

Personality Perceptions of Medical School Applicants

R. Blake Jelley, MA,^{*} Michael A. Parkes, MA^{*}, Mitchell G. Rothstein, PhD[†]

^{*} Department of Psychology

[†] Richard Ivey School of Business
The University of Western Ontario

Abstract: Purpose To examine the extent to which medical school interviewers consider perceptions of applicant personality traits during a semi-structured panel interview, the interrater reliability of assessments, and the impact of such perceptions on individual admission decisions.

Method Semi-structured panel interviews were conducted with applicants to the Doctor of Medicine Program at the University of Western Ontario in London, Canada. Interviewers also provided voluntary, “research only” ratings of applicants on nine relevant personality traits. Data from 345 applicants under consideration for admission were available for analysis.

Results Significant correlations were observed between personality ratings and important operational variables (e.g., interview scores). Applicants who were most likely to be admitted to the program were perceived as high on certain traits (i.e., Achievement, Nurturance, Endurance, Cognitive Structure, & Order) and low on other traits (i.e., Abasement, Aggression, & Impulsivity). Statistically removing variance shared with personality ratings from interview scores resulted in different admission decisions for over 40% of the applicants. Interrater reliabilities for personality perceptions were relatively low. However, interrater reliability of the panel interview used to make admission decisions was acceptable. Nonlinear relations between personality perceptions and interview scores were also explored.

Conclusion Some evidence was found that interviewers’ perceptions of applicant personality may affect their judgments when assigning interview ratings. Given that non-cognitive characteristics are perceived as important in the admissions process and that perceptions of personality traits have implications for decisions about which candidates to admit, suggestions for identifying desirable non-cognitive characteristics and for increasing the quality of assessments are offered.

Key Words: Personality; Perception; Interview; Selection; Admission; Medical Students; Medical Education.

In the process of determining which of many highly qualified candidates to admit to a program of medical education there is widespread concern for selecting those who will ultimately be effective physicians.¹ It is often held that various personal qualities, beyond academic achievement and ability, are important predictors of success as a medical student and of clinical competence after graduation.^{1,2,3} Adjectives like caring, persistent, orderly, and confident have been offered as reflecting some of the desirable characteristics of physicians. However, defining the qualities that make a ‘good doctor’ and applying them across medical specialties can be challenging.¹

Powis¹ noted that the explicit identification and definition of personal qualities is difficult and tends to be ignored in practice. Further, the task of choosing procedures to accurately assess such qualities can be even more taxing to a medical admissions committee. This is particularly true with respect to the use of psychometric tests.¹ Powis’ primary concerns regarding the demands (e.g., selection or design of

scales, validation) that accompany the appropriate use of psychometric tests are well founded. Indeed, even some of Powis’ more contentious comments (e.g., “psychometric tests are usually designed to assign people to groups rather than rank order them on a continuum” p. 460) serve to underscore the complexity of issues faced by Admissions Committees who opt to use personality inventories. For various reasons, Admissions Committees are unlikely to use psychometric tests for measuring personal qualities¹ despite some encouraging evidence about the use of such tests for predicting the clinical performance of medical students² and for predicting job performance.^{4,5,6}

Interviews may be used to assess various non-cognitive characteristics, including personality traits.^{1,3,4,7,8} The interview is all but ubiquitous for personnel selection in organizations and is similarly popular with respect to medical school admission.⁷ Its persistent popularity has become more palatable in recent years because “after decades of empirical embar-

assessment, recent meta-analytic research supports the belief that respectable levels of validity can be attained through responsible use of more structured interviews⁴ (p. 116). Benefits of more structured interviewing have been recognized in the medical education literature,^{7,9} yet questions as to the characteristics best measured by interviews persist both in medical education⁹ and in interview research.⁸ Interview researchers have started to direct greater attention to the constructs assessed in the interview, yet little progress has been made to date.⁸

It seems plausible that medical admission committees and interviewers could use interviews to assess applicant personality. Considerable research in personality and social psychology supports the notion that, in certain circumstances, observers can make reasonably valid judgments of others' personality, even when provided with little opportunity for direct interaction.⁴ Also, Tutton³ reported significant correlations between medical students' personality, as measured by self-reports to the California Personality Inventory (CPI), and interview ratings. Of course, the quality of information gathered during an interview is dependent upon interviewer(s) and the procedures they use. Structured interviews are designed to reduce idiosyncratic interviewer effects (e.g., subjectivity, bias) and increase reliability and validity.⁷ However, the majority of research on employment and medical admissions interviews has involved semi-structured interviews⁷ and there can be considerable resistance in developing, implementing, and maintaining a highly structured approach to interviewing.⁸ In practice, interviews may not be as reliable or free of interviewer effects as they could be. Nevertheless, there appears to be reason for optimism that the interview may be used to accurately assess personality constructs that are predictive of job performance.

Our primary purposes were to investigate (a) the extent to which medical admission interviewers evaluated specific personality traits of applicants in an admissions interview, (b) implications perceptions of personality might have on individual admission decisions, and (c) the interrater reliability of personality perceptions. This exploratory study reflects interest in non-cognitive predictors of clinical performance in medicine, the resurgence of interest in personality for personnel selection, as well as the modern construct orientation in selection interview research.

Methods

In accordance with standard procedures, semi-structured panel interviews were conducted with

medical school applicants. Each interview panel consisted of three members (1 physician, 1 community representative, 1 medical student). Numerous panels of interviewers conducted interviews with individual applicants who were randomly assigned to a given panel. Following each interview, interviewers independently completed rating forms that were used as part of the admissions process. Interviewers were also asked to provide "research only" ratings of applicants' personality, following completion of the standard rating forms.

Personality Perception Scores - The personality ratings employed a 1 to 5 Likert scale of the extent to which an applicant possessed each of the personality traits described (1 = "very slightly or not at all;" 5 = "extremely"). Personality traits, pertaining to the effective performance of physicians, were chosen on the basis of a review of the medical education and psychology literatures. Interviewers were provided with trait names and several adjectives that are descriptive of high scorers (listed below). The traits and adjectives were adapted from Jackson's¹⁰ Personality Research Form (PRF-E). The PRF is the product of a painstaking and sophisticated scale development program and is a widely used personality inventory.¹¹ PRF scales correlate with one or more of the Big Five personality factors and also provide important information that is sacrificed when only broad personality factors are considered.¹¹ Construct definitions and descriptive adjectives for both poles (high and low) of the traits are available in the PRF manual.¹⁰

The number of traits, rating method, and provision of brief descriptions were procedures employed to address a pragmatic concern for obtaining personality perception ratings in a minimally intrusive way. A given panel's mean rating of a given applicant on a given trait served as a personality perception score. The predicted directions of correlations between personality perception scores and interview scores (unknown to interviewers) are given in parentheses. These predictions were made on the basis of rational considerations and previous research on the personality traits of effective physicians.^{1,2,3}

- **Abasement:** meek, self-critical, humble, apologizing, deferential (predicted a negative correlation – applicants perceived as low on Abasement were expected to receive higher interview scores);
- **Achievement:** striving, accomplishing, purposeful, industrious, enterprising, self-improving, ambitious (predicted a positive correlation – applicants perceived as high on Achievement were expected to receive higher interview scores);

- **Aggression:** blunt, argumentative, pushy, easily-angered (predicted a negative correlation – applicants perceived as low on Aggression were expected to receive higher interview scores);
- **Cognitive Structure:** precise, perfectionist, accurate, avoids ambiguity (predicted a positive correlation – applicants perceived as high on Cognitive Structure were expected to receive higher interview scores);
- **Dominance:** controlling, domineering, forceful, leading, assertive, authoritative, powerful (predicted a positive correlation – applicants perceived as high on Dominance were expected to receive higher interview scores);
- **Endurance:** persistent, determined, tireless, energetic, dogged (predicted a positive correlation – applicants perceived as high on Endurance were expected to receive higher interview scores);
- **Impulsivity:** hurried, impatient, spontaneous, excitable, uninhibited (predicted a negative correlation – applicants perceived as low on Impulsivity were expected to receive higher interview scores);
- **Nurturance:** sympathetic, helpful, encouraging, caring, protective, comforting, consoling, charitable (predicted a positive correlation – applicants perceived as high on Nurturance were expected to receive higher interview scores);
- **Order:** neat, organized, systematic, well-ordered, disciplined, methodical, deliberate (predicted a positive correlation – applicants perceived as high on Order were expected to receive higher interview scores).

Applicants - Data were available for 345 applicants to the Doctor of Medicine Program at the University of Western Ontario, in London, Canada. These applicants met minimum admissions standards in terms of Grade Point Average (GPA) and Medical College Admission Test (MCAT) scores, and represented a subset of approximately 1600 applications. Personality rating forms were coded separately and were matched with other data by means of a common identification number.

Interviewing Procedure - All interviewers were required to attend a briefing session to review desirable and unacceptable activities with respect to selection interviewing. An interviewer manual, outlining interviewing goals and procedures, was given to all interviewers. Interviewers were informed that applicants were to be selected on the basis of personal characteristic, in addition to academic qualifications (i.e., GPA; MCAT). The interview rating categories used at the time of this study were: maturity; com-

munication skills; preparation for career; non-academic activities; awareness of Canadian and world affairs; human values; and overall impression (“suitability of the candidate to pursue a career in medicine”). Suggested questioning approaches for each category of this semi-structured interview were provided in the interviewer manual. Following interviews, rating forms were submitted to the Admissions Office where each interviewer’s interview score for each applicant was calculated based on a predetermined algorithm. The mean interview score (MIS) was the average of interview scores from all three sources: physicians; community representatives; and medical students.

Data Analyses - To answer the first research question, do interviewers seem to make judgments about applicant personality, we computed bivariate correlation coefficients between interview scores (e.g., MIS) and personality perception scores. A correlation between a given trait and MIS does not necessarily prove that perceived personality caused interviewers to assign a given interview score. However, failure to demonstrate a correlation between personality perception ratings and interview scores certainly calls into question the importance of perceived personality to MIS, at least with respect to the chosen traits. We expected significant correlations between the perceptions of applicants on the nine traits and MIS, as described previously.

The second research question focused on the extent to which personality perceptions affect admission decisions. The application score (APPSCORE) variable is particularly important with respect to this question because it serves as the basis for admission decisions. It is a composite of GPA, two MCAT scales, and MIS. One method for investigating this question was to correlate APPSCORE with personality perception scores. If no significant correlations were observed with the APPSCORE variable we would have little reason to suspect that interviewers’ perceptions of applicant personality were influencing admission decisions.

To further explore the second research question we dichotomized the APPSCORE variable and formed two groups: (a) First Choice applicants (i.e., top APPSCORES); and (b) admissible, but less preferred applicants. Applicants who did not meet minimum admission requirements were not interviewed and were, therefore, excluded from analysis. A multivariate analysis of variance (MANOVA) was conducted with the nine personality variables as dependent variables to address the question of whether there was some way of combining the personality

variables to distinguish First Choice applicants from other applicants. Independent sample *t*-tests were also conducted to examine mean differences between groups on each of the nine traits.

To more fully understand the impact that personality perceptions may have on admission decisions we also considered what would happen if perceptions of the nine traits were not allowed to influence admission decisions. Specifically, variance common to both personality ratings and MIS was removed from the latter variable. The "residual" MIS, independent of the nine trait ratings, was converted to the same scale as the original MIS and used to calculate residual APPSCORES. Applicants were rank-ordered on both the original APPSCORE and on the residual APPSCORE variables and a Spearman rank-order correlation was computed. We then examined the original ranks and the new ranks to identify the number of applicants for whom removal of personality information would result in different admission decisions (i.e., change in First Choice status). This statistical process was used to simulate what might happen to admission decisions if we were able to prevent interviewers from considering their perceptions of these traits. Its purpose was to show that perceptions of personality may influence which applicants are admitted to medical school.

With respect to the third research question, are interviewers' perceptions of applicant personality reliable, estimates of interrater reliability were obtained by calculating Intraclass Correlation Coefficients (ICC). ICC are ratios wherein the variability in ratings that is due to targets (applicants) is compared to the sum of this variability plus variability due to "error" (e.g., variability among raters and other sources of measurement error). We computed ICC for the MIS as well as for the personality perception ratings.

Finally, we conducted exploratory nonlinear analyses between personality perception scores and MIS. This follows a study by Shen and Comrey² wherein evidence was reported for quadratic relations between personality scales and medical school performance. For example, the highest "overall evaluations" of medical school performance were for students at a moderate or balanced point in terms of emotional stability. Good physicians neither suppress affect nor have drastic mood swings.² The Shen and Comrey study serves as a basis for some renewed interest in the possibility of nonlinear relations in the personality domain — relations that have been somewhat elusive in past research. It is also possible that interviewers look for "red flags" or extreme scores on certain traits (e.g., impulsivity) when

deciding which applicants should pursue a career in medicine.¹² In our analyses, each personality perception score and its squared and cubed counterparts were entered in successive regression equations to test for linear, quadratic, and cubic components, respectively.¹³ However, we would like to highlight the exploratory and tentative nature of our nonlinear analyses.

Results

Missing Data - There were no missing interview score data for the 345 applicants under consideration for admission. However, interviewers' personality perception ratings of applicants were voluntary, resulting in some missing data. On average, 90% of the members of an interview panel rated a given applicant on a given personality trait. Nine applicants did not receive any personality ratings from interviewers. As described in the method section, interviewers' mean personality perception ratings for each applicant on each trait were to serve as personality perception scores in this research.

Three methods of operationalizing mean ratings were considered. Means were calculated for a given trait if (a) all three, (b) at least two of three, or (c) at least one interviewer provided a rating of an applicant on a trait. All three methods resulted in similar patterns of correlations with other variables. Therefore, the mean of any and all available ratings (1, 2, or 3) was used to form the personality perception scores used in the analyses described herein. This option makes maximal use of the available data, a consideration that is particularly important for the analyses involving residual variables (description forthcoming).

Do interviewers seem to make judgments about applicant personality? - We computed bivariate correlation coefficients between mean personality perception scores and other variables in the data set (Table 1 in the appendix). Correlations between personality perception scores and interview scores offer some evidence as to the extent to which interviewers may be considering their perceptions of applicants on various personality traits when assigning interview scores. For eight of the nine traits considered in this study, significant correlations were observed with MIS. Applicants seen as low on Abasement (-0.30), Aggression (-0.27), and Impulsivity (-0.19) received higher MIS. Applicants perceived as high on the traits of Nurturance (0.69), Achievement (0.64), Endurance (0.58), Cognitive Structure (0.54), and Order (0.54) received higher MIS. All of these relations were in the predicted direction. Only Dominance (0.07) did not correlate

significantly with MIS. Similar patterns of correlations were apparent for all interviewer sources.

Do personality perceptions affect admission decisions? - Table 1 shows a pattern of correlations between personality perception scores and the composite APPSCORE similar to the pattern observed with interview scores. These correlations suggest that interviewers' perceptions of applicant personality may affect admission decisions. To further explore these data we considered differences in personality perception scores of First Choice applicants (top APPSCORES) versus other applicants who had been interviewed. A multivariate analysis of variance (MANOVA) was conducted with the nine personality variables as dependent variables to address the question of whether there was some way of combining the personality variables to distinguish First Choice applicants from other, minimally qualified, applicants. The answer was affirmative (Pillai's trace = 0.344, $F(9, 326) = 18.995$, $p < .001$, partial eta-squared = 0.344). Given the significant multivariate effect, univariate analyses (t-tests) of between group differences with respect to each personality trait were conducted. Significant differences were observed between groups on mean personality ratings for all traits except Dominance (see Table 2).

Another approach to understanding the impact of interviewers' perceptions of applicant personality on admission decisions was to statistically remove personality perceptions from MIS, use the resulting "residual" MIS to calculate residual APPSCORES, and compare the residual APPSCORES to original APPSCORES. The Spearman correlation between the

rank-ordered APPSCORE variables was 0.59. While the residual APPSCORES co-vary with original APPSCORES to a moderate degree, it is also clear that some changes in applicants' rank ordering occur when personality perception scores are removed from MIS. The original ranks and the new (residual) ranks for individual applicants were examined to identify the number of applicants for whom removal of personality information would result in different admission decisions (i.e., change in First Choice status). Of the 336 applicants with personality ratings, we observed 44 applicants who had original APPSCORES that earned them a place among the First Choice applicants who would not have been in this select group if the residual APPSCORES were used as the basis for admission decisions. In other words, over 40% of the First Choice applicants would lose their place in the program if perceptions of these traits were explicitly excluded from the interview process.

Are interviewers' perceptions of applicant personality reliable? - Estimates of interrater reliability for the personality ratings and MIS were obtained by calculating Intraclass Correlation Coefficients (ICC) (see Table 1 in the appendix). All values reported in the diagonal of Table 1 are based on the average of 3 ratings. The reliabilities of ratings of personality were less than desirable (0.45 to 0.68) and were probably due to the use of single-item ratings and unclear rating criteria. For example, with respect to Cognitive Structure only 45% of the variance in interviewer's perception ratings was attributable to applicants with over half of the variance due to variability among raters and measurement error. The ICC for the composite interview score used by the Faculty

Table 2 Personality Perception Scores of First Choice versus Other Interviewees.

<u>Personality Trait</u>	<u>First Choice Applicants?</u>		<u>Partial Eta-Squared</u>
	<u>No</u>	<u>Yes</u>	
Abasement	2.03 _a (0.78)	1.66 _b (0.57)	0.05
Achievement	3.62 _a (0.68)	4.34 _b (0.45)	0.22
Aggression	1.41 _a (0.58)	1.20 _b (0.32)	0.03
Cognitive Structure	3.13 _a (0.69)	3.74 _b (0.52)	0.16
Dominance	1.99 _a (0.83)	2.05 _a (0.69)	0.00
Endurance	3.30 _a (0.75)	3.95 _b (0.61)	0.15
Impulsivity	1.94 _a (0.81)	1.72 _b (0.63)	0.02
Nurturance	3.09 _a (0.78)	3.90 _b (0.60)	0.20
Order	3.29 _a (0.64)	3.87 _b (0.62)	0.15

Notes. Tabled values refer to group means (standard deviations) of personality perception scores on a 1-5 scale, and effect size estimates. First Choice Applicants = Yes; N = 99. First Choice Applicants = No; N = 237. Means in the same row with different subscripts differ significantly (2-tailed, independent samples t-test, equal variances not assumed, $p < .05$).

as part of the admission process was more encouraging (0.82).

Nonlinear Exploration. - Table 3 shows the squared multiple correlations (R^2) obtained when personality perception scores were entered as predictors of MIS. Linear, quadratic, and cubic models were examined in successive steps. Of particular interest is the significance of the difference in R^2 values within a given row (i.e., for a given trait) denoted by letter subscripts in Table 3. For example, Abasement perception scores were related to MIS in a linear model, accounting for approximately 9% of the variance. Adding quadratic or cubic terms to regression equations did not increase the proportion of MIS variance explained. Although

proximately 50%. A relatively strong positive, linear relationship was evident between Nurturance perception scores and MIS. However, candidates perceived as particularly low (i.e., around 1 on the 5-point scale) on Nurturance received lower MIS than would be predicted by the linear model.

Perceptions of applicants' on Cognitive Structure also exhibited significant linear ($R^2 = 0.30$), quadratic ($\Delta R^2 = 0.01$), and cubic ($\Delta R^2 = 0.02$) relations with MIS. Being perceived as high on Cognitive Structure was associated with higher MIS up to a point (approx. 4 on the 5-point scale) after which MIS leveled off and then tended to decline slightly. Order ratings correlated positively with MIS in a linear

Table 3 Exploratory Nonlinear Analyses Predicting MIS from Personality Perception Scores

Squared Multiple Correlations

	Linear Model ($Y' = b_0 + b_1X$)	Quadratic Model ($Y' = b_0 + b_1X + b_2X^2$)	Cubic Model ($Y' = b_0 + b_1X + b_2X^2 + b_3X^3$)
Abasement	0.09 ^{***} _a	0.09 ^{***} _a	0.09 ^{***} _a
Achievement	0.41 ^{***} _a	0.41 ^{***} _a	0.42 ^{***} _b
Aggression	0.07 ^{***} _a	0.08 ^{***} _a	0.08 ^{***} _a
Cognitive Structure	0.30 ^{***} _a	0.31 ^{***} _b	0.33 ^{***} _c
Dominance	0.01 ^a	0.03 ^{**} _b	0.03 [*] _b
Endurance	0.34 ^{***} _a	0.34 ^{***} _a	0.35 ^{***} _a
Impulsivity	0.04 ^{**} _a	0.09 ^{***} _b	0.09 ^{***} _b
Nurturance	0.48 ^{***} _a	0.49 ^{***} _b	0.50 ^{***} _c
Order	0.29 ^{***} _a	0.30 ^{***} _b	0.30 ^{***} _b

Notes. Letter subscripts (a, b, c) refer to tests of significance of incremental prediction (change in squared multiple correlation) of higher powers over lower powers (i.e., test of significance of the regression coefficient for the highest power in a given equation). Squared multiple correlations in the same row with different subscripts differ significantly ($p < .05$). Asterisks refer to tests of significance for linear, quadratic, and cubic regression equations. * $p < .05$; ** $p < .01$; *** $p < .001$.

the more complex Abasement equations were still statistically significant, overall, the regression coefficients for the highest power in each of the quadratic and cubic models were not significant. That is, increasingly complex models did nothing to enhance prediction of MIS over a linear model.

Similarly, for most other trait perceptions linear models were quite successful. Nevertheless, for six of the nine traits, quadratic and/or cubic models accounted for significantly more variance in MIS than did linear models (Table 3). However, the mathematical maximization inherent in multiple regression and small effects requires that caution be used when interpreting "curve fitting" results. A linear model using Nurturance to predict MIS accounted for approximately 48% of the variance, while a quadratic model accounted for approximately 49%, and a cubic model

explained some additional variance in MIS ($\Delta R^2 = .01$). As Order ratings declined from approximately 2.5 to 1 observed MIS were lower than would be predicted by the linear model. Although the quadratic model ($\Delta R^2 = .00$) for Achievement did not account for significantly more variance than did the linear model ($R^2 = .41$), the cubic model did explain significantly more variance in MIS than the quadratic model ($\Delta R^2 = .01$). MIS tended to level off as Achievement ratings declined from approximately 2.5 to 1 and this was not detected by the linear or quadratic models.

Some of the more interesting results in terms of substantive interpretations and effect sizes involved perceptions of Dominance and Impulsivity. Dominance did not correlate significantly with MIS in a

linear model ($R^2 = .01$). However, adding a quadratic component resulted in a significant multiple regression equation ($R^2 = .03$) that explained an additional 2% of the variance in MIS ($\Delta R^2 = .02$). That is, there was a slight tendency for those with the highest perceived Dominance ratings (i.e., > 3.5) to receive lower Mean interview scores than would be predicted by the linear model. A quadratic model for Impulsivity ($\Delta R^2 = .05$) accounted for more than double the variance (9% rather than 4%) in MIS than did the linear model ($R^2 = .04$). This reflected a sharp drop in MIS as Impulsivity ratings increased from approximately 2.5 to 5. Nevertheless, perceptions of Impulsivity explained little MIS variance.

Discussion

Personal qualities are seen as important predictors of clinical performance in medicine.^{1,2,3} Also, encouraging meta-analytic results^{5,6} have helped spur renewed interest in personality-job performance relations. Pragmatic concerns regarding the use of structured inventories for selection¹ provide an impetus for research on alternative (non-questionnaire) methods of assessing personality. There is the possibility that selection interviews may assess, or could be designed to assess, various non-cognitive characteristics.^{1,3,7,8} This notion also receives some support from interpersonal perception research.⁴

This study investigated issues surrounding the assessment of personality via an interview designed to assess personal characteristics, but not formally designed to assess specific personality constructs. Panels of three interviewers, each composed of one physician, one community representative, and one medical student, were assembled from a pool of interviewers. Applicants interviewed by a given panel may have been rated differently had a different panel interviewed them, especially considering the latitude given to interviewers with respect to specific approaches to questioning. However, the use of mean ratings from multiple interviewers would be expected to reduce idiosyncratic interviewer effects.

Applicants perceived as high on Achievement, Cognitive Structure, Endurance, Nurturance, and Order were more likely to be admitted to a Doctor of Medicine Program while those perceived as meek, aggressive, and impulsive were less likely to be admitted. The finding that scores on a semi-structured interview correlated significantly with eight of the nine traits provides some evidence that interviewers may be assessing applicant personality in the interview. Alternatively, interviewers may have provided personality ratings consistent with a "desirable applicant" stereotype. Regardless of which interpretation

is more appropriate, interviewers seem to think that personality is an important consideration in medical school admissions. The construct and predictive validity of personality perceptions were not assessed but are important considerations for future research. Reliability sets an upper bound for validity and in that regard there is room for improvement. However, use of single item rating scales in this research suggests that substantial improvements may be within reach (e.g., through the use of multi-item scales).

Statistically removing perceptions of these nine traits from MIS resulted in different admission decisions for over 40% of the applicants. The statistical procedures used to remove personality perceptions demonstrate that individual admission decisions are likely affected by interviewers' perceptions of applicant personality. It is unlikely that one would try to *prevent* perceptions of these traits from affecting the medical admissions process. Indeed, personal qualities are seen as important, so much so that considerable effort and expense is devoted to interviewing applicants. Having shown that personality perceptions may influence which applicants are given the opportunity to study medicine, we must consider what effect *better* measurement of these constructs would have on admission decisions. The most pertinent question is not what would happen if personality were removed from the admission process, but would better measurement of relevant constructs result in the admission of different people to medical school – people who might perform better therein, and in clinical settings as interns, residents, and practicing physicians? Better measurement of traits that are predictive of performance criteria could have important consequences for individuals (i.e., applicants) as well as for their patients and colleagues.

Assessment procedures must elicit from applicants relevant information and be scored appropriately to predict criteria of interest.

"The employment interview can be a uniquely fertile source of behavioral information from which to draw inferences about job candidates' personality functioning. Of course, this uniqueness can be both a strength and a weakness. Poorly designed interviews may gather such a smattering of behavioral information that they do not allow effective assessment of any meaningful psychological variables. The challenge is to design interviews that are coherent behavior samples from which to draw valid, job-related inferences"⁴ (p. 108).

Accumulated wisdom advises that one begin with a job analysis (e.g., analysis of successful medical students)⁷ and identify important worker characteristics, including personality requirements.⁴ Questions must then be designed to elicit behavioral information that serve as a valid basis for drawing inferences about the target traits. Behavioral information must be available for observation, diagnostic of target traits, and detected and correctly used by interviewers in order for personality ratings to be accurate.⁴ A high quality scoring system is also required to obtain reliable and accurate assessments for making important decisions.

We observed a reasonably high interrater reliability for the trait of Nurturance. This trait was likely perceived as an important consideration by the interviewers due to the nature of the work domain and was reinforced in several sections of the interviewer manual. For example, empathy was one of several personal qualities listed in a section dealing with the selection of students. It was also discussed under the rubric of communication skills in terms of active listening to patients' concerns and the importance of sincere attention and caring in that regard. Traces of this theme were also evident elsewhere in the manual (e.g., "is able to use personal strengths to encourage healing and provide moral support for patients"). In other words, interviewers were likely to carefully *solicit* and *attend to* information pertaining to candidates' standing on Nurturance. Further development of rating scales to increase the clarity of rating criteria should lead to more reliable and valid assessment of target constructs.

This study simulated important aspects of the process for obtaining quality, job-relevant assessments (e.g., information about the target job and personality requirements informed the selection of traits). Questions in the existing semi-structured interview were deliberately unaltered in order to investigate the independent contribution of personality judgments to the selection decision. Personality ratings reflected pragmatic constraints such as obtaining only one rating of each personality construct, which likely had a negative impact on the psychometric quality of these ratings. The success of this study suggests that it may be worth the effort to follow closely recommended assessment procedures to achieve even better results.

There are a number of areas where accumulated wisdom may need to partner with innovative research. Patrick and colleagues⁹ (p. 67) provided the following example of a question used in a structured

medical admissions interview: "Thinking back over the past few years, what is one experience you have had that influence or changed your life in a significant way." It is possible that such a question may tap a number of constructs depending on the experiences described by various candidates but does not promote clear assessment of any particular construct(s). Improvements in structured interviewing may be realized by paying greater attention to important, job-related constructs to be assessed rather than designing questions in isolation of such substantive considerations. This represents an alteration of typical structured interviewing techniques.

It may be advantageous to use construct definitions and descriptive adjectives of job-relevant traits from the personality literature to identify relevant traits (i.e., those worth pre-testing), design interview questions, and construct scoring procedures. Various structured interviewing strategies (e.g., behavioral interviews; situational interviews; work sample-type questions) could help to elicit information relevant to job-related trait constructs. For example, to assess Nurturance, a patterned behavior question might ask candidates to describe a situation in the past where they helped or consoled a sick or injured person. A situational interview question might describe a hypothetical scenario where candidates are faced with a dilemma and the "correct" answer is not obvious (e.g., to console a patient vs. to engage in another socially desirable, but less relevant, behavior). A work sample-type question could assess empathetic communication style by having candidate's convey news of a terminal illness to one of the interviewers in a role-playing scenario. Another possibility would be to use a role-playing scenario to assess the candidate's active and empathetic listening behavior.

Once questions are designed, structured interview scoring keys (see Patrick et al.,⁹ for a typical example) could be developed by considering construct definitions, descriptive adjectives, and questionnaire items. Scoring keys and rating scales may be designed to facilitate the use of elicited information to accurately assess job-relevant traits. Multiple ratings of particular items (e.g., separate ratings of several descriptive adjectives) could be aggregated to further increase the reliability of assessments.

The use of highly structured interviews for personality assessment is one possibility worthy of additional research. However, Binning et al.⁴ suggested that a moderate degree of interview structure (i.e., probing interviews) may be optimal for personality assessment, especially for the assessment of source traits (i.e., motives; emotions) rather than surface

traits (i.e., observable; interpersonal), and for assessing “dark” side personality characteristics that are difficult to assess with typical assessment procedures. Similarly Tutton,³ citing moderate correlations between CPI scores and interview ratings in that study, suggested that an inventory might be a useful adjunct to the process. The appropriate degree of structure, relative utility of assessing source vs. surface traits, and use of interviews and inventories alone or in combination are issues that await future research. Based on this study we suggest that interviews may be used to assess personality and that future research can improve the validity of interviews and provide information as to the constructs that are or that can be assessed effectively via interviews.

Our exploratory analyses of nonlinear relations revealed that linear models of personality perception and interview score relations are generally appropriate. Small but significant improvements in prediction may be obtained with more complex nonlinear models. Moreover, the nonlinear effects observed in this study also seem interpretable; post hoc (e.g., those perceived as extremely low on Nurturance tend to receive particularly low MIS). Along with the results of Shen and Comrey² these results lead us to suggest that researchers form and test a priori hypotheses with respect to how personality traits might relate to interview ratings or clinical performance measures.

Acknowledgments

We thank the U.W.O. Doctor of Medicine Admission Committee members and staff who made this research possible. In particular, we recognize Dr. Jim Silcox, Associate Dean – Admissions/Student & Equity Affairs, and Dr. Bertha Garcia, Admission Committee Chair, for their support.

Preparation of this article was supported by a grant (R2192AD6) from the Social Sciences and Humanities Research Council of Canada to Mitchell G. Rothstein.

References

1. Powis D.A. Selecting medical students. *Med Educ.* 1994;28:441-69.
2. Shen H., Comrey, A.L. Predicting medical students' academic performances by their cognitive abilities and personality characteristics. *Acad Med.* 1997;72:781-86.
3. Tutton, P.J.M. Medical school entrants: semi-structured interview ratings, prior scholastic

- achievement and personality profiles. *Med Educ.* 1993;27:328-36.
4. Binning J.F., LeBreton, J.M., Adorno A.J. Assessing personality. In: Eder R.W., Harris, M.M. (eds). *The employment interview handbook.* Thousand Oaks, CA: Sage Publications, 1999:105-23.
5. Tett, R.P., Jackson, D.N., Rothstein, M. Personality measures as predictors of job performance: A meta-analytic review. *Personnel Psychology.* 1991;44:703-42.
6. Tett, R.P., Jackson, D.N., Rothstein, M., Reddon, J.R. Meta-analysis of bidirectional relations in personality—job performance research. *Human Performance,* 1999;12:1-29.
7. Edwards J.C., Johnson E.K., Molidor J.B. The interview in the admission process. *Acad Med.* 1990;65:167-77.
8. Schmitt, N. The current and future status of research on the employment interview. In: Eder R.W., Harris, M.M. (eds). *The employment interview handbook.* Thousand Oaks, CA: Sage Publications, 1999:355-67.
9. Patrick, L.E., Altmaier, E.M., Kuperman, S., Ugonlini, K. A structured interview for medical school admission, phase 1: Initial procedures and results. *Acad Med.* 2001;76:66-71.
10. Jackson, D.N. *Personality Research Form Manual* (3rd ed.). Port Huron, MI: Sigma Assessment Systems, Inc., 1999.
11. Costa, P. T., Jr., & McCrae, R. R. From catalog to classification: Murray's needs and the Five-Factor Model. 1988;55:258-65.
12. Solomon, D. J. Personal communication. August, 2002.
13. Gardner, R. C. *Psychological statistics using SPSS for Windows.* Upper Saddle River, NJ: Prentice Hall, 2001.

Author Notes

Mr. Jelley is a Ph.D. candidate in the industrial and organizational (I/O) psychology program at the University of Western Ontario (U.W.O).

Jelley RB, Parkes MA, Rothstein MG. Personality perceptions of medical school applicants.

Med Educ Online [serial online] 2002;7:11. Available from URL <http://www.med-ed-online.org>

Lt. Cmd. Parkes is an officer in the Royal Canadian Forces and a graduate of U.W.O.'s masters program in I/O psychology.

Dr. Rothstein is an Associate Professor of Organizational Behavior and Director of the Ph.D. Program at the Richard Ivey School of Business, U.W.O.

Correspondence

Dr. Mitchell G. Rothstein
Richard Ivey School of Business
The University of Western Ontario
London, Ontario
CANADA, N6A 3K7

Telephone: (519) 661-3298
Fax: (519) 661-3485

E-mail: mrothstein@ivey.uwo.ca

Appendix

Table 1 Descriptive Statistics, Pearson Correlations, and Interrater Reliability Estimates.

	Mean	S.D.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
1 APPSCORE	80.80	4.77	--																				
2 First Choice?			.75	--																			
3 Residual APP-SCORE	80.88	4.53	.61	.44	--																		
4 MIS	74.43	12.91	.93	.66	.52	(.82)																	
5 Residual MIS	74.43	12.74	.46	.29	.92	.53	--																
6 Physician IS	73.99	15.12	.84	.59	.45	.88	.44	--															
7 CRIS	74.63	14.69	.79	.58	.48	.87	.51	.64	--														
8 MSIS	74.66	14.30	.82	.56	.45	.88	.45	.68	.65	--													
9 PPS Abasement	1.92	0.74	-.31	-.23	-.04	-.30	.00	-.21	-.30	-.28	(.55)												
10 PPS Achievement	3.83	0.70	.62	.47	.05	.64	.00	.60	.53	.55	-.33	(.63)											
11 PPS Aggression	1.35	0.53	-.26	-.18	-.01	-.27	.00	-.24	-.23	-.25	-.14	.05	(.55)										
12 PPS Cognitive Structure	3.31	0.70	.55	.40	.06	.54	.00	.50	.43	.50	-.11	.69	.07	(.45)									
13 PPS Dominance	2.01	0.79	.08	.04	.02	.07	.00	.08	.07	.04	-.29	.33	.64	.29	(.62)								
14 PPS Endurance	3.49	0.77	.56	.39	.04	.58	.00	.52	.49	.52	-.25	.79	.08	.65	.33	(.58)							
15 PPS Impulsivity	1.87	0.76	-.19	-.13	-.02	-.19	.00	-.13	-.16	-.21	-.07	.11	.53	.01	.51	.18	(.56)						
16 PPS Nurturance	3.33	0.82	.62	.45	.00	.69	.00	.65	.55	.61	.00	.39	-.40	.33	-.18	.38	-.21	(.68)					
17 PPS Order	3.46	0.68	.52	.39	.04	.54	.00	.49	.45	.47	-.09	.56	-.12	.66	.07	.51	-.20	.37	(.36)				
18 GPA (best year)	3.84	0.10	.24	.18	.16	.04	-.05	.02	.02	.07	-.05	.05	-.09	.10	-.03	-.01	-.13	-.01	.11	--			
19 MCAT verbal	10.06	1.03	.33	.36	.26	.06	-.02	.09	.03	.04	-.05	.09	.02	.11	.04	.11	.00	.03	.06	.06	--		
20 MCAT writing	11.63	0.81	.23	.22	.21	.02	-.01	.07	-.03	.00	-.07	.08	.00	.05	.07	.05	.02	-.05	.02	.03	.09	--	
21 MCAT Phys Sci	10.79	1.68	.06	.09	.06	-.07	-.07	-.05	-.05	-.09	.02	.00	.02	.05	-.02	-.09	-.13	-.09	.04	.29	.23	.10	--
22 MCAT Bio Sci	10.98	1.33	-.03	.04	.00	-.11	-.07	-.09	-.07	-.15	.11	-.06	-.02	.00	-.07	-.06	-.07	-.10	.00	.16	.16	.04	.53

Notes. N = 345. Means, standard deviations, and bivariate correlations involving all personality and residual variables are based on N = 336. Decimals omitted from Pearson and Intraclass correlations. Pearson correlations with absolute values of 11 (or greater) are significant at $p < .05$ and those with values of 15 (or greater) are significant at $p < .01$. Interrater reliability estimates are Intraclass Correlation Coefficients (ICC; one way random effects model) for the mean of 3 ratings.

APPSCORE = composite application score; MIS = mean interview score; Physician IS = Interview score assigned by a Physician; CRIS = Interview score assigned by a community representative; MSIS = Interview score assigned by a medical student; PPS = Personality perception score; GPA = Grade point average; MCAT = Medical College Admission Test.