The importance of race in medical student performance of an AIDS risk assessment interview with simulated patients

Andrew J Zweifler,[1] *Fredric M Wolf,*[2] *Mary S Oh,*[3] *J T Fitzgerald*[3] *& Laurie L Hengstebeck*[4]

Introduction This study was done to ascertain the effect of race on medical student–patient communication. The primary hypothesis was that interviewing performance scores would be higher when race of student and race of simulated patient instructor (SPI) were concordant than when they were discordant.

Methods Data obtained from student interactions with four Caucasian and four African American female SPIs participating in a case involving an AIDS risk assessment interview were analysed. Performance was assessed using two instruments: an 8-item behaviourally anchored interviewing skills scale and a 14-item checklist reflecting content relevant to sexual behaviour and AIDS risk. SPI groups were comparable and SPIs were trained to a high level of inter-rater reliability. Students (24 African American and 180 non-African American) were assigned to SPIs based on the spelling of the student's name. Performance was scored independently at the conclusion of each interview by both the SPI and the student her/himself.

Results African American students had lower scale scores than non-African American students in interactions with Caucasian (but not African American) SPIs; and all student scores, both on the skills scale and the content checklist, were higher with African American than with Caucasian SPIs (as assessed by both SPI and student). Women students had higher scores than men.

Conclusions Race of SPI has an influence on student performance of an AIDS risk assessment interview. Further studies focusing on racial interactions in the medical interview are required. It appears that race of SPI may need to be accounted for in the development and interpretation of SPI-based clinical competence exams.

Keywords AIDS, *aetiology; blacks; education, medical undergraduate; interviews; physician–patient relationship; *racial stocks; risk assessment; sexual behaviour; United States; whites.

Medical Education 2000;34:175–181

Introduction

Numerous studies have documented disparity in health status and health care between African Americans and Caucasians in the United States.[1] Among the explanations proposed to explain this phenomenon is that there are deficiencies in the quality of the relationship between doctor and patient when the patient is a member of a minority group.[2] Certainly there are substantial dissimilarities between members of varying racial groups which may impact the doctor–patient relationship.[3,4] Prominent among these are socio-economic, linguistic and cultural differences; however racism, the tendency of the members of the majority group to devalue members of minority groups, is logically another important contributor. Documentation and quantification of the impact of race in cross-racial doctor–patient interaction have considerable relevance to medical education, both from the perspective of curriculum development and that of the need for unbiased evaluation of clinical performance.

To analyse the effect of race on medical student–patient communication we used data obtained from a simulated patient instructor (SPI)-based exercise on AIDS risk assessment for second-year medical students at the University of Michigan Medical School.[5] Our hypothesis was that student performance, as assessed by both the SPIs and the students themselves, would be influenced by the interaction between race of student and race of SPI, and specifically: (i) that performance scores

[1]Department of Internal Medicine, University of Michigan Medical School, Ann Arbor, Michigan, USA
[2]Current affiliation: Department of Medical Education, University of Washington School of Medicine, Seattle, Washington, USA
[3]Department of Postgraduate Medicine, University of Michigan Medical School, Ann Arbor, Michigan, USA
[4]Medical School Administration, University of Michigan Medical School, Ann Arbor, Michigan, USA

Correspondence: Andrew J. Zweifler, 3918 Taubman Center, University of Michigan Medical School, Ann Arbor, Michigan 48109–0356, USA

would be higher when race of student and SPI were concordant than when they were discordant, and (ii) that correlations between student self-assessment and SPI assessment would be stronger when race of student and SPI were concordant than when they were discordant.

Data and methods

Data from student interactions with four Caucasian and four African American female SPIs portraying the same case over a period of three academic years (1993–95) were analysed.

Instruments

A behaviourally anchored scale and a content checklist to evaluate interviewing performance were utilized. These were modelled on instruments developed initially at the University of Arizona[6] and modified at the University of Rochester to evaluate preventively oriented simulated patient instructor interactions.[7] The instruments were designed to assess skills emphasised in Michigan's interviewing course for freshmen medical students and to reflect content relevant to AIDS risk and the sexual history.

Interviewing skills evaluated included: opening the interview; organizing the interview; probing and focusing; verifying data; listening and responding accurately (empathy); avoidance of projection of values; non-verbal communication; and closing the interview. Each item was scored on a 5-point behaviourally anchored scale with performance rated from poor (1·0) to excellent (5·0).

The content checklist was constructed with the assistance of experts in AIDS and infectious diseases, and included items related to sexual partners, sexual orientation, sexual behaviours and use of self-injected drugs. The 14 content items were rated on a 3-point scale (0 = not covered; 0·5 = partially covered; 1·0 = fully covered).

Simulated patient instructors (SPIs)

Simulated patients were recruited by advertisement and selected based on personal evaluation by an experienced trainer (LLH). SPIs ranged in age from 22 to 30 years. All had some college-level education. Three of the African American women and two of the Caucasian women were currently enrolled in university courses.

Each woman underwent 40–60 hours of training during which time each learned: (i) to portray the case (a woman with a history of having had unprotected sexual intercourse during a relationship with a drug-abusing man); (ii) how to evaluate student performance utilizing both the 8-item interview scale and the 14-item content checklist, and (iii) how to provide constructive feedback to students following the interview. Inter-rater reliability of SPIs was monitored during the training period utilizing a set of videotaped SPI interviews, with kappa coefficients previously having been demonstrated to exceed 0·80.[8]

Students

Second-year medical students were assigned to SPIs based on the alphabetical position of the student's name on a class list. Data were available for 204 students, 24 of whom were African American. There were 127 men and 77 women in this sample of students.

Curricular materials and syllabus

Reading material concerning how to obtain a reliable sexual history and the relevance of this information to the detection of risk for AIDS had been assembled.[9,10] This material, as well as copies of the interviewing scale and content checklist which would be used to evaluate student performance, were placed in a small syllabus which each student received at least 10 days prior to her/his scheduled session with the SPI.

Sessions

Students were ushered into an examination room by a coordinator and given approximately 20 minutes to conduct the interview. Following the interview, both the student and the SPI filled out identical evaluation instruments. They then met and the SPI provided item-by-item feedback to the student on his/her performance. The entire session was approximately 45 min in duration.

Data analysis

Score reliabilities for the behavioural interview scale and the content checklist were calculated using Cronbach's coefficient alpha.[11] Reliability refers to the consistency of the responses to the individual items within a multi-item scale. For each student, overall mean scores and standard deviations for the SPI and student's own assessments for the skills scale and the content checklist were derived by averaging ratings over the individual items.

Three-way analyses of variance of the mean ratings were used to investigate interactions among the three main demographic categories: SPI race (African American and non-African American), student race (African American and non-African American), and

	Non-African American students	African American students	Total
African-American SPIs			
Women students	4·24 ± 0·40 (35)	4·63 ± 0·45 (9)	4·32 ± 0·44 (44)
Men students	4·17 ± 0·46 (59)	3·95 ± 0·76 (5)	4·15 ± 0·48 (64)
Total students	4·19 ± 0·44 (94)§	4·38 ± 0·65 (14)§	4·22 ± 0·47 (108)*
Caucasian SPIs			
Women students	3·88 ± 0·80 (28)	3·30 ± 0·63 (5)	3·80 ± 0·80 (33)
Men students	3·44 ± 0·80 (58)	2·78 ± 0·19 (5)	3·39 ± 0·79 (63)
Total students	3·59 ± 0·82 (86)§	3·04 ± 0·52 (10)§	3·53 ± 0·81 (96)*
Total SPIs			
Women students	4·08 ± 0·63 (63)	4·15 ± 0·83 (14)	4·09 ± 0·67 (77)†
Men students	3·81 ± 0·74 (117)	3·36 ± 0·81 (10)	3·77 ± 0·75 (127)†
Total students	3·90 ± 0·72 (180)	3·82 ± 0·90 (24)	3·89 ± 0·74 (204)

Table 1 Simulated patient instructor (SPI) evaluations (scores) of students for interviewing skills scale items; data are mean ± SD followed by number of students in parentheses

*Significant main effect for SPI ethnicity, $F_{1,196} = 58·75$, $P < 0·001$.
†Significant main effect for student gender, $F_{1,196} = 10·31$, $P = 0·002$.
§Significant interaction for SPI ethnicity and student ethnicity, $F_{1,196} = 6·70$, $P = 0·01$.

student gender. Simple main effects[12] were calculated following determination of any significant interaction effects. Stepwise regressions of the overall means of the behavioural scale and the checklist were performed to investigate the contributions of various dichotomous representations of demographic features for students, SPIs, and similarity or difference of race between the students and the instructors. Student's *t* test was used to compare interviewing scale mean ratings by student race within SPI racial categories. Pairwise *t* tests were used to compare SPI ratings and student self-assessments by student gender.

Correlations between SPI and student assessments were calculated; the correlations of student self-assessment and SPI assessment by student race were compared within SPI racial categories using Fisher's *Z* transformation.

Results

Interview scale

SPI assessment
The scale is a reliable measure; the coefficient alphas[11] were 0·89 for the SPI evaluations and 0·83 for student self-evaluations. Student performance, using the mean score per SPI assessment on all scale items as a criterion (Table 1), was good (Table 1). For all students the mean score was 3·89 ± 0·74. Total scores for African American and non-African American students were similar, but women students scored significantly higher than their male counterparts (4·09 ± 0·67 versus 3·77 ± 0·75, $P = 0·002$). Scores with African American

SPIs were substantially higher than with Caucasian SPIs (4·22 ± 0·47 versus 3·53 ± 0·81, $P < 0·001$).

A significant ($P = 0·01$) interaction between race of student and race of SPI was observed (Table 1). Scores for African American students were lower than those for non-African American students in interviews with Caucasian SPIs (3·04 ± 0·52 versus 3·59 ± 0·82), while the scores for interactions with African American SPIs were almost equal (4·38 ± 0·65 versus 4·19 ± 0·44). The racial difference in student performance with Caucasian SPIs was statistically significant ($t = 2·06$, $P = 0·043$), while the racial difference with African American SPIs was not ($t = -1·06$, $P = 0·305$) (Fig. 1).

Stepwise multiple regression analysis, using various demographic variables including race and gender of students as well as race of SPI, revealed race of SPI to be the most powerful predictor of student performance (Table 2). Also important were concordant race of student and SPI (whether both African American or non-African American), and student gender (female). In this analysis, student race (African American) was also a predictor (negative) of scale scores.

Student self-assessment
Overall student performance, using the mean score per self-evaluation on all scale items as a criterion (Table 3), was very close to that reported by SPIs (3·91 ± 0·51 versus 3·89 ± 0·74). The smaller variability as compared with the SPI ratings in self-evaluated student scores is probably related to the fact that students are much less experienced as evaluators and tend to avoid grading on either extreme of the Likert scale. Scores for African American and non-African
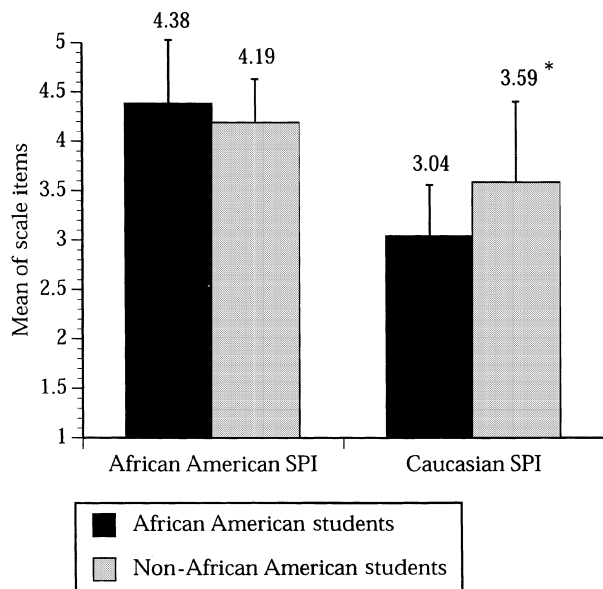
Figure 1 Scale scores (mean ± SD) for African American students ($n = 24$) and non-African American students ($n = 180$) by simulated patient instructor (SPI) ethnicity. For the Caucasian SPIs, a significant difference was indicated between the African American and non-African American students, ⋆$P = 0.04$.

American students were not different, nor were scores for men and women students.

As a group, students reported higher scores with African American SPIs than with Caucasian SPIs (4.05 ± 0.50 versus 3.76 ± 0.48). Race of SPI was the only predictor of student performance (beta $= 0.290$, $P = 0.000$) by stepwise multiple regression analysis. No significant interaction between race of student and race of SPI was observed.

Using SPI scores as a reference standard, women tended to underestimate their performance (4.08 ± 0.67 versus 3.94 ± 0.50; $t = 2.04$, $P = 0.045$) while men tended to overestimate their scores (3.76 ± 0.76 vs. 3.90 ± 0.52; $t = -2.12$, $P = 0.036$).

Finally, correlation of SPI-assessed versus student (self-assessed) performance was similar for Caucasian (0.4065) and African American (0.3654) SPIs, and there was no significant difference in correlation of SPI-assessed versus student (self-assessed) performance related to concordance or discordance of race of student and SPI (Table 4).

Content checklist

For the checklist, the alpha for the SPI evaluations was 0.73 and for the students' evaluations it was 0.66. Student performance using the mean score on all checklist items as a criterion was good, both by SPI assessment (Table 5) and by student self-assessment (Table 6). Significantly higher scores were reported by SPIs as well as students in interactions with African American SPIs as opposed to Caucasian SPIs. Women students had higher scores than men per SPI assessment but not by student self-assessment. No significant race or gender interactions were demonstrable.

Discussion

There are two major findings in this study of student performance on an AIDS risk assessment interview with simulated patient instructors (SPIs) of differing race. Based on analysis of interview scale data they are as follows: (i) there is a significant interaction between race of student and race of SPI, and (ii) all students, regardless of race, achieve higher scores with African American SPIs than with Caucasian SPIs. Although checklist scores suggest a similar interaction effect, this finding is not statistically significant; however, the data do reveal significantly higher scores for student interviews with African American as opposed to Caucasian SPIs. An additional, but previously reported observation,[5] is that female students achieve higher scores in this exercise than male students.

Variable	b Estimate	Standardized ß	t	P
Intercept	3.25	0.000	34.4	<0.001
SPI Ethnicity (Non-African American = 0, African American = 1)	0.851	0.578	8.27	<0.001
SPI/Student same ethnicity (Different = 0, Same = 1)	0.355	0.231	3.27	0.001
Student Gender (Male = 0, Female = 1)	0.275	0.181	3.00	0.003
Student ethnicity (Non-African American = 0, African American = 1)	−0.288	−0.126	−2.05	0.042

Table 2 Results of stepwise regression to analysis predict SPI mean scale rating $n = 203$

Dependent variable = SPI Scale mean, $R^2 = 0.296$, $P = 0.000$

	Non-African American students	African American students	Total
African American SPIs			
Women students	4·00 ± 0·44 (34)	4·22 ± 0·45 (9)	4·05 ± 0·44 (43)
Men students	4·05 ± 0·54 (58)	4·13 ± 0·64 (5)	4·06 ± 0·54 (63)
Total students	4·03 ± 0·50 (92)	4·19 ± 0·50 (14)	4·05 ± 0·50 (106)★
Caucasian SPIs			
Women students	3·77 ± 0·56 (27)	3·93 ± 0·50 (5)	3·80 ± 0·54 (32)
Men students	3·75 ± 0·46 (57)	3·65 ± 0·24 (5)	3·74 ± 0·45 (62)
Total students	3·75 ± 0·49 (84)	3·79 ± 0·40 (10)	3·76 ± 0·48 (94)★
Total SPIs			
Women students	3·90 ± 0·50 (61)	4·12 ± 0·47 (14)	3·94 ± 0·50 (75)
Men students	3·90 ± 0·52 (115)	3·89 ± 0·52 (10)	3·90 ± 0·52 (125)
Total students	3·90 ± 0·51 (176)	4·02 ± 0·50 (24)	3·91 ± 0·51 (200)

Table 3 Student self-evaluations (scores) for interviewing skills scale items; data are mean ± SD followed by number of students in parentheses

★ Significant main effect for SPI ethnicity, $F_{1,192} = 17.30$, $P < 0.001$.

Table 4 Correlations of SPI evaluations and student self-evaluations on interviewing skills scale items

	African American students	Non-African American students	P★
African American SPIs	0·37	0·38	0·99
Caucasian SPIs	0·68	0·41	0·30

★ Fisher's Z transformation.
Note: There are no statistically significant differences between any two correlations in this Table.

The study was designed primarily to test the hypothesis that there is an interaction between race of interviewer and race of respondent in a medical interview. This proved to be a valid hypothesis, but only in interviews with Caucasian simulated patients: African American students had substantially lower scores in these interactions than did non-African American students. On the other hand non-African American students did not get lower scores in interactions with African American SPIs. In fact both African American and non-African American students achieved relatively high scores in this interaction with an African American 'patient.' The explanation of this latter finding is not clear, and it should be interpreted with caution, given the limitations of our study: (a) the SPIs are a very small sample from which to generalize, all being relatively highly educated, female, and relatively young, and (b) the findings may be specific to the particular case which was utilized (AIDS risk assessment interview). Additional observations utilizing different cases, different tasks (e.g. provision of patient education) and a larger sample of SPIs, including men and individuals of varying educational attainment and socio-economic status, would be useful in determining the generalizability of our findings.

Our observation of the importance of race in student interactions with some 'patients' is consistent with evidence that differences in race as well as gender, ethnicity and socio-economic status may be barriers to medical communication.[3,13,14] Also it is quite clear from the survey research literature that differences in race of interviewer and respondent have a definite influence on the quantity and quality of responses to certain types of survey questions, particularly those that relate to opinions rather than facts.[15–17] In attempting to understand the finding that African American students get lower scores in interviews with Caucasian SPIs than do their classmates, there are at least two possible explanations. African American students may be especially challenged by the nature of this particular case; they may be intimidated by the requirement that they ask questions of a sexual nature of a white woman. Alternatively, it is possible that Caucasian SPI assessment of performance is biased against African American students. This is not likely given the extensive training which our SPIs receive, but it is conceivable. Further studies could be designed to shed light on this finding. For example, it would be of interest to conduct studies using the same case with SPIs who have been trained with videotapes of interactions between both African American and Caucasian students (to date no effort has been made to include interviews by students of different races in the SPI training programme at our institution).

The finding that all students, regardless of race, achieved higher scores in interviews with African American SPIs than with Caucasian SPIs was unexpected. Race of SPI proved to be the most powerful predictor of student performance, statistically more

	Non-African American students Mean ± SD (n)	African American students Mean ± SD (n)	Total students Mean ± SD (n)
African American SPIs			
Female students	0.89 ± 0.11 (35)	0.90 ± 0.08 (9)	0.89 ± 0.10 (44)
Male students	0.88 ± 0.11 (59)	0.85 ± 0.12 (5)	0.87 ± 0.11 (64)
Total students	0.88 ± 0.11 (94)	0.89 ± 0.10 (14)	0.88 ± 0.11 (108)*
Caucasian SPIs			
Female students	0.83 ± 0.15 (28)	0.79 ± 0.10 (5)	0.83 ± 0.15 (33)
Male students	0.76 ± 0.16 (58)	0.63 ± 0.12 (5)	0.75 ± 0.16 (63)
Total students	0.78 ± 0.16 (86)	0.71 ± 0.14 (10)	0.77 ± 0.16 (96)*
Total SPIs			
Female students	0.86 ± 0.13 (63)	0.86 ± 0.11 (14)	0.86 ± 0.13 (77)†
Male students	0.82 ± 0.15 (117)	0.74 ± 0.17 (10)	0.81 ± 0.15 (127)†
Total students	0.83 ± 0.15 (180)	0.81 ± 0.14 (24)	0.83 ± 0.15 (204)

Table 5 SPI evaluations (scores) of students for checklist items

*Significant main effect for SPI ethnicity, $F_{1, 196} = 31.06$, $P < 0.001$.
†Significant main effect for student gender, $F_{1, 196} = 6.38$, $P = 0.012$.

	Non-African American students Mean ± SD (n)	African American students Mean ± SD (n)	Total students Mean ± SD (n)
African American SPIs			
Female students	0.91 ± 0.08 (34)	0.91 ± 0.07 (9)	0.91 ± 0.08 (43)
Male students	0.87 ± 0.10 (59)	0.91 ± 0.09 (5)	0.88 ± 0.10 (64)
Total students	0.89 ± 0.10 (93)	0.91 ± 0.08 (14)	0.89 ± 0.10 (107)*
Caucasian SPIs			
Female students	0.81 ± 0.15 (28)	0.82 ± 0.09 (5)	0.81 ± 0.14 (33)
Male students	0.79 ± 0.13 (55)	0.71 ± 0.18 (5)	0.78 ± 0.13 (60)
Total students	0.80 ± 0.13 (83)	0.77 ± 0.15 (10)	0.79 ± 0.14 (93)*
Total SPIs			
Female students	0.86 ± 0.13 (62)	0.88 ± 0.09 (14)	0.87 ± 0.12 (76)
Male students	0.83 ± 0.12 (114)	0.81 ± 0.17 (10)	0.83 ± 0.13 (124)
Total students	0.84 ± 0.12 (176)	0.85 ± 0.13 (24)	0.84 ± 0.13 (200)

Table 6 Student self-evaluations (scores) for checklist items

* Significant main effect for SPI ethnicity, $F_{1, 192} = 33.81$, $P < 0.001$.

important than gender of student (which had been previously reported), and it was shown to be consistently present on analysis of both the interview scale data and the checklist data. At least two possible explanations need to be considered: (i) African American SPIs systematically grade students higher than Caucasian SPIs; (ii) students perform better with African American than with Caucasian SPIs. The first hypothesis is supported by the fact that there is less variability in the scores of non-African American students reported by African American SPIs than by Caucasian SPIs (Table 1) and it is plausible, since there is evidence that African Americans experience poorer health care and have lower expectations of health care providers than do Caucasians.[18] Because of this,

African American SPIs may have had lower initial expectations than their Caucasian counterparts when judging the performance of students in this exercise. There are two strong arguments against this hypothesis, however. First, student self-assessment results were similar to SPI results; students independently reported that they performed better with African American SPIs than with Caucasian SPIs, thus providing an internal validation of SPI-reported scores. Second, all SPIs were trained in the same manner, by the same trainer, to a comparable level of inter-rater reliability.

The second hypothesis may provide a better explanation of the finding that students in general achieved higher scores with African American SPIs. It is possible that students simply perform better in interviews with

African American SPIs than they do with Caucasian SPIs. This might occur because African American SPIs, in fact, behave differently from Caucasian SPIs, so that the interaction is a more comfortable one for students and student performance is enhanced. The nature of the specific case involved in this study may also be an important factor: (i) the case may be more 'real' to students when portrayed by an African American person, given the fact that AIDS is relatively more prevalent among African Americans; (ii) the case requires questioning about sexual practices which might seem more appropriate to students because of stereotypical assumptions about normative behaviour among African Americans as compared with Caucasians.

There may be substantial implications for medical education in the findings of this study, First, our data support the belief that race matters in doctor–patient interactions, and argue for inclusion of cross-racial SPI exercises in medical school curricula. As our society continues to diversify, the need for medical students to receive guidance in the clinical approach to patients of differing race, ethnicity and socio-economic status becomes more urgent. Our data indicate that this is true for students from both minority and majority groups. Perhaps of even greater importance is our finding that race of SPI is a powerful predictor of student performance in simulated patient-based exercises. Assuming that this observation is validated by additional research using different cases and a greater variety of SPIs, race of SPI will need to be recognized as a potentially confounding variable in the performance of subjects being evaluated by SPIs. Some of the reported variability in SPI-based examinations[19] may be related to this factor. Controlling for race of SPI may result in reduction of this variability, which could allow alteration of assumptions that have been made about the number of SPI interactions needed in order to obtain reliable results in such multistation examinations.[19] Finally, it appears that race of SPI will need to be considered in the development of clinical competence examinations which rely on SPIs for assessments. Our data suggest that African American examinees would be significantly disadvantaged if all SPIs in a multistation exam were Caucasian. Since the United States Medical Licensing Exam (USMLE) and Educational Commission for Foreign Medical Graduates (ECFMG) are moving in the direction of including SPIs in their examinations, this issue will need to be addressed in the near future.

Acknowledgements

References

1 Geiger HJ. Race and health care – an American dilemma? [editorial]. *New Engl J Med* 1996;**335**:815–6.

2 Nickens HW. A compelling research agenda [editorial]. *Ann Intern Med* 1996;**125**:237–8.

3 Kleinman A, Eisenberg L, Good B. Culture, illness and care: clinical lessons from anthropological and cross cultural research. *Ann Intern Med* 1978;**89**:251–8.

4 Quesada GM. Language and communication barriers for health delivery to a minority group. *Soc Sci Med* 1976;**10**:323–7.

5 Zweifler AJ, Dicken LL, Wolf FM. The importance of gender in the performance of an AIDS risk assessment. In: AR Rothman, R Cohen, eds. *The Sixth Ottawa Conference on Medical Education.* Toronto: University of Toronto Bookstore Custom Publishing; 1994: p. 460–2.

6 Stillman PL, Brown DR, Redfield DL, Sabers DL. Construct validation of the Arizona clinical interview rating scale. *Educ Psychol Measurement* 1977;**37**:1031–8.

7 Levenkron JC, Greenland P. Validation of the University of Rochester Risk Factor Interview Scale. *Am J Prev Med* 1987;**3**:152–61.

8 Mullan P, Zweifler A, Wolf F, Dicken L, Lenhard D. Evaluating medical students' sexual history performance. *Profess Educ Res Q* 1992;**13**:16.

9 Furner V, Ross M. Lifestyle clues in the recognition of HIV infection. How to take a sexual history. *Med J Aust* 1993;**158**:40–1.

10 Jewell ME, Jewell GS. How to assess the risk of HIV exposure. *Am Fam Physician* 1989;**40**:153–61.

11 Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika* 1951;**14**:297–334.

12 Winer BJ. *Statistical Principles in Experimental Design.* 2nd edn. New York: McGraw-Hill; 1971.

13 Waitzkin H. Doctor–patient communication: clinical implications of social scientific research. *J Am Med Assoc* 1984;**252**:2441–6.

14 West C. Against our will: male interruptions of females in cross-sex conversation. *N Y Acad Sci* 1979;**327**:81–97.

15 Anderson BA, Silver BD, Abramson PR. The effects of the race of the interviewer on race-related attitudes of black respondents in SRC/CPS National Election Studies. *Public Opin Q* 1988;**52**:289–324.

16 Schaeffer NC. Evaluating race-of-interviewer effects in a national survey. *Sociol Meth Res* 1980;**8**:400–19.

17 Schuman H, Converse J. The effects of black and white interviewers on black responses in 1968. *Public Opin Q* 1971;**35**:44–68.

18 Blendon RJ, Scheck AC, Donelan K, Hill CA, Smith M, Beatrice D, *et al.* How white and African Americans view their health and social problems. Different experiences, different expectations. *J Am Med Assoc* 1995;**273**:341–6.

19 Stillman P, Swanson D, Regan MB, Philvin MM, Nelson V, Ebert T, *et al.* Assessment of clinical skills of residents utilizing standardized patients. A follow-up study and recommendations for application. *Ann Intern Med* 1991;**114**:393–401.