

A FRAMEWORK FOR THE EVALUATION OF RESEARCH IN SOUTH AFRICAN
HIGHER EDUCATION INSTITUTIONS – Conceptual and methodological issues

Mochaki Deborah Masipa

Dissertation submitted for the degree, Doctor of Philosophy
(Social Science Research Methodology)
University of Stellenbosch



Supervisor: Professor Johann Mouton

December 2010

Declaration

I declare that the work contained herein and submitted electronically is my work for which I own copyright, unless otherwise explicitly stated and that I have not previously submitted the entire work or part thereof to obtain any qualification.

Signature..... *Date*.....

Copyright © 2010 Stellenbosch University
All rights reserved

Abstract

This study aimed at establishing whether or not an integrated and appropriate system exists for the evaluation of research in the South African higher education system. As background to the assessment of research in South African higher education, models of research evaluation from other countries were reviewed and served as reference to the discussions on the local efforts. In each case the higher education research systems were reviewed, including existing efforts of research evaluation that exists alongside the systems. The review followed a pattern that focuses on areas including the history and rationale, purpose (s) for research evaluation, political/transformation contributions and methodological issues for a clearer understanding of the contributions made by the efforts. The study followed a multiple-case study approach to review the models and the South African situation, with the local research evaluation efforts embedded within the study of South Africa as a case.

Five themes guided the reviews that were apparent for the final discussions of the study: the rationale and purpose of research evaluation, units of analysis used in the evaluation, dimensions/criteria used in research evaluation, governance and management of research evaluation processes and methodological issues related to research evaluation. The study revealed that none of the fragmented South African research evaluation efforts is suitable to deal with the transformation requirements expected of higher education institutions. This is mainly because of the voluntary nature of the current initiatives and their focus on the lowest level of units of analysis – the individual researcher. The one effort that would be better suited to meet the transformation imperatives – the HEQC institutional audits - does not concentrate on research exclusively but collectively addresses all core activities in institutions, reducing the attention necessary for research evaluation to make a meaningful contribution to higher education research.

The study suggested a comprehensive design for the framework of South African research evaluation. The purpose identified for the envisaged exercise is the development and improvement of quality research of international standards across the system of higher education in order for research to make meaningful contributions to national demands. Programmes/departments in the higher education institutions are suggested as the units of analysis in which quality, productivity, relevance and viability serve as criteria for evaluation.

Opsomming

Hierdie studie poog om vas te stel of 'n geïntegreerde en toepaslike stelsel bestaan vir die evaluering van navorsing in die Suid-Afrikaanse hoër onderwys stelsel. As agtergrond tot die beoordeling van navorsing in Suid-Afrikaanse hoër onderwys, word 'n oorsig verskaf van die modelle van navorsing evaluering van ander lande. Dit het gedien as verwysing vir die besprekings oor die plaaslike pogings. In elke geval is 'n oorsig gebied van die hoër onderwys navorsingstelsels, insluitend die bestaande pogings tot navorsing evaluering. Die oorsigte fokus op gebiede soos die geskiedenis en die rasionaal, doel van navorsing evaluering, politiese / transformasie bydraes en metodologiese vraagstukke vir 'n beter begrip van die bydraes wat gemaak word deur die pogings. Die studie volg 'n meervoudige gevallestudie benadering tot die modelle en die Suid-Afrikaanse situasie, met die plaaslike navorsing evaluering pogings onderliggend in die Suid-Afrikaanse gevallestudie.

Die oorsigte word gelei deur vyf temas: die rasionaal en doel van die navorsing evaluering, eenhede van analise wat gebruik word in die evaluering, dimensies / kriteria wat gebruik word in navorsing evaluering, beheer en bestuur van navorsing, en metodologiese evalueringsprosesse kwessies met betrekking tot navorsing evaluering. Hierdie temas is duidelik in die finale bespreking van die studie. Die studie het aangetoon dat nie een van die gefragmenteerde Suid-Afrikaanse navorsing evaluering pogings geskik is om die transformasie verwagtinge van hoër onderwys instellings te hanteer nie. Dit is hoofsaaklik as gevolg van die vrywillige aard van die huidige inisiatiewe en hul fokus op die laagste vlak van die eenhede van analise - die individuele navorser. Die een poging wat beter geskik sou wees die transformasiedoelwitte te ontmoet - die HEQC institusionele oudits - konsentreer nie uitsluitlik op navorsing nie, maar spreek gesamentlik alle kern aktiwiteite in instellings aan. Dit verminder die aandag wat nodig is vir navorsing evaluering om 'n betekenisvolle bydrae te lewer tot hoër onderwys navorsing.

Die studie stel 'n omvattende ontwerp voor vir die raamwerk van Suid-Afrikaanse navorsing evaluering. Die doel wat vir die beoogde oefening geïdentifiseer word, is die ontwikkeling en verbetering van die kwaliteit navorsing van internasionale standaarde oor die stelsel van hoër onderwys sodat die navorsing betekenisvolle bydraes kan lewer tot die nasionale vereistes. Programme / departemente in die hoër onderwys instellings word voorgestel as die eenhede van analise waarin gehalte, produktiwiteit, relevansie en lewensvatbaarheid dien as kriteria vir evaluering.

Table of Contents

Declaration	i
Abstract	ii
Opsomming	iii
List of figures	vii
List of tables	viii
List of abbreviations	x
PART ONE: INTRODUCTION AND BACKGROUND	1
Chapter 1 Introduction	2
1.1 The need for R&D	2
1.2 Statement of the problem	10
1.3 Aims and Objectives of the Study	10
1.4 Value of the Study	12
1.5 Chapter Outline and Break-down	13
Chapter 2 Research evaluation of university research	15
2.1 Introduction	15
2.2 Explaining / Understanding the research system	15
2.3 Origin and background of research evaluation	20
2.4 Roles and purposes of research evaluation	23
2.5 Units of analysis in research evaluation	26
2.6 Dimensions of research evaluation/research criteria	29
2.7 Methodological issues	32
2.8 Types of research evaluation methods	42
Chapter 3 The methodology of this study	46
3.1 Introduction	46
3.2 Study design	48
3.3 Models and cases selection	49
3.4 Methods of data collection/information gathering	54
3.5 Analysis	57

PART TWO:	INTERNATIONAL MODELS OF RESEARCH EVALUATION	60
Chapter 4	The Dutch model of research evaluation	61
4.1	Background	61
4.2	The Evolution of Research Evaluation in The Netherlands	61
4.3	Conclusion	87
Chapter 5	The United Kingdom (UK) model of research evaluation	93
5.1	Introduction	93
5.2	Historical background	93
5.3	The evolutionary process	96
5.4	Purpose of Research Evaluation (of the research assessment exercise)	97
5.5	Research Funding and Assessment	98
5.6	The research evaluation exercises	98
5.7	The introduction of performance indicators	122
5.8	Conclusion and summary	123
Chapter 6	The New Zealand research assessment exercise	137
6.1	Introduction	137
6.2	Background of the exercise	137
6.3	Purpose of the evaluation exercise	140
6.4	Methodology and strategies used in the assessment	143
6.5	Problems experienced	147
6.6	Conclusion	149
6.7	Summary	150
PART THREE:	SOUTH AFRICAN RESEARCH EVALUATION SYSTEMS	154
Chapter 7	The South African system of higher education research	156
7.1	Introduction	156
7.2	The history of the South African HE system and research (the old order)	157
7.3	The post-apartheid era	163
7.4	Policy and Procedures for Measurement of Research Output of Public Higher Education Institutions (DoE, 2003)	190
7.5	Summary	193
Chapter 8	The FRD/NRF rating system	203
8.1	Introduction	203
8.2	Historical background	203
8.3	The introduction of a rating system	206

8.4	Influences of the new dispensation	208
8.5	The purpose attached to the rating of researchers	211
8.6	Methodology for the evaluation and rating system	212
8.7	A critical look at the system	216
8.8	Advantages and challenges	218
8.9	Observations, implications and conclusions	221
8.10	Summary	223
Chapter 9	The Council for Higher Education – evaluation activities	228
9.1	Introduction	229
9.2	Background	231
9.3	Conceptualizing quality	232
9.4	The HE audience and other role-players	234
9.5	The CHE approach to evaluation (all HE functions included)	235
9.6	Synthesis of CHE/HEQC activities	256
Chapter 10	Discussions and Syntheses	264
10.1	Introduction	264
10.2	Synthesis of the three international models	264
10.3	Synthesis of the South African efforts of research evaluation	291
Chapter 11	Conclusions and Recommendations	320
11.1	Rationale and purpose of research evaluation	320
11.2	Units of analysis	322
11.3	Dimensions/criteria for research evaluation	323
11.4	Governance and management of research evaluation exercise	324
11.5	Methodology	326
11.6	Reflections and synthesis of the recommendations	328
Reference		330
Appendix 1	SAPSE 110 formula	341

List of figures

Figure 2.1	Placement of countries in relation to aggregation and/or steering of research	20
Figure 2.2	Units of analysis used in research evaluation	28
Figure 3.1	The framework used in the study displaying the three tiers from bottom to top	48
Figure 3.2	The multiple-case design with one case showing embedded cases	52
Figure 4.1	Summary of a cycle of events in the CF era	64
Figure 4.2	Summary of the VSNU cycle of events of assessment	73
Figure 5.1	A sketch of the input/output influence – a cycle of events as understood by the researcher	102
Figure 7.1	Indications of a decline in research publications in some top publishing universities in South Africa	176
Figure 9.1	Visual presentation of the audit process	246
Figure 9.2	Summary of a quality improvement sequence as interpreted by the researcher	248
Figure 10.1	An imbalance created by two opposing national imperatives	302
Figure 11.1	Indications of the interdependence of the national imperatives	321
Figure 11.2	Suggested structure for the management of the evaluation exercise	325

List of tables

Table 2.1	A summary of the types of research evaluation purposes	25
Table 2.2	A summary of the methods regularly used in research Evaluation	42
Table 4.1	A summary of the objectives in the evolutionary stages	88
Table 4.2	A summary of benefits and challenges	89
Table 4.3	A summary of processes of research evaluation	91
Table 5.1	Differences in participation between the two exercises in the UK (1996 and 2001)	118
Table 5.2	Dates and historical activities related to system initiation	124
Table 5.3	Objectives and processes in the RAE exercise	126
Table 5.4	Benefits/best practices and challenges/lessons learnt in the RAE exercise	130
Table 5.5	Differences and similarities between the Dutch and the UK models	134
Table 6.1	Summary of benefits and challenges of the exercise	153
Table 7.1	Enrolment discrepancies between HBUs and HWUs in 1992	158
Table 7.2	Categories (types) of research summarized according to SAPSE 110	161
Table 7.3	Distribution of research among universities in South Africa	163
Table 7.4	Allocations for the South African HEIs triennium, 2006/07 to 2008/09 in the new funding framework	170
Table 7.5	Governance structures in the new dispensation	195
Table 7.6	Summary of the old and new dispensation of research in HE	197
Table 7.7	A summary of challenges and solutions in the new dispensation	199
Table 7.8	A summary of the research reward system	202
Table 8.1	Classification of the rating categories	214
Table 8.2	A summary of the purpose/objectives, methods, benefits and challenges of the rating system	226
Table 9.1	A summary of the CHE programme of evaluation	260
Table 9.2	A summary of the outcomes and intended benefits of the CHE evaluation process	262
Table 10.1	A summary of the reasons for evaluation for the study models	273

Table 10.2	Criteria for research evaluation as used in the reviewed models	278
Table 10.3	Impressions of purposes of research evaluation – South African systems	306
Table 10.4	Criteria of research as used in the South African research evaluation systems	309
Table 10.5	Units of analysis of the South African research evaluation system	312

List of abbreviations

CF	Conditional Funding (Dutch)
CHE	Council for Higher Education
CHE M&E	Council for Higher Education monitoring and evaluation
CHEPS	Centre for Higher Education Policy Studies (Dutch)
Crest	Centre for Research on Scientific and Technology
CSD	Centre for Social Development
CSIR	Council of Science and Industrial Research
DES	Department of Education and Science (UK)
DENI/DEL	Department of Employment and Learning – Northern Ireland
DHET	Department of Higher Education and Training
DoE	Department of Education
DoL	Department of Labour
DST	Department of Science and Technology
FRD	Foundation for Research Development
FTE	Full-time Equivalent
GUF	General University Funds(ing)
HAIs	Historically Advantaged Institutions
HBO-raad	Hoger Beroeps Onderwijs raad
HBU's	Historically Black Universities
HDI	Historically Disadvantaged Institutions
HE	Higher Education
HEFC	Higher Education Funding Council (UK)
HEFCE	Higher Education Funding Council for England
HEFCW	Higher Education Funding Council for Wales
HEIs	Higher Education Institutions
HEMIS	Higher Education Management Information System
HEQC	Higher Education Quality Committee
HERO	Higher Education and Research Opportunities
HESA	Higher Education South Africa
HSRC	Human Science Research Council
HWUs	Historically White Universities

IHO	Inspectorate of education (translated from Dutch)
ISI	Institute for Scientific Information
JCR	Journal Citation Report
KNAW	Koninklijke Nederlandse Academie der Wetenschappen(Royal Academy of Arts and Science)
NFs	National Facilities (SA)
NRF	National Research Foundation
NWO	Nederlandse organisatie voor Wetenschappelijk Onderzoek(Netherlands Organization for Scientific Research)
NZ	New Zealand
OECD	Organization for Economic Co-operation and Development
OR	Operations Research
PBRF	Performance-based Research Funding (NZ)
PIs	Performance Indicators
PPMROPHEC	Policy and Procedures for Measurements of Research Output of Public Higher Education Institutions
R&D	Research and Development
RAE	Research Assessment Exercise (UK)
RIS	Research Information System
SA/RSA	Republic of South Africa
SAAIR	South African Association for Institutional Research
SAQA	South African Qualification Authority
SAPSE	South African Post-secondary Education
SHEFC	Scottish Higher Education Funding Council
Stats SA	Statistics South Africa
TEAC	Tertiary Education Advisory Commission (NZ)
TEC	Tertiary Education Commission (NZ)
TEOs	Tertiary Education Organizations (NZ)
UFC	University Funding Council (UK)
UGC	University Grants Committee (UK)
UK	United Kingdom
US	United States of America
VSNU	Vereniging van Samewerkende Nederlandse Universiteiten
WHW	Wet Hoger Onderwijs

PART ONE

INTRODUCTION AND BACKGROUND

Chapter 1 Introduction

1.1 The need for R&D

The evaluation of performance in higher education (HE) has become a global necessity (Stolte-Heiskanen, 1992). Such evaluation exercises mainly target public higher education institutions (HEIs), which governments in different countries regard as national responsibilities that have to be legislated upon and monitored. The focus is usually on the evaluation of the core functions of universities, of which research is one component.

As academic institutions that are responsible for the training of researchers and for undertaking research aimed at knowledge production, universities form a key component in the research productivity of a country. In South Africa for example, universities produce approximately 87% of the country's public research stock (Mouton *et al*, 2007). University academics also contribute to various public service programmes (consultations and contracts) and community engagement initiatives.

The HE system is central to the development and sustenance of research activities; this is even truer in developing countries where the university sector is the main producer of science. For this reason and others (to be discussed), research activities have to be constantly monitored and assessed so as to reflect on the level and extent of participation and the extent to which they contribute to the body of knowledge.

In most countries, the majority of universities depend on public funds. Campbell (Shapira and Kuhlmann, 2003) argues that research funded from general university funding (GUF) or a block grant has to be accounted for. If this is not done, such funds are vulnerable to “abuse” such as being deferred (used) for other operations in the institutions. Campbell therefore suggests a system of evaluation as an accountability strategy. The type of research supported by GUF is said to be curiosity-driven and basic; it makes a significant contribution towards the creation of a knowledge-based society. According to Campbell (Shapira and Kuhlmann, 2003:102), “the evaluation of university (basic) research marks an area of strategic relevance”. Of importance is Campbell’s (in Shapira and Kuhlmann, 2003) advice that “the whole university sector should be addressed”. This not only reveals the national level of research but also exposes fields of study that need to be developed.

In South Africa, research evaluation exercises of universities have been undertaken. Most of these evaluations, however, have been sporadic and commissioned for a particular situation. Examples include the study by Subotzky in 1997 on “The enhancement of graduate programmes and research capacity in historically Black universities” and that of Mouton and Gevers (2010) on the “R&D Evaluation: An overview of concepts, approaches and methods”. These studies unveil research activities carried out during a particular period. However, without a systemic and comprehensive evaluation process, there is no guarantee that a system of accountability will exist, or that a progressive improvement and development of university research will be achieved.

Research evaluations which are conducted occasionally are referred to as situational and are said to be common in developing countries especially where funding of research is allocated on a project competitive basis (Geuna and Martin, 2001). Campbell (Shapira and Kuhlmann, 2003) explains the tendency of situational research evaluation to evolve into a systemic/comprehensive type as conditions improve. This suggests that even though accountability may be conceived in a rather narrow manner, for example at a “case” level, this should be done across all institutions in order to build a more advanced and systemic form of evaluation.

The more regular efforts to evaluate research in South African universities are those conducted by intermediary bodies (state organs), which focus on evaluation that relates to the services they provide. This arrangement (of intermediary body participation) arises from the old system of HE in South Africa, which in the 1980s left the development and support of curiosity-driven research in the hands of the state. Pienaar *et al* (2000) noted that the Foundation for Research Development (FRD) was established in the early 1980 as a state body/organ to support HE research. The FRD developed its own strategies of research evaluation suitable for its own operations. Even though there was consultation with HEIs over time, the institutions and the entire system of HE had to fit within the FRD system. The FRD was later succeeded in the new dispensation by the National Research Foundation (NRF) to carry out the activities of promoting good research.

The NRF Institutional Review Panel (NRF, 2005) refers to the NRF as South Africa's "premier agency" because of its intensive involvement in research support. It conducts fund-driven research evaluation (based on project funding) in order to distribute state funds.

This is in line with the NRF's mission to promote and support research activities. The Foundation also promotes research capacity development, providing facilities and assisting with the retention of research expertise. Retention of skills, the Foundation states, is intended for improving the quality of the lives of South Africans. (NRF, 2002; NRF, 2005). The Foundation's mandate encompasses the entire system of research and innovation in the country, not only HE research.

The NRF commissions institutional assessment on a regular basis to measure its own performance (Krige, 2007). These include evaluating proposals for funding, evaluating the performance of national research facilities, and the rating of individual scientists.

The rating system was inherited from the FRD and was established to recognize and reward excellence wherein "quality" is regarded as the main component and is mainly measured against international standards. According to Pienaar *et al* (2000), the introduction of the "record-tracking" rating system of the FRD/NRF, following the advice of Professor Jack de Wet (a retired professor) in the early 1980s established a review culture in South Africa. The intention was to introduce a broad-based competitive atmosphere, in which researchers would aspire to achieve and maintain research excellence. This would in the long run hopefully lead to high standards of research and human resources across the HE system (Krige, 2007). Despite its setbacks especially related to its "discriminatory" treatment of previously/historically disadvantaged universities (HDIs), as cited by Krige (2007), this initiative was directly intended to improve curiosity-driven, self-initiated, basic research in universities and the entire HE system.

Positive results, including an increase in the number of 'good' researchers of international standing, have been reported to have emerged from the rating system (Pienaar *et al*, 2000). This developmental idea was nevertheless overtaken by national events which led to the crafting of a new NRF vision and mission, reducing the rigor and 'power' of the rating

system, especially in the determination of fund allocations. Until recently (2007), funding, which had been used by the FRD as an incentive to encourage researchers, was de-linked from the awarding of rates.

The other body that has been “indirectly” involved in university research evaluation activities is the Department of Higher Education and Training (DHET). The Department is the main stakeholder in public HE. During the apartheid era it supported institutions through the South African Post-Secondary Education (SAPSE) financing system. The SAPSE system made contributions through the general university funding (block grants). SAPSE administered a system of research subsidies which were based on the number of publications produced per institution (Melck, 1995).

In 2003 the Department revised the existing policy for research subsidy support - the “Policy and procedure for measurement of research output of public higher education institutions” (DOE, 2003) abbreviated in this study as PPMROPHEI

Whereas the NRF system is based on the voluntary participation of individual researchers (and later groups as accommodated by the NRF), the SAPSE system is aimed at the research efforts of a university as a whole.

With the introduction of the new system of higher education in South Africa in the mid-1990s, the White Paper 3 (DoE, 1997a: 54) commented on the existence of the fragmentation of efforts and suggested “the need for the co-ordination of research activities and their funding in higher education”. The White Paper on Science and Technology (1996:25) also suggested the introduction of a system that would co-ordinate research activities in order to reduce fragmentation, a system which, the White Paper believed, would enable the effective deployment of public resources. Although the establishment of the National Research Foundation (NRF) was meant to be the solution to the problem, the intermediary body continues to concentrate more on project specific funding and does not take responsibility for a systemic approach to research evaluation in the system.

Subsequent to the suggestions of the White Papers (White Paper 3) and the White Paper on Science and Technology), the National Plan for Higher Education (DoE, 2001) alluded to the

absence of an adequate information base to provide a clearer understanding of institutional trends and capacity constraints, the fragmented research funding and the absence of a clear national research plan. This, according to the National Plan, leads to the absence of clearly defined national research priorities. National priorities would define how HE researchers are expected to contribute through basic and innovation research.

The National Plan also recommended that the national research plan be linked with the national system of innovation (CHE, 2001), which, according to the DoE (2001:18), requires “the development of appropriate co-ordination mechanisms involving the different actors in the research system”, including HE. Consolidating the different research efforts in the HE system with a well conceptualized plan may lead to an inclusive, co-ordinated system of research evaluation in the system, which will make a unitary contribution. There is therefore an urgent need to review the system of research evaluation in South African HE with the intention of developing and implementing a more co-ordinated framework for evaluation.

More recently, the NRF Institutional Review Panel (NRF, 2005) also suggested that attention be given to the HEIs and the development of a possible model of research evaluation that would take cognizance of local circumstances of research and the needs of the HE system in the country. The panel was of the view that Higher Education South Africa (HESA) would be the appropriate structure to co-ordinate and oversee such a research evaluation exercise, with the inclusion of others as stakeholders. By implication, the HEIs or a representative body thereof are/is expected to exercise control over HE research evaluation activities and not act as stakeholders in a system for which they should be accountable.

HESA is made up of Vice-Chancellors and Principals of all South African public universities and may therefore be regarded as a representative body for public universities. The fact that HEIs in South Africa are autonomous, suggests the need for the HE system to be central to all their core functions including research and research evaluation. This may be achieved through the inclusion of other bodies as stakeholders, especially those that are already involved in similar exercises.

Different bodies/organs qualify as stakeholders by virtue of their involvement in research evaluation activities that affect or relate to HE research. The NRF, for example, is responsible for research funding and reports to the Department of Science and Technology

(DST). Government departments are also occasionally involved in research evaluation within their own operations. The fact that Departments such as Science and Technology sometimes contract university staff members to undertake their research and also to sometimes conduct their own (internal) evaluations makes the Departments suitable stakeholders. After all, all the organs focus on the same research, conducted mainly in order to address national imperatives and involve academic staff members.

The Council for Higher Education (CHE) has been assigned the responsibility of concentrating on issues of institutional and programme quality with international competitiveness as one of the main foci/intended outcomes (DoE, 1997b). In the process the CHE undertakes several quality control processes. In one of its programmes, the Council monitors and evaluates transformation trends in HEIs. The CHE also audits quality levels in individual institutions through internal evaluation and external validation by peers, and provides advice on improvements. Both processes occur across the entire system of HE and include all public HEIs. The two processes (monitoring and evaluation and the institutional audit programmes) serve as typical examples of a move towards central collation and broader/national evaluation strategies. The former system, however, deals with already available results from other researchers (secondary data) and therefore does not participate in or contribute towards planning these evaluation efforts. How the CHE evaluation systems make a contribution to a systemic development of research in HE is one of the important issues reviewed in this study. More information on the CHE's efforts is found in chapter 9 of this study

Following on the realization that all bodies/organs are meant to contribute towards research development in HE, there has been a need to analyze the similarities and differences existing between the bodies in order to find common ground. This would assist in establishing whether or not the fragmented efforts should be combined or whether the *status quo* should remain. It would also be necessary to establish whether or not a new research evaluation system would be the solution.

The existence of the different situational research evaluation efforts of different bodies/organs paints a picture of what Campbell classifies as a Type B model (the "pluralized" type, see chapter 2 for definition) observed in other countries such as Germany (Campbell in Shapira and Kuhlmann, 2003). According to Campbell, this Type exists where

different research evaluation efforts are conducted for different purposes. The author refers to this arrangement as “situational research evaluation”. If the different bodies discussed above (the NRF, CHE and the different government departments) are involved in research conducted for different purposes, then South Africa’s efforts fit the pluralized Type B model. Campbell observes that the pluralized Type has a tendency to evolve over time into a unitary system. If this is the case, it may require that plans and preparations for a systemic research evaluation system for the country be established, or, if one does exist, plans be made to improve it. Such plans or preparations need to be based on existing efforts for a suitable solution.

The suggestions made by the NRF Institutional Review Panel (NRF, 2005) should be taken into consideration. The Panel believed that the establishment of an evaluation system for HE would, while taking into account the purposes and objectives of internal systems of HE research such as scholarship and research excellence, also consider the imperatives of the country. After all, it is this scholarship that makes HE an important role player and the highest contributor to research productivity. While this study intends to develop a framework for such a system, care has been taken to leave the finer details to the different actors whose responsibility it will be to draw plans, and to implement and/or monitor the implementation of the exercise.

To lay a foundation for the South African situation, information on different models of research evaluation in HE is necessary for this study. This requires a detailed discussion of models used in other countries. Care was nevertheless taken to consider South Africa’s context within the complex situation in which evaluation has to take place. For example, the government of South Africa has been faced with the challenge of balancing international competitiveness in research against the need for transformation in the country (DoE, 2001).

In addition to national imperatives, the White Paper on Science and Technology (1996) together with the White Paper 3 (DoE, 1997a) also mention global participation and competitiveness as some of the national goals in this era of transformation. It is thus reasonable to expect that the evaluation of research in HEIs in South Africa will be inclined to, or even strive for international standards. International levels are directly or indirectly said to be used by intermediary organs and the state departments discussed above to standardize their research evaluation measures. With the South African HE system also

aspiring to acquire and maintain international standards in research, it is expected that efforts to keep up with such standards would require continuous assessment of some sort. As stated, this study was an attempt to establish the existence of, or to conceptualize a framework for such a system in case one does not exist.

If it is found desirable to establish a new system of evaluation, it would be necessary to first identify the main purpose and objectives before basing measurements on international standards. Otherwise, results may reveal one thing while the intentions meant the other. A typical example is the New Zealand experience during the country's first comprehensive research evaluation in which a multitude of purposes emerged, affecting the results of evaluation (Periodic research performance exercise report, 2004). The demands of transformation in New Zealand are very similar to conditions in South Africa at this point in time (McLaughlin, 2003). Both countries nevertheless seem to opt for international competitiveness even though they are faced with challenges in increasing the number of "world class researchers".

The other problems that relate to the establishment of the intended framework concern the existing disparities between universities. There is still a gap in research performance between the historically advantaged and disadvantaged institutions, a situation which, it seems, will persist for some time to come. The Higher Education Act 101 (DoE, 1997b) suggests that any efforts to respond to the challenge of inequalities should do so at the same time as attempting to meet international standards in order to 'kill two birds with one stone'. This study also intended to establish the extent to which the HE research evaluation efforts referred to above, as a collective or on an individual basis, have responded to the challenges with their different purposes.

The systemic/comprehensive research evaluation models of other countries alluded to previously and referred to later in this study (chapters 4, 5 and 6) have also been used for knowledge-base purposes. A comparison of the models broadened the base even further for the analysis and comparison of South African efforts and assisted in arriving at usable conclusions for the country. It was important to establish the purposes of evaluation and the methodological aspects, together with the reasons why these were selected and used by the different countries/models.

1.2 Statement of the problem

South Africa lacks a well co-ordinated and well-aligned framework and initiative for evaluating research in the HE system. The evaluation efforts that exist are manifold and situational and are not able to capture the trends of development and improvement necessary for the processes of transformation and for the determination of international standing for the HE system.

Whereas different bodies/organs such as the NRF, the CHE's Higher Education Quality Committee (HEQC) and the DoE (now DHET) engage in good practices of research evaluation, their efforts are varied, not well-aligned and do not directly reveal the internal research activities of universities or even the HE system as a whole. The result of this pluralized system is that the outcomes of each evaluation effort only serve its specific purpose. The CHE's Framework for Monitoring and Evaluation (CHE, 2004a) uses (or intends to use) the results of such situational evaluation outcomes. Meanwhile, this approach does not produce a comprehensive, easily accessible impression of the conditions of research necessary for accountability, internal regulation of research and for systemic policy making.

Therefore, there is currently no initiative that contributes directly toward the goals of a broad-based and periodic research evaluation process or towards an inclusive quality assessment/evaluation of the actual output of institutions across the system.

This background led to the formulation of the main research question of this study: Which model of research evaluation can best reveal the internal (intra-university) state and activities of research in South African universities for the purpose of systemic improvement of research across the entire system of higher education?

1.3 Aims and Objectives of the Study

This study reviews the diverse forms of HE research evaluation in South Africa. The purpose of the review was to establish whether or not these efforts are sufficient in achieving the aims of an integrated research evaluation system for HE in the country. The review therefore serves as an attempt to detect the purpose and in the absence of any, arrive at a framework that directly measures the activities of research in universities. In order to abide by the principle of best practice, it was necessary to study national models of research evaluation used in other countries and identify practices that would best suit the local conditions. The

conceptual, contextual and methodological lessons learnt from the models provided the base for the review of the South African efforts and guided the direction suggested in the recommendations of this study.

The more specific objectives of the study are therefore:

- To describe and more clearly understand the inter-relationship between the various components of HE research evaluation in the South African system. This has been done by responding to the following:
 - 1) the way different role players relate to each other ;
 - 2) the extent to which they have different evaluation purposes in mind for their exercises and
 - 3) with the result that they sometimes (even unintentionally) work against one another. This has been conducted through:
- A review of existing forms of research evaluation and review in South Africa. Some such forms are already in place – the DHET’s measurement of research output and research journal accreditation, the NRF’s rating of scientists, and the CHE’s institutional audit and its monitoring and evaluation systems. In each case the intention was:
 - To review the history and rationale behind the existing forms of research evaluation (research evaluation efforts).
 - To describe the research evaluation purpose(s) of the South African efforts.
 - To describe the methodologies and procedures and in some cases methods involved in each form.
 - To discuss the potential value and significance of the forms of research evaluation in line with transformation imperatives in higher education and training.
 - To review similarities, linkages and differences between the forms.
- To review three examples of structured models of research evaluation – The Netherlands, the UK and New Zealand – in detail in each case with the purpose being:
 - To describe the origin and rationale of each case

- To discuss in detail the purpose(s) of research evaluation in The Netherlands, the UK and New Zealand as well as the processes and procedures employed in each
 - To discuss criticisms that have been leveled against each system
 - To discuss similarities and differences and the implications thereof
 - To draw some lessons that would serve as a base in the review of South African systems of research evaluation
 - To draw lessons and make recommendations for South Africa on the future direction of research evaluation in HE
- To review the methodological aspects of research evaluation/assessment as used in the reviewed models and referred to in the literature
 - To draw lessons on methodologies for the review of the South African situation in order to make recommendations thereon
 - To make recommendations on the future of South African research evaluation

1.4 Value of the study

This study aims to describe the “pluralized” efforts of HE research evaluation in South Africa in order to develop a framework that could be used to evaluate the research activities of universities. As we have argued, there are currently different research evaluation systems and approaches at work in the South African research evaluation systems that have different origins, purposes and consequences. A clear understanding of the approaches would reveal the need for formulating a framework, and if necessary, guide the process of formulation. This was done with the intention of alerting the HE system of the need to take full responsibility for its own activities by formulating its own purpose for research evaluation and establishing ways (plans) of achieving such a purpose. Such plans would address both internal regulation of individual universities and also the activities of the entire system of HE research in the country.

It is therefore hoped that the study would lead to the establishment of a model that will assist universities to formulate intentions and measure their own performances. The HE system, on the other hand should be able to assist those institutions that are not meeting expected standards and be able to compare local achievements with those in the international arena.

After all, it is government's intention to have the HE system achieve international competitiveness.

It is envisaged that the system will engage more researchers in research evaluation and will afford them an opportunity to gain experience. When they display levels of international competence, researchers may be invited to participate in other countries, giving them even more exposure. World-class researchers may also be invited to South Africa to serve on local panels. The sharing of ideas with such skilled researchers will improve the knowledge base of research evaluation in the country.

If the HE system takes more responsibility for its own operations, the steering by government may be reduced and some level of responsibility may be shifted towards the new system, to the advantage of scholarship and the improvement of the level of scholarly contributions. If well planned and well calculated, the evaluation system may, while attempting to raise the numbers of 'good scholars', also address issues of inequalities among universities, which in turn may assist in increasing the number of researchers.

The above-stated reasons serve as drivers for this study. The motivation of a HE system of research evaluation comes at a time when some of the efforts have been in operation for a while. Although the CHE programmes are in their infancy, the DHET systems and the NRF rating system make the study easier as their processes and procedures are more visible. The establishment of this system is nevertheless overdue.

1.5 Chapter Outline and Break-down

The remaining chapters in Part one of the study are:

- | | |
|-----------|--|
| Chapter 2 | Research evaluation of university research |
|-----------|--|
- Introduction
 - Explaining/Understanding the research system
 - Origin and background of research evaluation
 - Roles and purposes of research evaluation
 - Types of research evaluation model
 - Methodological issues

Chapter 3 The methodology of this study

- Introduction
- Study design
- Models and case selection
- Methods of data collection/information gathering
- Analysis

Part Two of the study is devoted to a discussion of three international models of research evaluation.

Chapter 4 The Dutch model of research evaluation

Chapter 5 The United Kingdom model of research evaluation

Chapter 6 The New Zealand research assessment exercise

The three main current evaluation efforts in the South African system are discussed in Part Three of the thesis.

Chapter 7 The Department of Higher Education and Training

Chapter 8 The FRD/NRF rating system

Chapter 9 The Council for Higher education – Evaluation activities

The study concludes with two chapters in which the main findings of the study are consolidated (Chapter 10) and recommendations for a future integrated system are made (Chapter 11).

Chapter 2 Research evaluation of university research

2.1 Introduction

The chapter starts with a description of the research system before explaining the origin and background of research evaluation. Therefore, no explanation of the concepts of evaluation (in general) has been included other than those that relate to research evaluation. Research evaluation is then introduced through a historical overview. Thereafter, a theoretical framework for this study is proposed through the inclusion of important areas of focus; roles and purposes of research, units of analysis in research evaluation, timeframes, dimensions or criteria for evaluation and the methods used in research evaluation.

2.2 Explaining / Understanding the research system

Information on research and research systems serves as knowledge base for the study of research evaluation/assessment. This is guided by the fact that Rip and van der Meulen (1996) regard the evaluation of both entities as a way of understanding the functioning of a national innovation system of a country. In line with this understanding, Campbell (Shapira and Kuhlmann, 2003:99) emphasizes the “pivotal role” of academic research in the system of innovation, with its empirical and scientific nature. Rip (2003) agrees with Campbell that knowledge production from research and experimental development is associated with economic performance and the economic competitiveness of a country. Innovation is said to give more focus to the improvement of “the quality of life” and “policy relevance” (Rip, 2003), meaning that evaluation plays a strategic role in the system.

In an attempt to explain the research system, Rip and van der Meulen (1996) start by describing a “modern” research system and then suggest how a post-modern system may be arrived at. The authors identify four (4) important components that play an important role in a national research system; the researchers (or research community), the institutions where research is conducted, interactions and research processes (and procedures) and stress the importance of the existence of interdependence of the components to make up the research system. For example, a researcher should practice science in a scientific community (and not in a vacuum), with the intention of making a contribution to the body of knowledge.

Research and development require researchers, as a basic unit of the entire process. Campbell (Shapira and Kuhlmann, 2003) views academic institutions as places that provide a science base and science-induced activities. This makes them closely associated with the development of researchers, and thus contributing to the development of human capital.

The planning, process and even the products of such research are affected by features that may be beyond the researcher's control. According to Rip and van der Meulen (1996), these features, which are external to the researcher's environment include among others, the research institutions (including universities) and the state with their influential policies and procedures. For example, the university may require that research considers the institution's internal mission while the state may introduce incentives or punitive measures with the intention of reinforcing national priorities. This was observable during the World Wars when science and innovation gave more attention to the production of weapons, relating the role and relevance of science more to political needs. This was later interpreted by researchers as compromising the moral implications of science. In this way, science is linked to "ideologies and politics" (Arnold, (2004).

South Africa experienced a similar situation. The period 1960 to 1970 was dominated by operations research for military purposes. At the time, funding for university research was based on the general university funding (GUF) formula (SAPSE 110). In 1980, the government felt the need for the increased production of highly-skilled researchers and introduced a strategy (through the FRD) to fulfill this need (Pienaar *et al*, 2000; Krige, 2007). Coupled with such state demands has been the emergence of intermediary bodies, for example the research councils, used as mediators between the state and research organizations (universities), especially for research funding purposes.

The shift in science that involves different types of knowledge production also affected university research. This, according to Arnold (2004), led to the present interwoven nature of economic matters that relates institutional performance to other actors. Different authors attach different attributes to the shift to what Gibbons and colleagues (Pienaar *et al*, 2000) identified as two modes of science (Crest, 2001; Kuhlmann, 2003; Arnold, 2004). Mode 1 is associated with the production of basic research in institutions that rely on state funding such as universities and research councils. In this mode, knowledge production is related to some "entitlement" to state funds in which government is expected to provide what Arnold (2004)

terms “patronage”. Research in this mode is characterized by the traditional drive for quality research within a discipline, and is said to be homogeneous and hierarchical in nature.

Academic science is said to be undergoing a shift to Mode 2, which is said to be trans-disciplinary with quality control including social accountability. That is, knowledge production relates to application and those intending to apply such knowledge have to participate in the production. Care has to be taken when attempts of assigning university contributions to this mode are made. Also, whereas it is acceptable that the economy is dependent on applied science Campbell (2003:101-102) explains the importance of basic research conducted by universities as a guarantee for “long-term innovation”, making university research evaluation “an area of strategic relevance”.

The above discussion on Modes may be viewed in relation to discussions on the different funding mechanisms of universities. Different governments fund research for different reasons. These different reasons according to Arnold (2004) may compel governments to have different measures to manage the use of such funds. Arnold (2004) lists the following five roles of state-funded research:

- Developing absorptive capacity (creating a pool of scientists for employment)
- Promoting technology development (linkages between actors and systems)
- Funding of strategic research (increasing manpower for technological and economic needs)
- Funding basic research (create a pool of knowledge for strategic reasons)
- Bottleneck analysis (keeping check on the need for intervention)

These roles determine the direction to be taken when science attempts to fulfill national economic demands. Such funds may be distributed in different ways.

Campbell (2003) refers to two public funding modes for university research, the general university funding (GUF)/block grant and the earmarked funding. The GUF mode can further be divided into four types identified by Crest (2001) as follows:

- State fund allocated in part for research on the basis of evaluation outcomes
- Research allocations based on the size of an institution
- Research funds allocated on negotiation between universities and the state; and
- Allocations not directly linked to research assessment

The above funding arrangements influence the reasons behind research evaluation and provide the methodological direction to be followed in achieving such goals.

Rip and van der Meulen (1996) draw attention to a different scenario in the modern system of research that may be influenced by dependency and non-dependency on state funds. The authors introduce two dimensions of effect; that of steering and of aggregation. Through the dimension of steering the authors explain the behaviour of the state in influencing the research system with its own intentions. This, the authors indicate, is influenced through policies and procedures and the control/manipulation of government through the provided infrastructure and resources. In this way, the state acts as the principal and the audience at the same time. At the extreme end of the dimension, the principal may dominate and impose rules that the researchers (or research communities) have to abide by, such as contracts and sanctions.

Through the introduction of intermediary bodies the principal attempts to reduce dominance by mediating through the bodies. The intermediary bodies serve an ‘incentivizing’ responsibility and are expected to engage researchers/institutions through consultations. This is usually regarded by the principals as some kind of bottom-up approach to decision making. The UK system serves as a typical model in this regard (the model will be discussed later in this study). This portrays an atmosphere of what can best be expressed in Afrikaans as ‘samewerking’ (which may mean collaboration) between the two sides (principal and research institutes) even though the initiative considers (is based on) the aims of the principal. At the more moderate end of steering, the research community may be afforded more powers in which case the influence of the state becomes minimal.

Aggregation on the other hand, according to Rip and van der Meulen (1996), refers to institutional processes of research involved in “agenda building” together with infrastructure that supports such processes. It is important to note that agenda- building usually occurs in research communities, and may consider institutional aims, but may also take place at the level of intermediary bodies or even government departments themselves. In the latter instance aims and processes are determined by such government departments. Rip and van der Meulen warn of the extreme (highly institutionalized) form of aggregation in which the state’s aims may be undermined. When left to research communities, aggregation may even

be detrimental to the process it is meant to improve. For example when elitist behaviour determines the norms and standards for research competence, “young” researchers that cannot achieve the elitist status will be disadvantaged. In this way aggregation may have negative effects.

The two dimensions (of effect) usually co-exist, in which case they may exert the same effect in a system or may be influenced to operate at different levels, depending on national policies determining the research system. Rip and van der Meulen (1996) explain that when both steering and aggregation are low, competence is usually compromised through the exertion of less effort in agenda building and weakened intentions on both sides (researchers and the state). In cases where both components are at high levels, Rip and van der Meulen (1996) warn that benefits may not be feasible unless both sides (all role players) have a common interest. The authors advise that “exertion of same effect is only viable when there is no dominance by any of the role players” (state or researchers/research communities) and that this occurs “where there is both scientific (and) or societal interest” (Rip and van der Meulen, 1996:348).

In cases where there is high steering and low aggregation, in which case agenda building is surpassed by national aims, this may have long-term negative effects on research competence. For example, where high competition for resources and sanctions is imposed, aggregation may be so negatively affected that the system may have to be reviewed.

Rip and van der Meulen (1996) view a system with high aggregation and low steering as a legitimate process to allow heterogeneity in the system, and refer to such a system as a post-modern research system. The Dutch model serves as a typical example of such a system. It is believed that the system creates a conducive environment for the interdependency of role players, which is necessary for research expansion. Figure 2.1 below maps the steering/aggregation levels of countries studied by Rip and van der Meulen.

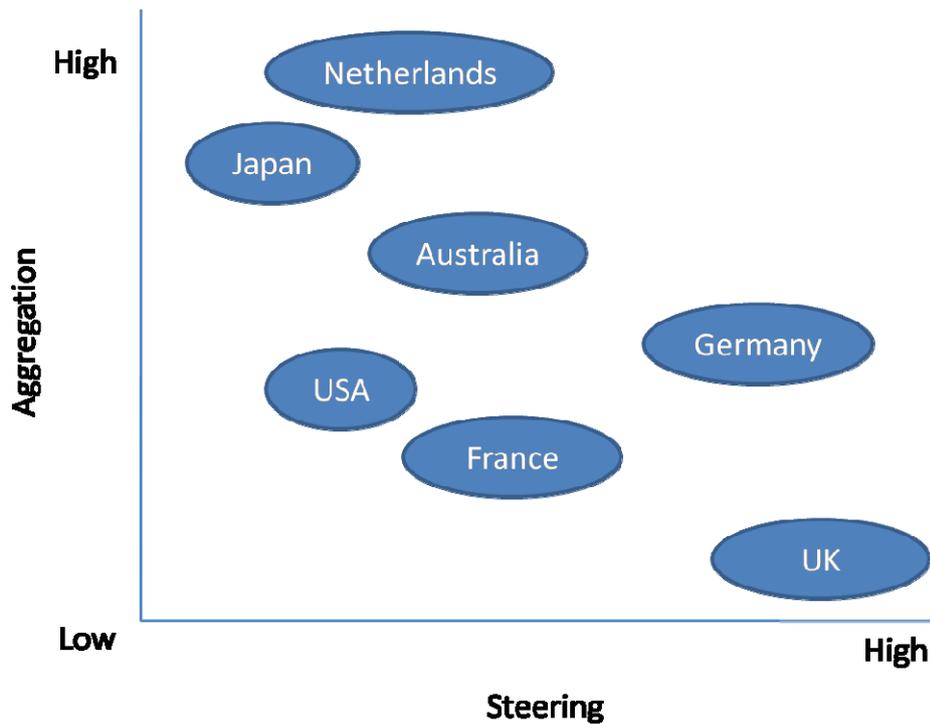


Figure 2.1. Placement of countries in relation to aggregation and/or steering of research adopted from Rip and van der Meulen (1996)

With this background knowledge of research systems, the following sub-section focuses on processes relating to research evaluation/assessment

2.3 Origin and background of research evaluation

Research evaluation has become a trend in many countries in recent years. Rip (2003:32) tracks research evaluation back to 1945 when the traditional ex-ante assessment was used as a tool to assess proposals in a bid to provide grants. Rip relates this input mechanism to the belief that “feeding the geese of science ... will produce golden eggs”. That is, researchers will continue to be productive as long as resources are made available. During this period, there were other evaluation exercises that were mainly linked to institutional management that did not make a significant impact at the time. In the post-war period (after 1945) government investment in research was affected by the weakened economy (Arnold, 2004). This also affected evaluation activities.

In the 1960 and 1970s, evaluation surfaced again and was conducted to establish whether the goals of research were achieved or not. At this time evaluation was not enforced at national level and several organizations were involved on a voluntary basis. Sizer (1988) cites the

Organization for Economic Co-operation and Development (OECD/CERI) as a typical example of an organization that laid a foundation for the evaluation process. In the late 1970s, as a result of the economic conditions of the time, governments introduced laws that compelled universities to account for their spending. This is the period when the relevance of research results became a matter of national concern. The emphasis of evaluation moved from mere allocation of funds to a focus on accountability. At the same time, in 1979, the UK introduced legislation that would initiate evaluation of teaching and learning, in order to reduce the spending of government funds (Geuna and Martin, 2001).

In the 1980s, the landscape of science policy changed (Rip, 2003), causing the ex-ante mechanisms to give way to ex-post strategies to assess strategic research (quality and impact). At this point, it was no longer only about how money was spent, and whether goals have been achieved but also about the appropriation of policies and how this could be achieved. In the early 1980s the UK commissioned an investigation into the future of research in universities and in 1985, the Jarrat report on 'Efficiency in Universities' sparked several policy documents that changed the future of state patronage of university research. In the Netherlands on the other hand, tracks of evaluation in the mid-1980s were guided by the need to relate autonomy to accountability of institutions. Whereas the intention to reduce dependency on state funds was and still is the focus for UK evaluation processes, the rationale in The Netherlands evolved several times. According to Weingart and Maasen (Whitley *et al*, 2007), Germany's involvement in exercises started after its Anglo-Saxon counterparts. Germany is said to be using a self-managed technique, which allows institutions to steer their own processes.

In response to shifts in trends in science and research, evaluation has addressed new challenges. A typical scenario is the continuous evolution of purpose in The Netherlands, which started with funding allocations in the 1980s, and changed to accountability between 1982 and 1992. As developments unfolded evaluation became an important strategic tool for policy development between 1993 and 2001 and now serves the formative and symbolic functions for institutions. In the UK, on the other hand, the purpose has not been altered. Different reviews of the evaluation programmes have changed methods to accommodate improvements and the changes brought about by new trends in science.

There are indications that the system of evaluation and the need thereof emerged from changes in science, technology and development, which are inherently based on research (Rip and van der Meulen, 1995). When these systems evolved, causing changes in the landscape of research, there was a continued need for evaluation and for evaluation to suit the evolving research systems.

As different landscapes of research exist in different countries, it is expected that the evolutionary processes and trajectories of research evaluation will also differ. These differences have been exposed to and are affected by the processes of globalization, which resulted in a tendency for comparison (between countries) and continuous adjustments in direction, swerving the different research evaluation ideas towards some common ground (Geuna and Martin, 2001).

Evaluation of research has now become well-established in many countries. Rip (2003:36) poses three challenges that have to be viewed in the context of recent developments, in which evaluation can “address strategic issues”, “improve national research systems” and “identify expected and unexpected impacts of open-ended R&D”. The information above is important as background information for better understanding when evaluation exercises are studied and for observing when new exercises frameworks are crafted.

Despite differences in the origins of research evaluation there are also similarities (commonalities) in the practices of research evaluation, especially those of ex-ante evaluation of proposals for funding. Although other countries evolved from this common system, traces are still deeply rooted in the research councils in most of the countries discussed by Geuna and Martin (2001).

For a better understanding of these differences and/or similarities of evaluation systems and for purposes of comparison, a framework has been designed for this study to provide areas of concentration in a nutshell. The framework consists of the following questions:

- What is the **purpose** of research assessment?
- What is being evaluated/assessed (**unit/levels** of assessment)?
- What is the **time frame** of the evaluation?
- What **dimensions** or aspects of research are assessed?
- What **methods** of R&D evaluation/assessment are used?

2.4 Roles and purposes of research evaluation

Different authors describe the significance of research evaluation in different forms. These descriptions (views) are mainly dependent on the environment in which such evaluation exercises takes place. For example, Rip and van der Meulen (1996) describe the purpose of research evaluation as explaining the performance of a country. Kuhlman (2003:352) on the other hand relates the shift in evaluation from the need to “demonstrate accountability, to the need to improve understanding and inform future policies.” This is in agreement with Campbell’s (Shapira and Kuhlmann, 2003:102) view of evaluation as being of “strategic relevance”. Therefore, the results of evaluation are used to inform the strategic formulation of policies wherein the allocation of research funds are internalized and meant for the improvement of research competence.

Although Rip and van der Meulen (1996) also view the role of research evaluation as an important aspect for national innovation systems, they associate such systems with the component of competence as a “key factor in explaining (such) performance”. This view possesses an inherent developmental approach to research which is characterized mainly by aggregation and not directly influenced by funding policies. Under these conditions, evaluation would be utilized to establish levels in institutional or national competences to enable the recommendation of ways in which these levels can be improved (Rip and van der Meulen 1996:344).

Geuna and Martin (2001) also emphasize the societal expectation for universities to be more efficient in utilizing public resources and the role that research evaluation plays as a basis for accountability of public funds spent in research. It is believed that evaluation related to accountability has its objectives in “measuring productivity”, providing “incentivizing” effects and utilizing powers in “controlling the academic workforce” (Weingart and Maasen in Whitley *et al*, 2007).

To emphasize this point of accountability, Campbell (Shapira and Kuhlmann, 2003) links the purpose of evaluation to the fact that university research funded from block grants, which he refers to as general university funds (GUFs), utilize public funds that need to be directly accounted for. The author compares GUF funding conditions with those of earmarked funding. The latter is said to come from different sources such as foreign funds, the private

sector and other special funds all of which have their own quality control mechanisms and expose universities to Mode 2 Knowledge production. The “curiosity-driven” GUF funded research, which still keeps universities within the Mode 1 framework, is said to have “no foreseeable near-application potential”, thus creating accountability challenges.

Present trends allow GUF-funded research to be judged on an ex-post basis. In his motivation for a university-wide research evaluation, Campbell claims that “there is a need for comprehensive institutional ex-post evaluation of university research” and emphasizes that “the whole university sector of a country should be addressed”. According to the author, “the greater the GUF funding component, the more there is a demand for such institutional ex-post research evaluation” (Shapira and Kuhlmann, 2003:106). Campbell can be interpreted as advocating the need for public accountability, advising that this can best be based on ex-post factors. Rip (Shapira and Kuhlmann, 2003:2-20 and 2-22), brings in another dimension of purpose and states that “evaluation is not just a check how money is spent (accountability) and whether the goals were achieved, but addresses questions (on) how appropriate the policy or program was and what must be a follow-up”. That is, “to improve understanding and inform future actions”.

Weingart and Maasen (Whitley *et al*, 2007) discuss yet another dimension of purpose that relates to competition. This symbolic purpose refers to universities in which acquisition of information on their relative standing becomes important. The authors relate the dimension of viability (dependent on foresight) to action that will or may be taken, which areas to pursue and how, and which ones to abandon. Different authors agree that all purposes of evaluation inclusively become part of the effort to strengthen the national research system (Rip and van der Meulen, 1996; Geuna and Martin, 2001; Campbell in Shapira and Kuhlmann, 2003).

In summing up the purposes of evaluation, Kuhlmann (2003:357) refers to the summative and formative poles of evaluation. When evaluation serves to legitimize performance as a measure of promotion, evaluation is said to serve a summative function. This concept maybe related to what Rip (2003:37-38) refers to as “decision support” as the results of evaluation provide the necessary evidence for action to be taken. If on the other hand evaluation serves as a “learning medium” for “future initiatives”, evaluation is said to serve the formative function, which Rip (Shapira and Kuhlmann, 2003) assigns the role for strategic change.

The above ideas provide a basis against which the purpose of research evaluation is grounded. How the results of evaluation are used mainly depends on the steering/aggregation intentions in a country. For example, the dimension of competence cited by Rip and van der Meulen (1996) is explained by such purpose as the improvement of the quality of science (intra-science quality) and this, according to Geuna and Martin (2001), should complement the internal efforts of universities on quality control. In other countries on the other hand, the results of a research evaluation exercise compare different levels of performance (of universities or programmes), leading to the phenomenon of “selective allocation of funds”.

While the dimensions of competition and national standing are also important, the goals of competence-improvement and accountability serve as the opposite ends of a continuum along which the purpose of research evaluation in different countries slides. This implies that at the one extreme, research evaluation serves as a measure of improving the quality of science, in which competence is an end while at the other extreme evaluation is a means to judge competency levels for the allocation of research funds. Placement of purpose along the continuum is the important factor in determining the type of research evaluation model used in a country. Such types are discussed later in this chapter. The purposes of evaluation are summarized in the table below:

Table 2.1 A summary of the types of research evaluation purposes

Purposes/functions	Explanation
1 Summative function	When the results of evaluation are used for legitimization/evidence purpose. The evidence may also be used to determine risks in a programme, policy or system
2 Formative function	When results as a guide to improve programmes, policies or a system
3 Strategic function	Evaluation may assist with information that contributes to strategic changes of a programme, policy or a system. Both this function and the formative function bank on results as a “learning medium”. The difference between the two comes

	with the extent of application of the medium.
4 Accountability function	When evaluations are used to establish efficiency for the utilization of state funds, results exposes the value for money of such programmes, policies or systems. Audit types of analyses are usually suitable for this function. Other forms of accountability may be for the fulfillment of national imperatives such as relevance.
5 Resource allocation function	Results of accountability are used to allocate resources. This ex-ante approach to resource distribution (based on ex-post information is observable in models with intentions to reduce state dependency and where resources are allocated on the basis of proposals, for example in research councils.
6 Quality control functions	The function aims at standardizing scientific quality and protecting it. The popular method to yield such results is the use of peer reviews meant to ensure acceptability. At larger scales, involvement of international peers has led to international acceptance.
7 Symbolic functions	Acceptance by others puts the programme, policy or system in good standing. This function relates evaluation to competitiveness and in some instances results for prominence and national or international rankings.

Source: Mouton (PowerPoint slides)

2.5 Units of analysis in research evaluation

Different models and exercises of evaluation use different units of analysis. Choosing a unit of analysis depends mainly on the purpose of evaluation. For example, when competency is measured departments/programmes/disciplines feature prominently as in the Dutch model. If international standing is important, as it was in the New Zealand exercise, then aggregation of scholars' performance across the university sector will most likely make the country serve as units of analysis. Ranking of universities, on the other hand makes institutional performance aggregates important and valuable for comparison.

Weingart and Maasen (Whitley *et al*, 2007) suggest that whatever the political influence, universities need to be supported in relation to their individual strengths. This is because according to the authors, each university has its own goals and special areas of offerings. These differences in landscape, together with the differences in quality and size of institutions impact not only on the grouping of study areas for evaluation exercises but also influence the choice of units of analysis.

The authors emphasize that the reputation of the department rests on the reputation of its prominent members who gain their reputation from research. According to Weingart and Maasen, this individual reputation depends entirely on those who control research funds, causing differences in reputations of different disciplines and determining stability or fluctuation of such reputations.

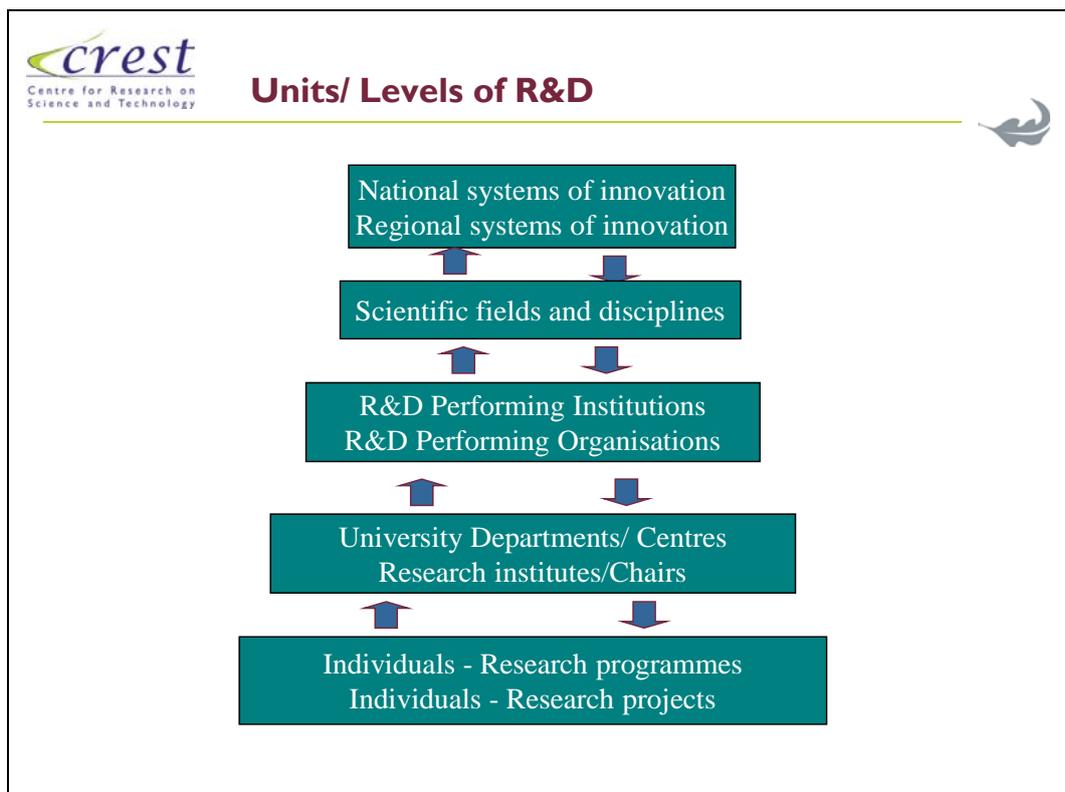
Of importance are problems related to the different levels of units of analysis. As indicated above, the reputation of a discipline may change when the performance of individuals is affected by extraneous factors such as funding alterations. Meanwhile, the ranking of universities maybe conducted at the expense of well performing departments/disciplines/programmes. That is, a high-ranking department will be affected by the low performance of other departments, disciplines/programmes if universities are regarded as units of analysis.

The above situation may be observed in a pluralized system of research where no one national system exists for university research and no policies exist across the entire system. In this case, the choice (of a unit of analysis) may depend on the evaluation to be conducted. Weingart and Maasen (Whitley *et al*, 2007) observe that researchers value the state of their department more than their universities and may attribute reputation more to disciplines than the entire institution. If no direction is given on the unit of analysis, researchers will most likely avoid being compromised by poor performers.

The allocation of funds also determines the choice of unit, depending on the level at which evaluation is conducted. For example, a research council may fund an individual or a programme, in which case the individual or programme will be the subject for evaluation. At a larger scale, for example when resources have to be allocated to universities, efforts are

aggregated and the university becomes the unit of analysis. This is observed in the UK's research assessment exercise (RAE) and the New Zealand exercise.

The diagram below (figure 2.2) indicates the different levels/types of units of analysis in a hierarchical manner. The bottom level is the lowest level in which assessment focuses on a researcher or a programme. This is more common when a proposal is being assessed for funding purposes and is a characteristic of the research councils. This unit may also be used when placement of staff is done in universities. Although the RAE exercise (UK) aggregates results of the research staff, staff performance is used as a sub-unit of analysis. The second level in the hierarchical form upwards, reflects the level at which departments, research units and research institutes are regarded as units of analysis. Studies that establish and/or rank universities per discipline, for example in the Dutch system of evaluation, fall within this category.



Source: Mouton (Power point slides)

Figure 2.2 Units of analysis used in research evaluation

The third level reflects the use of institutions and organizations. When the results of researchers or programmes of an institution or an organization are aggregated to reflect the performance of the entire institution or organization, the institution/organization serves as the unit of analysis. If on the other hand science fields or disciplines are assessed on a broader scale such as the national level, the field/discipline is regarded as the unit of assessment. The top level reflects the national system, the performance and /or standing of a country and is observable when countries are being compared. This level can also be used if countries are divided into regions and regional research and development has to be evaluated.

Care has to be taken when units of analysis are studied as some of the units may be used as sub-units and then calculated (aggregated) to reflect a higher unit.

2.6 Dimensions of Research evaluation/ Research criteria

In the process of measuring university research, the dimensions of research are used as criteria to define the parameters of evaluation. Whereas this chapter provides a simplified explanation of the dimensions, it should be noted that they may be used differently in different evaluation exercises. For example, in some countries, such as The Netherlands, a multi-dimensional approach to the quality criteria is used to explain competence in different areas, while in others for example the UK, one dimension of quality is perceived (Campbell in Shapira and Kuhlmann, 2003), for purposes of allocating funds. The Dutch system is seemingly more inclined to agenda building, and in the process defines quality in more dimensions than one in order to express their comprehensive system. For purposes of a better understanding of assessment strategies, the dimensions are generally explained below, as defined by different authors:

- **Quantity**

Quantity relates to the establishment of frequencies of research units. For example when outputs are counted without any consideration of inputs, then the total is referred to as the quantity of the results. This traditional method of determining outcomes may be used in the placement of individual researchers.

- **Quality**

According to Campbell (Shapira and Kuhlmann, 2003), quality relates to the level of scientific achievement of a product (innovation or publication), which is more inclined to

output. The Dutch Protocol (2003-2009) on the other hand, defines quality as a measure of excellence, relating this term to eminence of research efforts and the ability to perform at a prominent level and international standards. This ability is said to be realized through proficiency and rigor of research concepts and conduct together with success in scientific development. Acceptability by others brings in the notion of international standards.

The instrument used to measure quality, according to the Protocol (2003-2009), ranges from discussions (interviews) with research leaders to expert judgment (observations and document analysis) by members of the evaluation committees. The quality of facilities used for research also serves as an indicator of excellence. Therefore in most cases the measures of quality are based on the subjective judgment of those regarded as experts in the field.

- Productivity

This component is mainly based on numerical measures of variables used in the system. According to the Protocol (1998) productivity is better defined by relating output to the input of human resources. The Protocol (2003-2009) is more specific on the measure used and explains that quantitative information on publications and citations, in the form of bibliometrics are recorded as indicators of productivity especially in the natural sciences and life sciences. This is despite the problems attached to the Institute for Scientific Information (ISI) citation index, for example, the between-discipline impact scores. Technometrics (patents and citation patents) and socio-economic products of research are also counted as output.

As stated, the UK system relates quality to productivity, and this fulfils the purpose of fund allocations. Productivity and quality dimensions are often compared with the objectives and missions of the submitted research programmes and their differences are taken into consideration. This, in some sense, is the establishment of efficiency and encouragement for institutions to be accountable.

- Efficiency

Campbell (Shapira and Kuhlmann, 2003) relates research input to outputs to explain efficiency, while Sizer (1988:153) associates research excellence with the “value for

money” phenomenon. In both instances, the authors explain the dimension in terms of the use of resources, that is, how inputs are utilized to yield quality output in an economic manner. In these instances efficiency may be linked to productivity.

- Effectiveness and Relevance

Sizer (1988) defines the dimension of effectiveness as the achievement of objectives, strategies and action plans. This is based on the idea of responsiveness and therefore on relevance. Relevance is related to Rip and van der Meulen’s (1996) agenda building phenomenon. In this dimension, academic and social or professional impact anticipated from the research outcomes is assessed from the research group’s viewpoint (Campbell in Shapira and Kuhlmann, 2003).

The scientific, technical and socio-economic impacts of research define the dimension of relevance, achievement of which is regarded as effective. The possibility of expanding the definition in line with Rip and van der Meulen’s (1996) idea of “performance of a country” and Campbell’s (Shapira and Kuhlmann, 2003) “strategic relevance” was discussed by Verkleij (1998) who considered socio-economic relevance an important area in research. The same international concept stated in the Protocol (1998) reappears in the standard evaluation Protocol (2003-2009), with more emphasis on society-based, socio-economic and technological contributions. Both quantitative and qualitative measures are bound to serve the purpose of data collection and analysis.

- Viability, vitality and feasibility

According to the Protocol (1998), international competitiveness is regarded as a strong indicator for future standing in research. This is determined by the area(s) of study(ies) selected as well as approaches used. Consistency of programme elements and the availability of important features such as infrastructure are determinants of a viable situation.

Thus, the Protocol (1998) refers to the component as academic viability, which could be detected through different aspects such as infrastructure availability and other elements that would impact on the future position of the group.

Vitality and feasibility seem to carry the same meaning as viability, but expressed in a slightly different way. The choices made and the success rates of the project are the main features. Vitality and feasibility also refer to the flexible and sensitive ability of a group to discontinue non-viable projects and initiate new ventures (Protocol, 2003-2009). Measurement of professionalism in management and policy decisions in both the programmes and institutes are important, taking into account past performance and future plans. As changes in the institutions/departments/programmes may be affected by other uncontrollable (confounding) factors such as funding bodies' demands, a qualitative approach has to be used to gather information to enable detection of such factors.

A summary of the dimensions is given below:

- The **quantity/volume** of research – Frequency of products of research
- The **quality/excellence** of research - Research of an acceptable standard mainly determined through peer review. Some researchers believe that quality explains the visibility of a researcher within the body of science or the impact made by researchers in the field
- The **efficiency/productivity** - How much output is produced given certain input factors such as human resources and funds
- The **relevance/effectiveness** of research: This relates to the achievement of objectives, responsiveness to national agendas and contribution to agenda building
- The **viability/vitality** of research: This mainly refers to the sustainability of a field or programme of study. Viability is based on consistency and availability of important features and determinants of a viable situation. Foresight/vision, flexibility and sensitivity to discontinue non-viable projects and initiate new ventures are important here.

2.7 Methodological issues

Most authors agree that evaluation of research is a difficult task (Gordon and Purvis, 1991; Nederhof and Zwaan 1991; Nederhof and Noyons, 1992; Hemlin, 1996). This becomes more obvious when an exercise puts focus on a multiplicity of purposes or when evaluation has different audiences who are interested in different evaluation results. The purposes of R&D evaluation range from investigating 'excellence' (Gordon and Purvis, 1991), to the intention

to achieve quality research, which sometimes ends up being described in terms of ‘importance’, ‘visibility’, ‘impact’ and other such performance constructions. The confusion may be exacerbated by the lack of valid and reliable methods and instruments to yield the intended results.

In the process of trying to understand the measures/methods of research evaluation, a few issues may need to be clarified. For example, it is necessary to determine whether or not the main purpose of evaluation is the determination of levels of performance, in which case it is necessary to establish which objectives are intended in the study. If it is the competence of research that is at stake, the question is whether or not the popular measures such as peer review are independently valid and reliable enough to produce undisputable results. Competitiveness and international standing may also be established through the results of competency evaluation.

Incorrect selection of a method may result in incorrect and unreliable data, which in turn might invalidate results. In research evaluation in particular, unreliable data may create a wrong impression about a situation and this may in turn lead to incorrect decisions- and/or policy-making. In the process of obtaining more knowledge about the system of research evaluation, researchers have identified different methods, some of which have been exposed to tests of validity and reliability (Nederhof and Zwaan, 1991; Nederhof and Noyons, 1992; Hemlin, 1996). Meanwhile, it may not be easy to obtain a reliable measure against which results can be tested for validity.

Hemlin (1996) puts the blame for any possible confusion about the full reliability of evaluation processes on human judgment. This behaviour places the onus for the production of valid information heavily on the access of experts, a practice that is itself highly subjective. Unfortunately, at this point in time a reliable solution does not seem to be available. Publications and citations, which some of the evaluators prefer for the achievement of outcomes, are themselves products of a subjective peer review process and their quantitative nature does not discount the subjective effects.

Different authors place the methods of research evaluation into different categories suitable for discussion. For example, Arnold (2004) focuses on evaluation in management and refers to two traditions. The first category Arnold refers to as the “scientific peer review tradition”

that is being used to assess the quality of research. Results of peer evaluation may be linked to productivity and standing. The other category according to Arnold relates to ex-ante information when proposals are being evaluated for funding purposes (the input measures).

In their inherent forms, especially at micro-levels, the traditional approaches still provide the necessary information understood by scientists and therefore still serve the evaluation purpose. Arnold (2004:4) warns of the simple use of the methods at macro levels, for example at national levels where research is weighted against other national criteria such as “resources availability” and “socio-economic relevance”. Under such conditions the mixed method approach may be necessary, creating a third category that combines the other two.

The categories are explained below:

Peer Review Measures

Peer reviews in research evaluation involve judgments made by “expert” colleagues to evaluate research efforts or productivity, a research product, a researcher or even a researcher’s department. Hemlin (1996) regards the practice as expressing an operational opinion on quality from an intra-science perspective. The South African Policy and Procedures for Measurement of Research Output of Public Higher Education (DoE, 2003:4) views the measure as “a fundamental prerequisite of all recognized outputs” and as a “mechanism of measuring and thus enhancing quality”. Unfortunately, the document does not explain what is meant by the “fundamental prerequisite” in this sense and how it is expected to “enhance quality”.

Assessment or judgment is usually based on qualitative strategies such as document analysis, site visits and interviews. Depending on the design of evaluation, which itself relies on the evaluation purpose, these strategies may be used in a single-handed (solely) or in a combined form.

Similar to the input measures, the use of the peer review measure has drawbacks. In an attempt to detect the quality of research, the method has serious validity and reliability problems that cannot be overlooked, especially if the method is used to produce results that

determine benefits such as rating/grading and funding. Weingart (2003:4) draws attention to the loss of control of the scientific community (academics) over the measures that led to the loss of trust over its use by the public.

Despite its stated subjectivity, there are indications that aspects of research activities such as quality, productivity and impact relate to the peer review measure. This results from the fact that judgments (reviews) are made by acknowledged experts in the study area. According to Cave *et al* (1988), the interviewer (the expert) should be of the same status as those reviewed in order to qualify as a peer. Meanwhile, if the issue of “too senior” peers is at stake, one may also argue that “same status” may jeopardize judgment, a situation that one may be tempted to refer to as quality judgment. The purpose of evaluation and the decision context play a role in determining the “quality” or levels of panelists.

Data-based approach

These are commonly known as performance indicators (PIs). The use of these measures was made popular in the UK by the Jarrat Report of 1985 and was a topic of discussion throughout the 1990s. Just as quantitative measures are regarded as relatively objective by others, PIs have according to Weingart (2003) been used to compare/rank performance within and between institutions. PIs have also been used as a tool for resource allocation (Tognolini *et al*, 1994; Cave *et al*, 1996; van den Berghe *et al*, 1998).

Whereas in 1988 Cave *et al* defined PIs as ordinary ranking, the authors later aligned the measure to institutional goals, basing its operations on information management data (Cave *et al*, 1996). This emphasized the appropriateness of intra-institutional use (within institutions) where the mission is the same (Van den Berghe *et al*, 1998). As a result, the measures are said to be increasingly attractive to policy makers. Such operations also need to observe that the use of publications and citations in the PIs also succumb to problems related to these measures (Cave *et al*, 1988).

The use of PIs in research evaluation is not without its problems. Ball and Wilkinson (1994) observed that the PIs discount the differences in institutional missions, which also determine the research concentration of such institutions. Weingart (2003:2) indicates that the use of indicators especially to determine research quality is not yet generally accepted. Thus, the author warns of an “uncritical embrace” of the use of the measure.

Although Cave *et al.*, (1988) acknowledge the possibility of qualifying peer reviews as PIs, a quantitative approach to peer review can only expose the measure to the shortcomings experienced by quantitative measures other than the generic or even their inherent use as a qualitative measure. As a compromise, qualitative analysis with data grouped in themes may be used to enable easier interpretation of results by policy makers.

As already acknowledged, measures of research evaluation used to determine competency, are generally highly subjective. Westerheijden (1999:447) warns against a relaxed use of PIs in research evaluation, referring to them as “no more than signals that indicate where evaluators need to ask questions” and standing as “proxy” to unknown variables without any ability to judge quality. On the basis of the above motivation, it is important for research evaluation to be critical when the choice of methods is made, with an added level of competence of use.

Whereas other measures are easier to use and less costly rigorous scrutiny is relevant when applying measures that are meant to yield appropriate outcomes. As has been explained, PIs may be used to supplement other methods in research evaluation exercises (Weingart, 2003). Both the author and Campbell (Shapira and Kuhlmann, 2003) mention the use of bibliometric measures which have gained PIs more popularity among policy members. More information on the measures is explained in the relevant sub-sections below.

The choice of evaluation measure is dependent on the purpose of evaluation. The most commonly used indicators are; numbers of publications, numbers of citations, graduate (doctorate) throughput rates, patents and sometimes the ability to attract resources. All stated measures, with the exception of the ability to attract funds are output measures while the latter is an input factor. Weingart and Maasen (Whitley *et al.*, 2007) regard the use of extra-mural funds (an input measure) to be a determinant of “research strength”. Some of the measures, for example, citations are not instantaneously available and may take long before they are acquired. This requires a periodic output evaluation referred to by Campbell (Shapira and Kuhlmann, 2003). Campbell regards indicators as supporting peer review measures, with their objective nature used to ease tensions among the evaluation experts. Quantitative measures discussed in this study are the most commonly used.

The use of combined measures

In response to Arnold's concerns about the reliability of the methods, evaluation has moved to the engagement of more than one method. The mixed mode is also guided by the complexity of some of the objectives that can only be achieved through the collection of multiple data. In this approach, judgment-based approaches are combined with the input measures to improve results.

A composite discussion of the research evaluation methods

Martin and Irvine (1983) acknowledge the complexity of research evaluation. The fact that different institutions/organizations perform evaluation for different reasons makes judgment of the process even more difficult. What matters most is whether the results are valid and can be relied upon. Despite the inherent subjectivity of evaluation research (Rossi and Freeman, 1993), there has to be some level of assurance about the dependability of results. For this to happen, evaluators need to take into account the main intentions (purposes) behind the process of evaluation.

Verkleij's (1998) idea of a decision context provides a framework within which research purpose can be matched with the methods of assessment. The direction which the process of evaluation takes is therefore guided by both the intention of evaluation and the context within which the reasons for evaluation are formulated, with the latter forming the base for the process. Three main paradigms of purpose may be identified within such a context: the diagnostic, the prognostic and the comparative purposes of evaluation. The purpose may be diagnostic in nature, in which case levels of performance will be the focus of the study. In such cases, methods that determine the level of performance will be relevant. Meanwhile, it is necessary that the *status quo* be established prior to any evaluation.

Where evaluation has to be undertaken in order to predict future performance, other means may have to be built into the system to augment the methods so that predictions can be assured. Prognostic measures are more complex in that a lot of factors have to be considered or eliminated before any prediction can be made, and even when this happens, future events may not be guaranteed (Rossi and Freeman, 1993)

Where the different levels are utilized to compare performances, methods related to comparison assume the focal position. Methods such as those that are used to bring

performance to the level of comparison may not differ with the ones used in measuring the level of performance. In the former case, however, the different levels are compared, which implies a different design/approach (to the evaluation design).

A closer look at the ideas involved in commonly used methods, an examination of common areas of application together with a critical observation of advantages and disadvantages reveals important information that guide application of such methods. Conceptual information about the methods serves as a guide for matching methods to the purpose. Of the rest of the performance indicators usable in research evaluation, publications are most commonly used because they are easy to measure and somewhat objective. Counting frequencies from different sources is said to be less time consuming and economical. Research councils also utilize publications to reflect more involvement in funded projects (prognostic idea), which they believe reflects a high affinity for funds. In this sense, the measure serves an ex-ante role. Unfortunately, this audience-inclined idea excludes the outputs of authors involved in non-funded work. Publishing in “prestige” journals seem to improve the value of measuring quality.

Whereas output results from the use of the publication method may be used as input measures to predict whether a researcher will maintain the level of productivity or not, such a prognostic decision is based on a risk that different conditions may distort the process. For example, if the author was responding to demands that lead towards a promotion, research activities may decline as soon as such a position has been attained. This may also affect the productivity of the researcher’s department.

Secular demands such as national developmental conditions may also deter the process by shifting researchers’ focus on intra-science research (internal to science) to the more applied or even contractual research. In the latter instance, the agency that commissions the research may own the results and lay down conditions that prevent publication, and this may affect the rate of publications. Therefore, when publication counts are used for diagnostic purposes, the type of research conducted needs to be taken into consideration.

Citation analyses on the other hand, are said to reflect “preference” by others and an effort to contribute to the improvement of understanding and thus to the body of knowledge. The measure accommodates articles that are cited despite their negatively perceived impact.

Although this expression explains the notion of visibility or preference, care must be taken if it is used to mean quality. The diagnostic value of the measure is affected by the ‘window period’ of waiting to be cited. This makes the quantitative measure a typical ex-post measure, which despite the challenges experienced with its use, is preferred by both the policy makers and university management. The waiting period qualifies the measure for systemic evaluation. Prognosis and predictability may be affected by the unpopular Halo effect and other such factors that relate to the movement of scientists. In most of the research evaluation exercises citation analyses are coupled with peer reviews.

This expert judgment by colleagues (peer review) may be used at different levels of evaluation (see Table 2.1). Verkleij (1998) emphasizes the importance of the use of this measure for both diagnostic and prognostic purposes because of its contributory effect to academic development and improvement. This applies where results of judgment are used for quality improvement, to advise researchers on areas that need to be attended to in order to guarantee better results. It is for this reason that the Dutch system prefers this measure. Such advice is more useful at input and throughput levels, which allows the measure to serve an important purpose for the improvement of present and future performances. By implication, peer review is not only a measure of evaluation but also an important instrument for improvement of research quality and performance if appropriately utilized. Otherwise, the subjective status of the measure exposes it to serious biases (Matthew, Halo and Oxbridge effects).

The effectiveness of peer review in quality improvement depends entirely on the experts used in the evaluation panel. In this context, the word peer may be misleading, especially when it is interpreted to mean those of the same status as those reviewed (Cave *et al*, 1988).

Therefore, Campbell (Shapira and Kuhlmann, 2003:106) refers to peer evaluators as expert “scientists evaluating other scientists” and this word (expert) is preferred to avoid the different interpretations of the word “peer”. This resulted in the method being referred to as “expert judgment” by other authors/systems. The change of name does not eliminate the subjectivity related to the choice of such experts.

Another problem related to the use of the peer review method is the involvement of a broad spectrum of researchers. While this is usually taken care of by only involving experts in the same area, the uniqueness of other types of research (intra-disciplinary and inter-disciplinary)

approaches in comprehensive/large-scale research may affect the intended results. Weingart and Maasen (Whitley *et al*, 2007) also warn of “cognitive distance” among experts, which may result in non-correlation of their judgments with that of other scientists. Interviewing researchers and giving them the opportunity to comment on reports may help reduce the risk. These methods are preferred for their potential to include all parties. Cross-referencing among panel members is also helpful in evaluating inter-disciplinary research.

When a need arises to determine the ability to attract funds, the ex-ante research input measure becomes relevant. Output from a previous exercise may serve as an input variable for the next cycle/exercise. The method has been popular since the 1960s for purposes of comparison so as to allocate research resources, especially the regulation of block grants in universities. It is predicted that those who have a track record of awards would continue to be productive and thus deserve to get more funds. Cozzens (1995) refers to this as the “trust me phenomenon”. The phenomenon, which applied both to individuals and institutions, was affected by the new funding policies of the late 1980s and replaced by the “show me” phenomenon especially in the UK, the US and Australia (Cozzens, 1995) and lately, New Zealand.

A face-value look at the research input method may discredit its use as a prognostic tool. This is true even though the method is easy to use and data is easily accessible, as inputs may not necessarily guarantee outputs. The method may nonetheless be relevant when evaluation is meant to reveal other sources of income (for example in a research centre) such as contracts, which would otherwise not be easily detected unless broad range output measures are used (as used in New Zealand).

It has already been stated that secular conditions such as activities of change in government, transformation and development may compel researchers to take a detour from purely intra-science (curiosity-driven) research activities and participate in nation-building, applied research activities. The popular research types (designs) under such conditions are exploratory research or feasibility studies, and programme and policy evaluation, which are characterized by the submission of reports to those who commissioned the studies. The number of grants an individual has acquired may then be an indication of reliance on the researcher’s levels of engagement (performance).

In South Africa, research also has to contribute to development, nation-building and transformation activities, and this has to be considered when research is evaluated. The trust invested in input measures is not dependent only on the reports already submitted but also on the willingness of others to have invested on the researcher (evaluation by investors). Funding councils still tacitly apply the method as a criterion for their rigorous screening in the grants award exercises.

Verkleij (1998) condones the use of a measure or measures that carry the potential to improve performance. Verkleij asserts that performance will be enhanced if assessment is formative and is conducted with a view/intention to advise and not punish. For this purpose, expert judgment has an inherent ability to facilitate intra-science activities.

As indicated before, audience accountability also provides a direction and choice of measures. The figure below (2.1) provides a summary of a two-directional accountability process governing purpose for the evaluation of research. On one side, is the type that may be regarded as formative, which leads to development that it is implied is essential. One may assume that in this context, efforts will be made against all odds to strive for excellence. The context that is dominated by budget decisions seems to suggest that the levels of excellence have been attained and that either researchers stay within the expected levels, or they will perish. The model may be popular in countries where higher education subsidies have been drastically reduced.

The advantage of the formative assessment model is that it bears a fruitful path while the dominance of the budget-purpose approach seems to be making the purpose an end in itself. It is necessary therefore that the context be established prior to an attempt to determine which model is suited for a HE system.

As stated previously, the advantages and disadvantages of a measure serve as a guide for its selection. The main disadvantage of output measures is the inherent subjectivity posed by the measures. While a quantitative approach is used and is referred to as objective through frequencies of publications and citations, the processes that lead to article publication and those that are used to determine citation indexes are also subjective (peer reviews). Meanwhile, the subjectivity of peer review/expert judgment measures may be overcome, to

some extent, by the use of standardized measures. The measure is more effective when evaluation is conducted for purposes of academic improvement. On the other hand, input measures fail to predict intended outputs.

The methods used in the measures explained above are listed in Table 2.2 below. Individual methods have not been discussed in this study.

Table 2.2 A summary of the methods regularly used in research evaluation

Category/measure	Methods
Judgment-based approach	- Peer review - Reputational rankings
Data-base approach	- Bibliometric analyses (citations and co-citations/co-word mapping) - Econometric modeling - Surveys
Mixed-method approach	- Case studies / history

2.8 Types of research evaluation models

This section provides clarity on the different types of research evaluation models used in different countries. This will enable a better understanding of systems in South Africa and for comparison with other models. The types are presented as Campbell (Shapira and Kuhlmann, 2003) described them.

Campbell differentiates between two types of country evaluation models: Type A and Type B. He used levels and the extent of involvement of universities in different countries to arrive at the classification. Under Type A, Campbell classifies those countries that have been (or are) involved in a systemic and comprehensive research evaluation of institutions at national (inclusive) level and across all disciplines. In such a system, evaluation results are based on a publicised scale and common methodological standards are used across all disciplines based on evaluation cycles.

Campbell (in Shapira and Kuhlmann, 2003) also qualifies the Type A model as an ex-post research evaluation, based on peer-review and organized by an intermediary body, either assigned responsibility by the state, for example, the Higher Education Funding Councils (HEFCs) in the UK or by the science community as is the Vereniging van Samenwerkende Nederlandse Universiteiten (VSNU) in the Netherlands. One may conclude that the HEFCs are a direct influence of steering while the VSNU is a typical research communities' aggregation. The two bodies (HEFC and VSNU) are discussed in detail in chapters 4 and 5 of this study.

Campbell's Type B model is a more "pluralized" type in which evaluation is situational and independent of the grading system. This includes, among others, an individual research project or team evaluation, which is a typically characteristic of the type used by research councils and found in Finland and Austria. Another observable characteristic of pluralism is the comprehensive coverage of one discipline.

Reasons for the pluralized type are different in different countries and are a result of the differentiation of evaluation goals (Rip and van der Meulen, 1996; Geuna and Martin 2001; Campbell in Shapira and Kuhlmann, 2003). The differentiation of goals may be influenced by different factors. In Germany for example, operations in different regions create pluralized goals and thus a pluralized system of evaluation (Geuna and Martin, 2001). Some countries are highly influenced by their government funding types. In instances where funding allows a bigger portion to be allocated on a project competitive basis, like in Italy and Finland (Geuna and Martin, 2001), more authority is given to research councils, which in turn utilize project/programme evaluation results to allocate funds.

Also included is the fragmentation of responsibilities for science, which leads to the existence of different intermediary bodies (research councils). This differentiation may lead to a vast differentiation in funding type, for example other sponsors such as industry. According to Rip and van der Meulen (1996), this differentiated sponsorship creates a broader scope of interdependency (national collaboration) and introduces additional goals that not only reduce dependency (on state resources) but may also reduce the steering power of the state. The authors refer to this arrangement as an evolution of the modern research

system towards a post-modern one, and although the latter system is preferable, the managerial challenges of multi-sectoral systems under these conditions cannot be ignored.

Care has to be taken not to always relate state funding to the steering mode. Whereas the UK system (which has a Type A model) is mediating state objectives, it has been observed to be utilizing less state funds on research than some of the European countries that have less steering effects (Geuna and Martin, 2001). The authors allege that the Dutch system and in Japan, where state funds are said to dominate, other sources are more aggregated. This implies that even though the state is directly involved in the latter instance, both the state and the research communities may share common goals, causing the aggregation to neutralize steering effects. Such a condition is observable where research responds to the needs of a country and the component of relevance is introduced through common scientific and social interests (by the research communities and the state) as stipulated by Rip and van der Meulen (1996).

As already stated, in other instances aggregation may exceed steering influence, in which case the research communities are highly regarded by the state. Rip and van der Meulen (1996) warn of too much reduction of state influence and view it as abandonment of responsibilities or failure by the state to provide direction on imperative national issues. This not only denies the state an opportunity to directly benefit from research but also promotes inherent research dominance challenges that may obstruct new researchers from entering the arena. The 'absence' of the state leads to the absence of direction on research type.

In some countries such as Germany on the other hand, pluralism results from hostility towards envisaged competition inherent in a comprehensive evaluation system, and to some extent the level of dependency of universities on the state (Geuna and Martin, 2001). The steering-type of authority by the state, which may result in dependency of institutions, may determine the level of influence that the state has on them and the type of evaluation they should undertake. Therefore, a majority of Type B countries would be more dependent on research councils as a strong research funding stream.

The comprehensive coverage of one discipline found in Austria is said to already follow a sequence of (different) evaluations, with a tendency of moving towards a systematic evaluation approach (Geuna and Martin, 2001). Campbell (Shapira and Kuhlmann, 2003)

also highlights the element of interrelatedness and interdependence between countries brought about by levels of research-intensiveness and knowledge levels of some of the countries. There exists a tendency/opportunity for other countries to learn from others. This brings about the element of internationalization, with a tendency for other countries to be influenced in the direction of others.

Accordingly, Campbell (Shapira and Kuhlmann, 2003) postulates conversion from Type B to Type A, when research intensiveness improves, placing the Dutch and the UK systems among those ahead of others. The aspiration towards such systems occurs “in less developed countries”, in which “building up a modern research system is an important task” resulting from “ambitions to compete globally” (Rip in Shapira and Kuhlmann, 2003:2-29).

Unfortunately, the author warns that “by the time less-developed countries have emulated this model, the leading countries will have moved on” as a result of evolutionary processes. Notwithstanding this tendency of Type B countries to move towards a Type A model (Geuna and Martin 2001), other countries such as Germany are said to be delayed by some skepticism towards the competition involved in a systematic process of evaluation, as already stated.

South Africa does not have a unitary system of evaluation. It is therefore classified under Type B models, with different levels of research evaluation. This is the result of government initiating different exercises in order to address different purposes at different times. As this study particularly focused on the subject, information above served an important role in revealing the status quo in the country.

Chapter 3 The methodology of this study

3.1 Introduction

This chapter explains the methodology followed in the study including the selection of models/cases referred to, information gathering strategies and processes of analysis. The study uses an analytical, qualitative inquiry of a primarily non-interactive nature to create a basis for the understanding of the nature, events, practices and processes of research evaluation. This is accomplished through a series of benchmarks by using established, well co-ordinated systems (models) of research evaluation as points of reference. Care has been taken not to replicate information gathered from the models but to make context-bound generalizations arrived at through inductive processes and then translate the information into recommendations deemed suitable for South African universities.

The choice of an analytical style was informed by the explanation given by McMillan and Schumacher (1993) who define the research style as describing and interpreting the past and recent past from selected sources followed by the use of logical induction to qualitatively analyze documents preserved. The authors attach credibility of style to the procedures used that include the search and criticism of sources combined with the interpretation of facts, to improve external validity. This type of inquiry, oriented to the production of reconstructed understanding is said to replace internal and external validity by trustworthiness and authenticity (Denzin and Lincoln, 1998).

An *a priori* decision on the approach to be used was not finalized in the initial stages of planning and this allowed for the adaptation of strategies as the study continued to develop. Corroborating instances led to the other, with each level representing an increase in empirical content, constituting a constantly progressive theoretical underpinning. Therefore, a continued review of models and other sources and the analysis of information accrued, contributed to the continuous conceptualization of the study from the beginning to the end. This includes the design developed for the study (including strategies used to gather information and procedures followed) to arrive at the intended conceptual idea of research evaluation for the South African HE system.

The study starts in the introductory chapter, by providing a background of research evaluation efforts in South Africa together with some motivation for the need to focus on universities

(DoE, 1997a; DoE, 2001; Campbell in Shapira and Kuhlmann, 2003; NRF, 2005). This leads to the exploration of international systemic and comprehensive (large-scale research) evaluation practices in the form of benchmarks. The reason for the choice of systemic and comprehensive model is driven by focus the study has on a national HE system and how evaluation can contribute to the achievement of the systemic imperatives. Therefore, a non-fragmented approach, Type A model (Campbell in Shapira and Kuhlmann) would assist in creating an understanding of evaluation within the situation.

The information-seeking endeavors culminated in detailed descriptive interpretations of research evaluation in other countries, their analyses of worth and possible contributions of practices of the different processes. Focus has been placed on the origin, rationale, context, purpose and activities of the selected models. These type of studies McMillan and Schumacher (1993) classify as policy studies due to their nature of focusing on power and resources to enact on policies, the reverse of which is also applicable.

The study also looked at the relationships between the selected models, focusing on similarities and differences in order to promote understanding of the evaluation processes. It is acknowledged that the inherent features of the different contexts within and between models describe typicality of different situations over time and across locations/settings. To further promote clarity, an in-depth study was conducted on the methods used in research evaluation, concentrating mainly on the methods used in the three models referred to in the study.

The comparative nature of the study, which promotes interrelationship among models, was adopted merely to allow for some theoretical imagination of contexts and processes to create such typicality that served as an extension of understanding for the second level of the study, which is to formulate a new idea. The attempt to gain more understanding was designed to avoid uninformed moves which Prout refers to as “progression in an ocean of anomalies” and in keeping with Bohr’s idea of a consistent foundation for clarity (Worrall and Currie, 1995).

To avoid limitations in comparability, three models were reviewed, the Dutch, the UK and the New Zealand, for the purpose of contrasting dimensions so as to test and consider suitability and to generate an acceptable frame of operation / reference for the intended South African model. An analytical approach, as well as rigorous techniques were applied in the

search and criticism of sources and the interpretation of facts for possible causal explanations and logical induction arrived at in the three models. This assisted in the formulation of the knowledge base against which reference was made when studying the South African system. This process by analytical induction according to Finch (1986) offers the prospect of producing accounts which go a long way towards achieving reliable data.

3.2 Study design

Once the ‘primary’ knowledge base and background on research evaluation in general was established, it was necessary to study South African research evaluation efforts against the established information. For this purpose, the same design of a multiple case study approach was used for gathering information necessary to build on the ‘secondary’ knowledge base against which the intended framework could be conceptualized. The design therefore has four stages: the first, a review to create a background, a framework for the study and to maintain focus, the second to form a base and the last to formulate an idea. The whole idea resulted from the fact that the intended mechanisms and strategies are guided by benchmarks and experiences of other systems in other countries. An impression of the four stages is indicated in figure 3.1 below.

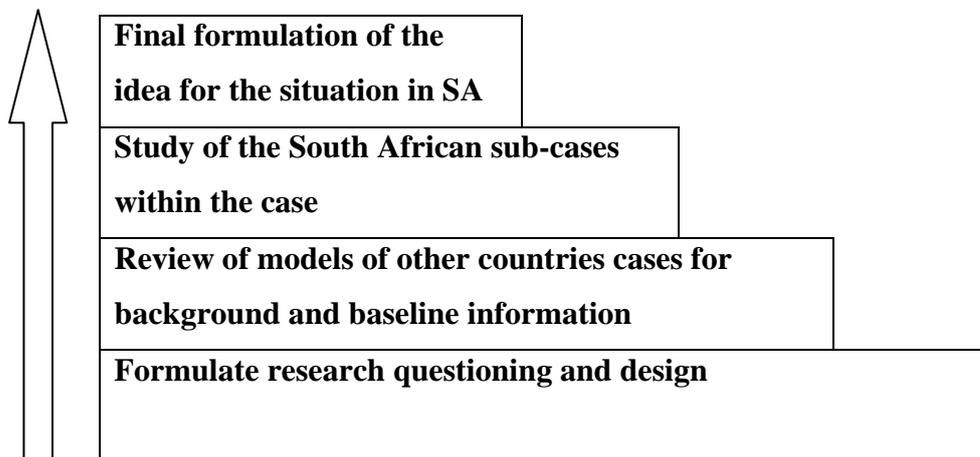


Figure 3.1 The framework used in the study displaying the four stages used from beginning (bottom) to end of the study (top) as formulated by the researcher

As already stated, the first tier displays a review in a form of a multi-case study made up of country models while the second is an in-depth study of South African research evaluation

practices referred to here as cases and serving as units of analyses. A more interpretative approach was used to fulfill intended goals at this level. This is revealed by indirect, non-intentional and non-deliberate differential comparison of phenomena and concepts between the cases within the broader framework, noting the differences between the models and between the models and the South African situation. Brochmann *et al* (1997) warn of the “fuzzy and messy” system problems attached to comparison in social sciences. By this the authors mean the complexity of observed systems and the “noise of non-systems” as a result of the varying integration of systems. Care was taken in this study to extract the necessary information to avoid this problem.

The review of the exemplary models created a frame of reference for the study approach. It was intended that cases on South African efforts at research evaluation would provide clarity on local practices, their origin, the contexts within which they operate and the goals hoped to be achieved through their application. The dimension was thus based on experiences and applications learnt from other models. Care was taken not to privilege the idea of universal standards, which may result in setting goals and norms that may be inapplicable to the South African context.

3.3 Models and cases selection

3.3.1 Selection of the models for review

Different countries use different models to review research activities in universities. . International models approaches were favoured because of their more advanced approaches that laid a foundation for the study. International participation and international competence are also regarded as important aspects of developments in South Africa.

A purposeful selection of models (McMillan and Schumacher, 1993) was guided by an extensive review of sources on research evaluation and research quality assessment coupled with intermittent verbal verification with different scholars at international conferences and centres. The reason for the selection was guided by the systemic nature of HE in South Africa. That is, that model to be studied would simulate systemic patterns and provide information on how evaluation of research can contribute to the improvement of research in such a system in order to address imperatives of such a system. The Dutch model was selected for its least differential nature with planned intentions for institutional accountability necessary for the autonomous South African Universities while the UK model shows a

different approach to this. Such information would guide the researcher on the choice of suitability of practices. Both models have a long standing history and are at advanced stages. The New Zealand model on the other hand is recent and it was hoped that with the history similar to that of South Africa, the model would reveal activities that are necessary for the local situation.

Some information had to be obtained prior to the final selection of models, to improve reliability on the models selected. The exploration was then followed by in-depth studies of a few models. This type of sampling is more appropriate to multiple-case designs studies than the maximization of variability, even though comparative analysis could not be avoided. The study of models followed a historical approach in order to expose the evolutionary processes and to reveal contextual differences at the different levels of development before a comparison could be drawn.

The choice of models was based on a ‘reputational’ approach informed by historic events and experiences accrued, especially at national level. While some countries have been observed to use pluralistic approaches to evaluation (Geuna and Martin, 2001), consideration was given to Campbell’s (Shapira and Kuhlmann, 2003) advice on the tendency of such approaches to evolve towards the systemic and comprehensive type when they develop. Countries that have been involved in research evaluation over a longer period of time, The Netherlands and the UK, stood better chances of being selected as they may be regarded as connoisseurs in the field (Campbell in Shapira and Kuhlmann, 2003, Lucas, 2004).

In both countries, research evaluation processes have been in existence since the early 1980s. All government subsidized (public) institutions participate in the exercise. This fitted with the intended focus of the study, South African public universities. The emergent model of New Zealand was included since, by context, it was perceived as being as close to the South Africa situation as possible. The New Zealand comprehensive/large-scale model is relatively new and is preferred as a guide for efforts to establish a new model.

Activities, practices and experiences of research evaluation prior to the introduction of the other countries’ new models are also reviewed in this study. This was done to minimize threats of reliability and validity that could occur as a result of the vast differences in experiences and contexts between the “connoisseurs” and South Africa. As previously

indicated, May's (1997) advise not to set standards beyond context and the level of judgment was taken into consideration. This is the main reason why some information had to be obtained prior to the final selection of the models. Care was also taken to study and conceptualize the selected models without any desire to generalize or even extrapolate beyond the limits of application.

The number of models selected was influenced by the availability of relevant information based on the main purpose of the study. Therefore, documents of models that had the logical potential to yield relevant information guided the selection exercise. The Dutch model was preferred for its evolving purposes, all geared towards high quality research productivity across the entire system. This would guide the South African imperative to reduce disparities. The UK model on the other hand, was selected because of its historic similarities with the S.A. efforts (purpose for funding) It should also be noted that most of the existing South African research evaluation efforts refer to the UK and the New Zealand models. A possible exclusion of similar models (systemic and comprehensive), if any, may be viewed as a bias towards the selected models, and this, if it exists, creates a setback for this study.

In systemic/comprehensive/large-scale research evaluation, governments are directly or indirectly involved to ensure processes of accountability and improvement (Campbell in Shapira and Kuhlman, 2003). Intermediary organizations are usually formulated in collaboration or in agreement with governments and assigned the responsibility to facilitate the process. In most cases if not all, the intermediary bodies engage advanced researchers. Examples of such organizations are the VSNU in The Netherlands, the HEFCs in the UK and the Performance-based Research Funding (PBRF) in New Zealand both of which are government bodies/structures. The NRF is a local example of such intermediary bodies.

In the United States, Canada and Germany 'federal' states operate autonomously with some guidance from the national department. In Canada, provincial operations have been reported whereby individual provinces may decide on how evaluation will be performed. In the US on the other hand, different constituencies are used for different purposes such as benchmarking and standards making the US system one of the most advanced differentiated higher education research systems. The exclusion of these models does not in any way suggest that they are unsuccessful but rather that this study focused on research evaluation of a single system of HE placed under a single DHET. The German situation illustrates how the Lander

(regional) setup still influences the operation of universities (Geuna and Martin, 2001). The fact that the South African HE system is a unified one led to the exclusion of these federal models. Although South Africa may learn from the federal state the intention of this study was designed to fulfill national imperatives. Therefore systems such as those in the US, that are highly differentiated were not the focus of this study.

3.3.2 Selection of the South African cases.

Relevance to the study has been the determining factor for the selection of sub-cases within the South African case. This followed Yin's (2009) embedded case study design, in order to accommodate comparisons between the South African Type B model with the other countries' Type A situation within the same phenomenon.



Figure 3.2 The multiple-case design with one case showing embedded cases (adopted from Yin (2009) and altered for this study

Efforts were made to identify bodies/structures in the country that are involved in research evaluation of universities in one way or another. For purposes of better understanding, the selected (participation) structures/bodies have been discussed in separate chapters in the study. Government documents and those of other parastatals involved in research evaluation exercises are mentioned below. Figure 3.2 depicts the embedded case study design as used in this study.

3.3.2.1 The Higher Education and related legislative documents

South African legislative documents relevant to HE research evaluation have been referred to in this study. To build on the history of South African research evaluation, documents of the old order (apartheid system) have been consulted. The emphasis, however, is on documents of the new system of higher education (from 1994) as indicated below:

- The White Paper 3 (DoE, 1997a) provides the conceptual framework for the transformation of HE in South Africa and led to the development and adoption of the Higher Education Act 101 (DoE, 1997b), which regulates the system and provides the legislative framework. The National Plan for Higher Education (DoE, 2001) is an operational document providing mechanisms for implementing ideas of the White Paper 3 and the Higher Education Act 101. All these are important in the regulation of activities of the public HE system which is the focus of this study.
- The White Paper for Science and Technology (1996) on the other hand, has been included for its role 'in shaping and steering research' in the country.

3.3.2.2 The Council on Higher Education (CHE) documents

The CHE and its committee, the Higher Education Quality Committee (HEQC), have been established and are regulated by the Higher Education Act 101. The inclusion of the CHE documents in this study was informed by the fact that its mandate is to improve quality in HEIs. The Council has employed different approaches to manage the situation and those that relate to or are involved in research evaluation have been included in this study. Those are:

- The Founding Document (2001)

This has been included to provide background information about activities of the Council.

- The Framework for Monitoring and Evaluation in Higher Education (2004)

In terms of legislation, the CHE is required to advise the Ministry of Education on the performance of HEIs for the purposes of efficient funding. Research is one of the components of institutions funded by the Department.

- The CHE Framework for Institutional Audits (2004)

The CHE, through its committee, the HEQC, audits the performance of institutions for quality purposes. These audits provide information to the Minister of Education. Although

one of the functions of the Audit wing is the accreditation of programmes, its involvement in internal quality management of institutions makes it relevant to this study.

- A Good Practice Guide for Quality Management of Research (2005)

This document of the CHE is closest to the aims of this study in that it focuses on HEIs research activities nationally except that developmental responsibilities are left to individual universities meaning that the stated activities in the Guide do not constitute a national research evaluation exercise. It's value and practices have nonetheless been studied to guide the current study's recommendations.

3.3.2.3 The Department of Education document (2003)

- The Policy and Procedures for Measurement of Research Output for Public Higher Education Institutions

This is the DoE's system of rewarding research productivity. Its direct involvement in research evaluation informed its inclusion in this study.

3.3.2.4 National Research Foundation (NRF) involvement

- The National Research Foundation Rating System (NRF 2002)

The NRF is an intermediary (parastatal) establishment responsible for developing research in South Africa. The rating system was intended to promote and support excellence in research and to support curiosity-driven research. The level of excellence has in recent years been used to encourage national and international competitiveness, an important transformation imperative.

The NRF has also been identified by the Department of Science and Technology as a body that should assist in the process of unifying research activities and the evaluation thereof.

3.4 Methods of data collection/information gathering

Processes of data collection and data analysis have been interactive throughout the study. The processes overlap and most of the time, display a tentative interpretation, resulting in a pattern used for corroboration. Information gathered and analyzed from one level

continuously served to provide a background and orientation for the next level, which also facilitated the continuous model/case selection exercise. The process of selection has been largely informed by analyses, interpretations and results of the other models/cases together with other sources. For Example, the study of the Dutch model guided the choice of the UK model while the latter led to the inclusion of the New Zealand model. Therefore, the study of one case triggered a sense that further information gathering (data collection) would yield more relevant information.

Creswell (1998) highlights the unique feature of a qualitative approach to research, enumerating the problems that may result from emerging issues that may develop in the field of study such as: how much information needs to be collected, how much theory should guide the study etcetera. These, Creswell warns, disable operations based on a detailed plan. He suggests the use of a general approach to a study. In this way, room is made for emerging knowledge to be tied to the context in which it is studied. Both the recording and analysis of information/data in this study are more narrative and descriptive.

While continued information gathering strategies may have an effect on validity, especially because it is informed by the researcher's 'sense', the information is regarded as trustworthy. Reliability is also not totally compromised as there has been a constant analysis of themes in the cases studied (models and cases). The descriptive nature of information recording used also helped ensure consistency.

McMillan and Schumacher (1993:446) advise that in analytical studies, "the search for facts begins with the location of sources", which requires the location of both primary and secondary sources. This study depended on both types of sources which are mainly documents. The primary documents referred to here are mainly legislative documents such as Acts, national policies, frameworks and other organizational policies and guidelines designed to direct and guide evaluation activities. A typical example is the Dutch Protocol, which is an organizational (VSNU) policy document and the National Plan for Higher Education (DoE, 2001) in South Africa, a government document. Intermediary organizations also have policy documents that guide their activities. Reports of evaluations were also reviewed as sources and these confirmed occurrences of the evaluation processes, indicating both the processes and the outcomes. The primary sources served as a basis for the approach to be used for documentation.

Secondary data has mainly been obtained from articles published in journals, conference papers, dissertations/theses and discussion documents. Some of the articles yielded fundamental information and were sometimes used as primary sources, for example, the explanation of concepts and historic periods. Most articles were nonetheless used for verification of opinions about situations with substantial interpretation. The articles played a major role during the analysis and interpretation of information and in this way served as secondary sources. These secondary sources have been used to guide the analytical processes of the study and assisted in developing a systematic analysis of concepts that allowed comparison of systems (models and cases) that are different in nature. The articles were also used to understand the different methods used in research evaluation.

Exploration was followed by rigorous attempts to obtain in-depth information on the identified systems. The attempts included visits to countries whose models were identified for the study (where possible), linking with scholars from other countries through meetings at conferences and/or referrals by others and through meticulous internet searches. Contact with those involved in similar exercises in their countries proved to be more fruitful as this both yielded relevant information available and created a platform for discussions which further improved understanding. The in-depth study of the Dutch model is a case in point. Members of the Centre for Higher Education Policy Studies (CHEPS) assisted in giving interviews and explaining the process. Some of the informants have been involved in the formulation and/or advising capacities of the models and policies used. These sources are in this case referred to as primary sources (McMillan and Schumacher, 1993).

As a centre of focus of the study, information was also gathered in South Africa. Whereas both primary and secondary data were obtained to provide background, the nature of the study compelled the researcher to concentrate more on primary sources. Legislative documents explain the rationale, intentions and context in each case, which were confirmed through other sources stated below. Documents of the intermediary bodies, on the other hand, provide information on practices in the country. Journals have been consulted for the opinions of scholars on the government documents and practices of research evaluation by other structures. Both information systems have been used by the researcher to determine the significance of the study and to obtain the direction that the study had to follow.

Only available sources were used despite intense efforts to acquire appropriate information and attempts to strengthen the trustworthiness and authenticity of the study. The probability exists that some important documents may have been omitted despite all these attempts. This is mainly the case with international sources where financial constraints prevented visits to relevant centres, which, it is believed, would have yielded more reliable information than did technological devices. This was especially true when electronically obtained documents could not give proper guidance in obtaining history-bearing documents that were necessary to provide the background information essential for the understanding of contexts, for example, the situation in the UK and New Zealand.

McMillan and Schumacher (1993) allude to Gottschalk's (1969) comments on dependence on "surviving records". The authors advise that necessary techniques for criticism be used to assess the authenticity and trustworthiness of the sources. This strategy was used in this study to improve the value of information gathered.

The conceptual framework on which the study is based reduces threats to reliability especially because document analysis bears low inference characteristics (McMillan and Schumacher, 1993). Reliability of information within a case was increased by verifying information through other documents especially where such information was found to be contradictory, to ascertain probability. In cases where triangulation encountered conflicting information, for example differences in time schedules, the primary sources were regarded as the correct sources, especially when consultation with informed sources could not be solicited. Although reciprocity was not intended in the study, communication with the scholars involved in similar studies also assisted with verification of information. The nature of information gathered shows activity trends that are somewhat historically based and therefore not a threat to internal validity.

3.5 Analysis

Analysis in this study followed the analytical induction process, with each step being informed and guided by information obtained from the previous steps. Qualitative methods of information utility and interpretation underpinned the analysis of data in this study. As previously stated, the search for information was coupled with criticism and analysis throughout the study, both on primary and secondary sources as suggested by McMillan and Schumacher (1993). The former (primary sources) mainly provided explanations of concepts

and processes while the latter was used to verify the relevance of sources. Both types allowed interpretations to create a basis for further investigations. At the initial stages of exploring information for the location of sources, care was taken not to base critical analysis on the assessment of authenticity but to use the exercise for location of authentic models/cases and identification of trustworthy sources that would provide reliable information to take care of reliability and validity effects.

The interpretation of information initially took precedence over the goal of causal analysis in order to understand each model/case with emphasis on particularity. This, Ragin (1987) advises, assists in treating each case as a distinct theoretical type. In this manner, evidence could be obtained in what Ragin terms a manner sensitive to historical chronology in an attempt to account for significant historical events.

Analysis nonetheless moved beyond the attainment of descriptive facts to the questioning of the reason for occurrences. Some causal explanations from scientific articles helped develop an understanding of past and present practices in different countries. At another level, interpretation of facts for causal conclusions was made through the review of legislative documents. Both the scientific and legislative documents were instrumental in creating an understanding of circumstances in which policies relating to research evaluation were developed and implemented, the feature which became the most important for the study.

The descriptive and narrative nature of concepts on research evaluation practices increased authenticity, while the simultaneous interpretative analysis of models/cases gave a broader understanding that guided the intentions of the study. Commonalities and differences of practices in different countries enriched data and maximized analysis, serving as directives for the analysis of local cases. For example, similarities and differences in aims / intentions for evaluation, contexts, manner and methodologies used and intended outcomes all set a pattern for information analysis.

The process then led to the shaping of perspectives and a conceptual framework for interpretation. In this way, some level of comparison unavoidably influenced understanding on analogical interpretations and helped determine plausible explanations for the similarities and differences of organized systems of information.

At the end, individual concept analysis within and between cases led to some causal inferences and generalizations in a collective sense and this informed the establishment of a concept that appeared more suitable to the South African situation. Conditional analysis strategies were used to study the national (South African) context of research and evaluation efforts. Differential comparisons with other models used in the study allowed association and the development of a framework.

PART TWO

INTERNATIONAL MODELS OF RESEARCH EVALUATION

(The following models are hereby referred to as international by virtue of them being models of other countries)

Chapter 4 The Dutch model of research evaluation

4.1 Background

The education system in The Netherlands is a unified one regulated by central government laws. While this applies across the public education system, partial autonomy is granted to universities, whereby government lays down a legislative framework within which executive boards and university councils operate (De Vries, 2001). University management structures look after daily management and administrative activities and are accountable to the government and to the Ministry of Education in particular.

The evaluation of subsidized HEIs in The Netherlands dates back to the mid-1980s. The process came about as an attempt and/or response by a collective of institutions to match the autonomy granted by government with expected levels of accountability that went along with processes of restructuring of HE (Maasen, 1997). A deal was struck with the Ministry of Education in which institutions promised to set up a system of quality assurance that would monitor internal systems, and then produce reports (De Vries, 2001). This resembles a situation that Rip and van der Meulen (1996) regard as high steering/high aggregation, in which the two bodies (systems) co-exist without dominating each other. With the South African HE system taking the same direction (of institutional accountability) it was decided to include a study of the Dutch model.

4.2 The Evolution of Research Evaluation in The Netherlands

The history of Dutch research evaluation is discussed in this section. Whereas discussions follow the historical pattern, focus is placed on the following areas:

- Origin and rationale
- Purpose for evaluation of research
- Processes and procedures followed
- Benefits
- Lessons to be learned on good practice and challenges.

In common with other countries, research evaluation in The Netherlands has always been conducted concurrent to the evaluation of teaching and learning (De Vries, 2001). This is

confirmed by CHEEC (1993), Maasen (1997) and van Vught and Westerheijden (Westerheijden, 1999), who state that research evaluations are coupled with those of teaching in so far as they are important for the teaching process and should therefore not be viewed as a separate entity. Procedures of evaluation that separate the two are, according to Westerheijden (1999), synchronized to avoid confusion and complications. The separation also enables faculties to gain a clear and overall view.

The university system went through stages of transformation of this systemic and comprehensive (large-scale) research evaluation with a somewhat continuous and developmental approach. The granted autonomy is characterized by levels of accountability to a quality improvement system. The system of evaluation qualifies as Campbell's (Shapira and Kuhlmann, 2003) Type A category, by virtue of being systemic and comprehensive and operating at a national level across all disciplines.

One important observation is that the aims of research evaluation in The Netherlands evolved over time, as did the role players. Verkleij (1998) attributes such changes to changes in the objectives and contexts within which such objectives are realized, which the author refers to as "decision context". The evolutionary stages of research evaluation are discussed below. Whereas reference is occasionally made to different authors in the discussions, more information was obtained from policy documents, referred to as "Protocols":

4.2.1 The Initial State-control era

During the state control era in the initial stages of the review system in the early 1980s, research evaluation was characterized by the intention to encourage research participation through the allocation of funds, based on the requirements of the research (inputs). This tendency to rely on pre-research (ex-ante) information is commonly used by research councils. It is necessary to note that the process of funding was still internally based (within the allocation system of institutions) and did not influence the allocation of funds by the government, especially because research funds were coupled with other allocations in the block grants. In this way, institutions were left with the responsibility to encourage research participation, with more funds allocated to those with higher levels of research activity. This traditional model was used mainly in European countries at the time, and is said to have had little, if any, influence on innovation (De Vries, 2001). The allocations were awarded without any proof of completion of research.

4.2.2. The Conditional Funding (CF) Era (1982-1992)

The need to move away from the previous governmental “give-away” model to the “value-for-money” philosophy in research evaluation became apparent around 1982. This conditional funding (CF) era was the first in the order of accountability to government funding with its accompanying level of autonomy (De Vries, 2001), although the “funding decision context” process was still based more on the idea of ex-ante requirements. The process was based on newly performed programmes, and did not consider priorities or differences in research aims. As previously stated, accountability for government funding served as the main goal - the practice being to fund “good” research for a period of five years, guided by the results of peer reviews.

The main purpose of government involvement at this stage was to ensure efficient and effective use of funds, where efficiency was measured on the basis of goal achievement. As stated, performance in this ex-ante model was judged on the basis of the achievement of research aims despite the lack of concentration on their differences. Good research would thus be qualified by peer reviewers on the basis of criteria determined for the purpose, similar to the arrangements of research councils.

As developments unfolded, performance was based more on research activities (process) and output, focusing mainly on publications. It was expected that problems related to publication-based assessment would be encountered and this system was no exception. The idea of basing performance on objectives and also evaluating through-put and output is a developmental one and when coupled with feed-back on assessment/evaluation from external visitors (reviewers), is said to contribute to the improvement of research activities. Figure 4.1 below gives a summary of activities regarded by Verkleij (1998) as the second leg of the conditional funding process.

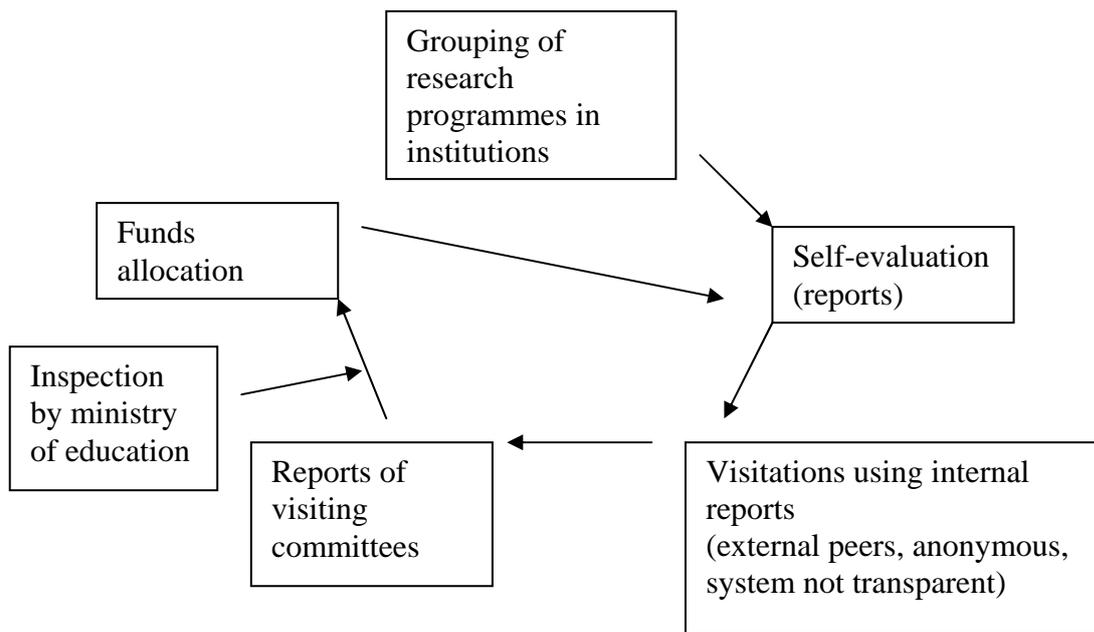


Figure 4.1 Summary of a cycle of events in the CF era (based on accountability of autonomous universities):

The establishment and utilization of the conditional funding system was guided by reasons/goals such as the promotion of quality, the encouragement of systemic discussions, the effective use of resources and for accountability to government (funding). The process of evaluation was carried out by a group of “anonymous” and local committees composed of members who were occasionally criticized as being non-critical. The Royal Academy of Arts and Sciences, known as the Koninklijke Nederlandse Academie der Wetenschappen (KNAW), which is regarded as “the most distinguished academic body” in The Netherlands (Jeliazkova and Westerheijden, 2004:7), was involved in selecting external review committee members (panelists) with the understanding that peer review is one of the most acceptable instruments in the evaluation of research performance.

Costs in executing the process were regarded as moderate and so were the benefits. One contributing factor to moderation of costs was the use of performance indicators, which are typically quantitative. Westerheijden (1999:447) warns against the use of performance indicators and refers to them as “no more than signals that indicate where evaluators need to ask questions” and as “not only partials, but also proxies to an unknown theoretical variable”.

According to the author, performance indicators can only generate data and not credible judgment that would lead to innovation.

A few advantages nevertheless accrued from the CF system. The process of external examination assisted in grouping similar research together into what was later referred to as “research programmes”. This was later used as a basis for internal management systems and served as a basis for the development of research policies by the Ministry of Education and Science (Jeliazkova and Westerheijden, 2004). Peer evaluation also contributed to the idea of nationalizing the quality of research.

The intention to encourage systemic discussions was also realized to some extent. For example, networks of post-graduate research assistants were created, which later constituted “research schools”. Accreditation of such a school by the Royal Academy of Arts and Science is regarded as prestigious in The Netherlands and the idea, even when it is fading, is still highly regarded because of its inter-university and collective nature.

Of importance was the reluctance of the review committees to qualify any of the reviewed (evaluated) programmes as inefficient and the failure of the committee to rule any of the programmes as excellent. This may be seen as accommodating researchers and programmes at different levels. While this has the potential to reduce the strain of a typical “value for money” model, and may act as encouragement for all efforts to be exposed to the process, there are problems attached. Universities would allocate funds without any criteria for control as reallocations were left to their discretion. The casual nature of evaluation, judgment and ruling made the funding of “good” research more difficult, defeating the initial intentions of the system.

The CF process did not last for long, but served as a learning experience for other initiatives that followed. A trend for a systemic and comprehensive/large-scale research assessment had been set, which would serve as reference for The Netherlands and other European countries.

4.2.3. The VSNU Quality Assessment of Research (1993-2001)

4.2.3.1 Background

Experience gained during the application of the Conditional Funding policies gave birth to an organization to take care of research evaluation processes for public universities that would

guarantee quality of performance. This idea came at the time when government strengthened its support for HE. It is notable that changes in the decision context were developmental and progressive based on experience, with one level informed by the activities of the previous one.

The Higher Education and Research Act of 1993, known as the *Wet Hoger Onderwijs* (WHW), was meant to regulate among other things, research in HEIs. The Act replaced the Higher Education Act and a collective of other regulations that preceded it. The 1993 Act defines government involvement in research evaluation as an ex-post (after performance) attempt to remedy irregularities in the institutions, with very minimal interference in HE activities. In return (for the minimal interference), institutions were expected to maximize and maintain quality of services through a process of self-regulation (Jeliazkova and Westerheijden, 2004).

4.2.3.2. Purpose

The purpose of research evaluation was becoming clearer at this stage. According to Protocol (1998), the main intention of the comprehensive evaluation of research in The Netherlands is the improvement of the quality of university research through processes of self-regulation within universities, faculties and research institutes across the country. Results of evaluations in this era were followed by advice on how such quality may be attained and enhanced. Most notable is the purpose of quality assurance fulfilled through the accountability procedures embedded within this notion of self-regulation. The notion is emphasized in order to stress that the act of placement of group efforts within the “context of group’s environments” is self-selective and not directed by opinions of others such as the external groups. Evaluation at this level includes review of the institutional, faculty (and group) missions for research and has the tendency to also contribute to the appraisal of what the Protocol (1998) refers to as “the state-of-the-art” in the discipline or academic area.

4.2.3.3 Operations

The initial body to take up the responsibility was the *Vereniging van Samenwerkende Nederlandse Universiteiten* (VSNU), referred to as the Sector Organization for the Traditional Universities, which was subsequently followed by the *Hoger Beroeps Onderwijs raad* (HBO-raad), referred to here as the Sector Organization for the Universities of Professional Education. According to de Vries (2001), the system of quality assurance was

based on the idea of internal assessment/evaluation (self regulation) and external quality regulation by panels. The latter process allows for the evaluation of similar groups of programmes to be conducted by review panels on the basis of self-evaluation reports and on-site visits.

According to the CHEES document (1993), the VSNU is owned and funded by the universities collectively, to take care of formative (process) and summative (ex-post) evaluation. This ownership is further explained by Lucas (2004:4) who states that the “VSNU represents the interests of the universities *vis a vis* political, government and community organizations.” That is, the organization exists to further the aims and interests of universities.

The summative level of evaluation has been suggested by Campbell (Shapira and Kuhlmann, 2003) as appropriate, as it gives enough time for the completion of basic research. In this process, the goals for formative evaluation are meant to surpass those of the latter. In 2001, the VSNU was in its third 6-7 year cycle of evaluation while the HBO-raad was in its second (independently in a binary system of HE). The bodies have responsibilities such as the formulation of the general protocol (which serves as a standard frame of reference), formulating plans of operation (including organization of panels) and implementation of the plans together with the publication of final reports.

In both sectors, all public institutions form part of this process and are exposed to the two processes of evaluation. Self-regulation relates closely to institutional autonomy and is believed to lead to outputs within “an acceptable range” (Jeliazkova and Westerheijden, 2004). According to the original plan, the two processes of evaluation were supposed to be “monitored” by the “Inspectorate”, referred to as IHO, who would represent government. According to sources (verbal consultations) this process of monitoring did not take place in research evaluation but did so in the evaluation of teaching and learning.

Unlike in the CF era where the purpose of government involvement was mainly to ensure efficient and effective use of funds, involvement during the VSNU period was aimed at monitoring and the assessment of improvement (on quality) processes. Had this process of monitoring happened, the change in intention to evaluate (decision context) would bring along stricter measures of evaluation and intensified responsibility levels. The intention was

that at this level government would be directly involved and under worse conditions of programmes, introduce sanctions. It is not clear why this plan for government intervention has not been implemented. This left the “sector organizations” with full responsibility shifting the governance continuum to more aggregated extremes. Inefficiencies, if established, followed by recommendations (for improvement) by the committees and such information is obtainable in the published reports.

The new idea also went beyond the idea of conditional funding by referring to dimensions that would define quality on which peer evaluation would be based. Although these dimensions have been outlined in chapter 2, this chapter highlights similarities and differences of definition over time (with evolution). The dimensions are briefly discussed below:

- Quality of output

According to the Protocol (1998) research/academic quality refers to, and is based on, the quality of output of the research groups in relation to dissertations produced, academic publications, professional publications (where necessary and contributing to the development of a discipline), patents and other similar academic research products. Participation in international co-operative projects and involvement in editorial boards, invitations to international conferences, academic awards, visiting professorships and acquisition of the Nederlandse organisatie voor Wetenschappelijk Onderzoek (NWO) funding all serve as criteria for evaluation and as indicators of quality.

It is important to note that the traditional outputs, especially articles in international / top journals and citations, which may be thought of as a disadvantage in the Dutch situation because of language problems (minority language) still served as an important measure in defining quality.

- Productivity

Productivity is defined by relating output to the input of human resources. Output is mainly characterized by the number of publications in total and per category. Productivity and quality are compared with the objectives and missions of the submitted research programmes and their differences are taken into consideration. Productivity

levels could be used for the establishment of programmes and for the encouragement for institutions to be accountable

- Relevance

Relevance is established from the group mission regarded as specifying context, academic framework and the line of science followed by the programme. Academic and social or professional impact anticipated from the research outcomes is assessed from the research group's viewpoint.

- Viability

International competitiveness is regarded as a strong indicator for future standing in research. This is determined by the area(s) of study(ies) selected as well as approaches used. Consistency of programme elements and the availability of important features such as infrastructure are determinants of a viable situation. The main idea is the determination of future positions obtainable in the groups' profiles.

The international academic arena features as a very important benchmark across the evaluation process as indicated in the above dimensions (aspects of evaluation). On defining quality, involvement in international projects, conferences and editorial boards are dominant determining factors. In this way, productivity would evade international influence. The Protocol (1998) does not relate output (the number of publications) to international journals, even though this can be detected when quality is discussed. Viability and relevance rely mainly on international comparisons, although reference is made to the research mission of the group.

4.2.3.4 Assessment systems of the VSNU

Two important systems of evaluation are identifiable in this process; the internal (self-evaluation) system, which serves as the initial level and provides information for the second level, the external evaluation system. The two systems are explained below together with the methods and strategies used in the systems.

4.2.3.4.1 The self-evaluation system

Maasen (1997) explains that the content of self-evaluation is dependent on the faculties and departments whose programmes are evaluated. The faculties decide which programmes should be evaluated and stress the points which they regard as important for the process.

An official format provided by the VSNU guides the process of self-evaluation and institutions are expected to provide logistic and financial support. Checklists are used as guidelines for comparison, and to obtain common patterns of submitted data across disciplines. Care is taken that submitted data is factual in order to avoid reverting back to the total/complete dependence on performance indicators, which were considered in the previous process. Data is also said to be neither purely quantitative nor qualitative, nor is it a list of minimal requirements. Instead, it uses the programme's own goals as a point of departure (CHEEC, 1993).

4.2.3.4.2 External evaluation

The VSNU committee operations are guided by rules laid down in the general Protocol. To prepare for the evaluation, the VSNU central committee constitutes discipline/area committees which then set their own frames of reference and elaborate on the standards set by the sector organization (VSNU). The discipline committees will prepare their own discipline specific Protocols. The discipline Protocols contain specific points of foci for the committees.

- Submissions and visits

Faculties are expected to submit data on their performance (description of results) over a period of five years for each programme. Submissions for each programme include a brief outline of its mission, describing the scope of research, goals and objectives/outcomes of the faculty/discipline, the nature (for example pure or applied science) and an overview of the programme. The latter provides information relating to the motivations for research and the direction which the research takes to address the identified needs. This includes the audience and stakeholders and the levels of relevance of the research for academic and community purposes. The objectives/outcomes should state the type of products anticipated from the research. Examples of the products are reports, articles, patents and others that may be accrued in the process or at the end. Submissions also include developments that may be expected to emerge within the context of the stated research.

Profiles of staff inputs are expected in a quantitative format and this is to be accompanied by five publications which are viewed by the faculty/discipline as essential. According to the Protocol (1998), lists of other publications by the faculties/disciplines must also be included. When the need for site visits arises, for example for verification of information, and when conditions are feasible, the visiting committees spend two to three days per discipline, meeting with individual programme leaders.

In evaluating/reviewing, the committees consider the position of the programme within the framework of its mission. An impression element is attached to the report and this is based on aspects such as the perspective of the research. Feedback serves as an important element of the evaluation process and as already stated, all results are available in published reports. While the process looks complex, time-consuming and expensive, it is more transparent and somewhat more democratic and it is not surprising that the results are relatively acceptable to institutions and researchers (Protocol 1998).

4.2.3.5 Methods of data collection and information handling

The following methods are used to collect data in this system:

- Document review of reports and written information submitted by faculties/research groups (both qualitative and quantitative) including mission, objectives and outputs;
- Visits by peer reviewers to faculties and laboratories for observation; and
- Interviews with research programmes leaders during visits.

4.2.3.6 Data analysis and recording assessment

In this system the committees are expected to rate programmes on the dimensions (quality, productivity, relevance and viability) of assessment. Scoring is performed on a five point scale ranging from 1, representing insufficient, to 5 representing excellent. Different disciplines (groups) of research are scored separately and submissions by faculties follow the system of programmes previously used by the CF system. A separate body is sometimes responsible for the bibliometric analysis of publication data.

In some instances committees may base their findings solely on document analysis while in others the other two methods may be used for triangulation purposes, depending on the level

of verification needed. Results are published nationally in reports and are mainly presented in numerical summaries for each programme, categorized into the four dimensions previously outlined. If a programme would be observed to be weak, a warning would be signaled and the institution would be expected to implement rapid improvements. The fear of consequences that may result from sanctions is in itself motivation for improvement and in most cases such improvements were effected. The purpose of improving research quality is thus emphasized.

4.2.3.7 Utilization of results

It is important to note that results (in the reports) do not relate directly to any reward or punishment (except when improvements are expected) and are therefore not necessarily utilized to allocate funds. It is also necessary to note that the monitoring system intended to be conducted by the Inspectorate (representing government) would not prescribe how institutions should respond to visiting committees' remarks, for fear of undermining the autonomous status of the institutions. What was intended as important to government was the drafting of a plan (by the affected institutions) for the implementation of the recommendations of the evaluation. In the absence of such a plan, authority would not be convinced that quality would improve. The Ministry of Education and Science would take the necessary steps after sending a warning signal (Jeliazkova and Westerheijden, 2004). As stated, the existence of the process involving government has been refuted by other sources.

Once the evaluation report has been produced, an inspection would be done, first to judge the quality of the report by the discipline-specific committees and published by the central VSNU committee and then to identify programmes said to have serious shortcomings. The initial plan was that once the reports have been published, the Inspectorate would conduct what de Vries (2001) refers to as meta-analysis, focusing on matters of accountability, improvement and processes of visitation. The report would then forwarded to the Ministry of Education with remarks from the Inspectorate. This process appears to have happened with the evaluation of other core functions and not research.

As already stated, in cases where a programme is identified as having "serious shortcomings" by the visiting panel, the programme management is expected to provide a plan of action to improve the situation. After a period of two to three years institutions have to show evidence of the implementation of the report's recommendations. If shortcomings have not been

“remedied” and no measures of improvement are visible, the programme is regarded as having “long-lasting serious shortcomings” with yellow card characteristics. The institution may submit another plan for improvement. If such expected improvements were not visible, possibilities of government subsidies withdrawals existed. This is one level at which the Dutch research evaluation is linked to funding. De Vries (2001) indicates that most such programmes are withdrawn by the institutions even before any form of intervention. Otherwise, rapid improvement is expected, which will lead to another process of self-assessment for the next round, completing the cycle of events of the process of evaluation portrayed in figure 4.2 below.

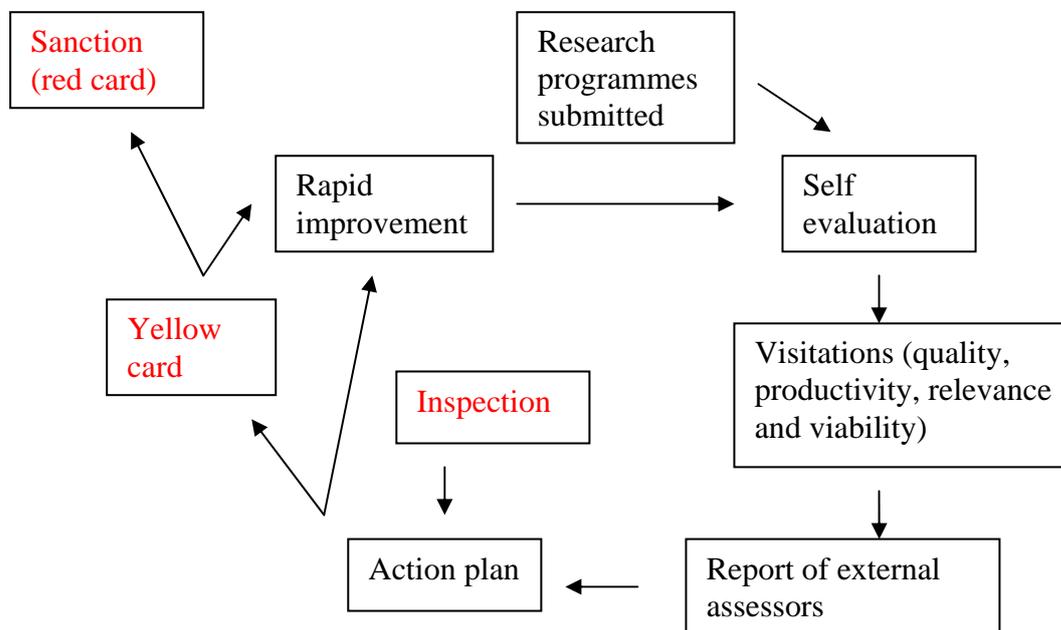


Figure 4.2 Summary of the VSNU cycle of events of assessment (the letter in red indicates services that would be implemented by government that are said not to have taken place):

4.2.3.8 Benefits

There are identifiable benefits of the programme. These include:

- The process of self-evaluation does not only prepare programmes for external assessment but also promotes internal quality management.
- External visits help the institutions to manage and improve on quality and this, de Vries (2001) says, is not well clarified.

- It is alleged (De Vries, 2001) that 60% of the recommendations are accepted and implemented. Programmes that have been identified as having “serious shortcomings” are more likely to implement the recommendations than those with established systems.
- There is a high level of acceptance of the process and report results.

Results are also said to have desirable effects such as external legitimacy and to assist the administrations of institutions by providing them with unbiased information on the status of their departments. The results may also be used by institutions for the differentiation of internal funding. An unfortunate situation arises when such research results are not used, especially in the absence of a link between the outcome of results and processes of decision-making.

Verkleij (1998) also outlines some of the factors that may contribute to acceptance of results. First the international component of the visiting committee is an important aspect as it contains global thinking and participation. International recognition is obtained when research output published in international journals is treated as a measure of good performance. This is important in the South African context which has international participation as one of its imperatives. Verkleij (1998) identified the involvement of faculties, especially deans, as an important contributing factor in the setting of terms of reference. Face-to-face interaction with assessment committee members, which combines assessment with advice, together with the multi-faceted nature of the assessment system, also contributes to the improvement of self-assessment skills.

Verkleij (1998) also alludes to the relationship between acceptance and intended goals of assessment (decision context of assessment). For example, if the goals are punitive, researchers will be negative and may choose to be defensive. A friendly approach is bound to improve levels of discussion between the committees and the researchers. The open sessions occasionally used by the visiting committees to reach out to all researchers during the visits are an example of a friendly approach.

Although this system of evaluation is said to be costly, it has been found to be more developmental and to yield more benefits. The contribution of evaluation to the effectiveness and efficiency of the system cannot be overlooked. A “clean system” which relates to a

system of selective funding where performance indicators are used and assessment is only based on past performance appears more efficient, but displays low features of quality improvement.

4.2.3.9 Challenges

Research in The Netherlands, like in other academic programmes, is suffering from what Jeliaskova and Westerheijden (2004:19) call global “small country complex” as a result of such factors as the “minority language” in the global sense. The inclusion of international scholars in the external visiting teams is an attempt to improve recognition by others as well as to expose and extend Dutch quality standards to the outside world. This is especially important because the country subscribes to the idea of international transparency and their quality policies are said to serve as examples for other European countries (Jeliaskova and Westerheijden, 2004).

Besides the “minority language” problem, the system encountered other problems commonly experienced in peer review. For example, Verkleij (1998) observed the tendency by external assessors to “make their own choice”. The following are some of the commonly observed challenges:

- Interest in assessors’ own fields.
- Fulfillment of audience/stakeholders’ interests by members of the visiting committee. Influential personalities such as principals are an example. There is a danger that the committee members may be more interested in pleasing the influential parties than in enhancing research.
- Problems of the assessment process itself that direct the selection of information to be sent to the external committees, which also affects modes (strategies) of communication between the researchers and the researched.
- In some cases there were detectable differences in the quality of judgment, a problem of inter-subjectivity within a committee.
- Impressions made by the Inspectorate are said to have contributed minimally in terms of standardizing judgment by the visiting committees.
- Research effects, which Westerheijden calls “visiting games” (Verkleij, 1998) also surfaced. This refers to strategies used by the researched to please a review

panel, how they handle interviews and how they tackle difficult questions in an attempt to retain autonomy and funds.

- Researchers may also tend to abide by assessment systems and just do what is expected of them, which compromises quality and improvement of products.
- The process is then blamed for lacking continuity and having too many successive visitation cycles.
- Difficulties in comparing reports across different programmes in different disciplines also create problems.
- The intended quality, which would be achieved through the involvement of international peers (the international component) is said to be disputable in some cases.
- There are often delays in the processes of inspection, especially in the compiling of meta-evaluation reports and Inspectorate follow-ups.

Verkleij (1998) also warns of possible problems that may result from the definition of the concept “research quality”. This is exposed by the use of “the number of international publications” as an indicator. The author also signals the dangers that may result from the inability of an evaluation system to handle diversity and differences in missions and standards between disciplines. Also, there may be problems when criteria used are not compatible with evaluation objectives. The latter may result when context changes and criteria fail to do so. For example, criteria for assessing quality will also be influenced by changes in the purpose of evaluation.

4.2.3.10 Composition of the visiting committees

The review panels in this cycle are made up of experts in the field of investigation including international connoisseurs. In most instances committees were composed of five members and occasionally included one foreign expert in the field of study, an educational expert (whose programme is not being reviewed), one or more members from the Dutch HE system, some members from professional bodies (CHEE, 1993) and the chairperson (who is regarded as independent). In some cases student representatives were involved but this has not been common. As previously stated, the selection of a foreign expert is an attempt to assure some international perspective. To make the process more effective, the expert should have knowledge of the Dutch language. The chairperson is an appointed member who should be of Dutch origin. The VSNU undertakes this appointment in consultation with the KNAW.

The members are proposed by the deans of the different academic/research disciplines and representatives of programmes and then nominated by the sector organization (in this case the VSNU) board.

All frames of reference are based on the standard framework set by the sector organizations involved (VSNU or HBO-raad). The committee focuses on problems encountered, the outlook, expectations and recommendations. In the case where multi-disciplinary or interdisciplinary research is involved in the assessed programme, a panel is allowed to consult with experts in other areas.

4.2.3.11 Government involvement

The information that follows has already been highlighted elsewhere in the previous sections and is included here for further explanation. As stated, it is important to note that the information is disputed by other experts to have taken place in research as it is mentioned in the stated referred to sources. It was nevertheless important to include this section as it gives a picture of a possible government intervention. If this process was planned to take place and did not, this study did not detect reasons for the exclusion. The absence of the process is itself a shortcoming that needs further investigation.

According to de Vries (2001), the Inspectorate of Education (IHO) in this system is left with the meta-evaluation task and follow-up responsibilities. The author enumerates these monitoring responsibilities as those of “guarding the guards” by making judgment on the reports, making follow-ups on the programmes after the results have been published and checking plans for improvement. The Inspectorate is also involved in approving the standard frame of reference and Protocol developed by the sector organizations and evaluates the activities of the committees and panels. Once the Inspectorate has communicated remarks to the Ministry of Education, the department may then communicate with the institutions, especially those with identified, long-term, serious shortcomings.

4.2.3.12 Analysis of the initial VSNU research assessment exercise

Different authors have analyzed the system of evaluation and anticipated changes that would be brought about by the progressive and developmental nature of the Dutch system of research evaluation with its changing decision contexts. According to Verkleij (1998), these exercises (including those prior to the VSNU operations) allowed institutions to initiate

internal evaluation and to recognize external evaluation expectations as part of their strategic plan. This enabled them to realize their strengths and weaknesses to the extent that they would be able to act independent of national intervention in future. In other words, self-evaluation seems to have potential to serve as a future trend.

The experience enabled institutions to use peer committees to also validate their internal processes which led to the improvement of the compilation of self-evaluation reports. In this way, individual institutional improvement would be guaranteed. The reports are said to be used as evidence to support applications for and to attract external funding. This is one level, other than the conditional funding level, in which funding is attached to the process. Verkleij (1998) predicted that the institutions would view quality improvement as their own responsibilities, which would translate the responsibility of the assessment committees into advisory roles. In such a system, de Vries (2001) says, the review panels would no longer uniquely be organized by the sector organizations.

Developments over time gave birth to the notion of “fitness of purpose” as a quality criterion, implying that support processes and policies added value to quality standards. Issues such as societal relevance became prominent foci as they prompted what Verkleij (1998: observation (d)) refers to as “ambitions for the imbedding of university research into societal priority scheme”. This idea is also relevant in South Africa, which prioritizes social and economic development.

The 1993-2001 evaluation cycle marked the beginning of a trend, which drove the system of research in HE in The Netherlands towards improvement and had institutions preparing for better performance in the next cycle.

4.2.4 The Era of the Standard Evaluation Protocol (2003-2009)

Information in this sub-section has been obtained from the Standard Evaluation Protocol (document). Planning has been based on the intention to improve research for the purposes of regular accountability. All research organizations are expected to follow the same procedure and autonomously define their units of evaluation, which may be the programme and/or institutes/research units. The “institute” is referred to by the Standard Evaluation Protocol (2003-2009:13) as “a group of researchers with a shared mission operating under the same management where different research groups can be under the same unit”.

Research/sector organizations are also expected to draft specific Protocols for internal and external evaluation, compose evaluation committees, design the evaluation programmes, finalize and publish evaluation reports and conduct meta-evaluation in this era. Information in this era excludes participation of government, as the entire process of evaluation and analysis thereof becomes a full responsibility of the research/sector organizations. Government receives accountability reports. In the plan, schedules are produced for all institutes, with a timetable for evaluation. Draft Protocols are discussed with the institutes before they are finalized.

4.2.4.1 Background

This is a new and current cycle of events initiated in the year 2003. In the light of development and progression on matters of research evaluation in The Netherlands, a cycle is usually followed by a review of activities to determine areas that warrant improvement. Following the 1993-2001 cycle, a report, “Kwaliteit verplicht”, referred to as “Quality obliges”, was compiled by the Kwaliteitszorg Wetenschappelijk Onderzoek (the Quality Assurance Scientific Research). The report outlined a new national evaluation system for public-funded research in the Netherlands. The initiative in this exercise was taken by organizations responsible for publicly-funded research namely, the Royal Academy of Arts and Science / Koninklijke Nederlandse Academie der Wetenschappen (KNAW), the universities (in the name of the VSNU) and the Netherlands Organization for Scientific Research (NWO), all referred to in the discussions as research organizations. Previously these research organizations were referred to as sector organizations.

While the KNAW featured in the previous cycle, it was playing an advisory role and the NWO was only cited as a criterion for prominence. It is notable that this cycle emphasizes the notion of accountability by referring to the involved organizations as those responsible for public funding. This era introduces evaluation of research beyond that of university levels. This study concentrates only on the evaluation of university led by the research organization (VSNU) and this does not reveal the functions of others (for example the departments and business and others)

Fourteen universities have been organized into units and each institution still decides on the programmes it wants evaluated – this responsibility being carried out by the university

bodies. The VSNU continues to be the main co-ordinator of the university evaluation exercise even though some overlap of programmes with other structures such as the KNAW has been noted.

4.2.4.2 Mode of operation

The six-year national evaluation process is being adhered to. This cycle/term had been planned to operate from 2003 to 2009. In the exercise, research units (institutes/departments) are given three years to conduct self-evaluations and produce reports. The reports serve two main purposes: as a mid-term review and also as preparation for the external evaluation to be conducted at the end of the term. Institutions are also expected to embark on a yearly review system to monitor data for evaluation in a systemic way. According to Lucas (2004:5) the “important feature of this system of evaluation is the development and reliance on self-evaluation whereby institutions can to some degree set the goals and expectations in keeping with their own particular priorities and missions.” It is at this level that competition and differentiation may be introduced in that internal systems would have been strengthened.

Past performance and future plans serve as determining factors. Past performance provides a future base and some level of confidence for prospective research, and this may be confirmed by results in the prospective analyses. Once a report is compiled the evaluation committee submits it to the board of the research organization (VSNU in this case). The organization’s board will then discuss the report with the institution and assist with policy decisions for that institution.

The evaluation results for each institution are determined from the submitted report together with the policy decisions arrived at in the discussions with the VSNU board. The research organization has a responsibility to periodically (annually) report to the Ministry of Education, Culture and Science on activities of evaluation, and reports and annual self-review results serve as evidence for public accountability. The evaluation results therefore serve as units for submission.

A situational analysis in the form of a SWOT analysis is also expected from institutes/research units and this may be used to draw conclusions on research programmes and future planning. This process attempts to reveal strengths, weaknesses, possible developments that are positive, and those that pose threats. Conclusions are then drawn and

these may influence a change in policies, objectives and strategies. Plans for the implementation of recommendations are also made.

4.2.4.3 Management of processes

Institutions and research units are grouped under one of the three participating research organizations. These bodies are responsible for the procedures of research evaluation of institutes/ research units falling within their jurisdictions. As already stated, the VSNU continues to coordinate the activities of universities.

Each institute is managed by a director or a board and in some cases a research leader. The institute may be composed of one or more research groups, wherein a group operates within a framework of research programmes under a leader. As university boards select the programmes for evaluation, the VSNU encourages disciplines and interdisciplinary synergy between institutions.

4.2.4.4 Aims and objectives of the system

A shift in decision context is the main determining factor for development and progression of the Dutch system (Verkleij, 1998). In this system, the main function of the research evaluation is to assist researchers and research units to make better decisions about future research management and policies, a strong component of quality improvement and viability. Lucas (2004) sums up the purpose of evaluation by referring to it as a tool to inform strategic policy in terms of priorities and use of resources.

The system intends to concentrate on the improvement of research quality with reference to international standards on quality and locally relevant.. This enables the system to contribute to the improvement of the quality of science. This implies that evaluation will be inclined to base judgment of the components on international standards, attainment of which is a benchmark set by the organization to serve as a requirement and criterion for accountability. As already stated, public accountability is now becoming a strong requirement and is said to also be an inherent element in “the improvement cycle in which this scheme of evaluation plays a dominant role” (Protocol 2003-2009:7). The socio-economic impact and the multi- and inter-disciplinary nature of research have now come to the fore. In addition, the process intends to focus on the improvement of levels of research management, leadership, strategies and policies.

Institutional missions are still considered, though in a much broader sense. For example, an institution may hold socio-economic goals in the highest esteem while others focus on technical or infrastructural functions. Differences are also expected in the types of research and patterns of communication. These all have to be assessed in terms of standards of quality and relevance. One way in which evaluation may acknowledge these differences is through the constitution of the evaluation committees or even during data collection processes. The research organizations intended to extend accountability by having the system accessible through the internet.

4.2.4.5 Evaluation criteria

Two primary levels of evaluation are identified in this system; the programme assessment, which focuses on the results and quality as well as the viability of research, and the institute assessment in which strategies and aspects of an organization are engaged. Criteria used to assess both levels are said to be the same and reflect the comprehensive performance of an institute and/or research groups together with their future potential. Intra-science differences in the use of criteria for evaluation between different research fields, for example natural sciences, social sciences and others, are noted and these inform the levels of elaboration of a criterion per field. For example, publications may be a very important indicator in one field while book authorship is important in another.

There exists a slight alteration in criteria from the previous system. Whereas most of the components are retained as criteria, meaning has been expanded to suit the new context. For example, the definition of quality is emphasizing international standards and recognition more and weighing these against innovative potential, while productivity still depends on scientific output. A slight emphasis is nonetheless attached to bodies such as the ISI which are major global references (quality research is said to be recognized through the ISI). Relevance, on the other hand refers to the impact of science on socio-economic systems. The component of vitality is added to viability and is expressed in the new Protocol as vitality and feasibility, directly linked to organizational matters especially those that deal with innovative potentials.

Scoring has not been altered although the five-point scale is still being used. The process takes care of inter-subjective differences within the committees by reporting differences in

judgment whenever they occur. Institutional missions provide direction which evaluation “thinking” has to follow. For example, nationally inclined research will differ from the research with international proclivity. Elaborations on the criteria/components are indicated below:

- Quality

The new Protocol (2003-2009) defines quality as a measure of excellence, relating this term to the eminence of research efforts and the ability to perform at a prominent level in terms international standards. This ability is said to be realized through the proficiency and rigor of research concepts as well as success in scientific development (improvement of the quality of science). The instrument used to measure quality ranges from discussions (interviews) with research leaders to expert judgment (observations and document analysis) by members of the evaluation committees. The quality of facilities used for research also serves as an indicator of excellence.

- Productivity

Quantitative information on publications and citations, in the form of bibliometrics are recorded as indicators of productivity especially in the natural sciences and life sciences. This is despite the problems attached to the ISI citation index, for example, the between-discipline in impact scores. Technometrics (patents and citation patents) and socio-economic products of research are also counted as output. Similar to practice in the previous exercise, output is reviewed in terms of human resource input. On the dissemination of the reports, the issue of internet publications is still being reviewed by the organization and may be used in future if considered credible.

- Relevance

The scientific, technical and socio-economic impacts of research define the component of relevance. The possibility of expanding the definition from the one used in the previous system was raised by Verkleij (1998) who considered socio-economic relevance an important area in research. The same international concept stated in the Protocol (1998) reappears in the standard evaluation Protocol (2003-2009), with more emphasis on society-based, socio-economic and technological contributions.

- Vitality and Feasibility

Protocol (1998) referred to the fourth component as academic viability, which could be detected through different aspects such as infrastructure availability and other research needs that would impact on the future position of the group. Vitality and feasibility seem to carry the same meaning, though expressed in a slightly different way. The choices made and the success rates of the project are the main features. The position is also explained by the flexible and sensitive ability of a group to discontinue non-viable projects and initiate new ventures. Measurement of professionalism in management and policy decisions in both the programmes and institutes are important, taking into account past performance and future plans. As changes in the organizations and institutes may be affected by other uncontrollable (confounding) factors such as funding bodies' demands, a qualitative approach has to be used to gather information to enable detection of such factors.

The following variables have been used to qualify the above components in the evaluation of programmes/research units:

- Research culture of the institution (in relation to the vision and mission to undertake research)
- Current research processes within institutions
- Team work *versus* individual research
- Processes relating to research activities (for example communication, assistance of junior staff, exchange channels, methods used for quality control)
- Doctoral degrees produced and training for supervision activities
- Objectives and report results
- Academic reputation of staff (bibliometrics, peer review activities, rewards and other awards)
- Collaboration with outside scientific committees
- Dissemination of research

4.2.4.6 Procedures of assessment

Most of the activities previously conducted by sector organizations (referred to as research organizations in the new Protocol) are retained. A few exceptions are procedures of reporting to government and responsibilities of conducting the meta-evaluation, both of which would then be carried out by the Inspectorate. An observable phenomenon is the high level of

autonomy and independence enjoyed by the research organizations in handling issues relating to the institutes falling under their jurisdiction. There are occasional overlaps where researchers fall under two different research organizations (boards) and the system is set to take care of such problems.

4.2.4.7 The evaluation committee

The VSNU board compiles an expertise profile of possible/potential members of the evaluation committees and then makes appointments after consulting with the management of the institutes. In the selection of the committee members, the board must ensure that the members are competent to perform assessment and independent from institute operations. The board is also responsible for ensuring acceptance of the committees by the institutes. The board may seek external advice on the profile of the committees both nationally and internationally. The process of appointment is kept transparent and institutes are informed of all activities until committees are duly constituted. The chairperson is appointed first and then the rest of the other members. A committee undertakes all evaluation activities and is accountable to the board which may in the process of finalizing reports ask questions (to such committee) for clarification. Same as in the previous era, committees are responsible for programmes (disciplines) submitted by the research institutes and are guided in their functions by the Protocol.

4.2.4.8 The research organization board and levels of accountability

This body is responsible for monitoring the process of reporting (by the committees) until evaluation reports are accepted. The board may refer back some of the questions raised by the research committees in their reports to the institutes (for verification of information) and responses to such questions usually form part of the appendices to the final reports. The reports are then submitted to the advisory boards of the institutes to be used for future planning.

The research organization board is accountable to the Ministry of Education, Culture and Science. The board reports on mid-term and external evaluation results. As already stated, the board is also accountable to the community and publishes information on the internet. Accountability is realized in processes such as the initial schedule of planned evaluation by the board, the institutes' annual reports, the annual reports of the research organizations

themselves and public meta-evaluation performed by an independently established committee.

Initiatives to organize independent committees (for meta-analysis) are taken by the KNAW. The role of the committee(s) is(are) to monitor the evaluation processes in order to check for compliance with standards, to monitor the evaluation committees' levels of involvement, to review levels of transparency of the evaluation process and to assess the impact of reports, recommendations made to institutes and decision-making for improvement. The meta-analysis process ensures the efficiency and effectiveness of evaluation.

4.2.4.9 The evaluation process

The research organizational board holds meetings to strategize on plans of action and on the compilation of the self-evaluation reports. The institutes submit the self-evaluation documents to the organizational board as input documents, and the board may either accept or decline the submission. In the latter case, explanations will be given with recommendations for improvement. The received documents are then served on the evaluation committee together with other documents such as the Protocol, procedures and board questions.

The committee then reviews the documents before the site visits to institutes about four weeks or more in the case of more institutes at the site. During the visits, the committee may see the director or board of the institute, research leaders, advisory committees and any other person that volunteers information or wants to be heard. Visits are negotiated with the institutes and additional information may also be solicited. The director of the institute is also expected to make comments on the draft report which will be finalized and then submitted to the organization board.

The evaluation committee can discuss the future of the research and a management letter can be compiled for the board to explain the situation. The letter may contain confidential matters, for example, the policy on personnel and other sensitive decisions and is therefore not open to the public. The two-pronged comprehensive evaluation process (internal and external) takes care of both research and management evaluation.

Publishing the report is an attempt to expose institutes to levels of accountability. The report should reflect on leadership strategies and policies, quality of resources and facilities, academic reputation, societal relevance and on strengths and weaknesses analyzed and formulated by the institutes. The results of each research programme are quantified in relation to quality, productivity, relevance and qualitative explanation of the quantified data is also provided. The latter contains a qualitative review of policies, quality of products and staff reputation, relevance and future perspectives of the programmes.

4.3 Conclusion

It took The Netherlands two and a half decades to come to terms with national needs and to operate at the level of international recognition by others..

This sub-section presented a summary of the Dutch model of research evaluation. The summary has been grouped into its origins, objectives, processes, benefits and lessons learnt.

4.3.1 Origin (History and Rationale)

- The restructuring of the system of education in the 1980s resulted in government reducing its steering powers and granting autonomy to the HEIs – legislated for the entire system.
- Evaluation dates back to the 1980s and has been evolving ever since, altering its purpose as a result of the changing decision-making context in a developmental way.
- Universities set up systems of quality assurance to monitor internal systems and report to government – accountable to the Ministry of Education.
- Evaluation has always been conducted concurrent to that of teaching and learning - the two were separated for convenience. This study only pays attention to research evaluation of universities.
- In 1993 the HE and Research Act defined government’s regulatory ex-post intervention, involving inspection to remedy irregularities until 2001, when responsibility was handed over to the research organization (VSNU).

4.3.2 Objectives, benefits and challenges and the summary of the research evaluation processes in The Netherlands (over time) have been summed up in the tables below (tables 4.1, 4.2, 4.3) together with the bodies responsible for processes

Table 4.1 A summary of the objectives in the different evolutionary stages, with responsible bodies for the stages

Stage	Purpose	Responsible body
Initial stage 1980	Universities were expected to encourage research participation through allocating funds based on research requirements	-1980- Universities submit reports to government -Internal GUF funds allocation
Conditional Funding 1982-1992	Government autonomy matched with accountability to funds - for ‘good’ research based on the achievement of aims, also meant to promote quality, encourage systemic discussions and effective use of resources	-Universities accountable to government -Local peer reviewers anonymous selected by KNAW -GUF funding control
VSNU quality assessment of research: 1993-2001	Comprehensive research intended to maximize university research quality through self-regulation and external evaluation and to improve academic standards	-1993- HE and Research Act regulations, ex-post intervention -VSNU owned and funded collectively by universities -Internal evaluation by universities, faculties and departments -Effective policies developed - Evaluation based on the General Protocol guidelines and the discipline protocols -Inspectorate does meta-analysis, evaluate panels, advise

		Minister
Standard evaluation 2003-2009	<p>Improve research for regular accountability. To assist researchers and units to make better future decisions on research and management policies for quality improvement and viability, with reference to international standards as criterion for accountability.</p> <ul style="list-style-type: none"> -Self-evaluation of the university system -Institutional mission related to national relevance 	<ul style="list-style-type: none"> -Institutions still select programmes -High levels of autonomy given to institutions -VSNU independent of government powers, drafts specific protocols for internal assessment, compose committees, conducts verifications publish reports, does meta-evaluation, monitors and reports progress to minister

Table 4.2 A summary of benefits and challenges experienced over the period of evolution

STAGE	BENEFITS	CHALLENGES
1980	<ul style="list-style-type: none"> - Research participation encouraged to those who have interest - Minimal costs 	<ul style="list-style-type: none"> -No criteria except research needs, no follow-up on outcomes
1982-1992	<ul style="list-style-type: none"> -Minimal costs -External reviews introduced -Grouping of similar programmes -Internal management system as base for policy -Introduction of peer review -This contributed to nationalization of quality research 	<ul style="list-style-type: none"> -No criteria for allocation of funds for good research - Casual rating not yielding excellence or effectiveness

	<ul style="list-style-type: none"> -Network on post-graduate research assistance -Introduction and accreditation of research schools (interuniversity collective nature) -Casual judgment used to encourage researchers -National trends of systemic research set. 	
1993-2001	<ul style="list-style-type: none"> -Birth of VSNU -Strengthening of government trust -Internal regulation improves -Building of international competence -Process accepted -Results followed by advice on improvement, useful and helpful -Self-reflection process initiated - Issues of national importance were introduced (relevance) - Quality relevance and recognition (international) by others - Productivity and viability observed - National evaluation of similar (common) programmes - 6-7yrs cycle initiated - Experience in using panels even for internal validation for future 	<ul style="list-style-type: none"> -System very costly -Minority language -Inter subjective difference -Minimal inspectorate contributions -Methodological problems (peer review) -Visiting games played by researchers -Some international reviewers not competent -Delays in meta-analysis reports
2003-2009	<ul style="list-style-type: none"> -Six-year national process adhered to -Institutions more responsible for internal regulation 	

Table 4.3 A summary of the process of research evaluation in The Netherlands

STAGE	PROCESS
1980	Ex-ante process based on research requirements. Universities produce reports.
1982-1992	First ex-ante information and later throughput and output (publications).
1993-2001	<p>Internal self-regulation and external evaluation by peers.</p> <p>Components of evaluation:</p> <p>Quality (output on; dissertations, publications, patents, citations, international participation and conferences, editorial boards, academic awards, visiting professors, NOW funding award)</p> <p>Productivity (output against inputs)</p> <p>Relevance (academic, social and professional)</p> <p>Viability (future standing)</p> <p>Internal regulation: Faculty selects programmes, institution provides logistic support, checklist from VSNU guide submission, no minimal requirements, data neither purely qualitative nor quantitative</p> <p>External evaluation: Method (document review, observations on visits 2-3 days, interviews with programme leaders)</p> <p>Submissions: Mission of faculty, nature of science, overview of research, intended outcomes, anticipated developments, staff profile, five important faculty publications and other publications, international performance</p> <p>Data analysis: Rating 1-5, bibliometrics done by a separate body</p> <p>Utilization: Poor performance if not improved leads to sanctions, no reward or punishment, results not attached to funds allocations, inspectorate not prescriptive, reports and remarks go to the Minister</p> <p>Committees: (one foreign expert knowledge of Dutch language, one educational expert, two members of HE system, one</p>

	member of professional body, chair –Dutch origin selected by KNAW. Inter-disciplinary consultation
2003-2009	<ul style="list-style-type: none"> - Internal yearly review for monitoring data (submission of reports) - Three-year self-evaluation, reporting as mid-year evidence and preparation for external assessment - Interdisciplinary synergy encouraged (submitted reports either accepted or rejected, with recommendations) - Ex-post performance serves as ex-ante (level of confidence) for prospective planning - Reports on internal assessment go to the VSNU board and discussed with institutions on policy decisions - VSNU sends periodic reports to Minister as evidence of institutional accountability as unit of submission - SWOT analyses are also used for future planning <p>Criteria: (field of publication, programmes, institutional assessment reports)</p> <p>Quality defined in terms of international standards and recognition, relevance (scientific, technical, socio-economic), viability (infrastructure availability, sensitivity to discontinue non-viable programmes and new ventures, managerial professionalism and funding organization effects), vitality and feasibility on innovation potential,</p> <p>Data collection and Scoring: 5 point scale (interviews with research leaders- additional information from volunteers, observation of facilities and report analysis)</p> <p>Products: Bibliometrics, technometrics, socio-economic products (current research processes, team <i>versus</i> individual work, communication and support, Doctoral dissertations, objectives and research results, academic reputation, external collaboration, publications). Evaluation looks at research, management and processes also</p> <p>Committees: Transparency in appointing, on expertise profile approval after consultation with institutions</p> <p>Reports: Published and confidential information withheld (report on leadership, resources, policies, academic reputation, social relevance, strength and weaknesses, quality of products and future perspective)</p>

Chapter 5 The United Kingdom (UK) model of research evaluation

5.1 Introduction

For a better understanding of the evaluation/assessment of research in the UK, this subsection first provides a background of the HE education sector in England, Scotland and Wales. The section deals with the historical background and discusses the evolutionary processes of research evaluation/assessment. Also discussed are the reasons for the promotion of evaluation processes. Finally, the section focuses on the purpose of evaluation and strategies used by the different bodies over time in some detail.

Similar to The Netherlands, the UK is classified under Type A countries by Campbell (Shapira and Kuhlmann, 2003) because of the systemic and comprehensive nature of its research evaluation model. The model is also regarded by Geuna and Martin (2001:6) as “one of the most advanced research evaluation systems in Europe”, and although there are signs of reluctance by other European countries to emulate it (Campbell (Shapira and Kuhlmann, 2003), it is still highly regarded. Besides, the New Zealand system, which is discussed later, is said to be “modeled to a large degree on the British research assessment exercise” (The 2003 Assessment, preface). Examining the system will hopefully promote understanding and provide information for comparison of the two models.

5.2 Historical background

Broad-scale evaluation of research in HE in the UK dates back to the 1970s. Participation in the research projects and workshops of the Organization for Economic Co-operation and Development (OECD/CERI) programme on institutional management in HE performance evaluation, at the time is said to have attracted minimal attention from the HEIs (Sizer, 1988). Future plans would nevertheless be based on these activities. In 1979, the government introduced the following two ‘requirements and expectations’ that affected the future existence and the direction to be followed by HEIs in the UK (Sizer, 1988; Johnes, 1996):

- Government reduction of public spending, coupled with a move towards a market economy, which influenced institutions to adopt a market-related focus in an effort to seek alternative funding. Government primarily hoped to reduce reliance on government funding and on a secondary note intended to ‘increase institutional autonomy’.

- HEIs were expected to be accountable for the services they provided in relation to resources allocated to them. Sizer (1988) lists some of the variables that would demonstrate accountability. These include; efficiency in the use of resources, effectiveness in the achievement of objectives, strategies and action plans, and responsiveness and relevance to societal needs. In research in particular, selectivity in the allocation of resources would be an important consideration. Of importance also is the need to improve the quality and excellence of teaching and research.

It should be noted that at this initial stage of planning the UK system planned a parallel establishment of evaluation for teaching and research.

Campbell (Shapira and Kuhlmann, 2003) confirms that the UK is one of the countries in Europe that has been reducing general university funding (GUF)/block grants and attempting to reduce dependency on government funding. This prompted the introduction and increase in the diversity of income to universities. The reduction in government funding together with a process to seek alternative funding may have been intended to guide institutions into providing teaching and research services that are in line with and relevant to the demands of those who provide resources. Fulfillment of such demands by resource providers would in a way affect institutional mission and the direction of science. Therefore, institutions were required to revise their missions and visions, supposedly in line with the new developments. With such a level of compromise, it is obvious that ‘selective’ funding became an important determining factor in the research assessment exercise (RAE) exercise.

Expectations of the system are that “institutions of higher education would increasingly have to compete for resources – not only among themselves within the sectors, but also with other institutions within the higher education sector – and must respond to demands for increased efficiency and value for money” (Johnes, 1996:18). Allocations of the research GUF followed the selectivity pattern and after the 1996 RAE, 98% of government research grants were allocated according to the recommendations of the research evaluation exercise (Barker, 2007).

According to Sizer (1988), the culture would change from professional lifelong tenure based on collegial means to an entrepreneurial model. One of the important government

recommended implementation strategies was to calculate and publish performance. This was intended to increase efficiency and to assist government to determine the true level of demand (Johnes, 1994).

Prior to this new idea, the university sector in the UK was predominantly state-funded (Johnes, 1994) with an academic self-regulation process. The new system was regarded by Bauer and Kogan (1997) as being imposed on the universities. The requirements triggered actions from different funding agencies, especially those responsible for state funding. In the early 1980s, a programme named the SRHE / Leverhulme Programme conducted a study on the future of HE in the country. This, according to Sizer (1988), put performance firmly on the agenda. Bodies such as the Treasury and the Committee for Vice-Chancellors were compelled to plan activities in tandem with performance evaluation. This extended to different academic levels (faculties) in the institutions. This resembles a typical example of what Rip and van der Meulen (1996) describe as government's extreme steering powers. The move was towards a somewhat prescribed, statutory state system that allowed universities to only control academia.

In the early 1980s, the University Grants Committee (UGC) introduced the reduction of funds to universities (the grant letter of July 1981). This was emulated by the Department of Education and Science (DES), which reduced research funding to public sector institutions, including universities. This compelled institutions to shift reliance to non-governmental funding, which would increase their research funding substantially. According to Sizer (1988) most organizations that responded to institutional financial demands laid down their rules of accountability in exchange to the assistance they provided.

The 1985 Jarrat Committee Report on 'Efficiency in Universities' (Sizer, 1988), the 1985 DES Green Paper on 'The Development of Higher Education into the 1990s' and other documents all contributed towards the 'value for money' philosophy. Of interest to this study, is the introduction of the UGC's Research Selectivity assessment in 1986 (the first in the cycle of events) and the publication of its results by the Secretary of State, which according to Sizer (1988) culminated in the DES White Paper on 'Higher Education, Meeting the Challenges' in 1987. The White Paper marked the start of a new era that "pledged" and affirmed the commitment of government to a system of institutional autonomy linked to accountability. This would be maintained through the alignment of institutional missions

with the production of teaching and research outputs, against which performance would be measured.

The following section focuses on the systems that led to the processes of evaluation.

5.3 The evolutionary process

The system of HE funding evolved with changes in government systems. Johnes (1996) attributed the changes to government's intention to increase the efficiency of the HE sector. This section provides a brief account of historical changes in the system of evaluation and its funding bodies. Initial discussions focus on activities from 1980 onwards, with the University Grants Committee (UGC) as the main funding body for universities in the early eighties. This body was later replaced by the University Funding Council (UFC).

With the restructuring of HE, which culminated in the abolition of the binary system in 1992 (combining polytechnics, colleges and universities), another funding structure emerged to replace the bodies responsible for the three levels of the UK HE system. Three Higher Education Funding Councils (HEFCs) were separately introduced for England, Scotland and Wales. The HEFCs became important bodies in HE representing Treasury and the Department of Education. The latter replaced the Department of Education and Science in 1993.

Since the introduction of the research assessment exercise (RAE) in 1986 the UK HE system has been involved in a systemic and comprehensive research evaluation. A second and third exercise followed in 1989 and 1992 respectively. The 1992 cycle was followed by one in 1996. A new cycle in 2001 cycle was followed by a review period and then the cycle in 2008 (Barker, 2007).

Lucas (2004) observed that improvements and modifications occurred from cycle to cycle. The availability of information and the quality of evaluation improved as the evaluation activities evolved. This is an important lesson for South Africa.

Similar to the Dutch model, Campbell (Shapira and Kuhlmann, 2003:98) classifies the UK exercise as Type A model, in which a comprehensive "ex-post" research evaluation covers and addresses all disciplines at national level. Lucas (2004) and Barker (2007) explain the

continuous reshaping and modification of the exercise through continued debate and discussion among universities and policy makers during and between cycles.

Research funding mechanisms are said to have remained unchanged throughout all the RAEs (Elton, 2000). Ironically the funding for teaching and research were treated differently despite the initial plan for similar treatment and the notion that the two core areas should support each other. Elton (2000) and Lucas (2004) conclude that the preference given to research compromised the quality of teaching.

5.4 Purpose of Research Evaluation (of the research assessment exercise)

In line with the new policy, emphasis was placed on quality improvements in both teaching and research. Performance evaluation (based on quality performance) would be conducted on individuals, departments and at the level of institutions. Johnes (1996) directly links the expected accountability to the creation of a platform for institutions to compete for resources. On research performance issues, Johnes (1996:18) states that “since the first attempt to rate the research performance of universities in 1987, resources provided by government for research purposes had been allocated on the basis of those achieving the highest rating receiving the highest share of funds”. The main purpose of evaluation was how to best allocate resources, a move which Johnes (1994:205-206) regards as “the form of more refined targeting of resources; in the place of block grants which fail to discriminate according to performance”.

Johnes (1994:206) also views the funding process as “designed so that individual researchers, departments and institutions receive research funding according to the perceived quality” and this would be based on past and future work. In this way, selectivity in the allocation of resources was linked to ‘perceived’ quality, reduction of reliance on government and diversification of income, on which evaluation is based.

Critics of the exercise have a slightly different view, although all agree on the purpose of the research funds allocation. Corfield (2006:2) regards RAE as “an enormous piece of bureaucratic machinery whose function is to measure the quality and quantity output of British university departments in order to calculate the level of research funding due in coming years”. This system, Corfield believes, has a tendency to obstruct important and

original ideas. He cites the non-counting of text books “written by people who do not do, or do not aspire to do research” as one of the oversights of evaluation (Corfield, 2006:3).

Lucas (2004) and Barker (2007:3) view the exercise as promoting elitism among researchers and within the HE system. Barker blames the exercise for “concentrating resources for research in a relatively small number of universities”. This is also confirmed by Geuna and Martin (2001). Barker (2007) adds that any institution that includes weaker staff in its submissions risks dropping grades, with negative funding consequences. This implies that to obtain grades, an institution would have to employ research active staff or forfeit subsidies. Therefore, the exercise is “a system, which demands success in research for institutional prestige and to potentially secure increasingly more scarce resources” (Lucas, 2004:1)

5.5 Research Funding and Assessment

Some of the initial criteria that informed the activities of selective funding are the perceived strengths and weaknesses in institutional research. It was in 1984 that the evaluation of input dominated the idea of performance measuring.

Johnes (1996) explains the two main systems of research funding in the UK. The first is the system of government subsidies for HEIs managed by bodies assigned such responsibility by government. Allocations in this system are directly linked to results (the ex-post) of the research performance assessment system (selectivity exercise). The second source is made up of research councils. Funding in this category is ex-ante and is done on a “project specific” basis. Peer review informs the process of allocation. Recently, in both cases, past performance influences allocations more, with an unavoidable confounding effect attached to the methods used. For the purposes of this study, evaluation of research funded through assigned (subsidiary) bodies is discussed in detail (as they evolved) in the following sections:

5.6 The research evaluation exercises

The system of evaluation/assessment continued to change to suit the changes in the HE system in the UK (Lucas, 2004). The initial exercise was conducted in 1986-87, covering a five-year period of data between 1980 and 1984. This assessment exercise was conducted by the-then funding body, the UGC. The second exercise was conducted by the UFC in 1989, covering the period 1984 to 1988. The HEFCs performed the third exercise in 1992. This exercise, which is regarded by Johnes (1996) as more extensive, included all universities and

assessed performance for the period 1988 to 1992. Other exercises that followed evaluated the entire system of HE as a unitary one. The activities of the exercises are discussed below.

5.6.1 The UGC research assessment exercise (1986-87)

This exercise was based mainly on the following methods (Johnes and Taylor, 1990):

- Informed peer reviews. Subject panels consulted well known academics in each subject.
- Inputs, calculated through research grants awarded over the period reviewed. Information for input was provided mainly by research councils.
- Brief account of research performance of a university (cost centre/research units).
- Five recent publications regarded by the department or cost centre as representing research work undertaken in the department/section. Very little emphasis was placed on the research output of an institution.
- Future research plans.

Results were used to rate the universities as one (1) to three (3) with 1 being the lowest and 3 the highest score. The exercise is said to have resulted in severe criticism especially for its use of inputs (ex-ante information) as a measure, which has a tendency to bias results in favour of larger departments. Other criticisms included the lack of clarity on the criteria used to evaluate research quality, the importance assigned to the anonymity of the selected assessors, that the methods used were not suitable for other subjects and the fact that different evaluation/assessment standards for different subjects were used. This last point made the rating very controversial. Universities, cost centres/research units and departments all served as units of analysis.

5.6.2 The 1989 Research Selectivity Exercise (UFC)

The second exercise was conducted by the UFC and attempted to respond to criticisms of the previous exercise by introducing transparency in deliberations and by improving on the type of data collected. A more output-orientated approach was adopted for data collection strategies and data. The methods and types of data are described below:

- Advisory groups / panels were formed for different subjects. These included foreigners. This may be interpreted as a build-up towards internationalization.

- Similar to the UGC exercise, informed peer reviews were used as a method for data collection and handling.
- The following information was provided to the subject panels by institutions and/or cost centres:
 - Research output over the period 1984-88
 - A brief report of departmental achievement in research
 - A brief description of future research plans
 - Summaries of publications and research reports
 - Up to two publications listed for each full-time academic staff member.
- An element of “attainable level of excellence” was introduced per subject to create some level of comparison across disciplines. This uses international standards as a benchmark and output was measured against this benchmark across all subjects. It is therefore the second level of international/global comparison (reference to others) experienced in the exercise. The scale used to measure the attainment of excellence refers to the quality of research and used both international and national ‘excellence’ as standards. A rating scale of 1-5 was introduced.

Research output is privileged as a measure of quality in this exercise. According to Johnes (1996), this may be calculated by taking a weighted average of the research rating awarded to university cost centres/departments/research units. The process of assessment was not without its critics. The following are cited as notable anomalies (Johnes and Taylor, 1990):

- An inadequate questionnaire used to collect data
- The process was conducted over a short period of time, that is, all subject areas evaluated over a period of three months across the UK, which is not sufficient.
- Over and under-estimation of output as a result of problems in the definition of research units. For example, the consideration of outputs of all full-time staff in the unit during assessment.
- Definition of what is regarded as output. For example, the system could not clearly differentiate between edited books and authored books, regarded published research reports as books and could not distinguish between articles published in refereed and non-refereed journals

- The use of output measures encouraged overproduction that leads to researchers compromising quality for quantity. This tended to undermine the system, and plans were made to address this issue in the next assessment exercise.
- The problem of inter-subject differences was not solved by the rating scale and this created dissatisfaction with the allocation of resources. There were also differences in the precision of assessment by the different panels.

The exercise conducted assessment on the basis that all universities are equal and operate at the same level. It failed to recognize the differences in institutional size and type or even the diversity within the HE system in the UK (Elton, 2000), all of which may have contributed to differences in performance. In order to award funds to a university, an arithmetic total was used of all disciplines added together (Campbell in Shapira and Kuhlmann, 2003) without consideration of these differences. This puts the larger cost centres on an advantage, a caution to those who would consider using the methodology/paradigm to evaluate research. The system continues to benefit institutions previously advantaged through the availability of resources to perform research, which gives an institution its character (Johnes, 1996).

Availability of resources alone will not necessarily lead to high productivity, especially because being competitive is by choice. Yet evidence reveals that inter-institutional variations per expenditure per staff member can be accounted for by variations in subject mix (Johnes and Taylor, 1990). The geographic location of an institution, the number and characteristics of research staff (culture of research) also contribute to the character of the university while the student/staff ratio usually reveals an inverse relationship with research performance (Johnes, 1996). Competitive universities are said to consider such factors when building and retaining research reputations. It is therefore tempting to conclude that the research culture of an institution is determined and shaped by the decision context which may be related to the mission and vision of the institution together with efforts made to fulfill these.

Where there is competition for research, there is a tendency for institutions to retain their position of advantage. This perpetuates the differences between institutions, Johnes and Taylor (1990) speak of this as the Oxbridge effect (after Oxford and Cambridge). One may also conclude that the ability of an institution to produce research output is in itself an input factor, which may create a rating problem in which incompatible institutions are compared

(comparing those with high with those with low inputs). This results from the tendency of input/output variables to form a cycle of events, especially in institutions where a research culture is a priority. Figure 5.1 portrays a situation (author's summary) in which a reach culture may sustain high output.

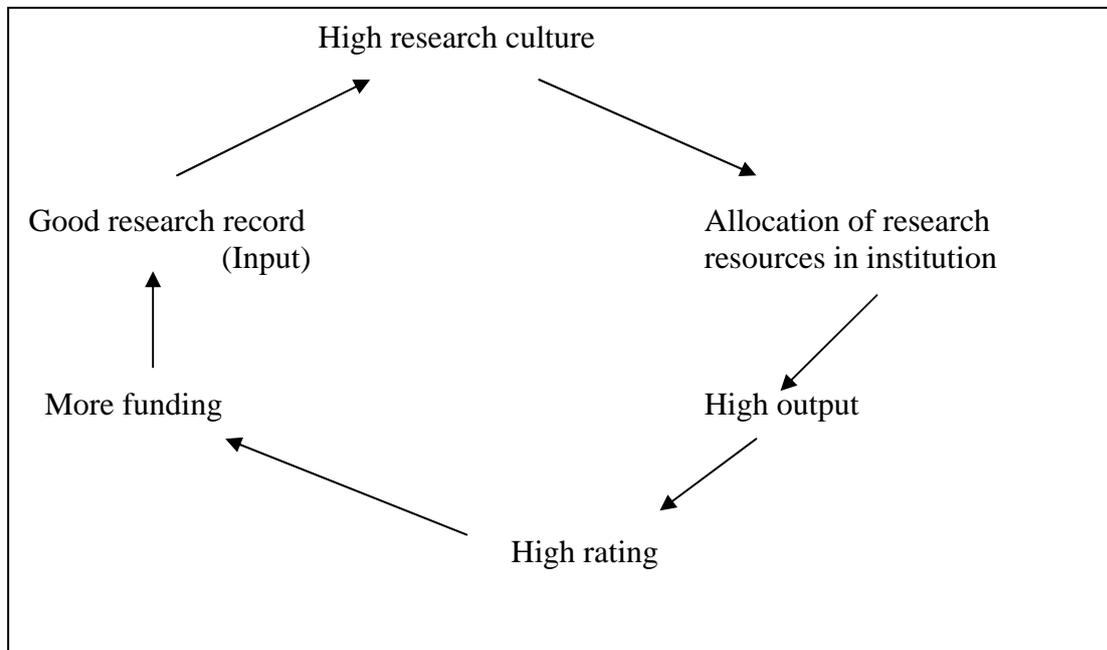


Figure 5.1. A sketch of the input / output influence – a cycle of events as understood by the researcher

Factors of interest that are said not to necessarily affect research performance directly but are necessary to note for evaluation purposes are: the average age of staff, the percentage of professors, the percentage of research students, and expenditure on some resources as a form of income and the conditions and status of a library. As previously stated, inter-university differences may be highly influenced by differences in input availability and these factors may negatively affect research rating.

5.6.3 The 1992 Research Assessment Exercise (RAE) – HEFC era

5.6.3.1 Background

The third research assessment exercise in the UK was performed by the HEFCs. Discussions at this level focused on the RAE in the British HEI and refer mostly to the Higher Education Funding Council for England (HEFCE). Records indicate an increase in the number and

variety of institutions submitting to the exercise. At all times, the main intention of the exercise remained “to inform selective funding of research” (HEFCE, 1997), which implies that government funding is made available for assessed research. Similar to the previous exercises, the idea of improvement of research was not obvious even though it was continuously referred to. Improvements in research productivity were stressed and there was heavy reliance on subjective peer review (with its shortcomings). Although other authors have been referred to, information on the UK from this era has been obtained from the HEFCE, RAE documents and the Higher Education and Research Opportunities (HERO) documents.

The exercise took place at a time when HE in the country was also affected by the development of national transformation policies. For example, the 1992 Further and Higher Education Act, which abolished the binary divide between the polytechnics, colleges and Universities (former UFC institutions), put all institutions at the same level of competition for funds. Therefore, some of the observed changes in HEI may not necessarily be attributable to the RAE. As a result of the new laws and systems, the exercise had to be brought forward by a year to accommodate changes brought about by transformation. This is believed to have affected preparation time for universities.

Several benefits accrued from the exercise. Of importance is the existence and development of an improved system of research management both at the institutions and within the funding organizations (Elton, 2000; Roberts, 2002; Barker 2007). This is said to have resulted from the exercise together with policies and the involvement of other funding agencies, which enabled institutions to pay attention to their research missions. The HEFCE (1997) claims that the idea of quality improvement gave birth to the establishment of graduate or even research schools. The system is also said to have encouraged interdisciplinary research which was originally not accommodated in the RAE in its panel formation (Elton, 2000; Barker, 2007)). Although the HEFCE (1997) claims improvements on the link between research and teaching, critics find that the RAE process caused research and teaching to drift apart (Elton, 2000; Lucas, 2004; Wikipedia, 2007) . Bauer and Kogan (1997) feel that the system was highly compromised by political nominations of non-academic personnel in the management and administration of the funding councils. These structures, it is alleged, concentrate on the market system and hardly pay attention to professionalism and the academia.

5.6.3.2 Context of the 1992 RAE

The fact that the exercise was conducted at a time when the binary divide in HE was abolished implied that all HEIs were placed on the same level to compete for funds – at least in principle. This meant bringing together institutions from completely different research cultures and backgrounds to compete: the university with a culture of basic research, funded mainly by the state, and the Polytechnics with an orientation towards applied, consultancy and contract research, mainly funded by the users of the research results. Because the exercise was fund-driven, it would be envisaged that the latter institutions would survive without being fully involved and that competition would not serve as a motivating factor unless the RAE policies restricted all funding avenues (including funds for consultancy and contract research). Otherwise the other secondary intentions of the exercise such as the improvement of quality may not be achieved across the institutions. The same level of competition would put the former Polytechnics at a disadvantage as far as ‘research development’ (of RAE expectations) was concerned.

The new HEFCE funding plan regarded the financial consequences of the exercise as modest especially because of limited public spending. The whole system was dominated by the phrase “funding for quality” and a smaller number of research-strong institutions continued to thrive on the funds than before (Johnes and Taylor, 1990). There are indications that a simultaneous increase in student numbers contributed to an increase in funding for teaching (HEFCE, 1997) and that this, together with attracted external funds, had an impact on the finance status and thus planning in HEIs.

5.6.3.3 Activities of the 1992 RAE

- Data Collection

It became apparent that ‘research audit’ types of evaluation and the consequential distribution of funds were preferred and perceived to be easier than the evaluation of teaching (Howie and Scott, 1993). This is because performance indicators were regarded as being the most objective indicators for the evaluation of research quality at the time (Tognolini *et al*, 1994; van den Berghe *et al*, 1998).

Therefore the exercise drew more attention from the politicians in the UK. The process of data collection focused on performance per subject and was based on the number of “research

active staff” per faculty/unit as an indicator of good performance, based on individual members’ performance over a period of four years. Research units/faculties served as units of analysis and were expected to select the members of staff whose work could be submitted (as indicating performance levels) as sub-units. Johnes (1994) explains that those members that were excluded from the submission list would receive no funding.

Credits given to submitted work of members benefited the institutions where they were employed at the time of the exercise. The credits were used to determine research funding for the institutions for the following four years. This means that the results for the 1992 exercise would inform research funding from 1993 to the next exercise around 1996-97. According to Johnes (1994), this funding is meant to sustain basic research and other research work declared important but may not easily access research council grants.

Data was collected in the following areas:

- * Publications
Each member of the active research staff was expected to submit two publications and two other forms of research output that would have been published
- * A publication count was submitted for each unit / faculty together with other quantitative materials.

Publications had to be those disseminated within a stipulated period (3.5 to 4.5) depending on the subject. Quality was judged against national standards using a five-point scale and international participation (participation in other countries) became an important factor.

Further submissions for the evaluation of quality include the following (Johnes, 1994; HEFCE, 1997; Oppenheim, 1997):

- Description of research plans (brief research statement concerning research activities)
- Number and value of research grants obtained within the research period
- Supporting quantitative materials, for example conference papers
- Post-graduate research activities (the number of research doctoral students supervised) as an indicator of research performance).

Peer reviews were used in the process of evaluation. Peer reviewers were appointed by the funding bodies for every subject area after consultation with professional bodies. Peer representation varied from subject to subject. Howie and Scott (1993) believe that measures were “tightly defined” to allow for quality judgment based on available information. Researchers were also observed attempting to find better ways of producing reliable results. Oppenheim (1997) suggested the use of citation counts as an alternative measure, which according to his investigation was a better determinant of quality than a simple publication count. Oppenheim argues that in the discipline of humanities in particular, citation counts correlates well with the rates awarded through peer reviews and that citation analysis is less time consuming and less costly. Oppenheim (1997) suggests that if citation analysis is not used as a replacement, the strategy can be used as a backup to peer review.

5.6.3.4 Structures involved in the exercise

Assessment panels were, according to Howie and Scott (1993), constituted for the subject structures with a total number of seventy two (72) structures. The subject structures served as units of investigation. This represented half of the number used in 1989 (HEFCE, 1997). The reduction was meant to increase efficiency by creating a platform for comparison and effectiveness, allowing the same panels to handle larger numbers of such structures. The research organizations, which also serve as funding bodies (HEFCs) determined the framework for the exercise, defining the subject structures and constituting panels, administrative and managerial structures. While the framework operated across subjects, subject-specific procedures were also allowed.

5.6.3.5 Publication of results

Results were published as soon as they were available. Some of the panels provided informal feedback to the research units while the responsible funding bodies were responsible for results in general (HEFCE, 1997). Rating concentrated on the weighted proportion of staff per unit graded as research active.

5.6.3.6 Funding implications of the exercise

The funding organizations calculated research funds according to performance (RAE rating) of different subject areas and made these funds available through block grants (together with funds for teaching). According to Johnes (1994), 94% of the HEFCE allocations were

directly influenced by the results. As previously stated, results for the 1992 exercise were utilized for allocations for the period 1993 until to the next exercise. Units rated 1 (see the five point scale at the end of the discussion on the UK) received no funding while those above this were funded accordingly (Johnes, 1994; HEFCE, 1997; Oppenheim, 1997). The process of allocation was criticized as complicated and non-transparent. Johnes (1994) reveals that HEFCE allocated an additional 3% of funds to reward institutions with a “good track record of securing research contracts”. While this may have been done to encourage institutions to seek external funding, it also has serious repercussions for institutions that, for some reason or other, cannot attract such funding. On a positive note, this may be seen as a way of rewarding and recognizing other forms of research that are not published, for example, applied research reports. The funding council also made an allowance for the development of research potential in the new universities. Not much was said on this matter and therefore this study could not gather relevant information.

5.6.3.7 Effects of the 1992 RAE

The emphasis on “research active” staff had its own shortcomings, some of which are:

- The system encouraged volumes of submissions (unit count of research active staff); this could compromise quality
- Fewer submissions of high quality could result in lower volumes, translating into less funding
- There were other factors relating to the sub-unit of analysis (indicator), “research active staff”, especially because of staff mobility that may take place during the stated period of funding.

The unit count of publications is a quantitative measure which is perceived as an important performance indicator for research activities in most countries and in the UK in particular. It is therefore difficult to disregard its use until a better and more reliable measure is established.

Other challenges experienced in the exercise are (Howie and Scott, 1993; Johnes, 1994; HEFCE, 1997; Oppenheim, 1997):

- How to treat basic and applied research equally. Submissions for the two areas were made separately. Treatment of applied research in the UK has been evoked by the need to connect research to utility in order to make it relevant. The RAE

exercise has been criticized for its inability to assess applied, professional and user-related research.

- Problems of the vast variations between types of institutions in terms of context (this includes important variables such as subject mix, finances, teaching / research balances and others) that relate to missions and visions.
- Publications of different quality levels and subjects counted on the same basis.
- The single-disciplinary approach disadvantaged inter-disciplinary output and this also affected funding.
- Questions arose relating to validity and reliability of assessment by the panels.
- Very little was known about the process of funding and assessment, together with the criteria used. The general report compiled and provided by the funding bodies did not provide reasons for rating and the report was seen as a non-transparent policy instrument.
- Funding did not immediately follow the rating results.
- As the exercise was the first after the restructuring of HE (abolition of the binary divide), some staff members, especially from the former Polytechnics were not familiar with the expectations.
- Not enough feedback was given for future planning and development. Only some of the subject structures gave informal but informative feedback.
- The process is time- consuming and uneconomical.
- The results created a tendency for institutions to recruit and retain research orientated staff for productivity. This may negatively affect the attention given to the training of lesser skilled staff.
- It was feared that more concentration on research would have a detrimental effect on other core functions, especially teaching
- The exercise is accused of having decreased the collegial status of research as it concentrated on the market related approach to funding
- The exercise concentrated more on the dissemination of research results than research activities. Some of the research activities are conducted for other developmental reasons and those would therefore not receive any subsidy.
- The unit costs of research output were estimated, resulting in generalizations that could deny potentially productive faculties an opportunity to be supported.

- The system allocated to institutions and left it to the institutions to deal with internal allocations. While this contradicts the intentions of evaluation, the process observes the principle of autonomy.
- Grant availability is an input factor (*ex-ante*) and publications an output (*ex-post*) one. Both these variables cannot guarantee the sustainability of activities of research and therefore cannot be regarded as good indicators/measures. It is not easy for past research to determine future allocations, as it may be difficult to determine the impact of input on a specific output.

Despite all the problems encountered, the exercise received the most support to date. There was potential to encourage performance in the not-so-well-performing institutions. The exercise would be least expected to offer encouragement, if any, to the well-established research institutions except to maintain the *status quo*. The contributions towards the shaping of research staff formations and a shift towards an increase in staff participation in research were later reported (HEFCE, 1997). Summative evaluation usually shows no sign of intrinsic stimulation and quality enhancement on individuals. Besides, not all social science research ends up being published. Burnhill and Tubby-Hille (1994) confirm the absence of encouragement of low grade researchers. Furthermore, Carpenter *et al* and Colman *et al* (Oppenheim, 1997) found no correlation between the RAE and other research evaluation measures.

The 1992 RAE can be seen as a turning point for the evaluation of research in the UK. There was general acceptance of the exercise by institutions, as a regular activity and institutions would immediately start preparing for the next cycle after submissions for the current one. More attention went in to strategic plans and internal reviews, including revisiting missions and consideration of relevance (to user) that would henceforth influence and shape the culture of research. The drive to obtain government funds also stimulated institutions to ‘solicit’ funds from other sources. This level of conformity, it is alleged, has a tendency to reduce the diversity of operations. The exercise is also said to have affected policies of other research funding bodies.

5.6.4 The 1996 Research Assessment Exercise

5.6.4.1 Background

This was the fourth in the series of research evaluation exercises in the UK, and the second after the abolition of the binary system in HE. It was therefore to be expected that the exercise would be more established and display improved practices and methods as a result of previous experiences. Changes from the previous exercises reveal progressive and developmental processes in the evaluation of research. In this exercise, the funding bodies are said to still dominate the provision of resources (income) in HE. The management of the project became the responsibility of HEFCE and this body was guided by the other funding bodies: SHEFC (Scottish Higher Education Funding Council) and HEFCW (Higher Education Funding Council of Wales).

The 1996 exercise was said to be a significant improvement on that of 1992. This is evidenced by the early preparations, improved plans and levels of transparency attained. Early efforts were made to familiarize institutions with the procedures of submission and evaluation and guiding documents were published as early as July 1995. The process was piloted and two of the 1992 panels were used as specimens to develop “key considerations” fundamental to all the panels. Panels published their detailed guidelines on criteria and working methods in their “statements of criteria” for assessment at the end of 1995. This was regarded as a breakthrough on transparency.

The RAE created a common understanding by introducing a standard definition of research that was meant to serve as a framework for all panels. Research was defined as:

“original investigation undertaken in order to gain knowledge and understanding. It includes work of direct relevance to the needs of commerce and industry, as well as the public and voluntary sector; scholarship; the invention and generation of ideas, images, performances and artifacts including designs, where these lead to a new or substantially improved materials, devices, products and processes, including design and construction” (RAE96 1/96, 1996:7).

Panels were expected to operate under this standard definition. It is reported (RAE96 1/97) that the definition raised problems of interpretation and lacked the guidelines necessary for awarding proportions on, for example, how to generalize and rate towards a standard. The 1996 RAE moved away from focusing on the volume of work done and concentrated on quality, based on ‘publicly available research outcomes’. The definition of national and

international standards seemed to have been well-perceived by panels, who were expected to strike a balance between being ‘reasonable and consistent and being unduly constrained in their judgments’. That is, operating within the same framework with an allowance for differentiation between subjects.

Some panels understood the meaning of standards at national and international levels to be more inclined to visibility than quality, especially because the citing of research output played a significant role. The ‘publicly available research outcomes’ of the UK system may also be interpreted as referring more to ‘visibility’ than to be defining the component of quality. The RAE defines quality of outputs through the medium of publication, linked to the acceptance of rigor in quality control of review of the publications, and thus dependent on subjective acceptance of the activities of journals. The definition is not very different from its use in the Dutch evaluation system. In some instances ratings were made from evidence of quality of output for individual researchers while other panels generalized on the work of researchers collectively, confusing the aspect of the sub-units of analysis.

5.6.4.2 Purpose of the exercise

The purpose of the exercise seemed to be similar to previous exercises (Elton, 2000). The RAE96 1/97 (1997:1) describes the purpose as the “rating of research quality for use by the funding bodies in allocating money for research in the HEIs which they fund”. The funding bodies execute this function on behalf of the Department of Education, and judgment of research quality remains in the hands of the experts (peer reviewers). The selective allocation of funds is based on the strength an institution can demonstrate in research. The presence of research active staff in an institution plays the most important role in the production of outputs, and the results of the exercise are purely summative.

It is important to note that one of the funding bodies’ responsibilities is to advise government on the funding needs of HEIs and allow government to make final decisions. In this exercise funds were given to deserving institutions in the form of block grants together with some guidelines for consideration. As a form of filtering down government’s steering powers, discretion was then left to institutions, which to some extent were guided by expected accountability for the use of such funds (Geuna and Martin, 2001). Funding provided by the funding bodies was meant for, among others the following activities:

- Salaries of permanent academic staff

- Computing costs
- Support for basic research
- Research training

5.6.4.3 Procedures

Most of the activities were based on the 1992 procedures, but in an improved form. In designing evaluation criteria for the panels, consultations were held with the academic community and research end-users. This allowed the panels to set additional criteria inherent to their subjects. The 1992 units of research were reduced from 72 to 69, with a merger of similar ones, reducing the panels from 62 to 60.

The main intention in the planning for the 1996 exercise was to provide the panels with raw data to work on, especially because clear-cut dates were set for activities. Data was checked for errors and verified with the HEIs. Most of the panel members were involved in the 1992 exercise and were already better experienced.

Submissions were based on researchers employed by an institution at the time of the census and research output was within the stated period. Similar to practices in 1992, institutions decided on who and what to submit. No more than four publicly available items per active researcher were expected and institutions could submit as many researchers as they had available in their different departments. In this exercise, basic and applied research output could be submitted together. Institutions could indicate in their submissions which panel their work was to be referred to. The HEFCE provided software for submission and all institutions were encouraged to utilize the device, and it facilitated the speed of data processing.

Submissions were called for in November 1995 and the closing date was set for the following year, to allow institutions to prepare. An increase in submissions was experienced with an additional 11% submitted for the 1996 exercise as compared to 1992. Criteria/cases for submission were the following:

- Names of active researchers selected for assessment and present at census period
- Four publications/publicly available items by each of the submitted researchers from 1990 to 1996 (citations)

- The number of research students and studentships in a department and income related thereto
- Statements of departments, research achievements and support, with external recognition.

Justifications for few or no publications for the submitted staff members were considered by the panels. In most cases the information would not be very helpful. Meanwhile, qualitative data was used to justify the quality of outputs and multivariate analyses assisted in building evidence towards a standard, as submitted work was used to support one another. According to Campbell (Shapira and Kuhlmann, 2003), ratings were given on a refined scale, from 1 to 5* with point 3 being further divided into two, making a total of 7 levels (see rating scale at the end of this section). The rates 1 and 2 did not attract any funding and in some cases different rates would be given to work of the same volume.

5.6.4.4 Panels and support services

Chairpersons were selected by the Chief Executives of the funding bodies. Most had served in the same positions in the 1992 exercise, or were recommended by the outgoing chairpersons of the panels. Chairpersons then consulted with subjects associations and other professional bodies for the selection of other members, considering eminence of an individual, coverage of the subject and geographical balances as important factors (RAE96 1/97). A total number of 560 members constituted the 60 panels, each with less than 10 members.

Most of the panels were responsible for single units of research while others were assigned more responsibilities. Panels were allowed to refer work to other panels for better judgment and even though this worked for some it was problematic in other instances where criteria for assessment were different. The panels met three to four times for periods of up to two days.

Panel Chairpersons attended several briefing sessions and continued to receive advice from the panel secretaries posted by the funding bodies. The secretaries also served as committee secretaries, advised on procedure and guidelines and provided administrative support. Some of the panels were assisted by assessors nominated by research councils and charities in order to guide assessment on the income criterion. Sub-panels were only used for consultation and were made up of stakeholders such as the research end-users. This assisted in relating the

stated quality to relevance. Panels were given full autonomy of operations and were encouraged to work independently to avoid the influence of others. This was perceived by some as an impairment of judgment.

In contrast to the previous exercises, institutions were assisted with their submissions. A seminar was held to explain requirements, the use of software and to define concepts. The HEFCE also put facilities in place for consultations and inquiries were made *via* telephone and email.

5.6.4.5 Publication of results

The HEFCE published the ratings and released the statistical data. Other panels offered developmental feedback in confidence to the heads of institutions on a separate submission. Some queries were received by panels from institutions and some panels agreed to give brief explanations. Some of the Chairpersons also discussed the ratings with departments in confidence. The Chairpersons reserved the right not to discuss the results.

5.6.4.6 Alterations to the 1992 RAE

As stated, the 1996 RAE reduced the number of units of research, thus reducing the panels. Each panel was given full responsibility for their units and was allowed to form sub-panels only for advisory purposes. The exercise dropped the idea of volumes of publications and decided on what the exercise perceived as quality, by assessing a maximum of four publications per submitted active researcher. Whereas the 1992 RAE rated results on a 5 point-scale, the 1996 exercise increased the scale to 7 points by adding 5* and dividing 3 into 3a and 3b. The latter exercise also emphasized the use of a common approach by the unit of research and criteria would differ only where it was necessary. In some cases the definition of units of research would change. In others, the same type of information could be submitted to different panels in the two exercises.

All the efforts to improve the RAE are said to have had impact. The collection of data through electronic means and the time allowed for preparations and for assessment are all said to have been very helpful (RAE96 1/97). A few practices were nevertheless identified that warranted improvement. For example the software needed improvements and time for preparations could also be improved.

Other problems that were identified were (RAE96 1/97):

- Some of the important events, for example, criteria for setting meetings and seminars, occurred at the end of the year
- Rating may not have been compatible across the unit of research as a result of differences in criteria
- There were variations in the preparation and management of submissions and not enough time was allotted for amendments. Some work could not be returned as one submission to one panel
- Corrupt data could not be detected prior to closing dates for submissions to be corrected on time
- Software was not adequately piloted
- Weaknesses in the definition of the word quality, that is, standard definition caused rating problems
- Referrals not easy to solve because of differences in criteria between panels
- In some cases, especially in art, it was not easy to detect when publication took place and there was no prior setting of criteria to take into account similar work
- No threshold levels for excellence categories, research students and research income and the problems of definition of research students
- Not enough information obtainable from the restructuring of research councils
- There was a problem of classifying other work such as conference papers which are circulating among delegates
- Some academics were claimed by more than one institution (poaching effects).
- Very few HEIs had received allocations from the research councils and some did not submit enough information on research plans. This created rating problems
- The RAE indirectly discouraged multi-disciplinary submissions
- Problems with benchmarking for international excellence
- Information about former staff not considered
- Failure to understand guidelines for submission
- Some submissions did not indicate sole-authorship, joint venture or contributions
- No enough space for additional information on circumstances of publication

These problems and others would serve as a base for improvement in the next exercise.

5.6.5 The 2001 RAE

5.6.5.1 Background

The exercise was the fifth in the series and was mainly conducted according to the same format as the 1996 RAE. Most of the rules were unchanged but were enhanced to incorporate recommendations. Four funding councils were involved: HEFCE, HEFCW, SHEFC and DENI / DEL (Department of Employment and Learning North Ireland). The latter body has been added to incorporate Ireland. While all of the funding bodies were fully involved in the process, the management and data collection processes were left in the hands of HEFCE (RAE/98/67, 1998).

After the 1996 exercise, the RAE team launched a broad consultation with HEIs, professional bodies and other structures interested in research, in preparation for the next exercise. Comments ranged from the request for more transparency, consideration for formative assessment, and early setting of criteria to the consideration of a long-term review of the 2001 exercise. Most of the comments are more about improvement than a complete change of purpose or design, placing more emphasis on the improvement of value and efficiency. An average of a five (5) year period between exercises was preferred in this instance, for the reason that the period would restrict the exercise to recent outputs and would allow enough time not to interfere with research activities. Previously no consistency existed regarding the periods between exercises. It was also important to review the definition of rating points for the purposes of better clarity and consistency of application across panels.

5.6.5.2 Purpose of the exercise

The purpose remained the same as that of the previous cycles: “to produce rating for research quality to inform allocation of funds for research” (RAE 1/98:2). Assessment was once again based on peer review and panels used professional judgment to formulate a view about quality of work submitted. The exercise still concentrated more on research active staff as a sub-unit of analysis/subject of analysis. Some important principles were attached to the activities of the exercise. Most of them had been applied previously, but they were improved in the 2001 exercise. These principles are: clarity, consistency, continuity, credibility, efficiency, neutrality, parity and transparency. All these were emphasized for the sake of consistency in evaluation and were to apply throughout the exercise (RAE 01/2004).

The same definition of quality used in the 1996 exercise was adopted. Some of the panels nevertheless added relevance as an important criterion. The notion of RAE judgment of

quality based on past performance being able to “take account of institutions’ investment in developing their staff and research culture” (RAE 1/98:17) was regarded as being questionable. Established staff members get recruited by other institutions and this encourages “poaching”, even though the RAE team claims that the developed culture creates a platform for new employment. This is an important point to note if benchmarking on this model is considered for the South African situation. One may positively conclude that the movement of researchers will ‘diffuse’ improvements across the country until the time when equilibrium has been achieved across all HEIs. How new staff happen to achieve the same standards is not indicated.

5.6.5.3 Methods of assessment

As already indicated, peer reviews are used within a common framework, with common sets of data and common data definitions. To achieve these, assessment regulations carried explicit key principles. In this exercise discretionary assessment of quality was reduced through an increase in consistency. Panels were nonetheless allowed to seek more information on the HEIs with the submissions. Panel-specific criteria were allowed and these had to be made explicit prior to submission. While there were recommendations for a move towards self-assessment, this area still needed further investigation. The seven-point scale used in the 1996 exercise was retained and description of the grades amended to make them easy to interpret. The promotion of consistency was also displayed through the clearer descriptors assigned to the points. Point 4 was upgraded with an element of “international excellence”. In the previous exercise this element was optional.

All the variables for assessment were retained: the listing of research active staff by category with payroll entry, research output for such staff (four items), research students and degrees awarded (showing their graduate records), research funding with grant information and a narrative description of research activities. All these still make the departments/research units the units of analysis. Whereas some researchers felt that numbers were important as a sign of good output, and that measures should take this into consideration, the RAE insisted on the quality notion and retained the maximum of four units (of research output) to avoid “premature publication”. In all fields except for the humanities, five years of work were considered that is from 1 January 1996 to 31 December 2000. Seven years’ work was assessed for humanities. There was an overlap of periods wherein some of the work submitted for the previous exercise was repeated in 2001.

5.6.5.4 Procedures of assessment

Improvements to the exercise resulted in a number of changes from the 1996 exercise. Planning for the 2001 evaluation began immediately after the 1996 exercise. Of importance is the early call for nomination of members of the assessment panels. Some of the sub-areas, for example Biochemistry, were combined with others and this reduced the number of units of research. The numbering (of the units of research) was nevertheless not altered since this would result in confusion.

Improvement of the framework was informed by the 1996 RAE Manager's report, the report on the 1996 data collection process and the 1997/98 consultation results. One of the most obvious changes is the inclusion of sub-panels for the evaluation of some of the areas, especially for Medicine and Education. Once more, applied work received equitable treatment, and the guidelines for submissions were greatly improved, with the addition of briefing notes. All panels developed draft statements of criteria and consulted before releasing the final document, which was submitted 16 months prior to the data submission date.

To enhance support and maintain consistency, umbrella groups were formed, especially for panels that shared common criteria. RAE managers also held briefing sessions with the panels. Institutional support was enhanced through the creation of a website where all circulars could be accessed, the introduction of an effective helpdesk and allowance for a contact person per institution. A series of seminars were also held for discussion of the framework and the process of data collection and a newsletter was published to keep all stakeholders informed. Submission date was set to be 30 April 2001. The differences between these dates and those of the 1996 RAE are summed up in Table 5.1 below (RAE 01/2004):

Table 5.I. Differences in participation between the two exercises (1996 and 2001)

Items	1996	2001
Submissions in total	2894	2598
Institutions participation	192	173

Work of category A	48072	48022
--------------------	-------	-------

The number of submissions has declined. Of importance is the decline in the number of institutions that participated in the exercise. Data was crosschecked and incorrect submissions were removed. Assessment took place from May to October 2001.

Several rigorous efforts were added to the assessment process. For example, some panels made efforts to read through all cited materials while others used samples for the exercise. A Central Resource Output Sourcing team was also established to assist with the acquisition of information and additional materials could also be requested from the institutions. Panel ratings were published and placed on the internet, showing all details except the personal particulars of researchers and strategic plans. Panels also wrote summary reports for institutions (given to heads of institutions) to explain reasons for the grades and a general overview report was compiled per discipline.

5.6.5.5 Assessment teams

In this exercise, all four funding bodies were fully involved in management. RAE management and the chairing of the data collection steering committee were alternated among members of these bodies on an annual basis. Other members of the four bodies served in different committees. The HEFCE was left with the responsibility of managing/overseeing data collection processes.

In this exercise, panel chairpersons were elected by members of the 1992 panels among themselves. The RAE called for nomination of panel members from the professional associations and learned societies. These formed a pool from which further nominations could be made. The chairpersons then listed the panel members following the guidelines used in 1992 and the funding structures considered. An additional call was made for research end-user nominees (Barker, 2007), since the research end-users formed 13% of the total panel members, especially in Engineering, Science and Medicine. The names of panel members were published in December 1999.

In the main, there was one panel per unit of research but conditions relating to multidisciplinary research and other factors required that 14 units of research be shared by

panels. Twenty-five sub-panels were also established out of 17 panels to assist with areas needing expertise, where disciplines overlapped, to assess emerging areas and to bring in a different perspective. A minimum of two members was allowed in a sub-panel and in most cases one of them would chair.

The panels were each assigned a secretary and a deputy, whose duties were the same as those performed in 1996. Most of the secretaries were provided by funding bodies from amongst their staff members and additional members were provided by the HEIs. Panels were also allowed to have advisors and a total of 350 experts were involved in the exercise. The 2001 exercise introduced the element of non-UK members to advise panels (Barker, 2007) on the allocation of points from 4 to 5*. Each panel would consult with five such members and a total of 300 were used for the benchmarking exercise for international excellence.

Improvements from the 1996 exercise caused an increase in financial costs by 68% from the previous exercise. The increase was mainly attributable to the increase in the size of panels, the introduction of sub-panels, the use of special advisors, the increased number of sessions, cross referrals and improved resources especially the central output system. All of these were introduced to improve the quality of the exercise.

5.6.5.6 Problems encountered

Fewer problems are cited for the 2001 exercise than was the case with the previous ones. There is evidence that the results of the exercise and the implementation of the funding process encouraged institutions to recruit research-credible (research active) staff. This process is referred to in the manager's report (May, 2002) as poaching, and as previously stated, led to movement of staff. The RAE put plans in place to compensate those institutions that lost staff through this process and hoped that they would be replaced with others that would be trained to the same 'heights'. While this seems to be unfair to the losing institutions, it has the elements of a long-term national research capacity building plan, as long as institutions do not get discouraged and are given resources to continue.

There were a few misinterpretations of the RAE guide which led to incorrect items being submitted. The submissions also failed to reveal the number of active staff per unit and thus ratings could not be guided by numbers. The ratings were also affected by instances where active staff members were listed by more than one institution. Data verification was

conducted to avoid problems of cheating. The search for output was limited to authored and edited books, journal articles and conference proceedings. This is possibly the reason for the drop in submissions. Other challenges experienced in the cycle are methodologically related. For example the inherent problems of peer review as stated in Chapter 2.

5.6.6 The post 2001 era

The next national exercise was to be completed in 2008 but has since been postponed a number of times. Although there were indications of the aging processes of RAE (Roberts, 2002), the executive of universities principals would not want to see the abolition of RAE without a viable alternative (Scott, 2007). The exercise is based on the same framework as the previous ones with some significant changes aimed at minimizing costs and the administrative burdens placed on the HEIs but maintaining high levels of effectiveness. Changes are informed by the 2001 exercise review together with stakeholder consultations. Such changes include the introduction of a six-year cycle exercise post 2008, the reduction in the number of panels and the retention of sub-panels, improvements in the weighting for grading joint submissions and the introduction of a “continuous graded quality profile for each submission, at the sub-panel level” (RAE 01/2004:2), which replaces the seven-point rating scale.

Although the purpose of the RAE remained the same across the series of evaluation exercises, the 2008 turnaround strategy is said to aim at provoking academics and managers “to think beyond lists of publications and force consideration of the conditions for undertaking research” (Barker, 2007:8). Assessment is based more on three indicators: research output, research environment and esteem indicators (Barker, 2007). The targeted research outputs include journal articles, conference papers and intellectual property devices. According to Barker (2007), environmental indicators involve: research students, research income and organizational structures, while esteem indicators include national and international recognition-prizes, fellowships and membership of advisory bodies.

The five point scale remains but has a higher level of quality classification moving from the previous 50% national, 50% international to world-leading in originality significance and rigor, raising the threshold of funding for the grades (Barker, 2007). The exercise is also more user-focused and although the overall ranking of the universities has not changed much

(Barker, 2007) universities continue to participate even though funding resources have been reduced. Barker (2007) also explains that the successive RAEs have improved the international quality of the UK universities, and agrees with Elton (2000) that the exercise has led to better planning and management of research within institutions.

In spite of all the benefits accrued from RAE, some authors identify unexpected consequences from the exercise in general. In order to allow researchers enough time to fulfill RAE requirements, temporary and junior staff members were deployed to relieve them of teaching workloads. This was at the expense of the notion that research and teaching should complement each other.

Lucas (2004) mentions that the RAE puts pressure on the UK universities to increase their funds and the hierarchical system imposed on them. In his analysis of the situation, Barker (2007:6) concludes that “universities and academics have been drawn into a strategic game, and invest tremendous efforts in trying to win”. The authors maintain that RAE demands should allow for diversity (a broader landscape) among universities, avoid excessive differentiation and the hierarchical system and change the funding policy.

Barker (2007) and Wikipedia encyclopedia (2007) advise that the post-2008 exercise will be highly metrics-based and will allow for disciplinary variations. The hardcore sciences (for example medicine) will need to be assessed on quality indicators while the social sciences and logical subjects (for example Mathematics) will stay with performance indicators. The following subsection briefly discusses experiences in the use of performance indicators in the UK. Other changes reported to have taken place in the 2008 exercise were the introduction of a “quality profile rather than a fixed seven-point scale” (www.rae.ac.uk/aboutus/changes.asp).

5.7 The introduction of performance indicators.

As government and funding structures embraced the new idea of evaluation of the HEIs in the 1980s in the UK, there emerged a need for the construction of indicators to measure performance. The introduction of performance indicators and the use of management statistics came as a result of recommendations by the Jarrat commission report (Sizer, 1988). The main purpose of the use of indicators in the UK was to evaluate performance of the

system of universities as a whole and individual institutions as part of planning for the allocation of resources.

Performance assessment of institutions, departments and individuals (academics) became an important dimension of management. This implies that the evaluation results would assist in making decisions and judgments “between competing and mutually exclusive alternatives” (Sizer, 1988:154). According to the author, implementation of the report recommendations met with problems, where accountability was reviewed narrowly. Johnes (1996) on the other hand feels that the importance of performance indicators was meaningful. The author nevertheless warns of the difficulty of constructing performance indicators that are able to reliably, quantitatively measure performance in HEIs, the latter being viewed as “a multi-product firm with multiple objectives”. He warns that “a reliable indicator for the research performance of universities should take into account the factors which influence the ability of each university to produce research output” (Johnes and Taylor, 1990:167). Rating may be done by comparing the expected (derived from estimated ability) with the actual, considering inputs and other characteristics of an institution.

The attempt to overcome the lack of agreement resulting from differences between institutions by introducing a multitude of indicators created problems of interpreting the broad spectrum of indicators. Johnes (1996:19) advises that “the use of performance indicators in aiding the decision-making process is limited if each indicator offers a conflicting view of an IHE’s efficiency or performance”. Sizer (1988) confirms that the use of performance indicators (PIs) to measure experience and effectiveness will be an oversimplification.

5.8 Conclusions and summary of the UK model

5.8.1 Introduction

The six RAEs discussed above are exercises that were undertaken over almost three decades. It is not surprising that the system is regarded as “tired”. Having survived its lengthy lifespan, it is either the case that its intentions should have been achieved by now or that a trend is set that would be able to sustain itself for an extended period of time. If neither of the two occurred, then serious alterations are called for.

The entire system has been summarized below, into the sub-headings of origin and rationale, purpose/objectives, processes of evaluation and benefits and challenges experienced in the exercises.

5.8.2 Origin and rationale (dates and events are indicated in point form)

The table below summarizes the historical activities in relation to the initiation of the UK research assessment system.

Table 5.2 Dates and historical activities related to system initiation

Date	Activity
1970s	Evaluation initiatives by OECD/CERI programme – management of performance evaluation in HE, on voluntary basis
Before 1979	All public HEIs predominantly state funded with an academic self-regulatory system
1979	Reduction of public spending and move towards market economy > expectations of institutional accountability to service provision (efficiency and effectiveness) and the improvement of quality of teaching and research This led to the introduction of competition for the GUF funds
1980	SRHE/Leverhulme programme put performance on the agenda Government steering power compelled universities and Committee of Principals to base activities on performance evaluation
1980	UGC Introduced reduction of funds to universities DES followed and reduced research funding to public sector institutions Universities resorted to non-governmental funding
1985	Jarrat Committee – on efficiency in universities led to DES Green Paper – Development of HE into the 1990s This introduced the ‘value for money’ philosophy
1985	Introduction of the UGC’ research selectivity exercise (first cycle)
1987	DES White Paper – HE, meeting the challenges

NB. Information on the series of exercises (including dates) is indicated in the subsections below.

5.8.3 Main funding bodies in the UK HE research system

- Government subsidy through block grants (GUF) > ex-post results
- Research councils > ex-ante

5.8.4 Objectives, processes, benefits/good practice and challenges/lessons learnt

The tables that follow provide a summary of the variables according to funding organizations and periods of occurrences. Table 5.3 indicates the dates of the cycles, the period for which data was required, the purpose of evaluation and activities according to the cycle periods. Table 5.4, on the other hand displays benefits/good practice and challenges together with lessons learnt throughout the series of research exercises.

Table 5.3 Objectives and processes in the RAE exercises

Funding Org.	Purpose	Process
<p>UGC – 1986 Data: 1980-1984</p>	<p>Resources provided by government for research purposes to be allocated on the basis of those achieving the highest rating receiving the highest share</p>	<p>-Informed peer reviews -Data: Inputs on research grants Brief account of university research performance Five publications seen as representing unit research work Future research plans -Rates 1-3</p>
<p>UFC – 1989 Data: 1984-1988</p>	<p>To best allocate resources – more refined targeting of research resources</p>	<p>-Peer reviews -Response to criticism of the 1st cycle -Transparency in deliberations -Output oriented data -Use of advisory groups -Data: Output, reports on departmental achievements, summary of publications and research reports, future research plans, up to two publications per fulltime research staff. -National and international standards introduced (level of excellence) -Rates: 1-5 -Weighted average awarded to cost centre make up allocations</p>

<p>HEFCs – 1992 Data: 1988-1992</p>	<p>Inform selective funding (allocate resources on the basis of rating received) with the following secondary intentions</p> <ul style="list-style-type: none"> -Improve quality of teaching and research -Purposes of accountability to state funds -Introduction of competition for resources -Funding given according to perceived quality 	<ul style="list-style-type: none"> -Subjective peer review based on available information -Audits, performance indicators -Research units as units of analysis -Number of research active staff per unit -Data: publications (two per research staff), other quantitative materials, description of research plan, research grants, post-graduate research activities, citation -Quality of national and international standards -Report published by HEFCE -Some of the panels provided confidential feedback
<p>HEFCs – 1996 Data: 1992-1996</p>	<p>Rating of research quality for use by funding bodies in allocating funds to HEIs. Results were also used to advise government to make final decisions</p>	<ul style="list-style-type: none"> -Mostly based on 1992 methods but improved -Standard definition of research (gaining of knowledge and understanding invention and generation of new ideas relevant to commerce, industry, the public and leading to substantial improvements) -Alteration of units of research (fewer) -Concentration on quality and not volume (public available outcomes) -Operation within same framework with allowance for differences between subjects

		<ul style="list-style-type: none"> -Data was submitted on software, making it easier to submit and easier to check -Block-grant funding (salaries of permanent academic staff, computing costs, support for basic research, research training) -Chairpersons appointed by CEO -Panels selected by academics, professional bodies, research councils and charity organizations -Panels allowed to refer work to other panels -Period of panel meetings four X two days meeting -Secretaries provided by research organizations -Inquiry made telephonically or by email -Scale 1-5* (seven points) -Data: names of active researchers, four publications, research students, research statement,
<p>HEFCs – 2001 Data: 1996-2000 (Humanities data for seven years)</p>	<p>To produce rating of research quality to inform allocation of funds</p>	<ul style="list-style-type: none"> -Ireland added (DENI/DEL) -Five- year period between cycles recommended -Peer review -Data (output): edited books, journal articles, conference proceedings, 4 items per research active staff, graduate students, research activity statement, funding (income) -Definition of quality more consistent -More information sort from institutions

		<ul style="list-style-type: none"> -7 point scale (1,2,3a,3b,4,5,5*), '4' upgraded to international -Fourteen UoA, combination of some units, inclusion of sub-units -Umbrella groups for panels to share criteria -Panel criteria drawn after consultation and released 16 months prior to exercise -Advisors allowed -Seminars held to discuss framework and data collection -Ratings published and reports submitted to institutions -Calls for nomination of panel members to form pool
HEFCs – Post 2001	<p>Produce rates for the allocation of funds</p> <p>(need to provoke thinking beyond publications)</p>	<ul style="list-style-type: none"> -Broader consultations conducted -Change to six-year cycle -Further reduction in panels -Grading of joint submissions -Indicators: research output (journals, conference papers, intellectual properties), research environment (students, income, structures), esteem (international recognition, fellowships and participation in advisory bodies) -Quality improved from 50% national/50% international to world class
HEFCs 2008	<p>Create quality profiles for each submission in order to assist in the determination of grants for 2009 to 2010</p>	<p>The use of a two-tier panel structure for standardization and clarity on criteria to accommodate the assessment of applied, practice-based and interdisciplinary research</p>

Table 5.4 Benefits/best practices and challenges/lessons learnt in the RAE exercises

Funding Org.	Benefits and good practice	Challenges/lessons learnt
UGC – 1986	<ul style="list-style-type: none"> -Systematic/comprehensive research evaluation had been introduced -Results used to rate universities 	<ul style="list-style-type: none"> -The use of input measures -Bias in favour of large department (data) -Lack of clarity on quality criteria -Anonymity of assessors, a problem to scholars -Methods not suitable to all subjects -Different evaluation standards
UFC – 1989	<ul style="list-style-type: none"> -Output used as a measure of quality (reliance on journals) -Introduction of level of excellence -Introduction of national and international standards 	<ul style="list-style-type: none"> -Questionnaire for data collection not adequate -Short period for evaluation -Problems with definition of research units -Problems with the definition of outputs -Outputs introduced cheating -Problems of inter-subjective differences -No consideration of institutional size and type, benefiting the previously advantaged – Oxbridge effect (Culture of an institution depending on institutional decision context, with production of output serving as an input factor)
HEFCs – 1992	<ul style="list-style-type: none"> -Turning point of evaluation in the UK -Reduction of reliance on government -Diversification of income 	<ul style="list-style-type: none"> -Promotion of elitism -High volumes of submissions from units -Fewer units submitted than expected

	<ul style="list-style-type: none"> -Securing increasingly more scarce resources -Acceptance of the exercise -Early preparations for the next exercise -Strategic planning for future exercises, revisiting of missions and visions -The drive to receive funds improved institutional performance (inspiration for better performance) especially in the not-so-well-performing institutions -Increase in research staff -Improved system of research management -Establishment of research graduate schools -Encouragement of interdisciplinary research -Institutions of different cultures compete -Improvement in the quality of research -Increase in the funding avenue 	<ul style="list-style-type: none"> -Creation of staff mobility (poaching) -Less influence on established research institutions and researchers -No encouragement for low grade researchers -Conformity to expectations reduced diversity in research -Low or no correlation between RAE and other research measures -Process of funding not transparent -A small number of institutions continue to thrive -Separation of teaching and research
<p>HEFCs – 1996 (Managed by HEFCE)</p>	<ul style="list-style-type: none"> -Submissions had increased by 11% -Exercise more established and improved as the 4th in the series -Early preparations of procedures and guidelines (published by 1995) -Process piloted 	<ul style="list-style-type: none"> -Some panels viewed the set standards to mean visibility and not quality -Quality depended on subject acceptance of activities of a journal -Software used not well prepared (corrupt data not detectable)

	<ul style="list-style-type: none"> -Panel criteria improved through consultation with academics -Basic and applied research submitted together -Institutions still made a choice on what to submit -Developmental feedback separately submitted to institutions and handled with confidentiality -Seminars held to explain requirements (e.g. software) 	<ul style="list-style-type: none"> -Rating not compatible across units -Varied submissions -Standard definition of quality across units affect rating -Referrals (inter-panel) affected different criteria between panels -Problems with classification of some variables (e.g. conference papers) by different panels -Pooling affected data -Sole authorship not considered in other panels -Not enough (input) information from research councils
HEFCs – 2001	<ul style="list-style-type: none"> -Exercise preceded by a broad consultation -Planning started as early as 1996 -Definition of quality more consistent -More information sought from institutions - Humanities data of seven years -Advisors allowed (including internationals) -Umbrella groups for panels to share criteria -Panel criteria drawn after consultation and released 16 months prior to exercise -Seminars held to discuss framework and data collection -Ratings published and reports submitted to institutions 	<ul style="list-style-type: none"> -Number of submissions declined -Increased finances to implement improvements (exercise more expensive) -Poaching leading to movement of staff -Incorrect submissions -Some staff listed by more than one institution

	<ul style="list-style-type: none"> -Calls for nomination of panel members to form pool -Poached staff replaced with others that would also be trained (departmental arrangement) -Data verified 	
HEFCs – Post 2001	<ul style="list-style-type: none"> -Further consultations conducted -Quality of research is said to have improved -Improvement in research management 	-Funding further reduced

General challenges: HEFCs

- Teaching suffered at the expense of research
- Extreme pressure for institutions to deliver
- Resultant hierarchy of universities
- Reduction of diversity as a result of conformity

General Benefits: HEFCs

- Introduced system of research management
- Establishment of research graduate school
- Encouragement of interdisciplinary research
- ‘Improvement’ in quality of research
- Increase in funding avenues.

5.8.3 A summary of differences and similarities between the Dutch and the UK models

The following table sums up differences and similarities between the two models in relation to Campbell's (Shapira and Kuhlmann, 2003) analysis:

Table 5.5 Differences and similarities between the Dutch and the UK models

Item	Dutch	UK
1. Responsible bodies	VSNU (Association of major universities)	HEFCs (Intermediary institution responsible for funding)
2. Linkages	Self-evaluation of universities Autonomy within the system of generous public funds	Close to public decision making concerning public funds
3. Duration/length of cycle	Six to seven years	Lately six years
4. Number of disciplines	34 Less and reducing all the time Inter-disciplinary research encouraged	1989 – 92 1996 – 69 2001 – 68 (too many, have institutional assignment problems)
5. Units of assessment	Research programmes (smaller than a department)	University department. Problems of inter-disciplinary research Sub-units introduced to solve problem
6. Characteristics of	-Review committee discipline based	-9-18 experts from a broad spectrum of institutions

discipline panels	<ul style="list-style-type: none"> -Panel chair is Dutch and nominated by institutional academic and selected on consensus -Other panel members are local and international -Evaluation carried out in English except for domestic oriented disciplines -Procedure standardized across disciplines and published in advance -Disciplines may adapt but not change methodology, therefore results may be compared across disciplines 	<ul style="list-style-type: none"> -Sub-panels used where necessary -Chairpersons nominated from panel members of previous exercise and appointed by CEO of funding organization -Panel members representing geographical areas and different characteristics of universities -Procedure standardized and published prior to evaluation
7. Content of submission	<ul style="list-style-type: none"> -General information: inputs, PhD thesis -Five best academic publications per research programme over a predetermined period -Faculty of departments (or institutions) forward information from internal assessment to panels 	<ul style="list-style-type: none"> -General information: staff information, research output, textual description and related data -Up to four items of research output per researcher over a predetermined period (output accessible to public) -Units submit to panels
8. Components	<ul style="list-style-type: none"> -Academic quality -Academic productivity (efficiency) -Relevance -Academic viability <p>The multi-dimensional definition highlights complexity</p>	<ul style="list-style-type: none"> -One quality dimension as an overall quality for research (reliance on decisions of journals and international standards) <p>Funding dominates university research</p>
9. Rating scales	1-5 the highest representing excellence	Revised at 1-4 for 2008

Chapter 6 The New Zealand research assessment exercise

6.1 Introduction

Information on this model was obtained from the 2003 and the 2006 periodic research exercise reports and only discusses operations of the one exercise. There is more discussion on the 2003 exercise. The 2003 periodic research performance exercise is the first comprehensive evaluation of ‘research quality’ in New Zealand. The exercise is said to have been costly, demanding and time consuming but systematic and authoritative. Activities of the 2006 exercise on the other hand were based on the experiences of the first with some added improvements.

The first section provides the background, based on the history and rationale for the model, followed by the purpose for the evaluation. The evaluation processes are provided under the section on methodology and these are followed by the challenges encountered in the model. The chapter ends with a conclusion and a summary.

6.2 Background on the exercise

The New Zealand tertiary education system has since the 1980s undergone several changes in relation to the changing economy of the country and to some extent changes in the ruling party (McLaughlin, 2003). The former resulted from the dependency of tertiary education on national funds. Changes in the economy were affected by the funding of the tertiary education system. In the 1980s, there was a demand for the tertiary education system to ensure “a knowledge economy” and “the need in the changing economy to provide tertiary education to a high proportion of the population” (McLaughlin, 2003:13). The increase in tertiary participation compelled government subsidy policies to shift from the “elite system” of subsidizing smaller numbers with higher amounts to a “mass system” of providing for more students with lower rates. This led to the reduction in the differentiation of institutions. Large numbers of very diverse students would be expected to have quality implications and as a result, the need arose for government to develop strategies on dealing with at-risk groups.

According to McLaughlin (2003), these demands have been imposed regardless of political party, philosophy or even policy approach. Whereas the idea of a broader access was initiated by the Labour-led government in the 1980s, the National-led government carried out its implementation, with a few additions such as the involvement of the private sector in the funding process. The National government also moved towards a competitive, market-based approach. When the pressure of student numbers and funding increased, government looked for alternative ways to solve the problems. One thing led to another, as government was caught up in the absence of accountability and capacity for governance issues. These challenges led to several tertiary education policy reviews in relation to the funding of the system.

A 1997 discussion paper and the White Paper of 1998 were formulated to guide policies to be more responsive to students needs. This would mean, among other things, that there would be improvement in accountability, improvement in governance and capacity building measures (for the competitive model), the introduction of quality assurance and audit mechanisms (for quality teaching and research) and alterations in research funding. Although government made repeated requests for inputs from the universities, the reviews were mainly government-driven, leaving the state with high steering powers.

In the late 1990s, New Zealand was characterized by changes in government, in which the National-led government gave way to the Labour-Alliance government (Tanczos, 2002, PBRF, 2003, Wrath, 2005). Prior to this change of government, the HE system had already experienced problems (stated above), which according to Tanczos (2002) would hopefully be solved by the new system.

Among other things, the system was blamed for policies that were incompatible with access to tertiary education (Tanczos, 2002). In research in particular the government was criticized for what was referred to as brain drain, which was interpreted as neglect “to nurture ‘a research culture with the right mix of funding, incentives, and devotion to the spirit of intellectual inquiry’.” (Wrath, 2005:1). Pressure was mounting at a time when government was losing power to the left-wing, the Labour Party.

According to Wrath (2005:1) the movement of tertiary education research staff to universities abroad was the result of the National-led government staff-exchange policy which was later interpreted as “the plight of ‘academic refugees’ driven abroad by economic need”. Unfortunately the policy opened access to information about the advantages of being a researcher abroad and created a path for academics to search for greener pastures. The solution to the brain drain problem would therefore have to reward those that participated in research in order to retain them.

Government was compelled to review its policies and in 1999 when the Labour Party took over, the Party used this opportunity and worked on a vision to strengthen the system of tertiary education research (PBRF, 2003). It is obvious that the changes in party leadership had some effect on the system. The competitive spirit of the parties nevertheless seemed positive in that the intention was attached to votes and thus meant to make an impression. The new government established the Tertiary Education Advisory Committee (TEAC) to help formulate long-term strategic direction for tertiary education (McLaughlin, 2003). The Committee’s are said to be more inclined to tertiary education systems than issues of access and student support. A manifesto was published that related HE research to national economic and social development. Another initiative of the new government was the Tertiary Education Strategy 2002/07. The following were advocated:

- The establishment of seven centres of research excellence
- The introduction of a programme to build research capacity in the Social Sciences
- The creation of a Performance-based Research Funding (PBRF) exercise

The PBRF is the centre of discussions in this section because it discusses issues of research evaluation. In the plan to implement the PBRF, government would rely on the results of investigations and advice by different bodies, for example; a series of audits commissioned by the Ministry of Education, the 2002 results of the PBRF working group, the Transition Tertiary Education Advisory Commission in 2001 and 2002 and a wide consultation with the Tertiary Education sector.

The Tertiary Education Commission (TEC) was assigned responsibility to manage the PBRF exercise. This was done through a constituted project team, responsible for policy, technical and administrative support. Panel chairpersons (sometimes referred to as ‘Chairs’) and panel members were appointed to take responsibility for evaluation. The reviewers received “some briefing” prior to the commencement of the exercise.

6.3 Purpose for the evaluation exercise

The multiple-purpose nature of the model is observable. This may be the result of the many reviews that were undertaken on the tertiary education system. For example, the report on the 2003 PBRF assessment exercise (PBRF, 2003) states that the aim of the exercise is to increase the average quality of academic research and to encourage research to support teaching. This is despite the fact that (as highlighted by Wrath in 2005) government’s intention was to provide incentives for researchers to stay in New Zealand. Research in the context stated by the PBRF (2003) is understood to contribute to the intellectual infrastructure of subjects and disciplines, the main purpose of which is believed to be advancing knowledge and the creation of a better living environment.

The increase in the quality of research and support of teaching activities would be achieved though making funds available to young research students and staff. According to the 2003 Assessment document (PBRF, 2003), this would improve research strength in the tertiary education sector. The report stresses that “if TEOs are not generating high quality research, this will have a detrimental impact on New Zealand’s overall research and innovation system” (PBRF, 2003:9), adding that research standards affect the quality of education. This is also emphasized in the 2006 report (PBRF, 2006). The term tertiary education organizations (TEOs), refers to both the tertiary institutions and research centres in New Zealand, while tertiary education involve universities, colleges polytechnics and ‘wananga’ (a Maori tertiary education institution in New Zealand). This is perhaps the reason why the system is generally referred to as tertiary education and not higher education (in order to accommodate the different types).

The introduction of the PBRF was therefore intended to assist in the “fair distribution of funds” (as a reward) for the production and attainment of the highest possible quality of research and teaching. Legislation in New Zealand requires that degrees be taught by well established scholars engaged in research, and there is a need for evidence of this. There is also a need for accountability for research funds. After some consultation with tertiary education institutions, there was evidence of a need for research to be funded on the basis of excellence, and also to encourage participation through reward. This UK-type purpose would be realized through the Centres of Excellence and then the PBRF. By implication, the results of the PBRF would reduce funding for poorly performing institutions.

Previous research funding was made available through public tuition subsidies (equivalent to a block grant) based on FTE weighted student numbers and funds from research foundations subsidized in terms of “top ups” for degree-level programmes. The two subsidies are vulnerable to fluctuations in student enrolments. The process of “top ups” would be phased out gradually from 2004 to be replaced by the PBRF funding by 2007.

Along with this thinking, the PBRF is said to intend to contribute to the general improvement of research performance in line with the Tertiary Education Strategies (2004-2007) and the Tertiary Education Priorities (2003-2004) by investing in excellence through a comprehensive process which respects academic traditions and refers to freedom and autonomy. The PBRF also intended to maintain consistency across subjects in line with international standards while maintaining continuity, differentiation, credibility, efficiency, transparency, cultural ‘inclusivity’ and be complementary to the new and existing policies in the country.

Conformity to international standards is an objective for global competitiveness and recognition. The 2003 Assessment report (PBRF, 2003:9) confirms this intention by stating that “a relentless focus on verifiable quality is essential if the tertiary education sector is to achieve and sustain internationally competitive levels of research excellence”. Verifiable quality implies recognition by those that are regarded as holding international standards (recognition). This is confirmed by the high involvement of scholars from other countries,

mostly engaged in the planning and assessment activities. Overseas scholars seem to qualify for the status by virtue of exposure to similar exercises.

Other secondary benefits intended through the PBRF are: the enhancement of accountability as a tool to motivate TEOs to review research and strategies, and encouragement of collaborative research and to motivate researchers to concentrate more on and improve research in areas of excellence. It was also hoped that the exercise would provide long-term benefits and that results of the 2003 exercise would serve as baseline data for future comparisons to track change. The exercise served more as some needs analysis study, testing the strength of academic research. Evidence was needed on whether or not New Zealand has academics of international standard. In other words, is the country a full participant in the global research community, respected inside and outside? A further purpose was establishing whether or not students are mentored and supervised by reputable researchers who undertake research of very high standards.

The results of the 2003 exercise reveal areas that require improvement of research, showing the need to nurture research and encourage them in order to attain world-class levels. The results provide a guide for the direction of resources that would improve knowledge. What is also evident is the testing of efforts for the country to undertake a periodic research assessment exercise upon which future improvements will be based.

The government of New Zealand hoped that with the introduction of the PBRF the performance-related distribution of research funds would provide incentives to invest resources strategically “in areas of strength, not weaknesses” (PBRF, 2003:viii) and to encourage specialization across the different disciplines of research through rewarding excellence. This implies investing in in-depth rather than broad areas of research to achieve international excellence in areas of specialization and interest in building profiles. Through this process, the government hopes to establish and maintain a steady balance between specialization and quality improvement.

The report on the 2003 exercise (PBRF, 2003: xii) shows satisfaction with the existence of world-class performers and research strength especially in the universities. Older universities, especially in the older subjects show good performance. One of the eight universities is said to “rank substantially higher when judged against international standards of research excellence”. If this achievement is important, then it confirms suspicions that international competitiveness and internationalization feature as some of the most important objectives for this research evaluation exercise. It will therefore not be surprising if the focus shifts towards pushing the eight universities to achieve international standards, as opposed to the elevation of disciplines.

The 2006 report showed improvement in the number of submissions and other performance variables as compared to the “base-line” results of 2003 (PBRF, 2006). Other revelations were confirmation of the 2003 results that New Zealand inherits an uneven landscape of research within the university sector leading to wide differentiation gaps between top performers and the low performers. Differences in performance have been observed across disciplines and between established and emerging researchers, which is normal and expected.

6.4 Methodology and strategies used in the assessment

6.4.1 Background

The exercise deploys a retrospective evaluation of research performance based on outputs produced over a set period of time with individual researchers as the main units of analysis. A mixed performance-assessment mode has been used, combining peer reviews with performance indicators. The methodology used is said to be inherited from other countries already involved and experienced in periodic large-scale evaluation of research. Hong Kong is simulated on the use of Evidence Portfolio (treating staff as units of analysis) while the UK is simulated in most of the procedures. Consultations have been made with the UK on design and implementation processes. Ironically, there are marked differences in purpose between the two countries.

The exercise is said to be a ‘periodic’ process and set for two phases; the internal assessment, wherein each participating TEO evaluates Evidence Portfolios presented by eligible staff

members and the TEC peer assessment exercise. The latter involves the use of expert panels to evaluate the contents of the Evidence Portfolios submitted by the TEOs. It is not clear who in the TEOs assesses the Evidence Portfolio and how this is done. The fact that a considerable number of the portfolios that were successful at the TEO level were not at panel levels raises doubts that similar assessment conditions apply across the two levels. As mention is made of the general guidelines, the discrepancy in judgment may have either resulted from the fact that the guidelines were not understood the same way, that the expertise levels of evaluation differ or that the guidelines cannot be relied upon (not the same at the two levels). While the panels are said to have attended a briefing session prior to commencement of the assessment exercise, nothing is said about the assessors at the TEO level. The 2003 report blames the following conditions for the differences in ratings between the TEOs and the TEC panels: subjectivity (especially at TEO level), time constraints, information added during the TEOs evaluation, which was not added to the Evidence Portfolios and lack of systemic examination of nominated research outcomes. All these show weaknesses in planning and preparation for the exercise.

The PBRF eligible staff's Evidence Portfolios are assessed in both cases and assigned categories, A, B, C or R, where A equals 5, B equals 3, C equals 1 and R is equal to 0. Categories A, B and C would qualify for submission to the TEC for peer review. Audits were conducted to confirm the data. Results indicate that attainment of an overall A rate has been difficult in that few researchers proved to be 'all-rounders' on the evaluation variables, leaving B and C rates for scholars said to have contributed significantly to the field of research.

The Evidence Portfolios contain research output of individual researchers for the period 1 January 1997 to 31 December 2002. Up to four nominated outputs and up to fifty (50) other research outputs could be submitted, bringing together the aspect of quality and quantity. Volume has been discredited by the RAE (UK) as encouraging overproduction, which then compromises quality. Research output is allocated 70% of the total Evidence Portfolio mark. Other variables used are peer esteem and contribution to research. Peer esteem, which is allocated 15% refers to recognizable achievements such as prizes and awards related to

research. Such important participation as “speaking” in conferences and other similar research related activities are also included. Eligible staff members are also expected to contribute to the research environment by supervising research students and attracting research grants. The latter two categories are together allotted 15%. External income is included as an indicator because it is believed to qualify as a “proxy” for research quality especially because such funds are usually released rationally. The measure of degree completion (by staff members) on the other hand is regarded as an indicator of whether or not TEOs fulfill their statutory obligations. As already stated, in New Zealand it is expected that research degrees are taught by people engaged in research.

At TEC levels, TEOs are expected to submit the Evidence Portfolios of members together with information on research degrees completed (Masters and Doctoral degrees) and external research income. Evidence Portfolios are worth 60%, degrees completed 25% and research income 15%. The last two variables have been based on 12 months of data, ending in December 2002. Panels took three to four days to complete their work and findings were reported to the monitoring / moderation panel. Cross-referencing was allowed and panels were free to determine the distribution of quality categories for their subjects in line with the evaluation criteria. Panel members had to excuse themselves when their colleagues’ portfolios were reviewed.

The actual focus of the 2003 assessment exercise was to compare capacity in the different TEOs and to compare levels of achievement in different subjects. If a researcher is poached by another institution 12 months prior to the census date, both institutions benefit proportionally. Eligible staff members who did not submit Evidence Portfolios received an R rating and this contributed negatively to their institutions and subjects. This R rating seems to indicate both non-compliance and poor performance and conclusions may be drawn that participation in the exercise is compulsory in the actual assessment exercise. This is despite the fact that final results reveal that the participation rate was 75%. Mentoring/moderation was conducted by a committee, made up of an independent chairperson and the chairpersons of the panels. The functions of the committee have been: to monitor consistency of framework application, to review standards and processes, to work out corrective measures

for inconsistencies and to advise the TEC. The TEC had a policy in place for data checking and verification.

6.4.2 Panels

The first PBRF cycle was implemented and managed by the TEC. Information about assessors in the first phase (TEO phase) is not clearly outlined in the report and efforts to obtain clarity have not been successful. The TEC structure, which is the second phase, has a project management team and 12 peer review panels, made up of seven to 20 members per panel, making a total of 165 members. Each panel is lead by a panel chairperson. Of the 165 members, 33 came from overseas. British experts were favoured for their experience in the RAE. The 12 panels were responsible for the assessment of forty-one (41) subjects (as units of research) and as already stated, panel chairpersons also served in the monitoring team.

Panel activities are guided by guidelines. Procedures also exists to solve conflicts of interest. Special advisors were utilized where necessary. Panels varied in scope and numbers of portfolios assessed. Some covered one subject while others were responsible for up to six subject areas.

6.4.3 Allocation of funds

The PBRF recognizes autonomy in the internal use of funds as is done by the HEFCs in the UK. On receipt of the PBRF funds, institutions use their own discretion for internal allocations. As in the UK, accountability is indirectly enforced through the expected responses to assessment results. Higher (Tertiary) Education patterns of funding used prior to 2004, the degree “top ups”, will gradually be phased out to be replaced by the PBRF funding (funding for strength), by 90% in 2004, 80% in 2005, 50% in 2006 and will completely be phased out in 2007. Disciplines (Subjects) are said to be funded directly through the research strength process.

The TEC avoided publishing individual research performance. This information was made available to staff members and their TEOs. It is anticipated that the new policy for

publication would be ready for use in 2006. The framework used and methodology applied all need to be improved (PBRF, 2003).

6.4.4 Participating Tertiary Education Institutions (HEIs)

Twenty two (22) out of 45 eligible institutions participated in all three components of the (TEC) assessment system (evidence portfolio, research degree completion, external research income). Two additional institutions participated in one component, one submitting on the 'degree completion' and the other on 'external research income'. The participating institutions are: eight universities, two polytechnics, four colleges of education, one 'wananga' and seven private training institutions. The number of portfolios assessed by peers was approximately 72% of the total eligible staff members.

Most of the universities, especially those with a long history of existence, achieved high scores. The non-university institutions on the other hand obtained very few A and B scores. The universities also performed at different levels, with the highest scoring 47.8% of the eligible staffs evidence portfolios on A and B rates and the lowest achieving 6.3%. The reverse was observed with R rates. The institutions that merged experience low scores while colleges of education scored the lowest.

The same observation was made with the age of the subjects. Older disciplines/subjects such as Psychology, Philosophy, Chemistry and others achieved higher scores than the others, with Medicine scoring the highest. As already stated, TEOs that chose not to participate in any of the components would not receive PBRF funding. It is believed that most of the TEOs that participated did so because of their research strength and it is feared that if all 45 participated, the average score of 2.59 (out of 10) could have been lower than it is.

Most of the scholars who rated R are at the early stages of their careers. It is therefore hoped that their performance will improve in the next round. Such improvements can be achieved over such short period of time (three years) if there are changes in the strategies of assessment, rather than rapid improvement of performance.

6.5 Problems experienced

The following problems have been cited by the PBRF (2003) report:

- There were no guidelines and therefore no control over the types of submissions
- Lack of clarity of assessment guidelines led to failure to comply in some cases
- Evidence Portfolios lacked professional appearance and quality for web publication
- Software problems
- Errors of auditing as a result of the tight schedule
- No audit on some of the work
- Evidence Portfolios lacked sufficient information. Most of them were not complete and many eligible staff did not submit. Some members felt that their work was not recognized as it should have been
- No follow-ups allowed – panels only used information in portfolios
- There were problems in creating comparable units
- It was not easy for staff to obtain an A rate since all the components of a portfolio would have to score high. As a result, some world-class, high-performers who did not fulfil all requirements of the portfolio failed to demonstrate standards in all the criteria (distortion of the real picture).
- Assessment period covered data for only six years and this would exclude major research output conducted prior to the period, leading to a decrease in the level of output.
- The number of participating TEOs cannot give a true reflection of a complete picture of quality across the whole sector of HE (the testing of strength)
- Assessment excluded part-time employees who could be making significant contributions
- Some subjects did not relate to well-established disciplines
- The main factors of assessment were ‘quality score’ rather than the number of researchers, which gave a wrong impression of research strength. Quality scores without consideration of eligible staff numbers only gives a picture, not the real situation in an institution. More areas have less than 10 As while others have no A at all.

- The impression given by R rates is that researchers could not achieve a C and yet some did not even submit
- Emerging researchers, who are most targeted for improvement, could not meet other criteria such as peer esteem
- Only about half of the portfolios nominated for A rate by the TEOs received the same rate from peer review, a sign of differences (discrepancies) in assessment between the two, the internal (portfolio) and the external (panel) evaluation stages.
- A high percentage, 17.3%, of submissions to panels with a C from the TEOs were assigned an R by the panel peer reviewers
- The clustering of subjects into areas makes it difficult to know which subjects within the areas are weaker.
- There are differences in the rating of staff in different subjects
- As funding is informed by results of PBRF and calculated over a period of time, it will be affected by fluctuations in the number of students

Most of the problems experienced in the 2003 exercise were refined in 2006.

6.6 Conclusion

The 2003 research performance exercise was the first of the PBRF periodic attempts. While the purpose of research may still need to be revised and methodology redesigned to meet such purposes, a few observations have been made from results of the exercise. There is a substantial number (5.7% of FTE weighted) of staff undertaking “word class research”. This serves as proof for international competitiveness. Yet care has to be taken not to extrapolate results to represent the research performance of New Zealand, taking into consideration the number that did not participate. There is also evidence of a wide gap between the well-performing and the poor-performing universities with one having achieved more than 50% of the good scores alone. Meanwhile, a quite sizable number (39.6% FTE) of staff do not meet the C quality category. The 2003 report attempts to attach this to young or emerging researchers, and believes that with their potential they will bring changes in the results of the future exercises.

It is interesting that some of the highly rated areas (on research output) achieved low scores. This may have been affected by the introduction of other categories in the assessment exercise. The rating at TEOs is only based on portfolios while the panels looked at a much broader scope.

For future improvements, the PBRF has contracted independent social scientists to evaluate the design and implementation of the programme. This process is supported by experts from the RAE in the UK. The period between 2003 and 2006 is nevertheless too short for any major improvements to take place. Government has the responsibility to improve capacity. The report indicates that after the 2006 exercise, systemic/periodic assessment will follow the period of six years as initiated by the UK. The third exercise is scheduled for 2012 wherein the PBRF' strength will be tested and the achievement of evaluation of objectives assessed.

6.7 Summary

The following points/tables sum up the chapter. As with the previous two models (chapters) the summary is divided into subheadings as follows; history and rationale of research evaluation, purpose, the evaluation process, benefits/lesson learned and challenges encountered. It should be recalled that only one (the first) research assessment is discussed in the chapter.

6.7.1 History and rationale of the study (point form)

- New Zealand HE (Tertiary education, as it is referred to) has been fully dependent on public funds for a while (at least since the 1980s), giving government all the steering powers over the system
- Changes in the ruling party led to continuous alterations to the system of tertiary education.
- In the 1980s (Labour Party-led) and the early 1990s (National Party-led) the focus was mainly on access to tertiary education
- This posed problems of the quality of students and thus education, leading to the need to improve the quality of teaching

- National-led government introduced staff exchange programmes which unfortunately opened access to information on research incentives (which did not exist in NZ) in other countries. This led to brain drain
- The late 1990s to early 2000 was characterized by strategic focus on the system of tertiary education (Labour-Alliance)
- In the early 2000, there was a need for changes in government policies on staff exchange to achieve staff retention through incentives and funding for research
- When Labour-Alliance took over in 1999, they introduced three strategic initiatives with the advice of the TEAC:
 - * the establishment of centres of research
 - * capacity building in social science research
 - * the PBRF exercise in the tertiary institutions (TEOs)
- The PBRF was implemented as the first periodic/systemic assessment of research in NZ in 2003.

6.7.2 Purpose

Several reasons emerge as purposes for the exercise that make it difficult to differentiate between the primary and the secondary purposes. The following have been identified:

- General improvement in NZ research performance
- Increase average quality of academic research by making funds available (quality improvement)
- Encourage research to support teaching (research standards are said to affect the quality of education)
- Advance knowledge and create a better leaving environment
- Increase research strength in tertiary education
- Fair distribution of research funds rewarding excellence (productivity of high quality of research and teaching)
- Accountability to research funds (reducing funds for poor performers)
- International competitiveness and recognition (proof for competitiveness)
- Encourage collaborative research

- Maintain consistency of acceptable standards across subjects
- Testing the country's efforts to undertake such an exercise periodically
- Provide incentives for resources to be invested strategically (depth not breath)
- Compare research capacity in the different TEOs

6.7.3 The exercise processes (point form)

- Retrospective (ex-post) research evaluation based on outputs over a set period of time
- Individual researchers are the units of analysis
- The use of outlined (standardized) guidelines
- Methods: Peer review and performance indicators
- Levels of assessment: evidence portfolios of staff in the TEOs
TEC peer review assessment
- Rating: A=5, B=3, C=1, R=0 (A,B,C in TEO qualify for submission to TEC for peer review)
- Period of data used, six years
- TEO data:
 - *70% allocated to output quality and quantity per portfolio (four nominated and up to 50 other outputs)
 - *15% for peer esteem (awards, prizes etc.) and contribution to research (speaking in conferences)
 - *15% for attracting grants (bring in external income), supervising research students and degree completion (statutory obligations of TEO)
- TEC data:
 - *60% evidence portfolio
 - *25% research degree completion (12 months data)
 - *15% external research income (12 months data)
- Panels:
 - *TEC manages the process
 - *5 570 staff members assessed (22 out of 45 institutions participated in all variables)
 - *12 review panels for 41 subjects (7-20members/panel) = 165 members (some panels covered up to six subjects)
 - *Cross-referencing allowed

*Panel chairs and members appointed (33 from overseas, especially the UK)

*Panels sat for up to four days and sent results to moderation panel

*Mentoring committee made up of neutral chair and chair of subject – monitor consistency review standards and advice TEC

6.7.4 Results

- Institutions: eight universities, two polytechnics, four colleges, one wananga, seven private training institutions
- Some institutions only participated in one variable, e.g. income
- Results not published but given to TEOs
- Universities fared better than other institutions
- Older institutions performed better than the rest
- Funding given to TEOs and used at own discretion

6.7.5 Benefits /best practice and problems/challenges

Information on benefits accrued from the exercise together with the challenges encountered is summarized in table 6.1 below.

Table 6.1 Summary of benefits and challenges of the exercise

Benefits/best practices	Problems/challenges
- Briefing of panels prior to commencement of exercise	- Amount of time spent preparing the portfolios
- Guidelines were outlined prior to exercise	- Generally low funds for research
- A substantial number of researchers undertake world-class research	- Poor planning and preparations
- Participation rate was high (75%)	- Differences in assessment standards between TEC and TEOs on the same data (differences in results)
- 22 out of 45 institutions participated in all variables	- No training mentioned for TEO evaluation
- Exercise accepted as a significant policy	- Open opportunities to alter TEO data prior to TEC submissions
- Movement of researchers from “drain to gain”	- Limitations on the time allocated for evaluation

<ul style="list-style-type: none"> - Research degrees to be taught by staff engaged in research - Cross-referencing allowed - Panel members excused themselves from assessing friends' work - If a researcher is poached 12 months prior to the census, both institutions benefit 	<ul style="list-style-type: none"> - Staff who did not submit also receive R rates – negatively affect institution (non-compliance or poor performance?) - Compulsory participation in this first exercise - Universities had advantage of receiving high scores - Big difference between good and poor performers - Not easy to obtain an A rate
---	--

PART THREE

THE SOUTH AFRICAN RESEARCH EVALUATION SYSTEMS

Chapter 7 The South African system of higher education research

7.1 Introduction

For the purpose of this study, discussions on the history of the system of HE research in South Africa in this chapter date back only to the times of apartheid. The reasons for the inclusion of the apartheid information was to build background on the status (similarities and differences) of the HE system in the country then and more recently in order to explain events relating to research and research evaluation in the universities.

Information includes the period from the early 1980s. This is despite the fact that Pienaar *et al* (20007) trace research studies in South Africa to the 1940s. The earlier period was not necessary for this chapter but will be discussed in chapter 8 under the history of the NRF rating system. The period discussed in this chapter was motivated by the information from Pienaar *et al* (2007), revealing that from the 1960s to the late 1970s universities were influenced to conduct operations research (for military support) as supported by government. It was only in the early 1980s that government started showing interest in the development of scholarship in universities.

By the 1980s, the entire system of HE was still racially fragmented. Throughout the era of apartheid HE systems, including research, were mainly funded by government (SAPSE 110). The more obvious government funding systems for research were the general university funds (GUF)/block grant subsidies, based on the SAPSE formula (output-based) and the funding from research councils. The latter is discussed in detail in chapter 8.

This chapter concentrates on research that is funded by the Department of Higher Education and Training (DHET). Information flows from historical events through to the operations of the new dispensation. Information on the latter mainly involves plans derived from legislative documents, including information on the *status quo* of HE research activities. This chapter is therefore divided into three parts. The first examines the history of the HE system dating back to the apartheid era, while the second provides information on the system in the post-apartheid period. The third part explores a policy developed by the DoE to reward research in HEIs.

7.2 The history of South African HE system and research (the old order)

7.2.1 Background

In the 1980s, the South African HE system was serviced by different departments (Melck, 1995). Universities that fell under the category referred to as White universities were catered for by the Department of National Education. The Indian and Coloured universities, which previously belonged to Indian and Coloured Affairs respectively, were assigned to the Department of Internal Affairs for current expenditure and the Department of Community Development for capital expenditure as government implemented changes within the apartheid system.

According to Melck (1995), the Black universities were divided into sub-categories of 'independent states' universities, self-governing states universities and those outside the other two categories (the national government universities). The latter group was made up of the original (old) Black universities which would have been served by the same Department as the White universities had it not been for apartheid, as would the Indian and Coloureds universities. The Department of Foreign Affairs was responsible for 'independent states' universities while the other two categories were the responsibility of the Department of Education and Training (current expenditure) and the South African Development Trust (capital expenditure).

These differences in departments had policy implications, resulting in the inequitable distribution of resources and serious disparities between the universities that government is presently battling to correct. The White Paper 3 (DoE, 1997a) clearly outlines the reasons that compelled the HE system to take the direction of transformation: the chronic mismatch between the output of HE and the needs of a modern economy. The White Paper also explains the unmet obligation for the 'foundation of critical civil society', imbalances and inequitable distribution of access.

Table 7.1 (from Subotzky, 1997) maps enrolment differences of post-graduate students at the historically Black (HBUs) and historically White universities (HWUs) caused by apartheid. The difference between the total number of Masters and Doctoral students marks the levels of

disparities at the time (12 years ago). The White Paper 3 also explains fragmentation in the system, with elements of insufficiency and ineffectiveness in some sectors and academic insularity and closedness. The latter already presupposes performance compatibility with other countries.

Table 7.1 Enrolment discrepancies between the HBUs and the HWUs in 1992

	HBUs		HWUs	
Honours	2, 385	54%	7,119	27%
Masters	1,688	38%	15,496	58%
Doctoral	318	7%	4,133	15%
Total	4,391	100%	26,748	100%

(Source: Subotzky – 1997)

The differences in areas of reporting also had other implications. Finance has obviously had more effect especially because South African universities like most in other countries depends mostly on government funding. Therefore, it is expected that assessment of performance in universities would be related to processes of accounting to government. This dependency is one way in which governments acquired all the steering powers and the rights to determine which university must fall under which department and the type and level of autonomy to be granted to institutions.

Although the public HE institutions in the country have always directly depended on the state subsidy, some level of autonomy has always been granted and related to internal regulations. This is similar to other countries discussed in this study, for example in The Netherlands. SAPSE 110 (:14) describes “university autonomy as a rightly cherished South African ideal”. Financially this means that a total block grant would be calculated per institution, with the discretion for use left to individual institutions. Allocations and expected accountability is discussed later in the chapter. The causes and effects of funding are discussed under the sub-heading of research funding.

During the late 1980s and the early 1990s sanctions were imposed on the apartheid government, which resulted in the SAPSE formula exceeding the amounts that Treasury could provide. This compelled government to introduce measures that would proportionately reduce the subsidy, the A-factor for current expenditure. This factor is explained later in this study.

7.2.2 Purpose for research evaluation in the old order

Research output of an institution was included under the SAPSE formula and calculated for purposes of funding. Submissions on research output were made by each university according to the criteria set by the Department. The Institution would then account by submitting an annual report. The internal counting of outputs for inclusion in the annual reports was a direct way of determining the research performance of a university, among other things, for funding.

7.2.3 University research funding process in the old order

Reinecke (1993) refers to the SAPSE 110 formula as a policy applicable from 1984 to 1992 for the allocation of research funds through the block grant. The formula indicates how institutions were rewarded for the number of articles published for both humanities and the natural sciences. The formula is listed in Appendix 1

It must be remembered that institutional autonomy granted each university “relative freedom to determine its own distribution priorities of the subsidy income between various expense categories” (Reinecke, 1993:17). Audits and annual reports to government served as tools for accountability within the framework of autonomy. Whenever government subsidies become inadequate, especially in the late 1980s and early 1990s, cutbacks would be calculated according to the percentage contribution of a university and relative to size (A-factor). When this was applied, no other consideration of differences would be made and all universities would be treated as if they were equal in all respects.

Financial pressures led to internal practices of “cross-subsidization” of programmes. This compromised programmes that would not be on the priority list of an institution. For example, if an institution did not value research highly, this could be compromised even when the published articles generated funds. This would then affect research activities, resulting in the reduction of output. When the difference between the research input and the subsidy formula is factored in, research output differences between universities that value research and those that do not is exacerbated. When the formula was revised in 1992 and in later years, the cutbacks indirectly affected research activities and research output was also affected. The same financial pressures affected the FRD rating system and thus incentives to undertake research. The rating system will be discussed in chapter 8.

As already alluded to previously, fragmentation of universities under the different departments of education had funding implications. Each department had its policies. For example, the universities under the Department of National Education would be allowed to accumulate reserves at the end of each financial year, while the Department of Education and Training surpluses were taken back (Melck, 1995). Historically disadvantaged institutions (HDIs), as Ishengama (2002) refers to the ‘non-White’ institutions/universities, would be expected to begin every year on a ‘clean slate’ and base all plans and developments on the new allocations. This further reduced the chances of development. It is not surprising that the output of journal articles was different between non-White and White universities.

According to SAPSE 110, not all HE programmes warranted government subsidy. Instructional programmes, research and public service programmes were significant for government funding (current expenditure). SAPSE 110 (15) classifies those that qualify for subsidy as those for which “significant public benefits are likely to ensure”.

7.2.4 The definition of research

For the purpose of this study, it is important to understand the definition of research at the time. SAPSE 110 (17-18) classified research as primary when it included activities that are “intended to produce one or more research outcomes”. The SAPSE 110 definition of primary research concentrates more on activities of outcome production, including the production,

recognition and application of knowledge. It is notable that SAPSE relates to the reward of output (productivity), especially of primary research.

Research falling under the abovementioned three categories of activities (by SAPSE 110) would be subsidized, while research including instruction and/or public service programmes would be subsidized under such categories (instructional and public service programmes). Research that relates to products such as artifacts and other similar products whose market price could not cover costs, received financial assistance outside the block grant (formula). According to SAPSE 110, other forms of research (falling under public service) with “no economic justification”, whose benefits are adequately reflected in the market price, do not qualify for subsidy and have therefore not been included in the subsidy (SAPSE formula).

SAPSE 110 cautions that research related to public services may be in the form of consultation, in which case the client may be responsible for expenses. There is therefore no justification for subsidies. Such research activities would be defined by other resource bodies, either business or other government departments, and this would be classified as consultations and/or contract research.

The above discussion classifies research into four categories, which may be utilized for the purpose of evaluation. The first primary research that produces publishable output (including the creation of knowledge), while the second category includes research related to instructional services. The third category relates to research related to public service programmes, including contracts and other forms that may not produce publishable outcomes. The fourth category relates to artifacts. See table 7.2.

Table 7.2 Categories (types) of research summarized according to SAPSE 110

Categories of research types	Source of funding
Primary research (publishable outputs)	Block-grant
Instructional and institutional public service research (operational and mostly not publishable outputs)	Block-grant

Artifacts and similar programmes	Special funding
Public service (consultations and contracts), reports	Business and government departments

The definition/classification of primary (subsidized) research also served as a guide for internal allocation of funds. As in the other countries discussed in this study, no formal directives are laid down on how the subsidized funds (primary research funds) have to be internally distributed to departments or the researchers that published the article. Although this leads back to the previous explanation of autonomy, some guidance, rather than instruction, would have provided some similarities in the prioritization of research in institutions.

7.2.5 Measure of research (methodology)

According to SAPSE 110, research output is regarded as the most appropriate factor to base subsidies on. The document (SAPSE 110) recognizes the fact that when a subsidy is provided for output only, this may imply that the “potential beneficial” outputs such as instructional and operations research that produces little or no publishable results will be discouraged. How the SAPSE formula accommodates published articles can only be explained through the “public benefit” rhetoric.

In order to accommodate all types of research, SAPSE 110 suggests that evaluation of research programmes be based on both input and output. This occurred when difficulties in quantifying research output were experienced and this led to the use of inputs as a measure to estimate productivity (SAPSE 110). It has nevertheless been difficult to determine subsidy based on such measures. For example problems arose in cases where it was assumed that money spent on research over a period of time implied that benefits are estimated to be equal to costs. Another measure that was considered with difficulties of reliability is the time devoted to research as opposed to other academic activities. According to SAPSE 110, usable measures for output are articles and books, in which the number of words written and the number of pages have been taken into consideration.

7.2.6 Dimensions of quality and productivity

Similar to challenges experienced in other countries such as The Netherlands, the dimension of quality and productivity needed to be clearly clarified. In defining quality, SAPSE 110 discourages the use of volume and suggests a few measures including the impact factor of a journal, visibility (as measured through citations) and the use of the list of accredited journals as approved and acknowledged by the Ministry of Education. The latter was introduced in 1985 and has been used as a measure of research output by the then Department of National Education to also determine productivity for SAPSE allocations (Mouton *et al*, 2006). The Department has always been responsible for the compilation and approval of accredited journals based on the reputation of a journal, regularity of publication and circulation, the status of the editorial committee (members) and the general status of the journal. Such accreditation has been done for both national and international journals for quality control. These measures used by government to allocate subsidies influenced their use by institutions for internal allocation, although not at the behest of the Department. Institutions would formulate their own internal research allocations policies and procedures.

7.3 The post-apartheid era

7.3.1 Introduction

The end of the apartheid era in South Africa in the early 1990s was marked by changes in the ruling party, characterized by the introduction of new laws and policies. A similar situation arose in New Zealand in the late 1990s, which brought about changes in paradigms and ideas in a bid for the new party to correct “disorders” perceived to be created by the previous ruling party (Mc Laughlin, 2003). As apartheid was perceived as the main cause of disparities in South Africa, eradication of the system and redress of its effects were the unquestionable fundamental objectives of the new regime.

Redress would not come easily as the foundations of inequality had already been laid, the most obvious being unequal standards (levels of operation) of education even among HEIs. If everything were neutral, the disparities would be remedied by holding the previously advantaged institutions at a development standstill to allow others to catch up. Steyn (2000) explains the difficulty of doing this by referring to the democratic rights of those previously

privileged to further learn and develop. This belief is contested by those who view it as a way of protecting turf (Christie, 1993 in Steyn, 2000).

Vibrant and drastic changes had to be implemented in order to turn around the entire HE system. One of the changes that had to take place was opening the doors of learning to all South Africans as stated in the Constitution of the country. Melck (1995) feels that this era led to increased access to universities. This practice of dealing with inequalities was also experienced in New Zealand (McLaughlin, 2003) at almost the same time. As in New Zealand, economic inequalities and inequities of educational standards posed problems. The South African government was faced with the task of rendering equal HE services to an 'unequal' population through unequal universities.

Towards the end of apartheid, the weakened economy had already compelled government to abandon the previously-used funding formulae and to resort to open-ended, demand-driven funding strategies in which allocations were made according to the "priority of officials" in the then Department of National Education (Melck, 1995). Although this is said to have led to some institutions not receiving any funds at all in some years, this did not contribute towards or hasten the process of equalizing (correcting) the system when apartheid ended. The elite phenomenon of the historically advantaged White universities had already been factored in. This would create performance differences between universities that would be difficult to bridge.

7.3.2 Plans in the new dispensation

A few attempts were made by the new government to work out a system for the creation of a system of HE in the country. After a series of gatherings that led to the production of different documents, the "White Paper 3 – A Programme for Higher Education Transformation" was promulgated (DoE, 1997a). The White Paper 3 outlines planned activities of transformation of the HE system in South Africa. This study took transformation as a point of reference in attempts to equalize and democratize HE. According to the White Paper 3, the most persuasive reasons for this direction of thinking were to 'redress past inequities', 'to serve a new social order', 'to meet pressing national needs' and 'to respond to

new realities'. That is, that the HE system would, while dealing with its own problems of equality, also be expected to "contribute to the building of a better life for South Africans". The White Paper 3 maintained that decision contexts in HE needed to be based on these intentions in order for the system to serve its purpose.

The point of departure according to the White Paper 3 was to treat the HE system as a single unitary system in order for it to contribute to the social, cultural and economic development of the country (DoE, 1997a). This amalgamation was also explained by the National Plan for Higher Education in South Africa (DoE, 2001). The single co-ordinated system would, according to the National Plan, be achieved incrementally, observing the missions of the institutions (in their three-year institutional rolling plans). The National Plan for Higher Education in South Africa (DoE, 2001) policy document was introduced to 'operationalize' the ideas of the White Paper 3.

The National Plan (DoE, 2001) was therefore meant to harmonize the HE system and address inequalities of the past. The 'new' National Department of Education intended achieving these objectives by strengthening the "weaker" institutions. The maintenance of standards of the better-performing institutions raised threats for the process of redress, especially in the short-term. Hopes were raised that in the medium- to long- term the process of strengthening the weaker universities would bridge the gap. Under these circumstances (of disparity) one would expect the fulfillment of some of the expectations of the HDIs to only be achieved through the implementation of a long-term plan.

The implementation of the National Plan was meant to allow the HEIs to meet the challenges that the system hoped to address through the production of graduates with high quality skills and competencies. The system intended to achieve this through the creation of a future generation of intellectuals and researchers of a representative demographic makeup, considering the fact that Black people were previously disadvantaged. It is not clear how the unequal standards could be equalized.

In order to “meet pressing national needs” and to “respond to new realities” the National Plan suggested that government maintain the standards of the better performing/developed institutions. Institutions would also produce research that would contribute to the economy and expose the country to global competition, creating a learning society that would advance, develop and enrich itself intellectually within the diverse system.

7.3.3 Research activities

It is necessary to review the role of HEIs in the new dispensation at this stage to establish the importance of their contributions in the field of knowledge production, as this area is expected to assist in meeting the challenges. The National Plan for Higher Education in South Africa (DoE, 2001) refers to research as “a principal tool for creating new knowledge” with a potential for inquiry and critical thinking, which has to be disseminated through “teaching and collaborative research tasks”. Institutions are also expected to establish a high-level research capacity that sustains application of research activities. These two expectations form the core of activities of research development in the new system. Other stated expectations involve the processes of establishing a more representative staff complement with well- developed capacity and securing sustainable funds shared by institutions to support national goals (DoE, 1997a).

Meanwhile, to fulfill national needs, the national policies and priorities charge the public funded HE system with accountability by creating a diverse system of a mix of institutional missions under a single co-ordinated system. Whereas the HE system is said to have the “caliber of research and teaching ... comparable to international best practices and standards”, the National Plan for Higher Education in South Africa (DoE, 2001) explains that the demographic profiles affect the “quality and quantity of research graduates and research output” as a result of the historic arrangements, continued racial divides and increasing competition between and among institutions. The White Paper on Science and Technology (1996) explains that the HE sector receives government funding through different systems to perform research. In turn, the sector is expected to provide talented human resources, meet local and international needs in knowledge production and create and sustain centres of excellence.

The White Paper 3 identifies redress of past inequities in research strength as the main challenge faced in the field of research. This is mainly because this component is essential for addressing national needs in partnership with stakeholders and to keep abreast of global trends.

Parallel to these expectations were the multiple challenges brought about by apartheid, which compelled redress through research capacity building, especially of the previously disadvantaged academic communities. The White Paper 3 refers to the existing link between research capacity and post-graduate studies as having been inadequate (in 1997) but necessary, especially in the historically disadvantaged institutions, characterized by imbalances of race and gender. Whether or not the link has now improved (in 2006/7) can only be established through research. Claassen (2006) nevertheless reveals that the same relations still exist in South African universities, which creates differentiation and continues to place the previously advantaged in an advantaged position. These are signs that a lot has to be done to reduce and ultimately close the disparity gap..

7.3.4 Global competitiveness in research

While the national issues form the base for development in the country, international competitiveness is said to be giving direction and setting the pace. The White Paper on Science and Technology (1996:4) argues that “the setting and realization of national goals cannot be undertaken without due regards to the implications of global imperatives”, and further elaborates that “the greater challenge will be to integrate successfully in the global systems and communities while addressing the local needs”. This global concern is also mentioned by the Higher Education Act 101 (DoE, 1997b) and is to be kept in mind when the objectives of knowledge production (research), together with efforts of assessing activities thereof are discussed.

This process emerges as part of the bigger picture of the national agenda for equality said to be embedded in the demands of globalization. The White Paper 3 also confirms that the national processes were set in place to operate against the reality of global imperatives.

Inclusion of the element of global operations adds a slightly different dimension to the national demands, which may affect the direction followed in the 'nation building' strategies. What this implies is that while the system serves the national needs, it also has to "respond to new realities and opportunities" (DoE, 1997a:7). This may also involve providing the necessary expertise for a modern economy while pursuing scholarship and intellectual inquiry.

The big challenge is the fulfillment of the obligations with such a 'non-equal' HE system and the fact that the bridging of inequalities and the response to new realities should run concurrently and should be based on global standards. This was the experience in New Zealand, where the plan for research evaluation had multiple purposes that all had to be observed at the same time.

The White Paper on Science and Technology (1996) believes that education will prepare researchers for more effective and innovative challenges, both locally and internationally. The White Paper claims that "a research mindset is essential in order to optimize the benefits of technology transfer from abroad" (White Paper on Science and Technology, 1996:10). It is for this reason, and the fact that the White Paper directly refers to involvement in research of the Department of Science and Technology, that it is included in this study.

In an attempt to meet global demands, the White Paper 3 intended to recognize the existence of institutions that achieved international standards in teaching and research capacity, thus influencing the direction of transformation towards the improvement of national development and standards for international challenges. This somewhat neglect of the reality of inequalities for the purposes of being competitive may have long term repercussions on the agenda of equality and the transformation process.

With the large number of previously disadvantaged HEIs, recognition of a few institutions that are competitive does not bring about an easy solution to inequality problems especially in research productivity. Meanwhile, the National Plan for Higher Education in South Africa (DoE, 2001) acknowledges that the HE system is not operating efficiently even in the 21st

century in terms of its core mandate. The production of knowledge and graduates needs to improve before embarking on new functions and roles (as confirmed by the White Paper 3).

The National Plan for Higher Education in South Africa (DoE, 2001) suggests that better performing institutions need to still be maintained and supported, and that those that are not be provided with opportunities to improve. The challenge, according to the National Plan, is to “nurture” a new generation of intellectuals, professionals and researchers, especially Blacks. When this is fulfilled, inequities will hopefully be redressed. The Department of Education believes that traditional HEIs that value scholarship (teaching and research) are better placed to play the role of knowledge generation and intellectual development for social interaction and engagement. One wonders where this idea places the other institutions..

Steyn (2000) classifies this problem as a quality/equality dichotomy. Quality in this sense relates to global competitiveness while equality refers to the unification of all the fragmented systems and making sure that all institutions enjoy equal opportunities. While equal standards are not emphasized here, it appears as though these may emerge as a product of quality and equality. Steyn nonetheless warns of the possible underlying conflicts between quality and equality, stating that unjustifiable inequalities will always be damaging not only to democracy but to the quality of education as well. As stated previously, on a narrow scale, whereas all individuals and institutions have the right to develop further, the national *status quo* (of inequalities) provides a tricky situation to deal with.

Claassen (2006) reveals that the top publishing universities and those with high enrolment of Masters and Doctoral students still remained the previously advantaged institutions by early 2004 (see table 7.3). According to the table, the Universities of Stellenbosch, Cape Town, Pretoria and the Witwatersrand continued to excel in publication output and the production of Masters and Doctoral degrees. Two out of the four are previously English universities while the other two were predominantly Afrikaans universities in the past. The University of Natal and Potchefstroom University have recently been merged with HDIs and although Claassen’s data reflects that of the old institutions, there is evidence that the new University of KwaZulu-Natal and the North-West University, as they are now referred to respectively still rank

relatively high (9 and 10) in the DHED 2007-2008 research output and even higher when other measures are considered (Mouton and Weingart conference report, 2010). One may draw conclusions that a lot has to be done to increase and broaden participation if national demands are to be met and to expand expertise for quality performance in order to best prepare for competition in the global arena.

Table 7.3 Distribution of research among universities in South Africa (Source Claassen 2006)

<p>◆ Research publication outputs: 70% of the publ. output units are produced by: 6 universities (SU, UCT, UP, Wits, Natal, UNISA)</p>
<p>◆ Master’s and doctoral degrees (general) Two-thirds of master’s and doctoral degrees are produced by (out of 22): 6 universities (SU, UCT, UP, Wits, NWU, UNISA)</p>
<p>◆ Master’s and doctoral degrees (in SET) Two-thirds of master’s and doctoral degrees are produced by: 5 universities (SU, UCT, UP, Wits, Natal)</p>

By implication, the demand for international competitiveness is another factor that poses a threat to the process of redress as it has introduced competition between and among local universities. The Department of Education fears that “the increased competition between HEIs has further fragmented the system and exacerbated the inequalities within the HE system” (DoE, 2001), research included. This is a realistic fear in that a competitive spirit would benefit those that have already improved, such as Stellenbosch, UCT and Pretoria, widening the gap in inequalities even further. Thus, competition, which is good for international comparisons, further threatens the equilibrium that the Department is attempting to achieve and maintain. The Department of Education welcomed competition that promotes

innovation and enhances quality (DoE, 2001), and suggested that such competition be regulated within a national framework for the purposes of sustaining the system. Once more, the challenge is to have a balance between international competitiveness and equalizing the levels of institutional plans on equity among staff and graduates.

In their three-year rolling plans (within their mission and vision) institutions are expected to indicate their goals and direction on the basis of analysis of their social/economic, political, intellectual and cultural context. Institutions are also expected to describe the “supply-side” factor (intellectual capacity in terms of staff, infrastructure and financial resources) against “demand-led” factors in the country. Information on these factors would assist in establishing the levels of efficiency and effectiveness of institutions. The National Plan for Higher Education in South Africa (DoE, 2001) acknowledges that some institutions lack planning capacity and end up imitating those that are regarded as “prestigious”.

The lack of or even poor capacity threatens the chances of development in those institutions whose improvement would narrow the inequality gap. The National Plan blames imitation on the lack of a “regulatory and policy framework designed to ensure diversity in the ...system” (DoE, 2001:2). The introduction of a funding and quality assurance framework, it is believed, would encourage “academic drift” and all institutions would then be assessed by the same performance benchmarks and compete for the same financial rewards from the same “often limited pool of resources”. In this sense (within a framework), competition encourages initiative . Again, this seems like a long-term goal, which requires meticulous and rigorous implementation to fulfill. Otherwise, the existing Oxbridge effect created by the apartheid system may be difficult to deal with, as New Zealand experienced (PBRF, 2006).

In attempting to achieve the goals of transformation (stated in the White Paper) and still maintain standards, the National Plan for Higher Education in South Africa (DoE, 2001) suggested coherence in the provision of HE with the limited resources used efficiently. This may be done such that the quality of teaching and research are improved across the system to address the demands of the country and uphold standards of academic quality. The urge to redress inequalities and to ensure “capacity of institutions to discharge their institutional

missions within an agreed upon national framework” (DoE, 2001) may be achieved through continuous collaborative efforts between the DoE and HEIs.

It is necessary to have a clearer understanding of expected contributions within the HE system in order to match expected participation with global activities. HE, according to the White Paper 3 (DoE, 1997a) is expected to mobilize human talent and potential, develop professionals and knowledgeable workers of international equivalence, and produce vibrant research and development systems. The system is also expected to integrate research with other services through well- planned research programmes, observe a democratic ethos and a culture of human rights and uphold rigorous standards of academic quality to sustain the quality of life. All core functions of HEIs are expected to fulfill these expectations as individual entities and as a collective. Therefore, research, which is a focus of this study, should strive to improve capacity to achieve these goals.

Unlike in New Zealand, the UK and The Netherlands, the White Paper 3 (DoE, 1997a: 54) argues that “good scholarship (which should characterize all HE academic staff) is not the same as research competence (which a minority of academics in the sector as a whole will either need or possess)”. The White Paper 3 claims that active research may assist in teaching but that teaching is not dependent on research experience. To this effect, government takes the route to utilize what the White Paper 3 labels “relatively scarce funds available for the development of research capacity” in a very conservative manner despite the demand for research development.

The White Paper states that “public funds for participating in research, whether basic or applied, should not be spread across all faculties or schools in all institutions but should rather be concentrated in those areas where there is demonstrable research capacity or potential”. This ‘rule’, as stated before, is meant to apply to all institutions as if they were the same in all respects. What is meant by ‘potential’ is not clear especially considering the fact that research is one of the core businesses of an institution and is internationally related to scholarship. One wonders whether the ‘rule’ does not symbolize the loss of the ‘redress battle’.

7.3.5 Defining dimensions of research in the new dispensation

7.3.5.1 Quality

The quality dimension is embroiled in the quality/equality dichotomy and has been observed to pose a threat to the process of transformation. In a bid to explain how quality fits within the process, Steyn (2000) uses words alike excellent, perfection, fitness for purpose, value for money and transformation to explain the concept. If all these are translated to explain research, the first three concepts relate more to some standard or level of performance (which may mean quality in this sense), while value for money would either refer to efficiency or ordinarily to the word 'importance'. It is difficult to conceive how transformation can actually mean quality in this type of a situation. Steyn insists that education can be assured to have quality if it provides opportunities for learners to develop their full potential, which the author attaches to transformation.

The White Paper 3 gives the impression of quality as academic and educational standards and refers to quality as some compliance with specific expectations and requirements in the sense of aims and ideals of excellence dependent on the specific purpose pursued. Application of quality, according to the White Paper 3 (DoE, 1997a: 2), involves "evaluation services and products against set standards with a view to improvement, renewal or progress". The White Paper 3 also alludes to the principles of effectiveness and efficiency as institutional functions to achieve desired objectives in terms of optimal utilization of available resources. These may be achieved in an atmosphere that allows for the advancement of scientific inquiry and in institutions that command self-regulation and independence of internal systems of teaching and research. The White Paper does not refer to transformation as explaining quality.

The White Paper on Science and Technology (1996) on the other hand, is confident in the quality of basic research produced in the country, stating that "South Africa has a proud record of quality basic research and" that "it is important to sustain" it. This is particularly important because of the need to preserve the "bridge that such research provides to the international scientific environment" (White Paper on Science and Technology, 1996:10). This idea to "sustain quality" is embraced by both the White Paper 3 and the National Plan for Higher Education. For this quality to be sustained, the White Paper (1996) suggests

incentives to encourage scientists. It is not clearly explained how those with untapped potential, especially Blacks, will be encouraged to increase participation. This idea holds more ground for established researchers involved in research with long-term benefits and those that are trained by them.

Quality of research in this source (White Paper on Science and Technology) is defined through “publishing in high impact journals” and “testing their (researchers’) ideas against international best practices by presenting their ideas in international fora” (White Paper on Science and Technology, 1996:10). According to the White Paper, it is the “tested process of bench-marking against the best in the field” that defines the quality of research work done. This ‘global’ definition is used by the Dutch, and in the UK and New Zealand. The Dutch arrived at and confirmed their stance through the evolution of their evaluation exercise. When imbalances of capacity created by apartheid are taken into consideration in South Africa, one may conclude that the process may first need a reasonable period of time to achieve equilibrium across HEIs in the country for it to acquire required the standards holistically. The Dutch strategy of evaluation may serve this purpose. This is important to note when evaluation is planned for.

7.3.5.2 Relevance

Relevance is sometimes referred to in the DHET documents as fitness of purpose. The White Paper on Science and Technology (1996) suggests that research aims be articulated to decision makers and the population at large. For this reason, research is viewed as an important responsibility of government, industry and science institutes in collaboration, with government taking the lead to ensure the realization of national goals and the HE system playing the role of undertaking research and producing researchers.

The White Paper on Science and Technology (1996) regards research as an important criterion in the development of a nation and for placement of a country in the international world. The White Paper cited the improvement of the National System of Innovation to enhance and sustain the quality of life, and to address the needs and aspirations of citizens (economic and social development) as an important contribution achievable through research. Also

important is the contribution of research to assist in the process of equitable development and to establish and sustain a democratic culture. Research is also expected to improve technology to 'relevant global trends' so that the country becomes economically competitive.

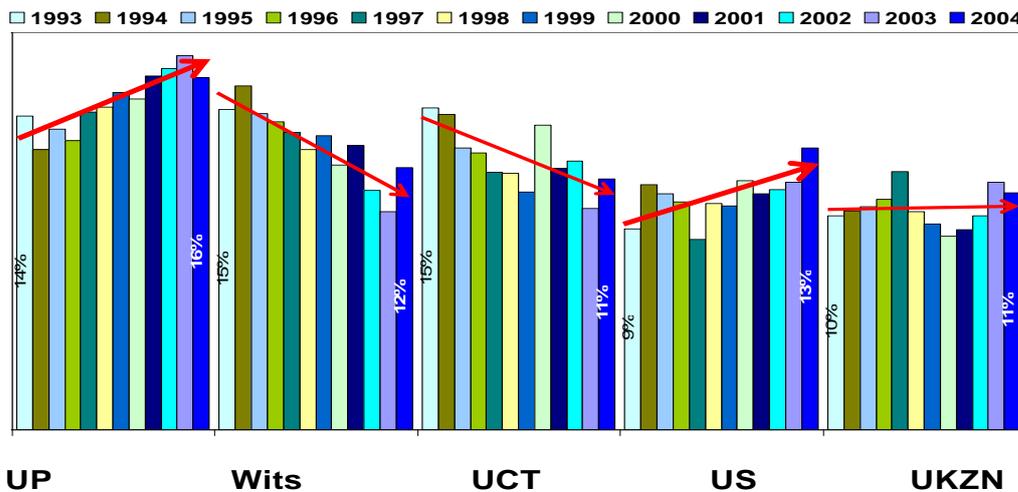
SAPSE 110 also referred to public service programme and contract research as a service provision to industry and government by HEI researchers. This is expected to continue until such time that other structures and departments view research as an/a imperative/necessity for the results of their work to appear in the public domain. It is also noted that government is responsible for policy setting, resource allocation and for legislating on regulations (White Paper 3) and therefore has to base decisions on empirical and scientifically determined data.

The demands for public service research, as previously indicated, became more relevant with the requirements of the new system and influenced the direction of researchers towards applied research. This decline in basic research and/or research output, as a result of the new system are acknowledged by the White Paper 3. As stated, both forms are equally important for development in South Africa. Whereas it is believed that applied research has the potential to "deepen democracy", basic research serves as a tool for "contributing to the global accumulation of knowledge" (DoE, 2001:16), and is said to be placing South Africa on the global map of knowledge production.

Basic research is also regarded as "crucial in nurturing a national intellectual culture" and is said to be contributing to improvements on discipline-specific human resources. This resource type research also creates opportunities for participation in international scientific development and thus innovation (DoE, 2001:16-17). Its importance should therefore not be compromised nor ignored and the National Plan proposes its promotion.

The National Plan for Higher Education in South Africa (DoE, 2001) acknowledged that there had been signs of a decline in research output probably as a result of the inclination towards applied research in recent years in the country. Such declines were observable at the time in some of the top contributing universities in the country such as the University of the Witwatersrand (Wits) and the University of Cape Town (UCT) as indicated in figure 7.1,

although other factors not indicated in this study may be at play. The decline has to be viewed as having been specific to the period, with a subsequent increase in research output from 2005. The increases in the University of Pretoria (UP) and Stellenbosch University (SU) have not been accounted for. The inclination towards applied research is said to be influenced by the sudden availability of resources for contract research, results of which may not be published and thus not made available for national and international scholarly consumption.



(Source Claassen – 2006)

Figure 7.1 Indications of a decline in research publications in some top publishing universities in South Africa

7.3.5.3. The sustainability and productivity dimensions

Sustainability in research is said to be dependent on the ability to respond to the opportunities and challenges provided by the global transformation in knowledge production and dissemination opportunities, which themselves are affected by the increase in global communication technology (White Paper on Science and Technology, 1996). The National Plan for Higher Education in South Africa outlines the goals of research development as securing and advancing high level research capacity to ensure continued intellectual inquiry and sustaining the application of research activities.

Similar to SAPSE 110, the White Paper 3 refers to research in South Africa as aiming to contribute to the production and dissemination of knowledge in collaboration with teaching

(through post-graduate teaching and training). To achieve this, the National Plan suggests an approach (objectives) that sustains current strength, promotes research and knowledge output to meet national developmental needs and enables the country to compete in a global context. To fulfill the objectives the system has to increase both research activities and the production of postgraduates. The process of research development, it is believed, is generated through postgraduate studies and collaborative research.

As indicated, output, on the other hand, is said to be concentrated in only a few institutions. This is because research and postgraduate training require a “critical mass of human and other resources” (DoE, 1997a:17). To achieve/expand this, the Department of Education placed hope on the intended development and improvement of the HDIs. Unfortunately, this may not be achieved in the pace that was expected.

Government addressed the decline in research participation by developing multiple sites for research and knowledge production (outside universities) to increase response beyond local imperatives. This matter is not addressed in this study. What is clear is the plan to improve productivity, to sustain existing output, facilitate collaboration and partnership and promote articulation between different areas of research.

7.3.6 Types of research

The White Paper on Science and Technology’s (1996) approach to research is twofold; for innovation, wherein basic research plays an important role, and the problem solving approach (applied research), dependent on multidisciplinary research. According to the White Paper 3 (DoE, 1997a), basic research is important for the support of the national intellectual culture and to generate high level and discipline-specific human resources towards global standards. Applied research on the other hand, is necessary and relevant for knowledge application and nation building. The latter approach is drawn from collaboration to produce application-based research. As advancement of knowledge through academic resources is also viewed as necessary for the co-ordination of democracy in the country, it is believed that both basic and applied research in the natural and social sciences contribute to innovation.

In order to improve research participation and activities in the HE system and the country as a whole, the White Paper 3 (DoE, 1997a) suggests an increase in resources (including the search for new sources of funding) and encouraging their effective use (for example the use of centres of excellence).

7.3.7 Other factors affecting research in HE

It is important to review other factors that affect research in HE to promote understanding of the causes or effects of occurrences that relate to such research. Issues relating to governance, institutional autonomy and accountability and those of funding are discussed hereunder.

7.3.7.1 Governance

Governance has been identified as one of the operations of HE that affect research activities. The White Paper 3 (DoE, 1997a) recognizes the principle of co-operative governance within the framework of institutional autonomy (within-public-accountability). In this system, government's influence occurs through policies and the involvement of the intermediary bodies that are said to be transparent. This type of governance excludes government from the micro-management of institutions and may be interpreted as an attempt by government to reduce its steering powers.

It should be remembered that transformation continues to feature as the main goal for government. With the CHE as an advisory body to the Department of Education, the Council may, for this reason, encourage the building of research strength. This may be done through post-graduate training, infrastructure improvement and strategic development especially for the HDIs. These three aspects have been identified as cornerstones for the eradication of inequalities in research performance, especially between the previously advantaged and those that did not receive equal attention in the past (DoE, 1997a).

According to the White Paper 3, the CHE may also develop or advise on the development of a performance assessment system to assist in the identification of progress and further needs. In this way, the White Paper relegates some of the responsibilities of evaluation to the CHE.

An advisory body brings hope, especially because discussions indicated the difficulties experienced when institutions have to handle the situation on their own.

The White Paper 3 also emphasizes the involvement of the NRF as an intermediary body to advise the Ministry of Education on the future needs of research infrastructure and capacity especially related to institutional improvement and redress of past inequities. Through the Foundation, the system intends to increase and mobilize both private and public funds for research through partnership. Support by the NRF is discussed in the next chapter.

Also important is the Branch of the Department of Education responsible for policy formulation within the system of transformation. The Branch has the responsibility to develop performance indicators to collect information from institutions. Such information may be passed on to the CHE and be used for research purposes (discussed in chapter 9) and/or for the allocation of funds (for example the general and the specific-purpose funds). The Branch, it is stated, may also commission research that would promote its functions. While the White Paper 3 attempts to define the duties of the CHE and those of the Department of Education Branch, it is clear that some overlap exists. The following quotes (DoE, 1997a: 41) are some attempt to clarify the situation:

“The Higher Education Branch and the Council on Higher Education will co-operate in undertaking their respective functions, sharing information and jointly participating in meetings and on projects as appropriate”.

The next quote explains the additional responsibilities assigned to the Branch:

“The Higher Education Branch will also be actively involved in inter-departmental activities that work to improve the coherence of the government’s overall approach to social and economic development, labour supply and skills formation, health and social justice.”

Thus, the Branch is the wing of the Department of Education that directly relates to the intermediary bodies. The bodies mentioned above are direct stakeholders in the system of

research in HE and are relevant to this study. Any plan for the evaluation of the system will involve these bodies in one way or another.

7.3.7.2 Institutional autonomy and public accountability

For a better understanding of the internal regulation system of universities (which serves the self-evaluation duties), it was necessary in this study to review the governance of the South African HEIs. It should be noted that the White Paper 3 (DoE, 1997a) and subsequently the Higher Education Act 101 (1997b) observe the internal governance system of HEIs. The White Paper 3 explains the position of the Department of Education, and its intentions not to micro-manage the internal activities of the institutions.

The internal Councils of institutions as the highest decision making bodies of universities, are expected to create good governance and be responsible for the mission, institutional culture, policies, performance quality and the reputation of their institutions. All the responsibilities stated here are important for this study, especially as far as self-regulatory processes are concerned. When good governance is in place, quality performance may result, enabling Councils to sustain public confidence, an area in which research plays an important role. Implementation of all the plans of the Councils is the duty of the executive management of the institutions. To this effect, HEIs have their own research management offices that are regulated internally to provide support for researchers, research students, research units and departments. Internal policies and procedures therefore govern these offices.

These policies and procedures are themselves based on the national policies and procedures. In this way, the HEIs are autonomous but accountable to the National Department of Education. It is through this system that inequalities within the HE system can be curbed. As stated, if some of the institutions are not able to manage their planning systems and instead imitate others, autonomy gets affected.

The Ministry of Education is said to only intervene in extreme circumstances to “restore good order and governance”. In order to sustain equilibrium across the system, the National Plan for Higher Education in South Africa (DoE, 2001) states that “Ministry will not hesitate to

intervene to ensure the implementation of national policies and transformation goals should this prove necessary”. This implies that autonomy is only recognized alongside accountability, wherein institutions “enjoy freedom and autonomy in their relationship with the state within the context of public accountability and national needs for advanced skills and scientific knowledge” (DoE, 1997a).

The National Plan for Higher Education in South Africa recognizes autonomy provided that it is not used to “prevent change and transformation”. It is expected that the transformation of HE exists within the national quality framework, within the sphere of their diverse programmes. Although this has always been stressed to equalize instructional programmes, not much emphasis has been placed on the research component, probably because of its individual-initiative nature (initiated on individual will). The CHE has attempted to provide some guidance through the “Good Practice Guide” discussed in chapter 9, for the purpose of improving the activities of research management in the universities.

Public accountability is exercised through making public funds for institutions conditional on their Councils providing strategic plans and reporting on their performance reference to their goals. It is expected that this will be done within a framework for continuous improvement based on quality assurance for the purpose of transformation and responsiveness to social and economic needs. These plans, the White Paper on Science and Technology (1996) suggested, need performance indicators against which they may be measured. By implication, there is a need for both internal and external assessment for which one purpose is accountability for the use of funds. Performance indicators were popularly used in the 1990s and the White Paper reflects this. If the system of assessment is not implemented yet, at least the need for it was raised by the White Paper (1996).

7.3.7.3 Funding

In the new dispensation, public HEIs continue to receive funds from the state (Pienaar *et al*, 2000). In addition, the Department of Education is said to have some responsibility to maintain growth, equity and quality objectives in the HE system, despite the stated institutional autonomy. According to the White Paper on Science and Technology (1996) the

Department of Education is responsible for research funding through the general university funds, which are grants based on publications in peer reviewed journals, full-time equivalence (FTE) funds, for which institutions have to account and innovative funds obtained through the NRF on the submission of proposals. The grant that is based on publications in peer reviewed/accredited journals is discussed in the section that follows (7.3.8) while the NRF funding is fully covered in chapter 8. This sub-section mainly focuses on the FTE funding system, together with public service programmes and grants.

While the Department of Education carries responsibility for the regulation of public funds, it is also responsible for the control of programmes and services, improvement of completion rates and regulation of labour-intensive strategies. The White Paper 3 (DoE, 1997a) observed that expansion of services without new investments results in overcrowding of facilities and functions, which usually lowers the morale of academics, leading to a reduction in research activities.

As government also intends to invest in postgraduate students to increase academic and labour manpower, this is done on the assumption that the present labour power (well qualified academic staff) is either aging or immigrating. The Department targets block funding to support research based postgraduate students and once more this is done 'in fields where institutions have demonstrable research training capacity'. Earmarked funds are also used, on request, for infrastructure to support postgraduate training.

In this period of transformation, which changed over time, South Africa experienced pressure on financial resources, with government expenditure (as a percentage of Gross Domestic Product abbreviated as GDP) showing unstable tendencies with time. Expenditure has increased from 0.72% (of the country's GDP) in 1995-1996 to 0.77% in 1996-1997, fluctuating between 0.75% and 0.77%, and then dropping to 0.73 in 2001. It further dropped to 0.72% by 2002 and was projected to be at 0.68 by 2004 (DoE, 2001). This came at a time of declining student enrolments, putting pressure on institutional operational budgets.

\

Some institutions, it is alleged, introduced efficiency measures while others expanded their income streams, which increased/created competition among institutions. Unfortunately, it is alleged that by 2001 more focus was placed on market-related (profitable) programmes, with little attention to quality and other national goals (DoE, 2001). The National Plan for Higher Education in South Africa (DoE, 2001) indicates that the future of public HEIs is to be channeled through national policies and should not be market-related or exposed to pressures of competition. This bid to obtain equality in some way restrains the autonomy of institutions.

Otherwise it may mean that government has to make enough resources available to avoid the competition said to exist in some universities. Funds are therefore secured for the following important research purposes stated by the White Paper 3 (1997a): to preserve and strengthen existing areas of research excellence, to develop new areas and centres of excellence, to develop research links and collaboration with industry, to encourage inter- and trans-institutional research and to facilitate collaboration with the Science, Engineering and Technology Institutions.

The new (post-apartheid) funding formula according to the White Paper 3 is understood to consider equitable allocations, promote excellence and seen to be widely acceptable within the principle of institutional autonomy and public accountability. The Ministry of Education uses different funding strategies as advised by the CHE. Firstly, goal-oriented strategies were used to improve access for students, to improve the quality of teaching and research and to improve completion rates and responsiveness to social needs. The second is the performance-related funding for performance related activities, as the department is being advised by the CHE.

According to the White Paper 3, the new funding framework (for public HEIs) entails block funding on triennial-rolling basis, while earmarked funds are meant for specific purposes such as research development (as stated above), libraries and information technology. Eighty five percent of the funds are said to be allocated on a subsidy formula while the other 15% is shared for urgent needs such as for transformation activities. Some such funds have been included in the block grants for the HDIs and initiatives for women in research.

Research output from these institutions has been said to be disappointing despite this intervention. This, according to the National Plan for Higher Education (DoE, 2001), may have come as a result of fewer resources (the critical mass) and in some cases the lack of incentives for postgraduate studies as a result of fewer skills or other unstated reasons.

The White Paper 3 advised that expenditure in HE in South Africa was relatively high by international standards. As a result, the new funding formula would be goal-oriented to support teaching, enrolment issues such as equity measures, research development and infrastructure, while the three-year forward estimated budget plans based on market demands and performance would be given as subsidies. The full-time equivalence (FTE) funds for research students include supervision, infrastructure and the consumables in research.

In order to ascertain that research funds are efficiently utilized, that is, not used for cross-subsidization of other programmes as occurred in the late 1980s and early 1990s (SAPSE 110), the blind component of research funding has been separated from the funding formula and redirected as earmarked funding. As was suggested by the White Paper 3 (DoE, 1997a), the redress measures to build capacity in research have mainly been driven by the Department of Science and Technology through the research councils, especially the NRF, to be discussed in the next chapter. The DHET also has its earmarked funds and these are allocated to universities for capacity building and for infrastructure. The earmarked funds, “would require the production and evaluation of competitive bids linked to business plans” which, it is believed, “would be onerous both for institutions and the ministry”.

Earmarked funds would also be assigned for training including that of research. Project costs, according to the White Paper 3 (DoE, 1997a), are the responsibilities of the research councils, for example the NRF, determined on the basis of peer review. Projects may nevertheless also be funded through the earmarked funds. As already stated, the earmarked funds are meant to increase the research capacity at high levels to address the needs of the country and to promote postgraduate enrolment (to increase scholarship) to take on a potential groups of recruits for the academic labour market. This is viewed as a step towards finding solutions for equalizing research strength wherever there is potential.

The CHE advised that research funding be dispensed thinly across a limited number of institutions, with demonstrable research capacity and potential, because of the financial constraints (DoE, 1997a; DoE, 2001). The Department of Education did not support the idea of concentrating resources in comprehensive research and postgraduate training institutions as this would promote elitism and perpetuate disparities between previously advantaged and disadvantaged institutions.

Ironically, the White Paper 3 advised that “the Ministry adopts a policy of concentration and selectivity in the funding of research and research training” (DoE, 1997a: 49). It was also suggested that negotiations for the funding of research students be done on competitive levels based on research capacity and performance of departments. This would be determined through assessment “informed by performance indicators of research capacity, competitive success and output of faculties”. This implies that emphasis would be laid on research strength and “within the funding grid, the Ministry would support and assist the expansion of postgraduate training in those parts of institutions (historically advantaged and historically disadvantaged alike) where there is demonstrable strength”. In this way, earmarked funds will assist in the development of institutional centres for postgraduate training where there is potential for success. “Demonstrable potential for success” sounds farfetched in the HDIs, unless there exist measures that can reliably and objectively determine such potential and evoke it.

In response to the White Paper 3’s advice, there exists a separate component of research in the new funding formula (see table 7.4), to ensure accountability and efficient use of limited resources. The formula is based on output and includes postgraduate publications. According to the Ministerial Statement on Higher Education Funding (2006), publication units and doctoral and masters (research) are the recognizable outputs for the calculation of research grants, based on triennial data acquired from the institutional audits and the Higher Education Management Information System (HEMIS). Non-research masters are calculated as teaching outputs (Ministerial Statement on Higher Education Funding, 2006). Table 7.4 below

indicates the allocation through the block grant, earmarked funds and for institutional restructuring, incrementally over a period of time.

Several ministerial statements have been issued, indicating the planned funding strategies of the Department of Education. Since 2003 several reviews and adjustments have been made to the HE funding framework, from the migration of funds to the three-year-rolling process based on FTEs and then the normative outputs based on the actual output of an institution (Ministerial Statement on Higher Education Funding, 2006). These adjustments continue to occur in a bid for government to effectively and efficiently fulfill demands according to national priorities.

The National Plan for Higher Education in South Africa (DoE, 2001) suggested that research be funded separately based mainly on research and graduate outputs. Whereas in the past research development funds were added to the block grants there is a move that beyond 2009 such allocations will be dealt with separately as research development grants. As a result of the mergers of institutions, different formulae are used for calculating the merged and the non-merged institutions. The Department of Education already opened this matter for scientists to assist on “appropriate additional indicators as part of its consultations” on the new formulae. This makes funding one of the important determining factors for evaluation.

Table 7.4 Allocations for the South African HEIs triennium, 2006/07 to 2008/09 in the new funding framework

Distribution of budget totals for 2006/07 to 2008/09							
	DISTRIBUTION OF ACTUAL BUDGET FOR 2006/07 (R MILLIONS)		PROVISIONAL DISTRIBUTION OF BUDGET		INCREASE ON BUDGET PROVISION FOR PREVIOUS FINANCIAL YEAR		
			2007/08 (R MILLION)	2008/09 (R MILLION)	2006/07	2007/08	2008/09
L. BLOCK GRANTS	9956	85%	84%	84%	8.87%	7.4%	8.37%
1.1 Teaching Inputs	6260	53%	53%	53%	6.9%	8.03%	8.92%
1.2 Institutional Factors	652	6%	6%	6%	6.9%	8.03%	8.92%
1.3 Teaching Outputs	1564	13%	13%	13%	6.90%	8.03%	8.92%
1.4 Research Outputs	1280	11%	11%	11%	6.90%	8.03%	8.921
EARMARKED GRANTS	1.231	10%	12%	12%	13.47%	19.91%	8.24%
INSTITUTIONAL	568	5%	5%	4%	3.27%	5.63%	0.00%

RESTRUCTURING							
TOTAL	11755	100%	100%	100%	9.05%	8.62%	7.96%

Source: Ministerial Statement on HE funding – 2006/7 to 2008/9 - Adjusted to only indicate totals and data relevant for this study

The White Paper on Science and Technology (1996) relates assessment to the implementation of a funding system “in which public funds are handled according to generally accepted accounting practices (introduction of accountability practices) and spent transparently to achieve desired outputs”. There is also a need to introduce a system of co-ordination to reduce fragmentation.

The HEQC quality audits on postgraduate programmes (serving as an important indicators) are meant to enable the targeting of resources (to potential centres) especially because postgraduates are regarded as a “highly qualified group”. Postgraduate scholarships are intended, in order to increase the pool of scholarships while gender and race differences are also addressed to curb inequities. In the medium- to long- term , an increase in graduate enrolment is targeted while in the short- to medium- term graduate outputs must be increased, and small enrolments and dropouts have to be discouraged. If this intention to boost the number of postgraduate researchers happens across all HEIs, there is hope that developments will take the direction of transformation.

In conclusion, as a solution to the quality/equality problem, it appears that the Ministry of Education is expected to protect existing capacity (short-term) in order to continue fulfilling research needs of the country. This should be geared towards the national research agenda (addressing inequalities especially on race and gender) and to meet intentions and demands. The process involves improving on postgraduate training, as already stated (not at the expense of existing capacity). Inter-institutional collaboration (nationally and regionally) is encouraged and together with earmarked funds (made available) is targeted for capacity building and for developing centres in the HDIs for research training.

7.3.7.4 Measurement for research performance

The White Paper 3 (1997a) referred to criteria for research recognition to the CHE and the NRF as advisory bodies to government for any inclusion and/or alteration of such criteria.

According to the White Paper, this is done in the light of international experience the two bodies possess and taking into account the specific conditions of research in the country. It is not clear how the consideration of international experience will take into account the 'specific conditions' nor have the 'specific conditions' been clarified. One may assume that all factors discussed in this chapter are implied. The advisors are said to have responsibility to establish such conditions (within the transformation processes) in line with international experience. The indication of willingness to accept advice 'in the light of international experience' and the fact that the White Paper 3 regarded such action as long overdue provides further motivation for this study and reference made to international models.

As stated, the White Paper 3 (DoE, 1997a) mentions government's commitment to the inclusion of research funding "for the recognition of measurable research output since such recognition has proven incentive value in stimulating high quality research and publication" (DoE, 1997a: 50), an enabling tool to monitor and compare performance criteria. The research output measure had previously been used in the SAPSE formula also because of its measurable ease for funding purposes (see appendix 1).

The White Paper on Science and Technology (1996) on the other hand, suggests the use of peer reviews, stakeholder inputs and efficiency audits in order to develop a formal system of performance measurement. This system according to the White Paper (1996) is linked to the output performance measures criteria and is an attempt to adopt good practice. At its inception, the White Paper confirmed that "a process to arrive at a set of performance criteria needs to be defined which will enable the effective deployment of public resources" (White Paper on Science and Technology, 1996:25).

As for the use of indicators, the White Paper on Science and Technology (1996) refers to output (ex-post) measures as opposed to input (ex-ante) measures as usable indicators for the nature and the contributions made to technology. This responsibility is assigned to policy research, especially for science and technology. The evaluation of research is expected to be measured against international best practices in the study areas. Other types of assessment

and management strategies are output audits, which are used for the allocation and creation of law to protect inventor rights.

The National Plan for Higher Education in South Africa (DoE, 2001) observed and cited limitations of policies used for measuring research output, which would later be addressed through the new “Policies and Procedures for Measuring of Research Output in Public Higher Education Institutions document (DoE, 2003). Unfortunately even in this policy some outputs (for example technical and policy reports) are not recognized. The document will be discussed later in this chapter. The National Plan for Higher Education in South Africa also observed that there is bias in outputs in some disciplines and that there exists a problem of outdated accredited journals. This has also been cited as a problem by Mouton *et al* (2006). The National Plan also raised concern about the lack of response to the development of new knowledge and modes of knowledge production. This may be related to concerns expressed by the Department of Education about the “decline in research output and capacity and the low postgraduate enrolment”. These are viewed as “symptoms of a broader problem that plagues the national research system, that is, its continual fragmentation and lack of co-ordination”(DoE, 2001:18).

7.3.7.5 Conclusion

The National Plan for Higher Education suggested that the introduction of a revised policy on the measurement of research output and making efforts to review the quality of postgraduate programmes would assist in the enhancement of research performance. The two areas (research output and postgraduate programmes) feature prominently and thus seem to serve as the main criteria (focus areas) for any planned evaluation of research. This is more inclined to policies that relate to funding and suits government needs (as it always has been). The tendency to concentrate on output measures neglects non-publishable forms such as instructional and public service research, which is relevant to national needs during this time of transformation. As stated, concentration of researchers on the latter services may be the cause of the decline in basic research. Therefore instructional and public service research, in which academic staff members are involved at one stage or another may also need to be recorded (evaluated).

With the development of a unified system that would fully concentrate on issues of research and co-ordinate them in relation to national imperatives and with action taken upon results of evaluation, improvements may be expected. The system should assist in the improvement of quality and equality in research activities.

The White Paper 3 (DoE, 1997a: 54) claims to recognize “the need for the co-ordination of research activities and their funding in higher education”. If discussions in the next two chapters do not reveal fulfillment of this recognition, the need may arise for the establishment of a system that will directly serve the interests of HE research (similar to the Dutch VSNU and the HEFCs in the UK). Recognition is given to efforts already made by the Department of Science and Technology (through the White Paper on Science and Technology) and the international expertise displayed by the NRF and the CHE. The White Paper 3 (1997a) suggested collaboration between the three and the DoE in the planning of assessment of research in HEIs. The latter two will be discussed in chapters 8 and 9 respectively.

7.4 Policy and Procedures for Measurement of Research Output of Public Higher Education Institutions (DoE, 2003)

7.4.1 Introduction

As indicated above, the National Plan for Higher Education in South Africa (DoE, 2001) cited limitations of policies used (then) for measuring research output. Concerns have also been raised on the problems of criteria for accreditation of journals (Mouton *et al*, 2006). This led to the establishment of this policy, based on Section 3(1) and 3(2) of the Higher Education Act 101 of 1997 and in consultation with the CHE. The policy has been in operation since 1st January 2005 (considering the 2004 output), and was intended to replace the “Information Survey Manuals – research output of the binary system of universities and technikons”. This was done in line with the new system of HE in the country, with its roots in the system of transformation, based on the White Paper 3 (DOE, 1997a) together with the National Plan for Higher Education (DoE, 2001).

7.4.2 Purpose

Based on the documents mentioned above, the policy intended to “sustain current research strength” and to “promote research and other knowledge outputs required to meet national development needs” (DoE, 2003:4). All these would be achieved through the encouragement of research productivity, marked by “rewarding quality research output”, enhancing “productivity by recognizing the major types of research output” and by using “proxies to determine the quality of such research output”. This reward system is not extended to commissioned and/or contract research output but remains with “original systematic investigation(s) undertaken to gain new knowledge and understanding”, self-initiated research. It is obvious that rewarding research relates directly to funding at ex-post levels. It is notable that this type of funding has been separated from the general university grant/block grant. The performance indicators used as “proxies” to determine quality are explained under the sub-heading of quality above (7.3.5.1).

7.4.3 Methodology

7.4.3.1 Mode of Operation

Outputs have to show proof of peer review and this, according to the policy document, is the mechanism used to “ensure” and “enhance” quality. In other words, peer review is the main determinant of quality of research output. Different modes of dissemination of output, including the electronic mode, are considered, provided stipulated criteria are met. Notably, the policy “does not support differentiation within type of inputs”.

7.4.3.2 System of reward

The system provides a subsidy that is based on “affiliation of the author”, qualified by positions of research or academic staff, research students or even visiting scholars. Submissions of authors belonging to the same institution acquire full subsidy while sharing occurs when authors belong to different institutions. Each output is claimed once and claims for the preceding year (n-1) are done under the stated conditions.

7.4.3.3 Types of recognized outputs

- Peer reviewed publications in accredited journals: These include articles, research letters, research papers and research articles in approved journals with an International Standing Serial Number (ISSN). A journal may appear in the ISI (the Science Citation Index, the Social Science Citation Index and the Arts and Humanities Citation Index), the IBSS and South African journals as recognized by the Department of Higher Education and Training (DHET). All journals have to meet the requirements of the DHET which is subject to annual review. In cases where languages are used other than English, abstracts have to be submitted in English. The subsidy for journal articles is a single (1) unit per article.

- Books: The books have to be reviewed by peers of the author and assessment documents have to accompany the book. The documents may be in the form of monographs, chapters and edited work. An evaluation panel of senior professionals is constituted by the DHET to evaluate the books. In cases where the book is written in another language, a page summary has to be written in English. A maximum of five (5) units may be allocated for books.

- Proceedings: Peer reviewed published conference, congress and symposium proceedings meeting specified criteria fall within this category of DHET subsidy. An abstract may be submitted if the proceedings are written in another language. Proceedings are allocated 0,5 points units. An ISBN number is necessary.

The list of approved journals is found on the DHET website. Journals may be included (added to), removed from or even amended on the DHET approved list. In all instances, procedures have to be followed. The DHET may remove any journal that fails to fulfill the expected criteria. Removal has to be done by 15 June of every year. Editors should provide information on any amendment made to a journal. The DHET may assist and support the inclusion of a South African recognized journal to the ISI and the IBSS, following the criteria provided for the purpose.

7.4.3.4 The subsidy process – submission of claims

The following points are stipulated guides for submission of subsidy claims:

- Annually for the n-1 (preceding year)
- Late submissions (n-2) will be considered when signed for by a university authority and with supporting evidence
- Audited claims are to be in by 15 May each reporting year

7.4.4 Conclusion

This peer review programme bases its operations on the award system, which depends fully on research output (ex-post system). The aim of the system is to encourage self-initiated research by providing resources to productive researchers. Although data accrued from the programme may be used to assess the performance of institutions, submissions are not compulsory and therefore not necessarily inclusive of all products. Also observable is the fact that research conducted for instructional improvement and the consultative type of research are not included in the system.

7.5 Summary of the chapter

7.5.1 Historical Background

7.5.1.1 The apartheid era

During this period disparities were caused by the fragmentation of the entire HE system of in which HEIs were reporting to different departments. The different departments had different policies that led to inequitable distribution of resources and thus mismatch of research outputs and differences in postgraduate education. This came as a result of the steering powers of government.

Institutional autonomy related to internal regulations and discretion to utilize allocated subsidies. Therefore, development depended on the planning abilities of institutions with no formal directives given. Accountability was done through audits and annual reports. Sanctions applied against the apartheid regime by the international community reduced basic scientific levels of engagement.

Government's financial constraints reduced allocations to universities through the SAPSE formula. As a result, cross-subsidization of programmes occurred within some institutions, compromising non-prioritized programmes. In the case of research, this would also compromise output. The block grant subsidy was only meant for programmes with "significant public benefits".

Research was based on publishable outputs related to the creation of knowledge (primary research). Instructional research (meant to study issues related to teaching) was funded with instructional funds and artifacts had special funding. Contracts and consultations would be funded by the 'public service' or recipient of the product and were therefore not included in the SAPSE (not necessarily for public benefit).

7.5.1.2 Post-apartheid era

The period around the early 1990s was marked by changes in the ruling party and thus laws and policies with transformation expectations. The recent unitary structure of the HE system was meant to eradicate inequalities, while redress is compared to the achievement of equilibrium across the entire HE system and is a very important transformation imperative.

There has been a need to address the imbalances of race and gender in and through research, which would be done incrementally over a period of time to allow for the maintenance of standards and continuity to provide national needs.

Inequalities would be solved by producing highly skilled and competent graduates and a new generation of intellectuals for research. There was also a need to strengthen weaker universities even though government still maintained the need to concentrate funds selectively on competitive levels. In contrast to this idea, the new funding formula considers equitable allocations while promoting excellence and public accountability. It not very clear how this would be achieved

Investment is concentrated on postgraduate education and the development of national policies, that is, to improve access, to improve the quality of core functions, to improve completion rates and to respond to social needs. Also important is the preservation and strengthening of existing areas of excellence and the development of new areas and centres. While South Africa is said to have a good quality record, incentives have to be increased to encourage scientists to sustain these levels.

Quality may be achieved through benchmarking against the best in the field. Academic and educational standards are to comply with specific requirements in the sense of aims and ideals of excellence depending on specific purpose. Evaluation is therefore to be set on standards in terms of the utilization of available resources (effectiveness and efficiency). Research here is understood as a tool for the creation of knowledge

Assessment is related to the implementation of the funding system, to make sure that public funds are handled according to acceptable practices. Assessment is based on the CHE and the NRF's criteria of research recognition and international expertise, while taking into account local conditions. The Department of Education is the main stakeholder and is responsible subsidizing the system. Table 7.5 summarizes the functions of other stakeholders.

Table 7.5 Governance structures in the new dispensation

Institution	Functions
Government: Main stakeholder	<ul style="list-style-type: none"> -Funding suits government needs (steering powers) -Improve completion rates -Regulate labour intensive strategies -Intervene in institutions in extreme circumstances to restore good order -Regulate public funds -Control programmes and services
CHE	<ul style="list-style-type: none"> -Graduate training -Infrastructure

	-Strategic development
NRF	Redress and capacity of research driven by the NRF
The HE Branch	-Policy formulation -Collect information from institutions (HEMIS) -May pass information to other bodies e.g. the CHE -Participates in HE projects
Institutional autonomy	-Create good governance -Responsible for missions and policies -Responsible for institutional culture -Responsible for performance quality -Responsible for institutional reputation -Take care of transformation and responsiveness to national needs

The system also has to collaborate with industry, work on inter- and trans-institutional research and forge collaborations with science councils. High levels of competition are to be avoided.

7.5.2 Evaluation purpose, measurements, definitions and funding mechanisms

These are summarized in table 7.6 below

7.5.3 Challenges and solutions in the new dispensation

The summary is provided in table 7.7

Table 7.6 Summary of old and new dispensations of research in HE

	Old dispensation	New dispensation
Purpose of research evaluation	For inclusion in the SAPSE formula as submitted through annual reports	Assessment is conducted to enable effective deployment of public resources. Recognition of measurable research output for funding is an incentive to stimulate high quality research. Publications are used as a tool to compare performance.
Measurement criteria	<ul style="list-style-type: none"> -Research output (ex-post measures) -Data from annual reports and audits 	<ul style="list-style-type: none"> -The use of ex-post measures. Publications are easy to measure for funding purposes -Based on self-regulation and for funding formula -Peer review, stakeholder inputs, efficiency audits and output audits -International standards -Increase in research activities (output), publication units -Postgraduate outputs (research graduates) -Measurements done through institutional audits and HEMIS
Defining dimensions	<p>Quality and productivity:</p> <p>Output from peer reviewed journal with reputation based on status of editorial committee,</p>	<p>Quality:</p> <ul style="list-style-type: none"> -Publishing in high impact journals -Testing ideas against international best (conferences) <p>Relevance:</p>

	circulation and status of journal	<ul style="list-style-type: none"> -Fitness of transformation purpose -Research important for equitable development -Research as an important tool for nation building -Research important for sustaining a democratic culture -Relevance for global competitiveness and accumulation of knowledge -Policy setting, new systems, resource allocation and logistic knowledge on regulations -Nurturing intellectual culture and innovation -Dissemination of knowledge
Defining research	<p>Primary research based on output production (publishable output) in relation to the creation of knowledge</p> <ul style="list-style-type: none"> -Research conducted for instructional improvement -Artifacts -Public service research (contracts and consultations) 	A tool for the creation of knowledge in order to meet pressing needs and to respond to new realities
Funding mode	Government subsidy (block-	Selective

	grant) as calculated on SAPSE formula	<ul style="list-style-type: none"> -Block grant (FTE for research students to promote enrolment), for supervision, infrastructure and consumables -Earmarked funds for development, e.g. research capacity (85% subsidy formula, 15% urgent needs, e.g. transformation) *Earmarked funds in the NRF (grants evaluated on competitive basis) *Earmarked for training and for special projects -Research reward system (DoE, 2003) -Merged and non-merged institutions calculated differently

Table 7.7 A summary of challenges and solutions in the new dispensation

Challenges	Solutions
<ul style="list-style-type: none"> -Integration into the global system while addressing local needs (national processes set to operate against global imperatives achievable through necessary expertise) - field of intellectual inquiry -Meeting such obligations (intellectual inquiry of global standards) in a non-equal system -Equalizing HE system at the same time as meeting the demands of a modern economy of global standards (concurrently) 	<ul style="list-style-type: none"> -Recognize and maintain institutions that achieve international standards -The traditional universities value scholarship and play the role of knowledge generation necessary for societal development -A unitary system of HE -Redress of inequalities

<ul style="list-style-type: none"> -Create a mindset for optimal benefits in the unequal environment -Research output from HDIs still disappointing despite intervention -Lessened incentives for postgraduate studies as a result of fewer skills in the HDIs -Nurturing a generation of researchers and intellectuals from HDIs -The top publishing universities also still produce the highest postgraduate output -The process of international competitiveness may lead to a conflict between quality and quantity. -The pressure may lead to high competition and thus further fragmentation (inequalities) -Poor capacity (in planning and copying from others) in the HDIs worsen the problem -Relatively scarce resources for the development of research capacity -Allocation of postgraduate training funds only in faculties with demonstrable strength (no good for HDIs) -Reports not recognized as output -General decline in research output and capacity (may be affected by a rise in applied research) -Methodological bias based on discipline, lack of response to new knowledge 	<ul style="list-style-type: none"> -Mobilize talent -Develop professionals and systems -Integrate research with other systems -Produce vibrant research, observing democratic ethos -Uphold rigorous standards of academic quality to sustain the quality of life -Creation of a research mindset to optimize benefits -Production of knowledge and graduates -Increase opportunities for the weaker institutions -Competition (healthy) within institutions has to be regulated within the national framework -Introduction of funding and quality assurance framework to improve poor capacity in the HDIs -Coherence encountered that may be achieved through continuous collaboration could also assist to equalize the system
--	---

<ul style="list-style-type: none">-Oxbridge/Matthew/Halo effects-Outdated accredited journals (including the award system)-According to the DoE, good scholarship is not the same as research competence (teaching not depending on research – DoE, 1997)	
---	--

7.5.4 The Research reward system (PPMROPHEI) (DoE, 2003)

The system is summarized in Table 7.8 below

Table 7.8 A summary of the research reward system

Introduction:	Funding included in the block grant
Purpose:	Encouragement of research productivity by rewarding quality research output, using proxies to determine quality of self-initiated research (original systemic investigation undertaken to gain new knowledge and understanding)
Mode:	Rewarding quality and enhancing productivity
Methods:	Proof of peer review
Reward system:	Research academic staff Research student Visiting scholars
Types of output:	Articles (1 point): Articles in approved journals, research letters, accredited journals (to fulfill requirements) Books (5 points): Monographs, chapters, edited work Proceedings (0.5 points): Peer reviewed published conference, congress and symposia papers
Panel	Senior professionals constituted by DoE

Chapter 8 The FRD/NRF rating system

8.1 Introduction

The NRF rating system has been included in this study because of its role in research support for HE and its strategies of assessment in providing such support. The rating system is one of the programmes of the NRF; an intermediary body legislated to execute the responsibilities of promoting research and development and facilitating the training of researchers. Both the White Paper on Science and Technology (1996) and the White Paper 3 (DoE, 1997a) refer to the NRF as a body that should guide activities of assessment of research in the country.

The rating system is also of interest to this study for its unique characteristics. Whereas the system applies peer review strategies used elsewhere in the world, the rating of individuals for the purpose of individual support is a unique phenomenon (Auf der Heyde and Mouton, 2007). Although there are some similarities in assessment strategies with such systems as the New Zealand system, the difference is that the latter operates on a large-scale (comprehensive) system. The system has been analyzed in this study to explore its contribution to the assessment of research evaluation in HE. This chapter includes a discussion of the historical background of the system, the purpose of rating, methodological aspects, and the problems experienced and good practices learnt from the system.

8.2 Historical background

According to Pienaar *et al* (2007) and Krige (2007) the history of research and research evaluation in the South African HE system, may be tracked back to the late 1940s, with the establishment of the Council for Scientific and Industrial Research (CSIR). Besides the duty to conduct research, the Council was mandated to support research in universities and museums through funding. From its inception, after World War II, the CSIR is said to have focused more on Operations Research (OR), especially research that supported military activities (Itmann *et al*, 2007).

Focus on applied research in universities became apparent in the period 1960 to 1970 when the CSIR established OR departments in universities and supported them. This establishment is said to have steered universities towards doing more strategic research. The internal and cross-boarder political instabilities of the 1970s and 1980s, that resulted in international isolation of South Africa, economic instability and concentration on military demands all compelled research to be relevant (applied) to the demands of what Krige (2007:9) terms “national self sufficiency”. Krige also views this arrangement as the then government’s way of using research to “bolster the apartheid system”.

International isolation also affected the involvement of researchers in international scientific debate, further lessening interest and participation in self-initiated (own choice) research. Itmann *et al* (2007) explain how the economic instability at the time affected the activities of the CSIR, causing it to shift to a market- oriented (consultancy) direction. This rendered the support of basic research even more difficult as a result of the fact that “the pressure to deliver to clients was not conducive to (such) research work” (Itmann *et al* 2007:3), thus affecting basic research productivity. It is obvious that the problem-solving research approach was preferred, to fulfill the “directed” research needs and obligations.

The influence of OR together with the burden of international isolation, created a situation in which the country started losing highly trained researchers. Therefore, the need arose for the promotion of scholarship, and at the time this could only be possible through the funding of self-initiated research. According to Krige (2007:12), this was associated with “the development and monitoring of skilled high-level researchers.” Krige explains how, in the early 1980s, the CSIR assigned a retired Dean of Science from the University of Cape Town, Professor de Wet, the responsibility to establish a foundation of research support in HE. Professor de Wet would also advise the CSIR on programmes for the funding of “good researchers”. De Wet’s “good research” is said to be determined through peer review evaluation of an individual researcher’s ability (proven ability).

After consulting with other experts locally and studying other models in other countries, Professor de Wet produced several reports proposing a “person-based” support model as opposed to the “programme-based” one. Some of the reports that served as a turnaround strategy for the support of basic research in South Africa include; “A new look at the Foundation of Research Development” and “A framework for the promotion of free research in universities”. By implication, free research results from own initiatives and relates closely to basic research. Krige (2007) explains the de Wet idea as that of awarding researchers based on merit, the full costs of research programmes payable for “selected research workers” while also improving on capital equipment for university research. Professor de Wet identified the CSIR as the funding body for research. The Foundation for Research Development (FRD) was founded in 1984 to oversee the research support.

The CSIR utilized the skills of the retired professor to reach out to the needs of academics (scientists). Although the universities are said to have been consulted in the finalization of the idea, the initiative was that of one scientist and the scientists themselves were not directly involved in the formulation. This somewhat imposed idea resembles the New Zealand and the UK’s government steered models and differs from the Dutch model, which was basically initiated by the universities trying to acquire independence from the state. The reason why the former countries followed the state initiated model is dependency on state grants and this has been also been the case in South Africa.

At the time, the need for “national self sufficiency” still existed. There was therefore a need to develop high-skilled researchers to perform both self-initiated and directed (applied) research in universities to maintain what Pienaar *et al* (2000:28) interpreted as an “investment in university research to develop scientific human resource”. This means that changes in research needs forced the CSIR and its subsidiary (the FRD) to continuously amend their research mandates “from time to time”. Alongside these changes, the evaluation and rating system was introduced and became a useful funding tool, which in fact was primarily meant to promote “curiosity-driven research” (self-initiated research).

It is this research evaluation and rating system that dominated discussions of HE research activities prior to 1999 in South Africa. It was in this era (the de Wet period) that the quality of research and that of researchers became the main focus. Pienaar *et al* (2000:29) cite the important fundamental principles for producing quality research resulting from Professor de Wet's advice, which is "to identify good researchers in whom to invest scarce resources". This advice formed a theoretical base that continued to dominate activities of research support, development and consequently research evaluation despite the changing mandates in the country.

8.3 The introduction of a rating system

For the entire history of the CSIR and its subsidiary, the FRD (and later the NRF), evaluation of research was attached to funding. Professor de Wet (Krige, 2007:14-15) emphasized the awarding of grants to promote "free research" (believed to be an important aspect in research training for excellence) in South African universities and encouraging inter-institutional collaborative research in "areas of national concern".

According to Krige (2007), self-initiated research was valued by Professor de Wet as stimulating and promoting the trademark of "good research". Professor de Wet believed that self-initiated research contributes to diversity, which makes researchers autonomous (Krige, 2007). This autonomy, it is believed, structures planned (then) person-based support (financial) that may lead to "good researchers" participating in research pertinent to national imperatives. In this way, 'relevance' is believed to also benefit from quality research. Accordingly, the de Wet principles were based on the belief that quality researchers produce good research, motivated by challenging research, conducted freely, with funds made available to such researchers. The sustainability of the process was based on the belief that "good researchers" when supported, would provide best training to others (colleagues) and to students, a philosophy that dominated the first comprehensive research assessment in New Zealand. The difference between the two models is in the unit of analysis in that in New Zealand the results of assessment of individual researchers cumulatively make up the performance of an institution.

The outcomes of the rating exercise according to Krige (2007) were (and still are) to obtain information and keep a track record of the research manpower in universities in the country, an exercise that should easily be obtainable through an audit. It is necessary at this stage to remember that participation in the exercise has always been voluntary, making it difficult for the process to reach out to all academics. This does not enable the ease to take stock of researchers in the entire system of HE in South Africa. In a comprehensive exercise such as New Zealand, the voluntary phenomenon works against those institutions where researchers abstain from participating.

As stated, the FRD was founded in 1984 to concentrate on research development and research support. Funds were made available for salaries, research equipment, assistance to postgraduate students and travel (for research). The NRF on the other hand, was established in the late 1990s to unite the efforts of the FRD and the Centre for Social Development (CSD). The FRD's support was "aimed at complementing the investment in research ... and enhancing the development of research excellence" (Pienaar *et al*, 2000:28).

It is important to note that the support system discussed in this section occurred concurrently with the support by the then Department of Education (DoE) for the HEIs. The DoE's support was directed through the block grants, with institutional autonomy to utilize the funds and research support through SAPSE publications (as discussed in the previous chapter). According to Krige (2007:44) the DoE's "publishing funds" (SAPSE publications) did not come about as a result of rigorous evaluation. The system took a different direction in defining the quality of journals and is said to have encouraged quantity of publications with emphasis placed on accredited journals. The slightly different emphasis on quality (rigor of assessment) makes the rating system different from the other systems that existed in the country. Although Auf der Heyde and Mouton (2007) compare the rating system to that of Mexico, its purpose and operations make it unique especially when compared to the other systems referred to in this study.

In the 1990s there was pressure for research councils (including the FRD) to be less reliant on government and to "become a market-driven organization". This obviously

impacted on the funding for what Krige (2007) refers to as ‘own initiated research’. This financial insufficiency (as a result of the deteriorating value of the local currency) became a problem when available support was no longer enough to cover all research needs, the most affected being the reading of papers in international conferences which is important for international linkages. This may have contributed to the delinking of funding from the rating system, leaving ‘prestige’ to be the main component attracting researchers to the system.

8.4 Influences of the new dispensation

In 1995, the FRD was compelled to respond to the national needs (of the new South Africa), in the same sense that it had to respond to the military needs of the late 1970s and early 1980s. Relevant expertise for science and technology and for the fulfillment of economic imperatives became important. This, Krige (2007:61) explained, led to a change in goals for the system. Research in science and technology together with research for industry (directed research) were supported as programmes. As stated, this responsibility was transferred to the NRF with its establishment in 1999 (Pienaar *et al*, 2000)

Although different reviews (assessment of the programme itself) have been conducted on the rating programme, no complete overhaul of the system was done, despite recommendations to that effect. What is obvious is the addition of programme funding as capacity building responsibilities that came along with the new mandates. Meanwhile, the idea of “good researchers” performing “good research” overshadowed the idea of “own-initiative”. Both the FRD and the NRF embraced the inclusion of social responsibility research within their own operations as research councils.

Whether or not this was a correct decision to make is a matter for concern (Pienaar *et al*, 2000), considering the objectives and impact of the evaluation and rating system on research in general. The rating system did not change the unit of analysis for rating although funding would be given to both individuals and programmes. The importance

of relevance marked a turning point and a base for redefining the relationship between rating and funding.

By the mid-1990s, the evaluation and rating of researchers strengthened the importance of research in the system of HE and even though it was partly dissociated from funding, the process was used to establish the presence or absence of good researchers according to peer review standards in institutions (Pienaar *et al*, 2000). The authors claim that the process had also been used to establish the research strength of participating institutions and re-introduce/keep HE on the international map (through the peer review standards). To encourage participation of staff members in all institutions, including the historically disadvantaged institutions (HDIs), the system was improved with intentions to identify young and other potential researchers (that have not been exposed to research) and to develop their track records. This type of a system, one envisages, rests heavily on the knowledge and interest of the management of such institutions. This is one stage at which the system would encourage involvement and participation of all institutions, irrespective of historical background, an important transformation imperative.

There were therefore improvements in the procedures of evaluation and rating to accommodate all academic levels (although not interdisciplinary research). Whereas at the initial levels researchers were classified into three categories of ratings the categories were altered from time to time. This was not without problems though. Through the expansion of rating categories, for example the introduction of the “late entry” (L), the system unfortunately took it for granted that graduate training that led to research excellence (or the buildup of potential) was equally sufficient in all institutions, including the HDIs where less research was conducted. Under these conditions, development programmes (of late entries) would be highly affected. Also, the system had inherent disadvantages, for example “good researchers” would be too occupied in maintaining their status to be able to allocate enough time to participate in such development programmes prior to the next cycle of evaluation and rating. The system obviously needed their assistance for the development of late entries.

As stated, the FRD resorted to associating “good research” with support by making rates a prerequisite for funding. By the post-apartheid period attention was given to the support of research proposals. The evaluation of proposals nevertheless continued to recognize the rating prestige and combined both the criteria of “good researchers” and those of relevance (in the evaluation of proposals). That is, in the evaluation of proposals, a high percentage (50%) of the points was allocated to researchers with good research standing (Krige, 2007) although the component of relevance of research was also given high priority. The implication of this is that rating would have to be conducted separately prior to the submission of proposals, which also considered relevance of such proposals to national needs (especially innovation – after the introduction of the White Paper on Science and Technology). At this point, the question is whether the *status quo* of development and research council operations (rating) provide the inclusive results of evaluation of research in HE. The NRF is playing both the roles of grading (through the rating system) and financial support, the latter being handled similar to research councils.

In this way, the NRF system is torn between upholding the goals of the rating system and fulfilling research obligations (needs) in the country. In addition, the objectives of research development have to be satisfied. A compromise was reached through the introduction of three systems of funding to accommodate individual rated researchers, projects and group funding (Krige, 2007). This process introduced fragmentation and altered the use of individual researchers as units of analysis (for example quality of a proposal became a category for allocating funds). It is at this point that the rating system lost impetus with time, with the rates losing influence on funding. Krige (2007) explains how proposal-based research surpassed the use of rates in the determination of support and how the rating system lost its initial priority level in the early 2000s.

The rating process continued despite the reduction of focus on individual researchers, especially the new entries, and continued to nurture established researchers and those with demonstrated potential to be established. To this effect, the evaluation and rating process was introduced to the social sciences and humanities in 2003 and this assisted in increasing numbers of those that volunteered to be rated. By 2005 evaluation and rating

remained an evaluation strategy for the NRF to “focus on track records of individual researchers rather than on individuals as members of research teams” (Krige, 2007:85).

The criterion of a track-record or demonstrable potential continued to serve as insurance for investors and even though delinked from funding, contributed to the determination of period and amount of a programme grant. The programme became the focus for funding, with individual researchers’ performance in the background. As stated, this is a system that is typical of research council evaluation. With the separation of funding from rating, the other advantage other than prestige remained the eligibility of rated researchers to receive NRF grants for an extended period.

The main strategy to assure quality of research remained the peer review system of research output, as determined by the predecessor (FRD). The NRF system, as will be discussed later in the methodology section, allowed for the nomination of peers, excluding those that have been recommended for exclusion by researchers, included national and international experts and considered performance over a period of time (five years of research output and post graduate training).

8.5 The purpose attached to the rating of researchers

Auf der Heyde and Mouton (2007) list the reasons that lead to the establishment of the rating system as the overwhelming strategic research that related to the military and energy projects at the time, which derived expertise from researchers in the universities. Concentration on strategic research is said to have affected participation in basic research, which is self-directed and is believed to be necessary for investment in scientific research (Pienaar *et al*, 2000).

As in New Zealand, South Africa is said to have experienced brain drain. Only in South Africa, this is said to have been aggravated by the isolation of the country due to apartheid policies.

In order to promote self-directed research, there was a need to encourage scholarship and retain research engagement. Auf der Heyde and Mouton (2007) refer to the recognition and rewarding of individual scientists as a way of strengthening basic science. The procedure to classify scientists into different levels of performance follows the assessment of past research performance and has been attached to monetary rewards. This process, according to Pienaar *et al* (2000:29), was invented in order to “improve the quality of research and research training” through the identification of “good researchers” that would produce “good research” of international standard and provide good research training. The belief behind the arrangement is that good researchers would continue to be active and productive when provided with support, therefore serving as some form of investment. The idea of “good researchers” producing “good research” and research training was one of the main objectives behind the initiation of the comprehensive assessment in New Zealand.

The assessment/rating of individual scholars on the basis of past performance and the awarding of rates make the system different from research councils’ strategic funding. The process of research evaluation and rating is discussed below, under the section on methodology for the system.

8.6 Methodology for the evaluation and rating system

8.6.1 Introduction

The concentration of the early 1980s on “proven ability” marks the beginning of the importance of ex-post criteria in the evaluation of research in South Africa. The model was replicated from the US and Canadian models that were preferred to those observed in Europe and the UK by the Professor de Wet group (Krige, 2007). At the time, international recognition became an indicator of good performance and the exercise was structured to include international reviewers with research expertise standing.

The idea of “good researchers” perpetuated the use of researchers as units of analysis, an arrangement preferred in the New Zealand large-scale (comprehensive) evaluation

exercise. From inception, the system created rates (used for rating) in terms of categories that differentiate awards according to researchers' levels of performance.

8.6.2 The rating system.

In the initial stages, rating was classified under three categories; the A, the B and the C rates (Auf der Heyde and Mouton, 2007). Category A would be associated with very high international standards and be recognized in the international community. How one proves that one is internationally recognized is a concern raised by Cherry (2008) and Auf der Heyde and Mouton (2007). Professor de Wet's category B, according to Krige (2007), distinguishes researchers with the ability and capacity to build team research and to support research training while also making contributions to the productivity of research. Category C researchers would have to attract project funding and strive to achieve the other two categories. Therefore, category C would at the time encompass what were later classified as "late entry" and "young Scholars".

In terms of support in the original state, categories A and B would receive support on free choice research for the period of five years, renewable. Category C would only receive basic research support (studentship and short-term assistance). Category C was also meant to provide opportunities for development. In order to sustain the performance of researchers, support has been subjected to review (Krige, 2007:18) and "if a researcher had not lived up to what they had promised"; support would either be reduced or discontinued. This means that a lapse could threaten the chances of sustainability of performance and thus of the system. Table 8.1 reveals the rating classifications as the system evolved and also explains expectations according to the categories.

It is upon these categories that the NRF's system is now based (after several evolutionary processes). The hierarchical funding (support) related to the categories was the origin of 'elitist' classifications in the evaluation and rating system. As the process evolved, more categories were included, for example, the "late entry" (L) was introduced. Later on the "young scholars" (Y) and the "presidential" (P) came to the fore, together with the categories D and E. As time went on, as a result of scarcity of funds, some of the categories that did not show any relation to the capacity of potential were dropped.

Table 8.1 Classification of the rating categories (obtained from Auf Der Heyde and Mouton 2007)

Rates	Qualifications
A	Researchers who without any doubt are accepted by the international community as being among the leaders in their fields.
B	Researchers not in category A but who nonetheless enjoy considerable international recognition as independent researchers of high quality.
C	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 not very high but who are actively engaged in scholastic activities.
D	Researchers who show promise of achieving the status of proven researchers in the foreseeable future, in terms of already existing track records, or proven researchers with established track records but currently less productive.
E	Candidates who, according to recent track records, are not likely to prove themselves or re-establish themselves as researchers.
Y	Young researchers normally less than 35 years of age who are highly likely to achieve C status by the end of the support period.
P	Researchers younger than 35 who have already obtained PhD degrees and who have shown exceptional potential as researchers.
Z	Candidates whose submissions for evaluation are regarded as premature
X	Candidates who cannot be evaluated as researchers

The assignment of ratings is based on recent research (output over five years). Evaluation was (and is still is) based on peer review of reports conducted by teams of specialists in the areas and as stated, with the inclusion of international experts, and site visits in the case of highly rated researchers. This is not surprising since the track records of good researchers and support thereof were the main intentions of the exercise.

Good research would be determined on merit through publication in reputable journals (as a measure of quality) and the training of graduate students. Applicants submit publication lists, a list of research active peers, a self-assessment statement, a list of

postgraduate students trained and participation in highly rated conferences nationally and internationally (Krige, 2007). Krige adds that as conditions improved, technical reports and patents were introduced. It should be noted that in the initial stages a rate would be afforded on the basis of the applicant's submissions on an annual basis.

In the early 1990s the evaluation and rating system seems to have been influenced by research evaluation in other countries, or as Krige (2007) suggests, influenced by the Hawkins Report, the recommendations of which may have been influenced by other countries. For example, the number of publications was limited to four (best recent).

Pienaar *et al* (2000:28) also highlight the assumption that “established researchers are mostly responsible and will do relevant research”. This allowed researchers to maintain their status funds, which they used at their own discretion. This thinking, according to Pienaar *et al* (2000), was maintained throughout the entire era (despite secular influences by the research council, the FRD). Efficiency of spending would be related to the effectiveness outcomes in the following five-year period of evaluation and rating, and this, according to Pienaar *et al* (2000), would be used to determine further funding (based on performance)

The FRD responded to the financial pressure (from increased demands) of the mid-1990s and the demands of the new system (new imperatives) and delinked funding from rating. At this level, a system of evaluation of projects (proposals) dominated the allocation of funds. As stated, although rating still continued, rates were merely used to credit a proposal submitted for evaluation. The delinking stimulated debates around the existence of the rating system and according to Cherry (2008:2) this was accompanied by a “significant increase in the number of researchers letting their rating lapse.” In 2007, the NRF announced the re-instatement of funding according to rates, despite the controversy attached to the system.

8.6.3 The assessment procedure

In the rating system, researchers voluntarily submit their documents, which include: a list of peer reviewed publications, any other form of publications, conference proceedings, reports, patents (if any) and a list of students supervised, all over the past five years. Also included is a self-assessment report, which should provide an impression of involvement in multi-authored work and any other form of involvement that cannot be disclosed through the above-named list. Researchers are also expected to identify up to 10 experts (peers) of their fields of study. According to Pienaar *et al* (2000), the assessment panel (made up of three to six experts) then have a responsibility to finalize the list of reviewers, three from the list submitted by the candidate and three independent ones. The panel takes into consideration the inclusion of international experts, as previously suggested by Professor de Wet (Krige, 2007).

Once the peer reviewers submit reports they are reviewed by both the discipline-oriented assessment panels and independent assessors. The two teams independently assign a rate. Other levels such as the executive evaluation committee and the appeals committee come in when there is dissatisfaction. Although the process has been blamed for being non-transparent, reporting back to the candidates and allowing for appeals strengthen the credibility of the process. The executive evaluation committee is also said to ratify rates. Site visits/inspection to the A and B rated researchers' place of operation gives an added layer to assessment.

8.7 A critical look at the system

The de Wet era marks the period of a more rigorous planning for research development through funding. Similar to all other programmes related to research development and evaluation, the system was (is) not without problems. Concentration was placed on 10 universities, referred to by Krige (2007:14) as 'the ten universities most active in research', which were later referred to as the historically advantaged institutions (HAIs). These institutions were continuously consulted and their suggestions considered in the finalization of important policy documents/reports. In other words, the HAIs have been somewhat involved. This denial of an opportunity for the historically disadvantaged

institutions (HDIs) to participate in the process of developing and establishing “good research” created the skills divide between the two categories of universities (HAI and HDIs), which the present government is struggling to bridge. When this is added to the weak support that was provided to the HDIs by their former departments of education (as explained in the previous chapter), disparities become even broader, making bridging the gap even more difficult.

The dominant involvement of the CSIR and its successors even on HE matters relates to what Waghid *et al* (2005) refer to as “power-over” (the HEIs) and the steering phenomenon and not a collaborative “power-with” related practice. Although Lange and Luescher (2003) claim minimal involvement of the state in the inclusion of intermediate organs such as the NRF and the CHE, the fact that the responsibilities of research support and evaluation thereof were (are) left in the hands of state organs with indirect but firm control over institutional self-regulation is a matter of great concern. The CHE, the FRD and later the NRF cannot be divorced from state operations and their functions cannot be regarded as “out of the hands of the state”. Therefore, despite these consultations, the status of the CSIR and its subsidiaries remained that of research councils, or what the 2005 NRF review report (NRF, 2005) refers to as “a national research agency”, in which participation was (is) voluntary.

As stated, it had already become clear that the system that demanded high quality publications and the production of postgraduate students did not formally, directly include the HDIs. Therefore the “emerging universities” (Krige, 2007:49) and the HDIs would probably not easily benefit from the evaluation and rating system that was meant to improve research participation and capacity for good research, unless such institutions appointed ‘reputable’ researchers . This was exacerbated by competitive nature of the rating system making it difficult for the few good researchers in such institution to have time to assist others (new entries). The system encouraged the best researchers to retain their high-rated status, and was “not sufficiently responsive to grassroots’ research” and to “changing research needs and directions” (Krige, 2007:55). It is not surprising that

even in 2004, top research universities still exhibit the same characteristics (Claassen, 2006) despite the HE transformation efforts of equalizing universities.

Some of the common problems encountered were those related to the researcher using a local language to write articles. The same experience is cited in The Netherlands, where Dutch-only publications require that reviewers be fluent in the language. Also problematic is the concentration on the number of submissions as opposed to the quality of the journal standards (although care was taken, in the acceptance of ‘reputable journals’). Waghid and le Grange (2003) view the methodology of the system of evaluation and rating as “elitist and subjective”, making it vulnerable to methodological effects such as the halo, Matthew and Oxbridge effects. It should nevertheless be remembered that Professor de Wet’s idea accommodated all of these effects in order to encourage those that were prominent to remain in the high echelons.

Although other problems existed that involved the absence of consensus on the allocation of rates in the evaluation and rating system, in some instances, the system also had its own advantages. Besides the claim made on the increase in good researchers, universities used results for different reasons such as profiling and placement in the hierarchy of achievers (researchers in good standing). The system was also used for placement of staff, appointments and for attracting external funding (Pienaar, *et al*, 2000; Krige, 2007).

8.8 Advantages and Challenges

Some of the advantages and challenges have already been referred to in the previous discussions (under 8.7). The rating system had advantages such as encouraging participation in quality research (especially self-initiated) by providing some research incentives and therefore building capacity. The system has been useful in the determination of the *status quo* of research strength in departments, universities and international circles and also placed the country on the international map through the use of peer reviews (Pienaar *et al*, 2000). The Hawkins report (Krige, 2007) claims an

increase in scientific publications especially for individuals supported by the FRD, an increase in the output of postgraduate students and growth in scientific papers read in conferences. In the current situation, one would think of other alternatives such as the then DoE's research rewarding policy (DoE, 2003) and the other NRF support programmes that would also make contributions.

The process may not have succeeded in eliciting some potential nor did it fully persuade all academic staff members to be involved in research or even the evaluation and rating system itself. Other challenges experienced are more related to implementation, for example, the diversity between disciplines (including conceptual and methodological differences) and questionable reviewers' reports (inappropriate reviews). Krige (2007) also alludes to the doubts raised about the credibility and value of the system, especially in fields other than the natural sciences.

In addition to all criticisms cited, as stated, the system continued to promote the halo, the Oxbridge and the Matthew effects and is therefore regarded as being "divisive" and elitist by others. This observation may imply that the process of evaluation and rating does not focus on group research capacity building and even when programme proposals were funded, there is no guarantee that the programmes would give attention to capacity building. This is because the acquisition and maintenance of international standards does not allow time for "good researchers" to concentrate on capacity building. There is also no emphasis on attention to national imperatives, which are in line with the NRF mission. Also expected are the 'after-effects' of rating mentioned by Krige (2007). Despite the increase in participation alluded to by Pienaar *et al* (2000), Krige (2007) warns that the controversial system started to lose participants because of its voluntary nature and reduced incentives. This is confirmed by Cherry (2008). Questions were also raised on the use of international standards as a measure of quality.

An important concern raised by Krige (2007) is the absence of a total review (interrogation) of the entire system even after the assessments of the system were conducted, change recommended and the mandate of the NRF altered. With the new

policies and revised national imperatives, substantial changes would be expected. Instead of completely revamping the system, the NRF responded by making alterations to the implementation process (number of categories, awards categories, selection of peer reviewers, treatment of peer reviewers' reports and terminology) and the direction researchers would take (involvement in group research).

Cherry (2008) views the neglect or reluctance to respond to review results by the NRF as some form of denial and a deliberate pursuit of an unjust cause by only responding to technical and procedural recommendations. This is despite the stated fact that some researchers allow their rates to lapse. Although other possible causes of the latter behaviour have not been investigated in this study, one feels tempted to agree with Cherry's (2008) "Hesa" remarks that perhaps managers and researchers have different views about the system. While researchers have some reservations, managers look at the advantages derived from the existence of the system.

Cherry's explanation in 2008 revealed that "barely 11% of scholars in South Africa's higher education sector are rated", which indicates the lack of enthusiasm and confidence by others, reducing credibility of the system.

The transformation of the NRF in early 2000 that delinked rating from funding may have been a plan to restructure, but this seemingly did not make any strategic changes to the system itself, as the rates would still serve as criteria in awarding funding for proposals. It would be interesting to establish whether or not the re-linking of individual support introduced in 2008 would improve the situation and reduce the pessimism displayed by those such as Cherry (2008). When this happens, the system will obviously only benefit those that have been advantaged.

The NRF strategic plan of 2001 was seen to be "asserting the advantages and validity of the system" (Krieger, 2007:89) and allowed the system to continue independently. This is

guided by previous experiences and the belief that high rates will position HEIs in the international research arena (idea of the UK and New Zealand).

This prestigious system should perhaps be assigned to serve a similar purpose as the Royal Academy of Arts and Science / Koninklijke Nederlandse Academie der Wetenschappen (KNAW) of the Dutch system of research evaluation to be able to retain its objectives. This would allow the establishment of a system that would oversee the entire system of research evaluation and would incorporate both research development requirements, and the research objectives relating to national imperatives and priorities. For this to succeed, the programme would have to be owned by the academic fraternity. Also, problems of inequalities of the HE system have to be addressed for the evaluation and rating process to become a priority focus of research evaluation in South Africa.

Despite the reviews conducted on the FRD/NRF evaluation and rating system and recommendations thereof, Krige (2007) places emphasis on the need for the evaluation of the impact of the rating system.

8.9 Observations, implications and conclusions

The FRD/NRF mandates oscillated between two poles, the de Wet model, in which support for research was to be earned, and the one in which support was aligned to national priorities. This oscillation has been a response to changes and trends.

The de Wet pole, wherein best performance of individual researchers was (is) important is based on ex-post information (research output) and uses researchers as units of analysis. This is guided by the belief that “best researchers” will continue to perform “best research” and yield best results. When this did not give attention to national priorities, the system of evaluation and rating was delinked from funding. The general research development and national priorities model on the other hand, relies on the quality of the proposal, which has ex-ante effects and is judged through the relevance of goals and the intended research contributions towards the redress of inequalities. The later uses proposals as units of analysis. Had de Wet idea been encouraged across all

universities, with some level of compulsion added to it, a situation similar to that of The Netherlands might have resulted. That is, a national developmental strategy would have applied across the university system, which would have geared researchers towards participating effectively even to fulfill the new demands (imperatives).

Time and time again, in an attempt to strike a compromise, a continuum would be created for operations to slide in between the two poles, and in some cases the two would be merged. This would occur with the understanding that a good proposal needs good researchers to conduct the research, and in this way rates would be considered. In this way, both the ex-ante and the ex post factors would be taken into consideration. Pienaar *et al* (2000) cite production of quality output, knowledge production and technology transfer as the determinants of research development. Therefore, funds should be made available for good researchers to undertake research that fulfills national priorities. In line with this idea, Pienaar *et al* (2000) draw attention to the contextual determination of quality research and take advice from Gibbons *et al* (Pienaar *et al*, 2000) that research evaluation should consider social responsibilities.

At its inception the NRF was particularly mandated to support and develop research through funding, human resources, the provision of research facilities and consciously or subconsciously maintain/improve international standing. Even when funding was disabled, the NRF pursued the evaluation and rating exercise for international reasons and used an extended period of funding to motivate researchers to stay within the programme. Consequently (at that point in time), the elite behaviour detached itself from social responsibilities and left it to those that evaluate proposals. The effects of the recent re-linking still have to be established.

With the NRF operating at the level of a research council, when assessing proposals and mainly basing rating activities on volunteer researchers, a more comprehensive and encompassing evaluation exercise is necessary for national research development. This should consider the problem of individual *versus* group outputs raised by Pienaar *et al* (2000:35) when they referred to the NRF evaluation and rating system. The authors stated

that “the evaluation of the outputs of an individual as opposed to the output of a temporary research group may be problematic”.

Besides the contribution to social responsibility, Pienaar *et al* (2000:35) advise that the evaluation of such a system should “serve all science cultures and take account of such factors as multi-disciplinary, diverse research outputs and teamwork”. For the ‘investment in human resources’, which is necessary for global economy participation, to contribute to the reduction of inequalities, there is a need to ‘exploit’ potential across all universities. Whatever other effort made by the NRF’s rating system, research evaluation activities can make a contribution towards development.

8.10 Summary

This section summarizes the chapter. First the historical background is summarized in detail, then the objectives. Methodology is then followed by benefits and challenges.

8.10.1 Historical background

South African university research has its origins in the 1920s and 1930s. As in other countries, more institutionalized forms of research received a major boost after the Second World War. In 1960 research, which mainly concentrated on weapons at the time, was introduced to universities with the necessary support. This introduced universities to strategic research and efforts towards self-initiated research deteriorated. Other factors that contributed to the deterioration are:

- The international isolation of the country, leading to national insufficiency
- The need for internal national self-sufficiency to boost apartheid activities
- International isolation, barring researchers from international debates
- A weaker economy that led to market-orientated research to deliver to clients

There was a need to improve the quality of research conducted in universities and this would be realized through the introduction of self-initiated research. There was a belief that good researchers, when supported, produce good research and provide best training

for others. The idea to encourage more self-initiated research came as a result of the following problems encountered at the time:

- The country was losing high-skilled researchers to other countries
- The need to promote scholarship
- The need to build more research capacity

The CSIR commissioned Professor de Wet to work out a strategy to promote and reward high quality research. De Wet introduced the rating system, which entails:

- Identifying good researchers and investing in them
- Person-based support as opposed to project-based (autonomy)
- Free research of own initiation (curiosity-driven)
- Awarding research, based on merit (in state grants)
- Volunteer participation

To implement the model and provide support for it the FRD was established as a separate unit within the CSIR in 1984. The Foundation provided support by making funds available for:

- Salaries
- Research equipments
- Research assistance
- Post-graduate travel for research

In the late 1980s, and early 1990s the rating system was affected by the country's financial insufficiencies. This led to the delinking of rates from funds and the major remaining award was prestige. In the late 1990s, the new government effected some changes to the programme by adding relevance to quality through:

- Introducing the evaluation of project proposals for funding

- Using ratings to determine a good proposal and for the allocation of period and amount of programme grant

In 2007 (implementation in 2008), funding was once more re-linked to the rating system without any serious alterations to the programme after that a few evaluation programmes had made some recommendations.

Table 8.2 A summary of the purpose/objectives, methods, benefits and challenges of the rating system

Purpose	Methods	Benefits	Challenges
<p>-To improve the quality of research and research training by identifying good researchers who produce own-initiated research and supporting them (building a track-record of good researchers)</p> <p>-“Good researchers when supported provide best training for others”</p>	<p>-Voluntary participation</p> <p>-Unit of analysis is individual researcher</p> <p>-Use of ex-post data for rating (later ex-ante information used for funding of proposals)</p> <p>-Peer review assessment on researcher’s proven ability</p> <p>-Good research is determined by publishing in reputable journals (quality)</p> <p>-Mode: Document analysis and site visits for A rated</p> <p>-Rating submissions: reviewed publications, other publications, conference proceedings, research reports, patents, students supervised, multi-authored work, any other related work, list of</p>	<p>-Encourages those that are prominent to stay prominent</p> <p>-Strengthens the importance of research</p> <p>-Build capacity and strengthen departmental and university research</p> <p>-Encourage participation in quality research</p> <p>-General increase in the maintenance of quality research</p> <p>-Help establish the presence or absence of good research</p> <p>-Help establish research strength of participating institutions</p> <p>-Increase scientific publications</p>	<p>-Disparities between HAIs and HDIs not considered</p> <p>-Only HAIs initially involved and consulted, causing the disparities</p> <p>-Competitiveness does not provide time for capacity building of new entries</p> <p>-Unit of analysis does not encourage departmental research</p> <p>-Voluntary assessment does not work for the entire system of HE as a high percentage of highly skilled researchers are not rated</p> <p>Not persuasive to all academic staff</p> <p>Does not consider group research capacity</p> <p>-Elitist phenomenon with high expectations not motivating new entries</p> <p>-Government steering (power-over) not favourable for scholarship</p> <p>-Diversity between disciplines including their methodologies not favourable to rating</p>

	<p>recommended peers</p> <ul style="list-style-type: none"> -Panels: discipline oriented, independent assessors, executive and appeal committees -Rates allocated into categories with high ones related to international performance, others to capacity building, productivity and received allocations for project funding (other new ones were withdrawn) -The high rates have been associated with prestige and NRF funding until the late 1990s when funding was attached to proposals (50% goes to rated researchers) 	<ul style="list-style-type: none"> -Growth in scientific conference papers -Keeps HE on the international map -Encourages involvement of all institutions -Help identify research potential (even at HDIs) -Overall improvement of post-graduate training -The use of peer reviews is still trusted in international circles. -Good rates are used to attract external funding -Rates are also used for staff placements 	<ul style="list-style-type: none"> -Delinking caused researchers to allow rates to lapse -System not responsive to grass-roots research or changing needs -Does not pay attention to national imperatives -Inability of NRF to review the entire system even with recommendations Proposal based funding: -Incentives related to rating reduced -Assessment of proposals affected the interest for rating -Proposal assessment resemble research council activities and therefore not necessarily good for system evaluation -Proposal funding not in line with rating objectives Methodological: -Problems of local language, halo, Oxbridge and Matthew effects, concentration on total publication and not quality -Subjectivity of reviewers
--	---	--	--

			-Absence of consensus and questionable reviewer credibility Some researchers query international standards as a measure
--	--	--	--

Chapter 9 The Council for Higher Education – Evaluation Activities

9.1 Introduction

The Council for Higher Education (CHE) in South Africa has as its main executive functions to ensure the quality of the provision of services in the HEIs and to advise the Minister of Education. The Council has been assigned this role by the White Paper 3 (DoE, 1997a). The Higher Education Act 101 (DoE, 1997b) has, through its legislative powers, implemented the recommendations of the White Paper 3 in providing for the CHE and the establishment of its committee, the Higher Education Quality Committee (HEQC).

Although the Ministry of Education hoped to ensure quality activities in HE through the CHE, the Ministry is responsible for the establishment of regulations to give effect to quality promotion and quality assurance in terms of the Act. This places the CHE in an intermediary position. Other responsibilities of the Council are to monitor and evaluate the realization of goals and objectives in HE, to promote development in HE (encouraging activities such as publications and conferences) and to consult with stakeholders (CHE, 2001).

According to Lockett (2007:1) “quality assurance in South Africa has been conceived as a means of furthering the state’s ‘transformation agenda’ for higher education”. Responsiveness to such societal needs may be addressed through research and innovation. According to Jonathan (2006), knowledge production and human capital towards which research contribute, are the “major factors” contributing towards growth, which itself is a strong basis for transformation processes. This requires engagement with local and international communities in which quality and standards are maintained.

In its commitment to fund research, the White Paper 3 (DoE, 1997a: 50) purported to stimulate high quality research and publications through the use of measurable research outputs as incentives. To fulfill this obligation, the intermediary bodies (CHE and NRF)

have been assigned advisory responsibilities in the establishment and development of the research assessment criteria.

As a way of taking stock and establishing the relevance of the CHE together with its committee, the HEQC for this study, it was necessary to establish the role played by these bodies in the evaluation of research in the HE system.

According to its Founding Document, the CHE was established to ensure high standards of education provision that will assist citizens to acquire the high level competencies and skills needed for the development of the country (CHE, 2001). Such skills need high quality education and training and a well resourced HE system. Taking into account the uneven quality assurance landscape in the HEIs, it is hoped that the system will contribute to the equalization of the entire system. In this way, the Founding Document places the operations of the CHE within the framework of the South African HE system. Within such a framework a quality assurance system is meant to aid the transformation and restructuring of the HE system. One other important consideration stated in the document is the internationalization of HE for easy mobility of graduates. Once more, this may be achieved by addressing the poor, uneven capacity within the “diverse social purpose ascribed to higher education in the current South African context” (CHE, 2001:2).

Since inception, the CHE and its committee the HEQC, have produced several guiding documents defining the roles and responsibilities of the CHE. The first document, the Founding Document discusses the origins, background and operations of the Council. Other important documents for this study are the Framework for Monitoring and Evaluation of Higher Education, the Framework for Institutional Audit and A Good Practice Guide for Quality Management of Research. All these documents include reference to some form of assessment strategies that are somewhat related to the expectations of the White Paper 3.

9.2 Background

The HEQC's mandate, according to the Founding Document, is to promote and audit quality assurance mechanisms in HEIs in compliance with the South African Qualification Authority (SAQA) policies. This means observing quality "as a critical principle for the restructuring of higher education" (CHE, 2001:3). SAQA registers bodies that monitor and audit achievement in line with established standards for the area of study. The Founding Document explains that such steps are taken to ensure that the standards are of international standard, which is also referred to by the White Paper 3. The provision of services and programmes are monitored and evaluated by the CHE to ensure effectiveness and efficiency of delivery, in line with the core functions of HEIs. Aligned to the pursuit of quality are the principles of equity, innovation and diversity, all of which form part of the process of transformation and are an integral part of the core-function service delivery on which the scope of evaluation may be based. The Founding Document explains that the CHE's mandate is quality promotion and while the principle covers more ground, there is no guarantee that evaluation thereof is inclusive of all core-function activities.

Of importance to this study is the statement from the White Paper 3 that "applying the principle of quality entails evaluating services and products against a set standard with a view to improvement, renewal or progress". With research being one of the components/core functionaries of HE, it is expected that the evaluation or audit of HEIs will have a bearing on research activities. Other functions of the HEQC are promoting quality, auditing HEIs activities and accreditation of programmes. Provision of services is based on a framework containing key areas such as quality goals, increased access, equity opportunities, responsiveness to local, regional and national needs, improved efficiency, an increase in the pool of basic and applied knowledge and social application.

According to the CHE Founding Document, the HEQC consults with HEIs in formulating criteria and procedures and in establishing a formative notion of quality assurance focused on improvement and development rather than punitive sanctions (CHE, 2001). This, it is indicated, is founded on the CHE's frame of reference, which is based on a

mixture of institutional self-evaluation and external independent evaluation, a framework similar to that is used by the Dutch VSNU.

One way to fulfill the plan is to encourage and orientate institutions to manage their own monitoring and evaluation activities in future. For example, the CHE believes that “self-accreditation is one of the HEQC’s key strategies for facilitating the move of the HE system towards a greater measure of quality assurance self-regulation”. This approach to quality assurance, according to the Framework for Institutional Audit document, “seeks to encourage institutions to contribute actively to the enrichment of the higher education sector and the broad society within which it operates.” (CHE, 2004b:4-5). That is, institutions are expected to enrich and add excellence to the HE system and society, produce a vibrant intellectual culture, and contribute to new ideas for innovation and promote quality.

These ideas form the framework within which the quality of universities’ functions are based. Any form of assessment operating within such a framework will not only be operating within a guided system but will also assist in the implementation of the idea. Guiding internal regulations, which may also impact on activities of assessment, is the fact that “the HEQC will uphold the accountability requirements of higher education provision within the two contexts of strong developmental and formative approach to quality” but continue to monitor provision of quality services. Also, the HEQC is said to “strive to complement and enhance the internal quality development initiatives of providers in order to encourage and support their search for continuous performance achievement” (CHE, 2001).

9.3 Conceptualizing quality

The autonomy of HEIs in South Africa is a *sine qua non* and to observe this, the CHE intends weighing quality provisions against the goals of the institutions. To this effect, a framework of accountability is said to be suitable, in which the active promotion of quality in the early stages of the HEQC’s operations will form the basis for developing the appropriate benchmarks for accountability. According to the CHE Founding

Document, accountability can be achieved through the setting of standards. In HE, this accountability is intended to instill public confidence, maintain levels of services relevant for nation building and for international credibility (CHE, 2001). It is envisaged that the achievement and maintenance of threshold levels will facilitate articulation between institutions and for support of excellence. The HEQC is said to be geared up to support the goals for quality improvement in a developmental way, and this requires continuous monitoring and evaluation.

The quality assurance system discussed here is meant for all three core functions, teaching and learning, research and community service. The Founding Document explains its intention to achieve quality in the three core functions against the principle of 'value for money' based on the HE purpose as outlined in the White Paper 3. This blueprint, it is said, provides an opportunity for evaluation. For this the Founding Document suggests external judgment that is rigorous and "non-reductive" and which will take into consideration the HE landscape in which quality will be judged. Attempts to observe the levels of responsiveness to the needs of the country are made within a 'fitness of purpose' framework based on the national priorities (goals, priorities and targets). In this way, the CHE attempts to look beyond the principle of quality only and bases its point of departure on some of the national imperatives that define the processes of transformation. The spectrum of components is necessary to understand when evaluation has to be structured.

It may be necessary at this stage to take note of the procedures followed in the 'operationalization' of the component of quality as this has implications for evaluation strategies. The Founding Document indicates that activities of quality assurance are undertaken in partnership with role players and implementation is undertaken in different phases. The first phase would lay the ground by preparing institutions and this would be followed by the testing of "offerings" on the grounds of accountability criteria. The evaluation process would be preceded by the establishment of self-evaluation capacity and where possible the acquisition of baseline information on quality assurance activities to be used in the monitoring process.

The stated mechanism of self-evaluation would produce information (in the form of reports) necessary for external validation. The CHE believes that systems such as HE (and its institutions) cannot completely self-evaluate, which makes external evaluation necessary. The latter process would also include site visits, peer review systems and the use of performance indicators. The Founding Document warns that this process of testing may evoke sanctions (CHE, 2001). Self-evaluation forms the base of the three comprehensive (large-scale) research evaluation models referred to in chapters 4,5 and 6 of this study.

9.4 The HE audience and other role-players

The Founding Document names the audience and stakeholders for quality assurance as the “learners” who need high quality HE providers, the HE providers who require an enabling environment, academics who expect to function in an environment that promotes teaching and research and employers of the graduates. The government and society are placed at monitoring levels, seeking accountability within the value for money principle. To fulfill these requirements, the component of quality needs to be observed and be transparent.

The Department of Higher Education and Training (DHET) is referred to as the main funding body for HEIs and as stated, is advised by the HEQC on the status and extent of service delivery. This makes the Department an important role player that will always be held in partnership on matters concerning HE. The CHE Founding Document ends by highlighting important contributions that the CHE attempts to make in HE. Some of these that may be relevant and guide this study are: the development of clear policies and procedures, identification of objectives, promotion of partnership with key role players financing the process, capacity building and the need for “an analytical and self-reflection approach” (CHE, 2001:16) based on a self-assessment and five year external evaluation cycle.

The Founding Document alludes to international comparison and expansion, already stated by the White Paper 3. On a “softer” note the CHE comments on benchmarking as “networks in order to share information as well as participate effectively in international debates and initiatives” (CHE, 2001:13). International systems directly influence activities of HE and thus serve as distant but important audiences. Benchmarking is also acknowledged where intentions to share innovative practices are indicated.

The Founding Document also reiterates and emphasizes the need for collaborative efforts for the evaluation of research. Similar to the discussions in the other legislative documents, the NRF is cited by the CHE as one of the suitable bodies for such an arrangement, making the NRF one of the more important role-players in HE research. It nevertheless emphasizes that the purpose of research evaluation when established, should be based on CHE policies. These expectations will be observed in this study.

9.5 The CHE Approach to Evaluation (all HE functions included)

As a body responsible for quality assurance in the HE system, the CHE does not concentrate on one function but generally addresses all core and other functions of the system. The two frameworks for evaluation/auditing used by the CHE together with the guiding document for the assessment of research management are the CHE’s policy documents relevant for this study. The first two discuss all functions while the latter only discusses research management. For a better understanding, these documents are discussed independently and later synthesized in order to reveal the relevant and important sections or even relatedness to this study. The documents are discussed below.

9.5.1 The Framework for Monitoring and Evaluation of Higher Education (CHE, 2004a)

9.5.1.1 Introduction

As one of the responsibilities of the CHE, monitoring and evaluation is a responsibility assigned by both the White Paper 3 and the Higher Education Act 101. The system, it is said, was meant to replace the South African Post Secondary Education (SAPSE) system previously used for the allocation of resources (although the present system is not

necessarily for the allocation of resources). The system is one of the fields in the CHE that would generate information useful for the Council when advising the DoE on reform and other issues and to inform the community on the performance of HEIs. Information is also made available to Parliament.

The Framework for Monitoring and Evaluation is based on the performance of HEIs in relation to their vision, policy goals and objectives. Also included are services with regard to transformation processes and local imperatives, mainly focusing on the trends of change and the fulfillment of national needs. It is intended that evaluation reports would also assist in the identification of key issues in the development of the HE system. The CHE describes and analyses the roles and efficacy of policies, structures, instruments, strategies and processes used for changing the HE system and to analyze the pace at which change takes place. Therefore, evaluation focuses on the stated factors and not so much on the general performance of co-functions.

The process of monitoring and evaluation (CHE Framework) starts by consulting with stakeholders and then negotiates for exchange of data with other providers. With the allocation of responsibility by the White Paper 3, the CHE is confident that the Council is the right place to house the process. According to the Framework (CHE, 2004a), the Council also acts as a mediator organization whose responsibility it is to oversee activities of institutional self-regulation. As an intermediary (parastatal) body, this responsibility would be equated to but not the same as that of the HEFCs in the UK and the PBRF in New Zealand, especially because the CHE is a statutory body.

9.5.1.2 Scope of the Framework

The system of monitoring and evaluation also reviews local imperatives against international trends and directions in HE. In the assessment of institutions, other different sectors of HE such as finances are also reviewed to enable advice on funding strategies. From the initial stages it was planned that monitoring indicators would be used for the core function areas while evaluation is based on the desired outcomes of transformation of the HE system. In this way, it is not as clear, as it is in the Founding Document,

whether the system is generally structured for the total evaluation of HEIs. While the system takes into account the institutional context within which implementation takes place, it is stated that consideration is made not to burden institutions with a demand for data.

Therefore, secondary data obtained from other departments such as the DoE and statutory bodies such as the NRF, the Human Sciences Research Council (HSRC) and Statistics SA and data from the HEQC are targeted. In this way the CHE only concentrates on analysis and interpretation of secondary data in line with its goals and hopefully translates data and results into credible information on the activities of HEIs. This plan to amalgamate and utilize secondary data does not qualify the Framework as a comprehensive/large-scale exercise. For example, the stated bodies produce data in relation to their own objectives and although the objectives are expected to be in line with the national imperatives, information may not necessarily fulfill the requirements for a holistic evaluation of the system. The NRF is mentioned here as providing secondary data, which may imply the possibility of the use of post-graduate research results or even rating data. All these stand a chance of not having available data depending on whether the NRF at the time undertakes such research evaluations. When this happens the system, which is undoubtedly the more economic one, may not be delayed or even disabled.

The CHE refers to different audiences in the Framework that will need the information produced from the evaluation activities. These include those involved in policy development, policy analysis, information interpretation and the management of finances that relate to HE activities with the hope that the information will reveal trends of success or deviation in policy implementation. Provision of such information will depend on the availability of information, which will also affect the regularity of such evaluation processes. The reliability of such data is determined by the “suppliers”. Other audiences are advisory bodies, implementation agencies, industry, sponsors and all users of HE. Through the Framework process, the CHE also intends to develop mechanisms and analyses in collaboration with stakeholders such as the science councils, state departments, and research organizations to produce results for different uses, keeping in

mind the idea of accountability and progression. The Framework does not specifically mention the NRF as a key participant in the venture although the section on research councils is regarded as inclusive to all.

Results are published through the media and existing CHE publications including the website for public consumption. It is hoped that institutions will also utilize the available results effectively for self-evaluation (self-regulation) within legislative parameters. As already stated, the academic community can also access the information.

9.5.1.3 Methodology

9.5.1.3.1 Areas and objectives of evaluation

The Framework explains the predicament of two focal points, the need to monitor and evaluate specific policy goals as opposed to the monitoring and evaluation of trends. The Framework states its objectives of evaluation as describing the state of affairs of transformation in the public HE institutions (trends), focusing on problems and on strategic ways to address them. This is done through comparing and creating a situation of sharing best practices, creating base-line information, encouraging public debates and rethinking the position of HE. The Framework is also intended to establish the direction of movement of the institutions, create awareness of the general trends and provide intellectual direction to the HE system, contributing knowledge for policy making and the understanding of change. To establish the efficiency and effectiveness of policies and strategies to implement them and to test instruments and structures used and to establish forms and pace of change are all the concerns of the monitoring and evaluation Framework.

Through all these objectives, the CHE intends to study how institutional goals translate into objectives, how policies relate to objectives and how to measure results and hopes to be careful about using sensitive enough indicators. This is despite the intentions to use secondary data. In their option to use indicators the Council refers to goals and objectives in the White Paper 3 and the National Policy for Higher Education documents by citing a whole range of domains/principles useable to explain transformation. As

stated, these are equity and redress (increased demographic representation) and quality (application and maintenance of academic standards) within the scope of the fitness for purpose (in relation to objectives), value for money and fitness of purpose (in relation to national goals) principles.

The Framework further refers to efficiency and effectiveness in the achievement of desired goals within affordable and sustainable limits and without any form of duplication. Also important is the responsiveness of HEIs to national goals, socio-political objectives and citizen formations together with performance of core functions in the development and the creation of knowledgeable and critical citizens. All of the domains are perceived within the operations of the core functions of the institutions (against which institutions are judged), yet information to achieve these evaluation goals depends on available secondary data.

The stated outcomes are important components of transformation and this process may be seen as responding to the system. The CHE further explains the sense of transformation as: redress of past inequalities, service for a new democratic order, addressing pressing national needs and a response to new realities in a global context. It is acknowledged that the move for evaluation is benchmarked against those countries which use evaluation results for accountability, allocation of resources, quality improvement and social transformation, even though such countries are not mentioned by the Framework. The latter imperative, although related to the others, is a South African priority. Emphasis is placed on quality, efficiency and responsiveness. The quality of data produced is very important if results have to be published, especially because publications are perceived as usable tools to promote democracy through transparency and to improve quality.

9.5.1.3.2 Methods used

The CHE is aware of the dangers of the use of quantitative measures such as performance indicators to measure a process such as transformation. This is consistent with Westerheijden's (1999) reference to performance indicators as proxies of a real situation. The Framework confirms this by stating that "implementation of policy by social actors is

not a linear process that can simply be read off indicators” (CHE, 2004a: 18). This is despite the fact that quantitative information may assist in deriving an impression of effectiveness and efficiency or even some aspects of quality provision for that matter. It is therefore stated that performance indicators, when used, can only give estimates.

A better understanding of HE processes and performance levels needs deeper studies. These deeper studies, according to the Framework, would allow for better understanding of institutional changes and their relation to societal and global trends. Evaluation research is best suited, it is felt, to pick these up and make sense out of results. How the CHE views all these to be achievable through secondary data remains to be seen.

9.5.1.3.3 Evaluation strategies

The CHE bases monitoring and evaluation on two frames of operation that relate to intra- and inter-institutional studies. First, that descriptive and explanatory information can be used for comparative and analytical purposes. This type of comparison is not aimed at promoting competition between institutions but to improve the situation. Secondly, the system utilizes what the Framework refers to as a “conceptual continuum between monitoring and evaluation activities” to track changes in the HE system.

It may be necessary to keep in mind the challenges faced by HE during this era of transformation. As these will inevitably affect implementation of objectives and thus the outcomes, the background information is necessary to assist in the interpretation of outcomes. The Framework mentions the merger of institutions (universities), programme rationalization, redress of past inequalities, the new democratic social order, pressing national needs and new local and international realities and opportunities as challenges expected in the process of evaluation. Some of these factors may also affect the availability of data meant for evaluation purposes.

9.5.1.3.4 Mode of monitoring and evaluation

According to the Framework, two processes are introduced in phases, within two processes feeding into each other. In the first phase “preliminary” performance indicators

are developed to hopefully provide estimates (CHE, 2004a). This is despite the criticism of the use of performance indicators by the Framework itself. At this level, it is stated, institutions are engaged to comment on the Framework. In the second phase, institutional goals are translated into indicators and more interaction will be made with key role players. Since the framework hopes to achieve all of these through the use of secondary data from other providers, it may be perceived that at this level negotiations for such data may be underway. This needs agreement with providers on operational definitions, criteria and standards for data sets and research collaboration. As indicated, data is sourced from the DHET, D of Labour (DoL), NRF, SAQA, HSRC, Stats SA and others in a form that will follow established protocol with the bodies.

Activities of monitoring and evaluation have been planned to run biennially from 2004 through the co-ordination of standardized monitoring and evaluation studies with other initiatives. The system uses co-ordinating routine descriptive and longitudinal monitoring systems. The selective in-depth case-studies are used to augment the above in a qualitative manner, the interpretative process and through the use of timelines for reporting. The use of case studies is unexpected in a process that bases its operations on secondary data, yet verification is a normal procedure in evaluation research and is therefore acceptable.

The CHE board is responsible for the process of monitoring and evaluation through a standing committee, which is also responsible for any negotiations for relevant data and for operation of the systems. Specialist research teams are from time to time considered (invited) to conduct research. For purposes of transparency, dissemination of results (in a joint production of publications) is done through *Higher Education Monitor*, *The State of South African Higher Education*, the CHE website and other media.

9.5.2 CHE Framework for Institutional Audit (CHE, 2004b)

9.5.2.1 Introduction

The audit is one tool used by the CHE as an attempt to respond to radical reconstruction within the complex system of HE. For example, in line with transformation, the audit

process observes ongoing reforms and restructuring, in an attempt to assist institutions to manage quality effectively. The audit process also “seeks to evaluate institutional initiatives to produce a vibrant intellectual culture both within the institution(s) and in society(ies) and acts as an incubator of new ideas and cutting edge knowledge as part of the national system of innovation.” (CHE, 2004b: 3). The process is viewed as an attempt to link the achievement of quality to transformation objectives (external evaluation). This has been established by the HEQC as a value-added system used to validate institutional information on the effectiveness of internal quality arrangements on development, enhancement and monitoring of quality in the core functions.

The HEQC runs a six (6) year cycle starting from 2004 to 2009, concentrating on institutional arrangements of assuring quality of core functions namely, teaching and learning, research and community engagement. The somewhat compulsory nature of the programme may be the reason why more attention is given by both the institutions and some HE organizations such as the South African Association for Institutional Research (SAAIR). Through the SAAIR Quality Institute the Association committed itself to preparations for the HEQC institutional audit.

The HEQC maintains the idea of institutional responsibility for credible self-evaluation and sustained improvement in its quality assurance venture. The function of the audit process is to take stock of the institutional arrangements and processes and produce results. As a new undertaking, the cycle initially conducted trial runs that were monitored closely with adjustments made where it was deemed necessary. The process has already been conducted in some universities, especially those that have not been affected by the merger processes.

9.5.2.2 The Audit process

9.5.2.2.1 Introduction

The CHE (2004b: 13) uses the terms ‘audit’ and ‘evaluation’ interchangeably (glossary-Audit scope). Institutional audit is defined as “improvement oriented external evaluation of institutional arrangements for quality in teaching and learning, research and

community engagement based on a self-evaluation conducted by higher education institutions” (CHE, 2004b: 15). According to the CHE (2004b: 1) “quality-related criteria constitute a crucial component in the execution of the HEQC functions, fulfilling the dual purpose of serving as evaluation tools for the HEQC’s audit and accreditation activities as well as setting broad benchmarks for quality management in higher education”.

9.5.2.2.2 Scope and areas of HEQC audits

Similar to the CHE monitoring and evaluation process, institutional audits focus on institutional policies, procedures, systems, strategies and resources for quality management of core functions together with academic support services. Criteria used in the audit process cover aspects of input, process and output.

The areas for audit involve the relationship between quality and fitness for purpose by assessing the extent to which an institution’s mission and academic activities take national priorities and needs into account. This also gives an impression of the extent to which institutions respond to regional and international imperatives. Therefore, audits are necessary for policy development, resource allocation, policy implementation, evaluation of extent and impact and identification of intervention for improvement. All other functions, for example finance and governance are considered when they make impact to quality objectives. Audits are used to “evaluate” such teaching and learning arrangements as RPL, moderation, assessment, and certification.

The first cycle (2004-2009) was planned mainly to focus on mission and planning together with arrangements to audit core activities. Results of the audit are used to prepare for the second cycle. Unfortunately, as will be discussed later, other factors influenced the completion of this cycle.

9.5.2.2.3 Criteria in HEQC audits

Audit criteria are used as a tool to identify important signals for quality as well as its management. National imperatives are also considered in the audit criteria for use by

institutions in self-evaluation. These include the national regulatory frameworks, institutional quality landscape (varying levels of capacity, for example, universities *versus* technikons), international trends and stakeholder comment. Other national imperatives of interest to auditing are the HE policies, the HEQC documents and the national benchmarks for comparability and articulation possibilities with reputable institutions abroad, for example, those of the DoE (on enrolment, output, research productivity, diversity, graduate rates etc). Benchmarking is used for goal setting and self-improvement. Also important are partnerships in HE (for example, inter-university), increases in cross-border provision and transformation aspects such as equity and access, and responsiveness to local regional and national needs.

As indicated under the sub-heading Scope above, the criteria are applied to areas including mission, context and objectives, goals and institutions' levels of development. These criteria serve as a standard tool usable by all institutions in their "quality management arrangements and as guides for self-evaluation. Therefore criteria assist the audit to focus on... indicators of quality management" (CHE, 2004b: 1).

Quality management means to standardize quality support and sustain existing levels, develop and enhance quality, monitor and evaluate and act on quality issues. As stated, criteria are based on the above understanding, focusing on input, process and output and evidence of self-evaluation has to be based on them. It is expected that quality management be integrated within institutional planning.

9.5.2.2.4. Audit methodology

9.5.2.2.4.1 Aims and objectives of the audit process

The objectives of the audit, according to the Framework, are to encourage and support the culture of continuous improvement in HE, and to validate systems of self-evaluation (CHE, 2004b). The audit also enables HEIs to develop reliable indicators for effective quality enhancement, and provides information and evidence, for example, of strength and excellence or weakness for improvement. Also important is the creation of baseline information on a common set of audit criteria for HEIs to disseminate good practice,

facilitate capacity development and create a platform for self-accreditation. These somewhat developmental intentions resemble the activities of research evaluation in The Netherlands in the late 1980s and early 1990s, whereby institutions were orientated into processes of self-assessment. These objectives may be used as assessment goals to find out if the existing efforts in South Africa are “good enough” for the situation in the country, benchmarking, using other models in other countries and fulfilling international comparison aims.

9.5.2.2.4.2 The audit process and methods

The CHE claims that the audit methodology is based on international standards consisting of institutional self-evaluation (audit portfolios with supporting information or any evidence of efficiency on core functions) and external validation by peers and experts (panel). The panels are appointed by the HEQC to also look at the internal quality arrangements, leading to the submission of a report to institutions. According to the Framework, “based on the panel’s findings, the HEQC’s report to the institution will identify areas of strength and good practice as well as areas in need of attention and provide commendations and recommendations for action” (CHE, 2004b:8). Institutions will then “implement appropriate follow-up strategies” in response to recommendations of the reports. A summary report is then made available for public consumption.

Unlike the CHE process of monitoring and evaluation, the audit process takes a holistic approach and directly acquires primary data from institutions. This approach enables the process to advise institutions on areas that need to be improved. It should also be remembered that the mandate of the HEQC relates to the building of an effective national quality assurance system. Two areas will be covered by the audits (as per audit arrangements). The first area covers foundation elements (mission, planning and management) and the second covers the core functions.

The audit process involves scheduling of visits – nine months’ notification on the extent of the visit and portfolio development (description, analysis, evaluation report). Self-

evaluation portfolios are to be submitted at least 10 weeks before the audit visit (portfolios to meet specific requirements).

A list of potential auditors already exists at the CHE. The HEQC sets up a panel for individual institutional audits. The size of a panel varies according to the complexity of the institution and the type of audit. The panel is made up of the chairperson, the audit officer (staff of the HEQC, principal point of contact) and other members drawn from different sources. Institutions may comment on the constitution of the panel to avoid conflicts of interest.

The audit process starts with a pre-audit workshop for familiarity and identification of focus area. This is followed by audit visits (validating claims in portfolios), which may take two to five days. During the visit, interviews are conducted with staff, students and other role players, available documents are scrutinized on site, tuition centres are inspected and additional information may be requested. The audit process is summed up in figure 9.1 below



Figure 9.1 Visual presentation of the audit process

- *- Self –evaluation reports are submitted 10 weeks prior to audit visits
- Pre-audit preparations include the constitution of panels by the CHE, institutions’ comments on the constitution, panels’ pre-audit workshops and the development of portfolios
- Two to five days visits

9.5.2.2.4.2 Reporting and the utilization of results

Quantitative and qualitative evidence of the adequacy and effectiveness of internal arrangements are compiled into a report and commendations and recommendations made. A draft is then presented to the institution for comments, followed by a final report. As stated, a summary report is made available for public consumption. It is important to note that results are not used for ranking of institutions but to encourage systemic quality improvement and institutions are to develop and implement quality improvement plans in relation to recommendations. To fulfill this expectation, the HEQC suggests explicit actions within a specific framework where areas of serious concern are identified. It is also important to note that audits do not directly affect funding and are not linked to any form of it, as funding is a responsibility of the DoE (recently named the Department of Higher Education and Training).

A further role of the HEQC is to monitor the progress of such implementation (mid-cycle progress) through the report from the institution (three years after the visit) and then forward results (observations) to the DHET for attention. The Committee (HEQC) is also involved in the monitoring of improvements in general. Figure 9.2 below portrays the sequence of events for quality improvement.

In the monitoring of improvements, the HEQC uses samples from universities and criteria are used as indicators in a form of checklist and supplemented by open-ended questions and answers.

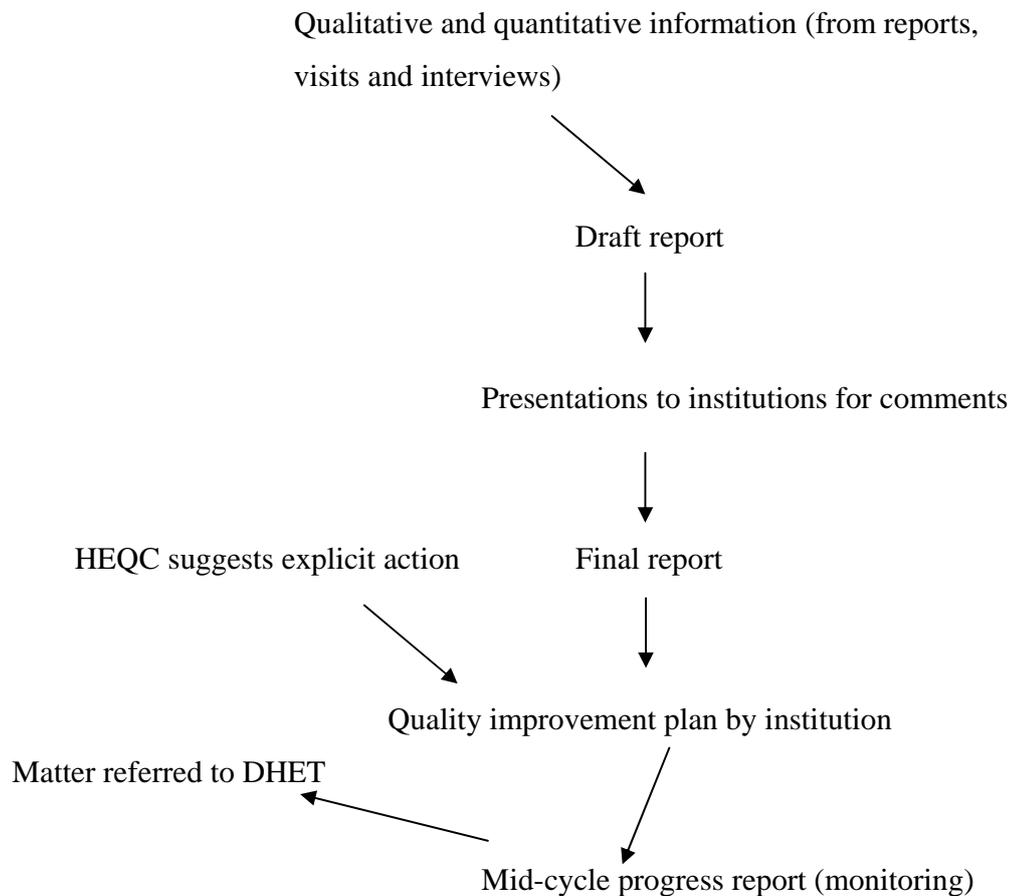


Figure 9.2 Summary of a quality improvement sequence as interpreted by the researcher

9.5.2.3. Criteria relating to research audit:

It is assumed that as research is a core function, the research audits will follow the same process as the rest of the functions. The emphasis with research is the enhancement of intellectual inquiry and application of research to social developments. The HEQC therefore focuses on the arrangements by institutions to ensure, develop and monitor the quality of research together with arrangements to ensure, develop and monitor the quality of postgraduate education. “It is assumed that some measure of research activity is underway at all higher education institutions” (CHE, 2004b) undertaken to inform improvements of quality of teaching and research and postgraduate education.

Quality related arrangements used for research as per the audit process relate to support of research functions and processes such as the availability of a research policy and plan, including the nature of research undertaken, adequacy of resources, plan for implementation, monitoring and evaluation and strategies for research development, capacity building and the implementation and monitoring thereof. Support also involves collaborative efforts, access to and allocation of funds and regulation of research output. Other quality related arrangements are the mechanisms for quality assurance of research including criteria to evaluate and approve research proposals and information systems to capture research capacity, funding and output. In short, the audit process is responsible for processes of improving the quality of research activities and not the output directly.

The CHE concludes the Framework by stating that “research functions and processes are supported and developed in a way that ensures and enhances quality, and increase research participation, research productivity and research resources” (CHE, 2004b: 17). In this way, it is hoped that performance of high standards will be achieved and enhanced.

9.5.3 A Good Practice Guide for Quality Management of Research (CHE, 2005)

9.5.3.1 Introduction and background

Similar to other CHE documents, the guide starts by highlighting the aims of change in South Africa and the goals outlined to achieve such aims. Democracy features prominently as a determining factor for societal development. Democracy is said to refer to equitable access and the enhancement of social responsibility. In this way, “the majority of the population can share in the wealth of the country” for the realization of “individual and collective potential” (CHE, 2005:1-2), to be achieved through a process of transformation.

In HE transformation is said to promote equality and access and to eradicate discrimination, and provide well co-ordinated teaching / learning and research programmes that include high-skill employment needs. Transformation would also imply

the promotion of democracy and installation of the culture of human rights, good values, critical and creative thinking and open doors for learning that advances knowledge and scholarship. Transformation can also be fulfilled by addressing the diverse problems of society and the country and through efforts that “uphold rigorous standards of academic quality”. This information forms the basis of planning and has to guide processes when the purpose of research evaluation is established.

Realization of the above mentioned responsibilities, the Guide states, depends on good planning, funding and quality assurance. The latter is referred to as the medium of transforming HE and is expressed in the form of principles, methodology and tools to monitor and improve the quality of academic provision. Quality of provision takes into cognizance the HE landscape which is affected by historical trajectories.

This Good Practice Guide is meant to assist with the development of the internal QA mechanisms of research management. As this is the programme/area that directly deals with research, it is discussed at length with the purpose of establishing its scope and contribution to this study. As research is recognized and highly rated in HE as playing an important role in knowledge production and the generation of scientists, the CHE found it necessary to give it treatment that would improve the quality of its management. This provides support to processes explained by the Institutional Audit Framework.

Similar to other core function in HEIs the field of research also contributes to social and economic development in the country within the democratic system. This is because research itself contributes to government and society through its knowledge production and dissemination effects. In order for the HE system to be internationally competitive and still meet transformation requirements, including redress of inequities, the fulfillment of national demands and the demands of science, several requirements have to be met. This has to be done despite problems of reduced funding. One of the transformation objectives is to have proper mechanisms of research support that may bring about opportunities for institutions to compete on equal grounds.

The whole notion of research development in institutions led to the establishment of research management fields to assist in the creation of a conducive environment for research by translating research policies and plans to be in line with those of institutions and to avail opportunities for researchers and post-graduate students. Therefore, research management creates an information system to monitor, evaluate and plan activities of research in institutions.

Of importance to the CHE is the role of research management to promote the production of quality research. This is achieved through the creation of ‘good’ policies, structures, processes, mechanisms etcetera, in order to provide support for the envisaged international competitiveness. This is done by creating an enabling environment for research. The process also encourages a culture that fosters imaginative, creative and innovative research. It is nevertheless acknowledged that “the production of excellent graduates, cutting edge research and innovative community engagement depend not only on the availability of efficient quality assurance mechanisms but also on the sustained nourishment of a community of students and scholars” (CHE, 2005:3). This is a huge task, which when managed effectively should turn around institutions through research and equalize research performance across the entire system of HE without tampering with the broad landscape. Once these are in place, comparisons are acceptable and evaluation is necessary to judge improvements and achievements.

9.5.3.2 Quality Management of Research

This Good Practice Guide adds to the criteria identified in the HEQC audit document by explaining good quality practice for quality research management (indicators) and to assist with the process of self-evaluation (internal quality management system) in line with institutional missions, identities and research aims. According to the Guide, these have to be achieved within the framework of institutional quality assurance, support, development and monitoring of research activities together with the allocation of resources.

The document (Guide) draws information from other HEQC documents especially on audits together with international practices on HE research management in relation to processes and institutional research performance within the context of operations (aligning the landscape with quality concerns). It nevertheless has shortcomings as far as the assessment of research performance is concerned. The following information gives the scope of the guiding document and also indicates areas that the document does not cover (CHE, 2005:6):

“Since the focus of the Guide is on mechanisms for quality management of research it does not consider research management in general ... does not include all the functions of an institutional research office or how it is managed ... it does not include suggestions for the quality assessment of the actual output of an institution”.

The latter statement is important in this study, especially because most of the models referred to use research outputs as a variable in their evaluation/assessment strategies. Therefore, the Guide is not inclusively taking care of the usual activities of research evaluation but focuses on developments (management) that may improve research activities. This opens a gap that may be filled through the development of a research assessment framework, which will guide both the internal (self-evaluation) and external evaluation of research.

The idea to create a conducive environment and avail opportunities for research and the training of post-graduate students creates a good foundation. Therefore, if assessment is conducted of these services, it will not only bring about improvement in the quality of research but will also directly or indirectly promote participation that may lead to an increase in output. Should the practice be conducted across the system of HE, there are possibilities of an across-the-system improvement. The latter is a transformation imperative, which when achieved, will serve as a tool for equitable access and thus the enhancement of societal development stated under the sub-heading of methodology below.

9.5.3.3 Methodology

The Good Practice Guide attempts to establish ways of assessing the research criteria of the audit manual, which refers to quality-related arrangement for postgraduate education and research. The document states as the objectives of the assessment, effectiveness of management as evidenced by input, process and output factors, data that would serve as evidence and reporting strategies usable. That is, questions such as ‘what assessment strategies are available?’, ‘which data should be used?’ looking at what is available’ (for example in HEMIS categories) can provide the necessary evidence and how a report should be compiled.

The management system includes the management of research policies, rules and regulations for researchers and postgraduate students, based on the development and evaluation of proposals, assessing resources, conducting research and making research public. The adequacy of resources and the consistency of implementation and monitoring of policies and plans feature most in this system (models). The Guide states that:

“We have to assume that any increase in research participation, output and funding is evidence of the effective implementation of the institution’s research policy and/or plan. To demonstrate any increase one obviously has to report on data over time,” (CHE, 2005:41).

By implication, periodic evaluation of the above-named criteria would provide an impression of the increase in levels of participation, output and funding. The model also includes the effectiveness of research functions judged through the quality management of structures (implementing, mentoring) on the above-named and the effective use of research information system models. These include data capturing on capacity, funding and output, clearly defined and comparable to other data in the national research system. Also included are support and development strategies including the support and development of post-graduate education and the impact of programmes.

On another note, the Guide refers to research output, the important indicators including publications, reports, patents and non-textual output in accredited journals, books, monographs, chapters, published conference papers and contract research per faculty/school and per gender and race over time. The other domain included is research funding and monitoring of research policies and plans (on review reports, records on policy decisions by research committees and/or senate).

9.5.3.4 Assessment and reporting on the effectiveness of quality management (internal self-regulation)

It was necessary for this study to have an understanding of the CHE self-regulation through the research management model, in order to establish the need for a new approach. The self-regulation tool is meant to assist institutions to report on their own efforts in the named criteria. Effectiveness in this sense refers to the enhancement of quality increase in research participation, research productivity and research resources. This definition is used in this Guide and needs to be read with the previously explained scope that, as stated, clearly excludes the quality assessment of the actual research output. This is especially because assessment usually implies gathering of information as evidence to support claims related to research support, development, participation, productivity and resources, which also relate to output measures.

The document identifies domains, which are thought to provide relevant information for the assessment of quality management of research. The domains include the following:

- Adequacy of the resources of research policies and/or planning. This refers to financial resources, human resources (proportions of academic staff on the total institutional staff, profiles of academic staff on qualifications and further studies and time devoted for research all per faculty) and equality based on race and gender.
- Consistency of implementing and monitoring of policies and plans. This is expressed through outcomes of the process of implementation. Consistency according to the Guide also includes research participation (capacity, increase in proportion of academics, changes in rank and qualifications) and increase in number of postgraduate students in relation to undergraduates per faculty. Participation refers to redress and equality (increase in the number of women and black researchers) and postgraduate student participation. Other criteria include the proportion of staff applying for NRF rating and success rates, proportion of staff applying for funding and success rates, the number of postgraduate fellows and the number of visiting researchers.
- Effectiveness of research functions as measured through structures and mechanisms. This refers to appropriateness of placement and authority, expertise to support and achievement of targets. This also includes the actual functioning of structures and mechanisms such as the holding of meetings and briefs and criteria used to perform functions.
- Effectiveness of implementation and monitoring of strategies for research report and development (including capacity development). Reporting at this level includes budget allocations for strategies, participation rates, success rates, performance indicators for measuring impact and programme focus.
- Effective use of comprehensive information systems (ISs) for planning and external reporting. This involves how the institution uses the system, reports on the update of information, retrieval of information (who and why, for planning, external reporting on capacity or performance). This is guided by

the fact that “if information captured in this system is infrequently used by the institution then it has to be assumed that the institution has a research information system that is inefficient and not comprehensive” (CHE, 2005:49).

- Effective implementation and monitoring of mechanisms and strategies of assuring quality, and supporting and developing postgraduate education. This refers to relationships between supervisors and students, ratios of supervisors to students per faculty, frequency of meetings between students and supervisors, the nature of such meetings (conceptualization, analysis or write-up) and feedback given to students. This level also includes to some extent the postgraduate research output on publications and conference presentations (including joint presentations).

The information above gives guidance to systems of internal research management. It is clear that the Good Practice Guide is an expansion of the audit process on research. The Guide provides detailed activities which when meticulously applied will guarantee quality management processes in research. This information is relevant to research evaluation in that it provides the profile and state of affairs of research activities in institutions. That is, the Guide provides information on the scope of internal-regulation and self-evaluation.

9.6 Synthesis of the CHE/HEQC activities

9.6.1 Introduction

This sub-section summarizes the systems/programmes of the CHE discussed above. Discussions reveal the mandate of the CHE and its committee the HEQC as the assurance of quality across the functions of the HEIs. Together with the NRF, the bodies are trusted to possess skills to enable the HE system to deliver on transformation, against a set of standards of international comparison.

Besides quality, other important transformation components aligned with the improvement of the HE system include equity, innovation and diversity. While innovation and diversity relate to responsiveness to national demands, high standards may be achieved through a HE centre (institution) that provides efficient and effective services. These demands hold the HEIs accountable and ‘compel’ them to deliver relevant services to fulfill the requirements of transformation and the CHE plays the supporting role.

9.6.2 Assessment/Evaluation

To establish levels of achievement of the components by the institutions, assessment is necessary. According to the CHE, this assessment system is best achieved through two main processes: self-regulation and external validation. It is also necessary for assessment to permit the process to flow in a formative manner to allow quality improvement to follow a developmental approach. The initial phase of assessment in the CHE programmes lays the ground-work and prepares institutions. This phase is also linked to capacity building for self-evaluation and the accumulation of baseline information for monitoring purposes. The second phase is attached to accountability criteria in a continuous and developmental system and may be altered with changing national imperatives.

9.6.3 Methodology

Table 9.1 below summarizes the CHE programmes under the methodological criteria of purpose, focus area and methods. Benefits and challenges of the programmes are indicated in table 9.2. Discussions of the summaries follow.

9.6.3.1 Purpose

The objective of the CHE evaluation to assess the alignment of HEIs’ activities with the transformation imperatives (CHE, 2004a), to assess quality and to examine if the activities are good enough to fulfill the national (transformation) imperatives (CHE, 2004b). These goals are to be achieved in different ways and in research in particular,

measures are in place to develop this core function (CHE, 2005). The transformation process to which all the programmes subscribe, is regarded as an attempt to equalize activities in the country, including the system of HE.

The purpose of creating a culture of continuous support stated by the Institutional Audit document contributes to the domain of sustainability of HE activities and the assessment programmes. That is, once institutions are involved and are able to manage internal assessment processes, self-regulation is initiated, which will hopefully bring about continuous development of activities and therefore quality.

Improvements in internal mechanisms are likely to but not necessarily lead to high levels of performance of intellectual competitiveness. However, continuous assessment of both the universities and at systemic levels not only provides information on trends and alignment (relevance) but also creates a good atmosphere of accountability with better proficiency.

9.6.3.2 Areas of focus

All three programmes in this section focus on the transformation criteria attached to levels of responsiveness to national needs (fitness of purpose), quality of services (fitness for purpose) and accountability for state funds (value for money). The domains/components are assessed, based on foundation elements and core functions of the universities.

It is necessary to remember that the Framework for Monitoring and Evaluation cuts across all core functions and systems of HE and has been included here for the reason that research is one of them. The Framework for Institutional Audits on the other hand, first includes the entire system and later (but briefly) focuses on research while the Good Practice Guide is fully focused on research management at a local university level. The intended outcomes for the latter two are related, which may either imply that the Guide is meant to 'control' the internal (self-) evaluation for the production of reports for institutional audits or that research assessment can be managed separately.

9.6.3.3 Methods

Whereas the CHE monitoring and evaluation programme has a similar focus as the other two programmes, the former establishes performance indicators to determine trends from secondary data to validate the system. In some cases, as indicated in the Framework for Monitoring and Evaluation, case studies may be conducted. The research management programme (Good Practice Guide document), on the other hand utilizes the internal efforts of the institution (self-evaluation) to create a culture of independence.

The Institutional Audit process uses a holistic approach and first involves institutions to compile portfolios and then engages panels for external validation. The process involves peer reviews, which includes report/documents analysis, site visits/inspections and interviews. The six-yearly programme is continuous and although in the beginning the acquisition of baseline information takes priority, the process reveals characteristics of a comprehensive arrangement observable in large-scale models discussed in chapters 4,5 and 6.

9.6.4 Conclusion

The CHE programmes reveal efforts prescribed by the Council to add value to the functions of HEIs in South Africa. While two of the programmes are not directly linked to research evaluation, the one that is (Good Practice Guide) only concentrates on the quality management of research. The most encompassing Institutional Audit programme gives guidelines as to how internal and external evaluation may be conducted but also concentrates on quality management of core functions. This opens an opportunity for further exploration of a model for research evaluation, taking into account the objectives, focus areas and the methods used in the programmes.

Table 9.1 A summary of the CHE programmes of evaluation

Programme	Purpose (Objectives)	Area of focus/components	Methods used
Framework for monitoring and evaluation	-To establish and monitor trends of the HE system alignment with the transformation system in SA	-Relevance/responsiveness (fitness of purpose) -Quality (fitness for purpose) -Accountability (efficiency, effectiveness and value for money) Focus: policies and goals, trends of transformation variables, levels of creation of knowledge.	-Conceptualize continuum between monitoring and evaluation in a longitudinal system -Use of secondary data obtained from other bodies (NRF, HSRC, Stats SA, HEQC, DoE, DoL, SAQA etc) -Select in-depth case studies to augment reports -Descriptive and comparative analysis -Biennial studies Assessment mode: CHE board and invited specialists
Institutional audit	-To find out if efforts of HEIs are 'good enough' to link quality to transformation imperatives (value-adding tool to validate institutional effectiveness) -Encourage and support the culture of continuous support	Take stock on: -Foundation elements: Mission planning and management (policies, resources, systems) -Core functions: Vibrant Intellectual culture, new ideas and cutting-edge knowledge (quality, responsiveness and accountability)	-Self-evaluation portfolios by institutions (descriptive analysis and reporting): *National imperatives criteria *Regulatory frameworks *Quality landscape *International trends *Stakeholder comments

		<p>-Research focus (quality of): Management, research, post-graduate education (policies, resources, evaluation of proposals, developmental strategies, capacity building, regulation of outputs)</p>	<p>*Variables: (enrolments, output, productivity, diversity, graduate rates etc) -External validation (CHE panels): Document analysis (reports), site inspections, interviews -Follow-ups by (CHE): Appropriate strategies</p>
Good practice guide	<p>-Assist with the development of internal QA mechanisms of research management -Promote the production of quality research -Provide support for international competition</p>	<p>Measure consistency of implementation of: -Policies -Structures -Processes -Mechanisms (Rules and regulations of post-graduate studies, evaluation of proposals, resources, publishing of results, scholarship)</p>	<p>-Primary data -Use of input-process-output data -Self-evaluation on focus areas</p>

Table 9.2 A summary of the outcomes and intended benefits of the CHE evaluation processes

Programme	Outcomes/benefits	Challenges
Framework for monitoring and evaluation	<ul style="list-style-type: none"> -The use of secondary data -Sharing best practices -Provide baseline information -Platform for public debates -Rethinking and determining the direction of HE (form and pace) -Create awareness -Contribute knowledge for policy making and understanding of change 	<ul style="list-style-type: none"> -Secondary data is not easily verifiable -Merger effects -Establishment of performance indicators for use with secondary data
Institutional audit	<p>General institutional:</p> <ul style="list-style-type: none"> -Create baseline information -Capacity for self-accreditation -Reveal good practices and weaknesses <p>Research specific:</p> <ul style="list-style-type: none"> -Enhance quality -Increase participation -Increase productivity -Improve resources 	<ul style="list-style-type: none"> -Focus mainly on the domain of quality

<p>Good practice guide</p>	<ul style="list-style-type: none"> -Enabling environment for research -A culture of imaginative, creative and innovative research (scholarship) -Production of excellent graduates -Increase in participation, output and funding -Establishment of research information system (RIS) -Increase in research resources -Monitoring plans and policies 	<ul style="list-style-type: none"> -Does not consider research management in totality -Does not include functions and management of research office -Does not include quality assessment of actual output of research -Only concentrates on the components of quality services -Does not take care of activities of evaluation especially of the entire HE system
----------------------------	---	--

Chapter 10 Discussion and syntheses

10.1 Introduction

This chapter first discusses and synthesizes information on the models of research evaluation discussed in this study for the purposes of forming a background. Later the chapter synthesizes and collates South Africa's pluralized efforts using the knowledge base drawn from the models mentioned above. The information is used to identify the similarities and differences between the different models and to understand the significance, effectiveness and relevance of the research efforts of South African higher education (HE). The information led to the conclusions and recommendations in chapter 11.

10.2 Synthesis of the three international models

10.2.1 Background

This sub-section synthesizes and discusses the models of three other countries referred to in this study namely: The Netherlands, the UK and New Zealand. The sub-section compares similarities and differences by using six descriptions to capture discussions and to create a better understanding. Discussions are presented according to the following headings:

- Origin, rationale and decision context
- Purpose of research evaluation
- Methodological issues
- Benefits and advantages accrued
- Problems and challenges encountered
- Best practices

10.2.2 Origin, rationale and decision context

The three systems of research evaluation fall under Campbell's (Shapira and Kuhlmann, 2003) Type A research evaluation model classification in that they are systemic and comprehensive/large-scale research evaluations of institutions at national level and across

all disciplines. The research evaluation models have different origins, rationale and contexts in which they were conceived.

The origins of the Dutch research assistant exercise date back to the early 1980s, where the state was willing to grant autonomy to HEIs provided that the institutions would in return demonstrate expected levels of accountability. From its inception, the decision context was inclined to institutional “agenda building” which Rip and van der Meulen (1996) refer to as “aggregation”. As a result of the lack of trust, government initially held to its steering powers through the intermediary funding organizations to check on accountability and then by introducing the monitoring of the systems. Monitoring by government was characterized by inspection activities and the monitoring of programmes to improve both the institutions and the process of evaluation as a whole. The meta-analysis/evaluation of the system by the Inspectorate is a typical indicator of the level of government involvement (monitoring). The purpose of government involvement was to ensure that the process complied with the expected levels of quality.

To meet government requirements and expectations the Dutch HEIs formed a unifying body, the VSNU to help build the expected levels of accountability. This was the only organization whose activities government would monitor. Therefore, the planning and implementation of evaluation became a responsibility of the institutions (through the VSNU).

As developments unfolded, all efforts remained with the research organization, the VSNU, which later also became responsible for meta-analysis/evaluation of the system. Rip and van der Meulen (1996) warn of extreme levels of aggregation and regard the move as abandonment of state responsibilities or even failure of the state to provide direction on imperative national issues. Also, there are fears that aggregation may introduce elitist behaviour detrimental to development (especially of new entries) in research. The Dutch system claims strength in the dimension/domain of relevance and intends to prove this, together with other dimensions through continuous publication of their performance.

The UK research assistant model dates back to the same period but in this case the motive behind the initiative was to reduce the HEIs' dependency on state funds. With the HEIs being predominantly state-funded, the funds could not meet the demands of the institutions in the UK and a need arose for a move towards a market economy, which would allow alternative funding in the system. Therefore, from the initial stages the system of research evaluation was at the steering end of the steering-aggregation continuum, with government "manipulating" the system through the provision of resources. Although the HEFCs served as intermediary bodies to reduce dominance, the main intention was to reinforce national aims. The operations of the intermediary bodies were also guided by state policies. Supposedly once the institutions achieved independence from state funds, the steering powers would automatically fall off.

It is necessary to note that in both instances (Dutch and UK) evaluation of teaching was undertaken concurrently but separated from that of research for the sake of convenience. Although in the UK there were fears that the pressure of research evaluation would compromise teaching activities, the exercise gained approval as it developed. As stated, while evaluation in the UK continued to be state-bound, the government in The Netherlands kept its distance with a critical close watch. Activities in The Netherlands started with the ex-ante evaluation of proposals for funding and later through the ex-post evaluation of output based on input.

Notably, both the Dutch and the UK HE systems were state funded. This meant that funding directly or indirectly emerged as a determining factor for the conditions in HE and thus the decision contexts for research evaluation at the time. In The Netherlands, the need for independence of institutions also indicated the influence of secular changes.

The UK and The Netherlands are geographically close to each other and it cannot be ruled that that they influenced each other. This is evidenced by the fact that in both cases accountability, which also affected the Dutch system in its initial stages, related mainly to the use of state funds. As developments unfolded, the Dutch decision context moved

from mere financial accountability to measures related to the improvement of the systems of research in HE. It is therefore not surprising that dimensions of quality, relevance, productivity and viability/vitality and feasibility became well-defined, guiding principles for evaluation in the country. Similar dimensions are referred to in the UK but without the same emphasis.

Both systems define quality through the medium of output and the acceptance of rigor of review of publications, a characteristic which is a subjective definition but a sign of interdependence of researchers, especially those that are regarded as more knowledgeable (experts) in their areas of research. As a result of this common understanding and possibly the geographic location, there emerged interdependence of the two countries on scientific experts, a practice that may have introduced the international dimension in the 20th century. Geuna and Martins (2001) view this as a tendency for comparison, which may swerve different research evaluation ideas towards some common ground.

The early 21st century is mainly characterized by the ideas of globalization/internationalization and research excellence. It is notable that in the two countries the idea has developed through a series of activities that raised the countries to high levels of competitiveness. This was done by involving experts from other countries as adjudicators and peer review experts in their systems of research evaluation and by encouraging local experts to be engaged internationally, in a bid to define quality. It is therefore not surprising that competence and standards make reference to 'other countries' (or even international standards).

Time and time again both systems refer to international standards, which seem to be intended for and targeted by both countries. It is nevertheless not clearly stated as to which country is referred to as the international referee in both cases. These standards are revealed through publications, conferences and involvement as international experts. Publication in journals that are acknowledged by 'others' and the sharing of scientific information with others in other countries are said to depend on a rigorous selection of

‘good’ papers. ‘Involvement as international expert by other countries’ is regarded as being significant/visible to those that extended the invitation.

Within the realm of internationalization, the Dutch system of evaluation takes precautions and protects the Dutch ‘minority language’ by involving experts, in their evaluation exercises that are of Dutch origin and/or speak the Dutch language. They are nevertheless also affected by the methodological confounding effects of defining quality through bibliometrics and similar variables which some of the panels in the UK refer to as a sign of visibility.

While the Dutch system has already moved to the institutional self-regulatory processes of research evaluation, which they built over a period of time, the idea has been proposed only recently in the UK. The Dutch self-regulation process became a trademark of the Dutch system in the endeavour to hold individual institutions accountable to quality services. The process has been initiated and developed over a period of time, through the external involvement of experts. It (the process) now involves such experts internally for mid-term evaluation and it is hoped that in future this will reduce the responsibilities of research organizations and thus the funding loads carried by the organizations. In this way, autonomy and accountability no longer rest with the entire system of HE only but also involve individual institutions.

New Zealand approached the last stages of the 20th century and the beginning of the 21st with a plan to improve the functioning of their HEIs (referred to as the tertiary education sector). As a result of changes in politics, the new government took the lead in aligning HE with national imperatives by creating centres of excellence, improving on social science capacity and establishing a system of HE research evaluation, basing funding thereof on performance. The latter initiative was intended to change the process of funding from the block grants and ‘top ups’ to a value-for-money, performance-based one.

There are clear indications that government steered the process to ensure accountability. Although, like the UK, they introduced an intermediary body the PBRF, the emphasis on funding is much stronger in New Zealand, and was used to compel researchers to perform or face exclusion from funding. Excellence was cited as an important value guiding activities and contribution to intellectual infrastructure for advancement of knowledge (contribution to the body of knowledge). The dimension/criterion served as a base/decision context for the first large-scale (2003) research evaluation exercise.

In common with the other two countries, autonomy of HEIs was observed in New Zealand. All three countries fulfilled their obligations by funding research according to evaluation results and leaving the internal distribution of funds to individual institutions. The UK associated autonomy directly with accountability and demanded that institutions report on the use of supplied funds. In New Zealand on the other hand, a somewhat compulsory evaluation process purports that low performance denotes lesser funding.

10.2.3 Purpose of research evaluation

It is necessary to briefly revisit theories in research evaluation prior to discussion of purposes in the three countries to promote understanding. As indicated in chapter 2, three descriptions are observed that explain why countries engage in comprehensive/systemic research evaluation processes. For Rip and van der Meulen (1996), the purpose of research evaluation is to explain the levels of performance of a country, which this study would classify under the diagnostic paradigm (see composite discussions in chapter 2), with the component of competence as the factor for explaining performance. Results of evaluation in this instance would be used to arrive at recommendations on how such levels can be improved, a characteristic of aggregation.

Weingart and Maasen (Whitley, 2007) relate the evaluation of research to the acquisition of information in relative standing for the sake of competition, a characteristic of the comparative purpose paradigm. Campbell (Shapira and Kuhlman, 2003) on the other hand, views research evaluation as a means to formulate policies to allocate research funds. Accordingly, the more the government subsidy the more the demand for research

evaluation to reduce the possibilities of abuse of funds said to be possible with curiosity-driven research. In most cases funds are allocated to researchers that display levels of competence (diagnostic) under the impression that they will use the funds effectively and efficiently, giving the evaluation exercise a prognostic overview. Competence in this instance is not an end but used as a means to judge how best funds are allocated.

The above information has been used hereunder to classify the purposes of the three models under discussion. Similar to the rationale and decision context, the three models display different purposes for research evaluation. In the Netherlands, the purpose had evolved from the encouragement of participation in research initially through funding allocations to the replacement of the 'give-away' model by the value-for-money principles. In the former input type model, institutions would obtain funds by clearly indicating needs. The ex-ante demands were later replaced in the value-for-money era by the need to prove products. At this stage, the need for quality promotion, encouragement of systemic discussions and the effective use of resources emerged.

Later on, there were clear intentions of the use of evaluation for the promotion of research quality (competence), a move that Rip and van der Meulen (1996) refer to as the improvement of the quality of science (intra-science quality) across the system of HE. The fact that no serious punitive measures for poor performance existed and no reward for excellence was granted indicates equal treatment and opportunities to all institutions. As evaluation became more internalized (in the institutions), there was a need to assist and encourage departments/research units to make better decisions about future research management. This practice is in agreement with Geuna and Martins' (2001) view of complementing the internal efforts of universities to improve quality. The level of competence across the system of HE shifted the focus of research evaluation, whereby international competitiveness and relevance have become the main objectives of research evaluation.

In the UK, on the other hand, selective allocation of resources provided the rationale for reducing government spending and reliance on government. Therefore, the purpose of

reducing state dependency was hidden behind accountability reasons. The objectives attached to the purpose are productivity (research output of good quality) and the ability to attract funding from other sources. Institutions would be funded on the basis of achievement, with those achieving high rates in research evaluation receiving the highest share of research funding. Therefore the existence of research active staff members determined the research strengths of an institution. This resulted in what is referred to as 'poaching' of staff members by the more financially able institutions in an attempt to stay relevant and attract more funds and a high rate of staff mobility was initiated.

Throughout the system, such movements were mainly internalized within the system of HE. With added efforts by government to assist institutions from which staff members have been poached, it is believed that over a period of time the country will achieve equilibrium and will have improved research skills and acquired knowledge of better quality across the system of HE to be able to acquire funds from other sources. The added assistance afforded to institutions that lost researchers is a unique feature of the UK system, which may assist in speeding up improvements and thus the acquisition of non-state resources. Otherwise, throughout the evaluation process, purpose in the UK remained 'to best allocate resources'.

Although the Dutch and the UK systems display different purposes for research evaluation, the component of quality based on international excellence in both models, emerged over a period of two decades and after immense effort. Unlike in the two other systems, it is not easy to arrive at the main purpose of research evaluation in New Zealand. A few very important aims can nonetheless be identified. In one instance, the increase in the quality of academic research that relates to Rip and van der Meulen's (1996) competence was mentioned as relevant for the improvement of the quality of teaching. This would mean that research evaluation would contribute to the improvement of science and thus research performance by "investing in excellence". Simultaneously, in line with the UK model, accountability for the use of state funds is also mentioned in New Zealand. This is related to excellence, but applied differently in that in New Zealand emphasis is on the reduction of funding for poorly performing institutions. This

requirement for accountability (participation through reward) is used as a tool to motivate institutions to review their research strategies. As already observed, the Dutch system intends to achieve this through an internal self-evaluation system.

Similar to the above two models (at their advanced stages), international competitiveness and the building of future profiles on this component is the purpose that is more visible in the evaluation results of the 2003 research evaluation exercise in New Zealand. This brought about the third component of the national standing in international competitiveness. The objective of identifying areas that warrant improvement in order to reach world-class levels is indicative of the awareness of global competitiveness. Since the exercise took place in the 21st century this behaviour is not surprising. The results of evaluation were also meant to serve as baseline information for comparisons with future evaluation results in the country. Data was used to compare subjects and to establish capacity levels in different institutions. The first exercise was a test of the strength of the country to undertake a comprehensive/large-scale research evaluation exercise. All these intentions are identifiable in this one (1) exercise.

Acquisition of baseline information, comparison of subjects (performance) and the weighing of strengths to conduct evaluation are reasonable reasons in the initial stages. Equally acceptable is the test for levels of international competitiveness, provided that the measures used are standardized and normalized. If all these intentions are addressed together the risk is that they will have a tendency to cloud one another. The purpose of allocating funds has gained momentum and has been successfully imposed, while the long-term intention to increase the quality of academic research may be viewed as being in the initial stages. If funding has to be based on quality productivity, which the exercise intends to increase, then the idea of reducing funding for poorly performing institutions may be seen as premature in this initial stage. Alternatively, lessons should have been learnt from the UK model and measures put into place to introduce institutions to the market economy for them to seek alternative funding and to provide them with well-planned opportunities for development.

In summary, the Dutch model is characterized by Rip's (Shapira and Kuhlmann, 2001) "competence-determination for improvement" and the UK is inclined to Campbell's (Shapira and Kuhlmann, 2003) accountability to the GUF funding idea. Rip and van der Meulen (1996) advise that benefits are better accruable when all role players have common interests. In the two models, relevance, which is usually attached to national needs, has been acknowledged. Although New Zealand ultimately utilizes results for funding, the model's multiple purposes oscillate between purposes of the above two models, showing characteristics of a plural system. This multiplicity is believed to create evaluation problems.

Table 10.1 A summary of the reasons for evaluation of the study models per purpose

Purpose	Dutch	UK	New Zealand
Summative (funding decisions/ resource allocation)	No	Yes	Yes
Formative purpose (improving science/ research)	Yes	Yes	Yes
Strategic (prioritizing of fields/ decision-making around staff/ recruitment)	Maybe	Yes	Maybe
Accountability	Initially	Yes	Yes
Resource allocation	No	Yes	Yes
Assurance of quality and excellence	Yes	Yes	Yes
Symbolic (recognition of reputation and prestige-rankings)	Partially (rate but do not rank)	Yes	Yes (rates of individuals aggregated to that of the institution)

10.2.4 Methodological issues

10.2.4.1 Design, approach and units of analysis

All the systems followed the systemic and comprehensive/large-scale evaluation design. Time and time again the literature refers to this as research assessment exercise or assessment of research performance. As stated, in the Dutch and UK systems, purpose and objectives have always been clearly outlined while the first exercise in New Zealand exhibits characteristics of feasibility or needs analysis study. The confusion results from the stated multiplicity of purposes in which each one of the purposes could constitute an own evaluation.

In The Netherlands different approaches appeared in the different periods of evaluation, wherein performance indicators used in the period 1982-1992 projected a more quantitative approach. Criticism of the reliance on non-descriptive data compelled the exercises to resort to both qualitative and quantitative approaches. This is correct in that data that has to inform developments cannot only be quantified (Cave *et al*, 1988; Johnes and Taylor 1990). The UK went through the same experience and whereas the quantity of research output per active staff (volume) was initially the main source of data, reviews diverted the exercise from volume determination to qualitative assessment.

New Zealand on the other hand, started with an attempt to use both approaches and directly referred to performance indicators as a strategy to collect data. Further discussion on the use of performance indicators will be provided later in this chapter. Notably, both the Dutch and the New Zealand models use two levels of data handling, the self-evaluation reporting and the evidence portfolios respectively as parts of the first level and the external peer review evaluation by panels as the second. The observable difference between the two models is that the Dutch model has reached a high level of formative evaluation through their internal system while in New Zealand the internal and external systems still have to be harmonized.

Both the Dutch and the UK models refer to departments/units of research as their units of analyses while New Zealand adopted the Hong Kong strategy and considered individual researchers on an institutional basis but resorted to assigning ratings to the institutions.

When one considers one of their purposes as revealing research strengths through the determination of world-class researchers, the choice of researchers as a unit of analysis does not come as a surprise. The assignment of ratings to individual institutions by the PBRF may be equated to ranking, which is expected to treat institutional aggregates as instruments for comparison. The effect of multiplicity of purposes on the unit of analysis has a bearing on the achievement of a purpose and as already stated multiplicity can cloud achievement of some of the intended objectives.

In the UK, research active staff members serve as sub-units of analysis with departments/units of research being awarded the points for their achievements. The Dutch system concentrates solely on the departments/units of research, which decide on which data to submit, in agreement with Weingart and Maasen (Whitley *et al*, 2007) who attribute research reputation to disciplines.

10.2.4.2 Target group and audience

All three systems target public HEIs although New Zealand also adds private training institutions. Whereas the calls for submissions of proposals/applications projects favour voluntary participation arrangements, the matter is handled differently in different countries. In The Netherlands it was initially intended that efforts by the VSNU to motivate participation would encourage faculty/research unit members to engage more in research. With the VSNU being composed of university scholars, public HEIs participate but departments/ research units reserve their right to select programmes for submission. In the UK, participation has been voluntary but state-funded institutions are compelled to be involved if they need to receive funding. The same compulsion occurred in New Zealand and is confirmed by the phasing out of other forms of subsidies. The difference between conditions in the UK and New Zealand is that the HEFCs (in the UK) also provide developmental funds and directly encourage institutions to seek external funding.

In the UK and New Zealand, governments play the major role, thus serving as the audience and therefore dictating purpose and measures used to achieve such purpose. In The Netherlands on the other hand the institutions are the role players and the exercise

allows the opinions of other stakeholders who may be the recipients of results to make inputs. Fortunately no earmarked funds are referred to in the exercises and thus decision-making is centralized (not fragmented to satisfy different audiences).

10.2.4.3 Variables /Criteria for evaluation

The component of competence featured most in the discussions of purpose of research evaluation above. Competence has been referred to as a means to determine the levels and standing of science in institutions or in a country. The component is also used as a means for allocation of funds. Whatever the purpose, it is important to have valid and reliable measures to determine the component because unreliable information may lead to incorrect decisions. Unfortunately, in research evaluation, according to Hemlin (1996) it may not be easy to obtain a reliable measure against which validity of results may be tested. Hemlin attributes this to the reliance on human judgment in peer reviews and other products of research (publications, citations and other measures are themselves products of peer review).

In the measurement of research in The Netherlands, different dimensions are used to explain competence while in the UK one dimension of quality is perceived. This results from the differences in the purposes of evaluation between the two countries. As a means for the allocation of resources, the UK classifies quality in relation to the scientific achievement of a product as revealed by Campbell (Shapira and Kuhlmann, 2003). Therefore, productivity of a set standard serves as an important measure in this instance. The Dutch system's intention is to acquire competence (as an end), thus the different dimensions to explain it. Quality of research according to the system relates to excellence, eminence and prominence and international standing. The quality of research facilities is also regarded in the measurement of quality. It stands to reason that subjective judgment will be used to measure quality as it is defined.

In all three countries productivity, which is the main criterion to explain quality in the UK, is based on numerical measures. These methods include bibliometrics, citations and

technometrics and others. Performance indicators are the main measures for data collection in New Zealand.

Responsiveness to needs and or missions, both academic and for social impact, is claimed by all models as relevant and effective. The UK aims at reducing state dependency (strategic relevance) through research evaluation while The Netherlands aims at the improvement of science. New Zealand on the other hand intends to, among other things, improve research quality for the sake of better teaching. Effectiveness is measured through both quantitative and qualitative measures. Efficiency in relation to the use of resources is an important dimension in the UK and therefore determines resource allocations.

The Dutch system also opted to assess viability, vitality and visibility dimensions. In this instance, international standing is regarded as a measure for future performance (Protocol, 2003-2009). Also important are the consistency of research programmes, availability of resources, and flexibility and ability of a research group to initiate new programmes and discontinue non-viable ones and professionalism. All these variables are used to determine the prognostic dimensions. Qualitative measures are used for assessment.

Although Campbell (Shapira and Kuhlmann, 2003) believes in ex-post data for institutional research evaluation exercises, some of the dimensions (especially viability and vitality) need the study of processes and may not necessarily be covered through ex-post information only. When evaluation is linked to funding, ex-post data in a systemic cycle will detect changes in research performance. All dimensions are observed in the three models (table 10.2)

Table 10.2 Criteria of research evaluation as used in the stated models

Dimension/criterion	Netherlands	United Kingdom	New Zealand
Quantity	Yes	Yes	Yes
Quality	Yes (research output and system)	Yes (research output)	Yes (research output)
Productivity/efficiency	Yes	Yes	Yes
Effectiveness/relevance	Yes	Yes	Yes
Viability/vitality	Yes	Yes (put more resources for improvement)	Yes

10.2.4.4 Assessment methods

To understand the methods used in the three exercises, background is necessary. Martin and Irvine (1983) acknowledge the complexity of research evaluation. To ensure some level of dependability, the purpose of evaluation has to guide the activities of evaluation. The purpose will then lead towards some decision context as a frame of operation.

Verkleij (1998) believes that the decision context serves as a framework within which methods of evaluation can be matched to purpose. For example where fund allocation is the main determining factor, evaluation is conducted for policy reasons. In this way, publications, which are referred to by others as audience-inclined measures, are regarded as objective and easy-to-use-measures and these would be popular in the UK and New Zealand. Under such conditions, quality justifications relate to “publishing in prestige journals”.

When the decision context relates more to intra-science competence (or what may be referred to as “good” research in the South African context), preference by others (scholarly visibility), displayed through citations and judgment by experts (peer review) both take the lead. The Dutch system is a typical example of this practice, with the latter measure used for quality improvement, depending on the expertise levels of the peers that provide advice.

Methodological approaches are also important for the choice of methods. For example, the Dutch and New Zealand exercises use both the self-evaluation and external evaluation systems. In the initial stages, in the 1980s and 1990s, the first level in The Netherlands, self-evaluation was merely a process of gathering information through the compilation of reports for external evaluation. It is not surprising that Campbell (Shapira and Kuhlmann, 2003) perceived self-evaluation as more inclined to monitoring through the use of quantitative measures (performance indicators). With time, data changed from staff profiles, inputs, outputs aims and objectives to research overview.

In the second level of the Dutch system, peer reviews were initially based on document analysis to determine ex-ante data (research needs). The period between 1982 and 1992 can be referred to as the first level of publication-based evaluation where reviews were based on input, throughput and output data. This process of using a group of experts (panels) to verify and analyze data submitted by the universities, moved from anonymity to transparency.

The assessment of submitted documents was augmented with site visits using observations and interviews as strategies for verification and triangulation of evidence. In the early 21st century, the system converted self-evaluation into a formative evaluation process in the form of annual and three-yearly reports. These are respectively regarded as the initial and intermediate levels of the six (6) year evaluation cycle. In the intermediate level (three years), which is referred to as the mid-term, the external reviewers are utilized in the self-evaluation to assist in the building up of valid and reliable reports. This move away from ordinarily collecting and submitting data is explained by Weingart and Maasen (Whitley *et al*, 2007) as institutional stimulation of peer review activities. The move is a step towards independence and the formalization of self-regulation, an outcome of the process of research evaluation, which reduces reliance on the VSNU, re-affirming the autonomy of universities. It should be recalled that the Dutch system allows universities to decide on programmes to be evaluated and that submissions are standardized across all programmes.

The evidence portfolios in New Zealand contain research output, recognizable achievements (peer esteem) supervision of research and attraction of grants. The four ‘dimensions’ are more quantitative and include both input and output measures. Through the attraction of grants some prognostic ideas may be reached that the researchers are reputable and have the capacity to produce output. The other measures are diagnostic in that judgments are based on outputs.

The UK on the other hand, fits into the second level of evaluation wherein data is derived from departments/research units and handed over to panels for peer review. Initially such information included input data (grants awarded), accounts of research activities, research output (publications) and future research plans. The latter is highly prognostic and as Rossi and Freeman (1993) advise future plans cannot be guaranteed. With experience and consultations data have been improved and more qualified to create consistency across the panels. It should be recalled that as a way of reducing state dependency, proof of independency, related to the ability to attract funds, the input measure is highly preferred in the UK.

As stated before, quality based on publication in “top journals”, became more important than volume of output. Such publications are subjected to peer reviews and according to Rudd (1988) and Hemlin (1996) are said to reflect quality. In this system reviewers were provided with data and document analysis strategies were used to formulate information into analyzable data. As the submitted document, be it publications or citations, is a performance indicators and thus easy to analyze, Westerheijden (1999) warns that PIs cannot be used to judge quality. Weingart (2003) suggests the use of PIs as a supplement to other methods. Site visits to institutions/research units have not been undertaken but more stringent methods have lately been used to ‘purify/clean-up’ data (by verifying with the institutions) prior to reduction and analysis. Similar to the other two exercises, panels are grouped into disciplines. The reliability of the data may be questioned since verification is still conducted with the institutions that submitted it.

In New Zealand, the TEC use evidence portfolios, information on degree completion (as supervised by individuals) and evidence of attracting external funding as data for external reviews (by panels). As indicated, these requirements are provided at subject level on the basis of individual researchers. The problem with the New Zealand exercise is that the internal evaluation by the TEOs that would have added information to their portfolios (which they analyze) that was not made available to the TEC. The latter is the body responsible for external peer evaluation. Therefore, there have been discrepancies between the two data sets and thus the results.

10.2.4.5 Analysis

In The Netherlands, Protocols (the guideline documents) guide the entire system of evaluation. As already indicated, the following four (4) evaluation criteria (dimensions/components) were referred to in the analysis of data; quality (defined as eminence to perform at prominent/international standards), productivity (publications, citations and output in terms of input), relevance (professional and socio-economic impact), viability/vitality and feasibility. Ratings would be allocated to departments/research units according to performance. In the UK on the other hand, the attainment of levels of excellence was used as the main criterion for comparisons across disciplines, benchmarked against international standards. Similar to processes in the other two countries, panel criteria/quality categories would be allowed to accommodate differences. A standard definition of 'research' was also introduced to guide panels.

The PBRF guidelines were used in New Zealand to analyze data and cross-referencing was allowed. Rates were awarded to individual researchers and cumulatively used to determine funding allocations and competency levels. In all three models, advisory services such as from professional bodies and/or international peers are utilized.

10.2.4.6 Reporting and dissemination of results

Initially the Dutch and the UK systems would not freely discuss results with institutions/departments/research units. As the research evaluation processes evolved reports were provided and discussed in some instances. Some of the panels preferred to

share information directly with departments and explain the allocated ratings. Formal reporting is handled by the research organizations/funding bodies (VSNU/HEFCs) in The Netherlands and the UK and publications/reports show the ratings. In the case of The Netherlands in particular, results are categorized into the dimensions/components/criteria used in the analysis.

Results of the first large-scale research evaluation in New Zealand were not published, pending the availability of publishing regulations. Reports were provided to institutions since they were necessary for the determination/allocation of funds.

10.2.4.7 Ethical issues

Reporting on the results of evaluation is mainly quantitative although explanations of ratings are provided. Therefore, details of vision, mission and future plans are not necessarily revealed. Publications are guided by legislation and as stated, in New Zealand publications were withheld as a result of the absence of a publishing policy. Individual institutional reports, in cases where feedback is provided by panels are submitted and discussed for purposes of improvements. Where necessary, information in all three systems was handled with confidentiality.

Several consultations with research end-users and professional bodies are mentioned in the Dutch and the UK exercises that contributed to improvements. This shows concern for contributions made by stakeholders. To improve on transparency, criteria for evaluation together with the members of panels were published.

10.2.5 Benefits and advantages accrued

It has been stated that comprehensive research evaluation has its own advantages over the pluralized system. Campbell (Shapira and Kuhlmann, 2003) has emphasized the need for a comprehensive ex-post research evaluation for the entire collective of HEIs in a country. Although different authors have different reasons to motivate for a comprehensive research evaluation, it is notable that all agree that this Type A model

inclusively becomes part of the effort to strengthen national research systems (Rip and van der Meulen, 1996; Geuna and Martin, 2001; Campbell in Shapira and Kuhlmann, 2003).

Although the pluralized (Type B) model is said to promote steering tendencies and thus dependency of universities on state funds, the model also has advantages. These include allowing differentiation of funding, the creation of interdependency for the benefit of national collaboration and allowance for additional goals (Rip and van der Meulen, 1996; Geuna and Martin, 2001). Information on the conversion from Type B to Type A when intensity improves and the ambition to compete globally exists seem to imply that Type A model may be the direction to follow. Precautions are nevertheless necessary if this option is seen as the answer to national needs.

A comprehensive/large-scale research evaluation trend has undoubtedly been set in motion in the three countries and other benefits have accrued for different reasons. For example, in The Netherlands in the early 1980s evaluation encouraged participation in research. The introduction of external evaluation and the face-to-face encounters with the external reviewers also made a significant contribution to acceptance of the processes, and this was further improved by the friendly non-punitive approach of the panel members. The VSNU further extended peer review services for internal evaluation, a process that will promote self-regulation for internal quality improvement.

The UK boasts of having created an awareness of research evaluation countrywide, which would improve systems of research management in institutions and the funding bodies. In all three models discussed, evaluation also motivated institutions and research units to solicit funds from other sources and results of the evaluation exercises were used as motivations to attract more funds. Results also revealed areas that warranted improvement and provided information on the direction for resource allocations.

All three countries refer to international legitimacy and in The Netherlands and the UK preparations towards the achievement of this component over a period of two (2) decades

have resulted in a series of evolving strategies. This applies both at the levels of research and experience in research evaluation. The involvement of British experts in the planning and evaluation of the New Zealand exercise is a typical example of their achievement of world-class excellence and is legitimizing the experts as being world-class. Whatever the achievements, these would not have accrued from effects of research evaluation. It is always good for a country to know its research strength especially since research is relevant for national innovation programmes and processes.

Other benefits involve the grouping of research and the formation of research schools, which occurred as a result of the need to form units of research (panel disciplines) for evaluation. These were reported in both the Dutch and the UK systems. Early preparations for evaluations, continuous improvement of plans and experience gained by panel members, all contributed positively to the development of research. Of utmost importance to the UK and New Zealand, the evolution processes enabled selective allocation of state funds and in the UK, reduction of reliance on funds. The Dutch HE system improved on what they refer to as “quality of research”. The latter outcome has also been observed in the UK and is intended in New Zealand.

10.2.6 Problems and challenges encountered

Not as much was reported on challenges in the Dutch system as was in the other two systems. The fact that the VSNU in the Netherlands was a body formed by universities may have resulted in continuous feedback and therefore continuous improvements leading to the understanding of needs and to acceptance by the institutions. Inputs would be made and challenges immediately attended to. Alternatively, the HE institutions might have been so desperate for autonomy that they chose not to do away with the programme. This does not imply that there were no complaints or criticisms but rather that the institutions supported it.

In the early stages, The Netherlands experienced methodological problems, especially with the use of ex-ante data, with the allocation of funds for research made without proof of completion. This was solved by the use of ex-post data based on inputs. Even with

this system, evaluation results, together with the expected improvements were affected by the experts' (Weingart and Maasen in Whitley *et al*, 2007) cognitive-distance within study areas, which may result in non-correlation of judgments.

Other inherent methodological challenges were also reported. For example, output in The Netherlands is affected by the problem of a 'minority language'. In The Netherlands the monitoring processes by the Inspectorate made minimal contribution as there were delays in the compilation of the meta-analysis reports. This responsibility has since been left in the hands of a quality committee with better levels of responsibility. Other problems experienced in the 1980s and the early 1990s were attended to in the subsequent exercises. Improvements nevertheless made evaluation more complex, more time-consuming and more expensive and this led the VSNU to consider intensifying self-evaluation.

The UK experienced problems of acceptance in the initial stages especially because of the steering powers of government through the intermediary funding bodies and the purpose attached to evaluation. Poor acceptance was also attached to concerns that the programme was giving research more attention and recognition and the other core functions, especially teaching were suffering as a result. Over the two decades, those who resisted have either given up, or contributions made by the RAE have become acceptable.

In this country also, methodological problems such as the use of input measures, the lack of clarity of evaluation criteria, anonymity of assessors, methods used and lack of standardization of criteria across panels were also experienced in the earlier stages and taken care of through improvements. The lack of standardization brought about problems of comparison and interpretation. In the late 1980s, problems were inherent methodological ones such as time schedules (evaluation period), the effect of the use of output, definition and the types of outputs and inter-subject differences. Inherent problems of expert judgment also affected the UK system. The halo, the Matthew and

the Oxbridge effects exhibiting the experiences and scholarly age of participants cannot be ignored.

Other non-methodological problems were experienced in the UK. For example, the abolition of the binary divide in the early 1990s placed institutions of different characteristics and research experience on the same competitive levels for resources. Comparisons were made without considering the differences in landscape and the fact that polytechnic and college staff members were not familiar with proceedings was not sufficiently considered. The drive to acquire more funds also caused instability in the HEIs and encouraged poaching of research active staff members by able institutions. As already mentioned, poaching increased staff mobility. At this level, the disciplinary approach to evaluation worked against inter-disciplinary research output. Some subjects were not fitted in to any of the disciplines and this caused confusion about the units of evaluation (disciplines/subject areas).

What complicated the system of research evaluation is the absence of correlation between the RAE and other research measures. Whereas the evaluation results were meant to guide the allocation of funds, the exercise would be followed by funding delays and the lack of provision of enough feedback for future improvements, which was time-consuming and uneconomical.

When quality is related to national and international standards this exposes researchers' work to peers, and the researcher may be seen as more visible. In the UK, this was also affected by inconsistencies in the units of research. In some cases quality was rated on an individual basis while other panels used research groups. This also affected work referred to other panels for cross-referencing wherein criteria of assessment would be different. Problems of data collection were also experienced with less space allocated for additional information on the data collection instruments. The RAE 2001 experienced a decline in the number of submissions with an accompanying increase in expenditure, the latter resulting from improvements in the process of evaluation (as it occurred in The Netherlands).

Serious problems were encountered with the New Zealand exercise of 2003. There exist considerable discrepancies between the ratings assigned by the TEOs and those assigned by the TEC, whereas the TEOs were feeding the TEC with data. This problem is blamed on planning and the use of additional data by the TEOs, which were not made available to the TEC. Those that are eligible but did not submit were equated to poor performers and allocated an R rating. This indirectly endorsed the notion of compulsory participation, which seems premature for a process that is still in the testing stage (acquiring base-line data). It is also alleged that the multiplicity of evaluation criteria made it difficult for researchers to obtain A ratings (the highest grade). The additional variables (citation and research output) nevertheless give the New Zealand exercise a trademark even though the use of performance indicators is criticized.

In general, research performance revealed by the process was rated below 50% (2.59 out of 10) a score that would probably be lower if all the 45 eligible institutions participated. Inconsistencies between grades were also reported in New Zealand. With no proper guidelines on submissions and software problems, some portfolios were not complete. It is also alleged that the portfolios were not of professional standards. Institutions that have merged and non-university institutions performed poorly as it would be expected. Other problems involved gaps between good and poor performance. For example, of those who participated, 39.6% did not make the C grade. The evaluation schedules were tight while the period for data inclusion (six years) is criticized for excluding important data. Data for part-time staff members were also excluded. It is not very clear how units of analysis can serve as a good indicator for research strength (quality *versus* numbers) as this has been used in these systems.

10.2.7 Good practices

The three models exhibit examples of good practice from which new ideas can be conceptualized. For example, in The Netherlands, no programme was declared inefficient or excellent (no punishment, no reward), and this was meant to encourage participation. In The Netherlands the system is owned and funded by a collective of

universities in which accountability (for socio-economic and professional reasons) is the driving force. Institutions decide on which programmes to submit and weaker programmes receive warnings and advice to implement improvements to avoid sanctions. This has been made possible by the fact that feedback was given on time and the fact that results were followed by advice on how to improve and enhance quality research.

As indicated, the general Protocol serves as a standard frame of reference and similar programmes are grouped and evaluated together. Self-evaluation reports are augmented by a qualitative approach to verify data and to detect the effects of confounding factors. These on-site visits by local and international experts to observe and interview researchers and research managers make the Dutch research evaluation process unique and indicate high levels of commitment to quality improvement. The VSNU also includes international experts that have knowledge of the Dutch language to avoid compromises. To emphasize this, the panel chairpersons have to be of Dutch origin. It should also be noted that institutions have an opportunity to select panel members through the involvement of the Deans in a more transparent and democratic way. Transparency is also revealed by discussions of Protocols with institutions.

The Dutch model allows panels to consult with other experts on multi- and interdisciplinary research. In the 21st century after reviews for improvement of processes, meta-analysis (previously conducted by the government inspectorate) became the responsibility of research organizations (with the help of a quality assurance body) and this may be regarded as an improved level of autonomy. Publishing of results is a good sign of public accountability.

In the UK, HEIs are state-funded but academic responsibilities and self-regulation lies with the institutions. The preservation of autonomy was further increased when state funds were reduced and the market economy introduced, forcing institutions to fend for themselves. Improvements in the evaluation exercises were boosted through reviews in response to criticism and changes were implemented in the cycles that would follow. For example, there were improvements on the type of data, from volume to quality, with

basic and applied research submitted together. Improvements were also observed in the manner of submission (software) and on strategies of evaluation with subject specific procedures based on the general framework. Quality of judgment was measured against national and international standards.

The appointment of reviewers was tightened to promote quality and such appointments are made by funding organizations. Panels were provided with raw data and end-users were occasionally consulted. Institutions received support through the use of websites, helpdesks and contact persons.

As evaluation evolved and institutions became more familiar with requirements, commitment also increased and researchers submitted more reliable data. This was improved by the early publication of guiding documents and responses to suggestions. In 1996 for example, the adjusted plans were piloted, resulting in the use of a common, standardized frame of reference and definition of research.

Similar to practices in The Netherlands, some of the panels provided feedback to departments/research units and results were published as soon as they were available. Unlike the moderate attitude of the Dutch model, The UK went to extremes to reward excellence. Institutions that were more active in research received an additional 3% for further development. This would undoubtedly increase the performance gap among HEIs. To prevent this, the system also provided development funds to those institutions that have been 'robbed' of their researchers. Such institutions are thus encouraged to improve on research strategic plans, missions and pay attention to the relevance of research. Although this has long-term effects, it is believed that in the long run this may serve as a country-wide research development strategy.

In the Dutch model, poor performing programmes are given advice and given the responsibility of showing improvement or losing the programme. Whichever way, both evaluation exercises have succeeded in improving research performance in their HEIs and the HE system as a whole, to a level, which others refer to as excellence and

international standards. It will not be surprising if ‘international standards’ are equated to their standards, especially now that their systems are used as benchmarks and their researchers as international/overseas experts by other countries. This gives the comprehensive and systemic evaluation exercises credit, and even though benefits are observable over some period of time, the inclusiveness of the entire system of HE provides a general and national improvement of research in HEIs. Also observable is the equalizing phenomenon of standards in HEIs, which may take long periods to accomplish especially in the UK.

The systems give full autonomy to panels and support them through the use of sub-panels composed of research end-users and members of the professional bodies who serve as consultants to the system to promote quality and relevance. The UK intends to reduce the units of research by reducing the sub-areas.

As a new development, a few good practices can be observed from the New Zealand system. Most of those are similar to practices in the other two countries. Practices such as the confirmation of data, briefing of panels prior to evaluation, the use of guideline documents, the use of international experts, cross referencing, awarding autonomy to panels and the use of special advisors are good examples. The system attempted, in its planning, to reduce mistakes already experienced by others. Similar to practices in the other two programmes (systems), panel members do not review colleagues’ work. New Zealand created yet another trademark by awarding benefits to the university that lists former employees in cases where a staff member is poached twelve (12) months prior to the evaluation period (census).

10.2.8 Conclusion

Campbell (Shapira and Kuhlman, 2003) states that comprehensive institutional research evaluation is usually ex-post. The three models discussed in this study (the Dutch, the UK and the New Zealand exercises) serve as typical examples of systemic/comprehensive/large-scale ex-post research evaluation models. Countries that use pluralized types of research evaluation such as Germany and Australia were not

discussed because of their tendency to simulate the countries involved in comprehensive evaluation, especially the UK and The Netherlands, as observed by Geuna and Martins (2001). Pluralistic systems show tendencies of evolving into comprehensive ones. In all three exercises referred to, evaluation encompassed all public HEIs.

Similar to all research evaluation processes, comprehensive research evaluation of institutions may be undertaken at different levels as governed by different purposes. In the UK for example, the main reason is accountability to government on spending, for the purpose of reducing dependency on government. Results are needed for the allocation of funds, that is, outputs determine future inputs. In the Netherlands on the other hand, the move is towards the improvement of the quality of research in institutions across the system (Sizer, 1988; Verkleij, 1998). In both cases, comprehensive/large-scale evaluation is conducted on all submitted research efforts in public HEIs over a set period of time. Both countries use a combination of research evaluation methods, guided by the decision context at the time of evaluation. The use of methods thus evolves with the evolving system of evaluation. The methods used under such conditions are also project-specific, with output measure dominating and directing the process of evaluate.

As has been stated previously in this chapter, governments may resort to intermediary bodies in an attempt to reduce government steering powers and thus involve academics (as experts) in their evaluation processes. Even under these conditions, governments' purposes continue to give guidance to the use of evaluation measures.

In conclusion, it is observed that problems encountered with the sole use of each measure (the absence of a fully reliable and valid method) seem to dictate the use of a combination of the methods. This is despite Campbell's (Shapira and Kuhlmann, 2003) suggestion that evaluation can be conducted without indicators. Weingart (2003) prefers the use of a combination of methods (peer reviews and indicators) for the methods to supplement each other in minimizing confounding factors. It is hoped that either the evolutionary processes of research evaluation or science will lead to the discovery of more reliable methods of data collection for the research evaluation exercises.

The Dutch and the UK built these experiences over a period of two (2) decades through trial and error, pilots, and adjustments and their programmes are still being improved. One may easily be tempted to imagine that copying them would be easier as new programmes would avoid their mistakes. The New Zealand model is a typical example of a new programme, which benchmarked to avoid mishap, but ended up with problems. Most importantly it has not been very clear what their main evaluation purpose is, which makes it difficult to know whether accrued achievements are anywhere near to intended outcomes or not. Otherwise, the process has been initiated and the strength to undertake a large-scale research evaluation tested. More can be learned from their successes and the failures and experiences of the other older models can serve as a foundation for future thinking and planning of a country's comprehensive/large-scale research evaluation for national improvement of research.

10.3 Synthesis of the South African efforts of research evaluation

10.3.1 Introduction

Discussions in this sub-section show some differences in the historic drive towards the initiation and sustenance of the processes that make up the pluralized evaluation of research in the South African universities. As a result of the historic differences of the efforts and different levels of importance at points in time, this study tracked events back to different periods of the efforts under study. For example, although the system of HE subsidy by the department (block-grant subsidy) has a much older history than the FRD rating system, the latter has been traced further back as it was necessary to understand the FRD's origin and rationale. This led to the tracking the FRD's history as far back as the 1940s while that of the department was only tracked back to the 1980s. The CHE on the other hand, is an initiative of the new system of government (post-1994) and therefore has a recent history.

10.3.2 Historical background and rationale for the existing research evaluation efforts

The histories of the systems exhibit concerns of the time that influenced the continuous, and sometimes, changing rationales of the systems of evaluation. The FRD rating system, for example, was conceived in the early 1980s as a device to improve strategic research by encouraging more basic research. The new programme was introduced to “maintain good research” despite international isolation. All these were government initiated and because it was during the apartheid era, only the HAIs were consulted. These were the main beneficiaries of the initiative. One may conclude that from inception, the programme contributed to the disparities that still exist between the HAIs and the HDIs.

Later, the process was negatively affected by resource constraints and then by the change in the political dispensation in the country. To a large extent, the two factors marked the turning point of the programme and downplayed the intentions of the rating system, which relied heavily on government funding for its sustainability. The rating system was detached from funding in the late 1990s with the establishment of the NRF, leaving prestige as the only motivating factor for voluntarily participating researchers to reapply for rating. Funding was ‘re-linked’ in 2007.

The fragmented system of HE during the apartheid era, affected the block-grant/GUF system of universities in their levels of development. Reporting to different departments of education exposed universities to different policies that led to the inequitable distribution of resources. This categorized the HE system in the country into two groups referred to as the HAIs and the HDIs. Besides the possible differences in the allocation of funds, the HAIs were allowed to retain reserves and utilize them to pursue developments they deemed necessary. This is one factor that exaggerated the differences in levels of scholarship (research productivity) between the two categories of universities. The uneven distribution and all the other negative effects led to the unevenness of research output and differences in postgraduate education. It does not come as a surprise that a high level of research output and highly rated researchers are still found in the HAIs even when only one HE system exists.

This unevenness also influenced planning abilities (differences) between HAIs and HDIs and together with the internal autonomy of institutions this resulted in different prioritization of programmes. This became more visible when SAPSE allocations were reduced as a result of sanctions. The discretion to utilize allocated subsidies in some universities, especially the HDIs led to cross-subsidization of programmes, resulting in a compromise of what may be referred to as non-prioritized programmes, with research and research training falling within this category.

The DoE-funded (block grant) research and the evaluation thereof were also affected by funding insufficiencies almost simultaneously with the FRD/NRF rating system. As a result, the funding formulae for university subsidies had to be adjusted as early as the 1990s.

Despite all the challenges, the departmental support for university research (block-grants) allowed for departmental and group research and provided a better opportunity to ‘groom’ new researchers. It is these researchers that would volunteer for the FRD rating.

10.3.3 Research and the research system in South Africa’s new dispensation

Research evaluation models are determined by the landscape of research. This was confirmed in chapter 2 wherein discussions displayed the importance of evaluation in understanding the functions of research and research systems. Understanding and conceptualizing South African research require a clear understanding of the South African research landscape. The following information reveals some of the challenges experienced after the apartheid era.

The new government introduced new laws in the mid- to late 1990s that would align services with new policies. With the transformation imperative, redress of inequalities would be equated to the achievement of equilibrium across the HE system. This would be achieved through the bridging of imbalances of race and gender, producing a high-skilled and competent generation of intellectuals for research across all universities and strengthening weaker universities.

Although it is arguable that good researchers of international standing are able to contribute to national needs, such as commissioned and contract research, the demands of international competitiveness may be compromised by concentration on these national needs. This is affected by the fact that productivity and quality (for international comparison) are based on publications, leaving little room for recognition of other forms of research output. Therefore a compromise may be made on international competitiveness if good researchers are charged with the responsibilities of redress unless such contributions are regarded as criteria for evaluation or some standards are set to attract benefits. By the same token, responsiveness of research to community needs may suffer as a result of the demands of international competitiveness. This contradiction may be the reason why some programmes (of evaluation) will have a tendency to fulfill one imperative at the expense of the other. Unless purpose is clearly stipulated and important objectives prioritized, some of the imperatives may be difficult to achieve.

Therefore the main challenge for South Africa is to strike a balance between strengthening the weaker universities and maintaining the standards of well-performing institutions. This dilemma of dual but contradictory expectations to improve research and create a balance in productivity levels across the system of universities and across race and gender creates problems for the concentration of government support as stated by the White Paper 3(1) (1997a:49). On the one hand, there is a need to “concentrate funds selectively on competitive basis” and on the other, the attempt to improve research of ‘weaker’ universities. The latter is intended by government to be achieved “incrementally over some (undefined) period of time”. There are fears that the longer the period the more difficult it may be to deal with inequalities and this may further widen the gap between the still better performing HAIs and the still underperforming HDIs as the former focus on further improvements at competitive levels. In short, this reveals the contradiction between achievement and maintenance of standards and bridging the inequality gap. As has been argued, attempts made through the DHET research development funding system have yielded disappointing results.

A model for research evaluation for South African universities should take this tension into account and attempt to strike a balance between the two intentions. At this point in time a brief review of other models is necessary to improve perception and conceptualization of the situation. Prior to the 2003 exercise, there were inequalities in the New Zealand HE system. Their system approached the problem through differentiation, penalizing institutions that did not have “good” researchers and were thus performing sub-minimally. It may be premature to reflect on the outcomes of the approach, although the 2006 exercise claims general improvements in research. The New Zealand government introduced the PBRF to manage the evaluation process. The system identifies good research (output) and aggregates it to represent a university.

The UK, on the other hand, needed to reduce dependency on state resources. All universities are expected by government (through the intermediary bodies, the HEFCS) to perform or loose support (perish). There is no intention to concentrate on poor performance. This is besides the abolition of the binary divide that took place in the early 1990s. Those universities that made attempts but lost good researchers as a result of poaching are compensated with funds to improve capacity, enabling research capacity development to continue across all universities. The system also uses research staff as sub-units and then aggregates results to make up the university total.

The Dutch universities initially encouraged research participation across the system and then moved towards maximizing the quality of research products in order to prove viable competence and to claim autonomy. Benefits have been accrued over a period of twenty (20) years, after which individual institutions are now encouraged to internalize the system of research evaluation. A body formulated of academics (VSNU) has been responsible for the system of evaluation.

While the South African national imperatives have been articulated, they are addressed at different levels in a way that is not co-ordinated even in the new dispensation. The efforts still operate at different levels with different units of analysis. For example, although the CHE’s monitoring and evaluation is meant to track transformation of the

system, the institutional audit focuses on individual institutional performance, the same as in large-scale research evaluation. The dependence of the monitoring and evaluation system on secondary information may obscure other activities that have not been researched. Otherwise the system may, in its operations include the meta-analysis for research evaluation.

Since the NRF maintained the element of prestige even when it was detached from funding, highly-rated researchers would still be found in the HAIs and this status quo seems to be lasting longer than expected. This is because the new proposal-based funding that replaced the person-based funding system of the FRD does not seem to succeed in attracting highly-rated researchers from these institutions to work within groups of emerging researchers from the HDIs, as this might compromise their research productivity. The intention of collaboration between the experts and the less skilled for the purpose of capacity building does not seem to go as planned, perhaps because participation is voluntary.

Despite the above-mentioned tensions, the new funding formulae (block grant) cannot avoid promoting excellence and public accountability. Assessment has shifted from making block grants (unrestricted) allocations (the use of which is still left to the discretion of universities) to relate more to monitoring of the use of public funds. This is because government involvement is obviously still predominant as the main source of funding.

The DHET research reward system has been retained and expanded through the new Policy and Procedures for Measurement of Research Output of Public Higher Education Institutions of 2003. As stated, the reward system also serves those researchers that are already productive. Therefore, any possible increase in the number of researchers together with the recent (2007) realignment of rating with funding would increase participation and most likely also increase the number of rated, competitive researchers. It is not clear how this would encourage “young researchers” and those in the underperforming institutions (HDIs).

The two identified systems of evaluation, the FRD/NRF rating system and the Department of Education and Training research reward system originated from apartheid system and contributed to the arrangement that created advantaged and disadvantaged HEIs. The two systems of research evaluation still (unintentionally) benefit those that were previously advantaged. The rating system further qualifies researchers into different levels, perpetuating inequalities and promoting elitism. The system does so without considering the causes of institutional and research performance differences. Therefore, evaluation processes have not altered the former government's intentions or succeeded in fulfilling the objectives of the new HE system (now also including training).

Attempts by the two systems of evaluation to bridge inequalities in research have not been successful. If such systems were considered as measures of a compulsory research evaluation basis, the negative effects would be worse. There are no observable indications of how both the NRF rating and the DoE research reward systems would, even with the efforts and plans put in place, close the disparity gap. High ratings and rewards are still received by the same apartheid-favoured individuals (or products thereof). Therefore, the research evaluation efforts in South Africa are not able to reverse disparities. Despite this, some institutions internally use the rates for promotions, appointments and even to reward excellence. Institutions that favour elitist differentiation also use the system to promote the idea.

The two efforts at research evaluation (NRF rating and DHET research reward systems) are suited to a situation in which parity already exists and research performance is equivalent across the entire system. In such a situation, differentiation of universities and competition are possible. Both the NRF rating and the DHET reward systems intentionally differentiate between skilled researchers and the unskilled, with the NRF displaying its elitist classifications by allocating grades to researchers. The existing differences/disparities in the HE system is not suitable for differentiation if the imperatives of equality have to be achieved.

The segregating phenomenon has potential to perpetuate inequalities in research productivity between academic staff members and universities that have had opportunities for research and research training and those that did not. This conclusion is motivated by the fact that rated researchers prefer to retain their rates and well-established research institutions intend to generate funds through the reward system rather than concentrating on the capacity building of less privileged and less productive researchers. No research evaluation effort exists in South Africa that determines and upholds South Africa's international standing and at the same time is able to strengthen weaker universities. The three exercises/models of research evaluation referred to in this study (Dutch, UK and New Zealand) reveal that national agendas are best addressed through a comprehensive system, a characteristic observable only in the CHE programmes in South Africa.

The CHE system on the other hand, is an invention of the new government system and structured to facilitate transformation imperatives. Although the entire system was aimed at ensuring quality across the HE system, especially core functions in the universities, other dimensions such as relevance (responsiveness to national needs), productivity and accountability are also addressed. All these have to be observed alongside the efforts of equalizing the system of HE (transformation). These demands may be the reason why the Council ended up with different frameworks/systems (the institutional audit and the monitoring and evaluation systems).

The CHE has not indicated why its systems have to be independent. Perhaps this was conveniently done to give specific alterations or additive information to each individual focus area. How and when the process will be customized to be a single process inclusive of all intentions has not been suggested. Quality and productivity are meant to address excellence and maintain standards.

It is not clear how under-performing institutions will be compelled to arrive at the expected levels. It is notable that besides the guide on the internal management of research, the CHE does not address the core functions (research and teaching) as separate

entities as is done in other countries. In The Netherlands, the UK and New Zealand, research evaluation is said to be separated from that of teaching in order to improve the attention and credibility of the entities.

10.3.4 Rationale and purposes for research evaluation in South Africa – the new dispensation

10.3.4.1 Introduction

Transformation of HE in South Africa forms the political crux for the development of the entire higher education and training system. The important transformation imperatives to be addressed are: the reduction of disparities within and between universities (and further education and training centres which are not included in this study), international competitiveness (standing) and responsiveness to national needs. These imperatives should guide the landscape of research and research systems in HE. Discussions therefore compare the purposes of the discussed efforts with the imperatives and use information on the international models as guiding principles. The following tensions exist that have to be clarified before the purpose of evaluation is conceptualized: the redress of inequalities *versus* excellence, selection of top institutions *versus* broad-base capacity building and world-class competitiveness *versus* response to national needs. These are discussed below.

10.3.4.2 Redress *versus* excellence

The idea of concentrating resources in areas where there is ‘demonstrable research capacity’ if not well conceptualized may counter that of ‘strengthening weaker universities’ and this is true when universities are the main focus of funding. It would be expected that either the South African efforts should address both requirements on an individual basis or do so in a collective form. The arrangement of concentrating resources in research-strong universities is unintentionally promoted by the rating and the DHET research award systems in that the two programmes concentrate on producing ‘good and productive researchers’. Even though the programmes allow all interested researchers to participate, it is obvious that only those that have capacity will be involved. It is for this reason that the rating system is labeled as elitist.

The high competition for rating or for funding from the DHET's research reward system does not facilitate capacity building for new entries. This should not be understood to imply that the DHET and the NRF have not put in place other efforts of capacity building but refer to activities of evaluation of the systems. Presumably, as long as resource allocation is a compelling reason for evaluation in these programmes and the unit of analysis is an individual researcher or a research output as it is in the stated efforts, it will be difficult to bridge the performance gap between the two ends of competency unless other compelling measures are put in place. Although the rewards for outputs of members of universities are aggregated into a university benefits it should be noted that analysis in the DHET award system concentrate on individual outputs as submits (same as in the UK model). The UK and New Zealand systems are based on resource allocation (financial benefit) yet both systems base their support on institutions (as units of analysis). The compensation system of the UK serves as an example of compelling measures.

Both the NRF and the DHET research reward systems efforts claim to aim at upholding good research for quality improvement and are suitable to a situation in which researchers are already better established across the entire system of HE for all universities to benefit. Universities that were previously disadvantaged may continue to 'miss out' and transformation may in this way be flouted.

Two scenarios present themselves within this challenge of staying excellent while improving the research of weaker institutions. The first is the need to concentrate on political requirements of 'equity' and compromise the 'excellence' that has been created historically while the other is a situation that accommodates both equity and excellence. The big question is whether or not the latter is possible. One extreme way to deal with the latter situation is to allow some universities the elite status and leave others struggling to acquire the status or perishing into teaching only institutions. The imbalance created by the two scenarios is depicted in figure 10.1 below, in which one component puts pressure on the other. If research in universities improves teaching as it is perceived to

do in New Zealand, then the latter institutions shall have been disqualified as universities. The situation may be worse than the apartheid arrangement. Transformation is aimed at removing inequalities and any resolutions unaligned to the paradigm, violate it. Thus, if it is transformation that is to be prioritized then equity will have to be the first to be addressed.

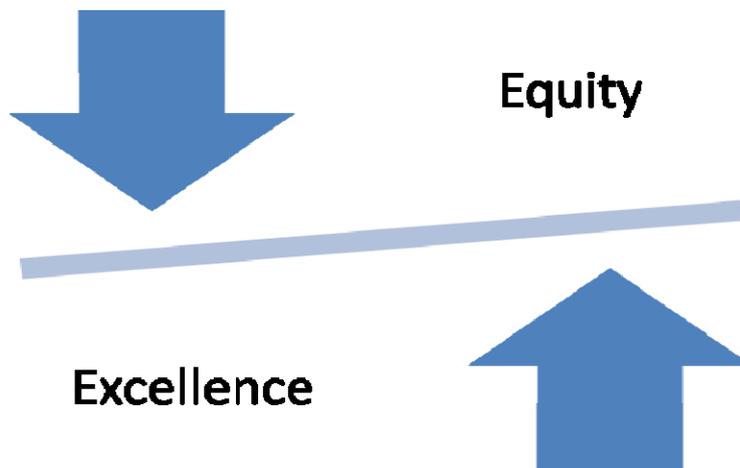


Figure 10.1 An imbalance created by two opposing national imperatives

10.3.4.3 Concentration in top selected institutions *versus* broad-base institutional capacity building

The two efforts (NRF rating and the DHET reward system) differ in intention/purpose of evaluation in that the NRF aspires to produce high quality researchers of international standing (through the rating system), motivated through funding. Despite the difference, both systems distribute resources by rewarding productivity. As organs of the state the efforts observe the principle of accountability to public funds. Both systems use results to distribute resources through a selective process similar to the UK and New Zealand models that align results to funding, concentrating funds on a competitive basis. Most of these researchers are still concentrated in the HAIs, leaving other low-performing institutions trailing when it comes to productivity. In the South African context funding

is more an incentive than compulsion to deliver outputs, while in the other two countries funding is used to compel productivity and competition.

As a result of its incentive nature, funding attracts good researchers to the evaluation process, occasionally promoting competition among unequal performers. As in New Zealand, the NRF and the then DoE believe that a good researcher will not only produce good research for nation building but will also promote scholarship and improve the training of research (post-graduate). By implication, those institutions that do not have good researchers may not improve in scholarship and research training unless proper planning is in place and the skills development programmes are utilized. Research training is important for continuity and sustainability and the absence of a good plan may lead to the perpetuation of imbalances across the system of HE.

The CHE's systems on the other hand, intend to improve the quality of the entire system of HE. This implies that evaluation will be without punitive or reward measures associated with funding. Unfortunately the CHE does not focus on research even though its processes utilize the Dutch approach by concentrating on systemic quality improvement (which is an agenda-building approach). The CHE's institutional audit programme uses the same unit of analysis as that of the Dutch system but the two differ in that the latter separates evaluation of teaching from that of research. Therefore, the CHE assessment operates similar to that of the relatively less differentiated Dutch model).

The CHE also strives for the improvement of internal regulatory processes, a step towards institutional independence. The two also differ in that the Dutch system is not driven by an intermediary body. This system has proven to be best suited for research needs in a transforming system of education (HE and Training in this case) in that it encourages development towards independence. Once all institutions perform at standard levels, competition may be encouraged. The Dutch and the CHE systems do not have punitive measures against poor performers and only make recommendations for the remedying of weaker institutions.

10.3.4.4 Tensions between world-class competitiveness *versus* responsiveness to national needs

The rating system and the DHET research award programme seem to concentrate more on the imperative of quality and international competitiveness and not much on the changing needs of the country. This is observable through the non-recognition of what SAPSE 110 refers to as ‘public service research’ by the research reward system.

Although the outputs of contract research may be submitted with the rating (NRF) submissions, the purpose of rating still relates to own-initiated research with the hope that the skilled researchers will develop interest in applied research.

Although the programmes have the potential and the ability to elevate competent researchers’ intellectual inquiry levels to global standards and to integrate the country’s best into the global system, this is taking place in a non-equal system. In the NRF, this may be blamed on the inability of the Foundation to respond to recommendations made by the reviewers of the programme. The efforts by the Foundation to call for proposals may have been a better idea if rated researchers responded to the call to develop entry researchers (mainly from the HDIs). With the realignment of rates with funding, it is not clear which direction the NRF wants to follow, making it seem that the original idea may be retained.

As stated, skilled researchers are capable of responding to national research needs, for example contract research. Unfortunately, if prestige is highly regarded, which it is in the NRF rating system, such researchers may not find enough time for such research unless it is publishable. This is despite the fact that the rating system does consider results of contract research to some extent. Even over the period of years of existence the rating system is not able to provide information on the national standing in research. This is because of its voluntary nature, allowing good researchers a choice not to participate.

10.3.4.5 Purposes of research evaluation in the South African efforts

In the post-apartheid era, accountability for use was declared in the university audit reports. Whereas the deployment of public funds is still a factor even today, the systems evolved to also include accountability (which is diagnostic) for the use of the funds, which leads to effective deployment (with prognostic effects). Whereas in the past (prior to 2006) government allocated research development funds for use on own discretion (block grant), this practice according to the 2006 funding framework would allow the submission of proposals for earmarked funds. This way of funding makes accountability for state funds much easier although ex-ante data also has its setbacks.

While in the past cross-subsidization of programmes was possible within the universities, the earmarked funds cannot be used to fund other programmes unless so stated in the proposal.

The encouragement of research productivity by rewarding quality research serves both the FRD/NRF and the DHET research reward systems (even though the quality control mechanisms differ). Both programmes endeavour to achieve high levels of research quality in the country. In this arrangement the system of HE is mainly the recipient of service as a result of its high involvement of its researchers and within these programmes, survives at the mercy of those who have the funds and decide on which criteria they use to qualify quality and standards.

Government nevertheless refrains from the UK's 'publish-or-perish' phenomenon but seems to leave the redress of research inequalities not well attended to. The quality-inclined CHE introduced a guideline for internal improvements of research, but falls short of explaining the manner of monitoring and evaluating the internal systems of research. Unless one wing of the audit process fully concentrates on research like it is done in the UK, New Zealand and The Netherlands, the benefits, if any, may be obscured. Otherwise, procedures of the audit process embrace the entire system of HE in the direction followed in comprehensive and systemic evaluation. Table 10.3 summarizes the purposes of the South African efforts.

Table 10.3 Impression of purposes of research evaluation - South African systems

Purpose	DHET output system (research reward)	CHE Audit (Improvement of quality)	NRF rating system (quality research of international standards)
Summative (funding decisions/ resource allocation)	Yes	No	Partially
Formative purpose (improving science/ research)	Yes (indirectly)	Yes	Yes
Strategic (prioritizing of fields/ decision-making around staff/ recruitment)	No	No	Yes
Accountability	Yes (funds)	Yes (quality)	Yes (funds)
Resource allocation	Yes	No	Yes
Assurance of quality and excellence	Partially	Yes	Yes
Symbolic (recognition of reputation and prestige-rankings)	Partially	No	Yes (rates of individuals)

When table 10.3 is compared with table 10.1, observable similarities exist between the Dutch model and the CHE institutional audit process. Also, the resource allocation-based UK exercise shows some similarities with the DHET research reward programme and to some extent the NRF rating system.

10.3.4.6 Governance issues

One historic effect still impacting on the functions of the programmes is government's involvement and steering powers. As a result of the 'power of funding' that the government possesses, the programmes that exist emerged from government intermediary bodies. Although there are claims that universities have always been consulted, there is little if any initiative or influence coming from the institutions. The rating system was initiated because government felt the need to have good researchers. Later on the NRF was introduced as an 'independent body', with its own objectives, but remaining as an intermediary body/parastatal and therefore a wing of government. The intermediary bodies are still the strong bodies through which government have steering powers over the universities.

The concentration, if any, on the efforts on the strategic purpose is the result of little involvement of academics in the research evaluation exercise. If the formative purpose (improvement of science) is a priority, participation of universities in the evaluation processes has to be strengthened. The voice of the universities is only heard when HESA is expected to make input into some of the programmes. As stated, this may result from dependency of institutions on government funding.

The existing level of institutional research participation in industry and government department research is mainly contractual, the output of which is usually not published. This is despite the investments that both industry and government put in the research activities in universities.

10.3.5 Dimensions of research recognized by the bodies

The dimension of quality of research is observed across all three bodies involved in research evaluation. From its inception, the FRD has been postulating the idea of identifying good researchers and supporting them (building a track-record of good researchers). The NRF relates quality research to international recognition in their awarding of rates and even utilizes international experts in their panels.

The DHET, on the one hand, recognizes publishable research output (diagnostic) and rewards this as an incentive to stimulate high quality research. The Department compiles a list of what they regard as recognizable (accredited journals), part of which is local journals that meet the requirements and the others internationally recognized journals. The move to fund/reward “successful” research proposal is also an indication of the acknowledgement and recognition of excellence (quality). The intra-science nature of the NRF rating system, as demonstrated by its rigor of assessment and the DHET reward system promote popularity of the systems among academics, even though their voluntary characteristics do not compel participation of new entries. The CHE on the other hand released the Good Practice Guide to assist with internal QA systems for the promotion of quality management and production of research and support. The CHE also defines quality in relation to international standards.

The DHET’s effective deployment of public funds is supposed to relate to the dimension of relevance. This is because research is recognized as a tool for the creation of knowledge in order for the country to meet “pressing needs” and to “respond to new realities” as stated by the National Plan for HE (DOE 2001) . The Department of Higher Education and Training will not only utilize earmarked funding to attempt to bridge the inequality gap but also require that such funds be accounted for. In the former instance the Department is geared to respond to a transformation imperative. Whether or not the imperative is achievable in this way still has to be proven.

The CHE’s Framework for Monitoring and Evaluation of HE keeps track of trends of transformation. How this can help track down research development within the transformation process and performance thereof is also not stated. One other component of relevance that relates to transformation is global competitiveness, which the efforts seem to consider effectively.

Very importantly, the CHE also links quality with relevance by validating institutional effectiveness. This approach is preferred and utilized by the Dutch in their recent alteration of their research evaluation system. In this way evaluation may assist in

determining whether or not research products in HEIs are effective enough to link quality to transformation imperatives.

The NRF recognizes productivity as one of the dimensions and this is mainly observed in line with the quality of products. Although the Framework for Institutional Audit mentions the regulation of outputs, the CHE documents mainly concentrate on processes of internal regulation, which when improved may lead towards better research output. There is no clear indication of the initiation of a process of assessing research output across the entire system of higher education yet.

Table 10.4 below gives a summarized impression of the criteria as observed with the South African efforts

Table 10.4 Criteria of research as used in the South African research evaluation systems

Dimension/criterion	DHET	CHE	NRF
Quantity	Yes	No	Yes
Quality	Yes	Yes	Yes
Productivity/efficiency	No	Not clear	Yes
Effectiveness/relevance	No	Yes	Partially
Viability/vitality	No	Yes	Partially

10.3.6 Methodological issues and procedures

10.3.6.1 Design and approach

As indicated above, on inception the different research evaluation programmes had different purposes that led to different approaches to the problem. These differences are notable in the different areas of methodology.

10.3.6.2 Units of analysis

With its mode of voluntary participation, the FRD had individual researchers as units of analysis around which assessment was concentrated. A good researcher according to the rating system is the individual researcher who is productive according to the requirements of the FRD/NRF and rates are allocated to individuals in a 'prestigious manner'. This individualization of assessment relates to that of the DHET research reward system although in the latter case more than one author of an article may be rewarded. This implies that the articles/books serve as units. Previously, the DoE rewarded universities for research on submission of research outputs for SAPSE allocations, using the institutions as the units of analysis. Whereas this arrangement continued with the block grant awarding of the HEMIS produced data, the plan for the period starting in 2007 was to have applications submitted for funding. This implies that the proposals would then serve as the documents (applications) of assessment and the researcher(s) as units, similar to the arrangement of the research councils.

The CHE mostly concentrates on institutional or systemic service delivery. The Framework for Monitoring and Evaluation of HE monitors trends of transformation alignment of the system of HE for example. The assessment results are expected to give an impression whether individual institutions are developing or not. In this case institutions are units of analysis. This concern is relevant to the idea of "strengthening weaker" universities. The Framework for Institutional Audits specifically concentrates on individual institutions, focusing on internal regulations and external validation, the principle used by the three models referred to in the study (Dutch, UK and New Zealand). In its report on the entire institution the process regards the institution as a performing agent.

The audit programme internal regulation and external validation approaches to all institutions may lead towards a systemic and comprehensive approach to research evaluation in line with the large-scale evaluation models. It should be noted that the audit process involves all functions of a university. That is, that no system exists in its mode and extent of operation that separates research from teaching. When assessment/audit is

conducted for research only, if it is so recommended by this study, the process would suit a large-scale research evaluation model. It is necessary under such conditions to decide whether it will be departments/faculties of individual researchers that will serve as assessment focus. The Good Practice Guide refers more to the internal management system such as internal assessment of academic research proposals and internal processes of research, and neither reveals institutional research performance that would be assessed or units to be assessed.

The choice of a unit of analysis relates more to purpose. Weingart and Maasen (Whitley *et al*, 2007) warn of the complexity of the choice of units of analysis in a pluralized system as a result of the different purposes. The FRD/NRF system and programmes of the DoE are influenced by funding similar to research councils and thus concentrate on funded individuals. To some extent this simulates the internal system of the New Zealand and UK models even though in these models researchers serve as sub-units and are then aggregated into university totals. The CHE on the other hand deals mostly with compliance of services to quality and relevance (to transformation) of such services, therefore taking institutional visions and missions into consideration and focusing on organizations/systems.

In The Netherlands, the improvement of competence in research programmes/departments is the denominator. The use of programmes/departments as units of analysis was therefore meant to prevent the effect that under-performing departments may have on good-performing departments when aggregation of results is used to represent an institution (when an institution is the unit of analysis). Also, this unit of analysis allows for the identification of needy research programmes/departments and improvements at this level, which may not be easily detectable when a university total is used. The differences in the intended outcomes of assessment lead to differences in the choice of units of analysis and this observation is important for this study. The table below (10.5) identified the units of analysis used by the efforts.

Table 10.5 Units of analysis of the South African research evaluation systems

Unit of analysis	DHET output	CHE Audit	NRF rating
National/systemic	No	Yes	No
Science field / discipline	No	Partially	No
Research institute / organization	No	No	No
Department/Unit of research	No	Partially	No
Individual research programme/project	Yes	No	Yes

10.3.6.3 Types of data utilized and methods used to collect the data

The rating system uses a detailed peer review approach to assess the researcher's credibility (proven ability). The review utilizes document analysis of the submissions made by applicants, which are mainly data-based. For confirmation of high level rates site visits are also conducted. The DHET research reward system also obtains data through submitted documents, which are then peer reviewed by experts of international standing. Although a researcher is regarded as a unit of analysis, the mode resembles that of ex-post data usage. The use of peer review serves as the system's strongest hold on research evaluation as the method relates to Hemlin's (1996) expression of opinion "from an intra-science perspective".

Although the NRF peer review system is highly regarded by academics and universities as more rigorous compared to the DHET reward system, in principle the two systems operate in the same way. The two efforts both depend on the (ex-post) data submitted by applicants for reviewing, both have panels for reviewing submissions and both release results for purposes of rewarding performance. The main difference between the two

efforts of research evaluation other than the level of rigor is the type of data commanded for review. Whereas the DHET reward system pays attention to the submitted item that is to be rewarded, that is, the article or book et cetera, that has been scholarly acknowledged, the NRF rating system holistically rates a scholar according to general performance in a number of areas and on a possible total number of items (submitted). The submission of items is a phenomenon related to performance indicators and is observed by Campbell (Shapira and Kuhlmann, 2003) as simulating quality measures. The DHET rewards the completion of doctoral research separately through the GUF funding.

The Institutional Audit programme on the other hand, studies submissions from self-evaluation portfolios of institutions/departments in the same manner as the above two systems do individual researcher's submissions (by panels) and then visits (peer review) the institutions for the verification of information, similar to the confirmation (triangulation) for highly-rated researchers in the NRF system. The broad focus of the institutional audit programme on all institutions, with the self-evaluation and external validation levels as used in comprehensive evaluation makes the programme more inclusive than the rest. The mode resembles that of the Dutch system in that site visits may be included. Its intention to have future follow-ups makes it continuous while systemic, bringing this programme as close to a large-scale evaluation system as the three referred to models in this study. As stated, the systems have different units of analysis as a result of the different purposes for evaluation.

It should be stated here that the rating system of the NRF runs parallel to the NRF's system of proposal funding, which utilizes the research council's strategy of ex-ante assessment of proposals. Whereas the DHET's old dispensation (block grant) depended on reports submitted by universities for SAPSE funding, the new funding framework proposes the submission of proposals (including discretionary and institutional factor grants) for the ex-ante data, all of which are documents. Therefore, whenever a proposal has to be submitted to qualify for funding, all systems use a similar pattern of approval.

As stated, the monitoring and evaluation process of the CHE deals with secondary data and therefore acquires data either electronically or on documents for viewing and sometimes analyzing if it has not been analyzed by the 'host'. The Good Practice Guide on the other hand, suggests ways of self-evaluation on research management mechanisms and not necessarily on performance.

A conclusion may be drawn, that utilization of items by the DHET rewards system makes the system a subset of the rating system, and that the rating system itself becomes the subset of the entire audit system considering the unit of analysis used (when only research was to be considered in the audit format). Therefore, with the exception of purpose and the units of analysis, the three systems use the same methods (just at different levels).

10.3.6.4 Focus areas (data)

This sub-section explains the information gathered through the different modes/methods of data collection. The NRF reviews data-base information by focusing on publications (the refereed and non-reviewed), conference proceedings, research reports, patents, post-graduate students supervised and any other relevant work conducted by the researcher. The DHET also depends on peer reviews, stakeholder inputs, efficiency audits and output audits. Submissions involves publishable outputs, postgraduate outputs as submitted by both the institutional audits and HEMIS, instructional research, public service research contracts and consultations) as well as artifacts. The Institutional Audit process of the CHE may be an attempt to revive the old DoE audit process that was helpful in providing data for the entire system of HE. Conference proceedings have always been regarded by the DHET as a test of international competitiveness. As stated, the DHET research reward system reviews the article, the monograph, the book, the chapter, edited work and/or conference papers, all on individual/unitary capacity of the submission.

As a recent system, the CHE's programme for Institutional Audit concentrates on management issues such as policies, strategies and processes. It should be noted that this programme simultaneously concentrates on all functions of an institution (especially the

core functions). Similarly, it is hoped that an inclusive programme of research evaluation will focus on variables that are the focus of performance directly related to research. When this happens, the system (exercise) will utilize the same data as the other two, but will produce results at a much broader (institutional) level than the two efforts and will thus be close to the comprehensive models. The monitoring and evaluation system of the CHE also studies policies and goals, but focuses on trends of development and transformation.

10.3.6.5 Benefits and the significance of the research evaluation efforts in South Africa.

The research evaluation efforts discussed above reveal advantages that may be similar in some of instances and different in others. In all instances, assessment contributes towards encouraging participation and productivity in different types of research. For example, both the rating system and the DHET's research reward systems strengthen the importance of self-initiated research. The result of this is participation in basic research. This assists with an increase in scientific publications and scientific conference papers. Good rates are also said to attract external funding.

The fact that the programmes refer to output as publications in accredited journals contributes to prominence and scholarship at high and international levels with skills to produce high quality/good research, train researchers and establish a high level of capacity that will deliver services even at national levels. The rating system specifically promotes eminence and upholds the standards to encourage it. International involvement is regarded as relating to quality and helps establish the presence or absence of good research for quality maintenance. This is observable in the constitution of peer review panels (for assessment), which includes international researchers. Encouragement for involvement in international conferences by the rating and the DHET reward systems also puts the country on the international map. Positioning on the international map also served as an added incentive for New Zealand's initial research evaluation exercise. The inclusion of public service or contract research data by the rating system adds the dimension/component of relevance and encourages applied research.

While the programme for the institutional audits demonstrates the most capacity to encourage involvement of all institutions with its somewhat compulsory nature, the NRF rating system and the DHET's research reward programme opens participation to individual researchers. The advantage of the latter is the ability to identify research potential even in the universities with low research output. The rating system and the DHET research reward system nevertheless have the potential to leave out researchers that are not volunteering to participate. Therefore, the country's entire research strength may not be revealed through the two systems. The CHE's Institutional Audit programme on the other hand, addresses the entire system of HE, thus providing an opportunity to reveal the weaker universities.

Although the CHE's Good Practice Guide provides information on how to raise research management standards and the fact that government provides development funds, the equalization of standards of research across all universities still needs attention. As stated, good research leads to rigorous standards of academic quality for the creation and sustenance of the quality of life. The results of the monitoring and evaluation of the programme of the CHE and that of the institutional audit process provide information for the understanding of change in the process of transformation. The programmes should be able to reveal whether or not there is improvement in the standards and quality of research in the weaker universities as well, and whether or not the disparity gap is being reduced (reveal good practices and weaknesses). The results of all the efforts if available may collectively contribute to knowledge for policy-making and these would illustrate the relevance of the programmes.

Both the Institutional Audit programme and the Good Practice Guide encourage capacity building for self-evaluation through their self-regulatory arrangements. Although the two programmes are not specific on variables, it may be concluded that once the processes are in place, evaluation criteria would be defined. The establishment or improvement of internal policies and procedures not only facilitates processes of monitoring activities but also assists in the increase and improvement of research resources. Therefore, the present

state of the programmes may be regarded as a preparatory phase for the future arrangement of a better-developed system. Meanwhile the involvement of the entire system of education (through the audit process) creates awareness and networks for sharing best practices. The internal regulation system referred to by the two CHE programmes, on the other hand, is used in The Netherlands, and not only makes the Dutch exercise unique but is also giving it strength in encouraging institutional autonomy while improving the aggregation system at the same time. The NRF ratings are also used by other institutions for staff placement.

Although the use of secondary data by the monitoring and evaluation exercise may not always obtain data/information for a systemic arrangement of evaluation, the system may, if undertaken, serve as a monitoring tool for the other systems (that provide the data) and if well planned, the idea may serve the purpose of meta-evaluation for a 'central' research evaluation system. Results of the monitoring and evaluation programme may also create a platform for public debate for the determination of the direction that the system of HE may take and therefore may influence policy issues.

The CHE's Institutional Audit programme is presently creating a baseline against which future evaluation structures may be based on and researched. How much this is done for research evaluation by the programme needs to be established. The programme for monitoring and evaluation also hopes to achieve this goal using secondary data. The Good Practice Guide suggests the establishment of a research information system (RIS). The DHET reward programme attempts to address the local language issue by allowing other items, for example books, to be written in any of the official languages in South Africa and have the abstract translated into English.

10.3.6.6 Challenges experienced by the systems

The advantages experienced by the research evaluation programmes mentioned above are not without challenges. Besides the fact that their efforts are independent and somewhat fragmented, each one of the programmes display some problems either in the past (historically) or at the present moment.

The other concern arises from voluntary participation and thus the units of analysis utilized by the rating system. As stated, the fact that researchers volunteer to participate leaves out highly-skilled researchers that do not volunteer, reducing evidence of the existence of good researchers. The non-compulsory nature of the rating programme does not encourage group research and thus does not necessarily contribute to capacity building.

The merger of institutions is one other problem affecting the programmes. For example, secondary data that was accrued prior to this process may be declared invalid for universities that merged if the data needed for use by the CHE's monitoring and evaluation programme relates to individual institutions. There has been a delay in the process of institutional audit for merged universities, and this makes it difficult to measure the effects of the programme and disrupts the systemic arrangement initiated by the programme.

The CHE programmes are mainly meant to improve the conditions of HE in the new dispensation (government) and therefore have a short history. The programmes still concentrate more on the processes and mainly reveal problems that relate to methodological issues. Whereas the institutional audit is said to concentrate more on the quality dimension, there are some indications that the programme also encourages productivity and relevance. The Good Practice Guide on the other hand, is blamed for not addressing issues of research management in totality. For example the functions and the management of the research office together with the assessment of actual research output are all not indicated. It is also notable that the document only concentrates on the components of internal quality services and does not link with activities of evaluation, especially of the entire system of HE.

The DHET reward system is also suffering from the problem of updating the ever-changing status of accredited journals, with the result that new entries may not have the opportunity while some outdated journals may stay in the system. This may be observed

as lack of response to new knowledge. The other problem encountered by the research reward system relates to the definition of scholarship, which may not necessarily be the same as research competence. This is in contradiction to the views of other countries, for example New Zealand, in which research evaluation is expected to reveal levels of research competency for better academic service delivery.

These methodological problems are similar to problems experienced elsewhere, including the countries referred to in this study. For example, the effects of ‘discipline bias’ experienced in the UK in the 1980s are another sign of the lack of response to new knowledge that may be interdisciplinary. The inherent subjectivity of the peer review programme and therefore its perceived bias has been cited as a problem encountered by the programmes. This has been observed as a result of the occasional absence of consensus and questionable review credibility among the NRF reviewers. Aligned to these are the halo, Oxbridge and Matthew effects associated with peer reviews. Some of the applicants for rating even question the notion of international standards as a criterion used to measure their performance.

As stated, secondary data is not easily verifiable, making the use of this data by the monitoring and evaluation programme (CHE) to be questionable, especially if used for systemic reasons. The programme has opted for the use of performance indicators, which themselves are referred to as proxy of reality of results.

CHAPTER 11 CONCLUSION AND RECOMMENDATIONS

This chapter draws together the main findings of this study. Recommendations in each of the sub-sections follow the conclusions, providing the conceptual basis against which a South African research evaluation framework can be assessed. The chapter starts with a discussion on the rationale and purpose of research evaluation, recommends the criteria best suited for different evaluation purposes, and then makes recommendations regarding units of analysis. The chapter also focuses on the management of the envisaged research evaluation framework and advises on methodological issues. Finally, the chapter consolidates and synthesizes the recommendations into a proposal for further study.

11.1 Rationale and purpose for research evaluation

Research evaluation in South Africa is premised on the transformation imperatives of the country and how these imperatives impact on or are impacted by the landscape of higher education (HE) research in the country. The conclusions hereunder centre on the three imperatives discussed previously, namely; the redress of inequalities, international competitiveness and responsiveness to national needs.

The study has found little evidence that the existing research evaluation efforts assist in acquiring information that would help to reverse the historical disparities across the HE and training system. In fact, based on their purpose, the efforts are more likely to produce results that expose continued inequalities. Voluntary participation in evaluation efforts does not compel or encourage academics, especially at the weak performing institutions, to engage in research. Therefore, competitive levels in research are still only evident at the previous high-performing institutions. The CHE's programme of institutional audits is structured to facilitate such change but combines the assessment of all core functions, reducing the concentration on the activities of research.

The existing disparities in research between institutions of higher learning do not only affect the efforts aimed at broadening the base of good researchers for international

competitiveness and standing, but also affect the base of talent (human resources) available to respond sufficiently to national needs.

Stemming from the challenges discussed in chapter 10, of tensions between equity requirements *vis a vis* the process of upholding excellence, this study concludes that the basis for evaluation is the prioritization of imperatives in order to build up a system that is not only sustainable but will also improve conditions of research in the South African universities. In an attempt to achieve this, the Dutch model is preferred for its systemic developmental approach, which encourages the increase in research participation in public HEIs. Institutions should in the process be orientated to accountability processes.

There are no doubts that quality (excellence) would be maintained and upheld because experts are involved in the exercise. It is admissible that in the initial stages the process will be time consuming and will require more resources. Therefore, planning has to take these factors into consideration and use experiences from other models (including the UK and New Zealand) to do so. Once participation has been increased, evaluation may then assist with the maintenance of quality research across the universities. At this later time, competition between programmes in universities and between the universities themselves may be encouraged and responsiveness to community research will also increase.

Figure 11.1 demonstrates the direction of flow as analysed and concluded in this study. This flow may be followed when the purpose of research evaluation is formulated and this is done in relation to the national imperatives. Redress of disparities forms the crux as it relates directly to the increase in participation, and this has to be regarded as the primary objective. It is this objective that, when effectively addressed and coupled with excellence will also improve international competitiveness in the entire system of HE. The internationally competitive researchers (good researchers) of a broad base will also be able to fulfill national demands.

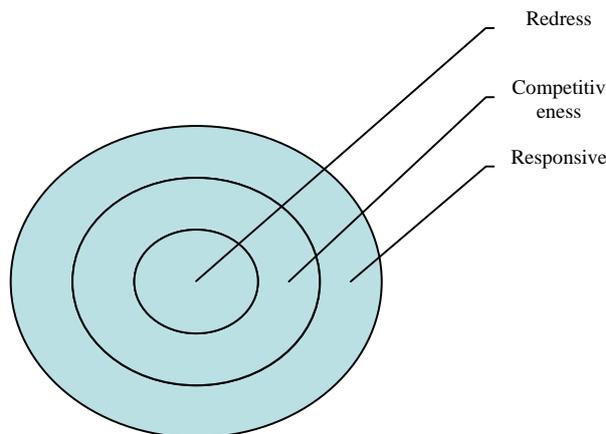


Figure 11.1 Indication of the interdependence of the national imperatives

With this background, the primary purpose for research evaluation becomes the development and improvement of quality research of international standard across the system in order for research to make a meaningful contribution to national demands. The purpose has to be formative, assure quality and be somewhat strategic in order for it to achieve the symbolic requirements. No resource allocation needs to be attached to the purpose for evaluation. The process has to compel engagement and productivity in research and put measures in place to do so without the use of punitive measures.

11.2. Units of analysis

The use of the lowest level of the units of analysis by the South African efforts (NRF rating and DHET's research award) does not provide information on any of the national imperatives. The aggregation of the total university output, if universities were to serve as units of analysis in this regard (using individual researchers or their outputs) could actually provide information about the standing of an institution or the country as a whole. Unfortunately voluntary participation means that other researchers may unintentionally be excluded. Also, the mere summing up of outputs falls short of establishing where the improvements are needed. The CHE's institutional audits use institutions as units of analysis to directly analyze individual institutions in order to

reveal levels of performance. Results are thus used to advise institutions on areas that require attention.

The use of programmes/departments as units of analysis as it is done in The Netherlands has the ability to focus deeper into individual institutions to identify programmes/departments that are performing poorly (that need assistance). This protects better performing departments from being compromised by the poor performers of the same institution. In this way, good researchers can be identified and retained in the programmes/departments.

The use of programmes or department as units is better suited for the South African situation in which parity is to lead to international competitiveness. The collection and the aggregation of data per discipline will initially serve as baseline information for future comparisons when improvements are determined.

11.3 Dimensions/Criteria for research evaluation

Quality is undoubtedly endorsed by the efforts of research evaluation in the country and this is observed in the assessment strategies and as a criterion for assessment. The criterion of quality is also used for the establishment of international standing. The same recognition is observed with productivity as it relates to accountability for resources. Relevance, which closely relates to national demands, is not prioritized by the research evaluation efforts, although determined by the institutional audit programme, meaning that this component is not encouraged. The component is useful when assessing whether or not research is responding to 'pressing national needs' and is responding to 'new realities'. As the present research evaluation efforts concentrate more on individual volunteer researchers, it may not be easy to judge levels of institutions' concentration on the imperatives and thus, viability may not be guaranteed.

It is recommended that all imperatives should be taken into consideration as all are necessary to contribute to transformation activities. That is, quality and productivity are necessary for both baseline and continuous assessment of research

programmes/departments in HE institutions. Information will then be used by institutions to improve, based on the recommendations of the assessment panels. The quality criterion will also be used for international standing and competitiveness, while relevance relates to transformation requirements and national developmental goals.

Viability is relevant for the determination of the continuity of the science field and is helpful when advice is given. If programme/departamental research output does not show signs of sustainability over time, even when improvement measures are imposed, such a programme is not productive and/or viable and may need to be discontinued. All these dimensions are necessary to assist evaluation results to be interpreted and for recommendations to be better understood for the planning of improvements. When the research in the programmes/departments is found to be responding to national needs this will signify its relevance to HE research.

11.4 Governance and management of the research evaluation exercise

There is no committee that is science-related (made up of academics) that is responsible for a broad-base research evaluation in the country. The present institutional research evaluation is conducted by intermediary bodies and is in one way or another linked to funding. The other intermediary body, the CHE has the potential for a comprehensive research evaluation exercise through the institutional audit process. However, the process does not separate research from other core activities, with tendencies of splitting attention and reducing focus on each one of the core activities.

There is an observable void, creating problems of accountability at national level as no structure takes full responsibility for the research activities of the autonomous HEIs. There is evidence of situational (sporadic) research evaluations conducted time and time again either when commissioned or when some researchers deem it necessary (referred to in chapter 1). This is a good sign that capacity exists among academic staff members to participate in such an exercise in addition to those of intermediary bodies.

The CHE's monitoring and evaluation programme has the responsibility of monitoring and evaluating transformation in HE in the country. The programme therefore has the potential to provide information on possible improvements if data is made available. That is, the programme can establish parity improvements and whether such improvements address national imperatives. In this way the programme will be determining the contributions of the research evaluation exercise (meta-analysis). The information is needed by both the departments of Higher Education and Training and Science and Technology.

It is recommended that a separate wing for the evaluation of research be established within the CHE, which, because of its potential to assess institutions, its relationship with the DHET and its status as an intermediary body is the most suitable for the activity. The wing may be formed as a new and standalone branch of the CHE/HEQC to concentrate only on research. The wing should be augmented in the planning and management of the evaluation exercise by representatives of universities, with the composition determined through HESA (or its sub-committees) in collaboration with university scientists. A sub-committee for research, if it exists, needs to be strengthened. If it does not exist, this may be initiated. The involvement of this non-statutory body would be to provide university system with an opportunity and right to be represented without fear of government prejudice.

There is also a need to involve experts in research evaluation, including both local and international scientists. These should include those researchers who are or have previously been involved in research evaluation through sporadic (situational) interventions or through involvement in the intermediary bodies. In this way, the expertise from the NRF rating and the DHET research award systems may be incorporated and adequately utilized. The combination of all these bodies (as indicated in figure 11.2) contributes to better placement in the steering/aggregation continuum. This added responsibility to the CHE may require legislation as the Council is a statutory body with statutory mandates. An alternative to this formation is a non-statutory body such as the VSNU in the Netherlands. Unfortunately, the formation of such a body has

challenges such as the commitment of universities, the non-formalised, non-defined differentiation existing between the universities, reliance on state funding and the fact that the CHE is in existence and its services may be made available.



Figure 11.2 Suggested structure for the management of the evaluation exercise

There is a need for accountability, the responsibility of which resides with government, to monitor and facilitate the implementation of national imperatives. The two departments (DHET and DST) have dual and shared responsibility for monitoring improvements and to conduct meta-analyses for the programme. An alternative to the formation is a committee formed of the two departments, utilizing the services of the CHE's monitoring and evaluation programme with a plan designed for the purpose and reporting to a committee made up of a union of the two departments (DHET and the DST).

11.5 Methodology

South Africa research evaluation efforts use designs that utilize quantitative data for assessment through peer reviews. The designs simulate those of research councils in that they concentrate on individual voluntary researchers and allocate benefits on this basis. The CHE's institutional audit programme, collectively analyses the core activities of individual institutions in a systemic and comprehensive manner.

The rigor of assessment used by the South African efforts (NRF rating and the DHET) is comparable with international standards, with the NRF applying more intense peer review strategies to further improve on quality. Similar to the international models the efforts use ex-post data. Data in both efforts is submitted on an individual basis (DHET submissions accumulated by universities). The efforts do not use internal evaluation by universities. In its analysis the CHE institutional audit handles data internally and externally.

A comprehensive and systemic research evaluation is recommended which has programmes or departments of universities as the units of analysis. The exercise has to allow institutions to participate in the collection of data, which has to be verified and then assessed externally by peers. Ex-post data may be augmented by the use of process information to capture activities of research in the departments. The initial assessment will collect base-line data for the determination of the state of affairs.

The management committee will facilitate the identification of panels and the appointment of panel members and plan training sessions, as it is done in the preparation of institutional audits. It is important to study all the problems experienced in other models to avoid problems. Data utilization is important for the achievement of the objectives.

Results of evaluation need to be made public. The first reports compiled by the panels have to be sent to institutions with panel recommendations and copied to the monitoring and evaluation team reporting to the DHET and DST departments. The reports would be followed by institutional plans on how such recommendations will be implemented. The plans should reflect requests for support from the institution. If developing a plan is the problem, a preliminary note may be sent to the monitoring and evaluation team for assistance to be provided.

The national report has to be transparent and popularised as much as possible. These would be based on the dimensions as recommended in the study and published per programme. The league-type of results sharing is not encouraged until such time that the purpose of evaluation includes competition (when the present imperatives have been achieved).

11.6 Reflections and synthesis of the recommendations

The suggested research evaluation model is based on the following framework:

Purpose

The primary purpose for research evaluation is to develop and improve quality research of international standards across the system in order for research to make a meaningful contribution to national demands.

Unit of analysis

Programmes or departments in HE institutions are identifiable as units of analysis and results have to be submitted to institutions reflecting such units.

Dimensions/criteria for evaluation

Quality has to be regarded consistently across the panels and common understanding has to be created within panels to avoid cognitive distance. Rigor of assessment should not be compromised. Productivity, on the other hand, will give an indication of levels of performance of the programmes/departments and will also be used to judge improvements over time. Guidelines for the achievement of national imperatives have to be determined in collaboration with the DHET together with the Department of Science of Technology. This will reveal the relevance of the programme. This will also assist in determining the viability of a programme.

Management and governance

The suggested research unit/wing of the CHE's institutional audit together with identified research evaluation specialists and representatives of the institutional scientists will form

a standalone programme management committee. All public universities should be compelled to participate in the programme. Results will be submitted to institutions with the recommendations and the committee will collaborate with the monitoring and evaluation programme to 'inspect' implementation of the recommendations. The monitoring and evaluation programme of the CHE (or a formation assigned such responsibilities) will also assess the achievement of the objectives over time and reveal possible improvements. Results will be submitted to both the project management committee and the Departments of Higher Education and Training and Science and Technology.

Methodology

A comprehensive and systemic research evaluation exercise is best suited for the purpose of research evaluation in this instance. It is therefore important that methods of data collection and procedure of assessment are clarified, and are transparent and user-friendly, with proper guidelines given to the institutions. It is important to have scientists participate in the panels, constituted in a transparent way.

Feedback to institutions is necessary and the implementation of recommendations should be enforced and supported. That is, support plans have to be put in place to assist poor performing departments.

Concluding reflection: Only if such an integrated research evaluation exercise is developed and implemented, will higher education research in the country be subjected to appropriate assessment criteria and processes and will the system be forced to face head on the dual demands of excellence and national imperatives.

REFERENCES

- Arnold E 2004 Policy Making. Evaluation research and innovation policy: a systems world needs system evaluation. *Research Evaluation*, vol 13:3-17
- Bauer M., Kogan M. 1997. Evaluation systems in the UK and Sweden: Success and difficulties. *European Journal of Education*, Vol. 32, No.2:129-143
- Ball R., Wilkinson R. 1994. The use and abuse of performance indicators in the UK higher education. *Higher Education* Vol. 22:417-427
- Barker K. 2007 The UK Research Assessment Exercise: The evaluation of a national research evaluation system. *Research Evaluation*, Vol 16 (1):3-12
- Bauer M., Kogan M. 1997. Evaluation Systems in the UK and Sweden: Success and difficulties. *European Journal of Education*, Vol. 32, No. 2: 129-143
- Brochmann G., Engelstat F., Kelleberg R., Leira A. and Mjoset L. (Ed.) 1997. *Comparative Social Research*. Jai Press LTD, London
- Brooks T.A. 1986. Evidence of complex citer motivations. *Journal of the American Society for Information Science*, Vol. 37:34-36
- Burnhill P.M, Tubby-Hille M.E. 1994. On measuring the relation between social science research activities and research publication. *Research Evaluation*, Vol. 4, No. 30:130-152
- Cave M., Hanney S., Kogan A. M. 1988. *Performance Indicators in Higher Education*. Jessica Kingsley Publishers Ltd. London
- Cave M., Hanney S., Kogan M, and Trevett G. 1996. *The use of Performance Indicators in Higher Education. A critical analysis of a developing practice*, Jessica Kingsley Publishers, London
- Claassen W. 2006 *The Higher Education Research Landscape in South Africa*. Presentation slides, Bibl_simp.ppt
- Council for Higher Education. 2001. *Higher Education Quality Committee, Funding Document*, Pretoria

- Cozzens S. 1995. US Research Assessment: Recent Developments. *Scientometrics*, Vol. 34, No. 3: 351-362
- Creswell J.W. 1998. *Qualitative inquiry and Research Design*. Sage Publishers, United Kingdom
- Denzin N.K., Lincoln Y.S. (Ed) 1998. *The Landscape of Qualitative Research*. Sage Publishers, London.
- Der Heyde A and Mouton J 2007 Unpublished review report, Crest
- De Vries O. 2001. *Higher Education Quality Assurance in the Netherland: the road towards accreditation*, Utrecht.
- Geuna A and Martin B.R 2001 *University Research Evaluation and funding: An international comparison*. Science and Technology Policy research. Electronic working paper series
- Hemlin S. 1996. Research in research evaluation. *Social Epistemology*, Vol. 10, No. 2: 209-250
- Howie J.G.R., Scott N.C.H 1993. The university research assessment exercise 1992: a second opinion. *Lancet*, Vol. 342:665-666
- Ishengoma J.M. 2002 *Financing higher education in Post-apartheid South Africa: Trends, developments and challenges ahead* (unpublished)
- Jeliazkova M., Westerheijden D.F.(Forthcoming chapter, 2004) *National System of Higher Education in the Netherlands. Accreditation in the Framework of Evaluation Activities: A Comparative Study in the European Higher Education Area. Country Report the Netherlands*.
- Johnes G. 1994. Research Performance Measures: What can international comparison teach us? *Comparative Education*, Vol. 30, No. 3: 205-216
- Johnes J. 1996. Performance assessment in higher education in Britain. *European Journal of Operational Research*, Vol. 89: 18-33
- Johnes J., Taylor J. 1990. *Performance Indicators in Higher Education*. The Society for Research into Higher Education and Open University Press.

- Krige S 2007 An Historical Review and Analysis of the NRF Rating System 1983-2005
- Lange L, Luescher T.M 2003. A monitoring and evaluation system for South African higher education: conceptual, methodological and practical concerns. South African Journal of Higher Education, Vol. 17, No. 3:82-88
- Lucas L. 2004. Linking Research and Teaching: The significance of research funding and evaluation policies. Conference Paper: Research and Teaching: closing the divide? An international colloquium, Marwell Conference Centre, Winchester
- Luckett K 2007 The introduction of external quality assurance in South African higher education: an analysis of stakeholder response. Quality in Higher Education, Vol 13 (2):97-116
- Maasen P.A.M. 1997. Quality in European Higher Education: recent trends and their historic roots. European Journal of education, Vol. 32, No. 2: 111-127
- Martin B.R. and Irvine J. 1983. Assessing basic research. Some partial indicators of scientific progress in radio astrology. Research Policy, Vol. 12: 66-90
- Martin B.R. and Irvine J. 1983 . Assessing basic research; the case of the Isaac Newton telescope. Social Studies of Science, Vol. 13: 49-86
- May T. 1997. Social Research 2nd Ed. Open University Press, Philadelphia.
- McLaughlin M. 2003. Tertiary Education Policy in New Zealand report, sponsored by Ian Axford (New Zealand)
- McMillan J.H., Schumacher S. 1993. Research in Education, a conceptual introduction. Harper Collins College Publishers
- Melck A 1995 University Financing. South African Journal of Higher Education. Research Supplement. Vol 9 (2):128-133
- Moravesik M.J., Murugeson P. 1975. Some results on the function of quality of citation. Social Studies of Science, Vol. 5: 86-92
- Mouton J, Dowling Z, Eigelaar I 2001 R&D Evaluation: An overview of concepts, approach and Method. CREST, University of Stellenbosch

- Mouton J., Hackmann H. 2006 The National Research and Technology Audit. Survey on Scholarship, Research and Development. Centre for Interdisciplinary Studies, Stellenbosch
- Mouton J and Weingart P 2010 Powerpoint Slides. Workshop on R&D Evaluation, Stellenbosch
- Nederhof A.J., Noyons E.C.M. 1992. International Comparison and Departments' Research Performance in the Humanities. *Journal of the American Society for Information Science*, Vol. 43, No. 3: 249-256
- Nederhof A.J., Zwaan R.A. 1991. Quality Judgments of Journals as Indicators of Research Performance in the Humanities and Social and Behavioural Sciences. *Journal of the American Society for Information Science*, Vol. 42, No. 5: 332-340
- Noble J.H. (Jr.) 1974 Peer review: quality control of applied social research. *Science* 185: 916-921
- Oppenheim C. 1997 The correlation between citation counts and the 1992 Research Assessment Exercise Rating for British Research in Genetics, Anatomy and Archaeology. *Journal of Documentation*, Vol. 53, No. 5: 477-487
- Over R. Anonymous. Use of peer review by the Australian Research Council. *Australian Universities' Review*: 31-35
- Pienaar M, Blankley W, Schirge G.U and van Gruenewaldt G 2000 The South African system of evaluating and rating individual researchers: its merits, shortcomings, impact and future. *Research Evaluation*, volume 8 (1):27-36
- Ragin S.C. 1987. *The Comparative Method: Moving Beyond Qualitative and Quantitative Strategies*. University of California Press, Los Angeles.
- Reinecke C.J 1993. The present subsidy formula and the subsidies generated by the Humanities and Natural Sciences, *Africa* 2001, vol 2 (1):16-18
- Rip A and van der Meulen B.J.R 1995. The Netherlands. The patchwork of the Dutch evaluation system. *Research Evaluation*, volume 5 (1): 45-53
- Rip A and van der Meulen B.J.R 1996 Policy approaches. The post-modern research system. *Science and Public Policy*. Vol 23 (6) 343-352

- Roberts G. 2002 Leap over the RAE to a new way of assessment. RA review www.ra-review.ac.uk
- Rossi P.H., Freeman H.E. 1993. Evaluation. A Systemic Approach. Sage Publications, London.
- Rudd E. 1988. The evaluation of quality of research. *Studies in Higher Education*, Vol. 13, No. 19: 45-57
- Scott S. 2007. Peer review and the relevance of science. *Science Direct*, volume 39, issue 7:1-20
- Shapira P and Kuhlmann (Ed) 2003 Learning from Science and Technology Policy Evaluation – Experience from the United States and Europe. Edward Elgar, USA
- Sizer J. 1988. In Search of Excellence – Performance Assessment in the United Kingdom. *Higher Education Quarterly*, Vol. 42, No. 2: 153-161
- Steyn 2000 Quality Education and quality in education: a dilemma for the democratic South Africa? *South African Journal of Education*, 20:46-49
- Stolte-Heiskanen V. 1992. Research Performance Evaluation in the Higher Education Sector: A grass-root perspective. *Higher Education Management*, Vol. 4, No. 2:179-193
- Subotzky G. 1997. The Enhancement of Graduate Programmes and Research Capacity at the Historically Black University. Summary of final research report, EPU, University of the Western Cape, Bellville
- Tognolini A., Adams N., Hattie J. 1994. A Methodology to Choose Performance of Research Attainment in Universities. *Australian Journal of Education*, Vol. 38, No. 2: 105-117
- Van den Berghe H., Houben J.A., de Bruin R.E., Moed H.F., Kint A., Luwel M., Spruyt E.H.J. 1998. Bibliometric Indicators of University Research Performance in Flunders. *Journal of the American Society for Information Science*, Vol. 49, No. 1: 59-67

- Verkleij A.C.L. 1998. Because every situation is different ... A contribution to the discussion of the pros and cons of large-scale research evaluation. Confederation of the European Union of Rectors Conference, Brussels
- Waghid Y., Berkhout S., Taylor D., de Klerk J. 2005. In defense of institutional autonomy and academic freedom: Contesting state regulation of higher education. *South African Journal of Higher Education*. Vol. 19, No. 6:1177-1195
- Waghid Y., le Grange J. 2003. Research and development in higher education: rating or not? *South African Journal of Higher Education* Vol. 17, No. 1:5-8
- Weingart P 2003 Knowledge and inequality. Bielefeld University
- Westerheijden D.F. 1999. Innovation Indicators in Science and Technology Evaluation: Comments from Higher Education point of View. *Scientometrics*, Vol. 45, No. 3:445-453
- Whitley R et al (other authors not qualified on document) 2007 *Yearbook Sociology of the Sciences*, Springer
- Worral J., Currie G. (Ed). 1995. *The Methodology of Scientific Research Programmes*. Cambridge University Press
- Wrath K. 2005. Ramping up research - Turning drain into gain
unlimited.net.nz/unlimited.nsf/unid
- Yin R.K 2009 *Case Study Research Design and Methods*. Fourth Edition, Sage, London

Organisational Documents:

- CHE 2001. Higher Education Quality Committee (HEQC) Founding Document
- CHE 2004a. Towards a Framework for the Monitoring and Evaluation of South African Higher Education
- CHE 2004b. Framework for institutional audit
- CHE 2005. Good practice guide for quality management of research
- Committee for Higher Education in the European Community (CHEEC) 1993 Quality management and quality assurance in European higher education. Methods and mechanisms. Compiled for the Commission of the European Communities.
- CHEMS 1996. The Real World of Performance Indicators. A review of their use in

- selected common wealth countries. Hobbs the Printer, Tottan
- DoE 1997a. White Paper 3. <http://www.polity.org.za>
- DoE 1997b Higher Education Act 101 Government of South Africa. Government Gazette, Vol 443, No. 23415:1-48
- DoE 1997b. Higher Education Act 101. Republic of South Africa, Government Gazette. Vol. 1196, No. 21784:1-10
- DoE 2001 National Plan for higher education.
<http://www.polity.org.za/html/govdocs/misc/higheredu2.htm>
- DoE 2003 Policy and Procedure for Measurement of Research Output in Public higher education institutions
- Higher Education Funding Council for England 1997a. The impact of the 1992 Research Assessment Exercise on Higher education institutions in England.
http://www.miss.ac.uk/education/hefce/pub97/m_97.html :1-23 (HEFCE, 1997 - m6/97))
- Higher Education Funding Council for England 1997b. Data collection for the 1996 Research Assessment Exercise: Review by Page E.S
http://www.miss.ac.uk/education/hefce/pub97/m2_97.html :1-15 (HEFCE, 1997 - m2/97)
- Higher Education Funding Council for England 2002. HEFCE announces major review of research assessment. www.hefce.ac.uk
- Higher Education & Research Opportunities (1996) 1996 Research Assessment Exercise: The Outcome. http://www.hero.ac.uk/sites/hero/rae/rae96/cl_96.html : 1-16 (RAE 1/96)
- Higher Education Funding Council for England (1998) Funding higher education in England: How the HEFCE allocates its funds.
http://www.miss.ac.uk/education/hefce/pub98/98_67.html : 1-22 (HEFCE 98/67)
- Higher Education & Research Opportunities (2001) Research Assessment Exercise in 2001: key decisions and issues for further consultation.
http://www.hero.ac.uk/sites/hero/rae/miss/1_98.html : 1-14 (RAE 1/98)
- Ministry of Education 2006, Ministerial Statement on Higher Education Funding
- NRF (2002) The Evaluation and Rating of the Research Performance in South Africa

NRF Institutional Review Panel report 2005

Periodic research performance exercise report (2004). Evaluating Research Excellence, Tertiary Education Commission, Wellington, New Zealand

Protocol – 1998. Assessment of Research Quality <http://www.vsnu.nl>

Research Assessment Exercise (1997) Conduct of the Exercise: RAE Manager's Report. http://www.miss.ac.uk/education/hefce/rae96/cl_97.html : 1-22 (RAE 96 1/97)

South African Post-Secondary Education (SAPSE-110) An Investigation of Government Financing of Universities. Department of National Education, Pretoria

Standard Evaluation Protocol (2003-2009) for Public Research Organizations <http://www.vsnu.nl>

Research Assessment Exercise (2002) Research Assessment Exercise: manager's report. http://www.hero.ac.uk/sites/hero.rae/pubs/other/2001_managers_reportfinal.doc : 1-22

RAE 01/2004

Tertiary Education Commission (2004) Performance-Based Research Fund, Evaluating Research Excellence: the 2003 assessment. Tertiary Education Commission, National Office, Wellington, New Zealand

White Paper on Science and Technology 1996. <http://www.polity.org.za>

VSNU Protocol. 1998. Assessment of Research Quality <http://www.vsnu.nl>

Reviewed but not cited:

Bickman L and Rog DJ (1998) Handbook of Applied Social Research Methods. Sage Publications, London.

Christiansen I.M., Slammert L. 2005. A multi-faceted approach to research development (1): Address the myths. South African Journal of Higher Education, Vol. 19, No. 6:1011-3487

Christiansen I.M., Slammert L. 2006. A multi-faceted approach to research development (11): Supporting communities of practice. South African Journal of Higher Education, Vol. 20, No 1:15-28

- Chubin D.E., Moira S.D. 1975. Content analysis of references: adjunct or alternative to citation counting? *Social Studies of Science*, Vol. 5:423-441
- Garfield E. 1979. Is citation analysis a legitimate evaluation tool? *Scientometrics*, Vol. 1: 359-375
- Gillet R. 1989. Research performance indicators based on peer review: a critical analysis. *Higher Education Quarterly*, Vol. 43: 21-38
- Gordon M.E., Purvis J.E. 1991. Journal publication records as a measure of research performance in industrial relations. *Industrial and Labor Relations Review*, Vol. 45, No. 1: 194-201
- Griffith J. (1995) Research assessment. Council for Academic Freedom and Standards report No4
- Harley S. and Lee F. (1995) *The Academic Labour Process and the Research Exercise: Academic Diversity and the future of Non-Main Stream economics in the UK in the universities.* (Leicester Business School)
- Hay 2000. Quality research in South African higher education: illusions, imperatives and ideals. *South African Journal of Higher Education*, Vol. 14, No.1:53-60
- Humphrey C., Moizer P. and Owen D. (1995) Questioning the value of research selectivity exercise in British Universities accounting. University of Leeds.
- Johnes G. 1988. Research Performance Indicators in the University Sector. *Higher Education Quarterly*, Vol. 42, No. 1: 54-71
- Johnes G. (1990) Measures of research output in university departments of economics in the UK 1984-88. *Economics Journal* 100:556-560
- Johnes G. (1992) Performance indicators in higher education: a survey of recent work. *Oxford Review of Economic Policy*, 8(2):19-34
- Le Granje L. 2004. Western science and indigenous knowledge: Competing perspectives or complementary framework? *South African Journal of Higher Education*. Vol. 18, No. 3:82-91
- MacLeod D. 2006. Research exercise to be scrapped. *Guardian Unlimited* education.guardian.co.uk/RAE/story

- Mac Roberts M.H and Mac Roberts B.R. 1986 Quantitative measures of communication in science: a study of the formal level. *Social Studies of Science*, Vol. 16: 156-172
- Martin B.R., Skea J.E.F., Ling E.N. 1992. Performance indicators for academic scientific research. Report submitted to the Advisory Board of the Research Council and to the Economic and Social Research Council, SPRU, University of Sussex
- Neale-Shutte M., Fourie J. 2006 Challenges to internationalization in African higher education. *South African Journal of Higher Education*, Vol. 20, No. 1:118-142
- Nederhof A.J. (1989) Books and Chapters are not to be neglected in measuring research productivity. *American Psychology*, 44:734-735
- Porter A. L. 1977. Citation analysis: queries and caveats. *Social Studies of Science*, 7: 257-267
- Prinsloo P., Louw H.A. 2006. Being educated in the twenty-first century: An exploration. *South African Journal of Higher Education*, Vol. 20, No. 2:288-298
- Rudd E. 1983. The problem of allocation of funds for university research and assessment of its quality. University of Essex, Colchester
- Rushton J.P., Endler N.S. 1977. The scholarly impact and research productivity of departments of Psychology in the United Kingdom. *Bulletin of the British Psychological Society*, Vol. 30: 369-373
- Ruskin and the RAE 2006. Philosophy of real Mathematics.
www.corfield.pwp.blueyonder.co.uk
- Smith D.M. 1986. UGC research ratings; pass or fail? *Area*, Vol. 18: 247-250
- Smith A.G Benchmarking Google Scholars with the New Zealand PBRF research assessment exercise
- Spee A. and Borman R. (Eds) (1991) *Performance Indicators in Government Institutional Relations* (Paris OECD)
- The Higher Education Qualification Framework. 2004. Department of Education, Republic of South Africa. *Government Gazette*, Vol. 469, No. 26559:1-29

- University Funding Council research Assessment Exercise 1992, the outcome, London.
 University. Funding Council Circular 26/92, 1992
- Van Raan A.F.I. (1993) Advanced bibliometric methods to assess research performance and scientific development: Basic principles and recent practical application, *Research Evaluation*, 3: 151-166
- Wallmark J.T. and Sedig K.G. 1986. Quality of research measured by citation methods and by peer review – a comparison. *IEEE Transactions on Engineering Management*, Vol. 33: 218-222
- Webster D.S. 1981. Methods of assessing quality. *Charge*, Vol. 13, No. 7: 20-24
- Westerheijden D.F. 1997. A solid base for decisions. Use of the VSNU research evaluation in Dutch universities. *Higher Education*, Vol. 33: 397-413

References unaccounted for:

Cherry	2008
Corfield	2006
Gottshalk	1969
Elton	2000
Itmann <i>et al</i>	2007
May	2002
Tanczos	2002

Appendix .1 SAPSE 110 formula (taken from Reinecke, 1993)

$$FG = (K1.SH + K2.SN) + (K3.AH + K4.AN) + (K5.IH + K6.IN)$$

Where:

SH = subsidy for students in humanities

SN = subsidy for students in natural sciences

AH = number of articles published in humanities

AN = number of articles published in natural sciences

IA and IN = increase in potential effective subsidy students in humanities and natural sciences respectively

K = Coefficient determined by factors of cost units such as service workers and equipments.