

Reliability of interviewing in medical student selection

P RICHARDS, I C McMANUS, S A MAITLIS

Abstract

Six interviewing panels which had assessed candidates for places at a medical school in 1984 were reconvened in 1987 to assess videotapes of the interviews conducted by themselves and by other panels. Overall recommendations made by panels showed excellent reliability within panels and high reliability between panels.

Introduction

The interview is a central part of student selection in all London University medical schools and in many other British medical schools. Interviews have, however, been criticised as a form of assessment because of poor reliability (consistency between interviewers) and poor validity (the relation between the assessment

and the candidate's true ability and potential).^{1,4} We previously described the interview procedure at our medical school in the University of London⁵ and noted relatively high correlations between the independent judgments of three interviewers who were present at the same interview, implying a reasonable degree of reliability.

The opportunity for this study arose when the BBC began a long term television project in which the selection, training, and eventual practice of a cohort of medical students at St Mary's would be monitored over many years. As a result 56 selection interviews carried out at St Mary's during the autumn of 1984 were filmed in whole or in part by professional cameramen as unobtrusively as possible. After the first programme was screened (*Horizon: Doctors To Be*, 30 June 1986) several of the interviewers expressed the wish not only to discuss in more detail what was taking place in interviews but also to see some of the interviews in their entirety. About 25 interviewers therefore spent an afternoon watching and then discussing four complete interviews. Three of the four cases produced little dissent concerning the final decision that was reached, but in one case most of the interviewers, as well as the chairman of the original panel, thought that the decision had probably been wrong. How stable, therefore, were interviewers' assessments? The present study was carried out to answer that question. Six interviewing panels were reconstituted to reassess not only their own interviews but also those of other panels.

Throughout this paper we report only on the reliability of

St Mary's Hospital Medical School, London W2 1PG

P RICHARDS, MD, FRCP, dean and professor of medicine

I C McMANUS, MD, PHD, lecturer in psychology as applied to medicine

S A MAITLIS, BSC, research assistant

Correspondence to: Professor Richards.

assessments—that is, how much they agree with one another. We are not discussing validity—the degree to which assessments correspond with the true description of the candidates.

Method

Of the original nine interviewing panels that had been filmed, it was possible to reconvene six in March-May 1987. On each study day two panels were reconvened, each with its chairman and two interviewers. The two panels were seated in adjacent rooms, each equipped with a video monitor, and each watched both its own interviews from 1984 and those of the other panel. Half of the panels saw their own interviews first and the other half saw the interviews of the other panel first. As far as possible the interviews replicated the process that had occurred three years earlier. Interviewers were provided with Universities Central Council on Admissions forms, and with similar forms to those used in 1984.

The tape for a candidate was played from the moment the candidate entered the room until he or she left when the tape was stopped. Interviewers then made their own independent assessments of the candidate, after which the panel as a whole discussed the candidate and reached a final recommendation. The panel did not see any of the discussion from their previous deliberations nor were they given any information concerning their judgments in 1984. Candidates were omitted from the study if they had been included in the broadcast television programme or had been a part of the afternoon's discussion described earlier.

Statistical analysis was by means of contingency tables and by the analysis of Pearson's correlation coefficients, so that using a structural model described elsewhere⁶ it was possible to calculate intrainterviewer reliabilities—that is, whether the same interviewer assessing the same candidate made the same judgment—and interinterviewer reliabilities—whether two separate interviewers given the same information (in this case in the form of a videotape) made the same judgment. Similar calculations were made for intrapanel and interpanel reliabilities.

Results

Six panels, made up of two chairmen and 11 other interviewers, were reconvened and reassessed a total of 31 interviews. Table 1a shows the judgments made by the original interviewers in 1984 and the same interviewers in 1987. Interviewers used the same scale as in 1984, on which there were four steps: A (definite offer), B1 (probable offer),

TABLE 1—Overall recommendations of panel for each of 31 separate interviews

	1987			
	A	B1	B2	C
<i>(a) Comparison of original decision in 1984 with decision of same panel in 1987 after watching a videotape of the interview:</i>				
1984:				
A: Accept	8	2		
B1: Probable offer	3	4	4	
B2: Waiting list		2	2	1
C: Reject			2	4
<i>(b) Comparison of decision in 1987 by original panel with decision of other panel in 1987, each watching videotaped interview:</i>				
1987—original panel:				
A: Accept	6	4	1	
B1: Probable offer	2	4	2	
B2: Waiting list		2	2	3
C: Reject		1	2	2

B2 (waiting list), and C (reject). Of 31 interviews, 18 produced an identical recommendation, and the remaining 13 produced a decision that differed by only one step from the original, with as many candidates moving up (seven) as moving down (six). No recommendations differed by two or more steps. Pearson's correlation between the two judgments was 0.825. Table 1b shows a similar comparison between the recommendations based on the videotaped interviews by the reconvened original panel and by the other panel. Of 31 recommendations, 14 are identical, 15 differed by one step, and two (7%) differed by two steps. No candidates were recommended for definite acceptance by one panel and rejection by the other. The correlation

between the two judgments is 0.673. Taken together these two correlations allow the calculation of an intrapanel reliability of 0.908 and an interpanel reliability of 0.901.

Since each panel contained three people a total of 93 separate sets of assessments were made of candidates. By including only forms with complete data 79 complete sets of judgments were left. Table II shows correlations and reliabilities within and between interviewers for each judgment. ("Health" was omitted since in almost all cases this was rated as

TABLE II—Pearson's correlations between judgments of original interviewer in 1984 and same interviewer in 1987 (84-87) and between original interviewer observing videotape in 1987 and another interviewer watching videotape in 1987 (87i-87ii), as well as calculated intrainterviewer and interinterviewer reliabilities

Scale	Correlations		Reliabilities	
	84-87	87i-87ii	Intrainterviewer	Interinterviewer
Academic ability	0.177	0.199	0.421	1.060*
Personality	0.559	0.436	0.748	0.883
Contribution to life of school	0.558	0.524	0.747	0.968
Potential	0.589	0.472	0.767	0.895
Recommendation	0.769	0.528	0.877	0.828

*Indicates theoretically inadmissible value owing to sampling variation which should be interpreted as unity.

"excellent.") Differences between and within individuals in their judgments were greater than the differences between panels, but this did not prevent the high degree of intrapanel and interpanel reliability. Separate analyses of chairmen and other interviewers suggested that chairmen were more consistent in their individual judgments.

Discussion

Taken overall the recommendations made by the panels of interviewers are very reliable, not only with their own previous judgments but also with those of other interviewers now observing the same interview. It is possible that that reliability comes in part from the particular questions that were asked by a panel and that interviewers asking different questions might come to a different conclusion. That possibility cannot be assessed here.

Reliabilities were higher for chairmen than for other interviewers on all scales. Each chairman attends interviews weekly during the autumn term over a period of years and is thus far more experienced than other interviewers, who interview only twice or three times a year. Additional experience may account for the increased reliability of chairmen's judgments, which suggests that additional training accompanied by a standardisation of criteria might increase the reliability of the judgments of less experienced interviewers. Previously we showed that the judgment of the chairman has a disproportionate influence on the overall recommendation of the panel as a whole.⁶ If the judgments of chairmen are more reliable then the situation becomes defensible, provided, of course, that their judgments are valid.

We are grateful to the applicants to St Mary's who agreed to be filmed while undergoing their selection interviews and to the interviewers who helped us by re-viewing those interviews. We also thank the BBC TV *Horizon* programme for providing us with videotapes of the interviews and for helpful discussions about the project.

References

- Arvey RD, Campion JE. The employment interview: a summary and review of recent research. *Personnel Psychology* 1982;35:281-322.
- Schmitt N. Social and situational determinants of interview decisions: implications for the employment interview. *Personnel Psychology* 1976;29:79-101.
- Ulrich L, Trumbo D. The selection interview since 1949. *Psychol Bull* 1965;63:100-16.
- Herriot P. *Down from the ivory tower: graduates and their jobs*. Chichester: John Wiley, 1984.
- McManus IC, Richards P. Audit of admission to medical school. II. Short-listing and interviews. *Br Med J* 1984;289:1288-90.
- McManus IC. Medical students: origins, selection, attitudes and culture. University of London: London, 1985. (MD thesis.)

(Accepted 26 January 1988)