DOCUMENT RESUME

ED 185 114 TM 800 132

AUTHOR Markert, Ronald J.; Shores, Jay H.
TITLE Assuring Fairness in the Medical School Admission Interview Through Analysis of Rater Difficulty and Consistency.
PUB DATE Apr 80
NOTE 12p.; Paper presented at the Annual Meeting of the American Educational Research Association (64th, Boston, MA, April 7-11, 1980). Table 1 may be marginally legible.

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Admission Criteria; *Bias; Higher Education; *Interviews; *Medical School Faculty; *Medical Schools; Personality Assessment; Predictive Validity; Rating Scales; Reliability
IDENTIFIERS College of Osteopathic Medicine TX; *Interrater Reliability

ABSTRACT

Although unreliable and not predictive of medical school performance, the admission interview continues to be used extensively to collect noncognitive data about medical school applicants. A procedure is suggested for assuring that a group of interviewers assigned to an applicant does not unfairly help or hinder the candidate's ratings. At least 3 medical school faculty interviewed and rated 231 applicants as exceptional, acceptable, minimally acceptable, or unsuitable; reliability between raters was low. To minimize this problem, a difficulty-consistency index was developed for classifying interviewers, and a rule-of-thumb was proposed for assuring fairness in assigning interviewers. Data from one year's applicants show that 48 of 148 candidates who were actually admitted to the class of 1983, were interviewed by a group of interviewers with an unsatisfactory difficulty-consistency total. (Author/CP)

# Assuring Fairness in the Medical School Admission Interview Through Analysis of Rater Difficulty and Consistency

Ronald J. Markert, Ph.D.
and
Jay H. Shores, Ph.D.



Dr. Markert is assistant professor and director of evaluation and Dr. Shores is associate professor. Both are with the Office of Medical Education, Texas College of Osteopathic Medicine in Fort Worth.

Presented at the Annual Meeting of the American Educational Research Association, Boston, April 1980.

Appreciation is extended to Dr. Michael Budd, Assistant Dean for Student Affairs, and the Office of Admissions for providing the researchers with access to the data upon which this study is based.

The medical school admission interview has been shown to correlate poorly with performance in medical school (1-7). In addition to lack of evidence for predictive validity, critics have noted that medical school admission interviews are unreliable (8) and expensive in terms of professional time (9).

Nevertheless, the interview continues to be widely used. Poorman (10) reported that the 1975-76 Medical School Admission Requirements Handbook indicated that 104 of 109 U.S. medical schools providing data utilized the interview in the admission process. Cough (7) observed that proponents of the medical school admission interview support its use on two different unverified bases. First, the interview can be used to identify and reject applicants for reasons which cannot be discerned through examination of data submitted by applicants. Second, the interview to some extent assures the applicant that he is receiving individual attention. It is the authors' view that other means (i.e., structured personality tests, letters of recommendations, essays) of assessing non-cognitive information (e.g., personality characteristics, attitudes, motivation, interests) have not proven satisfactory. Lacking a viable substitute, the interview has persisted in large part because of the appeal of personal involvement. Whatever the justification, Poorman (10) is no doubt correct when he states that in medical schools "the interview is here to stay; if not by reason, then by tradition." (p.301)

Problem

Since the interview will remain an important part of the medical school admission process despite its drawbacks, what can be done to assure fairness in its use? Medical school admission committees have expressed concern that an applicant's chances of being selected often depend upon the interviewers to whom he is assigned and not necessarily

upon the noncognitive criteria which the interview is to assess.

This study addresses the issue of fairness by examining two important components of interview fairness - the difficulty and consistency of interviewers - in the context of one medical school's admission process. Difficulty refers to the tendency of an interviewer to assign higher or lower applicant scores in comparison with other interviewers. Consistency refers to the degree of agreement among interviewers rating the same applicant.

Setting

The Texas College of Osteopathic Medicine admits its entering class by assessing applicants on a variety of criteria: premedical grade point average, scores on the Medical College Admission Test (MCAT), quality of experience in health-related work, exposure to and understanding of the osteopathic profession, likelihood of fulfilling the college's mission of providing general practitioners for the state of Texas, premedical academic and other honors, and suitability for the practice of medicine as judged by letters of recommendation and one-on-one admission interviews.

An Admission Committee establishes substantive areas in which interviewers are to judge applicants. For the class entering in the fall of 1979 (Class of 1983) these areas were problem-solving, life problem-solving, human interactions, responsibility, social sensitivity and awareness, osteopathic motivation, and self-appraisal. The faculty interviewers participated in a September 1978 workshop prior to the October to March interviewing period. The workshop stressed the substantive areas mentioned above, techniques for interviewing, and trial assessments.

Data

During the five-month interview period 231 applicants were invited for interviews. Each applicant was interviewed individually by at least three faculty members. The applicant was rated either exceptional (3), acceptable (2), minimally acceptable (1), or reject (0). In 24 of the 708 interviews raters chose a rating between two categories (e.g., 1.5 - between minimally acceptable and acceptable). Of the 231 applicants interviewed, 216 received three interviews while 15 received an additional or fourth interview. The fourth interview was employed when an applicant received an exceptional (3) and reject (0) rating among his original three ratings.

Table 1 presents data related to the 30 faculty interviewers who conducted 10 or more interviews. Fifteen faculty interviewers who conducted less than 10 interviews were excluded from the study.

The columns headed group refer to the mean and standard deviation for the rotating group of two or three colleagues who rated the same applicants as the interviewer. The mean absolute difference is the average difference between the interviewer and the group rating. The mean +/- difference is an index of an interviewer's tendency to be a difficult rater (minus) or an easy rater (plus). The correlation coefficient is a Pearson r between interviewer and group ratings.

Results and Discussion

Inspection of the interviewer mean column in Table 1 reveals that an applicant's total rating could vary dramatically depending on the interviewers he was assigned. (Interviewers were assigned on a convenience basis and not with a random procedure.) While three "average" interviewers would result in a total rating of 5.52 (1.84 x 3), other combinations would result in very different totals. If rating an "average"

Table 1: 9

RATER COMPARISON FOR ADMISSION INTERVIEWS - CLASS OF 1983

| INTERVIEWER NUMBER | NUMBER OF INTERVIEWS | RATINGS 0.0 | 0.5 | 1.0 | 1.5 | 2.0 | 2.5 | 3.0 | INTERVIEWER MEAN | INTERVIEWER STD DEV | GROUP MEAN | GROUP STD DEV | MEAN ABS DIFF | MEAN +/- DIFF | CORR COEFF |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2 | 10 | 1 | 0 | 9 | 0 | 1 | 0 | 0 | 1.84 | 1.039 | 1.56 | 0.612 | 0.96 | 0.28 | 0.158 |
| 4 | 33 | 3 | 0 | 12 | 0 | 10 | 0 | 8 | 1.70 | 0.937 | 1.65 | 0.742 | 0.58 | 0.07 | 0.545 |
| 5 | 29 | 1 | 0 | 16 | 1 | 10 | 1 | 0 | 1.38 | 0.567 | 1.87 | 0.737 | 0.69 | -0.42 | 0.311 |
| 6 | 20 | 0 | 0 | 6 | 3 | 4 | 4 | 3 | 1.97 | 0.722 | 1.84 | 0.634 | 0.34 | 0.04 | 0.734 |
| 7 | 36 | 0 | 0 | 17 | 1 | 12 | 0 | 6 | 1.68 | 0.727 | 1.95 | 0.803 | 0.58 | -0.27 | 0.537 |
| 8 | 24 | 1 | 0 | 3 | 2 | 9 | 1 | 8 | 2.10 | 0.803 | 1.96 | 0.731 | 0.66 | 0.15 | 0.486 |
| 9 | 10 | 2 | 0 | 0 | 0 | 6 | 0 | 2 | 2.20 | 0.980 | 1.97 | 0.557 | 0.97 | -0.17 | 0.017 |
| 10 | 37 | 3 | 0 | 5 | 0 | 12 | 0 | 17 | 2.16 | 0.945 | 1.82 | 0.673 | 0.75 | 0.34 | 0.435 |
| 11 | 35 | 0 | 0 | 8 | 0 | 16 | 0 | 11 | 2.09 | 0.732 | 1.89 | 0.719 | 0.52 | 0.19 | 0.521 |
| 12 | 32 | 0 | 0 | 9 | 0 | 16 | 0 | 7 | 1.94 | 0.734 | 1.90 | 0.728 | 0.69 | 0.04 | 0.216 |
| 13 | 20 | 2 | 0 | 5 | 0 | 4 | 1 | 8 | 1.97 | 1.020 | 2.10 | 0.469 | 0.73 | -0.12 | 0.430 |
| 14 | 10 | 0 | 0 | 1 | 0 | 3 | 0 | 6 | 2.50 | 0.671 | 1.99 | 0.961 | 1.06 | 0.51 | -0.209 |
| 15 | 14 | 1 | 0 | 1 | 0 | 6 | 0 | 6 | 2.21 | 0.860 | 1.99 | 0.586 | 0.55 | 0.23 | 0.574 |
| 16 | 19 | 0 | 0 | 5 | 0 | 7 | 0 | 6 | 2.00 | 0.790 | 1.94 | 0.497 | 0.44 | 0.11 | 0.617 |
| 18 | 28 | 0 | 0 | 3 | 0 | 20 | 1 | 4 | 2.05 | 0.506 | 1.86 | 0.558 | 0.42 | 0.19 | 0.237 |
| 19 | 21 | 4 | 1 | 5 | 2 | 6 | 0 | 3 | 1.43 | 0.959 | 1.79 | 0.867 | 0.68 | -0.33 | 0.606 |
| 20 | 21 | 1 | 0 | 6 | 0 | 6 | 0 | 8 | 2.00 | 0.926 | 1.86 | 0.929 | 0.67 | 0.14 | 0.554 |
| 21 | 15 | 3 | 0 | 6 | 0 | 4 | 0 | 2 | 1.73 | 0.943 | 1.56 | 0.647 | 0.63 | -0.28 | 0.467 |
| 23 | 17 | 0 | 0 | 2 | 0 | 6 | 0 | 9 | 2.41 | 0.691 | 2.02 | 0.766 | 0.60 | 0.33 | 0.504 |
| 24 | 14 | 1 | 0 | 6 | 0 | 5 | 0 | 2 | 1.97 | 0.821 | 1.92 | 0.539 | 0.92 | -0.34 | -0.109 |
| 25 | 12 | 0 | 0 | 3 | 0 | 7 | 0 | 2 | 2.17 | 0.731 | 1.75 | 0.672 | 0.86 | 0.50 | 0.193 |
| 26 | 26 | 1 | 0 | 6 | 0 | 6 | 0 | 11 | 2.12 | 0.801 | 1.72 | 0.635 | 0.83 | 0.39 | 0.250 |
| 27 | 14 | 2 | 0 | 3 | 0 | 7 | 0 | 2 | 1.86 | 0.995 | 2.10 | 0.582 | 0.99 | -0.54 | -0.256 |
| 28 | 16 | 5 | 0 | 4 | 0 | 2 | 0 | 5 | 1.44 | 1.223 | 1.74 | 0.872 | 0.89 | -0.30 | 0.438 |
| 30 | 13 | 4 | 0 | 4 | 0 | 3 | 0 | 2 | 1.23 | 1.049 | 1.69 | 0.773 | 0.62 | -0.46 | 0.799 |
| 32 | 23 | 2 | 0 | 9 | 0 | 8 | 0 | 4 | 1.61 | 0.972 | 1.80 | 0.644 | 0.68 | -0.19 | 0.468 |
| 33 | 10 | 1 | 0 | 2 | 0 | 3 | 0 | 4 | 2.00 | 1.000 | 2.00 | 0.837 | 0.80 | 0.00 | 0.299 |
| 34 | 14 | 2 | 0 | 5 | 0 | 5 | 0 | 1 | 1.82 | 0.921 | 1.68 | 0.554 | 0.68 | -0.25 | 0.460 |
| 38 | 15 | 3 | 0 | 4 | 0 | 4 | 0 | 4 | 1.80 | 1.083 | 1.72 | 0.529 | 0.86 | -0.12 | 0.139 |
| 39 | 20 | 4 | 0 | 5 | 0 | 9 | 1 | 2 | 1.47 | 0.942 | 1.64 | 0.665 | 0.77 | -0.17 | 0.372 |
| TOTALS(*MEAN) | 622 | 47 | 1 | 170 | 9 | 219 | 9 | 167 | 1.8**** | 0.911* | | | 0.72* | -0.02* | 0.364* |





BEST COPY AVAILABLE

applicant, the three most difficult interviewers (nos. 5, 21, and 30) would yield a total of 3.94 (1.38 + 1.33 + 1.23). The three easiest interviewers (nos. 14, 23, and 25) would yield a total of 7.19 (2.50 + 2.41 + 2.28). Based on the cutoff level established by the Admission Committee, the latter applicant would be a viable candidate for admission, the former would not.

Thus, given that the ratings are of questionable reliability, how can fairness be assured in circumstances where the interview process must utilize a cadre of interviewers who are not professionally trained for interviewing and whose academic heterogeneity and personal perspectives on the definition of a physician differ greatly?

One often suggested solution to the problem outlined above is to adjust each interviewer's mean rating in accordance with the group mean. Thus, interviewer no. 4 might have .07 subtracted from each rating since he is .07 above the group mean (1.70 - 1.63). This procedure assumes that applicants are randomly assigned to interviewers. The assumption is made that all interviewers in the long run are given applicants of equal merit to interview. This, of course, is not the case. The medical school interview process is reliant upon the scheduling contingencies of both interviewers and applicants.

An alternative procedure which is less dependent on the assumption of random assignment uses a two-variable approach to categorize interviewers. Table 2 categorizes the 30 interviewers on the dimensions of difficulty and consistency. Rational judgment was used by the authors in establishing three categories of difficulty and consistency. (Note: Statistical tests to ascertain differences in difficulty and consistency on the basis of sex, age, and academic department were nonsignificant.)

Table 2: Difficulty-Consistency Classification of Interviewers*

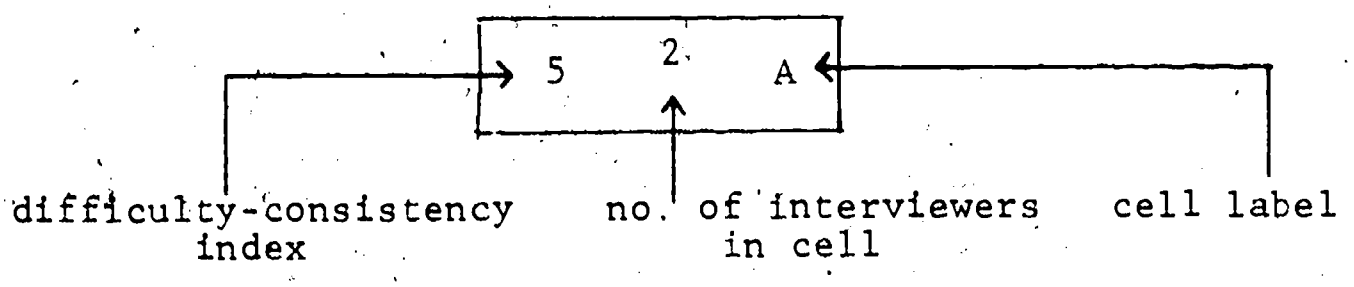
| Difficulty | Consistency (Pearson r): less than .30 | .30 - .49 | more than .49 |
|---|---|---|---|
| 1.60 or less | 5 2 A | 3 5 D | 1 2 G |
| 1.61-2.05 | 4 6 B | 2 2 E | 1 4 H |
| more than 2.05 | 5 3 C | 3 2 F | 1 4 I |

Cells G, H, and I would be assigned a difficulty-consistency index of 1; cell E would be assigned an index of 2; cells D and F would be 3; cell B would be 4; and cells A and C would be 5. On the 1 - 5 continuum a 1 represents the most desirable interviewers (i.e., those most in agreement with their group of interviewers) while a 5 represents the least desirable (i.e., those that are either difficult and not consistent with their group or easy and not consistent with their group).

An admission committee might be advised to establish a rule-of-thumb by which an applicant would not be interviewed by three interviewers who total to more than eight. Thus, if an applicant were assigned a 5 (easy or difficult and inconsistent), he would also have to be assigned a 2 and a 1 or two 1's, that is, two of the better interviewers.

---

*Table 2 should be read as follows:



If this rule-of-thumb were applied retrospectively to the Class of 1983 interviewers, how many applicants would have been "fairly" treated (i.e., assigned a group of interviewers whose difficulty-consistency classification totalled to 8 or less). Table 3 reports these data.

Table 3: Number of Applicants in Each Difficulty-Consistency Classification

| | Difficulty-Consistency Classification For Interview Groups* | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
| No. of Applicants | 15 | 14 | 15 | 20 | 19 | 17 | 13 | 6 | 10 | 14 | 3 | 2 | 0 |

Mean = 7.34
S.D. = 2.93
N = 148

One hundred of 148 (67.6%) applicants were interviewed by a group of three interviewers whose difficulty-consistency classification was 8 or less. Forty-eight (32.4%) applicants were interviewed by groups who totalled 9 or more. Thus, nearly one-third of the Class of 1983 applicants was interviewed by a group of interviewers whose difficulty-consistency total was unsatisfactory.

---

*Of the 231 applicants, 15 were not included in Table 3 in that they had four interviews and 68 were not included in that they had one or more interviews by the least active interviewers not included in Table 1.

Summary

Although unreliable and not predictive of medical school performance, the admission interview continues to be used extensively as a means of collecting noncognitive data about medical school applicants. This study suggests a procedure for assuring that a group of interviewers assigned to an applicant does not unfairly help or hinder the candidate's ratings. The difficulty-consistency index was developed for the purpose of classifying interviewers, and a rule-of-thumb proposed for assuring fairness in assigning interviewers. Data from one year's applicants show that 48 of 148 candidates were interviewed by a group of interviewers with an unsatisfactory difficulty-consistency total. The procedure described was developed in the practical setting of one medical school's admission process. It is suggested that continued practical research combined with experimental study of the fairness issue offers promise for improvement of the much-maligned medical school admission interview.

## References


1. Richards, J.M., Jr., and Taylor, C.W. Predicting Academic Achievement in a College of Medicine from Grades, Test Scores, Interviews and Ratings. Educ. Psychol. Meas., 21:987-[illegible], 1961.

2. Johnson, D.G. A Multifactor Method of Evaluating Medical School Applicants. J.Med. Educ., 37:656-665, 1962.

3. Gough, H.S., Hall, W.B., and Harris, R.E. Evaluation of Performance in Medical Training. J.Med. Educ., 39:679-692, 1964.

4. Lief, V.F., Lief, H.I., and Young, K.M. Academic Success: Intelligence and Personality. J.Med. Educ., 40:114-124, 1965.

5. Mensh, I.N. Orientations of Social Values in Medical School Assessment. Soc. Sci. Med., 3:339-348, 1970.

6. Murden, R., Halloway, G.M., Reid, J.C., and Colwill, J.M. Academic and Personal Predictors of Clinical Success in Medical School. J. Med. Educ., 53:711-719, 1978.

7. Gough, H.S. How to Select Medical Students. Medical Teacher, 1:17-20, 1979.

8. Gordon, M.J., and Lincoln, J.A. Family Practice Resident Selection: Value of the Interview. J.Fam. Pract., 3:175-177, 1976.

9. Kelly, E.L., A Critique of the Interview. J.Med. Educ., 32 (No. 10, Part 2):78-84, 1957.

10. Poorman, D.J. Medical School Applicant: A Study of the Admission Interview. J.Kans. Med. Soc., 76:298-301, 1975.