

Final examination performance of medical students from ethnic minorities

I C McManus, P Richards, B C Winder & K A Sproston

Imperial College School of Medicine at St. Mary's, London, UK

SUMMARY

It has recently been claimed that UK medical students from ethnic minorities perform less well than White students in final undergraduate examinations and that this results from discrimination in clinical examinations. In this paper the authors examine the performance of students in two prospective studies of 1981 and 1986 entrants to London medical schools taking finals in 1986, 1987, 1991 and 1992. Overall, 18.2% of students in the survey failed one or more examinations at finals. UK ethnic minority students were 2.09 times more likely to fail one or more examinations than were UK White students. Male students were also 1.65 times more likely to fail an examination than female students. More detailed analysis of individual marks confirmed that UK ethnic minority students performed less well than UK White students, although non-UK ethnic minority students performed *better* than UK White students. The difference between UK White and ethnic minority students could not be explained by differences in educational achievement, study habits or clinical experience. The poorer performance of UK ethnic minority students extended across multiple-choice questions (MCQ), essay, clinical and oral examinations. Ethnic minority students performed particularly poorly on examinations in medicine and surgery as compared with pathology and clinical pharmacology. Although UK-born ethnic minority students perform less well in final examinations than UK White students, this is unlikely to be explained by racial discrimination as the difference is present in MCQ examinations marked by machine and as non-UK ethnic minority students perform *better* than UK White students. The cause of the difference in performance is therefore not clear and requires further study.

Keywords

*Educational measurement; educational status; *ethnic groups; Great Britain; *minority groups; prospective studies; *students, medical

INTRODUCTION

Medical students and doctors from ethnic minorities (Smith 1987; McKenzie 1995) are disadvantaged, but the reasons for this are not yet clear. Recently, it emerged that at the University of Manchester Medical School, 'male medical students with Asian surnames are more likely to fail their final clinical examinations than other students' (Dillner 1995). The reasons for the relatively high failure rate in Manchester are not clear: the A-level grades of those failing were equivalent to those achieved by other students, and the discrepancy was not found in written examinations. We are able to study two cohorts of medical students admitted to London medical schools in 1981 and 1986. In particular, we have looked for systematic differences between White and non-White students in final examination performance, determined whether this is restricted to specific modes of examination or particular examination subjects, examined possible differences in background factors, including O- and A-level performance, Second MB performance, intercalated degrees, different study habits or different clinical experience, and have compared ethnic minority students who are UK nationals with those who are not.

METHODS

In 1980 and 1985 we began prospective studies of medical student selection, studying all applicants for admission to St. Mary's Hospital Medical School in 1981 and 1986, constituting 13% and 25%, respectively, of all applications in those years. Detailed results of those studies have been reported elsewhere (McManus & Richards 1984; McManus *et al.* 1989); St. Mary's is a small school and most of the successful applicants in the survey were admitted to other schools. We had detailed information readily available on the O- and A-level examination results of all these students, and in the 1986 cohort we also had a measure of study habits (a shortened version of the Biggs Study Process Questionnaire; Biggs 1978). Performance in Second MB examinations and whether or not students had taken an intercalated BSc degree were ascertained

from medical school registrars. Students were subsequently followed up in the last few months of their final clinical year with a questionnaire asking, *inter alia*, about experience of a wide range of medical conditions seen, operations witnessed, and practical procedures carried out (McManus *et al.* 1993), as well as study habits (Biggs 1978).

All students who had entered London medical schools and took University of London final examinations, either school based or university based, were included in the study. Students at St. George's Hospital Medical School and the Royal Free Hospital School of Medicine who took school-based examinations were included only in the overall analysis, and in those parts of examinations (particularly MCQs) which they took in common with those taking the university-based final examinations. The examinations of the University of London (then) had a similar structure, all schools participating in the university-based examinations. In 1988 and in 1994 we obtained detailed examination results on all students in our study from the records kept by the registry in Senate House. In this study we consider all the students in the two cohorts for whom results of those final examinations were available.

The examination was in five distinct sections, most of which, with the exception of Pathology, were taken in the final year of the course. In all five sections (i.e. medicine, including psychiatry, paediatrics and public health); surgery; pathology; clinical pathology; clinical pharmacology; and obstetrics/gynaecology) there were MCQ and essay examinations; in addition, in all of the subjects except pathology there was an oral examination (*viva voce*), and all except clinical pharmacology included a practical examination – for medicine, surgery and obstetrics/gynaecology in the form of a clinical examination. In each subject a pass/fail decision was made separately, and students were divided into those who had failed one or more subjects and those who had failed none; it should be noted that because of rules concerning compensation it was possible in some circumstances to fail a subject and yet pass the examination overall. All of each candidate's individual marks were also collected, and a principal component analysis used to calculate, separately for each cohort, a single score that best described all of the examination results – the 'overall score'. This method ensured that different methods of marking and scoring could be combined, and that high correlations between subcomponents did not unduly weight some components. Additionally, all marks were converted to standardized scores (z -scores) with a mean of 0 and a standard deviation of 1. Marks were calculated for the five separate subjects (medicine, surgery, pathology, clinical pharmacology and obstetrics/gynaecology)

and for the four modes of assessment (MCQ, essay, oral and clinical/practical examinations) by averaging across all appropriate z -scores and then restandardizing the new scores so that the standard deviation remained at unity across the group as a whole. As all marks were approximately normally distributed, all of the derived marks were also distributed normally to a first approximation.

The ethnic origin of the students was classified in slightly different ways in the two cohorts. In the 1981 cohort the students were divided into those with and those without a non-European surname (McManus & Richards 1985), a method that has high validity for identifying students from ethnic minorities (McManus *et al.* 1990); the vast majority of students with non-European surnames were of Asian origin. Students in the 1986 cohort were classified on the basis of their own self-described ethnic origin (McManus *et al.* 1989). In this paper we will distinguish between White students and all others from ethnic minorities. We also distinguish between UK and non-UK citizens.

Statistical analysis was carried out using the Statistical Package for the Social Sciences (SPSS), using t -tests, analysis of variance (ANOVA) and covariance and multivariate ANOVA (MANOVA) as appropriate. Missing values in multivariate analyses were replaced by mean substitution. A difficult problem in any statistical analysis is the setting of an appropriate value for α , the probability of a type I error, which is conventionally set at 0.05 – the 5% level. Sometimes it is argued that in situations in which the problems associated with a type I error are high, as could be argued in the present case, then α should be set at some lower level, say 0.01. The problem with that is that as α is set lower and lower so β , the probability of a type II error, inexorably rises (Howell 1992). In the present case we tried to solve this dilemma by keeping our *a priori* α level at 0.05, but in the Discussion section we consider to what extent our results would have been different if α had been set at a lower level such as 0.01.

RESULTS

Of a total of 330 students from the 1981 cohort and 361 students from the 1986 cohort who had taken London final examinations, ethnic origin and nationality were known in 671 cases: 86 were from ethnic minority groups, of whom 62 (72%) were UK nationals. Only 9 of the 585 White students were non-UK nationals, a number too small to compare statistically with other groups; for the remaining analyses they are included within the White group. Measures did not differ significantly between cohorts, and therefore the two cohorts were combined for further analysis.

The initial analysis compared those students who had passed all their final examinations (including pathology) with those who failed one or more examinations at finals; 18.2% (126/691) of students were in the latter group. Among UK nationals, 16.7% (96/576) of White students failed one or more examinations compared with 30.6% (19/62) of ethnic minority students. Among non-UK students, 33.3% (3/9) of White students failed one or more examinations, as did 16.7% (4/24) of non-White students. Of the male students, 21.6% (85/393) had failed one or more final examinations, as compared with 13.8% (41/298) female students. Multiple logistic regression of all candidates showed that men were significantly more likely to fail an examination than women ($P = 0.0181$, odds ratio = 1.653, 95% CI = 1.090-2.507) and that ethnic minority students were significantly more likely to fail an examination than White candidates ($P = 0.0274$). However, the last result is difficult to interpret because of an almost significant interaction of ethnic origin with nationality ($P = 0.0827$). The analysis was therefore re-run, but restricted to UK nationals. The effect of sex was unchanged (odds ratio = 1.660, $P = 0.0202$); in addition, ethnic minority students were significantly more likely to fail an examination than White students ($P = 0.0136$, odds ratio = 2.09, 95% CI = 1.164-3.766).

A more detailed analysis was possible using overall finals score, which is a continuous measure approximately normally distributed, which was analysed by ANOVA with between-subjects factors of ethnic origin, nationality and sex. Overall, women performed better than men ($F(1,663) = 9.84$, $P = 0.002$); there were also significant main effects of ethnic origin ($F(1,663) = 7.11$, $P = 0.008$) and of nationality ($F(1,663) = 10.73$, $P = 0.003$) and an interaction between ethnic origin and nationality ($F(1,663) = 4.29$, $P = 0.039$). Figure 1 shows that ethnic minority candidates of UK nationality performed less well than White UK nationals, whereas non-UK students from ethnic minorities performed better than other groups, in each case the difference being about 0.45 standard deviations. Restricting the analysis to UK nationals confirmed that the performance of ethnic minority UK students was less good than that of White UK students ($F(1,637) = 10.09$, $P = 0.002$).

Assessment of the prior educational achievement of UK ethnic minority and White students showed no significant differences on O-level, A-level or Second MB examinations, and similar proportions took an intercalated BSc degree (Table 1). Analysis of variance in the UK nationals after removing as covariates the effects of O- and A-level grades, Second MB performance and taking a BSc showed that, although the covariates were jointly highly significant in predicting

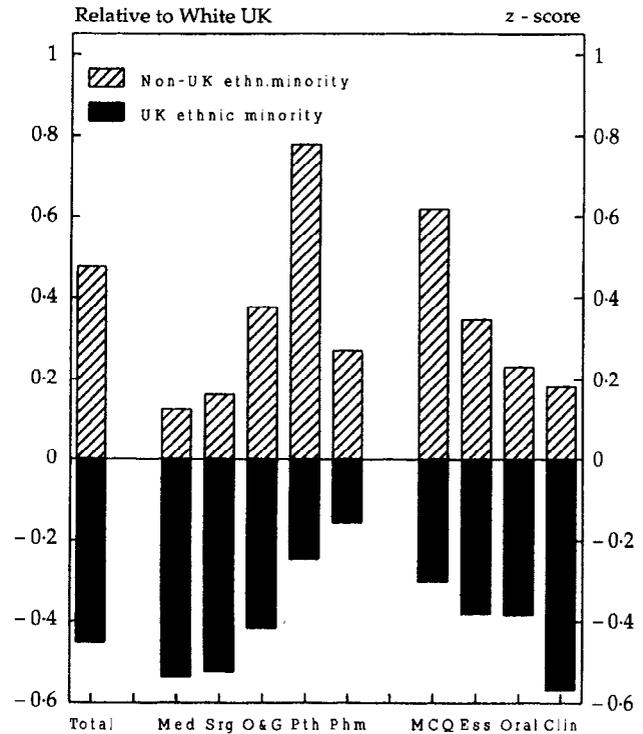


Figure 1 The performance of UK and non-UK ethnic minority students, relative to UK White students. All measures are in z-score units, with a standard deviation of one for the whole population.

finals performance ($F(4,614) = 36.04$, $P < 0.001$), ethnic minority and White students were still significantly different ($F(1,614) = 36.04$, $P < 0.001$); the sex difference also remained significant after taking the covariates into account ($F(1,614) = 4.98$, $P = 0.026$). The differences between UK ethnic minority and White students cannot therefore be explained in terms of differences in previous educational achievement. The clinical experience and study habits in the final year (and, in the 1986 cohort, at application) showed no significant differences between White and UK ethnic minority students; analysis of covariance showed that, although these covariates were jointly significant in predicting final examination results ($F(9,625) = 4.07$, $P < 0.001$), the effect of ethnic origin still remained highly significant ($F(1,625) = 8.74$, $P = 0.003$), as also did the sex difference ($F(1,614) = 8.21$, $P = 0.004$). Differences between UK and White ethnic minority students in final examinations cannot therefore be explained in terms of previous examination achievement, study habits or clinical experience.

As students took examinations in five different subjects, and were examined by four different methods, we used MANOVA to assess whether the profile of achieve-

Table 1 Demographic, educational, study habit and clinical experience measures in White and ethnic minority students

Mean (SD; N)	White (UK and non-UK)	Ethnic minority		Significance	
		UK	Non-UK	White vs. UK ethnic minority	White vs. non-UK ethnic minority
<i>Demographic measures</i>					
Female	44.6% (261/585)	32.3% 20/62	37.5% (9/24)	$P = 0.062$	NS
Mature entrants (age ≥ 21 at entry to medical school)	9.1% (53/582)	11.3% (7/62)	8.3% (2/24)	NS	NS
<i>Educational achievement</i>					
Mean O-level grade (A = 5, B = 4, C = 3, D = 2, E = 1, O/F = 0) (McManus & Richards 1984; McManus <i>et al.</i> 1989)	4.27 (0.53; 571)	4.28 (0.39; 61)	4.44 (0.43; 20)	NS	NS
Mean A-level grade (A = 5, B = 4, C = 3, D = 2, E = 1, O/F = 0) (McManus & Richards 1984; McManus <i>et al.</i> 1989)	3.99 (0.74; 583)	3.99 (0.75; 61)	4.48 (0.71; 24)	NS	$P = 0.002$
Second MB (4 = Honours or Distinction; 3 = Pass first time; 2 = Pass with resits; 1 = Fail) (McManus & Richards 1986)	2.82 (0.54; 585)	2.77 (0.55; 62)	3.04 (0.69; 24)	NS	$P = 0.058$
Intercalated BSc taken	26.8% (157/585)	25.8% (16/62)	20.8% (5/24)	NS	NS
<i>Learning style (Study Process Questionnaire; Biggs 1978)</i>					
Application (1986 cohort only):					
Surface learning	12.81 (3.58; 274)	13.91 (3.09; 36)	11.22 (3.31; 9)	NS	NS
Strategic learning	23.07 (4.17; 281)	23.89 (4.23; 37)	24.44 (3.17; 9)	NS	NS
Deep learning	21.52 (3.98; 278)	22.00 (4.42; 39)	23.86 (3.19; 7)	NS	NS
Final year (1981 and 1986 cohorts):					
Surface learning	12.87 (3.85; 305)	14.47 (3.93; 19)	13.83 (4.24; 12)	NS	NS
Strategic learning	14.71 (4.89; 342)	15.25 (4.76; 20)	17.42 (5.09; 12)	NS	$P = 0.061$
Deep learning	18.27 (4.41; 339)	17.1 (4.79; 20)	19.5 (5.16; 12)	NS	NS
<i>Clinical experience (McManus <i>et al.</i> 1993)</i>					
Medical conditions	33.39 (6.48; 353)	33.45 (5.37; 22)	34.08 (4.48; 12)	NS	NS
Surgical operations	45.10 (7.25; 354)	46.59 (7.25; 22)	43.42 (3.87; 12)	NS	NS
Practical procedures	41.39 (3.91; 354)	41.55 (3.83; 20)	41.92 (2.64; 12)	NS	NS

ment in different subjects or modes of examination differed between UK White and ethnic minority students. A graph of performance in the different types of examination and the different subjects (Fig. 1) suggests that UK ethnic minority students may perform particularly poorly in clinical examinations (which included the practical examination in pathology); however, MANOVA found no evidence of a significant interaction between

ethnic origin and examination type ($F(3,1902) = 1.32$, NS) or between ethnic origin and examination subject ($F(4,2536) = 1.90$, NS). Exclusion of the pathology practical examination from the clinical examinations produced a similarly non-significant interaction between ethnic origin and examination type ($F(3,1902) = 1.26$, NS). Analysis of all students (including non-UK students) found no evidence of interactions between

examination mode and ethnic origin or nationality or an examination made by ethnic origin by nationality interaction; that is the profiles shown in Fig. 1 are statistically flat, and there is no evidence that ethnic minority students perform significantly more poorly in any particular mode of examination. Among the UK students the difference between White and ethnic minority students on the MCQ examination was statistically significant in its own right (ANOVA: $(F(1,634) = 4.31, P = 0.038)$).

Analysis by examination subject did find that the subject by ethnic group interaction was significant ($F(4,2652) = 2.98, P = 0.018$), although there was no interaction with nationality ($F(4,2652) = 2.13, NS$). Figure 1 suggests that ethnic minority subjects perform less well at the more clinical examinations of medicine and surgery in comparison with subjects such as pathology and clinical pharmacology, although the above findings show that this difference cannot readily be explained by the mode of these examinations.

DISCUSSIONS

To prove racial discrimination unequivocally is not straightforward and rarely has the elegance of a controlled, experimental study (Esmail & Everington 1993). The mere demonstration of differences between White and ethnic minority groups is not sufficient in itself as ethnic groups differ in many ways from each other. Thus, the Commission for Racial Equality found that the main reason for a low proportion of Black students on nursing courses was actually a low application rate (Commission for Racial Equality 1987), and that fewer Black applicants were accepted for social work owing to poorer examination results and less appropriate experience (Gardiner 1985). Nor can differences always be explained by discrimination at face-to-face interviews; in the case of senior house officer (SHO) selection the difference arose at the shortlisting not the interview stage (McKeigue *et al.* 1990), and in the MRCGP examination the difference was present in the MCQ and not in the oral examination (Wakeford *et al.* 1992).

This prospective study of students taking the final medical examinations of the University of London confirms the suggestion that UK students from ethnic minorities perform less well than do White students. As the study is prospective, it can also confirm that UK ethnic minority students are equivalent to White students on measures of previous examination performance (O-levels, A-levels, Second MB and the taking of an intercalated BSc). Also, there is no evidence that UK ethnic minority students differ from White students in their

study habits, either at the time of selection or in the final year, or in their experience of medical and surgical conditions, or practical procedures.

The hypothesis that the poorer achievement results from overt and covert ethnic discrimination, as implied in the Manchester report (Dillner 1995), meets several difficulties. Firstly, we note that it is only UK ethnic minority students who perform less well: non-UK ethnic minority students perform *better* than White UK students. Unless examiners can assess not only the ethnic origin of students but also their nationality then this result is difficult to ascribe to discrimination. More problematic is the fact that, among the UK students, the ethnic minority group performed less well in all types of examination, particularly in the MCQ examination, which is marked by machine. It has also been reported for the MRCGP examination that ethnic minority candidates perform less well on the MCQ component of the examination (Wakeford *et al.* 1992), although that difference was not present in UK-trained Asian doctors, but only in those who had trained outside the UK. Although, overall it seems clear that the UK ethnic minority students in our study, as in the Manchester study, performed less well than White students in final examinations, that difference cannot readily be explained as a result of discrimination, and at the moment is unexplained.

As stated earlier, the a priori α -level for our analysis was set at 0.05. In view of the potential importance of our results, we reconsidered them setting α instead at the more stringent level of 0.01. To some extent this is, of course, an artificial exercise as it is clear in advance that if we set α at some very extreme level, say 10^{-30} , then almost no difference would ever be found between any groups, and almost the whole of the world's scientific literature would be 'non-significant', and likewise would be ridden with type II errors. Be that as may be, the effects of a more stringent 0.01 α -level do not dramatically alter our conclusions. The results of the logistic regression using pass/fail as the dependent variable are now 'non-significant', but that may be expected as logistic regression of a binary dependent variable is inherently less powerful than statistical procedures that use continuous measures, such as those used in the remainder of the paper. The important thing is that the *pattern* of the results is consistent with the analysis of the more sensitive continuous measures. Considering the continuous measures of examination performance – which more nearly represent the true situation in which examination candidates vary from the extremely good through to the extremely bad, through all gradations between and in an approximately normal distribution – then almost none of the results are altered by using $\alpha = 0.01$. UK ethnic minority candidates still perform significantly less well

than White candidates, and that result is still valid after taking all relevant covariates into account. The only substantive result to change is that the individual analysis of the MCQ results no longer shows a significant difference between White and ethnic minority students. However, it is still the case that there is no overall interaction between ethnic group and examination type, so that the general conclusion must be that, overall, UK ethnic minority students perform less well than White students, and that there is no evidence of differences between different types of examinations. We believe therefore that our conclusions are robust against any reasonable change in α -level.

In summary, our study suggests that UK ethnic minority students perform less well at final examinations than do White students, and that this difference cannot be readily explained in terms of previous educational experience or in terms of differential performance on particular types of examination.

ACKNOWLEDGEMENTS

Follow-up of the 1981 cohort study and the initial 1986 cohort study was funded by the Economic and Social Research Council; follow-up of the 1986 cohort was funded by the Leverhulme Foundation. We thank the staff of the Senate House Registry, and in particular Ms Cristiana Hill and Mr Stephen Brooke, for their assistance. Finally, we thank the students in our survey for their help in completing our questionnaires.

REFERENCES

- Biggs J B (1978) Individual and group differences in study processes. *British Journal of Educational Psychology* **48**, 266-798.
- Commission for Racial Equality (1987) *Ethnic Origins of Nurses Applying for and in Training - a Survey*. Commission for Racial Equality, London.
- Dillner L (1995) Manchester tackles failure rate of Asian students. *British Medical Journal* **310**, 209.
- Esmail A & Everington S (1993) Racial discrimination amongst doctors from ethnic minorities. *British Medical Journal* **306**, 691-2.
- Gardiner D (1985) *Ethnic Minorities and Social Work Training*. Central Council for Education and Training in Social Work, London.
- Howell D C (1992) *Statistical Methods for Psychology*. PWS-Kent, Boston.
- McKeigue P M, Richards J D M & Richards P (1990) Effects of discrimination by sex and race on the early careers of British medical graduates during 1981-7. *British Medical Journal* **301**, 961-4.
- McKenzie K J (1995) Racial discrimination in medicine [Editorial]. *British Medical Journal* **310**, 478-9.
- McManus I C & Richards P (1984) An audit of admission to medical school. 1. Acceptances and rejects. *British Medical Journal* **289**, 1201-4.
- McManus I C & Richards P (1985) Admission to medical school [letter]. *British Medical Journal* **290**, 319-20.
- McManus I C & Richards P (1986) Prospective survey of performance of medical students during preclinical years. *British Medical Journal* **293**, 124-27.
- McManus I C, Richards P & Maitlis S L (1989) Prospective study of the disadvantage of people from ethnic minority groups applying to medical schools in the United Kingdom. *British Medical Journal* **298**, 723-26.
- McManus I C, Maitlis S L & Richards P (1990) Identifying medical school applicants from ethnic minorities. *Studies in Higher Education* **15**, 57-73.
- McManus I C, Richards P, Winder B C, Sproston K A & Vincent C A (1993) The changing clinical experience of British medical students. *Lancet* **341**, 941-4.
- Smith R (1987) Prejudice against doctors and students from ethnic minorities [Editorial]. *British Medical Journal* **294**, 328-9.
- Wakeford R, Farooqi A, Rashid A & Southgate L (1992) Does the MRCGP examination discriminate against Asian doctors? *British Medical Journal* **305**, 92-4.

Received 11 September 1995; accepted for publication 27 October 1995