Science in South Africa: have Black Africans, Coloureds and Indians prospered since the ending of *apartheid*?

Grant Lewison[1] and Daisy Jacobs[2]

[1] *grantlewison@aol.co.uk* and *g.lewison@ucl.ac.uk*
[1] Department of Information Studies, University College London, Gower Street, London WC1E 6BT, UK

[2] *daisymjacobs@yahoo.com*
[2] Department of Library and Information Science, University of Zululand, Private Bag X1001, KwaDlangezwa 3886, South Africa

Abstract

Since the ending of *apartheid* in 1994, the South African government has made big efforts to involve non-white South Africans in many activities formerly denied to them, or severely restricted. We wondered how much success it had had in increasing the participation of blacks, coloureds and Indians in research, and in which fields they were working. We did this through an analysis of the surnames of South African scientists, characterising them as English, Afrikaans or other White on the one hand, or Black African, Coloured or Indian. We looked at South African papers in the Web of Science for 1988-89, 1998-99 and 2008-09, and found that the non-white fractional presence (as a proportion of the total with classified names) rose from 10% to 14% and then jumped to 23% in 2008-09, with black Africans accounting for more than half of the non-white total. Non-whites were most prominent in chemistry, physics and maths (>22%) and least in earth & space, and biology (~12%). They collaborated somewhat less than white scientists with Europe, but much more with the USA, and Africa – especially the black South Africans. South Africans also now publish less in national and more in international journals.

Introduction

Science in South Africa during and after apartheid

Research and publication of results by scientists and researchers are of vital importance in any given geographical region including the African continent which is challenged with so many development needs. So urgent are Africa's needs that the effective participation in research and dissemination of this knowledge is to a certain extent a matter of life and death. Most of the research in science in South Africa was previously carried out by senior white colleagues and in historically developed universities. With the advent of democratisation and freedom for all in 1994, the new government provided increased support for scientific research and created new funding organisations such as the National Research Foundation (NRF) so as to encourage South Africans from all higher education and research institutions.

South Africa has a population only one third that of Nigeria, the most populous country in Africa, but a GDP two thirds higher (figures for 2007). It has by far the largest scientific output in Africa. This tends to correlate with national wealth, as measured by GDP, see Figure 1, where the spot for South Africa (ZA) lies just above the trend-line.

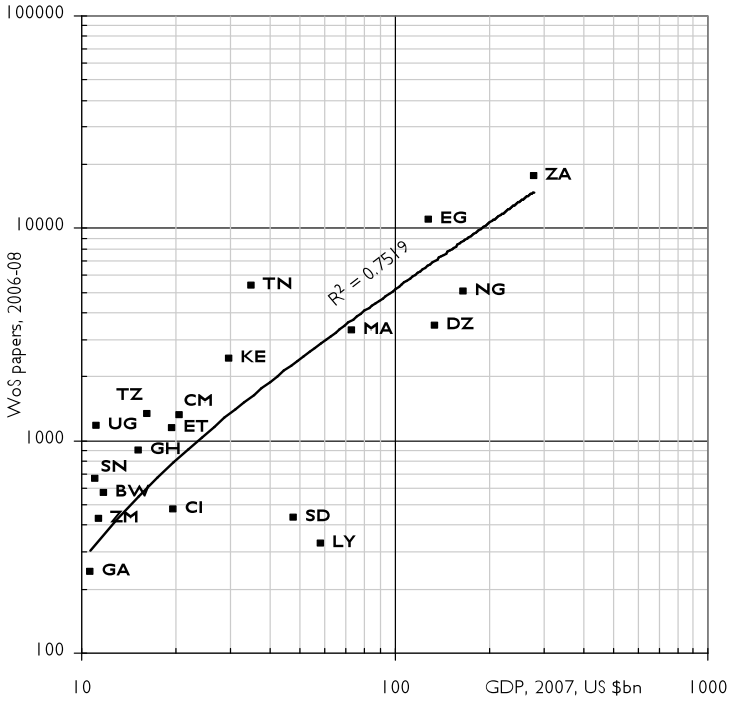

Figure 1. Output of papers (articles, proceedings papers and reviews) from 19 African countries in the SCI-E and SSCI, 2006-08, as a function of GDP in 2007 (US $bn).

BW=Botswana, CI=Cote d'Ivoire, CM=Cameroon, DZ=Algeria, EG=Egypt, ET=Ethiopia, GA=Gabon, GH=Ghana, KE=Kenya, LY=Libya, MA=Morocco, NG=Nigeria, SD=Sudan, SN=Senegal, TN=Tunisia, TZ=Tanzania, UG=Uganda, ZA=South Africa, ZM=Zambia

As a country, South Africa underwent a major change in 1994 when democratic elections with universal suffrage led to the election of a government led by the African National Congress, and the swift repeal of the laws enforcing the system of *apartheid*, or separate development. This had provided for very unequal treatment for the different racial groups making up South Africa. There are four main such groups: black Africans (formally described as Africans, but hereinafter as Blacks) comprising 39.7 million in 2010 out of a total of 50 million; Whites (Afrikaners, English-speakers, and others, many of them immigrants from southern and eastern Europe) comprising 4.6 million; Coloureds comprising 4.4 million; and Indians (many descended from workers brought to Natal in the 19th century to work on sugar plantations), comprising 1.3 million.

South Africa has a well-developed university system, although under *apartheid* there were only seven universities open to non-whites. The two oldest are the University of Fort Hare, founded in 1916, which was historically black, and the University of the Western Cape, founded in 1959 as a college for Coloured students. There were also five others: the University of Transkei – now renamed Walter Sisulu University – the University of Venda, the University of the North West, the University of Durban Westville – for Indians – and the University of Zululand. Currently, the various Higher Education Institutions including the former Technikons have been grouped into 21 universities, listed in Table 1 with their outputs of papers in 2008-09 (integer counts). They vary greatly in scientific size, as can be seen. Trigraph codes have been appended to their names for use in a later figure.

Table 1. List of 21 South African universities with their output of WoS papers in 2008-09.

University name	*N*	*University name*	*N*
Univ Cape Town - UCT	2229	Univ Ft Hare - UFH	201
Stellenbosch Univ - UST	1620	Univ S Africa - USA	189
Univ Pretoria - UPT	1546	Tshwane Univ Technol - TSH	167
Univ Witwatersrand - UWW	1542	Univ Limpopo - ULP	149
Univ Kwazulu Natal - UKZ	1478	Durban Univ Technol* - DUT	136
Rhodes Univ - RHU	498	Cape Peninsula Univ Tech - CPU	120
Univ Free State - UFS	411	Univ Venda - UVD	72
North West Univ - NWU	404	Walter Sisulu Univ - WSU	43
Univ Western Cape - UWC	355	Vaal Univ Technol - VUT	20
Univ Johannesburg - UJB	348	Cent Univ Technol - CUT	17
Nelson Mandela Metr. Univ - NMM	209		

* includes Univ Zululand

Besides the universities, there are two major hospitals that have published many papers: Groote Schuur in Cape Town (affiliated to the University of Cape Town), and Tygerberg in the city of the same name in Western Cape province (affiliated to Stellenbosch University). The other major contributor to South African science has been government labs, run by the Agricultural Research Council (notably the Veterinary Institute at Onderstepoort), the Council for Scientific and Industrial Research (CSIR) and the Medical Research Council. The latter has a number of free-standing units, and also ones within some of the leading universities.

There is quite an extensive literature about the South African science system and its outputs. Many of the earlier papers write pessimistically about the relative decline in output and in citation performance (Jacobs & Ingwersen, 2000; Pouris, 2003; Ingwersen & Jacobs, 2004), the low rate of inter-institutional collaboration (Mouton, 2000) and the paucity of R&D expenditure (Blankley & Kahn, 2005). This mood changed during the last few years, with more optimistic assessments of South Africa's output by Jeenah & Pouris (2008), and of its growing tendency to collaborate internationally (Sooryamoorthy, 2009; Sooryamoorthy, 2010). However we are concerned mainly with the ethnic composition of its scientists.

The use of names to infer ethnicity

Surnames are often very revealing of their bearers' origins and ethnicity, although of course some are ambiguous, such as Lee, which is common in China, England and Korea. But their analysis can be used for different purposes, such as investigations into consanguineous marriages and their possible deleterious consequences (Crow, 1980; Bittles *et al*., 1993; Biondi *et al*., 1993); the identification of individuals in particular ethnic groups for research purposes (Tjam, 2001); for checking to see if medical care is being dispensed equally to all racial groups (Lagerberg et al., 2005) or how many immigrants are living in an area (Nicoll *et al*., 1986; Piazza *et al*., 1987; Chakraborty *et al.*, 1989; Startseva *et al*., 1994; Biondi *et al*., 2001). They can equally be used for the analysis of scientific output and the contributions of different ethnic groups (Lewison & Igic, 1999; Webster, 2004) and patterns of migration (Lewison & Kundra, 2008). These investigations depend on the assignment of names to different groups, which is inevitably difficult and somewhat subjective, although it can be assisted by comparison between names in the target group and names on selected lists. This was the method that was adopted here.

Methodology

Selection of papers and grouping of names

We wished to select papers from the period before the end of *apartheid* in 1994, from the period shortly afterwards, and recently, so as to see whether the ethnic composition of the South African research community had changed as a result of new policies. In order to obtain enough papers for analysis, we selected articles, notes and reviews from the Web of Science (WoS), both the Science Citation Index Extended and the Social Sciences Citation Index, for three two-year periods: 1988-89, 1998-99 and 2008-09 with an address in South Africa; there were 25,047 papers in total. The full bibliographic details of the papers were downloaded to sets of files, 500 at a time, and these files were then opened successively and the paper details combined in a single Excel file. The authors' names were formatted in the style used on the CD-ROMs (*viz*. Smith-RA) and the addresses were similarly treated, with all the addresses in the WoS combined and duplicates (if any) removed. The addresses of the journal publishers (coded as "PA" in the WoS) were also added to the file and the country of publication listed separately.

The addresses on the papers were parsed to provide the fractional count of each country, including that of South Africa, and the authors of all papers with no international collaboration were then listed in a single column (N = 15,951). Their initials were removed, leaving only the surnames, and they were listed in descending order of frequency of appearance – the leaders were van Wyk and van der Merwe, with Smith in fifth position. There were 41,543 authors with 9409 different names. We decided to group them into eight categories, shown in Table 2 with their unigraph codes used in the subsequent tables.

For some of the groups, reference was made to existing lists of names, such as surnames in the 1881 UK census [ones (n = 3000) with at least 1500 individuals were used], or to names used by scientists from India (for the Indian names) or from Malaysia (for the Coloured names) in order to categorise them. But the large majority had to be assigned individually to groups on the basis of the authors' experience. Some names defied categorisation but they were a small minority, accounting for only 2125 authors (5%).

Table 2. List of eight ethnic groups and codes used for the analysis of South African names.

Code	*Group*	*Notes*
B	Black	Includes other African names
C	Coloured	Includes Muslim and Arabic names
E	English	Includes Welsh, Scottish and Irish names
H	Han	Includes Chinese and other East Asian names
I	Indian	
K	Afrikaans	Includes Dutch names
W	White (other)	Includes West and East European, Greek, Russian names
U	Unknown	

Assignment to major fields and other analyses

The assignment of papers to fields was performed with the thesaurus provided by CHI Research Inc., which was subsequently updated to reflect additional journals included in the WoS. These were assigned to fields on the basis of their names, or the titles of papers that they contained. The advantage of this system is that each journal is assigned to only one field, although some assignments could be disputed. Twelve major fields were used, listed in Table 3 with the trigraph codes used subsequently in the tables.

Table 3. List of major fields used for the analysis, with short codes.

Field	*Code*	*Field*	*Code*
Biology	BIO	Health Sciences	HSC
Biomedical Research	BMR	Mathematics	MAT
Chemistry	CHE	Physics	PHY
Clinical Medicine	CLM	Professional Fields	PRO
Earth and Space	EAS	Psychology	PSY
Engineering and Technology	ENG	Social Sciences	SOC

As mentioned above, a geographical analysis was undertaken to give the fractional counts of each country on each paper so that co-authorship patterns could be established. The country of publication of each paper was also recorded: most were in just four countries, the USA, the UK, South Africa and the Netherlands. Finally, the document type was noted in order to show if non-white authors from South Africa were being invited to write reviews – an indication that they were regarded as senior researchers (Lewison, 2009).

Results

Outputs of papers overall and by the different groups

South African scientific output remained almost constant from the mid-1980s to 2000 (Pouris, 2003) but has since increased sharply, see Figure 2. So evidently the decade after the end of *apartheid* was one of scientific stagnation, but since the early years of this century, matters have much improved – this is reflected in the literature, as mentioned above.

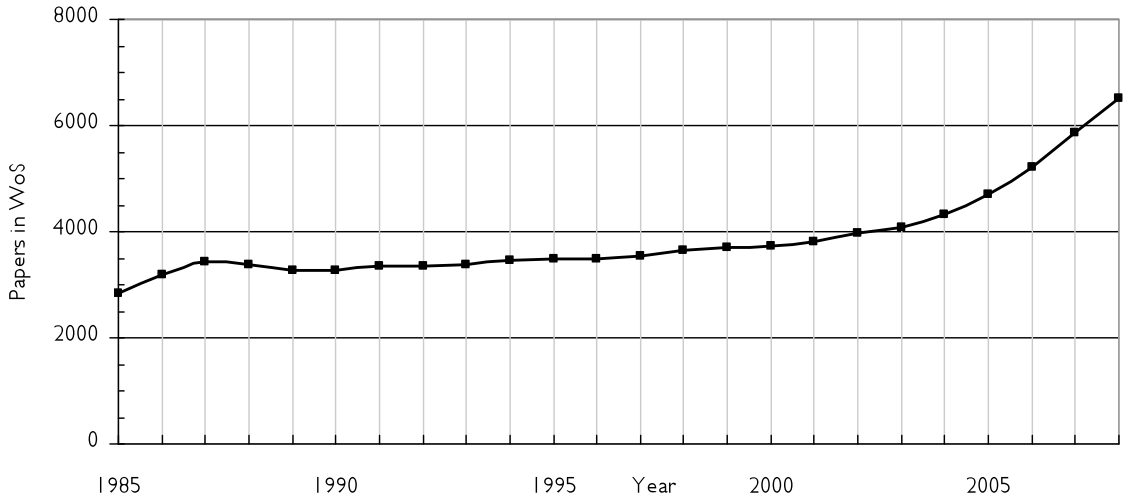

Figure 2. Output of papers (articles, notes, proceedings papers and reviews) from South Africa in the SCI-E and SSCI, 1985-2008 (three-year running means).

The racial composition of South African scientific output in the three two-year periods is shown in Figure 3.

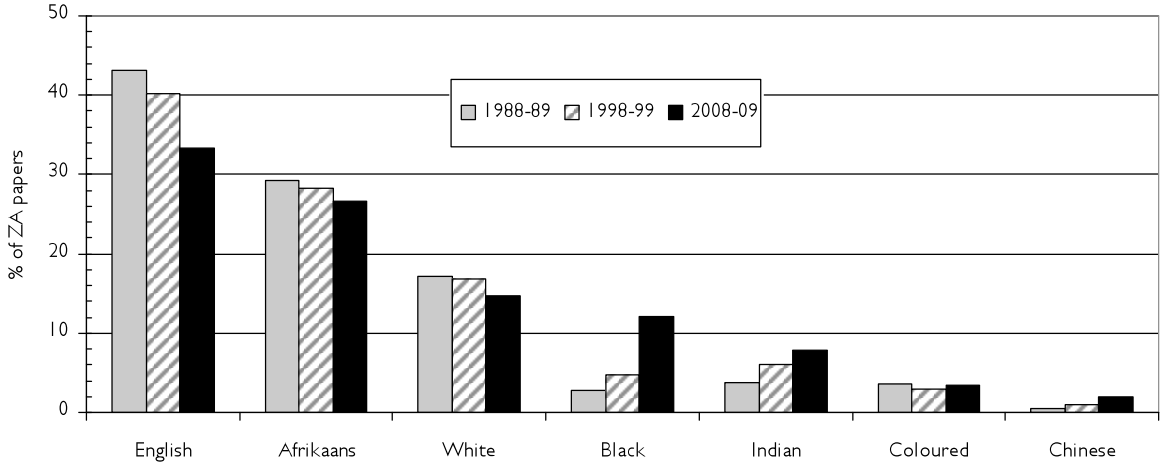

Figure 3. Racial composition (fractional counts) of authors of South African papers in the WoS in three two-year periods (authors of unknown origin excluded).

This figure shows, as expected, that the presence of the three white groups (English, Afrikaans and others) has steadily declined and that of the blacks and Indians has increased. The Coloured representation has remained almost static in percentage terms, although in 2008-09 with a larger total output, their number of papers increased from 164 in 1998-99 to 284. In fact, because of the rapid expansion in overall output since 2000 seen in Figure 1, the actual numbers of papers from the three white groups have also expanded.

Racial preferences for the major fields

Table 4 shows the preferences of the white and non-white groups for different major fields of science over the whole period.

Table 4. Major fields and the ethnic (white/non-white) representation within each of them for South African scientific output; 1988-89, 19998-99 and 2008-09. For codes, see Table 1.

Field:	*Total*	*EKW*	*BCI*	*BCI, %*		*Field:*	*Total*	*EKW*	*BCI*	*BCI, %*
CHE	1832	1144	352	23.5		*CLM*	6660	4315	1020	19.1
PHY	1233	618	185	23.0		*BMR*	2999	1965	386	16.4
MAT	855	498	144	22.4		*SOC*	747	498	80	13.8
HSC	298	153	42	21.5		*PRO*	572	405	58	12.5
PSY	468	300	81	21.3		*EAS*	2399	1502	213	12.4
ENG	1558	995	243	19.6		*BIO*	5411	3912	507	11.5

So the non-white groups preferred the "traditional" fields of chemistry, physics and mathematics, whereas the white groups remained strong in biology, earth & space, professional fields and social sciences.

Countries of publication and international co-authorship

Figure 4 shows the preferences of South African researchers for journals published in different countries. There is a progressive decline in South African journals, and this may account in part for the lack of growth in its scientific output in the 1990s as some of them were removed from the SCI.

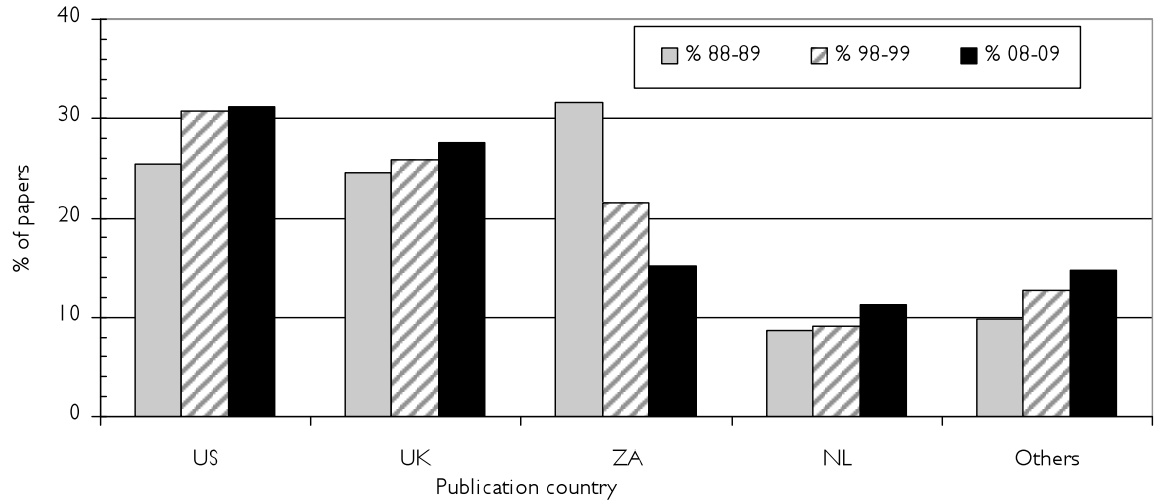

Figure 4. Publication country of journals in which South African scientists published their papers, 1988-89, 1998-99 and 2008-09.

The non-whites preferred to publish in international journals (18% overall) whereas the whites used national journals rather more (22%). This may reflect their relatively greater involvement in biology (much of which was concerned with the conservation of South African wildlife) and earth sciences (mostly involving mining and geology within the country).

Figure 5 shows the amount of international publication by the two main groups of non-white and white authors. Both are expanding their international presence, but the whites are expanding faster, probably because they have greater economic possibilities to travel abroad to find appropriate partners.

Within the non-white authorship, there were noticeable differences in their willingness to collaborate with different countries, as shown in Table 5.

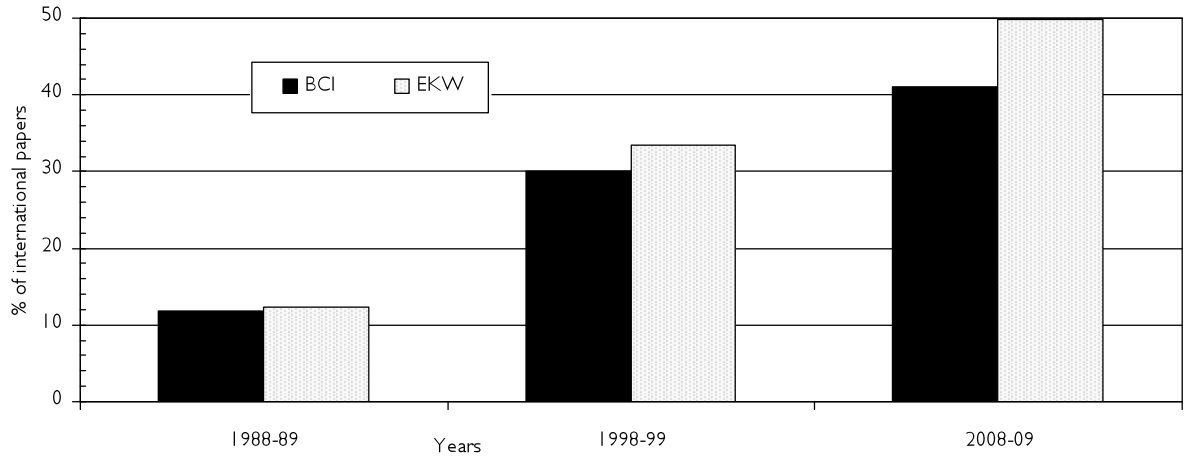

Figure 5. Extent of international co-authorship in scientific publications by non-white authors (BCI) and white authors (EKW) in South Africa; integer counts.

Table 5. Percentage of co-authorship of the three non-white groups of South African authors with 10 European countries, the USA and three African countries (Kenya, Nigeria and Zimbabwe), 1988-89, 1998-99 and 2008-09 combined.

Group	*Int'l*	*EUR10*	*USA*	*KE, NG, ZW*	*% EUR*	*% US*	*% KE, NG, ZW*
Black	272	89	138	66	33	51	24.3
Coloured	135	46	91	16	34	67	11.7
Indian	290	106	164	17	36	56	6.0
All	5255	2065	1277	169	39	24	3.2

Collaboration with the Europeans is slightly less than the average for South Africa as a whole, but it is much greater with the USA – especially for Coloured authors – and again much greater with other African countries – and the black South Africans are the most involved with them and the Indians the least.

Percentage of reviews – a measure of esteem

Because reviews tend to be commissioned from, or submitted by, relatively senior researchers who have an established reputation, they provide a simple indicator of the esteem in which a country's (or an institution's) senior scientists are held by journal editors (Lewison, 2009). Table 6 shows that, although the percentage of reviews written by both the white authors and the non-whites have increased, the former have increased faster. This may simply reflect the greater willingness in recent years for journal editors to work with South Africans, and that there has not been enough time since *apartheid* ended for non-white authors to gain a similarly high scientific profile. But it is a sensitive indicator and one that is easy to check.

Table 6. Numbers and percentages of reviews written by non-white (BCI) and white (EKW) South African authors in three two-year periods (integer counts).

Period:	*BCI all*	*EKW all*	*BCI rev*	*EKW rev*	*BCI % rev*	*EKW % rev*	*Ratio*
1988-89	1334	6223	22	111	1.6	1.8	0.92
1998-99	1621	6088	44	190	2.7	3.1	0.87
2008-09	4348	9647	263	696	6.0	7.2	0.84

The changing role of universities

The ethnic composition of the authors of papers from each of the 21 universities listed in Table 1, and their predecessors (some of which were called Technikons), at the three dates was determined as the percentage of Blacks, Coloureds and Indians in the total, excluding names of authors in the Unknown group. As would be expected, this percentage has risen steadily for almost all universities, but the rate of progress, and the final situation, have varied greatly, see Figure 6. [The universities have been ordered by the mean percentage of non-white authors.] The University of Fort Hare and Walter Sisulu University rank highest, although the latter has recently been overtaken by the Durban University of Technology, including the University of Zululand. The overall average for all the universities (this will include some double-counting) has risen from 10.6% in 1988-89, to 14.3% in 1998-99 and 24% in 2008-09. These figures are very close to the average for South Africa as a whole, showing the universities were not leading the way in the promotion of non-white researchers, as might have been supposed, and the non-white researchers are not availing themselves of the opportunities now open to them.

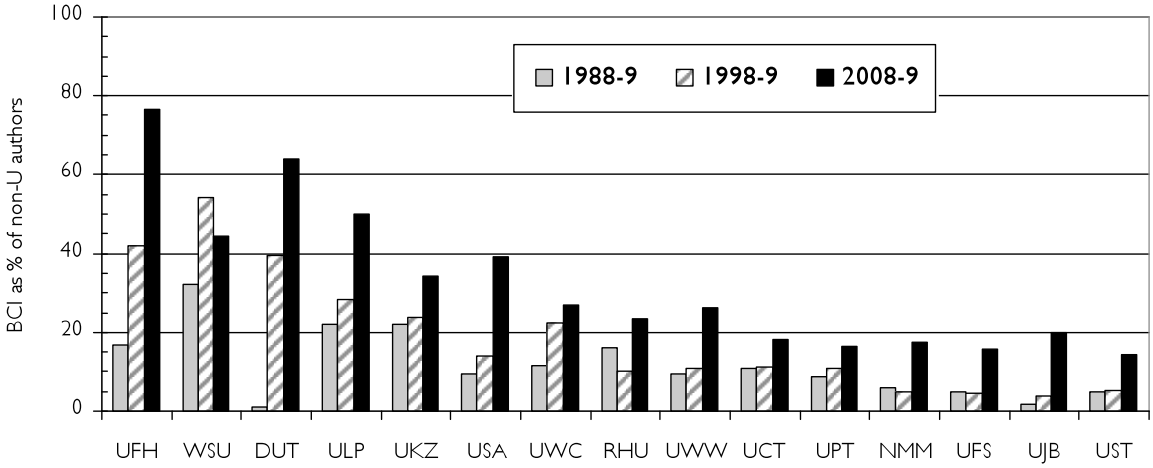

Figure 6. Percentage of authorship (excluding unknowns) by non-white authors (BCI) in 15 universities in South Africa; fractional counts. For codes, see Table 1.

The results for the two large university hospitals (Groote Schuur and Tygerberg) are in line with those of their parent universities, with the former having almost twice the percentages of non-whites among its researchers as the latter.

Discussion and Conclusions

Comparison of results with other data sources

It is difficult to find data on the numbers of the different racial groups in employment as researchers, as the South African labour force surveys do not distinguish this job type, and the R&D surveys do not distinguish racial groups. However, there are data on the composition of South Africa's leading science academy, the Academy of Science of South Africa (ASSAfr), whose members are listed on their website (http://www.assaf.org.za/?page_id=176). We were able to identify the racial groups of almost all the members (342 out of 353) by their names, and so to compare the percentage presence of each group among both publishing scientists and distinguished researchers (many of whom were described as "emeritus", and so possibly no longer active). The comparison is shown in Figure 7, and it is clear that the composition of the academy reflects very well the composition of the research community, except that Afrikaners are somewhat under-represented – possibly because they have tended to go into other professions. This graph also shows that, although the number of names of members of the academy is quite small (and their first names were also given) so that most could be individually allocated to a group by inspection, the group distribution agrees well with the allocation of very large numbers of names to groups by a variety of means and a mechanical process.

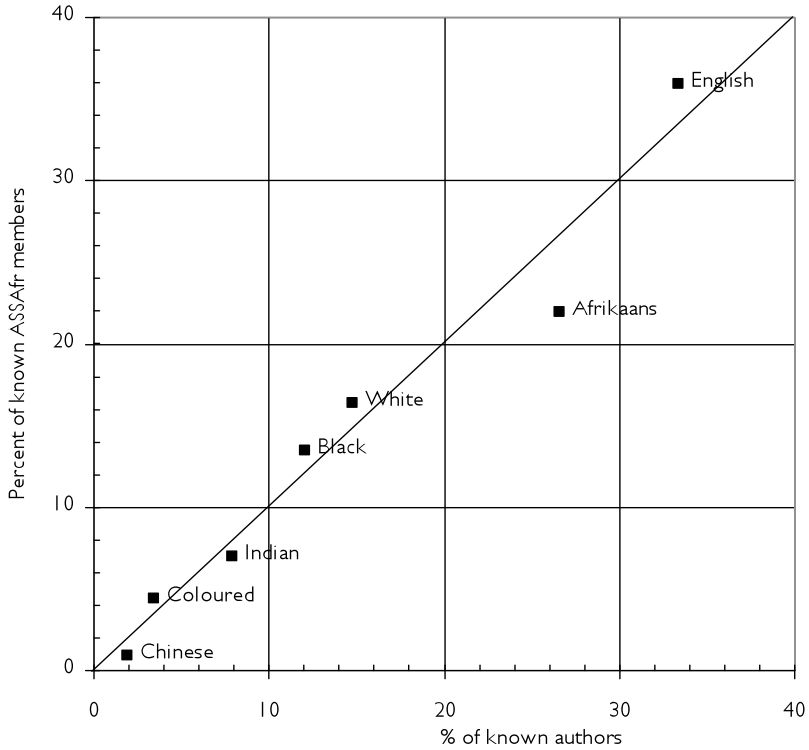

Figure 7. Percentage of membership (excluding unknowns) by scientists of different racial groups in the Academy of Science of South Africa, integer counts, and comparison with their presence in South African research outputs, 2008-09, fractional counts.

Conclusions

It appears that the identification of South African scientific authors by population group on the basis of their surnames (or family names) works satisfactorily and that therefore conclusions can be drawn on the status of the different groups at different times, pre- and post-*apartheid.* The change in the law has clearly made a big difference to the representation of the non-white groups in South African science, particularly to that of Black Africans, but equally clearly, much remains to be done as they account for four fifths of the population but only one eighth of the researchers. This slow process is almost inevitable, given the generally poor quality of schooling available under *apartheid* and the time needed to go from schoolgirl or schoolboy to researcher. The universities themselves need to instil in the undergraduates the need for doing research and for accessing information independently. Most black students coming from rural schools lack information-seeking and language skills, and so are spoon-fed with study notes, which add to their disadvantage. This dependence on help somehow destroys the students' sense of initiative. A further constraint has been the government's desire not to rush the change, and destroy what good schools there are in order to spread the available money for education equally among the whole population.

More effort is needed to provide well-qualified and dedicated teachers in all schools, especially in science and mathematics. While better school education will be needed to bring on a larger cohort of non-white researchers, particularly Black Africans, some steps might be taken now to encourage those who are already qualified. For example, universities with low scientific output might consider hiring distinguished scientists from abroad who could stimulate the South African researchers, and also provide international contacts for them, especially in Europe where it appears that non-white researchers are less at home than in the USA. Encouragement for attendance at international conferences would also be helpful, as would the holding of events in these universities. The ISSI conference to be held at the

University of Zululand in association with other campuses of the Durban University of Technology is an excellent step in this direction.

Acknowledgment

The analysis was made possible by the use of Excel macros written by Dr Philip Roe and Professor Judit Bar-Ilan, for which we are most grateful.

References

Biondi, G., Lasker, G.W., Raspe, P.D. & Mascie-Tayor, C.G.N. (1993) Inbreeding coefficients from the surnames of grandparents of the schoolchildren in Albanian-speaking Italian villages. *Journal of Biosocial Science,* 25, 63-71.

Biondi, G., Raspe, P. & Mascie-Tayor, C.G.N. (2001) Migration through surnames in Campobasso Province, Italy. *Journal of Biosocial Science,* 33, 305-310.

Bittles, A.H., Grant, J.C. & Shami, S.A. (1993) Consanguinity as a determinant of reproductive behaviour and mortality in Pakistan. *International Journal of Epidemiology,* 22, 463-467.

Blankley, W. & Kahn, M. (2005) The history of research and ex perimental development measurement in South Africa and some current perspectives. *South African Journal of Science*, 101, 151-156.

Chakraborty, R., Barton, S.A., Ferrell, R.E. & Schull, W.J. (1989) Ethnicity determination by names among the Aymara of Chile and Bolivia. *Human Biology*, 61, 159-177.

Crow, J.F. (1980) The estimation of inbreeding from isonymy. *Human Biology*, 52, 1-14.

Ingwersen, P. & Jacobs, D. (2004) South African research in selected scientific areas: status 1981-2000. *Scientometrics*, 59, 405-423.

Jacobs, D. & Ingwersen, P. (2000) A bibliometric study of the publication patterns in the sciences of South African scholars 1981-96. *Scientometrics,* 47, 75-93.

Jeenah, M. & Pouris, A. (2008) South African research in the context of Africa and globally. *South African Journal of Science*, 104, 351-354.

Lagerberg, D., Magnusson, M. & Sendelin, E. (2005) Surname as a marker of ethnicity. A study from child health services shows that immigrant respective Swedish fmilies seem to be isolated in different ways. *Lakartidningen*, 102, 2145-2148.

Lewison, G. & Igic, R. (1999) Yugoslav politics, "ethnic cleansing" and co-authorship in science. *Scientometrics*, 44, 183-192.

Lewison, G. & Kundra, R. (2008) The internal migration of Indian scientists, 1981-2003, from an analysis of surnames. *Scientometrics*, 75, 21-35.

Lewison, G. (2009) The percentage of reviews in research output: a simple measure of research esteem. *Research Evaluation*, 18, 25-37.

Mouton, J. (2000) Patterns of research collaboration in academic science in South Africa. *South African Journal of Science*, 96, 458-462.

Nicoll, A., Bassett, K. & Ulijaszek, S.J. (1986) What's in a name? Accuracy of using surnames and forenames in ascribing Asian ethnic identity in English populations. *Journal of Epidemiology and Community Health*, 40, 364-368.

Piazza, A., Rendine, S., Zei, G *et al*. (1987) Migration rates of human populations from surname distributions. *Nature*, 329, 714-716.

Pouris, A. (2003) South Africa's research publication record: the last ten years. *South African Journal of Science*, 99, 425-428.

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

Sooryamoorthy, R. (2010) Science and scientific collaboration in South Africa: *apartheid* and after. *Scientometrics*, 84, 373-390.

Startseva, E.A., Elchinova, G.I., Mamedova, R.A. *et al*., (1994) The use of the migration index, the parameter of surname diversity, and the values of entropy and redundancy of surname distribution for description of population-structure. *Genetika,* 30, 978-981.

Tjam, E.Y. (2001) How to find Chinese research participants: use of a phonologically-based surname search method. *Canadian Journal of Public Health,* 92, 138-142.

Webster, B.M. (2004) Bibliometric analysis of presence and impact of ethnic minority researchers on science in the UK. *Research Evaluation*, 13, 69-76.