

Learning in practice

Effect of ethnicity on performance in a final objective structured clinical examination: qualitative and quantitative study

Val Wass, Celia Roberts, Ron Hoogenboom, Roger Jones, Cees Van der Vleuten

Department of General Practice and Primary Care, Guy's, King's, and St Thomas's School of Medicine, London SE11 6SP

Val Wass
senior lecturer in general practice
Roger Jones
Wolfson professor of general practice

Department of Education and Professional Studies, King's College, London SE11 6SP

Celia Roberts
senior research fellow

Department of Educational Development and Research, University of Maastricht, Maastricht, Netherlands

Ron Hoogenboom
research assistant

Cees Van der Vleuten
professor of education

Correspondence to:
V Wass
valerie.wass@kcl.ac.uk

BMJ 2003;326:800-3

Abstract

Objective To assess the effect of ethnicity on student performance in stations assessing communication skills within an objective structured clinical examination.

Design Quantitative and qualitative study.

Setting A final UK clinical examination consisting of a two day objective structured clinical examination with 22 stations.

Participants 82 students from ethnic minorities and 97 white students.

Main outcome measures Mean scores for stations (quantitative) and observations made using discourse analysis on selected communication stations (qualitative).

Results Mean performance of students from ethnic minorities was significantly lower than that of white students for stations assessing communication skills on days 1 (67.0% (SD 6.8%) and 72.3% (7.6%); $P=0.001$) and 2 (65.2% (6.6%) and 69.5% (6.3%); $P=0.003$). No examples of overt discrimination were found in 309 video recordings. Transcriptions showed subtle differences in communication styles in some students from ethnic minorities who performed poorly. Examiners' assumptions about what is good communication may have contributed to differences in grading.

Conclusions There was no evidence of explicit discrimination between students from ethnic minorities and white students in the objective structured clinical examination. A small group of male students from ethnic minorities used particularly poorly rated communicative styles, and some subtle problems in assessing communication skills may have introduced bias. Tests need to reflect issues of diversity to ensure that students from ethnic minorities are not disadvantaged.

Introduction

Students from ethnic minorities seem to perform less well overall than white students in both undergraduate and postgraduate medical examinations.¹⁻⁴ Any form of potential racial discrimination within our examination systems is a cause for concern.^{5,6} Problems with complex discourse may disadvantage students in oral examinations who have been trained overseas, but

there is little further published work on the impact of differences in ethnicity on performance in examinations.⁷ This is becoming an increasingly important issue in undergraduate assessment. Fairness and consistency of assessment across UK medical schools is crucial.⁸ We need to understand any source of potential bias that may lead to racial disadvantage when developing tests for these skills.

When looking for potential discrimination within examinations, standardisation is a key issue; the more standardised the content, the less the potential for bias. Objective structured clinical examinations are currently most often used to assess undergraduate skills and include standardised simulated scenarios to test communication skills.^{9,10} Yet it is still difficult to achieve true objectivity.^{11,12} However carefully designed, the scenario presented to students will vary since neither simulated patient nor student is speaking from scripts. Examiners and simulated patients make judgments based on an impression of how well the student managed the consultation. This judgment, in turn, will be informed by their assumptions about what makes an effective consultation.¹³

We aimed to investigate whether students from ethnic minorities are disadvantaged by a bias in marking in a final year objective structured clinical examination, with a particular focus on stations assessing communication skills.

Methods

Our study took place in June 1999 during the final MBBS examination of the then Guy's and St Thomas's medical school. This comprised a three and a half hour objective structured clinical examination conducted over two days, consisting of two stations for history taking of long cases (21 minutes each) and 20 stations (seven minutes each) for clinical examination (nine stations), communication skills (six), and practical skills (five). The stations were similar but not identical on the two days. Simulated patients were professionally trained to standardise the scenarios used on communication stations.

A different examiner marked each station against a checklist and gave a final five point global rating for overall clinical competency. Simulated patients awarded a five point global rating for overall



Extracts from transcripts appear on bmj.com

Table 1 Communication scenarios

Day 1	Station 8	Station 18
Day 1	Explain to a rather obstinate, elderly, middle class white woman that her chest x ray film showed a possible lung metastasis. Bronchoscopy had been recommended. She denied that the cancer may have returned and just wanted antibiotics for her cough	Take a sexual history from a young Muslim student who had had unprotected casual sex at a party. She was concerned she might have caught something. She felt very upset about this accidental break with her cultural tradition and loss of her virginity
Day 2	Assess a Chinese businessman who has come for the results of liver function tests. These indicate he may be drinking too much (this role player was asked to present himself as not entirely fluent in English)	Negotiate with a young Afro-Caribbean man who wants a methadone prescription because he says he has lost the one given to him at the drug rehabilitation centre

communication skills, independent of the examiner. The examiners and simulated patients had been briefed on the procedure. A minimum competence score for each station was set in advance, using the Angoff standard setting method.¹⁴

Each day we selected two communication scenarios, using role players from different ethnic backgrounds (table 1). The students gave informed verbal consent for video recording the scenarios. The local research ethics committee approved our study. Details of the students' ethnicity were made available after the examination.

The students were grouped as white, south Asian (Indian, Pakistani, Bangladeshi, Chinese, and Asian other), Afro-Caribbean, and other. For the purpose of our study, all students from ethnic minorities were categorised as one group, "ethnic minorities" (82 students) and all other candidates as "white majority" (97 students).

Quantitative analysis

We analysed the mean performance for stations for each day on all 22 stations, stations grouped by communication, practical, clinical skills, and long cases, and the specific study stations. We used an independent two sample *t* test to examine relations between student performance and ethnicity. We regarded *P* values greater than 0.01 as non-significant. The reliability of each objective structured clinical examination was calculated with Cronbach's α .

Qualitative analysis

All video recorded encounters were viewed as well as recorded comments made by simulated patients and examiners after students had left the station. The duration of the interaction, student's ethnicity, and observations made during the viewing were recorded on a standard form. An assessment was made of the extent to which simulated patients and students established a relatively patient centred encounter. Any potential misunderstandings, false assumptions, or explicit discriminatory behaviour were noted. The analyst (CR) viewed the encounters "blind," allocating grades and comparing these after the viewing with the ratings from the simulated patients and examiners. Discrepancies between analyst, simulated patient, and examiner were recorded.

We used these records to select specific interactions for detailed transcription. We identified recurring themes, which acted as background to the detailed discourse analysis. We used these to clarify the complexity of the doctor-patient consultation and the communicative demands it placed on students.¹⁵

Results

We excluded four of the 179 students, as their ethnicity was undeclared. Table 2 gives the ethnicity, sex, and age

Table 2 Analysis of ethnicity by sex and age for 175 students of known ethnicity

Ethnicity	Male	Female	Mean age (years)	Age range	Total (%)
White	50	47	24.6	23-32	97 (55)
South Asian*	31	27	23.0	22-30	58 (33)
Afro-Caribbean	1	4	25.6	23-32	5 (3)
Other	6	9	24.2	23-29	15 (9)
Total	88	87	24.4	22-32	175 (100)

*Indian, Pakistani, Bangladeshi, Chinese, and Asian other.

of the remaining 175 students. Seventy eight (45%) were from ethnic minorities. All but two students had received secondary school education in the United Kingdom.

Table 3 shows the mean performance of the students in the overall examination and in the specific study stations. Mean performance on communication stations was significantly higher for white majority students than for students from ethnic minorities on both day 1 (72.3% (SD 7.6%) 67.0% (6.8%); $P=0.001$) and day 2 (69.5% (6.3%) and 65.2% (6.6%); $P=0.003$). The Cronbach α reliability of the objective structured clinical examination was 0.74 and 0.76 on days 1 and 2, respectively.

We were unable to assess 49 (14%) of the video recorded interactions due to technical faults. We found no explicit examples of breakdown in communication or of discriminatory behaviour in the remaining 309

Table 3 Mean scores, standard deviations, and T and P values (independent two sample *t* test) for students from white majority compared with students from ethnic minorities for components of objective structured clinical examination on days 1 and 2

Exam component (No of stations)	White majority group		Ethnic minority group		Comparison of means*	
	No of students	Mean (SD) score (%)	No of students	Mean (SD) score (%)	T value	P value
Total examination (22):						
Day 1	49	74.1 (4.3)	35	70.6 (5.1)	3.43	0.001
Day 2	48	72.5 (4.6)	43	70.0 (4.3)	2.70	0.008
Communication (6):						
Day 1	49	72.3 (7.6)	35	67.0 (6.8)	3.31	0.001
Day 2	48	69.5 (6.3)	43	65.2 (6.6)	3.08	0.003
Clinical (9):						
Day 1	49	72.3 (5.5)	35	70.7 (5.1)	1.33	0.190
Day 2	48	72.1 (5.0)	43	70.0 (4.5)	2.19	0.030
Practical (5):						
Day 1	49	79.5 (5.7)	35	74.5 (7.5)	3.45	0.001
Day 2	48	76.7 (6.9)	43	75.1 (7.6)	1.07	0.290
Long cases (2):						
Day 1	49	74.1 (7.6)	35	70.8 (8.7)	1.86	0.066
Day 2	48	72.7 (7.5)	43	71.3 (7.3)	0.92	0.359
Station 8:						
Day 1	49	67.8 (12.7)	35	63.8 (13.9)	1.38	0.172
Day 2	48	65.2 (11.1)	43	65.1 (7.8)	0.07	0.947
Station 18:						
Day 1	49	77.4 (10.3)	35	68.1 (16.1)	3.22	0.002
Day 2	48	72.6 (13.4)	43	64.5 (16.4)	2.55	0.012

*Independent two sample *t* test.

interactions. Neither simulated patients nor students showed, through talk or bodily movements, any expression identifiable as a negative response to the other's ethnicity.

For detailed discourse analysis we transcribed 28 (9%) interactions, representing a range of scores from high to low and including students from both groups. Two main findings emerged.

Firstly, students created different interactional climates. Those receiving high grades were relatively empathetic, responsive, and persuasive, building a joint problem solving framework with the patient. Conversely, some failed to build this framework, displayed various moves to distance themselves from patients, and were given low grades by both examiners and simulated patients.¹⁶ Students from both groups failed to create this interactive framework (see box and [bmj.com](#)). Relatively more male students from the ethnic minority group were in this category: Fifteen (12 male) of the 22 students scoring below minimum competence were in the ethnic minority group. In these instances there were no obvious cultural and linguistic differences, although these students were more likely to have pronunciation, word stress, and intonation influenced by their heritage language.

Secondly, there were instances where the examiner gave top marks but the simulated patient from an ethnic minority gave a lower mark. These students tended to use a style in which explicit guidance was deferred, there was more talk about the nature of the consultation, and there was more talk about talk—for example, “But first I'd like to know a little more about you” (see [bmj.com](#)). Although this style fitted well with the white examiners' textbook notions of a patient centred consultation, it was not rated so highly by some simulated patients from ethnic minorities.

Discussion

We found differences in the mean performance of students from ethnic minorities on communication

Framework for good and poor communicative styles

The framework for communicative style consisted of four levels:

Performance factors—these included clarity, slips of the tongue, hesitations, voice quality, and aspects of non-verbal communication

The design of questions and responses—for example, the ways in which students showed that the patient's problem needed to be jointly managed, or the ways in which students were sensitive to the needs of the patient; or by contrast the negative labelling of the student or the use of “trained” empathy

The overall thematic staging of the consultation—for example, a student who had to resist giving a methadone prescription to a drug addict managed the consultation so as not to either give in or refuse too early on; or, by contrast, a student shifting rapidly from one topic to another and preventing the simulated patient from following the student's line of reasoning

Ideological positioning of the student—for example, how much to rely on personal authority and how much on medical authority

What is already known on this topic

UK medical schools are concerned that students from ethnic minorities may perform less well than white students in examinations

It is important to understand whether our examination system disadvantages them

What this study adds

Mean performance of students from ethnic minorities was significantly lower than that of white students in a final year objective structured clinical examination

Two possible reasons for the difference were poor communicative performance of a small group of male students from ethnic minorities and examiners' use of a textbook patient centred notion of good communication

Issues of diversity in test construction and implementation must be addressed to ensure that students from ethnic minorities are not disadvantaged

stations in the final year objective structured clinical examination. Although we saw no obvious evidence of breakdowns in communication or of discriminatory judgments, two subtle differences emerged: a particularly poor communicative style that may have distanced some students from ethnic minorities from the simulated patient¹⁶ and instances where the examiner's assumptions did not match the expectations of the simulated patients.

We do not want to claim too much from these two points: it would be wrong to reify ethnicity and assume problems and differences in communicative style on the basis of an individual's ethnicity alone. Similarly, our comments on the differences between examiners and the simulated patients from ethnic minorities are speculative and need further investigation. In combination, however, these two factors may account for at least some of the differences in the ratings of white students and students from ethnic minorities.

The style of some students to distance themselves from patients reflects a medical model of consultation rather than a more social one preferred by examiners. Students from ethnic minorities might be more likely than white students to use this style because a medical model is less demanding of communication skills and may be perceived as appropriate. In this case, differences in motivation and learning styles also need to be considered.¹⁷ So several complex factors, styles of communication, values, and ways of learning may all be important and may be related to the ways in which students are socialised into medical school culture.

Students who live outside the medical school and have networks of family and friends where there are quite different communicative experiences from those in the university may be less exposed to the informal and social talk around medicine occurring in institutional life.^{18 19} They may therefore have less

opportunity to tune into the current institutional norms about what counts as a good consultation.

There is a case for developing a wider repertoire of communicative styles for setting the stations. Important questions have been raised for educators in cross cultural communication in medicine, but this needs to be addressed at the local level—for example, by examiners working with a range of non-traditional students and simulated patients from ethnic minorities to build up a wider repertoire of styles sensitive to the diverse backgrounds of patients.¹³ Institutions may also not be aware of hidden processes that reward some students and penalise others in final examinations.

We thank Professor Gwyn Williams and Dr Charles Twort for approving this study, Stevo Durbaba for his technical assistance, Phil Doulton of Professional Role Play, Nora Edmead for her administrative support, and the students and examiners for their cooperation.

Funding: The King's Fund.

Competing interests: None declared.

- 1 McManus IC, Richards P, Winder BC, Sproston KA. Final examination performance of medical students from different ethnic minorities. *Med Educ* 1996;30:195-200.
- 2 Dillner L. Manchester tackles failure rate of Asian students. *BMJ* 1995;310:209.
- 3 Wakeford R, Farooqi A, Rashid A, Southgate L. Does the MRCGP examination discriminate against Asian doctors? *BMJ* 1992;305:92-4.
- 4 Jolly BC, Cohen R, Rothman AI, Ross J. Graduates of foreign medical schools: demographic and personal predictors of success on an OSCE-format internship programme entrance examination. In: *Proceed-*

ings of the 27th annual conference of the Association of American Medical Colleges. Washington, DC: Association of American Medical Colleges, 1988:234-9.

- 5 Smith R. Prejudice against doctors and students from ethnic minorities. [Editorial.] *BMJ* 1987;294:328-9.
- 6 McKenzie KJ. Racial discrimination in medicine. [Editorial.] *BMJ* 1995;310:478-9.
- 7 Roberts C, Sarangi S, Southgate L, Wakeford R, Wass V. Oral examinations—equal opportunities, ethnicity, and fairness in the MRCGP. *BMJ* 2000;320:370-4.
- 8 Catto G. Education, education, education. *BMJ Classified* 23 Jun, 2001:2-3.
- 9 Fowell SL, Maudsley G, Maguire P, Leinster SJ, Bligh J. Student assessment in undergraduate medical education in the United Kingdom. *Med Educ* 2000;34(suppl 1):1-49.
- 10 Wass V, Vleuten C van der, Shatzer J, Jones R. Assessment of clinical competence. *Lancet* 2001;357:945-9.
- 11 Vleuten CPM van der, Swanson DB. Assessment of clinical skills with standardised patients: state of the art. *Teach Learn Med* 1990;2:58-76.
- 12 Colliver JA, Verhulst SJ, Williams RG, Norcini JJ. Reliability of performance on standardised patient cases: a comparison of consistency measures based on generalizability theory. *Teach Learn Med* 1989;1:31-7.
- 13 Skelton JR, Kai J, Loudon RF. Cross-cultural communication in medicine: questions for educators. *Med Educ* 2001;35:257-61.
- 14 Angoff WH. Scales, norms, and equivalent scores. In: Thorndike RL, ed. *Educational measurement*, 2 ed. Washington, DC: American Council on Education.
- 15 Erickson F, Shultz J. *The counsellor as gatekeeper: social interaction in interviews*. London: Academic Press, 1982.
- 16 Roberts C, Wass V, Jones R, Sarangi S, Gillett A. A discourse analysis study of "good" and "poor" communication in an OSCE: a proposed new framework for teaching students. *Med Educ* 2003 (in press).
- 17 McManus IC, Richards P, Winder BC, Sproston KA. Clinical experience, performance in final examinations, and learning style in medical students: prospective study. *BMJ* 1998;316:345-50.
- 18 Gumperz, J. A discussion with John Gumperz. In: Eerdman S, Previgagn C, Thibault P, eds. *Discussing communication analysis 1 John Gumperz*. Lausanne: Beta Press, 1997.
- 19 Atkinson P. *Medical talk and medical work: the liturgy of the clinic*. London: Sage, 1995.

(Accepted 20 November 2002)

My most unfortunate mistake

The heights of folly

My house jobs had finished, and I had just arrived in Huaraz, northern Peru, for a trek around Mount Huascarán with Andy, my fellow senior house officer and flatmate from medical school.

After 72 hours of acclimatisation at an altitude of 3500 m, we set off for the five day adventure. As dawn broke, we started to ascend, my only symptom being a mild headache. After four hours of climbing the near vertical mountain footpath, the headache had matured into a persistent thump with each heartbeat. Further water and painkillers weren't having much effect. My pace had slowed, much to my annoyance, and we were still some distance from the first pass at a height of 4900 m. My shortness of breath was becoming increasingly apparent, and I was unable to progress more than 10 paces at time. We both knew the effects of altitude were becoming more evident.

However, the top of the pass was now in view, and, after a further two hours of painstakingly slow progress, we reached the top, where I proceeded to vomit. After 15 minutes of retching, I was able to take in the breathtaking view, but a rush of diarrhoea swiftly curtailed any further opportunity for admiring the scenery. I needed to drop altitude, so we proceeded over the pass and down the mountain, away from civilisation. After a drop in altitude of about 500 m and multiple stops for retching, ataxia developed. I was by now a hopeless wreck, incapable of standing let alone walking, retching, while my headache had now reached a hammer-like quality with each heartbeat. At this point all I wanted was for the suffering to end.

After making a crude shelter for me, Andy twice set back up the mountain to find help, and at the second ascent he managed to intercept a truck and persuade the driver to divert to where I was camped. "Ah la soroche [mountain sickness]," stated the driver and gestured that I needed to drink a small bottle of his

aftershave, which I dutifully did. I was then placed in the back of his truck and was driven down to a collection of houses at about 3500 m, where I found the vomiting had started to abate. We stayed there for the night, during which I had supraventricular tachycardia and Cheyne-Stoke's respiration. The next day I was placed on a truck and taken down to sea level for recuperation.

By not descending when the symptoms started to develop, we had failed to acknowledge the fundamental importance of recognising and acting on the clinical signs of mountain sickness. Undoubtedly I owe my life to two people: firstly, Andy, who, after his experience in managing a difficult hypoxic patient, has become an anaesthetist; and, secondly, the truck driver, for rapidly transporting me to lower altitudes. Ingestion of aftershave probably had no effect on my recovery, and alcohol would not be recommended for treating patients with symptoms of mountain sickness.

For my part, the take home message is clear: when travelling in the mountains, listen to your head, or, failing that, at least travel with an anaesthetist.

Jonathan C R Bowling *Chelsea and Westminster Hospital, London*

Andrew John Lee *Derriford Hospital, Plymouth*

We welcome articles up to 600 words on topics such as *A memorable patient, A paper that changed my practice, My most unfortunate mistake*, or any other piece conveying instruction, pathos, or humour. If possible the article should be supplied on a disk. Permission is needed from the patient or a relative if an identifiable patient is referred to. We also welcome contributions for "Endpieces," consisting of quotations of up to 80 words (but most are considerably shorter) from any source, ancient or modern, which have appealed to the reader.