

Global excellence at the expense of local relevance, or a bridge between two worlds? Research in science and technology in the developing world

H. Barnard^a, R. Cowan^b, M. Müller^c

^a*Gordon Institute of Business Science, University of Pretoria, 25 Fricker Road, Illovo, Johannesburg, 2146 South Africa, barnardh@gibs.co.za*

^b*BETA, Université de Strasbourg, France; UNU-MERIT, Maastricht University, Netherlands*

^c*Chair of Systems Design, ETH Zurich, Switzerland*

Abstract

Do world-leading researchers from developing countries contribute to upgrading locally, or do they disengage from the local context? The paper investigates the scientific collaborations of university-based science and technology researchers in the database of the South African National Research Foundation (NRF), and analyses the co-authorships of researchers who were ranked by the NRF during the 2001 to 2007 period. To establish the extent to which a researcher can access knowledge outside the South African academic science and technology research community, and share it inside that community, we develop a measure of ‘gatekeeping’. The evidence suggests that there is not a local/global trade-off in knowledge creation in academia in the developing world, and that the world-leading researchers in developing countries may play an especially important role as conduits of new knowledge in their country.

Keywords: Science and technology, research, upgrading, developing country, collaboration

1. Introduction

This paper documents the role of world-leading local researchers in the science and technology system of a less developed country. Using data from South Africa, the paper finds that world-leading researchers help accelerate the introduction of new knowledge into their country because they are better connected to external sources of knowledge than less competent local researchers. Additionally though, while they

are not better connected locally than other local researchers, there is no evidence that world-class researchers are worse connected locally than other researchers within their country. This allows for the knowledge created by the world-leading researchers of a country to spread through the local research community. This combination of strong external and strong internal connections means that world-leading researchers are central in acquisition and distribution of frontier knowledge.

The literature consistently points out that upgrading in developing countries relies on knowledge networks that are connected both globally and locally (Lall, 2001; Marin and Bell, 2006; Narula and Dunning, 2000, p. 200). Developing countries benefit from access to more advanced sources of science and technology through global connectedness, provided there are mechanisms to allow the new knowledge to be shared locally within the country (Pack, 2000). World-leading researchers in developing countries seem ideally placed to play such a connector role: They are based in the developing country, but their international standing suggests that they are also engaged at the frontier of knowledge creation. However, the literature documents several reasons why globally connected but locally disconnected enclaves may develop around the most competent researchers in a developing country. Such enclaves have been previously documented in international business contexts in less developed countries, especially in those cases when quality is of particular importance (Akbar and McBride, 2004; Bowen et al., 2002; Feinberg and Majumdar, 2001).

Given the scarce resources in developing countries, the world-leading researchers may well receive research resources at the expense of less recognised local researchers who could, either because of their more limited access to resources, or because of perceptions of unfairness, disengage from the research process. Second, the research priorities of a developing country do not necessarily coincide with those of the international research community, and researchers who can be described as ‘world-leading’ may have achieved their international recognition at the expense of locally relevant research (Leung, 2007). Finally, it requires time and effort to collaborate, and in some cases co-authorships with less competent partners are a drag on researcher productivity (Lee and Bozeman, 2005). World-leading researchers from the developing world who are engaged in the leading global knowledge networks of their field may because of practical considerations not have time to collaborate with local researchers.

However, there is also evidence suggesting that enclave formation is not likely to be a major issue. Allison and Long (1990) point out that research productivity and recognition tend to follow from favourable departmental conditions rather than the reverse. This implies that there is likely to be fairly equitable access to research resources in a department even if researchers have differing levels of competence. The creation of breakthrough knowledge is characterised by the presence rather than the absence of tensions between different knowledge domains, provided that knowledge from different technological or geographical origins can be integrated. This means not only that the local/global tension may be very productive, but also that there is an incentive for researchers to work towards connecting rather than disconnecting the worlds within which they work. Finally, Zuckerman (1967) finds evidence that more competent researchers are also better collaborators. If the world-leading researchers in a less developed country are indeed particularly skilled at collaboration, they should be able to connect with researchers both locally and abroad.

This paper uses concepts from social network theory to better understand connectedness in the academic science and technology system, both internally, i.e. among researchers in the system, and externally, with academics in other countries and with industry. It contributes theoretically to the sociology of science in developing countries, and specifically provides evidence that world-class knowledge creation in under-developed contexts does not take place at the expense of local concerns. Instead, the paper provides evidence of the local value of world-leading researchers in the research network of a developing country and in so doing, contributes to the debate about ‘local relevance’. The paper suggests that policymakers need not be too concerned about the risk that expectations from the developed world may detract from the usefulness of research for a developing country (Leung, 2007).

The paper investigates the scientific collaborations of university-based science and technology researchers in the database of the South African National Research Foundation (NRF), and analyses the co-authorships of researchers who were ranked by the NRF from 2000 to 2006. To establish the extent to which a researcher can access knowledge outside the South African academic science and technology research community, and share it inside that community, we develop a measure of ‘gatekeeping’ which considers both internal and external connectedness. The evidence suggests

that there is not a local/global trade-off in knowledge creation in academia in the developing world, primarily because the leading researchers in developing countries are very productive. Because of their large number of collaborative projects, where both international and local researchers participate, the world-leading researchers in developing countries are likely to play an especially important role as conduits of new knowledge in their country.

2. Science in emerging economies

The importance of science and technology in driving economic upgrading has been recognised at least since the rise of the Asian Tigers (Bell and Pavitt, 1997; Kim and Nelson, 2000; Lall, 2001; Lundvall and Borrás, 2005). Nelson (2005) argues that the role of science, and concomitantly, the role of public research institutes and universities, is becoming even more important in the twenty-first century. He argues that they not only provide training for increasingly science-intensive industries, but also that ‘having a domestic base of good scientists provides the basis for breaking into the international networks where new technologies are being hatched’ (2005:25).

Nelson’s statement reflects a leitmotif in the work on science and upgrading: Scientists need to be connected both locally and globally to achieve local upgrading. Global connectedness allows researchers to accelerate their learning by tapping into advanced knowledge creation processes, while local connectedness allows for their learning to be widely disseminated. But there are three challenges that could prevent the goal of both local and global connectedness from being achieved: Inadequate local expertise and capacity, a tension between local relevance and global excellence, and practical constraints limiting the number of collaborators that a given researcher can have.

With regard to local expertise, the argument on ‘absorptive capacity’ (Goode, 1959; Cohen and Levinthal, 1989) suggests that to the extent that the gap between local and global scientific capacity is too great to be bridged, countries are unlikely to benefit from international contact. Pack (2000) points out that decades of heavy investment in R&D in India and Latin American countries like Argentina did not result in economic upgrading, and instead did little more than create “pockets of high capacity” (Pack, p. 75) or local research excellence, as the local skills base was

not adequately developed to absorb and disseminate the newly created knowledge.

Organizations can employ a variety of strategies to overcome the constraints of an underdeveloped local capacity base. Research in international business provides evidence that world-class institutions in a developing country, especially when knowledge creation and quality are critical concerns, can easily develop as essentially disconnected enclaves, and that they provide only limited benefits to the country in general. This can be the case both for subsidiaries of foreign MNCs (Akbar and McBride, 2004; Bowen et al., 2002) and in local industries and leading firms (Barnard et al., 2009; Feinberg and Majumdar, 2001).

What are the implications of these macro-level concerns for the individual researcher working in a developing country? In developing countries, problems are experienced with a range of (sometimes fairly basic) elements of the research process: Contacting people, transmitting information, keeping information secure and keeping others informed of progress (Duque et al., 2005). In a middle income country like South Africa more sophisticated forms of technology, e.g. specialised scientific instrumentation, are often quite expensive and not widely available.

To enable researchers in developing countries to work at the global frontier of knowledge, their home institutions must deliberately invest in infrastructural capacity for research. Broad-based research support could also benefit researchers working behind the technological frontier. However, because of the resource constraints typical of developing countries, a significant prioritisation of support seems likely. Targeting the bulk of support at world-leading researchers may appear to be a fair way to prioritise scarce resources, whether or not their colleagues benefit from that decision. If investments take place at the cost of other researchers, for example if non-leading researchers are marginalised in the allocation of scarce funds or skilled human resources, the formation of disconnected enclaves is again possible. But if equipment were procured and made available to all researchers in a unit, leading and non-leading researchers alike are likely to benefit.

Prioritising support for leading researchers also introduces concerns of fairness: Evidence from the US suggests that equity in work arrangements matters to academics. In their study of wage inequality in US universities, Pfeffer and Langton (1993) find that a high dispersion in salaries is associated with lower research pro-

ductivity overall and fewer collaborations. Researchers in a less developed context may have a greater tolerance for unequal work arrangements, but to the extent that they do not, there is additional reason to expect that world-leading researchers in the developing world may ultimately work in enclaves.

A second consideration that complicates achieving both local and global connect- edness is the tension between local relevance and global excellence. The work that is important and needed in a developing country may or may not be relevant in the scientific community in general. For example, Nelson (2005) points out that in fields like agriculture and medicine, differences in soil and climate conditions or prevalent diseases require the development of different technologies — and therefore a research agenda that does not reflect global’ preoccupations. Duque et al. (2005:769) put it this way:

[S]cientists in developing regions face another issue, unknown to those in devel- oped regions. Do they strive for recognition in the international scientific community? In this case, what matters is publication in international journals, the ‘gold standard’ for tenure in universities or high performance evaluations in national research insti- tutes. Or do they focus on publication in local outlets, invisible to scientists in the international community, but important for the dissemination of information in the regional context?

This tension between local relevance and global excellence could be actual or perceived. For example, Leung (2007) expresses a concern that the dominance of a Western paradigm may hamstring Asian-focused research in management studies. Whether subsequent empirical work bears out that belief or not, may not be that important. Instead, it matters whether researchers in a country behind the techno- logical frontier believe that they need to minimise their engagement with local issues in order to advance to the frontier in their field. If researchers believe that local en- gagement could be at a cost to global recognition, whether their belief is well-founded or not, their interaction with other local researchers and impact in their own country are likely to be limited. Therefore the (perhaps actual, perhaps perceived) tension between global and local needs could limit the benefits of world-class research for a

developing country.

Additionally, there are the ‘transaction costs’ (Lee and Bozeman, 2005) associated with collaboration:

Staying in touch by various media, engaging in social ingratiation, waiting for others to comment, respond, or do their part of the research — these are just some of the factors taking time and energy even in the best collaborative relationships. (2005:674).

These elements of collaboration do not require skill as much as time, and there is therefore little reason to expect that leading researchers will be exempt from these types of costs. In addition, researchers from developing countries with a less reliable scientific, internet and mail infrastructure often have a greater burden of these types of transaction costs. Face-to-face contact becomes more important, but researchers must travel longer distances to attend conferences, use facilities or meet with co-authors in (typically) the developed world. If relationships with less competent (local) colleagues “are a drag on the productivity of the more experienced researchers” (Lee and Bozeman, 2005:674), researchers may prioritise with whom they enter into research relationships and limit the amount of work they do with local colleagues.

However, counter-arguments to each of these three points suggest that fears of enclave formation around world-leading researchers in developing countries may be exaggerated or even unfounded.

As regards concerns about local capacity, Allison and Long (1990) find evidence that favourable departmental conditions are important drivers of research productivity and prestige (rather than the reverse, where prior productivity provides access to greater departmental resources). While they recognise the possibility of a virtuous cycle — where world-leading researchers attract additional resources — Allison and Long argue that departmental resources include intellectual stimulation and especially research facilities, in other words, the types of resources that could benefit the entire complement of researchers in a unit. In this context, concerns about fairness and disengagement by other local researchers are less likely to be relevant.

The argument about the tension between global excellence and local relevance does not take into account that breakthrough knowledge emerges particularly when there

is a recombination of hitherto disconnected technological or geographically located knowledge fields (Phene et al., 2006; Rosenkopf and Nerkar, 2001). In other words, although there is a local/global tension, it could be a highly productive tension with the potential for globally groundbreaking insights. This does not mean that it is easy to harness the potential of that tension; indeed, it seems likely that only the most competent researchers would be able to meaningfully link quite different knowledge worlds. But there is consistent evidence in the literature that knowledge creation takes place best if there is some ‘middle ground’ — a context where knowledge is new but not entirely unfamiliar (Fleming, 2001; Katila and Ahuja, 2002; Miller et al., 2007). This creates a strong incentive for leading researchers in the developing world to try and create a ‘middle ground’ between local and global knowledge practices. Researchers in a developing country are likely to find it easiest to create such a shared context by ensuring that local colleagues are familiar with global knowledge bases and research practices (rather than the reverse; trying to interpret local specificities for the global research community).

As regards the more pragmatic consideration that constraints on the number of co-authors may force leading developing country scholars to work only with foreign partners, Zuckerman (1967) provides evidence that leading researchers are more inclined to collaborate and have more productive research collaborations than other researchers. Zuckerman compared Nobel prize winners with another group of researchers who could be seen as world-leading, which makes generalization a bit tricky. However, this work does raise the possibility that world-leading researchers may be especially collaborative in their research orientation. This raises the possibility that the leading researchers in developing countries may feel the resource constraints on collaboration less severely than do their colleagues, so they may be able to collaborate with both local and international research partners and therefore act as local/global connectors in the knowledge network.

In short, the literature provides reasons to expect that world-leading researchers in developing countries could be particularly important connectors between local and global knowledge networks, but also that they may be connected to global research networks at the expense of connectedness to their home research environment. The goal of this study is to provide empirical evidence about these contradictory pre-

dictions, and to highlight the local consequences of the creation of global frontier knowledge in a scientifically and economically lagging country.

3. South Africa

3.1. International standing

The study investigates the patterns of collaboration among researchers in the academic community in South Africa, and specifically in science and engineering. Because the role of academic gatekeepers is arguably different in a world-leading versus a lagging research context, it is important to assess the international standing and impact of South African research. Citation rates are a commonly-used measure of the impact of research and used to establish quality in a variety of knowledge production contexts (DuBois and Reeb, 2000; Jaffe, 1989; Starbuck, 2005).

The relative ISI citation ranking reveals clearly that most of South African science is behind the technological frontier (see Table 1). The impact of most South African research is similar to the impact of work done in peer countries, other leading emerging economies like Argentina, Brazil, India and Mexico. Even in the best performing disciplines, the work of most South African researchers remain quite far from the frontier, and has a comparable impact to that of researchers in the non-core disciplines of small late-coming economies like Ireland, Israel, Finland and Taiwan.

Similarly, a review of the average citation rate of publications by South African authors in ISI relative to the average citation rate of ISI publications overall for 1995-2008 reveals that the citation rate of South African papers is about 74 percent of that of the rest of the world. (See Figure 1.) The only discipline in which the South African average citation rate is above the average for ISI is Computer Science, where South African papers received on average 3.1 citations, and papers in ISI overall 2.76.

The evidence from ISI data suggests that none of the world-leading South African researchers work in a discipline where South Africa is a global leader, and that they are unlikely to have many colleagues whose performance is, by global standards, above average. South African researchers — both the world-leading researchers and their less skilled local colleagues — are likely to benefit from being connected to researchers in the rest of the world and the cutting-edge knowledge developed elsewhere.

Table 1: South African research competitiveness by discipline: relative ISI ranking (2008)

Discipline	National citation ranking
Multidisciplinary	16 (Israel)
Environment and ecology	22 (India)
Plant and animal sciences	23 (Finland)
Geosciences	24 (Finland)
Immunology	24 (Ireland)
Psychiatry/psychology	25 (Brazil)
Microbiology	28 (Taiwan)
Clinical medicine	31 (Wales)
Space science	32 (Ukraine)
Agricultural sciences	34 (North Ireland)
Pharmacology and toxicology	37 (Ireland)
Biology and biochemistry	38 (Turkey)
Molecular biology and genetics	38 (Argentina)
Computer science	38 (Mexico)
Neuroscience	39 (Chile)
Mathematics	41 (Ireland)
Chemistry	43 (Slovakia)
Engineering	44 (Slovenia)
Materials science	44 (Slovakia)
Physics	47 (Iran)
All fields	38 (Argentina)

Note: Similarly ranked countries in parentheses

Source: ISI (now Thompson Reuters) Web of Knowledge (accessed March 2009)

3.2. National rating of researchers

This study relies on data on ‘rated researchers’ from the National Research Foundation (NRF) of South Africa. The NRF (www.nrf.ac.za) is a state agency that has as its mission the provision of leadership in the promotion of research and development of research capacity. One of its key roles is to facilitate the ranking of researchers at universities. The ranking process is similar to the tenure process at North American universities, but considers only research (not teaching or service to the institution), is centrally administered (not by the institution) and is valid only for a set period (the exact period has changed over time, but is around five years). Once rankings lapse,

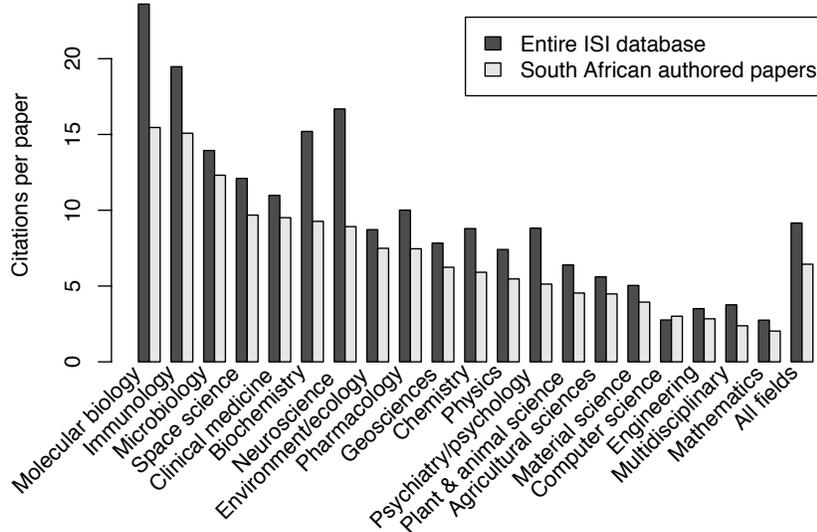


Figure 1: Average citation rate of South African-authored ISI publications relative to the average disciplinary citation rate overall (2008). Source: ISI (now Thompson Reuters) Web of Knowledge (accessed March 2009)

researchers must reapply for ranking, which may be higher, lower or the same as the previous round. ‘Rated researchers’ qualify for preferential access to research funding and also receive institutional recognition, as NRF-ratings are viewed by universities as an objective inter-institutional indicator of research capacity. There are therefore strong incentives for active researchers to apply for rating.

Becoming a ‘rated researcher’ is an achievement in itself, but within the ranking system there is also a clear hierarchy. There are three categories for established researchers: world leader (A), world participant and local leader (B) and local participant (C). For young researchers, i.e. within five years of obtaining their doctorate, there are also three categories: potential world leader (P), potential local participant (Y) and latecomer (L), a type of affirmative action category where providing a development plan for the researcher is a precondition for ranking.¹ See Table 2 for the

¹Our data contain these 6 categories, and two more: Rating Unsuccessful, and Not Processed.

NRF definition of each category.²

As part of the application process, researchers provide the NRF with evidence of research outputs. Although assessment is based primarily on peer-reviewed publications, an application would detail journal articles, books, conference presentations, patents, policy or technical reports, and publications in the public press. It is important to note that rankings are not based solely on volume of outputs or citation rates.³ A specialist review panel consisting of local subject experts chooses three reviewers from a list of six to ten reviewers suggested by the applicant, and three additional reviewers. Each reviewer assesses the research outputs and both an independent assessor and the specialist review committee will consider the six referee reports before meeting to assign a rating. When consensus cannot be obtained, or when the recommendation is to accord an A or P (i.e. world-class or potential world-class) ranking, the decision is escalated to an executive evaluation committee. Finally, there is an appeals process for the cases where a researcher disagrees with an assigned ranking.

NRF rankings are a useful indicator of the quality of researchers in the academic community in South Africa because of the rigour of the review process, the fact that the NRF filings are detailed with little missing information, and because nearly all active researchers in the South African academic research community are NRF-ranked. The NRF categories also explicitly relate the primary contribution of researchers to either the global or local context. A-rated and to a lesser extent, B-rated researchers make a contribution to frontier knowledge internationally, whereas the focus of C-rated researchers is primarily local. The data are therefore well-suited to explore the role of world-leading researchers in a technologically less developed context.

3.3. Co-authorship network

The study focuses on researchers working in scientific fields at South African universities. All NRF filings from the beginning of 2000 until the end of 2006 are available

²The NRF ranking system was developed in 1982 for the engineering and scientific fields in South Africa, and the system has evolved over the past three decades: the social sciences were included from 2002 and the L category was introduced after the end of Apartheid. The L category is currently in the process of being phased out. The policies reflected in this article are those that were in place for the period of data gathering, 2000 to 2006.

³Some scholars have received B and even A-ratings on the basis of a small but compelling body of work, and sometimes on as little as a single groundbreaking publication.

Table 2: Definitions of research ratings

Category	Definition
A	Leading international researcher Researchers who are unequivocally recognised by their peers as leading international scholars in their field for the high quality and impact of their recent research outputs.
B	Internationally acclaimed researcher Researchers who enjoy considerable international recognition by their peers for the high quality of their recent research outputs.
C	Established researcher Established researchers with a sustained recent record of productivity in the field who are recognised by their peers as having produced a body of quality work, the core of which has coherence and attests to ongoing engagement with the field as having demonstrated the ability to conceptualise problems and apply research methods to investigating them.
P	NRF President's Awardee Young researchers (younger than 35 years of age), who have held the doctorate or equivalent qualification for less than five years at the time of application and who, on the basis of exceptional potential demonstrated in their published doctoral work and/or their research outputs in their early post-doctoral careers are considered likely to become future leaders in their field.
Y	Promising young researcher Young researchers (below 40 years of age), who have held the doctorate or equivalent qualification for less than five years at the time of application, and who are recognised as having the potential to establish themselves as researchers within a five-year period after evaluation, based on their performance and productivity as researchers during their doctoral studies and/or early post-doctoral careers.
L	Late entrant into research (category terminated at the end of 2009) Persons (normally younger than 55 years) who were previously established as researchers or who previously demonstrated potential through their own research products, and who are considered capable of fully establishing or re-establishing themselves as researchers within a five-year period after evaluation.
R.U.	Rating unsuccessful
N.P.	Not processed

Source: <http://www.nrf.ac.za/projects.php?pid=34> (accessed August 2010)

in electronic form and used for the analysis. One file corresponds to one application of a researcher for rating. Each file includes socio-demographic information (i.e. age, sex, race, and marital status) as well as professional information (i.e. title, scientific field, and list of publications) on the researcher and the rating by the NRF. The NRF assigns a unique identifier to each applying researcher. By the end of 2006, 1338 active researchers were rated by the NRF, which corresponds to nearly the whole active South African academic research community in those scientific fields. Because applications need to be renewed every five years, 38 researchers applied twice within the observation period. In these cases, we include the latest submission in the sample which yields the latest NRF-rating and the most comprehensive list of publications.

An application for rating always includes a list of publications provided by the researcher. Each publication entry includes information on the type of publication (e.g. conference proceeding or peer-reviewed article), year of publication, and the names of the co-authors. We select all 21403 peer-reviewed articles published within the observation period beginning 2000 to end 2006. A peer-reviewed article by definition codifies a unique piece of knowledge and therefore identifies a single collaboration project. The restriction to peer-reviewed articles avoids double counting of collaborations as they might also have manifested in other publishing activity such as conference proceedings.

Some collaborations which actually led to publications are missed because articles published after the latest date of submission of the researchers are not included in the NRF-filings. Rankings are based purely on publications at the time of ranking: reviewers are provided with copies of the articles of a researcher being ranked. However, missing articles due to an early year of latest submission might bias our analysis in case the dates of submission and the rating are not independent. The contingency table, Table 3, shows the latest year of submission by the ranking of researchers. A chi-square test including rated researchers (categories A-L) obtains a significance level of 0.57 and, thus, does not reject the null hypothesis that year of submission and rating are independent (Sheskin, 2007). Therefore, we can be confident that missing articles do not bias the analysis severely.

The co-authorship network is constructed from the list of publications provided by each applying researcher. The network is constructed using co-authorship data. Two

Table 3: Year of latest submission by rating

	A	B	C	P	Y	L	R.U.	N.P.	Sum
2001	8	76	107	1	26	10	12	6	246
2002	0	0	10	0	0	0	6	1	17
2003	9	56	119	2	26	7	17	7	243
2004	7	51	103	1	29	4	18	8	221
2005	7	49	108	2	34	5	34	10	249
2006	13	80	171	1	36	11	26	16	354
Sum	44	312	618	7	151	37	113	48	1330

Note: A chi-square test over categories A-L does not reject the null hypothesis of independence at a significance level of 0.57. A chi-square test which includes not only rated researchers (A-L) but also those which have been unsuccessful and not processed rejects the null hypothesis at a significance level of 0.11.

researchers are said to be connected if they have co-authored at least one paper. The weight of the connection is determined by the number of papers they have written together, and by the number of other researchers involved in those papers. The next paragraphs give a detailed description of how the network is created, and discuss respective methodological decisions.

All non-ranked co-authors are considered to be external to the South African academic research network. They are subsumed in one node; the external source of knowledge. Given the centrality of the ratings system among South African university-based researchers, those external knowledge sources are likely to consist either of non-South African academics or possibly South African non-academics. Sooryamoorthy (2009) documents a general preference for international over domestic collaboration among South African researchers. He also documents the overwhelming contribution of the academic environment to published research in South Africa, with industry contributing only 2% of publications, and a similarly small share of co-authors. In a larger-scale study, Bernardes and da Motta e Albuquerque (2003) highlight that countries at a midlevel of development generally have very few links between industry and academia, and it is therefore likely that the external network of South African academics comprises mainly of foreign academics.

The linking of applicants to their co-authors relies on the matching of the co-

authors by their names. In order to obtain a valid matching, the referencing format for all authors was standardized, so that co-authors were not double counted because of different styles of references (John Smith versus Smith, J. versus J. Smith etc.). Where necessary, manual checking was conducted to validate names. In eight cases the same standardized name refers to two different NRF-rated researchers, mostly father and son. In these cases the son is deleted from the sample and all co-author links are assumed to refer to the father.⁴ Thus, 1330 researchers remain in the sample.

Each ranked researcher corresponds to one node in the South African research network, and two researchers are linked if they have jointly authored at least one paper. The value of a paper is weighted by the number of co-authors which contributed to it. This reflects the idea that the fewer co-authors contribute to one paper, the more intense the collaboration between any two co-authors is likely to be. Specifically, we follow the weighting approach of Newman (2005). He argues that a researcher, a priori, divides her collaboration time on a particular project equally among her collaborators. For example, a paper published by three authors connects those authors with intensity $1/2$ because each author spends half her collaboration time with each of her collaborators. In general, if n authors contribute to one paper, they are all connected by links with weight $1/(n - 1)$. A technical attraction of this weighting rule is that the weighted degree of a researcher equals the number of co-authored papers. The weighted link between two researchers is determined by summing the co-authorship weights of all the papers on which they appear as co-authors. This yields the network accumulated over the years 2000 to 2006, used in the analysis. Table 4 provides summary statistics on the co-publication network. Excluding external links and only considering links among South African scientists, we observe a large connected component which connects 877 out of 1330 South African researchers. The rest of South African researchers are connected in groups smaller than or equal to six persons or are isolated within the South African research system. A minority of 89 researchers has neither international nor South African co-authors.

South African researchers collaborate frequently with non-South African researchers. Out of 1330 researchers, 1230 researchers have external collaborations within the ob-

⁴Alternatively, deleting both father and son from the sample does not affect the analysis.

Table 4: Network statistics of co-publication network

Statistic	Value
No. of researchers in scientific fields	1330
No. of links	3651
No. of researchers in main component	877
No. of researchers in 2nd largest component	6
No. of isolates	278
No. of links to external source	1230
Average shortest path in main component (weighted Dijkstra)	9.3

servation period and, on average, a South African researcher collaborates three times more with external researchers than with South African researchers.

4. Analysis of gatekeeping

4.1. Measures

To what extent is a researcher important in bringing knowledge into South Africa? The central argument of this paper is that it will depend both on the researcher’s access to outside knowledge, and his or her ability to spread it within the South African research community. The higher the researcher’s external weight, the better he or she is able to access outside knowledge; the lower the distance of the researcher to other South African researchers, the faster and more effectively the new knowledge will spread within the country.⁵ We create a measure, “gatekeeping”, to reflect this dual process, and apply it at the node level:

$$Gatekeeping = \frac{External\ weight}{Average\ shortest\ weighted\ path}$$

⁵The idea here is that knowledge passes through scientists collaborating, and the fewer steps between a “sender” and a “recipient” and the denser each collaborative step, the more effective and efficient will be that transmission. This assumption maintains the focus on one type of knowledge transmission, and ignores others, namely conference or seminar participation and journal publication. Access to journals is typically made at the institutional level so there should be little difference in that regard among researchers of different rank. Additionally, there is often a considerable lag between discovery and publication, so much “recent” knowledge is likely to be communicated through other channels.

External weight is the weight of the link to the external source, i.e. the intensity of collaboration with non-South African scientists. The average shortest weighted path is calculated among the South African researchers in the largest connected component of the network using the Dijkstra algorithm where the weight of one link corresponds to the inverse of the collaboration intensity between two researchers.⁶

The NRF rating is the central independent variable. The main interest of this paper is the extent to which researchers with higher global relevance (e.g. with A or B ratings) also act as sources of novel knowledge and as gatekeepers in the South African research system generally. The number of papers (*Articles*) a researcher publishes is a common productivity indicator. Because research output and scientific impact often coincide, the number of articles may be considered as an alternative to the NRF-rating. However, in our context the number of articles is interpreted not only as an additional proxy for researcher quality. Because the co-authorship network is constructed from publications, the articles of a researcher directly generate the outcome variable, gatekeeping.

Controlling the rating variable by articles therefore is likely to have a strong effect. On the one hand, it subtracts the rate of scientific output from the scientific impact as it is rated. On the other hand, it somewhat normalizes gatekeeping by the number of articles. A researcher with many articles will, on average, have many external co-authors and many internal co-authors. The former increases the numerator of gatekeeping, the latter is correlated with a small denominator. What including the number of articles as an explanatory variable does is to reveal the effects of the structure of a researchers collaboration network: whether, for example, A researchers have specific locations in the network (lying between two groups for example) that

⁶Average weighted shortest path is defined in the following way. Between any two researchers i and j are one or more paths in the collaboration network. A path is a list of researchers, (i, k, l, \dots, m, j) where consecutive researchers in the list have collaborated at least once. The shortest weighted path between i and j is the path which minimizes the sum $1/w_{i,k} + 1/w_{k,l} + \dots + 1/w_{m,j}$, where $w_{k,l}$ is the link weight (collaboration strength) between k and l . The idea here is that pairs of researchers who have many collaborations will constitute effective channels of transmission and we are interested in the most efficient channel between two agents. We take the inverse of the link weight to generate a ‘minimum cost’ rather than a ‘maximum efficiency’ (the latter being undefined in most efficient algorithms). The average shortest weighted path for agent i is simply the mean of the shortest weighted paths between i and all other connected agents.

permit them to transmit knowledge effectively.

Further control variables used are the scientific field of the researcher, *Title*, *Race* (given the racialised history of South Africa), *Gender*, and *Age*. These are all factors which may influence both the NRF-rating as well as the gatekeeping score. For 15 researchers not all covariates are available and these are excluded from estimation.⁷ Table 5 provides the bivariate Pearson correlations for variables used in estimations.

4.2. Estimation and testing

Network data violates the assumption of independent observations on which standard asymptotic inference relies, because individuals are connected and therefore explicitly not independent from each other. We take care of this issue in two steps. In a first step, a linear regression model with autoregressive disturbances is estimated. The coefficient estimates are the same as those of ordinary least squares but with reliable standard errors. The estimates of the standard errors hinge on assumptions of the autoregressive effects. Therefore, in a second step, testing is based on a permutation test which does not invoke such assumptions. The following gives a more detailed account of the linear regression and the permutation test.

The model we estimate is a standard model in spatial econometrics where it is commonly referred to as the spatial autoregressive error model (see for example Anselin, 1988):

$$y_i = \mathbf{x}_i' \alpha + \mathbf{z}_i' \beta + u_i \tag{1a}$$

$$u_i = \rho \sum_{j \in N_i} u_j + \epsilon_i, \tag{1b}$$

The first equation models the dependence of gatekeeping (y_i) on rating (the dummy vector \mathbf{x}_i) while controlling for several individual characteristics (\mathbf{z}_i). In the second equation, the disturbance of agent i , u_i , is modelled as a linear combination of the disturbances of its neighbours, N_i , plus an independent and identically distributed

⁷More specifically, for five researchers *Title* and for ten researchers *Age* is missing. Alternative estimations which include these 15 observations and instead exclude the two variables yield the same general insights as the results presented.

Table 5: Mean, standard deviations and Pearson correlations for variables used in estimations

	Mean	S.D.	1	2	3	4	5	6	7	8	9	10
Gatekeeping	0.75	1.24										
A-rated	0.03	0.18	0.20***									
B-rated	0.24	0.43	0.22***	0.10***								
C-rated	0.46	0.5	-0.10***	-0.17***								
P-rated	0.01	0.07	0.02	-0.04	-0.52***							
Y-rated	0.11	0.32	-0.07*	-0.07*	-0.20***	-0.07*						
L-rated	0.03	0.16	-0.06*	-0.03	-0.09***	-0.15***	-0.03					
Rating Unsuccessful	0.08	0.28	-0.11***	-0.06*	-0.17***	-0.28***	-0.02	-0.06*				
Not processed	0.03	0.18	-0.10***	-0.04	-0.11***	-0.18***	-0.01	-0.07*	-0.05			
Biochemistry	0.04	0.19	-0.03	-0.01	0.01	-0.06*	-0.01	0.07*	-0.03	-0.06*		
Cell Biology	0.02	0.14	<0.01	-0.03	-0.02	-0.01	-0.01	<0.01	0.02	-0.02	0.05	-0.03
Chemistry	0.09	0.29	0.04	-0.03	-0.01	-0.06*	-0.02	-0.02	0.05	0.07**	<0.01	-0.06*
Earth Science	0.08	0.27	-0.02	-0.01	0.05	0.01	-0.02	-0.03	-0.03	-0.03	-0.03	-0.06*
Engineering	0.13	0.34	-0.14***	0.02	-0.03	0.01	<0.01	0.03	-0.01	-0.03	0.02	-0.08**
Health Science	0.16	0.37	0.05	-0.02	-0.09***	0.07*	0.02	<0.01	0.07*	0.07*	-0.02	-0.09**
Mathematical Science	0.12	0.32	-0.17***	0.02	0.01	<0.01	-0.03	-0.04	0.03	<0.01	0.01	-0.07**
Microbiology	0.05	0.23	0.10***	0.01	-0.01	<0.01	-0.02	0.03	0.02	-0.01	-0.03	-0.05
Physics	0.09	0.28	-0.08**	0.03	0.10***	-0.06*	0.01	-0.01	0.02	-0.07**	0.02	-0.06*
Plant Science	0.07	0.26	0.11***	<0.01	0.04	-0.02	0.02	<0.01	-0.03	<0.01	-0.04	-0.06*
Dr.	0.28	0.45	-0.16***	-0.12***	-0.22***	-0.13***	0.02	0.38***	0.07*	0.07*	0.15***	0.02
Coloured	0.03	0.18	-0.06*	-0.03	-0.08**	0.03	-0.01	0.09***	0.02	-0.01	-0.01	0.01
Indian	0.06	0.23	-0.02	-0.03	-0.06*	-0.01	-0.02	0.03	0.02	0.08**	0.02	<0.01
White	0.83	0.37	0.11***	0.07**	0.17***	0.03	-0.02	-0.10***	-0.07*	-0.13***	-0.09***	0.01
Male	51.34	9.56	0.03	0.16***	0.26***	0.08**	-0.09***	-0.48***	-0.02	-0.01	-0.07*	-0.04
Age	0.78	0.41	-0.02	0.08**	0.08**	0.04	-0.01	-0.14***	-0.12***	-0.02	0.01	-0.06*
Articles	16.88	17.3	0.88***	0.28***	0.25***	-0.11***	0.02	-0.09**	-0.08**	-0.13***	-0.13***	-0.05*
(12)	-0.05											
(13)	-0.04											
(14)	-0.06*	-0.09**										
(15)	-0.06*	-0.12***	-0.11***									
(16)	-0.05	-0.14***	-0.13***	-0.17***								
(17)	-0.03	-0.08**	-0.07*	-0.14***	-0.16***							
(18)	-0.04	-0.10***	-0.09**	-0.09***	-0.11***	-0.09***						
(19)	-0.04	-0.09**	-0.08**	-0.12***	-0.13***	-0.12***	-0.07**					
(20)	0.02	0.02	0.06*	-0.09**	-0.10***	-0.10***	-0.09**	-0.09**				
(21)	0.03	0.05	-0.05*	-0.02	<0.01	-0.03	0.06*	0.01	0.10***			
(22)	-0.04	-0.01	-0.03	0.13***	0.05	0.01	-0.04	-0.03	-0.03	-0.05		
(23)	0.01	-0.11***	0.09**	0.02	-0.10***	-0.05	0.03	0.03	-0.08**	-0.41***		
(24)	<0.01	-0.01	0.01	0.04	-0.08**	0.02	-0.01	-0.37***	-0.13***	-0.08**	0.18***	
(25)	0.01	0.05*	0.11***	-0.25***	0.06*	-0.06*	0.08**	-0.17***	0.04	-0.01	-0.06*	0.17***
(26)	-0.01	0.02	-0.03	-0.10***	-0.10***	-0.07*	-0.03	-0.20***	-0.06*	-0.02	-0.09**	0.06*

Note: N = 1315 researchers. **, *** signify 5%, 1% and 0.1% rejection levels of significance.

(i.i.d.) error term ϵ_i , with $E[\epsilon_i] = 0$ and $E[\epsilon_i^2] = \sigma_\epsilon^2$. The coefficients (α, β, ρ) and the variance of the error term (σ_ϵ^2) are estimated via (quasi) Maximum Likelihood.⁸

The validity of the estimated standard errors hinges largely on the assumption that the errors ϵ_i are independently distributed. We test this assumption with Moran’s I test statistic, the most common statistic to test for spatial and network autocorrelation (see for example Anselin, 1988; Arbia, 2006). In short, it measures the extent to which a random variable is correlated over neighbours.⁹ Under the null hypothesis that errors are independent over neighbourhoods of the network, the I -statistic of regression residuals is asymptotically normally distributed with known formulas for mean and standard error (Arbia, 2006:91). This yields a model specification z-test which is presented jointly with the regression results.

The main hypothesis is that researchers of higher rating take more central gatekeeping positions in the collaboration network. Under the null hypothesis (H_0) the gatekeeping score is independent of rating. The alternative hypothesis (H_α) is that when two researchers are differently rated but otherwise identical, the higher rated researcher can be expected to have a higher gatekeeping score. For example w.r.t. the rating categories A and B , we write this one sided hypothesis as $H_0 : \alpha_A = \alpha_B$ and $H_\alpha : \alpha_A > \alpha_B$. The familiar Wald-test statistic of this hypothesis is

$$W = \frac{\hat{\alpha}_A - \hat{\alpha}_B}{\hat{\sigma}_A^2 + \hat{\sigma}_B^2 + 2\hat{\sigma}_{A,B}^2},$$

where the hat indicates that coefficients and standard errors are estimated. Under the null hypothesis W is asymptotically chi-square distributed with one degree of freedom (see e.g. Cameron and Trivedi, 2005:225).

However, the Maximum Likelihood estimate of the coefficients’ standard errors assumes a certain dependence structure, i.e. a linear effect of the first order neighbourhood. If this assumption does not hold, standard errors are not consistent and

⁸In order to formulate the conditional likelihood of the sample, the errors ϵ_i are assumed to be normally distributed. The term “quasi” reflects that the estimator is likely to be consistent and asymptotically efficient also in case the normality assumption is violated (Arbia, 2006:64). Similar to the unit root restriction in time series analysis, a necessary condition for identification is that $|\rho| < 1$ (Anselin, 1988:63).

⁹Moran’s I is defined as $I = \hat{\epsilon}' A \hat{\epsilon} / \hat{\epsilon}' \hat{\epsilon}$, where $\hat{\epsilon}$ denotes the estimated error vector and A is the weighted adjacency matrix of the co-publication network (Arbia, 2006:91).

the distribution of W in fact is not Chi-square.

Therefore, we apply a permutation test to obtain empirically the sampling distribution of W . Under the null, we permute many times the rating of, say, A and B researchers and calculate each time the Wald-test statistic W^* . This generates the distribution of test statistics that would be observed if the null hypothesis were true, that is, if rating does not matter to gatekeeping. The null hypothesis is then rejected if the reference statistic W calculated on the real data is unlikely to be drawn from the distribution obtained via this permutation. Thus, the significance level of rejection of the null hypothesis is the fraction of times that W^* is larger than W , obtained from the observed data.¹⁰

For this procedure the estimated variances do not need to be consistent. They only scale the difference of the coefficients of interest. Therefore, we may use Ordinary Least Squares (OLS) estimates of equation 1a as well as the Maximum Likelihood estimates of equations 1a and 1b jointly. Because Maximum Likelihood is computationally costly, we use OLS for the permutation test.

In principle, permutation tests may also be based on the permutation of the dependent variable or of the estimated error terms (discussed in e.g. Anderson and Legendre, 1999). We refrain from doing so because both would affect network dependencies which we would like to keep constant. On the other hand, permutation of the rating breaks the correlation between rating and further covariates. The interpretation is similar to a treatment effect. Under the null hypothesis, networking behaviour and characteristics of the researcher is given and the NRF-rating neither informs us nor has an effect on gatekeeping.

4.3. Descriptive statistics

Table 6 gives an overview on publishing activities of researchers by rating category. Out of 1330 researchers, 73% are experienced (A,B,C), 15% are young (P,Y,L) and 12% are not rated (Rating Unsuccessful or Not Processed). Whereas the average number of co-authors per article is similar for all categories, the number of peer-

¹⁰We use the Wald-test statistic W rather than the statistic $\hat{\alpha}_k - \hat{\alpha}_l$ because the former is asymptotically pivotal whereas the latter is not. That this reduces the coverage error (i.e. nominal minus true coverage) has been shown for example for bootstrap methods (see Hall, 1992:12).

reviewed articles published in the observation period differs. Within the group of experienced as well as within the group of young researchers, on average higher ranked researchers publish more.

Table 6: Publishing activity of NRF-ranked researchers in science and technology

Rating	No. of researchers	No. of co-authors per article [Mean (S.D.)]	No. of articles [Mean (S.D.)]
A	44	2.60 (1.41)	40.61 (37.02)
B	312	2.77 (2.17)	23.46 (21.87)
C	618	2.64 (1.74)	14.26 (10.61)
P	7	3.71 (1.55)	21.14 (9.59)
Y	151	2.8 (1.59)	11.95 (9.49)
L	37	2.46 (1.56)	8.65 (7.02)
R.U.	113	2.16 (1.73)	8.67 (10.55)
N.P.	48	2.05 (1.8)	4.83 (4.14)

Note: Peer-reviewed articles published between 2000 and 2006.

Figure 2 shows the boxplots of the distribution of the gatekeeping score by NRF rating. The comparison of the distribution of gatekeeping for the different rating categories shows that the ordering of the median gatekeeping score coincides with the order of the NRF-rating. Therefore, only considering the rating categories we expect a higher gatekeeping score for researchers with higher rating.

A further analysis of the gatekeeping scores suggests that the increased gatekeeping score is the result not of shorter paths internal to the network (in other words, greater connectedness in the South African scientific university-based research community), but rather a larger number of external collaborations. (See Table 7.) Higher rated researchers are engaged in more collaborations with scholars outside their local research community, but do not seem to differ substantially in terms of their local connectedness to other South African researchers. However, the very large standard errors found in the distribution itself indicate that there is large heterogeneity within rating categories.

The econometric results presented in the next section describe how the differences between rating categories are significant, taking into account the dispersion of gatekeeping and the systematic effects of further covariates.

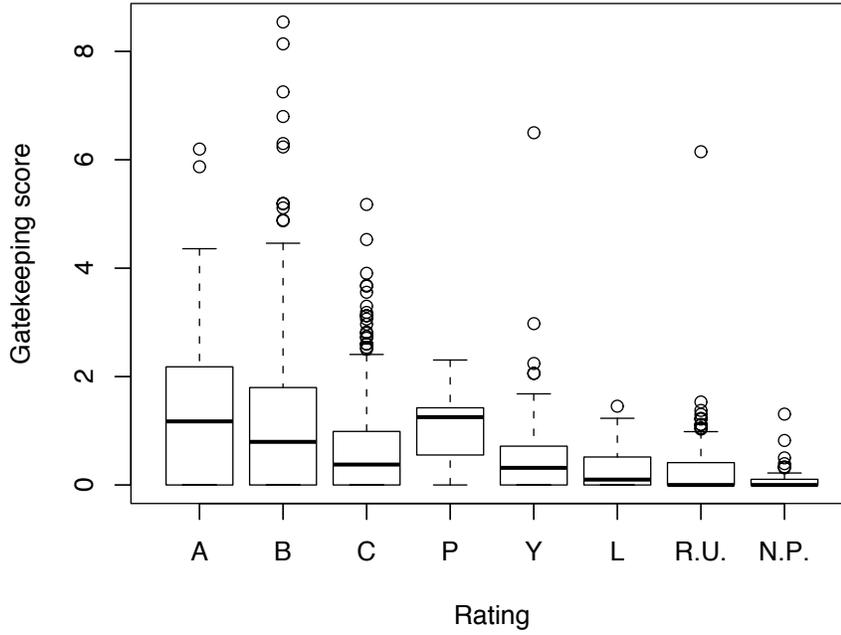


Figure 2: Boxplot of gatekeeping score by NRF-rating. The boxes comprise the lower and upper quartile of respective distributions, the median is marked with a bar, hinges extend the boxes by at most 1.5 the interquartile range. Three researchers, i.e. two A and one B researchers with gatekeeping score larger 10, are not displayed.

5. Results

The main result is that higher ranked researchers have higher gatekeeping scores, and this difference is due to the number of publications rather than due to the pattern of collaboration. Both the regression estimates and the permutation test results show this in a coherent way. This section describes the results.

Table 8 displays the regression results of four models.¹¹

Model 1 includes only the ranking as an explanatory variable. As expected the

¹¹All regression models have a significant network effect in the disturbances ($\hat{\rho}^2$) and the test of Moran's I is never able to reject the hypothesis that the estimated residuals ($\hat{\epsilon}_i$) are independent over neighbourhoods. Therefore, the standard errors obtained can be assumed to be unbiased.

Table 7: Gatekeeping score and how it is decomposed

Rating	Gatekeeping score [Mean (S.D.)]	Average shortest path in network [Mean (S.D.)]	External collaborations [Mean (S.D.)]
A	2.11 (3.15)	10.06 (10.02)	29.4 (25.79)
B	1.23 (1.63)	10.32 (8.32)	18.2 (16.73)
C	0.62 (0.78)	9.12 (7.64)	10.81 (8.87)
P	1.07 (0.66)	10.59 (4.33)	16.04 (6.77)
Y	0.51 (0.72)	9.92 (7.42)	7.86 (7.19)
L	0.34 (0.43)	10.81 (9.36)	5.65 (5.22)
R.U.	0.3 (0.67)	7.32 (7.45)	6.24 (8.56)
N.P.	0.1 (0.23)	5.91 (7.57)	3.52 (3.81)

coefficients mimic the gatekeeping scores seen in Table 7. The order of the average gatekeeping score corresponds to the order of rating within the group of senior (A,B,C) and junior (Y,P,L) researchers.

Model 2, which includes only control variables, shows that the scientific field significantly affects the researchers' gatekeeping score. Because all fields obtain negative coefficients we conclude that researchers of the base category Veterinary Science are most connected both internationally as well as in the national science system. In addition we find that Professors obtain significantly higher gatekeeping scores than PhDs (category Dr. is significantly negative) and white people are more likely to have a higher score than black people. Controlling for these characteristics, the age of the researcher is insignificant.

Model 3 includes both rating and controls, we see again that the rating is a good predictor of gatekeeping, even controlling for many properties of the researcher: higher rated researchers have higher gatekeeping scores. With respect to the control variables, the coefficients of scientific fields and title are largely unaffected when the researcher's quality, indicated by rating, is included. However, racial differences disappear and the age of the researcher becomes negative implying that within one rating category younger researchers are better connected.

A question that naturally arises is whether the strong gatekeeping performance of the highly rated researchers is driven by specificities of their network positions, or

Table 8: Linear regression with autoregressive residuals of rating, publications and covariates on gatekeeping

	Model 1	Model 2	Model 3	Model 4
Intercept	–	1.051 (0.22)***	–	–
A-rated	1.975 (0.158)***	–	3.151 (0.287)***	-0.244 (0.161)
B-rated	1.078 (0.059)***	–	2.183 (0.241)***	-0.213 (0.132)
C-rated	0.553 (0.041)***	–	1.614 (0.227)***	-0.188 (0.123)
P-rated	1.064 (0.381)**	–	1.886 (0.401)***	0.012 (0.214)
Y-rated	0.348 (0.082)***	–	1.407 (0.212)***	-0.162 (0.114)
L-rated	0.244 (0.171)	–	1.447 (0.269)***	-0.072 (0.144)
R.U.	0.274 (0.096)**	–	1.349 (0.239)***	-0.092 (0.128)
N.P.	0.082 (0.147)	–	1.308 (0.257)***	-0.053 (0.137)
Biochemistry	–	-0.526 (0.162)**	-0.517 (0.153)***	-0.063 (0.082)
Cell Biology	–	-0.378 (0.207)	-0.265 (0.195)	-0.11 (0.103)
Chemistry	–	-0.231 (0.123)	-0.182 (0.117)	-0.069 (0.063)
Earth Science	–	-0.461 (0.127)***	-0.505 (0.12)***	-0.121 (0.065)
Engineering	–	-0.887 (0.111)***	-0.846 (0.105)***	-0.313 (0.058)***
Health Science	–	-0.286 (0.108)**	-0.224 (0.102)*	-0.21 (0.056)***
Mathematics	–	-0.986 (0.115)***	-0.978 (0.108)***	-0.323 (0.06)***
Microbiology	–	-0.404 (0.148)**	-0.394 (0.14)**	-0.013 (0.076)
Physics	–	-0.676 (0.124)***	-0.795 (0.118)***	-0.351 (0.064)***
Plant Science	–	-0.113 (0.132)	-0.157 (0.124)	0.037 (0.067)
Dr.	–	-0.478 (0.069)***	-0.289 (0.069)***	-0.012 (0.037)
Coloured	–	0.062 (0.185)	-0.028 (0.176)	0.035 (0.093)
Indian	–	0.198 (0.157)	0.132 (0.148)	0.082 (0.079)
White	–	0.368 (0.111)***	0.191 (0.106)	0.1 (0.057)
Male	–	0.111 (0.072)	0.053 (0.068)	-0.022 (0.036)
Age	–	-0.004 (0.003)	-0.015 (0.003)***	-0.001 (0.002)
Articles	–	–	–	0.066 (0.001)***
$\hat{\rho}^2$	0.015 (0.001)***	0.014 (0.001)***	0.014 (0.001)***	0.024 (0.001)***
$\hat{\sigma}_\epsilon$	0.988 (0)***	0.986 (0)***	0.929 (0)***	0.486 (0)***
N	1315	1315	1315	1315
R^2	0.36	0.36	0.43	0.84
Moran's I p-value	0.48	0.45	0.43	0.43

Note: *, ** and *** signify 5%, 1% and 0.1% rejection levels of significance respectively. Baseline categories are Veterinary Science (scientific field), Prof (title), Black (race) and Female (gender).

simply be their being more productive. Model 4 addresses this question by adding the number of papers published over the 7 year observation period (variable Articles) as another control. What we observe is first, that the Articles variable has a highly significant positive coefficient, and that the adjusted R^2 increases from 0.43 to 0.84, indicating that productivity matters. The second observation is that with the addition of publications as an explanatory variable, the coefficients on ratings become very small (relative to Model 3), negative, and insignificant. This strongly suggests that what drives the difference in gatekeeping activity among researchers of different ratings is not that they have peculiar positions in the science network, but rather that they are more productive.¹²

These results are consistent with the permutation tests presented in Table 9.¹³ Table 8 shows that in Models 1 and 3 we should reject the hypothesis that ratings are irrelevant to gatekeeping. However, in Model 4, when we control for productivity, we see that rating has no significant impact on gatekeeping.

The explanation for the role of productivity can be seen by examining Table 7. There it is clear that differences in gatekeeping by ranking are driven by external weight rather than by internal connectedness. It seems reasonable to assume that a scientist with more publications will have more external co-authors, and indeed the correlation between number of publications and the (weighted) number of external co-authors is 0.94. This intuition is confirmed by regression analysis, using the two components separately as dependent variables, and the same controls as in Table 8. The results for external weight mimic those for gatekeeping: rating has explanatory power, but when productivity is added as an explanatory variable, coefficients on

¹²Further regressions, not displayed in this paper, including the position of researchers in the South African co-authorship network in terms of degree centrality, shortest path betweenness centrality, clustering, average nearest neighbour connectivity, and/or main component membership fully confirm this argument.

¹³This table is read in the following way, row by row. Row A-B, for example presents Wald statistics from the initial estimates of Equation 1a and the p-values for those statistics as derived from the permutation of A and B researchers. That is, ratings A and B were re-assigned randomly among the population of A and B researchers (keeping numbers of As and Bs constant). Equation 1a is re-estimated and the Wald statistic for A versus B is calculated. This is repeated 1000 times to generate a distribution of Wald statistics under the hypothesis that there is no difference in gatekeeping between A and B researchers. From this distribution we can generate a p-value for the Wald statistic of the real data.

Table 9: Permutation tests of the hypothesis of equal gatekeeping scores for pairs of rating categories (Wald-statistic (p-value))

	Model 1	Model 3	Model 4
A-B	24.75 (0.007)	2.85 (0.001)	-2.11 (0.951)
A-C	44.73 (<0.001)	5.1 (<0.001)	-2.24 (0.967)
B-C	92.52 (<0.001)	2.39 (<0.001)	0.02 (0.469)
P-Y	2.76 (0.055)	1.01 (0.070)	-0.09 (0.532)
P-L	3.34 (0.003)	1.07 (0.026)	-0.86 (0.773)
Y-L	4.52 (0.017)	0.26 (0.345)	-1.21 (0.873)

Note: The permutation test is based on OLS coefficient estimates which are slightly but not significantly different from the estimates in regression table 8. The sign of the Wald-statistic might therefore differ from the sign implied by the regression table.

rankings are significantly different neither from each other nor from zero. By contrast, ranking does not “explain” average path lengths. Here coefficients on rank are not significant and are not different from each other. Adding the productivity measure as an explanatory variable has no change when average path length (or its inverse) is the dependent variable. We can safely conclude that differences in gatekeeping scores are associated with differences in researcher rating, and are largely determined by the number of articles produced rather than with the patterns of research collaboration.

6. Sensitivity analysis

A sensitivity analysis shows that our main results are robust to the data treatment, i.e. handling of outliers, or our definition of the gatekeeping score. Detailed results of the sensitivity analysis may be obtained from the authors upon request. This section provides a qualitative assessment.

Because results from permutation of raw data might be especially sensitive to outliers (Anderson and Legendre, 1999), the analysis has been repeated with outliers removed from the sample. As outliers, we identified six researchers with exceptionally high gatekeeping scores compared to other researchers of same rating.¹⁴ In addition two researchers with age greater than 100 years have been removed from the sample.

¹⁴These are two A, two B, one Y, and one unsuccessful researcher.

Removing the outliers changes neither the ranking of rating categories by gatekeeping nor the overall testing results. Removing the two A researchers with the highest gatekeeping score affects the results most because there are relatively few A researchers. The average gatekeeping score of A researchers drops from approximately 2.1 to 1.5. Consequently, the permutation test rejects that A and B researchers have the same expected gatekeeping score only at a 5% significance level which is higher than the 0.7% level presented in the result section. Removing outliers had no further substantial effects.

We chose the gatekeeping measure for the main analysis because it combines in a simple fashion the two crucial aspects we are concerned with, namely, access to external knowledge and diffusion within the research system. However, our measure of gatekeeping captures knowledge diffusion within the research system by the shortest paths between agents.

Similar to betweenness centrality based on shortest paths, this is vulnerable to the critique that knowledge might diffuse along any path and not only the shortest. Newman (2005) has proposed an alternative to betweenness, which he bases on random walks. A random walk on a network starts at a source node and moves to further nodes by following links uniformly at random until the sink node is reached. Random walk betweenness centrality ascribes a high centrality score to nodes which are on many random walks between any two other agents and is calculated exactly as the electricity flowing through nodes of an electrical grid. Most of the current flows through nodes on the shortest path, the path with the lowest resistance. Nodes on longer paths still carry current but less so as the resistance increases with the length of the paths. In order to assess the robustness of our findings with respect to the dependent variable gatekeeping, we can adapt random walk betweenness centrality, simply by considering not a random walk between all pairs of nodes, but rather random walks in which all internal nodes serve as targets but in which the sole source is the external node. This describes the distribution of knowledge created outside the system (the source) to all nodes within it.

Regression and permutation test results for established researchers (A, B, and C researchers) do not change with the alternative dependent variable; especially significance levels of the permutation tests all remain below the 0.1% level. Again,

higher ranked researchers are better at distributing externally generated knowledge, but this is driven largely by the fact that they produce more articles. However, differences among young researchers (P, Y, and L researchers) become insignificant when using random walk betweenness instead of gatekeeping. The reason is that young researchers differ mainly in their number of external collaborations which has a large effect on their gatekeeping score but less so on random walk betweenness.

7. Conclusion

The paper provides evidence that in a country that is behind the technological frontier, world-class researchers act as important conduits of knowledge into the local academic research community. The higher the competence of researchers, the better connected they are beyond the local research community in a global network.

The central difference between the higher and lower rated researchers is the fact that higher-rated researchers are more prolific producers of articles than their counterparts. The NRF rating process is designed to avoid the use of productivity measures (for example, citation counts) as a proxy for quality, but still our econometric analyses suggest that the higher productivity of world-class researchers is the core driver in their being nodes connecting global and local knowledge creation processes. Because they publish more articles overall, there are more opportunities for them to collaborate with colleagues both locally and abroad.

The NRF-ratings are derived not from the status of the journals in which researchers publish, but are based on the reviews of a specialist committee that is given copies of all their publications.¹⁵ This makes the NRF-rating a more reliable indicator of research quality than, for example, the use of impact factors of journals. Based on the NRF-rating, we find a co-occurrence of high quality and high quantity which suggests that there is no trade-off between quality and quantity in research.

The literature highlights three areas where it is unclear whether the presence

¹⁵This removes the significant problem in using, for example, impact factors of journals to estimate quality of individual researchers. It is well known that paper quality is highly skewed within any journal, and so using the mean ‘quality’ of a journal to estimate the quality of a paper, which impact factors implicitly do, is likely to be a poor measure of the quality of an individual researcher’s output. See Wall, 2009.

of world-class researchers in a less developed context would be beneficial for the local research community: Potentially disconnected local versus global knowledge worlds, the fact that research resources in less developed countries are especially scarce with the ramifications of potentially “unfair” allocation decisions, and finally the practical demands — “transaction costs” (Lee and Bozeman, 2005) — of maintaining collaborative relationships. In all three areas, there is an assumption that some form of trade-off is necessary, and that when world-class researchers opt for global connectedness, local connectedness is sacrificed.

Yet the evidence suggests a very different logic: World-leading researchers in a developing country are particularly prolific, which minimizes their need to make trade-offs. Previous work has demonstrated that highly skilled researchers are notably competent at managing collaborations (Zuckerman, 1967), and this paper suggests that this skill may have a particular manifestation for world-leading researchers in developing countries. By expanding the number of projects they work on, higher rated researchers also expand their pool of collaborators. The evidence suggests they are able not only to manage relationships with collaborators outside of the country, but also to build and maintain local relationships at a level similar to that of their local colleagues.

The opportunity to work with a world-leading researcher represents a resource in and of itself to local colleagues. Because there is adequate evidence of the world-leading colleague collaborating locally, local colleagues are likely to be generally supportive when there is some form of preferential allocation of scarce resources to A-rated researchers. In addition, the productivity of the highest ranked researchers makes the allocation of such scarce resources far less contested and less likely to elicit concerns about fairness.

The analyses conducted in this paper do not allow us to comment on how the A and B-rated researchers manage the challenge of potentially disconnected global and local knowledge worlds, other than to comment that they are clearly able to interest a substantial number of colleagues in their work. It seems likely that they are able to integrate those worlds (to the extent that the two worlds are disconnected) — the work of a recognised researcher is likely to fall within a general area of interest and not be divisible into ‘local’ versus ‘global’ projects. However, further research is

needed to establish how the scientists manage the tensions imposed by operating in two sometimes quite different knowledge worlds.

Indeed, much work remains to be done on how world-leading researchers contribute to knowledge in the local knowledge network in the developing country. Internal linkages can occur through (essentially hierarchical) supervisory relationships, through (more equitable) collegial collaboration or some combination of the two. Whether a researcher works mainly with (ex-)students or collaborates more generally with colleagues is likely to have different effects on the nature of knowledge flows in the academic science research community overall, and it would be useful to track with whom collaborations take place.

Finally, the evidence from this paper is based on research in science and technology. Advances in science and technology are particularly important for economic advancement, but to understand the extent and nature of the local/global tension in knowledge creation, it is important to also consider social sciences and the humanities. The role of context is likely to be more pronounced in the social than in the natural sciences, and additional work is needed to tease out how frontier knowledge in the social sciences and humanities is created and disseminated in a local network.

Even though additional work is needed, the evidence from this paper is robust enough to conclude that it is useful for the science and technology network of a developing country to have world-leading researchers. In addition, it seems that one of the key abilities (and arguably even tasks) of such world-leading researchers — much more than in the developed world — is to act as a connector between different knowledge worlds and different communities of researchers.

They are able to do so because not only their intellectual acuity, but also their research productivity exceeds that of their colleagues. Because they are more productive, many of the trade-offs presumed to govern their local/global interactions are less severe. However, it is important to note that trade-offs do not disappear, but are simply mitigated. Understanding the mechanisms that allow researchers to increase their productivity therefore seems particularly important.

There is a very limited population of world-leading or A-rated researchers — only 44 in the period under study. By better understanding the mechanisms that allow for the greater strength of collaboration (more articles), and by better understanding

the structure of collaborations (e.g. whether collaborators are local or global, and students or peers), it may be possible to better use the larger population (312) of ‘internationally acclaimed’ or B-rated researchers to connect local and global knowledge networks.

This also requires a better understanding of collaborators’ past experience on future work. Although the paper claims that collaborations allow for knowledge diffusion, the diffusion is not investigated. For example, it will be important to establish under which conditions authors produce more articles or better articles after having collaborated with highly experienced world-leading researchers to establish whether the internal, local environment of the world-leading researchers does catch up.

A number of important and intriguing questions remain about the nature of interaction that world-leading researchers in a developing country have with their peers locally and abroad. However, this paper provides conclusive evidence that world-leading research done in a developing country has a direct benefit for other researchers in their home country. In a country with the level of research capacity of South Africa — just slightly below the mean of the world — concerns about the emergence of enclaves or parallel knowledge systems seem unnecessary.

References

- Akbar, Y. H., McBride, J. B., 2004. Multinational enterprise strategy, foreign direct investment and economic development: the case of the Hungarian banking industry. *Journal of World Business* 39, 89–105.
- Allison, P. D., Long, J. S., 1990. Departmental effects on scientific productivity. *American Sociological Review* 55 (4), 469–478.
- Anderson, M., Legendre, P., 1999. An empirical comparison of permutation methods for tests of partial regression coefficients. *Journal of Statistical Computation & Simulation* 62, 271–303.
- Anselin, L., 1988. *Spatial Econometrics: Methods and Models*. Kluwer Academic Publishers, Dordrecht.

- Arbia, G., 2006. *Spatial Econometrics*. Springer-Verlag, Berlin, Heidelberg.
- Barnard, H., Bromfield, T., Cantwell, J., 2009. The role of indigenous firms in innovation systems in developing countries: The developmental implications of national champion firms' response to underdeveloped national innovation systems. In: Lundvall, B.-A., Joseph, K., Vang, J., Chaminade, C. (Eds.), *Handbook of Innovation Systems and Developing Countries*. Edward Elgar, Cheltenham: UK, pp. 249–279.
- Bell, M., Pavitt, K., 1997. Technological accumulation and industrial growth: contrasts between developed and developing countries. In: Archibugi, D., Michie, J. (Eds.), *Technology, Globalisation and Economic Performance*. Cambridge University Press, Cambridge and New York, pp. 83–137.
- Bernardes, A. T., da Motta e Albuquerque, E., 2003. Cross-over, thresholds, and interactions between science and technology: lessons for less-developed countries. *Research Policy* 32, 865–885.
- Bowen, J. T., Leinbach, T. R., Mabazza, D., 2002. Air cargo services, the state and industrialization strategies in the Philippines: The redevelopment of Subic Bay. *Regional Studies* 36 (5), 451–467.
- Cameron, A., Trivedi, P., 2005. *Microeconometrics: Methods and Applications*. Cambridge Univ. Press., New York.
- Cohen, W. M., Levinthal, D. A., 1989. Innovation and learning: The two faces of R&D. *The Economic Journal* 99, 569–596.
- DuBois, F. L., Reeb, D., 2000. Ranking the international business journals. *Journal of International Business Studies* 31 (4), 689–704.
- Duque, R. B., Ynalvez, M., Sooryamoorthy, R., Mbatia, P., Dzorgbo, D.-B. S., Shrum, W., 2005. Collaboration paradox: Scientific productivity, the internet, and problems of research in developing areas. *Social Studies of Science* 35 (5), 755–785.
- Feinberg, S. E., Majumdar, S. K., 2001. Technology spillovers from foreign direct investment in the Indian pharmaceutical industry. *Journal of International Business Studies* 32 (3), 421–437.

- Fleming, L., 2001. Recombinant uncertainty in technological search. *Management Science* 47 (1), 117–132.
- Goode, R. 1959. Adding to the Stock of Physical and Human Capital. *American Economic Review* 49 (2), 147–155.
- Hall, P., 1992. *The Bootstrap and Edgeworth Expansion*. Springer-Verlag, New York.
- Haslam, N., Laham, S. M., 2010. Quality, quantity, and impact in academic publication. *European Journal of Social Psychology* 40, 216–220.
- Jaffe, A. B., 1989. Real effects of academic research. *The American Economic Review* 79 (5), 957–970.
- Katila, R., Ahuja, G., 2002. Something old, something new: A longitudinal study of search behavior and new product introduction. *Academy of Management Journal* 45 (6), 1183–1194.
- Kim, L., Nelson, R. R., 2000. *Technology, Learning, & Innovation*. Cambridge University Press, Cambridge.
- Lall, S., 2001. *Competitiveness, Technology and Skills*. Edward Elgar, Cheltenham, UK.
- Lee, S., Bozeman, B., 2005. The impact of research collaboration on scientific productivity. *Social Studies of Science* 35 (5), 673–702.
- Leung, K., 2007. The glory and tyranny of citation impact: An East Asian perspective. *Academy of Management Journal* 50 (3), 510–513.
- Leydesdorff, L., Amsterdamska, O., 1990. Dimensions of citation analysis. *Science, Technology, & Human Values* 15 (3), 305–335.
- Lundvall, B.-A., Borrás, S., 2005. Science, technology, and innovation policy. In: Fagerberg, J., Mowery, D. C., Nelson, R. (Eds.), *The Oxford Handbook of Innovation*. Oxford University Press, Oxford and New York, pp. 599–631.

- Marin, A., Bell, M., 2006. Technology spillovers from foreign direct investment (FDI): the active role of MNC subsidiaries in Argentina in the 1990s. *Journal of Development Studies* 42 (4), 678–697.
- Miller, D. J., Fern, M. J., Cardinal, L. B., 2007. The use of knowledge for technological innovation within diversified firms. *Academy of Management Journal* 50 (2), 308–347.
- Narula, R., Dunning, J., 2000. Industrial development, globalization and multinational enterprises: New realities for developing countries. *Oxford Development Studies* 28 (2), 141–167.
- Nelson, R., 2005. The roles of research in universities and public labs in economic catch-up. In: Santangelo, G. D. (Ed.), *Technological Change and Economic Catch-Up*. Edward Elgar, Cheltenham, UK, pp. 19–32.
- Newman, M., 2005. A measure of betweenness centrality based on random walks. *Social Networks* 27 (1), 39–54.
- Pack, H., 2000. Research and development in the industrial development process. In: Kim, L., Nelson, R. R. (Eds.), *Technology, Learning, and Innovation*. Cambridge University Press, Cambridge, pp. 69–94.
- Pfeffer, J., Langton, N., 1993. The effect of wage dispersion on satisfaction, productivity, and working collaboratively: Evidence from college and university faculty. *Administrative Science Quarterly* 38 (3), 382–407.
- Phene, A., Fladmoe-Lindquist, K., March, L., 2006. Breakthrough innovations in the U.S. biotechnology industry: The effects of technological space and geographic origin. *Strategic Management Journal* 27 (4), 369–388.
- Rosenkopf, L., Nerkar, A., 2001. Beyond local search: Boundary-spanning, exploration, and impact in the optical disk industry. *Strategic Management Journal* 22 (4), 287–306.
- Sheskin, D., 2007. *Handbook of Parametric and Nonparametric Statistical Procedures*. Taylor and Francis, London.

Sooryamoorthy, R., 2009. Collaboration and publication: How collaborative are scientists in South Africa? *Scientometrics* 80 (2), 419–439.

Starbuck, W., 2005. How much better are the most-prestigious journals? The statistics of academic publication. *Organization Science* 16, 180–200.

Zuckerman, H., 1967. Nobel laureates in science: Patterns of productivity, collaboration, and authorship. *American Sociological Review* 32 (3), 391–403.