



Does examiner bias in undergraduate oral and clinical surgery examinations occur?

Douglas Stupart, Paul Goldberg, Jake Krige, Delawir Khan

Objective. Oral and long case clinical examinations are open to subjective influences to some extent, and students may be marked unfairly as a result of gender or racial bias or language problems. These concerns are of topical relevance in South Africa. The purpose of this study was to assess whether these factors influenced the marks given in these examinations.

Methods. Final-year surgery examination results from the University of Cape Town from 2003 to 2006 were reviewed. These each consisted of a multiple choice paper, an objective structured clinical examination, a long case clinical examination and an oral examination.

Results. The marks of 604 students were analysed. Students who spoke English as a home language performed better in all examination modalities. Female students scored slightly

higher than males overall, but they scored similarly in the clinical and oral examinations. There were significant differences in the marks scored between the various population groups in all examination modalities, with white students achieving the highest scores, and black students the lowest. These differences were most marked in multiple choice examinations, and least marked in oral and clinical examinations.

Conclusion. We could find no evidence of systemic bias in the oral and clinical examinations in our department, which reinforces the need for ongoing academic support for students from disadvantaged educational backgrounds, and for those who do not speak English as a home language.

S Afr Med J 2008; 98: 805-807

Numerous examination modalities are used to assess theoretical knowledge and competency in medical students. Oral and long case clinical examinations are particularly open to criticism as they are inherently subjective to some degree and may also partly involve the assessing of language skills rather than examining students' grasp of the curriculum.¹ Furthermore, the examiner may have a conscious or unconscious bias that could influence certain students' marks; this is a particular concern in South Africa, where racial classification and prejudice have played a significant part in the country's history and politics. The demographics of staff and students within the Department of Surgery at the University of Cape Town are disparate in that most lecturers (and examiners) are white males, while the majority of students are female and of other races. Furthermore, the medium of teaching and examining (and the home language of most examiners) is English, whereas many students speak another first language. We wished to assess whether systemic bias according to language, gender or population group has influenced the marks given in oral and clinical examinations in our department.

Ethical approval for this study was granted by the research ethics committee of the UCT Faculty of Health Sciences.

Department of Surgery, University of Cape Town

Douglas Stupart, MB ChB, FCS (SA)

Paul Goldberg, MMed, FCS (SA)

Jake Krige, MB ChB, FCS (SA), FRCS (Ed), FACS

Delawir Khan, ChM, FCS (SA)

Corresponding author: D Stupart (D.Stupart@uct.ac.za)

Methods

The University of Cape Town (UCT) final surgery examination consists of four modalities: (i) 400-question multiple choice (MCQ) paper; (ii) objective structured clinical examination (OSCE) consisting of 20 clinical stations each with short questions about a clinical problem such as X-ray interpretation, recognising photographs of clinical signs, etc.; (iii) long case clinical examination; and (iv) oral examination.

The MCQ papers are marked by computer. At each of the 20 OSCE stations, students give a short written answer. The set of answers for each station is marked by a different examiner. Marking of the OSCE is not blinded as students' names are on the answer sheet. During the clinical examination, each student is required to take a history from, and examine, a patient and present the case to two examiners. The examiners assess the ability of the student to accurately assess the clinical case, and ask relevant questions related to that patient. The unstructured 20-minute oral examination, with two examiners, examines the student on a wide range of general surgical topics.

We reviewed students' final-year examination marks from 2003 to 2006. We analysed the marks allocated for each component of the examination (MCQ, OSCE, oral and clinical) separately. All marks were equally weighted and are stated as percentages. Pearson correlation coefficients were calculated between the four examination modalities, and consistency between them was calculated using Cronbach's α coefficient (a widely used psychometric statistical tool for assessing whether a number of different tests measure a similar construct²).

The self-declared home language, gender and population group of each student were obtained from UCT records. The



terms 'white', 'coloured', 'Indian' and 'African' are used as they were population groups as defined by racial classification legislation formerly in South Africa. We do not consider these terms legitimate, except to acknowledge that the different experiences of people so categorised have led to persistent inequities.

Statistical analysis was done using MedCalc (Mariakerke, Belgium) software. Continuous variables were analysed using the Student's *t*-test (when two sets of data were compared) or Kruskal-Wallis test (when more than two groups were compared). Categorical data were compared using the chi-square test. In all instances where parametric statistical tests were used, the data fulfilled the required criteria of normality and equal variance. All results are stated as mean (95% confidence interval (CI)) unless otherwise stated.

Results

Between 2003 and 2006, 694 students sat the final surgery examinations. Demographic data were incomplete for 90 students and were excluded from the study, leaving 604 for analysis, comprising 369 female and 235 male students. One hundred and seventy students described themselves as African, 99 as coloured, 102 as Indian and 233 as white. English was the first language of 437 students, while 167 spoke another home language. Population group was significantly associated with home language. English was the home language of 222/233 white, 96/102 Indian, 96/99 coloured and 23/146 African students ($p < 0.0001$).

The highest correlation coefficient between examination modalities was between the MCQ and OSCE examinations (0.53). The lowest correlation was between the oral and clinical examinations (0.31). The correlation coefficients between the other combinations of examination modalities were all between 0.34 and 0.36. The *p*-values for correlation were < 0.0001 in all cases. Cronbach's α coefficient for consistency between the examinations was 0.74, which shows adequate consistency.²

Students whose home language was English scored significantly higher marks in each of the examination modalities compared with students who spoke another home language. The mean marks were 63.5 (62.8 - 64.3) v. 57.3 (55.9 - 58.6) in the MCQ, 68.8 (68.1 - 69.5) v. 63.1 (61.9 - 64.3) in the

OSCE, 67.4 (66.8 - 68) v. 64.2 (63.1 - 65.2) in the oral and 68.6 (68.1 - 69.1) v. 65.1 (64.1 - 66.1) in the clinical examination; and *p*-values were < 0.0001 in all cases.

There was a small but statistically significant difference in the overall marks of female and male students (67.9 (67.3 - 68.5) v. 64.9 (64.2 - 65.6), $p = 0.02$). Female students scored significantly higher marks in the MCQ (62.5 (61.6 - 63.4) v. 60.6 (59.5 - 61.8), $p = 0.01$) and OSCE (68.1 (67.3 - 68.9) v. 65.9 (64.8 - 66.9), $p = 0.001$) examinations, but there was no significant gender difference in performance in the oral (66.8 (66.1 - 67.4) v. 66.0 (65.2 - 66.9), $p = 0.17$) and clinical (67.9 (67.3 - 68.5) v. 67.1 (66.4 - 67.9), $p = 0.1$) examinations.

The mean marks for the four examination modalities in each of the population groups are illustrated in Fig. 1 and Table I. There were significant ($p < 0.0001$) differences in the marks scored between the population groups in all examination modalities, with white students achieving the highest scores and African students the lowest. The difference in marks scored according to population group was most marked in the MCQ examination, and least marked in the oral and clinical examinations. When analysed according to population group, but including only those students who spoke English as their home language, these trends persisted but were less marked (Table II).

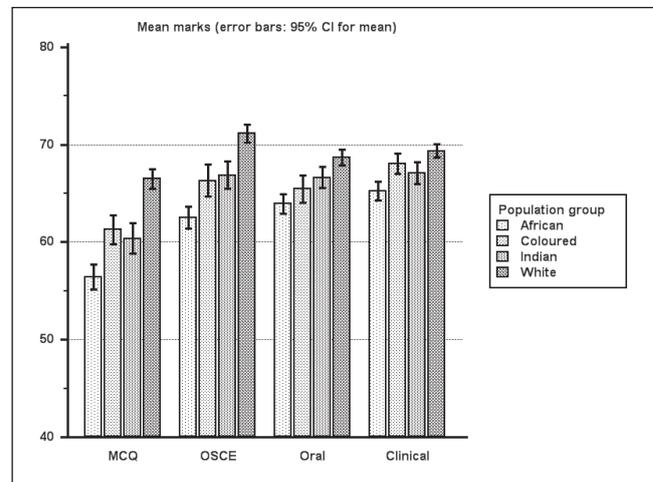


Fig 1. Marks achieved by students from different population groups in their final MB ChB surgery examinations.

Table I. Marks (stated as mean (95% CI)) achieved in the four examination modalities by students of different population groups

	Population group				<i>p</i> -value
	African	Coloured	Indian	White	
MCQ	56.4 (55.2 - 57.7)	61.3 (59.8 - 62.8)	60.4 (58.8 - 61.9)	66.5 (65.5 - 67.5)	< 0.0001
OSCE	62.5 (61.4 - 63.6)	66.3 (64.7 - 68.0)	66.9 (65.5 - 68.3)	71.2 (70.3 - 72.1)	< 0.0001
Oral	63.9 (64.3 - 66.2)	65.5 (64.1 - 66.9)	66.6 (65.6 - 67.7)	68.7 (67.9 - 69.5)	< 0.0001
Clinical	65.3 (64.3 - 66.2)	68.0 (67.0 - 69.1)	67.1 (66.0 - 68.3)	69.4 (68.7 - 70.0)	< 0.0001
Total	62.0 (61.3 - 62.8)	65.3 (64.3 - 66.2)	65.3 (64.4 - 66.1)	68.9 (68.3 - 69.5)	< 0.0001



Table II. Marks (stated as mean (95% CI)) achieved in the four examination modalities by English home-language students of different population groups

	Population group (English home-language students only)				p-value
	African	Coloured	Indian	White	
MCQ	59.1 (56.4 - 61.7)	61.2 (59.6 - 62.7)	60.2 (58.6 - 61.8)	66.4 (65.4 - 67.5)	<0.0001
OSCE	65.7 (62.9 - 68.5)	65.9 (64.3 - 67.5)	66.7 (65.2 - 68.1)	71.2 (70.3 - 72.2)	<0.0001
Oral	65.9 (63.9 - 67.9)	65.4 (64.0 - 66.9)	66.8 (65.6 - 67.9)	68.6 (67.8 - 69.4)	<0.0001
Clinical	67.6 (65.5 - 69.7)	68.1 (67.1 - 69.2)	67.0 (65.8 - 68.2)	69.5 (68.9 - 70.1)	0.002
Total	64.6 (62.9 - 66.2)	65.2 (64.2 - 66.1)	65.2 (64.3 - 66.0)	68.9 (68.3 - 69.6)	<0.0001

Conclusions and discussion

Oral examinations and long case clinical examinations have been a part of our assessment of medical students for many years, but the fairness of such examinations, which rely to some extent on a linguistically and culturally determined discourse between student and examiner, has been questioned.¹ This is of particular relevance in South Africa, with its cultural and linguistic diversity. We are not aware of previous studies that have attempted to detect examiner bias in oral and clinical examinations at a South African medical school.

Poorer performance of male and ethnic minority candidates in medical school examinations has been noted,³⁻⁸ and also in postgraduate medical examinations.⁹ Persistent inequities in South African society result in marked differences in schooling quality and economic status between population groups, which may manifest in differing performance in university examinations.¹⁰⁻¹²

Our study showed consistent, notable differences in the marks achieved in final-year surgery examinations at UCT between students from different population groups, and between English home-language students and those who spoke other home languages. These differences in performance were evident in all examination modalities. Female students performed slightly better overall in the examinations.

Language and population group were linked in this student population. The majority (414/434 (95%)) of white, coloured and Indian students spoke English as their first language, compared with only 23/170 (14%) of African students. English home-language students scored higher marks than other students in all examination modalities. As with other studies,¹ the differences in performance between population groups may, at least partly, be ascribed to differences in language ability.⁴ However, differences in performance between population groups persisted, but were less marked when only English home-language students' marks were analysed. This is similar to the findings of a study of two London medical schools' third-year examination results, which noted underperformance of Asian students, including those who spoke English as a first language.⁵

There was adequate consistency between the examination modalities, despite the subjective aspects of the oral and clinical examinations. Differences in performance by population group,

language and gender were greatest in the MCQ (which was marked by computer, and therefore effectively blinded to the students' ethnicity, gender or home language), and least evident in the oral and clinical examinations. Therefore, we could not identify any evidence that the differences in performance were the result of examiner bias in the oral and clinical examinations. However, our inability to detect such bias may reflect the methodological limitations of this purely quantitative study. Direct, qualitative studies of the interactions between examiners and students, using ethnographic or socio-linguistic tools, may detect more subtle differences in the ways that different students are examined.¹

This study highlights the need for ongoing academic support for students from educationally disadvantaged backgrounds who may have difficulties studying in a language other than their home language. Universities should also encourage their academic staff to learn an African language, to communicate better with their students. We also need to assess whether questions in the MCQ and OSCE examinations are sufficiently clearly worded for non-English-speaking students.

We thank Brenda Fine and Karen van Blerk for their assistance in retrieving the data.

References

1. Roberts C, Sarangi S, Southgate L, Wakeford R, Wass V. Oral examinations - equal opportunities, ethnicity, and fairness in the MRCP. *BMJ* 2000; 320: 370-375.
2. Bland JM, Altman DG. Statistics notes: Cronbach's alpha. *BMJ* 1997; 314: 572.
3. McManus IC, Richards P, Winder BC, Sproston KA. Final examination performance of medical students from ethnic minorities. *Med Educ* 1996; 30: 195-200.
4. Wass V, Roberts C, Hoogenboom R, Jones R, Van der Vleuten C. Effect of ethnicity on performance in a final objective structured clinical examination: qualitative and quantitative study. *BMJ* 2003; 326: 800-803.
5. Haq I, Higham J, Morris R, Dacre J. Effect of ethnicity and gender on performance in undergraduate medical examinations. *Med Educ* 2005; 39: 1126-1128.
6. Woolf K, Haq I, McManus C, Higham J, Dacre J. Exploring the underperformance of male and minority ethnic medical students in first year clinical examinations. *Adv Health Sci Educ Theory Pract* 2007; DOI 10.1007/s10459-007-9067-1.
7. Kay-Lambkin F, Pearson SA, Rolfe I. The influence of admission variables on first year medical school performance: A study from Newcastle University, Australia. *Med Educ* 2002; 36: 154-159.
8. Lumb AB, Vail A. Comparison of academic, application form and social factors in predicting early performance on the medical course. *Med Educ* 2005; 38: 1002-1005.
9. Bessant R, Bessant D, Chesser A, Coakley G. Analysis of predictors of success in the MRCP (UK) PACES examination in candidates attending a revision course. *Postgrad Med J* 2006; 82: 145-149.
10. Alexander R, Badenhorst E, Gibbs T. Intervention programme: a supported learning programme for educationally disadvantaged students. *Med Teach* 2005; 27: 66-70.
11. Perez G, London L. Forty-five years apart - confronting the legacy of racial discrimination at the University of Cape Town. *S Afr Med J* 2004; 94: 764-770.
12. Iputo JE, Kwizera E. Problem-based learning improves the academic performance of medical students in South Africa. *Med Educ* 2005; 39: 388-393.

Accepted 7 July 2008.