive was the Foundation for Research Development (FRD)/National Research Foundation (NRF) to review and evaluation?
- To what extent were changes and continuities responses, both positively and negatively, to the observations and recommendations of the formal reviews undertaken?
- What other factors may have been involved?
3.1.2 Approach and methodology

The Krige Report is essentially a historical study and employs standard historical methods. The bulk of the research involved analysis of primary and secondary materials including materials provided by the Evaluation Centre of the NRF. In addition, a small number of personal interviews with key informants were conducted and reported as personal communications. Furthermore, a pre-final draft of the final report was circulated to a number of key stakeholders for comments which were subsequently incorporated in the final version.

The report is, as one would expect from a historical review of the NRF rating system, essentially descriptive, narrative, rich in detail and well-documented. The organisation of the report is clear and logical.

The report has set itself the aim to cover the following themes and topics:

- To provide the necessary context of the genesis and original conceptualisation of the rating system
- To describe in some detail the original thinking by the main actors in the process (Garbers, Arndt, De Wet) including the philosophy underpinning the inception of the rating system, the stated aims of the new system and the main features thereof.
- To describe the main changes that have been made over the period 1985 – 2000. These changes affected most aspects of the system including the eligibility criteria, the rating categories and their descriptors, the peer review process (use of local and international referees), the assessment and appeals process, and the matter of providing feedback to participants in the process.
- To discuss the internal and external review studies that have been undertaken since 1985 as well as the FRD/NRF’s responses to these studies.
- To comment on the correctness and appropriateness of criticism levelled at the system in these reports as well as of the NRF responses to such criticism.

The Krige Report is in general well-researched and the findings are appropriate and mostly well-substantiated. We would list two criticisms of the report – however, neither of these would necessarily have impacted negatively on the main findings presented in the report:

- There is very little mention made of broader science policy developments in the early periods. The lack of reference to the link between the NRF (and the rating system) and major events in the science policy domain, especially in the 1980s and early 1990s is...
somewhat surprising. It would have been useful to have seen references to the role of the National Department of Education and the introduction of the 1986 subsidy system, the relevance (if any) of the Scientific Advisory Council as well as the influence of the findings of the IDRC Commission on Science (1983). There is also no discussion of the debates that were conducted with the formation of the NRF in 1999 and the possible inclusion of the Agency Funding of the MRC at that time. It was widely expected at the time – also because it was foreseen in the White Paper on Science and Technology in 1996 – that the health scientists would be included in the agency role of the NRF. This would have had the consequence that health scientists would also have become part of the rating system and thus created a more inclusive system. Their current exclusion from the rating system is a continuous point of contestation and has major implications at the institutional level.

- The discussion of the unintended consequences of the rating system and the way in which institutions have responded to the rating system overlaps with the much more detailed brief of another report in this study (Lombard). We have, therefore, not incorporated any part of this rather brief section in the historical review of the system as it is covered in detail under a separate heading.

Our overall assessment of this report is positive and we would suggest that consideration be given to publishing an edited version of this report in its own right.

### 3.1.3 Key findings and conclusions of report

The key findings of this historical report on the NRF rating system will be summarised and discussed under the following headings:

- The origins and rationale for the introduction of the system
- The core philosophy that underpinned the introduction and initial design of the system
- The main phases and shifts in the implementation of the system
- Key issues that emerged from various reviews of the system of the years

#### 3.1.3.1 Origins and rationale for the introduction of the system

The history of institutionalised government funding for research and scholarship at South African universities is a fairly recent event as is internationally the case as well. The National Science Foundation in the USA was only established as a national funding agency in 1951 following on recommendations contained in the very influential *Science: The Endless frontier* report by
Van nevar Bush in 1950. In South Africa the CSIR, which was established in 1945, was tasked with university funding in 1952.

Given the socio-political history of South Africa, it is not surprising that funding for research during the period 1960 to 1980 was dominated by strategic research in service of the nationalist government. It is now well documented that under the apartheid government, R&D expenditure was very much focused on national defence, energy and other strategic areas. As the Krige Report shows, these factors affected the work of the CSIR directly as it increasingly devoted its efforts and funding to working with industry and the national defence industry to solve immediate applied problems. It also led to increased collaboration – in the area of strategic and applied research-between the science councils, industry and the universities. This is evidenced, amongst other things, by the establishment of the Cooperative Scientific Programmes (CSP) in 1975.

One of the results of the increasing collaboration between the CSIR and Industry as well as the increasing isolation of South African science because of international sanctions and boycotts was a deliberate shift towards more market-driven principles. This is well captured in the following extensive quote from the Krige Report:

By May 1980, when Dr Chris Garbers took over the reins from Dr Brink, difficult external factors made the need for change more apparent. These included increasing international isolation, persistent inflation, worldwide recession, the aftermath of the energy crisis, the weakening price of gold and the onset of severe drought. After 1976, ongoing internal resistance to apartheid and increasing South African military and other involvement in cross border incursions and wars made for political stability and violence. In 1983, the country experienced its first State of Emergency since 1960….Over the next seven years, the CSIR itself went through a process of major overhaul, to become a market-driven organisation. Significant influences were the 1979 findings of the Riekert Commission and Wiehahn Commission (which represented attempts to reform the labour-related aspects of the apartheid system) and the 1983 Klue Report which preceded the White Paper on an Industrial Development Strategy (1985) (p. 11, R2.1).

The upshot of all of these forces was that funding and support for self-initiated, basic research suffered. At the same time the country also experienced an increasing loss of some of our best scientists and scholars because of the deteriorating political situation. In a personal communication by the then President of the CSIR, as cited in the report Chris Garbers indicated that “while he had been concerned about the inadequacy of funding for university and museum-based research, his particular focus had been the position of “outstanding researchers” and the paucity of postgraduate bursaries to nurture brilliant young students (p. 13, R 2.1).

To put the matter of national research funding on a more sound footing, Dr Garbers and Dr Arndt (Deputy President) subsequently co-opted Prof Jack de Wet, retired Dean of Science at the
University of Cape Town (UCT), to investigate options for research funding in higher education and to advise the CSIR in indicating the role the Research Grants Division (RGD) could play to satisfy these research funding needs. Prof De Wet commenced his work on a new funding mechanism of university research in early 1983 and conducted a wide range of consultations with national stakeholders and representatives from universities. On the 8th of April 1983 a draft - *A new look at the role of the Research Grants Division in the promotion of research in the South African Universities* - was sent to universities for comment. A final document was submitted to the Executive Committee of the CSIR Council on 9 May 1983.

The CSIR Council accepted in principle the four major elements of the report:

- awards should be based on the merits of the applicant as a proven research worker;
- the RGD should be responsible for the payment of full costs of research programmes of selected research workers;
- research units currently supported by the RGD should continue to receive support under the new scheme; and
- capital equipment requirements of university researchers should receive urgent attention.

(p. 14, R2.1)

In May and June of 1983 Prof De Wet and Dr Arndt visited western Europe and the USA and Canada to investigate university research funding systems. Two months later, in August 1983, Prof De Wet submitted a foundational report: *A framework for the promotion of free research in universities and an integrated research effort in the areas of national concern between the universities CSIR institutes and other bodies* - attached to which was *An addendum to a framework etc.*

### 3.1.3.2 The core philosophy that underpinned the introduction and initial design of the system

[S]elf-initiated or curiosity research is what people do to satisfy their own curiosity. It is what turns them on and for good people to do what turn them on is both highly satisfying for the researcher and often the most productive use of a good man’s (sic) time. What they do often seems quite useless in the eyes of others until, often very much later, the seminal nature of their work becomes abundantly clear to everybody. The moral is to support and encourage the good researcher to do his own thing.

(p. 16, R2.1)

Prof De Wet’s ideas on the autonomy of the individual “good researcher” as quoted above are reminiscent of Vannevar Bush’s “endless frontier” style of research but were not necessarily congruent with the new priorities outlined by the CSIR in the early 1980s. However, as the Krige
Report shows, he argued that this autonomy in both choice and methodology would establish world class researchers who could and would meet the needs of the national interest, under the umbrella of one autonomous research funding body.

The core philosophy of De Wet’s approach to the new rating system which underpins his whole approach is captured well in the following quote: “A good researcher is one who is evaluated by his (sic) peers, not on proposed projects but on proven ability”, on a selective non-uniform basis. (p. 16, R2.1)

This emphasis on past performance and recognition of established excellence is the key pillar on which the new rating system would be built. In the mid-eighties the FRD’s concerns were clearly primarily directed at rewarding proven research excellence rather than recognising future potential and capacity. At the point of its inception, it is quite clear then that this was a system that was designed primarily for the top (elite) scientists in the natural and engineering sciences. The discourse on research capacity building would only enter into these debates much later.

However, some provision was made for “brilliant young researchers” in the first versions of the system. De Wet proposed at the time that

> [t]he new system will therefore be structured in such a way as to make for provision from RGD for the needs of (i) the proven researchers in the universities of high quality and (ii) brilliant young members of University staff [who are] not yet proven researchers but who, on the basis of their academic records, should be given every opportunity for a limited period to become active and productive researchers. The researchers, not quite of the high quality envisaged in (i) above, will get basic support from RGD and for funding of particular projects, but with the expectation that other projects would be funded by the University. The University will have the responsibility to those members of the University staff who are not provided for in the above. (p. 17, R2.1)

From the above it is quite clear that the rationale behind the proposed rating system found its justification in a philosophy that prioritised the individual scientist and scholar. The aim would be to recognise and reward the individual scientist and in doing this, strengthen the basic science base of the country and (hopefully) reverse the increasing brain drain of top scientists. Recognition of scientific excellence would be based on an assessment of the past performance of the individual scientist. Based on this assessment, which would be quantified in a rating score, a monetary reward commensurate with the rating would then be disbursed to the individual.

Two main features of the original intentions of the rating system, therefore, need to be highlighted and kept in mind in any interpretation of its later history:
First, recognition and reward was intimately linked to each other in the original conceptualisation and design of the system: reward to the individual scientist was automatically linked to the recognition of degree of scientific excellence.

Second, the system was developed to address two very different but related systemic challenges: the lack of adequate support for fundamental self-initiated research and the increasing loss of high-level human capital in the country at the time. It follows that any assessment of the future role or nature of the rating system has to take into consideration the extent to which these conditions still apply to our current science system or whether new demands and challenges need to be addressed.

3.1.3.3 The inception and first years of the rating system

The very first version of the new system as proposed by Prof De Wet comprised a fairly simple classification based on three categories:

Table 1: Categories identified by De Wet

<table>
<thead>
<tr>
<th>Category</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Researchers who by the very highest international standards have established for themselves recognised positions in the international community of workers in the fields as one of themselves – a member of the club</td>
</tr>
<tr>
<td>B</td>
<td>Researchers of considerable distinction around whom a team of workers and graduate students in training could be built, able to make a significant contribution to research in South Africa</td>
</tr>
<tr>
<td>C</td>
<td>Proven researchers who had not made it into the above categories, including “younger men (sic) whose talents have not yet been developed to the full”</td>
</tr>
</tbody>
</table>

In addition De Wet also put forward a scheme intended for young members of staff who were holders of PhDs of no more than three years’ standing; who had excellent academic records and had started doing research but had not yet become fully established as independent researchers. This category later became the Y category (denoting “Young” researchers).

The other features of the initial version of the system were the following:

- **Linked funding**: A funding package allocated to the rated scientists to provide them with enough support for an initial period of five years.

- **Self-directed research**: The candidate’s research programme was his/her own free choice without any regard to content or focus.
Differentiated funding values: A differentiation in funding values with the expectation that researchers in Category C would not receive anything more than basic support for their research.

Possibility of renewal: Rigorous assessment after five years of each rated scientist could result in discontinuation or continuation of funding and even improvement of rating category.

But the rating system needed an “institutional vehicle”. De Wet’s influence on the direction of national policy is evidenced by the fact that his arguments for one research funding agency eventually carried the day and in 1984 the Main Research Support Programme (MRSP) and the CSP were combined to form the CSIR Foundation for Research Development (FRD), headed by Dr Reinhardt Arndt.

The main aims of the MRSP were as follows:

[T]o promote “own choice” research at universities and museums as well as the training of research students. Implicit in a support programme in which research grants will be allocated solely on the merit of applicants as researchers, is the absence of favourite fields of research. Researchers are free to choose their own fields of research and to pursue research in the way they think best. The MRSP aims at the support of “own choice”, in the belief that the best climate for good research is one in which researchers decide on their own research preferences. (p. 23, R2.1)

With the official establishment of the FRD the rating system could be implemented and by August 1984, four months after the distribution of the revised RGD Awards Guide in April, close on 900 evaluation statements had been received. It seems that evaluation statements had begun to come in as early as February 1984. This would indicate that institutions had been thoroughly canvassed and appraised of the draft guidelines developed by de Wet, and felt confident enough to submit applications prior to the publication of the RGD Awards Guide.

Between 1984 and 1987 various documents saw the light which informed the first years of the implementation of the system. These included the following internal reports:

First Research Grants Division Awards Guide (January 1984)


A new look at the Foundation for Research Development (May 1986)

The Krige Report documents in great detail the shifts and changes in the rating system as captured in these various documents. During this period no major shifts in the basic philosophy and assumptions occurred. But various incremental and smaller changes were made to the classification system. By 1985 already new categories A to E and Y were added. As the Krige Report indicates, there are no indications as to how these descriptions were arrived at. By 1987 the following classification system was being used:

**Table 2: Rating classification system**

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A</strong></td>
<td>Researchers who without any doubt accepted by the international community as being amongst the leaders in their fields</td>
</tr>
<tr>
<td><strong>B</strong></td>
<td>Researchers not in category a but who nonetheless enjoyed considerable international recognition as independent researchers of high-quality</td>
</tr>
<tr>
<td><strong>C</strong></td>
<td>Proven researchers who have maintained a constant high level of research productivity and whose work is regularly made known internationally, or proven researchers whose current research output is less but who are actively engaged in scholastic activities</td>
</tr>
<tr>
<td><strong>D</strong></td>
<td>Researchers who show promise of achieving the status of proven researchers in the foreseeable future, in terms of an already existing track record, or proven researchers with an established track record who are currently less productive</td>
</tr>
<tr>
<td><strong>E</strong></td>
<td>Candidates who, according to a recent track record, are not likely to prove themselves or re-establish themselves as researchers</td>
</tr>
<tr>
<td><strong>Y</strong></td>
<td>Young researchers normally less than 35 years of age who are highly likely to achieve C status by the end of the support period</td>
</tr>
<tr>
<td><strong>P</strong></td>
<td>Research is younger than 35 who have already obtained a PhD degrees and who have shown exceptional potential as researchers</td>
</tr>
<tr>
<td><strong>Z</strong></td>
<td>Candidates who submission for evaluation is premature</td>
</tr>
<tr>
<td><strong>X</strong></td>
<td>Candidates who cannot be evaluated as researchers</td>
</tr>
</tbody>
</table>

According to the Krige Report, “the changes in description and numerical assignations to the categories were probably the result of internal non-public discussions at all levels of the process, but particularly at the Award Committee level. This refining of the wording and scoring of the system has taken place on an almost annual basis. Without an in-depth study of committee documents and interviews, it is difficult to establish the reasons”. (p. 30, R2.1)

**3.1.3.4 Key issues that emerged from the Hawkins and MacKenzie reviews of the system**

In its review of the period after 1987, the Krige Report discusses at length various issues, concerns and challenges that emerged as the system evolved, demands and expectations in the external environment changed and reviews pointed to matters that required resolution. Some of these issues appeared and re-appeared in two extensive external reviews that were published in 1991 and 1996 respectively:


Despite the difficulties and problems raised by the two reports and which are discussed below, it is important to point out that the Krige Report indicates that both these reviews in essence reviewed the rating system positively.

The Krige Report writes the following on the Hawkins review:

> Both Hawkins’ sources of information lead to the conclusion that the system had been successful: that it was a substantial improvement on the previous one, and the system of recognizing excellent performance had created a positive research climate and generally good morale amongst academic researchers. This had led to substantial increases in all types of research output. (p. 41, R2.1)

And its summary of the main conclusions of the MacKenzie review highlights the following positive conclusions:

> The MRSP and Core Programmes stimulated and expanded research, cultivated research ethos where there had been none, created centres of excellence, trained highly skilled students and catapulted South African science into the international arena. It must be acknowledged that much of this was due to the evaluation principle which created an element of healthy competition and identified a hierarchy of researchers in whom the FRD placed its trust through the flexibility and freedom in the operation of its rolling support grants system. (p. 53, R2.1)

We have summarised the main findings and conclusions of the Krige Report with regard to these two reports thematically rather than chronologically as many of the same issues were raised repeatedly. We have identified seven major themes:

- The emphasis on quality rather than quantity of output.

- The meaning of the rating descriptors.

- The nature of knowledge production.

- Allocation of funds between different rated classes.

- Funding of scientific equipment.

- Problems experienced with referees and composition of the subject committees.
The necessity for transparency and continuous feedback.

The emphasis on quality rather than quantity of output

A recurring theme in many reports was a concern that researchers would not recognise that an emphasis on quality in journal selection is much more important than mere volume of output. This issue was already raised in an internal 1987 report which emphasised that the assessment of the research calibre of an applicant was not primarily based on publication count, but rather on research achievement as seen in the actual content of papers reflecting innovation and sustained activity.

In the first major external review by the mathematician and statistician, Professor Hawkins (1991), he pointed to “the imperfections in the measurement of output” in relation to tangible outputs such as publications, conference papers and graduate student production. He was particularly concerned not to equate quantity with quality. He went into a considerable amount of detail about the statistical methods by which he came to the conclusion that published papers had increased in both quantity and quality. He cautioned that

... correlation is not causality. A demonstrated increase in output could not be taken as definitive proof of the effect of the change. An important factor here was the change in University subsidy formula including an explicit component for research publications. The simultaneous timing of these changes made it effectively impossible to quantify the separate effects of these changes. (p. 42-43, R2.1)

Hawkins noted difficulties with so-called “publication money” from the Department of Education subsidy, which were not the result of the rating system but might impact upon it. He noted that the formula took little account of the quality of the paper and of the different difficulties of publishing and paper lengths typical of different disciplines; and that there was very little consistency across institutions in the way publication money was spent.

The meaning of the rating descriptors

It is rather surprising that the meanings of the descriptors of the rating criteria did not receive much attention in the various reviews of the NRF rating system. Given that so much depends on the ability of different reviewers and panels interpret and understand the meaning of an A or B in a similar manner, it is interesting that this matter was only explicitly raised in the Hawkins Report (1991). Hawkins raised this issue in the context of briefing peer reviewers, but it is clear that his scepticism is related to a more general approach and philosophy that requires more objective and quantifiable means of rating individuals. He writes:
What exactly do we mean by the descriptive wording used to assign applicants to the A, B or C or no rating? What does a reviewer really mean by saying that Dr Bloggs is known internationally for his work? One possibility would be to ask for a quantification: at what percentile of post-PhD workers around the world would you put Dr Bloggs’ for research novelty? (p. 45, R2.1)

It is a pity that not more attention has been given in the various external reviews to this central issue of the meaning of key descriptors in the rating system. As the Krige Report indicates, the FRD at the time unfortunately also did not engage with the core of the Hawkins Report at all:

In terms of the evidence available from the documents analysed, it does not seem that the FRD engaged with the substantive and sophisticated criticisms of the rating system which Hawkins explored at length. And, perhaps more importantly, the FRD did not seem to engage with the alternatives he proposed. (p. 51, R2.1)

The lack of engagement by the FRD and its external reviewers over the years to the obvious challenge of achieving clarity in the meaning of key words such as “international standing” or “proven track record” or “considerable international recognition” is a weakness that needs to be urgently addressed in the future. It also seems that there was no recognition of the fact that these words are not only understood differently by different people in the same field, but that different fields of science would describe and understand them very differently. With the expansion of the system to the social sciences and humanities in 2003 these issues would be foregrounded again.

**The nature of knowledge production**

The peculiar demands of those who undertake research of an applied nature and who work closely with industry and who invariably publish much fewer articles in peer reviewed journals have raised debate from the earlier reports onwards. It has been argued in a number of reports that those who participate in contract work where they are involved in technology transfer – like engineers – are to some extent penalised by the rating system and its emphasis on international journal articles.

In his 1991 report Hawkins already noted that academic engineers do most of their creative work in the form of consulting which is written up as internal technical notes. This could not be accommodated by the rating system with its emphasis on international quality of work. Younger engineers seemed to be happy to publish in refereed literature; there remained “a lost generation of older academics” who had little record of publication in refereed media. He concluded that engineers did not feel their interests were being sufficiently represented by the FRD. (p. 48, R2.1)

The MacKenzie review (1996) also referred to the issue and according to the Krige Report indicated that “with regard to scientists and engineers, the system did appear to favour scientists over engineers. The large disparity in ratings between engineers and scientists warranted more careful analysis”. The MacKenzie Report also noted that
the Core Programme promoted effort in subjects which would result in research publications rather than in the general needs of the engineering community; and

- since engineers tended to work as teams together, it was difficult to identify internationally recognised engineers as individual contributions could not be easily extracted.

It could be argued that the issue of differences in modes of research and knowledge transfer that are illustrated by the discussion around scientists and engineers, have only become more prominent since the mid-nineties when these sentiments were expressed as cited above. Since that time, Mode 2 forms of knowledge production (Gibbons et al.) have undoubtedly increased its share in university research. The exponential increase in funding of more industry-type applied research through THRIP and the Innovation Fund and the general increase in third stream monies to the universities has meant that huge numbers of engineers and applied scientists in the life, agricultural and other sciences have turned their efforts and energy away from fundamental and basic research. With the inclusion of the social sciences and humanities in the rating system, yet other modes of knowledge production require recognition.

**Allocation of funds between different rating categories**

The Krige Report has documented very clearly how the principle of “funding differentiation” was an integral component of the rating system since its inception. In 1991, Hawking identified a persistent problem – the underfunding of researchers below B level. He argued that funding for C and Y was far below what was needed for meaningful research. Since C researchers were the backbone of the whole system it was “myopic to short-change them”. The funding was unfairly low, given that the per capita output of C researchers was nowhere as small a ratio as that of funding for As and Bs. He noted that Cs’ outputs only differed from As’ and Bs’ in foreign keynote conferences, while they had the same postgraduate outputs. They needed a level of about half or more that of Bs. (p. 44, R2.1)

The MacKenzie Report of 1996 also urged the FRD to reconsider the C rating, as it had demonstrated the cost effective nature of C and Y ratings. The Report acknowledged that many interacting factors and intangibles were involved in assessing productivity/ cost effectiveness – and the review committee decided to attempt to measure output per FRD Rand input. They found that C rated researchers were most productive, whilst P rated researchers were least productive, “with Y counterparts outperforming them ‘in each costed output’…. The ‘raw data’ argue in favour of a reduction of the differentials in funding between the categories”. (p. 55, R2.1)

The sharp differentials between C and A and B levels of funding were source of grievance and frustration, but the review argued that “The raw data cost-effective analysis indicates that the smoothing of these differentials may be carried out without any significant deterioration of the
quality of the research”. The funding to P grant holders came under the spotlight for the first time – with the review deeming it to be excessive.

In fact, the MacKenzie review encouraged the FRD to consider abolishing the C rating. Indeed, it wondered whether the “new Open and Directed Programmes [presented] an opportunity to replace the present complex evaluation process by a simpler ‘accreditation scheme’ which evaluated whether an individual was a proven/able researcher or not”. The levels of funding of the proven researcher and/or research team would be determined by the “quality, feasibility and relevance of the proposal” (emphasis added). (p. 55, R2.1)

The issue of differentiated funding became irrelevant when the “automatic” linking of rating to funding was abolished in 2000. However, should the NRF consider continuing with the system and re-instate a linking to funding, this matter would again require urgent attention.

**Funding of scientific equipment**

The issue of funding for running expenses, especially equipment, was raised by Hawkins in 1991 who argued that all researchers spent the same amount of money on items of relevant equipment. The original argument was that C rated researchers would borrow equipment from A and B rated researchers with whom they were working, and that this would encourage collaborative work, especially among C and Y rated researchers. However many of these researchers worked in environments where there were no A and B rated researchers, and so had no equipment.

Hawkins recommended that capital and running costs be assessed separately – and that there should be a willingness to provide more expensive capital items to C and Y rated researchers. He favoured “some further flattening of the curve in relation to running expenses”, using new funds, not reducing those available to A and B researchers – but also cautioned that the FRD should avoid duplication particularly in expenditure on expensive equipment for A and B rated researchers. (p. 44, R2.1)

Although not discussed in any great detail, the problem of funding research equipment reared its head again in the MacKenzie Report of 1996 where the case was made for a more careful analysis of the cost of theoretical and experimental research as experimental research required a good deal more funding in terms of equipment and running expenses. (p. 57, R2.1)

**Problems experienced with referees and composition of the subject committees**

Given the centrality of the peer review mechanism to the overall rating system, it is not surprising that the quality of peer reviewers (the referees) would surface in the very first external review in 1991. Hawkins identified a number of problems related to referees’ reports: superficial analysis;
overtly critical statements or fulsome praise; delayed responses due to temporary absences; and refusal to participate because of claimed incompetence in judging specialist contributions. (p. 32, R2.1)

It is also clear from the Krige Report that the FRD had recognised this as a potential problem quite early on and had in fact started in 1987 already to rate its referees/reviewers. In 1987 the following guide for selection committees for rating the reviewers was produced:

| 1a | Good |
| 2a | Satisfactory |
| 3a | Poor |
| a  | indicates appropriate reviewer |
| i  | indicates inappropriate reviewer |

In 1990 assessment categories for reviewers’ reports were changed, suggesting the beginning of the unexpected challenges which the selection committees faced in utilising peer reports. The guidelines were as follows:

| 1  | Excellent |
| 2  | Satisfactory |
| 3  | Unsatisfactory |
| a  | Appropriate |
| i  | Inappropriate |

The following could apply for an “inappropriate” reviewer:

- that a referee is not active in the field of research of the applicant
- that a referee has a less favourable FRD status than the applicant in the particular research field
- that a referee’s report is not meaningful, too praiseworthy or contains no factual information regarding the applicant as a scientist.

Reviewer/referee reports seemed to vary significantly – ranging from issues such as how differently or similarly local or international peers understood the rating categories, to lack of appropriate input often linked to qualifications, or “simple lack of application to the task at hand”. The problem with peer review reports increased the amount of time and administration needed to evaluate applicants, as well as causing delays with rating outcomes.

Hawkins made two suggestions to deal with these issues:
• Use fewer foreign reviewers, with each reviewer rating more applicants so that the more reliable relativities of the candidates rather than absolute assessments could be obtained.

• A set of associate reviewers be created, each handling assessment of applicants in their particular area. This could operate mainly by correspondence, and so the people involved would not need to be in South Africa. This will build up a base of reliable reviewers. (p. 47, R2.1)

A related problem that speaks to the heart of the peer review process is the composition of the subject committees. Again the Hawkins Report seems to have addressed this issue in more detail that other reports. Hawkins felt that the subject committees were seen as exclusive and not representative of all institutions involved. He made the following points.

• There was a need for impartial informed subject committees in some areas, with some foreign members. Foreign members could be involved through correspondence. This would help to avoid perceptions of subject committees as exclusive cliques.

• The FRD should, without sacrificing academic qualification, take some note of the spread of members, in terms of institutional, geographical and subject specialty. Conscious effort should be made to locate potential members in universities and museums that have not recently been represented on subject committees. (p. 47, R2.1)

The necessity for transparency and continuous feedback

Various reviews pointed to problems related to (lack of adequate) feedback to applicants over the years. In evaluation terminology, this is the demand of the formative intent of all assessment system. As the Krige Report correctly emphasises, any assessment or evaluation system should be geared to a long-term relationship with people who will be applying and reapplying for funds. And Hawkins makes the important point that the need for written summaries of reviewers’ and committees’ perceptions of the applicants’ relatively strong and weak areas is essential as part of the overall goal of improving the science system. Evaluations and assessments which are only summative in nature (judgemental of quality and outcome) and do not also include a formative element (the need to improve and develop the system further) and doomed to failure or apathy in the long run.

In this regard, Hawkins described “the implementation of assessment procedures as unnecessarily secretive and mysterious; that there was a ‘deficiency in information flow’. There was a need for a plain focused and matter-of-fact inexpensively produced newsletter for researchers monthly or bimonthly, which could include background stories including sketches of members of subject committees”. (p. 46, R2.1)
The issue of secrecy and transparency was also taken up in the MacKenzie Report in 1996 who noted the following problems with regard to transparency and feedback:

- a lack of transparency with regard to criteria for full funding allocations;
- absence of clear guidelines as to favourable/unfavourable factors for evaluation; and
- inadequate feedback on evaluation. (p. 54, R2.1)

### 3.1.3.5 Facing the Challenge of a New Political Dispensation 1995-2000

The Krige Report chose, correctly in our view, to discuss the further evolution of the rating system after 1994 in a separate section. The report clearly identifies a number of significant events in the mid-1990s which would ultimately also impact on the rating system itself. These included:

- The new political dispensation that came about with the democratic elections in 1994.
- The new science policy context in the country (with the publication of the White paper on S&T in 1996); and then later.
- The establishment of the National Research Foundation (out of the former CSD and FRD) in 1999.

Our approach again is to summarise the main developments as well as issues and challenges faced by the rating system thematically rather than chronologically. The Krige Report highlights at least six themes:

- The impact of the new political dispensation on the rating system.
- The introduction of a “dual” assessment element through the linking of individual rating with project evaluation and the gradual delinking of funding.
- The increasing steering of research through the introduction of directed themes and eventually focus areas (1996 – 2000).
- The increasing bureaucratisation of the system.
- The inclusion of the social sciences and humanities in a unified rating system in 2003.
- A growing disillusionment with the system amongst certain academic and scholarly circles.

Two key reports figured during this period. These are:
Review of the NRF Rating System: Synthesis Report

- **Evaluation of the programmes and activities of the former CSD and FRD.** The report of the Review Panel aimed at stimulating and informing the discussions regarding the way forward for the NRF. April 2000.


1. The impact of the new political dispensation on the rating system

The new political dispensation would ultimately impact on various aspects of the science system and this would be no different for the NRF. In general it would require the NRF to ensure that it meets the demands and needs of historically disadvantaged scientists and researchers. In addition to historical concerns with excellence and proven performance, new demands relating to capacity building and redress were now firmly on the agenda.

It is then not surprising that the FRD in 1995 acknowledged “that experts and expertise were drawn from only one racial group”. In a new strategic document (*Facing the Challenge, 1995*), the FRD committed itself to giving “considerable attention to activities aimed at negating the legacy of racial discrimination in SET research, education and training”, also highlighting the importance of increasing access for women. “Corrective action” rather than “affirmative action” was named as the strategy for negating “the consequences of past policies that emphasised university education and training at the expense of technical, technological and vocational education and training”. (p. 60, R2.1)

One of the first practical consequences of this “new” philosophy was the introduction in 1995 of the L category to “accommodate those researchers who experienced impediments that prevented them from establishing themselves as researchers who could produce outputs of the quality, quantity and regularity required for individual peer evaluation”. (p. 62, R2.1)

The SETI Review Report of the CSD and FRD (1997) also spoke directly to the issue of how these agencies were addressing the historical imbalances of the past and emphasised that the historically disadvantaged institutions had not benefited sufficiently from the previous system. In particular the report noted that

> though patterns of research support have been changing to recognise redress and equity, there continues to be a chronic under supply of black grant holders, black bursary holders at the higher postgraduate levels, research grants located in HDIs [historically disadvantaged institutions] and female grant holders. The need to improve the distribution of access, opportunity and participation has to be reconciled with the mission to fund research of the highest quality, which is presently based at established research institutions. (p. 69, R2.1)
2. The introduction of a “dual” assessment element through the linking of individual rating with project evaluation and the gradual delinking of funding

In the 2005 Institutional Review of the NRF the panel writes as follows:

On the question of whether to re-instate the link between rating and funding, opinion was divided. Some supported the idea, as a way of rejuvenating interest among researchers in being rated while others opposed it, on the grounds that re-instating the link would undermine the NRF’s efforts to spread research funding more widely across historically disadvantaged groups and to younger scholars who had not yet become established researchers. (p. 85, R2.1)

The Krige Report does not demonstrate unequivocally that the (gradual) delinking of the funding component from the rating was in fact primarily to due considerations of greater inclusiveness of scientists and scholars. In fact the SETI Review Report (1997) pointed out that the amounts of money available for funding individual scientists (especially in the C categories) was so inadequate that it was defeating the purpose of the system as a whole.

By 2001 the link between rating and funding has been completed severed. This period also coincide with the introduction of the Focus Area Programmes and the new philosophy of the NRF to increasingly steer research in the national system of innovation.

Another significant development occurred during this same period (1995-1996): the additional requirement that a research plan be included as part of the evaluation process. This is clearly illustrated in the details of the assessment tool (the Multi-Criteria Decision-Making Tool) in 1996 and in the 1997 Guide to Research Support in which

the entry requirements to qualify for FRD funding in the Open Research Programme and programmes within most of the Directed Themes are:

- a valid rating in one of the categories A, B, C, P, Y or L; and

- acceptance of a research plan that outlines the proposed research and the anticipated outcomes of the research within a funding cycle. (p. 63, R2.1)

As the Krige Report notes:

All research plans would be assessed in accordance with the criteria of the MCDM and the scores allocated to research plans would be used to prioritise funding. By implication, “a rated researcher is not guaranteed access to funding if, in the assessment process, his/her research plans is not accorded a high priority”. (p. 63-4, R2.1)
In summary: As aptly commented by the 2005 Institutional Review Report, with the delinking of funding and ratings, “the rating status is largely honorific, with the one important exception that eligibility for long-term (5 year) grants from the NRF is limited to rated scientists after an initial maximum of three two-year grants”. (p. 84, R2.1)

3. The increasing steering of research through the introduction of directed themes and eventually focus areas (1996 - 2000)

It was inevitable, after 1994, that the new government would wish to steer the science system more actively and deliberately in alignment with the national goals as expressed in various documents (RDF, GEAR). This is explicitly stated in the White Paper on Science and Technology in 1996 and even more clearly in the National Research and Development Strategy of 2002.

Already in 1996, the FRD inaugurated a new system that would allow “for the integration and synchronisation of activities designed to address the corporate goals”. This focused on three core categories:

- competitive research in SET (Science, Engineering and Technology)
- corrective actions
- academic-industry cooperative research

These core categories not only underpinned the new research programmes of the FRD but also had to be addressed in any documentation submitted as part of the application process. (p. 60, R2.1)

This initial trend towards more directed funding would culminate in 2002 with the implementation of the Focus Areas Programme. As the Krige Report states: “The Focus Area Programmes represent an important facet of the NRF approach to the steering of the research system in the direction of its missions”. (p. 72, R2.1)

4. The increasing “bureaucratisation” of the system

In various reports from 1997 onwards comments emerged on the increasing bureaucratisation of the system. More specifically, it was pointed out that the growth in the administration of the system was disproportionate to the decline in the monetary value of the grants. This lead the compilers of the Review Panel Report (2000) to conclude that the “amount of administration, which might have been acceptable five years ago, is now no longer tolerable with respect to the real value of the return, and leading researchers are wondering whether the gain is worth the effort”. (Review Panel Report, 2000)
The Krige Report also points to the fact that the guidelines to committee members for the assessment of reviewers’ reports were adjusted almost annually and became even more detailed between 2000 and 2005. According to the report this made an already overburdened system more complicated. They continue: “It seems that reviewers themselves were showing signs of fatigue”. In 2005, guidelines for specialist (selection) committee members were extremely detailed in outlining the possible issues which might arise in peer review reports.

Specialist Committee members must play an interpretive role when they assess reviewers’ reports. For example, if a reviewer states that the applicant is publishing in top journals, yet the journals are in the bottom tier of journals in the field, then the validity of the reviewer’s report must be called into question. Similarly Specialist Committee members should recognise the weakness of reviewers’ reports which overly praise the importance of the applicant’s work and where the work concerned is clearly not the calibre suggested in the report. (p. 81, R2.1)

We return in our concluding assessment to this issue as this is a classic case of a system producing diminishing returns for the effort and cost invested in it.

5. The inclusion of the social sciences and humanities in a unified rating system in 2003

After approximately two years of consultation, the NRF decided in 2002 to expand the evaluation and rating system to scholars and researchers in the social sciences and humanities. As the Krige Report shows, during the consultation process, various concerns and issues were raised about the feasibility of this decision. In a report on a workshop held in February 2002 on this issue, at least four “critical factors” were identified that would need to be addressed:

- the very large range covered by social sciences and humanities;
- the question of what comprises research in music, art and design;
- the conceptual and methodological diversity; and
- the permeability of disciplinary boundaries. (p. 77, R2.1)

In the 2005 Institutional Review the authors make an important point that speaks directly to the assessment of social scientists. They argue that the current system is seen by many scholars in the system as valourising international at the expense of local research. “A high rating requires evidence of the international standing of the research. Yet, in many disciplines, established researchers have stressed to us the importance of contributing to and shaping local debates, engaged with critical developmental issues. With leading researchers writing for local journals, the international reach and impact of a piece of work may not be an appropriate yardstick of its scholarly quality. The case of law was often mentioned in this regard.” (p. 86, R2.1)
One gets a clear sense from the Krige Report that the peculiar challenges that the inclusion of the social sciences and humanities in the rating system would pose to the NRF have not yet been adequately considered, discussed and addressed. There are significant differences in the research and scholarship practices of the different disciplines in the social sciences and humanities respectively. It is inevitable that social policy research, research on local educational problems, investigations of the South African legal system and interpretations of South African literary texts (to name a few) are primarily undertaken to impact on local rather than international events and dynamics. It is now generally recognised in social studies of science, that the pursuit of “local knowledge” is a legitimate enterprise. Considerations of universal application and international reputation are much less appropriate to these endeavours than in some fields of the natural sciences.

These issues are similar to our comments in Section 3.1.3.4 (3) on different modes of knowledge production. In that section, we referred to the earlier discussions on how the engineering sciences were accommodated in the rating system. The argument here is essentially the same.

6. A growing disillusionment with the system amongst certain academic and scholarly circles.

The Krige Report shows that most of the early reviews (Hawkins and MacKenzie in particular) were generally positive in their overall assessments of the rating system despite criticisms of specific aspects thereof. However, more general criticism of the system started to surface in the 2000 Review Panel Report. The report raised questions as to whether the researcher rating system had any real impact on either the amount of funding granted or the outcomes of the research. (p. 71, R2.1)

But the most pervasive criticism of the rating system would only appear in the NRF Institutional Review Report (2005). The compilers of this report argue that the system is now the “subject of intense contestation and debate in the higher education sector, varying from advocacy by institutions developing a research profile (as an indicator of progress, for example) to active opposition and doubts on the part of serious scholars in the system (p. 83, R2.1). They continue with a very negative assessment:

We have found the rating system to be divisive in the scholarly community; have posed questions about its alignment with the NRF missions and the criteria used in the NRF’s research funding system, and have heard from many groups that the rating system was increasingly irrelevant because its only link to funding was eligibility for longer-term support; believe there is unevenness in the instruments used for rating and problems with the “one size fits all” assumptions; and are concerned about the system’s inherent bias and excessive operational complexity. (p. 83, R2.1)
Their final conclusion is rather damning: In short the rating system was no longer seen as part of “a robust and credible instrument for assessing research quality”. (p. 84, R2.1)

**Concluding comments**

One of the issues that the Krige Report comments on throughout its documentation of the history of the rating system is the responses of the NRF to the various reviews and reports that had been produced over this period. In general, the NRF has always maintained that the system is defensible and appropriate “because it was a constant within a rapidly changing set of programme and evaluation approaches, procedures and tools”. (p. 88, R2.1)

The Krige Report comments as follows on this:

> A consequence has been that the system was adjusted rather than interrogated. Changes took place in three areas: the award categories, in terms of numbers of categories, terminology and numeric allocation; in the number of peer reviewers selected, and in the rating of peer reviewers’ reports. In general, however, the reviews from 1991 did not seem to have a serious impact on the NRF’s direction in terms of the rating system. (p. 88, R2.1)

We conclude with their assessment of what they regard as the most significant event in the history of the rating system:

> Possibly the most significant change to the system was provoked by the restructuring of the NRF itself in response to major political change. The subsequent implementation of the MCDM tool began a move away from individual researcher evaluation and marked the beginning of the delinking of individual rating from funding. At the time of writing funding and rating have been completely delinked, though, as the NRF itself points out, project proposals which involve the participation of rated researchers are more likely to be successful in being granted funding. Nevertheless, this indicates a significant move away from Prof De Wet’s hierarchical rewarding of excellent researchers supported in doing “free research”. (p. 88, R2.1)

### 3.2 Summary of and comments on Report 2.2 (Mapping the formal and informal use of the rating system over time by various institutions; C. Lombard)

#### 3.2.1 The brief and questions addressed

*The Brief: Mapping the formal and informal use of the rating system over time by various institutions (including their research offices, academic faculties, departments and programmes, human resource departments, individual researchers, etc), government departments like the Department of Education (DoE), institutions such as the Council on Higher Education – Higher Education Quality Committee (CHE-*
HEQC), science councils and private companies with large research and development (R&D) units like SASOL, ESKOM, etc. This mapping should also indicate the purposes for which the ratings are/were being used by the various groupings. This project would involve collecting primary data from the institutions and organisations.

The author further explicates the above by stating that the purpose of the study is “to investigate and, through evidence-based conclusions, determine the formal and informal use of the rating system over time”. (p. 4, R2.2)

3.2.2 Approach and methodology

The primary approach was a qualitative one, informed by wide-ranging individual and group interviews, as well as through written responses to mailed questionnaires, and a background literature review. The interviews were semi-structured and covered the following stakeholder groups: executive research managers, research administrators, functional managers of researchers (heads of department, deans), and in some cases researchers. The scope of institutions providing individual or corporate input included 21 universities (of which the author visited 15), 4 museums and 12 science councils or other public research-performing institutions, with more than 160 individual inputs. In addition, research-promoting institutions such as the national Department of Science and Technology and seven large private or parastatal corporates also provided responses. Written responses were solicited through a web-based questionnaire that was disseminated to all rated researchers; 194 responses were received to the 1700 questionnaires distributed – although being “disappointed” at this low rate of return, the report considers that the responses “represent a sample that can provide useful information”. (p. 25, R2.2)

Although a response rate of 11% may not seem so low as to exclude fairly robust conclusions, it may suggest little enthusiasm for or capacity to engage with the survey on behalf of rated researchers; views of non-rated researchers were not canvassed in this exercise. Lombard, however, does not come back to this point in the subsequent analysis of her findings. In addition to narratives, the report also includes some cleverly-designed tabular graphs that present the key

---

3 Earl Babbie (2000) *The practice of social research*. “A quick review of the survey literature will uncover a wide range of response rates. Each of these may be accompanied by a statement like “This is regarded as a relatively high response rate for a survey of this type.” (A U.S. senator made this statement regarding a poll of constituents that achieved a 4-percent return rate.) Even so, it’s possible to state some rules of thumb about return rates. I feel that a response rate of 50 percent is adequate for analysis and reporting. A response of 60 percent is good. And a response rate of 70 percent is very good. You should bear in mind, however, that these are only rough guides; they have no statistical basis, and a demonstrated lack of response bias is far more important than a high response rate”. As an acknowledged expert in survey methodology, Babbie’s views are influential. However, it is also fair to say that lower response rates (between 30 and 50%) for postal surveys are also accepted by other experts as still being able to produce credible results.
findings in a very succinct fashion. Finally, primary data were also collected from numerous relevant corporate documents such as policies and procedures.

The report is divided into three broad categories:

- the first (sections 2 and 3) respectively presents separate corporate perspectives on the use of the rating system within universities and other institutions (spanning both research-performing and research-promoting institutions) and the impact of this use;
- the second (section 4) presents findings on the perception of individual researchers about the impact the rating system has had on their careers; while
- the third category (sections 5 and 6) assesses several contextual matters impacting on the use of the rating system.

Because over 95% of rated researchers are employed at universities, most of the emphasis in the report is on this sector.

Addenda include lists of acronyms and abbreviations, references, and participants in interviews, as well as guidelines for the semi-structured interviews, and the questionnaire to researchers. In addition, the author includes an appendix with comments on the NRF rating system and its practices, informed by the free-format inputs she had received in the written responses and the interviews, but whose contents strayed beyond the brief for the report.

Overall the report is well-considered and structured, and within the constraints of its brief methodologically sound. On most occasions Lombard mentions obvious limitations to the certainty of her conclusions, and the only comments we would make here are offered mainly for the sake of completeness: Lombard’s report lists several criticisms made of the rating system (p. 31, R2.2) without, however, conveying an exact sense of the degree to which they are held among the 194 respondents. Although methodologically correct, based on our own experience, we expect such a quantification would merely have revealed more clearly the unorganisable spread of opinions against the system, without adding much more information.

The only other constraint on her conclusions, which Lombard has also pointed out, is the low level of response from institutions and organisations other than universities.

This is a very interesting systematisation of the uses to which the rating system are being put – and no longer being put. Lombard has documented and described in a methodologically-sound manner a wide array of uses to which the rating system is put, and has provided comprehensive

---

4 This theme is also probed in the Marais report (see section 3.3 of this report)
discussions of these, as summarised in the next section which highlights what we consider the key findings.

3.2.3 Key findings and conclusions of report

3.2.3.1 USE OF RATING SYSTEM BY INSTITUTIONS

The report documents extensive formal and informal use by universities of the rating system across a number of their internal operations and functions. Largely, the rating system is used more broadly in the domains of Science, Engineering and Technology (SET) than in the Social Sciences, Humanities and Law (SSHL), though in some universities the rating system is also used strongly within the latter disciplines. The fourteen identified uses of the rating system are the following:

- **Strategic goals** – eleven universities indicated that improving the number of rated researchers constitutes an institutional strategic objective and plays a role in strategic planning.

- **Research management** – eight reported using the rating system in their internal research management activities, for example as a criterion in the definition of research centres or niche areas.

- **Recruitment** – most universities utilise the rating system formally or informally to suggest levels of preference in their recruitment processes especially in SET, and some even acknowledged using the NRF listing of rated researchers for headhunting exercises.

- **Selection and appointment** – the rating system is widely used as an indicator or recommendation in selection and appointment of academic staff, though it is never used formally; again its use is stronger in SET than SSHL.

- **Promotion** – as in the previous category, although here the formal use was counter-indicated in most cases on the basis that the system is not equally transferable across different disciplines, and merger-related inequities in qualifications also compounded its use in promotion exercises.

- **Retention** – although pre- and post-retirement retention of leading researchers is a concern across the sector, very few universities’ formal or informal retention procedures explicitly factor in the rating system.

- **Performance management** – only three universities indicated the formal use of the rating system in performance management, though perhaps its use in strategic management (and goal setting, therefore) may extend its formal application; informally it finds use in career mapping.
Remuneration – twelve universities indicated existing or pending formal use of the rating system in remuneration of staff through either once-off bonuses or salary increases, though opinions on this matter are strongly divided across the sector.

Internal funding – eleven institutions indicated formal or informal use of the rating system to guide internal grant allocations.

Awards – many universities use the rating system to inform decisions about special internal (research) prizes.

Benchmarking – although a very high level of formal use of the rating system in strategic planning (and presumably the formulation of Key Performance Indicators) was evident, no university reported its formal use in benchmarking, though there is wide informal use.

Profiling and marketing – 14 universities indicated formal or informal use of the rating system in their promotional and marketing materials and efforts.

Research capacity development – in a fair number of universities the rating system is used as an important benchmark in guiding the development of researchers via the Thuthuka programmes.

Other funders/Networks – several universities indicated that reference to the rating system was used tactically in attempts to leverage more private resources; interestingly, there is a wide-spread belief among researchers that industry has no interest in the rating system, though this was proven false through feedback from large corporate research promoters and sponsors.

The extent to which the rating system is used for the above purposes across the university sector is depicted (mapped) graphically in Table 3, below, which was included as Table 2.1 in the report.

Uses are categorised as either “Formal” (i.e. included in documentation and policy) or “Informal”. “Informal: High” indicates that the university and/or faculty attaches high importance to the system as an indicator though it might not be included in documentation; “Informal: Low” indicates that there is low importance attached to the importance, though it is still used. (p. 7, R2.2)

The following key is used to indicate the different levels of use:

Formal: Red
Informal: High: Green
Informal: Low: Blue
In some cases an institution scores at two levels for a specific “use”. This system has been used to indicate that possibly a particular faculty might use it formally but the university does not; or that the university uses the system both informally and formally for that specific use.

Where the score is marked with an asterisk (*), this indicates that the university has definite plans to include the rating system as a criterion/indicator (either formally or informally), but plans are not yet approved.

In summary, the report has identified the following four clusters of universities differentiated by the extent of utilisation of the rating system; those that use it (p. 15, R2.2):

- Extensively, strategically and to a large extent formally:
  University of Pretoria, North-West University, Tshwane University of Technology, University of Stellenbosch and the University of the Witwatersrand.

- Those who use it fairly widely – informally and formally:
  Nelson Mandela Metropolitan University, University of Johannesburg, University of Cape Town, University of the Free State, UNISA.

- Those who use it largely informally:
  Central University of Technology, Durban University of Technology, Cape Peninsula University of Technology, University of KwaZulu-Natal, University of the Western Cape.

- Those who use it in a very limited way:
  University of Limpopo, Rhodes University, Mangosuthu Technikon, University of Zululand, Vaal University of Technology, Walter Sisulu University.

In concluding this section, the report notes the fundamental shift in the utilisation of the rating system: from an FRD/NRF-internal indicator of research funding with little external use, toward an externally-utilised indicator of research development largely uncoupled from funding decisions.
## Table 3: Map of uses made of rating system by universities

<table>
<thead>
<tr>
<th></th>
<th>CPUT</th>
<th>CUT</th>
<th>DUT</th>
<th>Mangosuthu</th>
<th>NMU</th>
<th>NWU</th>
<th>Rhodes U</th>
<th>Stellenbosch U</th>
<th>TUT</th>
<th>UCT</th>
<th>UFS</th>
<th>UF</th>
<th>UFS State</th>
<th>UJ</th>
<th>UKZN</th>
<th>UJ</th>
<th>UNISA</th>
<th>UJ</th>
<th>U Pretoria</th>
<th>UWC</th>
<th>U J Venda</th>
<th>U Zululand</th>
<th>VUT</th>
<th>WITS</th>
<th>WSU</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Strategic Goals</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Formal</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
</tr>
<tr>
<td>Informal: H</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
</tr>
<tr>
<td><strong>Research Management</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Formal</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
</tr>
<tr>
<td>Informal: H</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Informal: L</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td><strong>Recruitment</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Formal</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td>Informal: L</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Selection &amp; Appointment</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Formal</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Informal: H</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Informal: L</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Promotion</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Formal</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Informal: H</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Informal: L</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Retention</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Formal</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Informal: H</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Informal: L</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Performance Management</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Formal</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Informal: H</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Informal: L</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Remuneration</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Formal</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Informal: H</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Informal: L</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 3 (cont.): Map of uses made of rating system by Universities

<table>
<thead>
<tr>
<th></th>
<th>CPUT</th>
<th>CUT</th>
<th>DUT</th>
<th>Mangosuthu</th>
<th>NMMU</th>
<th>NWU</th>
<th>Rhodes U</th>
<th>Stellenbosch U</th>
<th>TUT</th>
<th>UCT</th>
<th>UFree State</th>
<th>UJ</th>
<th>UKZN</th>
<th>U Limpopo</th>
<th>UNISA</th>
<th>U Pretoria</th>
<th>UWC</th>
<th>U Venda</th>
<th>U Zululand</th>
<th>VUT</th>
<th>WITS</th>
<th>WSU</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Funding</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Formal</td>
<td>I</td>
<td>I</td>
<td>H</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
</tr>
<tr>
<td>Informal: H</td>
<td>D</td>
<td>I</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>D</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
</tr>
<tr>
<td>Informal: L</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
</tr>
<tr>
<td><strong>Awards</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Formal</td>
<td>I</td>
<td>I</td>
<td>D</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
</tr>
<tr>
<td>Informal: H</td>
<td>I</td>
<td>I</td>
<td>H</td>
<td>L</td>
<td>D</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>D</td>
<td>L</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
</tr>
<tr>
<td>Informal: L</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
</tr>
<tr>
<td><strong>Benchmarking</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Formal</td>
<td>I</td>
<td>I</td>
<td>D</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
</tr>
<tr>
<td>Informal: H</td>
<td>I</td>
<td>I</td>
<td>H</td>
<td>L</td>
<td>D</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>D</td>
<td>L</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
</tr>
<tr>
<td>Informal: L</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
</tr>
<tr>
<td><strong>Profiling: Institution</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Formal</td>
<td>I</td>
<td>I</td>
<td>D</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
</tr>
<tr>
<td>Informal: H</td>
<td>I</td>
<td>I</td>
<td>H</td>
<td>L</td>
<td>D</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>D</td>
<td>L</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
</tr>
<tr>
<td>Informal: L</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
</tr>
<tr>
<td><strong>Research capacity development</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Formal</td>
<td>I</td>
<td>I</td>
<td>D</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
</tr>
<tr>
<td>Informal: H</td>
<td>I</td>
<td>I</td>
<td>H</td>
<td>L</td>
<td>D</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>D</td>
<td>L</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
</tr>
<tr>
<td>Informal: L</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
</tr>
<tr>
<td><strong>Other funders/networking</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Formal</td>
<td>I</td>
<td>I</td>
<td>D</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
</tr>
<tr>
<td>Informal: H</td>
<td>I</td>
<td>I</td>
<td>H</td>
<td>L</td>
<td>D</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>D</td>
<td>L</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
</tr>
<tr>
<td>Informal: L</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
</tr>
</tbody>
</table>
The report next turns its attention to mapping the extent of use of the rating system within museums – part of the original focus of the de Wet/Arndt rating system. Unfortunately only four of the eleven institutions contacted provided a response; in these cases, the formal or informal use of the rating system is limited to only eight or nine institutional functions, whereas in universities 14 such functions were identified; Table 4 maps the findings of the study in the same manner as outlined above for universities. (p. 17, R2.2)

**Table 4: Uses made of the rating system by museums**

<table>
<thead>
<tr>
<th></th>
<th>Albany</th>
<th>Bayworld</th>
<th>Durban</th>
<th>East London</th>
<th>IZiko</th>
<th>McGregor</th>
<th>Natal</th>
<th>National</th>
<th>Northern Flagship</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recruitment</td>
<td>Formal</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
</tr>
<tr>
<td></td>
<td>Informal: H</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
</tr>
<tr>
<td></td>
<td>Informal: L</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
</tr>
<tr>
<td>Selection &amp; Appointment</td>
<td>Formal</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Informal: H</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Informal: L</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Promotion</td>
<td>Formal</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>Informal: H</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td></td>
<td>Informal: L</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td>Retention</td>
<td>Formal</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Informal: H</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Informal: L</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Performance Management</td>
<td>Formal</td>
<td>R</td>
<td>R</td>
<td>R</td>
<td>R</td>
<td>R</td>
<td>R</td>
<td>R</td>
<td>R</td>
</tr>
<tr>
<td></td>
<td>Informal: H</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
</tr>
<tr>
<td></td>
<td>Informal: L</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
</tr>
<tr>
<td>Remuneration/Awards</td>
<td>Formal</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
</tr>
<tr>
<td></td>
<td>Informal: H</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td></td>
<td>Informal: L</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Benchmarking</td>
<td>Formal</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
</tr>
<tr>
<td></td>
<td>Informal: H</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Informal: L</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Profiling: Institution</td>
<td>Formal</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Informal: H</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Informal: L</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other funders/networking</td>
<td>Formal</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Informal: H</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Despite the relatively small number of rated researchers (18) in the museum sector, the report maintains that the rating system “is quite extensively used”, outlining several specific formal and informal applications to internal functions.
The use of the rating system in Science Councils, National Facilities and other research institutions was also explored by means of 13 responses to 19 institutions approached. Within Science Councils the use of the rating system appears to be very limited, though the report outlines a few instances of its use. Importantly, the HSRC apparently believes the rating system to be “biased towards university-based researchers who tend to become experts in a relatively narrow field of scholarship, as opposed to other parts of the National System of Innovation where applied and team-based research is more generally undertaken” (p. 20, R2.2). Also, the Medical Research Council does not take into account the rating of an applicant in its funding decisions. The one possible exception is the Agricultural Research Council, which seems to attach a fairly high degree of importance to the rating system, though it does not actually use the system in its internal functions.

Only two of five National Facilities (NFs) approached provided responses, so it is difficult to draw conclusions, but the report nevertheless expresses surprise at the seemingly low implementation of the rating system by the NFs – especially considering they report within the NRF. Several other public research organisations were also canvassed but revealed very little, if any, utilisation of the system.

Lastly, the report explores the use of the rating system within seven private research sponsors including Sasol, Sappi, Denel, Anglo American Research, etc. Where companies had been aware of the system, they often utilised it informally as an indicator of research quality in funding decisions. Government departments that responded to the survey (DST and DoE; only two out of seven contacted) do not utilise the system in their business.

The report concludes this section by pointing to the potential for increased use of the rating system within this group of public research organisations, and encourages the NRF to market the system more effectively.

### 3.2.3.2 Use of Rating System by Individual Researchers

The findings of the report in respect of the use of the rating system by and its impact on individual researchers are drawn from a survey of the 194 responses received to 1700 distributed questionnaires, these responses emerging from individuals in all rating categories, distributed roughly in accordance with the number of rated researchers in each of the categories. The spread of respondents across the disciplines is also roughly similar to that of rated researchers.

Although all efforts were made by the author to restrict responses to those issues explicitly covered in her brief, respondents evidently felt compelled to “provide detailed opinions on the rating system in general and their experience of it in particular” (p. 25-6, R2.2) – the author reports on these comments in Addendum 4 of the report, and encourages the NRF to consult those comments.
The survey explored the perceived impact of the rating system (positive; negative; no impact) within the following five dimensions:

- **Career advancement** – 58% indicated a positive impact, the largest positive experience coming from SET and SSHL programs, the lowest from the Health and Medical sciences (the low impact of the rating system in these disciplines may be linked to the fact that the Medical Research Council makes no use of the rating system). Importantly, the perceived impact of the rating system on career advancement seems directly proportional to the seniority of the rating attained: thus A- and Y-rated scientists generally perceived the most positive impact of all researchers – i.e., the lower the rating of the researcher, the lower the impact of rating on their career development. This effect is independent of discipline.

- **Increased funding** – on average only 16% of respondents felt that rating had generated positive impacts in regard to funding, but 40% of those in the P-category had experienced positive benefits in this regard – i.e., in higher-ranked categories there is a better chance of securing high funding, but on average little positive impact is perceived to flow from rating in respect of funding. This effect is independent of discipline.

- **Access to research community** – 75% of respondents indicated that they had perceived no benefits in regard to increased access to the South African research community as a consequence of being rated. However, most A-rated scientists felt that being rated had indeed improved their access to the local research community. This effect is independent of discipline.

- **Access to international research community** – similar results to those above were obtained for this dimension, and again it is independent of discipline.

- **Increased profile** – just under half the respondents indicated that rating had improved their profile in the broader community, with some variation in the effect across the disciplines, though. Interestingly, 100% of Fine Arts respondents indicated rating had improved their profile, and again the highest-rated respondents (A and P) indicated the greatest positive effect on their profile, those with C-ratings the lowest.

The report concludes that researchers do not perceive significant benefits from being rated in regard to the perceptual issues explored. (p. 30, R2.2)

Finally, Lombard observes a disjuncture in attitude of rated researchers between the “fairly high level of ambivalence regarding the rating system” uncovered in her survey, (p. 30, R2.2) and the “findings of a Corporate Stakeholder Survey conducted in January 2005 which indicated that 73.4% of grant-holders were positive that the NRF “provided a useful rating of individual
researchers (49.8% agreed with the statement and 23.6% definitely agreed) with only 10.9 disagreeing. 15.7% were neutral. This point is not explored further by Lombard, however.

### 3.2.3.3 Institutional Context of Rating System

Institutions’ reasons for utilising the rating system for their internal functions largely and generally relate to the widespread perception (perhaps more strong among research managers than researchers?) that the system is rigorous, robust and transparent, and cannot easily be replicated in the institutions (p. 31, R2.2) – this is a fairly commonly-held view, and clearly represents the justification for the range of uses to which the rating system is put in the institutions, especially the universities. On the other hand, the number and range of criticisms of the rating system reported in the Lombard Report illustrates little consensus around objective concerns and many misperceptions. For instance, apparently many researchers believed that rating is the “final arbiter” on a wide range of matters, without being able to offer any objective grounds for this belief (p. 31, R2.2). In Addendum 4 Lombard presents an analysis of the various free format responses, inputs and opinions she received, which the NRF is advised to study if the rating system is to continue being utilised.

The report then reflects on several constraints to the wider implementation of the rating system, several of which seem peripheral and need not be commented on here. However, the more important ones include the following:

- **The debilitating effects of mergers** – apart from sapping energy, the institutional mergers have in most cases required a complete re-write of the full spectrum of university policies and procedures and that in this urgent need frameworks promoting the use of the rating system generally were at the back of the queue.

- **Low levels of research development within (some) institutions** – where there are very few rated researchers because the institutional level of research is comparatively underdeveloped, the implementation of the rating system as some internal discriminator of management tool makes little sense.

- **The perceived lack of incentives (since funding is no longer linked to rating)** – critically, the effective delinking of research grants from the rating system as a consequence of the growth in funding requests considerably exceeding the growth in available funds has “had an extremely negative effect on everyone’s perception of the rating system” (p. 32-3, R2.2). Researchers are subject to competition in a dual process now – rating and project evaluation – but only the latter actually brings with it measurable rewards.

---

5 NRF Institutional Review – Corporate Stakeholder Survey. January 2005 (p. 27)
Scepticism about the scientific validity of the system and its equitable application across all disciplines – there continues to exist a strong divide between the SET and SSHL disciplines over the acceptability of the rating system (although in a few universities this is not the case). Many in the SSHL domain remain convinced that there is a fundamental discontinuity between the rating system and the prevalent system of science in this area because much research in SSHL is (a) multidisciplinary whereas the rating system seeks depth rather than broadness of scope, (b) locally applied, and therefore unsuitable for international peer reviewed academic journals, a key benchmark of the rating system, and (c) many of the disciplines in this domain have little or no counterparts outside South Africa, thus making meaningless the rating system notion of “international standing”. Broadly-speaking, researchers in SSHL believe the rating system to have been crafted around the scientific practices of SET, and its applicability to SSHL is therefore in question. In the health sciences a specific hurdle is the conflict between most academics’ dual appointment as a clinician (focused on patient care) and an academic (expected to do research).

Lastly, Lombard makes an interesting observation in that there is no clear link between the number of rated researchers and extent of utilisation of the rating system in an institution: at the extremes, UCT with the largest number makes little apparent use of the system, whereas TUT, for instance, with a small number of rated researchers makes extensive use of the rating system.

The report ends with a brief summary of the key points, several conclusions and suggestions.

Finally, there are four addenda: (i) a list of participants providing written feedback, group, individual interviews, and telephone interviews; (ii) guidelines for semi-structured interviews; (iii) the questionnaire for individual researchers, and (iv) an analysis of free-format comments on the NRF rating system by those participating in the survey.

3.2.3.4 SUMMARY OF KEY FINDINGS

Perhaps the most important matter is Lombard’s articulation of what most researchers and research managers in the country had already intuitively concluded – namely that the purpose of the rating system is now almost exclusively external to the NRF, rather than internal. Perhaps this is still a perceptual matter, but certainly there is not much external evidence or knowledge of any specific utilisation of rating within NRF processes, the last vestige of its use having been eroded in grant award letters thanking rated researchers for good proposals, but (unfortunately) turning down or severely cutting their requests due to funding constraints.

In a sense, researchers have run into the law of diminishing returns: they need to invest more and more effort to be and remain competitive on paper, but for less and less returns in terms of research grants. As with economic activities this is an unsustainable trend that can be expected
sooner or later to reach a critical point of inflection where the effort is no longer justified by the rewards. On a few occasions Lombard’s report hints that this point is being reached quickly or may already have been reached; for example: “However a much deeper level of interrogation of the systems and processes is required. Without the support of the primary stakeholders the system will continue to experience declining credibility.” (p. 41, R2.2)

But a systematic assessment of this question was outside her brief and therefore not undertaken; neither was it part of any of the other reports.

A second and important question raised in the report is whether the NRF is monitoring progress (throughput) of individual researchers through the ratings? One would expect this given the developmental nature of the rating system. Although making reference to it, Lombard’s report does not explicitly explore this point; but given our own knowledge of the NRF and its processes we can say we are unaware of a systematic tracking and management of rating candidates through the different rating categories. In part, this would assist in providing the rating system with a purpose once more. In his report Marais indicates that of 327 Y-rated researchers approximately 33% later obtained at least a C-rating and approximately 6% attained “considerable recognition and became international leaders”; in his view this constitutes “a satisfactory return on investment”. (p. 32, R2.3)

Thirdly, if the rating system is to continue, its application within the NSI needs to be optimised in order to leverage maximum impact of the system – specifically, this means that all public research organisations (whether performers or funders) should adopt the system; certainly it is surprising that the National Facilities are apparently not utilising the system in a material way despite the fact they report within the NRF! The MRC’s apparent lack of interest in utilising the (or a similar) system in its funding decisions further dilutes the credibility; this is probably true also of the WRC.

Fourthly, in reconceptualising the future of the rating system, the NRF should bear in mind Lombard’s findings that rating has only had a positive impact on the personal profile of A-rated researchers, with possible marginal positive impact on P- and B-rated ones: in other words, researchers with lower ratings could legitimately ask the question: why bother?

Lastly, if the NRF decides to continue the system in its current or even some pared-down form and reestables a transparent and reliable link to funding, there would be a need (a) to promote greater insight into the system among all stakeholders, and (b) greater conceptual buy-in and support especially within the SSHL. Moreover, a review should be undertaken of the applicability of the system to multi-disciplinary research and to applied research whose output addresses more directly policy needs than peer-reviewed publications.
3.3 Summary of and comments on Report 2.3 (Impact of the NRF evaluation and rating system: A review; H. Marais)

3.3.1 The brief and questions addressed

The Brief: The impact of the rating system within specific disciplines or fields of scholarship on the scholarly productivity of South African academics. In other words, this study should address the social impact of the rating system, i.e. how the introduction of the rating system changed researchers’ attitudes to research and their focus within fields of research. The review panel would be interested in an analysis of the effect of the rating system on attitudes to undertaking research by staff of universities and the transformation of universities into research intensive institutions.

It may be useful to undertake a comparative analysis of fields such as engineering (which have a significant professional orientation) with other less applied areas of research. While it is acknowledged that the evaluation and rating of individuals in the humanities, social science and law has had a relatively short history, the service provider will be expected to comment on the short-term and anticipated difference the system has made/will make on these fields. The service provider will be expected to identify indicators such as quantity and quality of research outputs, internationalisation of research findings, networking, student training, etc, that will be useful for providing evidence of the impact of the introduction of the rating system. Evaluation and rating data at the NRF will be made available to the service provider.

This brief presents itself as multifarious and complex: for instance, the first sentence links the key notion of “impact” to “scholarly productivity”, but thereafter it is linked to “social”, to “researchers’ attitudes”, to “researchers’ ... focus within fields of research”, and then the notion of “effect” is introduced adding further complexity. The penultimate sentence requires the author of this report to generate indicators (metrics) that can be used to provide evidence of impact, but it is not clear at face value whether those metrics are meant to allow measurement of future impact, or whether they are meant to inform the earlier (historic) assessment of impact(s).

In addition to the conceptual complexity of the above brief, its clarification was furthermore complicated by the evident need for Marais to negotiate the content of his report with the HESA/NRF consortium that commissioned it, as evidenced, for example, by his comment that the “original proposal for the current project included a survey of university researchers’ attitudes and experiences, but that component was not supported by (the) HESA/NRF task group, since ... signs of respondent resistance were detected”. (p. 8, R2.3)
Against this background it is understandable that Marais spends some time unpacking this brief into specific research questions and deliverables, this exercise also being strongly influenced by a very formal approach to the establishment of an appropriate methodological framework.

Eventually the following questions emerged in response to the brief and subsequent negotiations with the commissioner of the report:

- Has the NRF rating system influenced the attitudes and orientations of academics as individuals (also as groups/teams/other types of units) toward research and funding sources for such research?
- Has the NRF rating system influenced the outputs (productivity) of academics as individuals?
- Has the NRF rating system influenced any aspects of the fields of knowledge production?
- Has the NRF rating system had any influence at an institutional level (i.e., on the organisational structures of universities)?
- What are the perceptions of R&D communities, particularly in the higher education sector, of the NRF rating system?

The potential for some overlap with the Lombard Report is evident; this will be discussed further in subsequent sections of this report.

### 3.3.2 Approach and methodology

In contrast to some of the other reports we have reviewed, Marais’s approach to the formulation of a methodological framework was extremely elaborate and highly formal. Eventually a mix of qualitative and quantitative investigations was undertaken, the quantitative aspects drawing data from the NRF and the SA Knowledgebase held at CREST at Stellenbosch University, as well as a secondary analysis of data compiled during the previous NRF Stakeholder Survey in 2004/5. Qualitative information was gleaned from individual interviews (with 10 “key role players in the rating system” [p. 11, R2.3]) and focus group interviews (with 37 participants almost evenly distributed across six regional interview session). While Marais lists the backgrounds of the 10 participants in the individual interviews (8 academic and institutional managers, only 2 “senior” scientists), he does not provide a similar insight into the participants in the group interviews; this is a pity since Lombard’s report suggests managers/administrators and researchers may have different perceptions of the merit and impact of the rating system – the former appreciating it for its

---

6 Institutional review of the National Research Foundation, Pretoria: NRF, 2005
utility value in strategic and performance management, the latter seeking a meaningful purpose in it.\footnote{Indeed, Marais confirms this point himself on p. 34 of his report.}

The database compiled during the NRF Stakeholder Survey included responses from NRF grantholders, panel members, reviewers, research organisations, and NRF staff and board members, but the variables analysed during that survey and those Marais’s intended to probe did not coincide in all respects, and for this reason he cautions (p. 13, R2.3) that the former “should be seen as proxies at best”, based on the “reasonable assumption that if a respondent reacted positively toward an item dealing with rating-related matters, it would indicated a positive attitude towards the rating system itself” (p. 31, R2.3). The latter assumption seems potentially (perhaps fatally) flawed when one considers the following analogy: a positive response to an item dealing with policing-related matters need not necessarily reflect a positive attitude toward the system of policing.

The secondary analysis involved correlating responses to rating categories which were combined, though, because of the small number of responses to the original survey. Specific parameters probed included “the performance of the NRF … (in) … increasing the quality and quantity of researchers and students, facilitating the utilisation of knowledge, adherence to quality, and provision of useful rating of individual researchers” (p. 14, R2.3).\footnote{Lombard has already made reference to the results of this survey, having found inconsistencies between its (fairly favourable) findings and her own.} Mann-Whitney $U$ tests and a Kruskal-Wallis analysis were applied to the data.

Publication patterns of researchers were extracted from the SA Knowledgebase. Four broad disciplinary domains were investigated: natural and agricultural science; engineering and applied technologies; health sciences; social sciences and humanities. 3684 rated researchers were tracked, but a further 567 could not be matched to entries on the database. It is not clear how many unrated researchers constituted the database. (p. 15, R2.3)

In addition, Marais highlights eight published documents that were also consulted during his review.

From a quantitative perspective, Marais’s sample size is very moderate when compared to Lombard’s; moreover, as will be seen in the discussion of his findings, Marais’s approach suffers from several intrinsic constraints that ultimately limit the report’s use in evaluating the impact of the rating system on the development of South African research.
3.3.3 Key findings and conclusions of report

3.3.3.1 IMPACT OF RATING SYSTEM ON RESEARCHERS

Chapter 3 of Marais's report is titled “Attitudes and orientations of researchers”; we interpret the link between this theme and the brief to mean that Marais is assessing the impact of the rating system on researchers by means of their attitudes and orientations, though we missed this conceptual link being made explicit. These attributes are traced via the quantitative secondary assessment of previous survey results (as outlined above), and through the individual and group interviews. Of course, a study of the attitudes of researchers towards the system, is not really an impact measure. An impact assessment would seek to understand the benefits to and effects on a system or people in the system. At best, a survey of attitudes would reveal whether the researchers are positively or negatively orientated towards the system which MIGHT be a function of whatever benefits they had gained prior or not.

Unfortunately, in reporting on his quantitative analysis Marais omits to mention what number of the researchers in the database are rated/not-rated; seemingly neither is an indication of this provided in his outline of the methodological approach on the secondary analysis. The closest one gets to this is in his cryptic reference to the respondents all being “involved in the rating system” (p. 32, R2.3). In addition, of the four findings he records only one is relevant to the brief for this report (in his own words it is the “most directly relevant”): (a) that 65% of researchers agreed that “the NRF provided useful ratings to researchers”, and (b) that the most highly-rated ones agreed with this statement most strongly. Finding (a) had already been alluded to by Lombard, who pointed out that this to some extent deviated from her own findings, though she declined to explore this further. The value of this finding is hard to determine, though, since the phrase “useful ratings” is highly ambiguous. Marais’s second finding correlates with that of Lombard who found that in general only A-rated researchers had experienced positive effects flowing from their rating. The other three findings simply reflect on researchers’ attitudes toward the NRF and its beneficial effects on the quality and quantity of research and researchers.

The usefulness of Marais’s qualitative assessment is also mixed, unfortunately. They are useful because to a large extent they also correlate with Krige and Lombard’s findings (and associated conclusions) in respect of the following:  

- while there is some measure of support for the rating system within the (university) system, there is also disillusionment and “cynical rejection”; (p. 33, R2.3)

- the delinking of funding from rating-status is the one key demotivator impacting on attitudes toward the rating system; and

---

9 In other words, a measure of triangulation has been achieved.
the dual application process of rating and proposal evaluation has significantly increased administrative demands on researchers while being accompanied by a significant and real decrease in funding support (we have alluded to the law of diminishing returns).

Where Marais’s qualitative analysis becomes shaky is when it begins to suggest quantitative conclusions, especially in view of his much more limited direct respondent base than Lombard’s (whose interview process drew on more than 160 people and who directly surveyed 194 individuals): in fact, Marais records qualitative assessments on the basis that they were held by “more than one” participant in the focus groups! (p. 36, R2.3.) Most problematic is his conclusion (p. 35, R2.3) that “researchers and university management support the system and would like it to continue” subject to certain conditions – this conclusion is ungrounded because (a) there is no clear separation of researchers from management in his analysis and assessment of results (and it seems his focus groups were also mixed), and (b) there is simply no quantitative basis for such a very generalised statement.

In fact, the one persuasive and substantiated issue raised in this chapter suggests exactly the opposite to Marais’s above conclusion: for four consecutive years of the last five there has been a steady and significant increase in the number of researchers whose ratings have lapsed, as illustrated in the Table below, which maps the lapsed ratings for UCT, UKZN, UW, SU, and UP (“Top five”), the remainder of the university sector (“Rest”), and some other research institutions utilising the rating system (“Other”). (This table was adapted from Table 3.1 on page 37 of the Marais Report). Doubtless some of this increase can be accounted for by natural attrition, but it is difficult not to be struck by especially the radical increase in 2007.

**Table 5: Lapsed ratings of researchers**

<table>
<thead>
<tr>
<th>Year</th>
<th>Type of Institution</th>
<th>∑</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Top five**</td>
<td>Rest</td>
</tr>
<tr>
<td>2003</td>
<td>43</td>
<td>24</td>
</tr>
<tr>
<td>2004</td>
<td>23</td>
<td>15</td>
</tr>
<tr>
<td>2005</td>
<td>25</td>
<td>11</td>
</tr>
<tr>
<td>2006</td>
<td>33</td>
<td>22</td>
</tr>
<tr>
<td>2007</td>
<td>81</td>
<td>54</td>
</tr>
<tr>
<td>∑</td>
<td>205</td>
<td>126</td>
</tr>
</tbody>
</table>

Sources: NRF (2007) and NRF communication
Notes: ** UCT, US, Wits, UP, UKZN
* Other: Museums, science councils, unknown
3.3.3.2 Impact of Rating System on Research Output

In Chapter 4 Marais turns his attention to “Output Productivity”, through which, in part, he intends assessing the impact of the rating system on the quantity of research output – the other part constituting an analysis of its impact on knowledge production, i.e., the nature of the research produced (Chapter 5, “Fields of Knowledge Production”). The approach is largely quantitative and reveals numerous interesting and new data.

The first metric probed by Marais is the average per capita article output of rated and unrated researchers over the periods 1990-3, 1994-7, 1998-2001, and 2002-5; data for this analysis were extracted from the SAKnowledgebase with the assistance of CREST staff. (In considering this data readers should bear in mind that rating only became a feature of the SSHL domain from about 2000 onwards.) Table 6, below, extracted from Table 4.1 of the Marais Report (p. 42), elegantly illustrates

(a) the significantly higher average per capita publication rates of rated over non-rated researchers;
(b) a general trend of decreasing scientific productivity over the 1990-2005 period; and
(c) decreasing productivity across the domains of natural sciences, social sciences and humanities, engineering and technology, and health sciences.

Table 6: Average per capita article count: rated/non-rated researchers, 1990-2005

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Natural Sciences</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rated researchers</td>
<td>2.73</td>
<td>2.63</td>
<td>2.48</td>
<td>2.42</td>
</tr>
<tr>
<td>Non-rated researchers</td>
<td>0.41</td>
<td>0.38</td>
<td>0.35</td>
<td>0.30</td>
</tr>
<tr>
<td>Engineering and Technology</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rated researchers</td>
<td>1.68</td>
<td>1.60</td>
<td>1.56</td>
<td>1.55</td>
</tr>
<tr>
<td>Non-rated researchers</td>
<td>0.47</td>
<td>0.44</td>
<td>0.42</td>
<td>0.39</td>
</tr>
<tr>
<td>Health Sciences</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rated researchers</td>
<td>1.61</td>
<td>1.63</td>
<td>1.67</td>
<td>1.49</td>
</tr>
<tr>
<td>Non-rated researchers</td>
<td>0.33</td>
<td>0.31</td>
<td>0.31</td>
<td>0.26</td>
</tr>
<tr>
<td>Social Sciences and Humanities</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rated researchers</td>
<td>2.47</td>
<td>2.42</td>
<td>2.52</td>
<td>2.18</td>
</tr>
<tr>
<td>Non-rated researchers</td>
<td>0.80</td>
<td>0.77</td>
<td>0.72</td>
<td>0.63</td>
</tr>
</tbody>
</table>

The heightened productivity of rated researchers is also demonstrated in Table 7 (below, extracted from Table 4.3 of the report, p. 46). Broken up according to the disciplines listed, these data
examine the number of researchers with more than 3 articles since 1990\textsuperscript{10} (column 2), the number (column 3) and percentage (4) of those authors who are NRF rated, and the percentage of rated researchers among the top 100 most productive researchers. The data clearly demonstrate the above average productivity of rated researchers within all domains.

Table 7: Relative publication performance of rated researchers

<table>
<thead>
<tr>
<th>(1) Scientific field*</th>
<th>(2) No. of researchers with &gt; 3 articles</th>
<th>NRF-rated researchers</th>
<th>(5) % NRF-rated researchers among top 100 most productive researchers</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>(3) No.</td>
<td>(4) % column 2</td>
</tr>
<tr>
<td>Health Sciences</td>
<td>1 437</td>
<td>304</td>
<td>21</td>
</tr>
<tr>
<td>Engineering</td>
<td>463</td>
<td>248</td>
<td>54</td>
</tr>
<tr>
<td>Natural Sciences</td>
<td>2 651</td>
<td>251</td>
<td>47</td>
</tr>
<tr>
<td>Social Sciences</td>
<td>2 802</td>
<td>557</td>
<td>20</td>
</tr>
</tbody>
</table>

* “Natural Sciences” include the agricultural disciplines, while “Social Sciences” include Humanities and Economic Sciences.

Next, Marais investigated the breakdown of the heightened productivity of rated researchers to their rating category. Specifically, he examined whether an upward mobility within the rating system (for example from B- to A-rating; from C- to B- rating, etc.) is linked to increased productivity, as one would expect if A-rated researchers are assumed to be more established and (hence) more productive than ones with lower ratings. His findings show that “in the case of researchers that received their rating in the period 2002-6, upward mobility was associated with increased article productivity” but that productivity declined for “immobile” and “downwardly mobile” researchers during the same period (p. 43, R2.3). Specifically this means that as researchers progress up through the rating system they can be expected to become more productive.

Over the period of investigation (1990-2005) both rated and un-rated researchers increased their rate of ISI-accredited publications while decreasing that in South African journals, but although this is interesting, it says nothing about the potential impact of the rating system as the trend is independent of rating status. Similarly, as measured by international co-authorship both rated and un-rated researchers increased significantly the extent of international collaboration in their

\textsuperscript{10} A threshold criterion of 3 papers over the period 1990 – 2005 was included in order to ensure some comparability between the rated scientists and the non-rated. The argument advanced by CREST was exactly that if someone had not published at least 3 papers in 15 years, it would be someone who is likely to be a student (who co-published with the supervisor) or alternatively a very inactive scientist who most likely never submitted for rating.
publications over the period under review: again this is interesting though the apparent jump after 1994 is, of course, not surprising, but as the trend is not unique to rated researchers it provides little insight into the possible impact of the rating system on the observable. Marais also shows that A-rated researchers (as of the period 2002-2006) have the highest rate of international co-authorship, but because international co-authorship is not exclusively linked to the potential impact of the system, this is an interesting but not germane observation. Lastly, Marais demonstrates that international co-authorship is linked to upward mobility within the rating system – as it is fairly much also to those who remain immobile in the system – but that for those who are downwardly mobile the relationship is neither positive nor negative. Again, because non-rated researchers showed the same increase in international co-authorship over the period of review, Marais's findings bear little relevance to the issue at hand: the potential impact of the rating system on international co-authorship.

Lastly, Marais sought to explore the potential impact of the rating system on other types of research outputs and outcomes, including patents, artefacts, and students. While some of the information he presents is anecdotally interesting, it adds little or no value to the purpose of his report, because in part the data are self-selected (presented by the researchers in their funding applications and not independently vetted after the fact), and because in all cases no benchmarks or comparisons with unrated researchers are presented.

In conclusion, the obvious point can be made that all of these data are typical correlational data. The fact that there are correlations between output and increased co-authorship and increased publication in ISI journals can easily be explained by other causes, such as the liberalisation of the science system in South Africa after the apartheid isolation years, increased collaboration, and networking. And, of course, generally increased output is probably more related to the DoE subsidy system than anything else.

Hence, the linking of rating to research productivity merely demonstrates that the rating system is effective in identifying productive researchers – as measured against a specific yardstick (that of accredited articles and some qualitative assessment of impact of the researcher). Marais himself commented on this fundamental philosophic constraint as follows:

As already noted several times in this report, one should be cautious in interpreting relationships as definitive impacts proven by data in the tables. The question remains whether the data demonstrate the “impact” of the rating system on publication behaviour or the external validity of the rating system. Distinguishing between cause and effect is not possible with the available information, as presented here, but it would be fair to conclude that this analysis showed a strong relationship between ratings and publication behaviour and at least does not invalidate the *NRF Rating System*. (p. 47, R2.3)
Any direct, positive impact of the rating system on the publication “behaviour” of scientists would most likely have persuaded and incentivised researchers to change their behaviour manifesting itself in observables such as:

- more scientists would have been rated more highly (getting an A-rating then becomes the immediate goal);
- more scientists would have moved through the ranks over the years (from C to B to A);
- more scientists would have submitted for rating and more would have been successful.

But the only reference to this is that indicating an approximately 33% throughput rate from Y-rated.

At an even more systemic level, it would be reasonable to expect the rating system – in recognising and rewarding excellence – would have increased South African science’s overall performance, in term of its overall share of world output (which has actually declined), increased number of highly-cited papers (no information was presented on this) and so on.

### 3.3.3.3 IMPACT OF RATING SYSTEM ON DISCIPLINES AND KNOWLEDGE PRODUCTION

In Chapter 5 (“Fields of Knowledge Production”), Marais seeks to map the relationship between the rating system and (i) the 22 traditional disciplines within which it operates, and (ii) knowledge production more generally. He reminds the reader that the thrust of most national science and technology (and related research) strategies is toward Mode 2 research characterised strongly by multi- and interdisciplinary as well as applied research, and although he hints at it, he does not significantly probe the apparent discontinuity that might arise between this strategic imperative and the rating system, which seems more discipline-bound, promoting fundamental as opposed to applied research. Instead, after a few more sentences Marais concludes that the rating system has “had at least two effects at disciplinary level: firstly, the rating system has served as a promoter of quality research in which fundamental research is given due recognition, and, secondly, it has added to the ambivalence around the continuum between fundamental and applied sciences” (p. 53, R2.3). Unfortunately we are sure of neither the source nor the purpose of these statements.

Thereafter Marais presents numerous empirical observations and statistics about the distribution of rated researchers across five categories of disciplines (natural sciences, social sciences, humanities, health sciences, and engineering), their respective rating categories, and their “R&D Expenditure” – the exact definition of the latter category is not clear: it could mean the level of NRF grants awarded in these categories, the amount of NRF funds expended (which is not the same as the former), or the R&D expenditure as per the national R&D survey (which includes all direct and indirect costs), but in this case it is not clear how this could be linked to only rated researchers. In other words, this category is ill-defined. Moreover, while some of the data are interesting, their
merit in terms of the purpose of the Marais Report is not clear, given lack of comparison with non-rated researchers and a historical analysis: the data are time-bound and descriptive, not historic and analytical, and have no apparent link to trends or impact.

However, in the section entitled “Participation in directed funding programmes” some interesting and revealing rated/un-rated comparisons come to the fore. Figure 1 (below, extracted from Figure 5.2 of the report, p. 56) depicts the source of and amount of funding received by rated and un-rated researchers, respectively. The figure illustrates that both in proportion to their respective Focus Area grants and in actual financial terms unrated researchers benefit far more from THRIP and Innovation Fund support than do rated researchers. The former derive 19% (equivalent to R214m), while the latter derive only 14% (R99m) of their funding from directed sources. Marais discusses this diagram and concludes

- “rated researchers participate extensively in NRF funding programmes” (but so do unrated ones);
- that perhaps rated researchers apply more to undirected programmes because they appreciate the greater degree of freedom granted in research conceptualisation; and
- that research projects undertaken by rated researchers may not have significant “impact” on key strategic imperatives.

The latter, especially strikes us as a particularly pertinent observation in regard to the potential impact of a rating system on knowledge production, but unfortunately this point was not explored further. It would have been interesting, for example, to consider the corollary of this observation and its implication for the rating system: as judged by the extent of support for applied and strategically important (albeit utilitarian) knowledge production, unrated researchers may be far more efficient in addressing national priorities than are rated researchers!
Finally, Marais concludes this section with a reflection – influenced by qualitative input received during his interviews – on whether the rating system might have influenced “the orientation of rated researchers towards disciplinary research” (p. 57, R2.3). We are unsure about the validity of this discussion because of (i) the very limited number of individuals participating in his survey, (ii) the absence of clarity on how many of them were researchers, and (iii) uncertainty in our reading that in respect of this theme Marais is only reporting on qualitative feedback from researchers – as opposed to administrators.

Moreover, internal discontinuities further undermine the validity of this section. For instance, Marais reports that “comments were made that the rating system served as a strong spotlight on a research proposal” (p. 58, R2.3), but fails to outline how or whether he probed this type of feedback further: the rating system and the process of formulating and submitting a proposal are operationally, conceptually and administratively divorced from each other, and the above statement consequently sheds little light on the stated theme. Such imprecision creates further uncertainty about the merit of the qualitative discussion in this report.

Nonetheless, Marais lifts one perceptual issue that had also surfaced in other reports (for example, in Lombard’s) that needs to be mentioned here: the rating system operates within narrow
disciplinary boundaries – this may be beneficial to the/some disciplines, but undermines inter-disciplinarity (and applied and policy-oriented research).

3.3.3.4 IMPACT OF RATING SYSTEM ON INSTITUTIONS AND THE RESEARCH COMMUNITY

While conceding that this theme is the subject of the Lombard Report, Marais nonetheless explores a few possible impacts of the rating system on the institutional environment, commenting on the implications for the internal policy environment, and presenting an analysis of the relationship between NRF funding for and the number of rated researchers at universities; the information for this was drawn from von Guenewaldt, and is graphically illustrated in Figure 2 (extracted from Figure 6.1 of the report, p. 65). Although the graph suggests a strong correlation between these two parameters, as such graphs often do it also reveals instances where universities with fairly similar number of rated researchers are the recipients of significantly different levels of NRF support (e.g., UWC and UJ – the universities of the Western Cape and of Johannesburg), or vice versa (RU and UF – Rhodes University and the University of the Free State). This chapter presents some other interesting statistics, but as in the case just discussed, the data present snapshots of the system from which little unequivocal information can be extracted about the impact of the rating system on institutions and the research community.

Figure 2: Relationship: Number of rated researchers and NRF funding leveraged


### 3.4 Summary of and comments on Report 2.4 (The NRF evaluation and rating system in the world context; A. Pouris)

#### 3.4.1 The brief and questions addressed

**The Brief:** The system of rating individuals for funding purposes is often regarded as a uniquely South African innovation. A comparative study of other national systems for the evaluation of individuals is required. In this study it will be important to compare “apples with apples” since the NRF rating system is based on the rating of individuals whereas other systems may be rating proposals, departments and/or institutions. In this case the unit of analysis will be the rating of the individual or the individual’s contribution to the rating of the department/institution/proposal.

As stated in the Brief, the overall aim of this study was to compare the NRF Rating system with other similar research evaluation and assessment systems. No additional specific research questions were formulated in response to this brief. Presumably such a comparison could have addressed questions such as the following:\(^\text{11}\):

- What are the main similarities and differences between the NRF Rating system and other evaluation systems that also involve a (numerical) rating or scoring of individual scientists?
- What were the original reasons and motivations for the introduction of such systems and how does this compare with the introduction of the NRF rating system?
- Do these other systems have similar aims and expected outcomes compared to the NRF system?
- Have other rating systems changed fundamentally over time and, if so, what have prompted these changes and what have been the effects of such changes?
- Where such systems differ from the NRF system, what are the main reasons behind such differences?
- Have there been any regular and recent reviews of such systems and, if so, what have been learnt from these reviews?

#### 3.4.2 Approach and methodology

This study consisted essentially of a desktop study and primarily used internet materials. In addition to the use of internet materials, the author also consulted with a few key informants at the

---

\(^{11}\) The author of this report did not explicitly formulate any questions for his review. The questions listed here are our additions.
NRF and other funding councils. With one exception, no other studies were consulted. In our own brief review of the New Zealand and Mexican funding systems, we included references to additional internet sources.

According to Pouris, the objective of his report was “to outline a number of approaches similar to the peer review system used by the NRF (at a national level), compare and contrast them and develop relevant recommendations where the NRF approach appears to deviate from international best practice (p. 10)”. He continues:

While there are a large number of organizations using peer review for supporting research (e.g. Defense Advanced Research Projects Agency (DARPA), Department for Environment, Food and Rural Affairs (DEFRA), Deutsche Forschungsgemeinschaft (DFG), Research Council of Norway, Wellcome Trust and others) we chose to focus on the systems of the National Science Foundation in the USA; the Australian Research Council [ARC] and the Research Councils in the United Kingdom. These systems were chosen because of the expressed interest of those organizations to follow international best practices. In addition, we refer to a number of schemes which focus specifically on evaluation/rating of individuals (to the exclusion of projects) such as the “Performance Based Research Fund” in New Zealand, the National Science Council’s “Research Outcome Award” in Taiwan and the “National Researchers System” (SNI) in Mexico. (p. 10)¹²

Unfortunately the approach followed by the Pouris Report and captured in the quote above is flawed in two respects. First, the decision to compare the NRF rating system with other evaluation systems that use good or best practice in assessment does not address the issues at hand. Most research evaluation approaches in the top science systems of the world employ some mode of peer assessment. This is true of the DFG (Germany), the NWO (The Netherlands), the Canadian Research Council, the CNRS in France, all the research councils in the Scandinavian countries and many more. The Brief explicitly required a comparison with evaluation systems that include some (quantitative) rating of individual scientists. For this reason alone, the discussion in a large part of the report to the NSF, the ARC and the ESRC in the UK is irrelevant to the Brief.

Second, Pouris unfortunately does not clearly distinguish between the fundamental differences that characterise the different funding agencies and approaches in a dual funding system. The most obvious example is the UK system where a dual system of funding is in operation (as in South

---

¹² Since the Project Brief clearly excludes a comparison with systems that are not aimed at generating individual ratings at all, it is not clear why the Pouris report included the NSF, UK (RAE) and Australian evaluation systems. These discussions, which constitute the bulk of the Pouris report, are simply devoted to the standard peer review processes involved in the evaluation of grant proposals. The sections on the New Zealand and Mexican systems in the report are the sections that are appropriate to the Brief but are unfortunately also very brief (only 6 pages). We also do not understand why the Taiwanese outcome grants are discussed in the report as no rating or scoring is involved in the approval of these grants. For the purposes of this synthesis report we have therefore gathered additional information on the New Zealand and Mexican systems in order to arrive at a more informed comparison with the NRF system.
Africa). The funding councils in the UK (such as the Higher Education Funding Council of England) are tasked to allocate block funding on behalf of the government to all universities. Since 1986 this allocation has been based on a performance-based system – the UK Research Assessment Exercise. The UK RAE which is conducted every 3-4 years involves a very rigorous assessment of all academic departments who submit themselves for assessment. Departments (not individuals) are ultimately rated on a 7-point scale. These departmental scores are aggregated to produce an institutional score which determines the amount of funding that the university will subsequently receive from the respective funding council. The second component of the dual funding system in the UK is represented by the respective research councils (Economic and Social Research Council, Engineering and Physical Science Research Council, Medical Research Councils). These councils operate like most national funding agencies and allocate resources on a competitive basis after peer assessment of project proposals.

It should be clear that these two “streams” in the UK science system perform very different functions. The UK RAE is a mechanism to allocate block funding to higher education institutions to fund research and is based on peer review in more than 60 field-specific panels. The research councils constitute a second channel of targeted, project funding on a more regular basis and is, of course, also based on peer review.

The South African system also involves a dual funding approach: the performance-based research funding mechanism of the Department of Education where institutional allocations are based (since 2003) on research outputs (publications and post-graduate students) and the system of funding agencies (primarily the NRF and MRC but to some extent also the WRC).

What is unique to the South African system (with the possible exception of the Mexican system) is that the NRF employs two different assessment systems: the standard project review process which is comparable to what most research councils worldwide do and the individual rating system.

This rather lengthy discussion is necessary to explain why the approach followed in the Pouris Report is misleading and why the findings of his report were unfortunately of limited use to the compilers of this report.

3.4.3 Key findings and conclusions of report

Research evaluation systems and approaches differ from each other in at least three crucial ways:

- The unit of the evaluation or assessment (the what).
- The purposes of the evaluation or assessment (the why).
- The mechanism or process of the evaluation system (the how).
Against this background it is possible to position the NRF Rating system in comparison with other research evaluation systems. First, it is quite evident that most – if not all – national funding agencies utilise some form of peer review mechanism. This is especially prevalent where research proposals are submitted for evaluation. In these cases peer review is the only generally accepted mechanism (e.g. NSF, UK Research Councils, DFG, Australia and so on). The brief for this project, however, clearly excluded these “standard” forms of peer review and required that the NRF rating system be compared to other systems where some form of rating or scoring is applied to individual scientists. Table 8, below, presents a “mapping” of different evaluation systems that employ some form of rating or scoring. It also includes, for the sake of completeness, other levels of evaluation (departments/programmes) as well.

**Table 8: Evaluation systems that involve "rating"**

<table>
<thead>
<tr>
<th>Evaluand</th>
<th>Rating Goal</th>
<th>System</th>
<th>Evaluation purpose</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Individual scientist</td>
<td>Intrinsic (end in itself)</td>
<td><strong>Mexico (SNI)</strong></td>
<td>Reward/recognition Incentive schema for individuals</td>
<td>Ongoing</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>South Africa (NRF)</strong></td>
<td>Reward/recognition of individuals</td>
<td>Ongoing</td>
</tr>
<tr>
<td></td>
<td>Extrinsic (means to a higher end)</td>
<td><strong>New Zealand (PBRF)</strong></td>
<td>Funding allocation to institutions</td>
<td>Every 3 yrs (change to every 6 years soon)</td>
</tr>
<tr>
<td>Department of academic unit</td>
<td>Extrinsic</td>
<td><strong>UK (RAE)</strong></td>
<td>Funding allocation to institutions</td>
<td>Every five years</td>
</tr>
<tr>
<td>Research programme</td>
<td>Intrinsic</td>
<td><strong>Netherlands (VNU)</strong></td>
<td>Funding allocation to institutions</td>
<td>Every five years</td>
</tr>
</tbody>
</table>

*“Rating”: i.e. assigning a numerical value to the quality or worth of the evaluand (unit of evaluation)*

Against the background of our discussion in Section 3.4.2 above, the table captures the distinction between rating systems that form part of institutional funding allocations (such as the UK RAE and the New Zealand PBRF) and rating systems that culminate in a rating of a scientist for the purpose of individual recognition and reward. We refer to this as a distinction between the *intrinsic* and *extrinsic* goals of individual rating systems. Where the system’s primary aim is to generate a (numeric) rating of the individual as a “stand alone rating”, we will refer to this as an intrinsic rating goal. Where the system’s primary aim is to generate a (numeric) rating of the individual as part of or as a means to a higher-order aggregate score, we will refer to this as an extrinsic rating goal.
As is indicated below, this is one of the crucial differences between the NRF rating system, which is motivated by an intrinsic evaluation goal, and the New Zealand PBRF where the aim is to rate individual scientists in order to produce scores which are subsequently aggregated to the levels of subject areas and academic institutions. In the latter case the point of generating individual ratings is to ultimately arrive at an assessment (and even ranking) of subject-area and university for funding purposes. In the case of the NRF, no further aggregation of individual scores is undertaken but the individual rating is undertaken as the intrinsic goal of the system.

Our discussion of the Pouris Report is thus confined to the results of his discussion of the New Zealand and Mexican systems as the only systems that involve individual ratings. In our discussion of these two systems, we have used the following headings:

- Origins and rationale for the system.
- Purpose of the system.
- Evaluation mechanism.

### 3.4.3.1 Performance-based Research Fund (PBRF) in New Zealand

#### The origins of and rationale for the PBRF

The New Zealand government’s decision in mid 2002 to introduce the PBRF marked the culmination of many years of vigorous debate over the best way of funding research in the country’s tertiary education sector. In 1997, the previous National-led government proposed a new system for research funding and subsequently appointed a team of experts to consider the options. For various reasons, little progress was made. In 2001, the Tertiary Education Advisory Commission (TEAC), which was appointed by the Labour-Alliance government, recommended the introduction of the PBRF as a central component of a new funding regime for the tertiary sector. In essence, TEAC recommended a mixed model for assessing and funding research: on the one hand, the proposed model incorporated an element of peer review (as used in the British and Hong Kong research assessment exercises [RAEs]); on the other hand, it incorporated several performance measures (as used in the Australian and Israeli research funding models).

The proposed measures were external research income and research degree completions. In response to the TEAC Report, the government established a working group of sector experts in mid 2002, chaired by Professor Marston Conder. This group worked with the Transition Tertiary Education Commission and the Ministry of Education to develop the detailed design of a new research assessment and funding model for the tertiary sector. The Report of the Working Group on the PBRF — Investing in Excellence — was published in December 2002 and approved by the Cabinet.
In brief, the working group endorsed the key elements of the funding model proposed by TEAC, including the periodic assessment of research quality by expert panels and the use of two performance measures. It also supported TEAC's idea of using individuals as the unit of assessment, rather than academic units as in Britain. It did, however, recommend that the funding formula have different weightings from those proposed by TEAC – and it developed a comprehensive framework for assessing the research performance of individual staff.

It is interesting to note that two very practical concerns motivated the establishment of this study of institutional assessment. The writers of the 2002 document believed that the results of the PBRF would enable “students to make decisions about where to study; and the private sector and community to identify appropriate research partners”. In addition it was anticipated that “the publication of quality scores will promote further concentration of funding and people around areas of excellence”.

These comments provide further clarification and substantiation of the points made above, viz. that the PBRF is essentially a system by the New Zealand government to allocate institutional funding in a fair and systematic manner. In order to arrive at such a fair disbursement of allocations, individuals and departments are rated and their scores aggregated to produce a total institutional score.

**Purpose of the PBRF**

The original motivation to address the challenges posed by weak research programmes and the need to develop a system that would address strengthening programmes and concentrations of research capacity is affirmed in the 2006 report by the Tertiary Education Commission (TEC):

The implementation of the Performance-Based Research Fund (PBRF) acknowledged that TEOs had been heavily dependent upon EFTS funding in order to support their research activities. This meant that certain research programmes were vulnerable to large shifts in student demand. It also meant that the volume of research in particular subject areas was determined more by the pattern of student demand than by the quality of research being undertaken. In the late 1990s, a portion of the EFTS subsidies for degree-level programmes was notionally designated for research in the form of degree “top ups” and the subsidy rates for different course categories were adjusted. This did not, however, alter the fundamental nature of the research funding system in the tertiary education sector; nor did it address the underlying weaknesses.

But it is also clear that the aim of the PBRF was also to recognise and reward excellence, not merely to strengthen pockets of excellence:

From 1999 onwards, significant efforts have been made to improve the tertiary funding regime in the interests of encouraging and rewarding excellence. The first major step in this process was the government's decision in 2001 to fund the creation of a number of centres of research excellence
(COREs) within the tertiary sector. To date, seven COREs have been established; a further selection round is in progress. A second key step was the establishment of the PBRF as a funding programme that entails the periodic assessment of research quality together with the use of two performance measures. All the funding that earlier had been distributed via the degree “top ups” has now been transferred to the PBRF; and, in 2007, more than $67 million (including GST) additional funding will be available. (TEC, 2006)

These aims are elaborated upon in the latest TEC Report on the system (TEC: 2006) “The government’s main aims for the PBRF are to: increase the average quality of research; ensure that research continues to support degree and postgraduate teaching; ensure that funding is available for postgraduate students and new researchers; improve the quality of public information about research output; prevent undue concentration of funding that would undermine research support for all degrees or prevent access to the system by new researchers; and underpin the existing research strengths in the tertiary education sector”.

The evaluation mechanism

The PBRF is governed by the following set of principles:

- **Comprehensiveness**: the PBRF should appropriately measure the quality of the full range of original investigative activity that occurs within the sector, regardless of its type, form, or place of output;

- **Respect for academic traditions**: the PBRF should operate in a manner that is consistent with academic freedom and institutional autonomy;

- **Consistency**: evaluations of quality made through the PBRF should be consistent across the different subject areas and in the calibration of quality ratings against international standards of excellence;

- **Continuity**: changes to the PBRF process should only be made where they can bring demonstrable improvements that outweigh the cost of implementing them;

- **Differentiation**: the PBRF should allow stakeholders and the government to differentiate between providers and their units on the basis of their relative quality;

- **Credibility**: the methodology, format and processes employed in the PBRF must be credible to those being assessed;

- **Efficiency**: administrative and compliance costs should be kept to the minimum consistent with a robust and credible process;

The PBRF has three components:
a periodic Quality Evaluation using expert panels to assess research quality based on material contained in Evidence Portfolios;  

- a measure for research degree completions; and 

- a measure for external research income.

In the PBRF funding formula, the three components are weighted 60/25/15 respectively. The PBRF is managed by the Tertiary Education Commission Te Amorangi Maturanga Matua.

In summary: The PBRF is a “mixed” performance-assessment regime in the sense that it employs both peer-review processes and performance measures. There are three elements to its assessment: periodic Quality Evaluations: the assessment of the research performance of eligible TEO staff, undertaken by expert peer review panels; a postgraduate “research degree completions” (RDC) measure: the number of postgraduate research-based degrees completed in participating TEOs, assessed on an annual basis; and an “external research income” (ERI) measure: the amount of income for research purposes received by participating TEOs from external sources, assessed on an annual basis. For funding purposes, the weightings given to these three elements are: 60% for the Quality Evaluation; 25% for RDC; and 15% for ERI.

In the 2007 funding year, the funding allocated by means of the three PBRF performance measures was almost NZ$231m (based on current forecasts) and is derived from 100% of the former degree “top up” funding, together with additional funding from the government totaling NZ$67m per annum. Performance in the 2006 Quality Evaluation determined the allocation of 60% of this funding until the next Quality Evaluation (planned for 2012). Overall, the PBRF will determine the allocation of approximately NZ$1.5 billion over the next six years.

Under the approach adopted, the maximum quality score that can be achieved by a TEO, subject area or nominated academic unit is 10. In order to obtain such a score, however, all the PBRF-eligible staff in the relevant unit of measurement would have to receive an “A” Quality Category. Given the nature of the assessment methodology adopted under the 2006 Quality Evaluation, and the very exacting standards required to secure an “A”, such an outcome is extremely unlikely.

The standards required for achieving an “A” Quality Category, as stated in the PBRF Guidelines 2006 and applied by the 12 peer review panels were exacting. Many staff who produced research outputs of a world-class standard did not secure an “A” because they did not demonstrate either

---

13 Evidence Portfolios: Collection of information on the research outputs, peer esteem, and contribution to the research environment of a PBRF-eligible staff member during the assessment period that is reviewed by a peer review panel and assigned to a Quality Category.
the necessary level of peer-esteem or a contribution to the research environment to the standard required.

Two other factors also contributed to some high-caliber researchers receiving a “B” rather than an “A”:

a) The assessment period covered only six years. In some cases, major research outputs were produced just before, or just after, the assessment period, with the result that the researcher in question received a lower score for their Research Output component than might otherwise have been the case.

b) The Evidence Portfolios of some high-caliber researchers did not provide sufficient detail of their Peer Esteem and/or Contribution to Research Environment.

The end-result of the PBRF is a ranking of institutions in terms of their overall aggregate score. This feeds into the final funding allocation to each institution is illustrated by the 2007 results in Table 9.

**Table 9: 2007 PBRF Indicative Funding**

<table>
<thead>
<tr>
<th>TEO</th>
<th>Quality Evaluation</th>
<th>Research Degree Completions</th>
<th>External Research Income</th>
<th>Total</th>
<th>Percentage of Total PBRF Funding</th>
</tr>
</thead>
<tbody>
<tr>
<td>University of Auckland</td>
<td>$37 442 726</td>
<td>$19 265 406</td>
<td>$13 153 591.00</td>
<td>$69 861 723</td>
<td>30.28%</td>
</tr>
<tr>
<td>University of Otago</td>
<td>$30 944 018</td>
<td>$9 502 337</td>
<td>$8 051 667.00</td>
<td>$48 496 022</td>
<td>21.02%</td>
</tr>
<tr>
<td>Massey University</td>
<td>$20 122 794</td>
<td>$9 964 081</td>
<td>$4 494 520.50</td>
<td>$34 581 396</td>
<td>14.99%</td>
</tr>
<tr>
<td>University of Canterbury</td>
<td>$14 468 664</td>
<td>$6 984 796</td>
<td>$1 958 699.38</td>
<td>$23 412 159</td>
<td>10.15%</td>
</tr>
<tr>
<td>Victoria University of Wellington</td>
<td>$13 492 715</td>
<td>$5 057 858</td>
<td>$2 123 526.75</td>
<td>$20 674 100</td>
<td>8.96%</td>
</tr>
<tr>
<td>University of Waikato</td>
<td>$8 840 939</td>
<td>$4 076 049</td>
<td>$1 910 509.13</td>
<td>$14 827 497</td>
<td>6.42%</td>
</tr>
<tr>
<td>Lincoln University</td>
<td>$4 323 681</td>
<td>$1 179 007</td>
<td>$2 110 840.75</td>
<td>$7 613 528</td>
<td>3.30%</td>
</tr>
<tr>
<td>Auckland University of Technology</td>
<td>$3 797 089</td>
<td>$1 042 630</td>
<td>$488 716.31</td>
<td>$5 328 435</td>
<td>2.31%</td>
</tr>
<tr>
<td>Unitec New Zealand</td>
<td>$2 154 291</td>
<td>$218 203</td>
<td>$77 689.67</td>
<td>$2 450 184</td>
<td>1.06%</td>
</tr>
<tr>
<td>Otago Polytechnic</td>
<td>$462 783</td>
<td>$83 185</td>
<td>$15 700.10</td>
<td>$561 668</td>
<td>0.24%</td>
</tr>
<tr>
<td>Waikato Institute of Technology</td>
<td>$335 576</td>
<td>$90 458</td>
<td>$63 158.25</td>
<td>$489 192</td>
<td>0.21%</td>
</tr>
<tr>
<td>Manukau Institute of Technology</td>
<td>$411 272</td>
<td>$0</td>
<td>$20 843.03</td>
<td>$432 115</td>
<td>0.19%</td>
</tr>
<tr>
<td>Christchurch Polytechnic Institute of Technology</td>
<td>$347 531</td>
<td>$0</td>
<td>$5 655.80</td>
<td>$353 187</td>
<td>0.15%</td>
</tr>
<tr>
<td>Christchurch College of Education</td>
<td>$192 109</td>
<td>$26 857</td>
<td>$7 613.17</td>
<td>$226 579</td>
<td>0.10%</td>
</tr>
<tr>
<td>Te Wananga O Aoteaora</td>
<td>$170 794</td>
<td>$27 589</td>
<td>$10 560.64</td>
<td>$208 943</td>
<td>0.08%</td>
</tr>
<tr>
<td>Open Polytechnic of New Zealand</td>
<td>$161 503</td>
<td>$0</td>
<td>$45 384.50</td>
<td>$206 887</td>
<td>0.09%</td>
</tr>
<tr>
<td>Te Whare Wananga O Awanuiarangi</td>
<td>$184 321</td>
<td>$0</td>
<td>$5 729.86</td>
<td>$190 051</td>
<td>0.09%</td>
</tr>
<tr>
<td>Whitecliff College of Arts and Design</td>
<td>$31 426</td>
<td>$117 880</td>
<td>$0.00</td>
<td>$149 306</td>
<td>0.06%</td>
</tr>
<tr>
<td>Eastern Institute of Technology</td>
<td>$147 566</td>
<td>$0</td>
<td>$710.48</td>
<td>$148 276</td>
<td>0.06%</td>
</tr>
<tr>
<td>Dunedin College of Education</td>
<td>$62 511</td>
<td>$33 441</td>
<td>$6 800.94</td>
<td>$102 753</td>
<td>0.04%</td>
</tr>
<tr>
<td>Nelson Marlborough Institute of Technology</td>
<td>$78 873</td>
<td>$0</td>
<td>$0.00</td>
<td>$78 873</td>
<td>0.03%</td>
</tr>
</tbody>
</table>
**Concluding comments**

Although the New Zealand PBRF involves individual ratings of scientists, the system is still fundamentally different from the NRF rating system in at least three ways:

- The individual ratings are used as further inputs to generate aggregate scores for subject-areas and higher education institutions as a basis for national funding allocations.

- The PBRF as a system aims to assess the overall strength (and weaknesses) in the science system in order to improve research excellence and also encourage high-level skills development (PhDs).

- The intensity and costs of the system has dictated that such a review is only undertaken every three years (in fact as from 2006 this will only be done every six years) and is therefore not an ongoing and continuous process as is the case with the NRF rating system.

Overall, it is evident that the New Zealand PBRF has more similarities with the UK Research Assessment Exercise than the NRF Rating system.
3.4.3.2 THE NATIONAL RESEARCHERS SYSTEM (SNI) IN MEXICO

The origins and purpose of the SNI
Since the 1980s, the Mexican government has taken steps to strengthen quality assurance in its higher education sector. This has been the result of severe financial crises in the 1980s. The crisis caused a 50 percent decline in the purchasing power of faculty salaries, forcing many qualified academics to quit their jobs or to take on additional employment. This resulted in severe staffing problems and deterioration in teaching conditions at a time of increasing enrollments. This led to public concerns and government demands for improving the quality of higher education.

Mexico has initiated several quality assurance approaches for its higher education system since then. In the public sector, institutions have had some form of internal review since the early 1990s, initially through annual self-assessment and later through more detailed institutional development plans. At the same time, mechanisms for external evaluations based on external peer reviews of academic programs have also been put into place. In several professional specialties, accreditation councils have also been established.

In July 1984 a National System of Researchers (SNI) was created with the aim of acknowledging and rewarding the work performed by researchers in the country, whether at public universities, public research centres and some private universities having an agreement with the National Council on Science and Technology (CONACYT).

The evaluation mechanism
SNI is administered by the CONACYT - the most important public organisation in the country promoting and supporting science and technology activities. It reports directly to the President of Mexico and is responsible for coordinating, orienting, systematising and promoting scientific and technological activities.

The SNI identifies two categories: Candidates and Researchers. The first category is made up of students in the last year of their doctoral studies and students who have recently completed their doctorates.

The second category – researchers – is divided into three levels. The first level includes researchers with doctorates who have already demonstrated their productivity and are involved in innovative, high-quality research projects. The second level is made up of researchers who have consistently carried out research recognised for its originality, whether as an individual or as part of a group. Finally, the third level is reserved for researchers who have made important contributions to the fields of science or technology, the value of which has been recognised by the national and international academic community and who have also done outstanding work as educators at the highest level.
In all cases, the SNI provides some degree of funding support to beneficiaries that allow them to devote themselves full-time to their work in science or technology, without having to become distracted from this fundamental task. Benefits are multiples of the official minimum salaries, graded according to level and are tax free\(^\text{14}\). The SNI classifies its researchers into four knowledge areas:

- Area I, physical and mathematical sciences;
- Area II, the biological, biomedical, and chemical sciences;
- Area III, social sciences and the humanities; and
- Area IV, engineering and technology.

Selection occurs through a peer review system and maintaining membership is based on continuing productivity. Membership in the SNI system confers prestige in addition to providing more income. Figure 3 and Table 10, below, show the evolution of the SNI and highlights the significant growth in the number of members it has had over the last few years.

![Figure 3: Members of the SNI, 1984–2005, Mexico](image)

Source: CONACYT (2004) and (2005a)

**Figure 3: Members of the SNI, 1984–2005, Mexico**

---

Table 10: Members of the SNI by category and level, 1992–2005

<table>
<thead>
<tr>
<th>Year</th>
<th>Candidates</th>
<th>National Researcher by level</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>I</td>
<td>II</td>
</tr>
<tr>
<td>1992</td>
<td>2 655</td>
<td>2 860</td>
<td>779</td>
</tr>
<tr>
<td>1993</td>
<td>2 274</td>
<td>2 810</td>
<td>797</td>
</tr>
<tr>
<td>1994</td>
<td>1 683</td>
<td>3 012</td>
<td>807</td>
</tr>
<tr>
<td>1995</td>
<td>1 559</td>
<td>3 077</td>
<td>839</td>
</tr>
<tr>
<td>1996</td>
<td>1 349</td>
<td>3 318</td>
<td>862</td>
</tr>
<tr>
<td>1997</td>
<td>1 297</td>
<td>3 546</td>
<td>952</td>
</tr>
<tr>
<td>1998</td>
<td>1 229</td>
<td>3 980</td>
<td>1 032</td>
</tr>
<tr>
<td>1999</td>
<td>1 318</td>
<td>4 191</td>
<td>1 159</td>
</tr>
<tr>
<td>2000</td>
<td>1 220</td>
<td>4 345</td>
<td>1 279</td>
</tr>
<tr>
<td>2001</td>
<td>1 128</td>
<td>4 682</td>
<td>1 556</td>
</tr>
<tr>
<td>2002</td>
<td>1 324</td>
<td>5 385</td>
<td>1 729</td>
</tr>
<tr>
<td>2003</td>
<td>1 631</td>
<td>5 784</td>
<td>1 898</td>
</tr>
<tr>
<td>2004&lt;sup&gt;p&lt;/sup&gt;</td>
<td>1 876</td>
<td>5 981</td>
<td>2 076</td>
</tr>
<tr>
<td>2005&lt;sup&gt;p&lt;/sup&gt;</td>
<td>n.a.</td>
<td>n.a.</td>
<td>n.a.</td>
</tr>
</tbody>
</table>

Source: CONACYT (2003) and (2005)
Notes:
- n.a. = not available
- <sup>p</sup> Preliminary data.

**Concluding comments**

In our assessment the Mexican SNI comes closest in evaluation model to the NRF Rating system. It does involve individual ratings of scientists and these ratings (like the original NRF system) are directly linked to a monetary reward. Interestingly enough, the Mexican SNI system was designed and implemented primarily as a human resource development strategy as it was aimed initially to encourage Mexican scientists to remain in the country through a monetary reward system. In effect, it could be interpreted essentially as a brain gain initiative within a system of market failure where poor academic salaries were not adequate to retain the best scientists in the country. This is not to ignore the fact that the rating system also confers prestige on the highest rated scientists.
and that it also acts as an incentive system to improve research productivity. In essence it remains, however, a system of salary augmentation.

3.5 **Summary of and comments on Report 2.5 (Review of processes used to map the rating of individual researchers; M. Madikizela)**

### 3.5.1 The brief and questions addressed

**The Brief:** A review of the processes used to manage the rating system in the past five years. The NRF system of evaluation and rating has different sub-processes in which various groups of individuals are involved, i.e. applicants, research offices of the institutions to which the applicants are affiliated, peer reviewers and members of panels at three different levels within a hierarchy. These processes will be analysed, commenting on their efficiency, fitness-for-purpose, effectiveness, transaction costs, etc. This study should also comment on the strengths and weaknesses of the overall evaluation and rating processes.

The author summarised this brief as follows: “[T]he overall aim is to assess whether these processes are working in the most effective and efficient way possible within available resources” (p. 3, R2.5). The Executive Summary already makes mention of the fact that not all processes were examined due to time constraints. The methodological approach taken by the author in addressing the brief is outlined below.

### 3.5.2 Approach and methodology

Data and information for the report were sought and compiled from printed, published and web-based NRF materials, as well as from numerous, fairly structured interviews with both employees of the NRF (16 staff within the Evaluation Centre and associated senior managers) and with (named) external stakeholders including chairs and members of assessment panels (13 in all), as well as (unnamed) Designated Institutional Authorities (DIAs, 14 institutions listed). Written material included various NRF web-based documents, as well as a host of internal process documents from the Evaluation Centre (EC), as well as externally-disseminated guidelines for applicants, for example. Numerous documents are listed as references, and several are also cited in the body of the report.

The report maps 13 steps in the process stream from the point of invitation for resubmission or the receipt of evaluation applications (it is not quite clear which of these is chosen as the starting point) to the management and finalisation of appeals; this is a fairly comprehensive spread, and in our view the exclusion of the step involving the publication of a general call for rating applications is an immaterial oversight since the rating and evaluation system is well-known among its target...
audience at least in the universities. Next it attaches priorities (low, medium, high) to these steps on the basis of two criteria, the first being

- the extent to which the specific process is considered to “cause problems based on initial discussions with management”, (p. 17, R2.5) the second being

- each step’s ranking within the “Keen Process Matrix”. (p. 18, R2.5)

Although no details are given in regard to the latter, it is presumed to refer to what is called the “Salience/Worth Matrix” as developed in the book *The Process Edge* by Peter Keen.¹⁵ No details are given on how the ranking was achieved, and in the context of the report this issue did not seem sufficiently material to warrant obtaining further details from the author of this report. Of the 13 steps, three are classified as low, two as medium, and eight as high priority.

Partly on the basis of this ranking, four steps or functions are identified as “core processes”, another three as “support processes”:

### Core processes
- Screening of applications
- Assessment of applicants
- Feedback to applicants
- Appeals process

### Support processes
- Appointment of reviewers
- IT system
- Promotion of rating system

Of these seven items, four had been classified as high priority and one as medium, while two of the “support processes” (IT System and Promotion of Rating System) had not been included in the analysis of priorities, and it is therefore not clear how they were identified; in any event, it is doubtful whether they constitute a “process” within the rating and evaluation system, though their potential impact on the efficiency of the system is obvious.

Subsequent to this analysis, the report then deals in more detail with these seven items in turn, for each identifying strengths, weaknesses and suggested improvements. The process “Screening of applications” is reviewed in two categories (screening by the nominating institutions and screening by the NRF), as is that of “Assessment of applicants” (rating of peer reports by Specialist Committees, and rating decision by Assessment Panels). The discussion of the key processes identified in several instances includes presentations of relevant statistics captured by the EC.

Mostly the analysis is qualitative but at several points the report lists quantitative data relating to EC statistics.

---

Finally, the report concludes by summarising the key conclusions already presented during the process analysis for efficiency, and strength and weaknesses. Several appendices include: the brief for the report; documents and references; a list of interviewees and institutions consulted, and the questionnaires utilised for interviews with EC staff, Assessment Committee members, DIAs and IT managers; Swim-lane diagrams for the process, covering responsibilities and functional operations of seven different structures (DIAs, EC, Specialist Committees, Reviewers, Assessment Panels, Executive Evaluation Committee, Appeals Committee); and the NRF definitions attached to the different rating categories.

The author prudently lists upfront several limitations that constrained the content and scope of his report (p. 19, R2.5):

- (initial) confusion about the processes to focus on;
- limited sample size of interviewees in some cases;
- the timing of the report during the mid-year holidays;
- no interviews with rated researchers; and
- no contact with the steering committee as a result of the resignation of the appointed “driver”.

However, several additional and inherent constraints in the approach and analysis undertaken in this report are evident. Firstly, it is not clear on what basis the author identified the respective strengths and weaknesses associated with the seven key steps in the evaluation process that he focused on: of course, to a large extent this is informed by subjective and qualitative evaluation, but the evaluation itself must be informed by one or more inputs such as comments by interviewees, statistical information, fundamental characteristics of the process, and so on – in no case is it clear what precisely informed the author in his conclusion about the respective strengths and weaknesses. Several EC statistics are listed, but their relevance to the evaluation of the strengths and weaknesses is seldom clear. On several occasions the report includes phrases like “no weaknesses … were mentioned”, (p. 33, R2.5) which inevitably begs the question “by whom?”; unfortunately this phrase also suggests that strengths and weaknesses listed in the report were often those proffered by interviewees – the extent to which they result from critical and independent reflection by the author is not clear.

Secondly, in several instances there exists a confusion between inherent weaknesses or strengths within the processes (controlled by the NRF), and (external) risks associated with them. For example, the fact that institutions often do not have sufficient time to screen applications because
applicants submit their request late is a risk of the system, not an inherent weakness of the processes – if everyone did what they were supposed to, the system might function well. Similarly, the fact that good training is provided to new members of the EC is not an inherent strength of the screening system within the EC.

Thirdly, although several quantitative data are presented in the analysis of the crucial steps of the rating process, they are never used as a measure of efficiency; this is understandable to a point because there are no comparable ready benchmarks, but one would have expected at least some subjective evaluation of the quantum of the data presented as an indicator of efficiency. Little effort seems to have been made in extracting really useful and applicable data, a specific case in point being the report’s mention of the financial cost of the rating system – it acknowledges that this goes considerably beyond the R2.3m operating costs of the EC, but shrinks from trying to estimate those additional (external) costs on the basis they would need to be obtained from the institutions. The latter is true for an exact cost estimate, but a reasonable ball-park figure could have been obtained by assuming an average number of specialist/peer reviewers per rating application (say 4), multiplying it by an average amount of time spent on reviewing the applications (say 10 hours) and an average salary cost (say R300 per hour), and adding to it a similar estimate for the time costs of panel members. Proceeding in this way the possible external costs of the system will very quickly escalate in excess of R5m – this consideration might have impacted on the author’s conclusions.

Another example of inadequate interrogation of quantitative data can be found with regard to the average time lag between the closing date for applications and the announcement of the outcome of a rating application, which in 2006 constituted 8.4 months. There is no attempt to assess the historical trend in this number (the NRF should have a good 15-20 years’ worth of data), nor to establish what portion of this lag lies within and outside the NRF EC. In view of these outstanding questions, it is therefore difficult to understand how Madikizela can conclude that “the overall cycle time of about 8.4 months is acceptable” (to whom and compared to what?). (p. 7, R2.5)

Fourthly, although there certainly is much anecdotal evidence that the internal administrative processes in the EC are efficiently managed16 (this being confirmed by our own experience of the EC’s operations), the approach taken in this report insufficiently probes those parts of the rating system that operate outside – and, importantly, beyond the control of – the EC. Inasmuch as the sustainability of the rating system is critically dependent on the support and buy-in of the country’s researchers and their ability to sustain it, their almost complete absence in this study constitutes a significant constraint to the usefulness of its findings.

16 See, e.g., also Lombard’s reference to “professional work” in Section 7.2, on page 42 of Report 2.2.
3.5.3 **Key findings and conclusions of report**

The table below lists the main strengths and weaknesses identified for the key procedural steps analysed in this study, as well as relevant recommendations on how to improve the performance of the steps. For reasons of space, the descriptions used in the table below are generally shorter than those in the report, and a few of the purported strengths or weaknesses listed as such in the report have been excluded from this table, either because they arise from anecdotal observations or because they constitute external risks, not internal procedural ones (a point raised again below).

At a conceptual level, however, the account below seeks to reflect as accurately as possible the content of the original report in order to provide those who have not read it as direct an impression of its content as possible.

**Table 11: Strengths and weaknesses in key procedural steps**

<table>
<thead>
<tr>
<th>Process step</th>
<th>Strength</th>
<th>Weakness</th>
<th>Recommendation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Screening of applications</strong></td>
<td><strong>- by institutions</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Effective filtering mechanism</td>
<td>• Late submission leads to insufficient</td>
<td>• Applications should be opened earlier</td>
</tr>
<tr>
<td></td>
<td>• Active participation increases knowledge</td>
<td>screening time</td>
<td>• Application form should be shortened</td>
</tr>
<tr>
<td></td>
<td>of system</td>
<td>• No standardised process</td>
<td></td>
</tr>
<tr>
<td><strong>- by EC staff</strong></td>
<td>• Second quality step</td>
<td>• High workload, especially involving</td>
<td>(No recommendations made)</td>
</tr>
<tr>
<td></td>
<td>• Rigorous internal processes and procedures</td>
<td>training</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Competent staff</td>
<td>• Inherently cumbersome process</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Good training provided</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Appointment of reviewers</strong></td>
<td>• Input by candidates on selection of</td>
<td>• Time-consuming</td>
<td>• Poor performing reviewers should be excluded</td>
</tr>
<tr>
<td></td>
<td>reviewers</td>
<td>• Reviewer fatigue</td>
<td>from NRF database</td>
</tr>
<tr>
<td></td>
<td>• Detailed guidelines</td>
<td>• Small pool of reviewers for highly</td>
<td>• Choose reviewers from among those citing work</td>
</tr>
<tr>
<td></td>
<td>• Specialists’ own network</td>
<td>specialised candidates</td>
<td>by the candidate</td>
</tr>
<tr>
<td></td>
<td>complements EC’s</td>
<td>• NRF database for identifying reviewers</td>
<td>• List of “citors” to be supplied to Specialist</td>
</tr>
<tr>
<td></td>
<td>• Mature process for identifying reviewers</td>
<td>not tailored to areas of applications</td>
<td>Committees as potential reviewers</td>
</tr>
<tr>
<td></td>
<td>• Bias reduced through international</td>
<td>• Difficult to find reviewers for early-</td>
<td>• Compensation for reviewers</td>
</tr>
<tr>
<td></td>
<td>spread of reviewers</td>
<td>career applicants</td>
<td></td>
</tr>
</tbody>
</table>

(No recommendations made)
<table>
<thead>
<tr>
<th>Process step</th>
<th>Strength</th>
<th>Weakness</th>
<th>Recommendation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assessment of applicants</td>
<td>• Screening of reviewers’ reports</td>
<td>• Variable quality of reports</td>
<td>• Feedback to applicants</td>
</tr>
<tr>
<td>- rating of peer reports by</td>
<td>• Sufficient number of usable reports</td>
<td>• Reviewers not reading guidelines</td>
<td>• Reviewers to suggest rating categories</td>
</tr>
<tr>
<td>Specialist Committees</td>
<td>• Good training of specialist committee members</td>
<td>• Assessments without motivations</td>
<td>• Greater quantification of the evaluation</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Excessive bias in reports</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Mismatch between reviews and expected key words</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(None mentioned)</td>
<td>(No recommendations made)</td>
</tr>
<tr>
<td>Feedback to applicants</td>
<td>• Applicants assessed in two stage process</td>
<td>(None mentioned)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Composition of panels add credibility</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Floating assessor ensures consistency across different disciplines</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Appeals Process</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- appeals against Assessment</td>
<td>• Different committee to that which made the rating decision</td>
<td>(None mentioned)</td>
<td>(No recommendations made)</td>
</tr>
<tr>
<td>Panels</td>
<td>• Consistency assured by cross representation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- appeals against EEC decisions</td>
<td>• EEC constitutes a third committee</td>
<td>(None mentioned)</td>
<td>(No recommendations made)</td>
</tr>
<tr>
<td></td>
<td>• No legal flaws evident in appeals process</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IT Support Processes</td>
<td>(None mentioned)</td>
<td>• Not user-friendly</td>
<td>• A new rating management tool is under development</td>
</tr>
<tr>
<td>Promotion of the rating system</td>
<td>• Multiple marketing channels</td>
<td>• Lack of interest by staff</td>
<td>• Promotion needs to draw more strongly on Research Offices</td>
</tr>
<tr>
<td></td>
<td>• Direct approaches to institutions</td>
<td>• Lack of marketing strategy</td>
<td>and rated researchers</td>
</tr>
<tr>
<td></td>
<td>• Relevant training</td>
<td></td>
<td>• Perception survey should be conducted</td>
</tr>
<tr>
<td>Cost of the rating process</td>
<td>(None mentioned)</td>
<td>(None mentioned)</td>
<td>• An overall cost estimate should be obtained.</td>
</tr>
</tbody>
</table>

Essentially the report concludes that the processes supporting the rating system are mostly efficient and effective, although a few modifications are recommended in some cases. Perhaps the recommendations will prove useful in strengthening the existing system insofar as EC-internal procedures apply, but given the methodological shortcomings outlined above, and the less
intensive scrutiny of processes outside the NRF (and the consequent uncertainties about the efficiencies and sustainability of the system as a whole), the usefulness of these limited modifications may well be very limited; this point is discussed in more detail below.

This brief was a difficult one to service, requiring a familiarity with administrative processes that is not readily available, and complicated also by the absence of readily obvious benchmarks. However, more rigorous reflection on available anecdotal and statistical information may have yielded deeper and more original insights. For instance, the 95%-plus concurrence of assessment panel members (evaluators) on the rating recommended for applicants is excellent, (p. 6, R2.5) and may indeed indicate extremely precise guidelines as suggested in the report – but it could also be the result of pre-screening by the Evaluation Centre (EC) which cuts out about 1/3 of postal reviews (p. 5, R2.5). In a few instances it is conceivable that this highly effective pre-screening and exclusion of apparent outliers and poor returns may reject a review that provides a significantly different perspective on a specific application/ant that could add considerable value to the discipline or knowledge in general, but in the (likely) absence of such discipline-specific skills within the EC this point would be lost; perhaps all excluded reports should also be tabled at the Assessment Committee meetings, merely to rule out such a possibility.

Another example illustrating the inherent constraints of too narrow a focus on EC-internal processes relates to the report’s conclusions relating to the process “Appointment of reviewers”: among others, the report notes the constraint of “reviewer fatigue” and also refers to “a lack of commitment from the specialist committees” but nonetheless concludes that “a sufficient number of reviewers can be obtained and relied upon” (p. 5, R2.5). This conclusion is very important in that it relates fundamentally to the sustainability of the rating system. Unfortunately, though, the level of reviewer fatigue and lack of commitment is not explored or quantified in any way, nor is any attempt made to critically evaluate the historic level of effort required by the EC to secure sufficient reviewers: is an increase in reviewer fatigue (in what is, after all, a fairly small National System of Innovation) coupled to increasing difficulty in securing reviewers? And what about perceptual issues relating to the extent of conceptual support across the research community? While it may not have been practically feasible within the brief for this report to have addressed these issues, the implication is

(a) the conclusions of the report need to be dealt with circumspectly, and

(b) perhaps further enquiries should be considered by the NRF prior to the finalisation of decisions about the future of the rating system.
This investigation into the processes that sustain the rating system suffers from two major shortcomings – the first is that processes outside the EC were not probed at much depth, and the second is that there are no obvious benchmarks against which to evaluate the performance of those processes, especially those internal to the EC. The latter constraint is intrinsic to the task set the author of the report, and although the absence of external benchmarks provides a significant conundrum, the report does not illustrate significant attempts to address it. For this reason too, therefore, it is difficult to assess the performance of the entire scope of processes and we have already suggested that the assessment of the rating system’s cost effectiveness has been insufficiently elaborated.

In conclusion, therefore, the Madikizela Report’s usefulness to the evaluation of procedures supporting the rating system is significantly limited by intrinsic and methodological constraints.
Appendices

A) List of abbreviations and acronyms

ARC Agricultural Research Council
CNRS Centre National de la Recherche Scientifique (National Scientific Research Centre), France
CONACYT National Council on Science and Technology, Mexico
COREs Centres of Research Excellence
CREST Centre for Research on Science and Technology
CSD Centre for Science Development
CSIR Council for Scientific and Industrial Research
DARPA Defense Advanced Research Projects Agency
DEFRA Department for Environment, Food and Rural Affairs
DFG Deutsche Forschungsgemeinschaft
DIA Designated Institutional Authority
DoE Department of Education
DST Department of Science and Technology
EC Evaluation Centre
ERI External Research Income
ESRC Economic and Social Research Council, UK
FRD Foundation for Research Development
GEAR (Macroeconomic strategy for) Growth, Employment and Redistribution
HDIs Historically Disadvantaged Institutions
HESA Higher Education South Africa
HSRC Human Sciences Research Council
IDRC International Development and Research Centre
ISI Institute of Scientific Information
MCDM Multi-Criterion Decision-Making Model
MRC Medical Research Council
MRSP Main Research Support Programme
NFs National Facilities
NRF National Research Foundation
NSI National System of Innovation
NWO De Nederlandse Organisatie voor Wetenschappelijk Onderzoek (Netherlands Organization for Scientific Research)
PBRF Performance-Based Research Fund
RAE Research Assessment Exercises
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>RDC</td>
<td>Research Degree Completions</td>
</tr>
<tr>
<td>RGD</td>
<td>Research Grants Division</td>
</tr>
<tr>
<td>SET</td>
<td>Science, Engineering, and Technology (sometimes including Health)</td>
</tr>
<tr>
<td>SETI</td>
<td>Science, Engineering and Technology Institutions</td>
</tr>
<tr>
<td>SNI</td>
<td>National Researchers’ System, Mexico</td>
</tr>
<tr>
<td>SSHL</td>
<td>Social Sciences, Humanities, and Law</td>
</tr>
<tr>
<td>SU</td>
<td>Stellenbosch University</td>
</tr>
<tr>
<td>TEC</td>
<td>Tertiary Education Commission</td>
</tr>
<tr>
<td>THRIP</td>
<td>Technology and Human Resources for Industry Programme</td>
</tr>
<tr>
<td>TUT</td>
<td>Tshwane University of Technology</td>
</tr>
<tr>
<td>UCT</td>
<td>University of Cape Town</td>
</tr>
<tr>
<td>UKZN</td>
<td>University of KwaZulu-Natal</td>
</tr>
<tr>
<td>UP</td>
<td>University of Pretoria</td>
</tr>
<tr>
<td>UW</td>
<td>University of the Witwatersrand</td>
</tr>
<tr>
<td>WRC</td>
<td>Water Research Commission</td>
</tr>
</tbody>
</table>
**B) Terms of reference of review**

The section below outlines the Terms of Reference for this synthesis report.

3. **Consolidated report**

The Review Steering Committee would now like to appoint a service provider to write a report consolidating the findings of the five individual studies listed above. This consolidated report will be submitted to the Review Steering Committee for consideration prior to finalisation and submission to HESA and the NRF. The review is scheduled for completion in September/October 2007.

The final consolidated report should include the following:
- an executive summary
- background to the review and the five studies
- questions addressed in each of the five studies
- approaches and methodology used in each of the five studies
- key findings of each of the five studies
- conclusions
- appendices containing, for example, the terms of reference of the review, lists of acronyms, etc

Please note: The service provider/person selected for the task will be given an opportunity to discuss in more detail the parameters of the consolidated report.

4. **Deliverables and timeframes**

4.1 The service provider selected for the consolidated report will be required to

- give verbal feedback on key issues arising from the consolidated report to members of the Review Steering Committee if requested;
- liaise with the service providers conducting the five individual studies if clarity is needed on aspects of their individual reports;
- submit a draft report two weeks before the due date of the final consolidated report;
- give a presentation on the final draft consolidated report; and
- submit the final consolidated report by the mutually agreed date, in September/October 2007.

4.2 It is estimated that the compilation of the consolidated report will take about four weeks.