The Effect of Blinded Versus Nonblinded Interviews in the Resident Selection Process

Lois E. Brustman, MD
Fern L. Williams
Katherine Carroll, BA
Heather Lurie, MD
Eric Ganz, MD
Oded Langer, MD, PhD

Abstract

Objective To evaluate whether resident applicants' academic performance biases the assessment of nonacademic qualities.

Methods In this prospective, descriptive study, 2 blinded (personal statement only) and 1 nonblinded (application) 30-minute interviews were compared for candidates ranking into Top 10, Upper Third, Middle Thirds, Lower Third, and Do Not Rank classes.

Results A total of 234 candidates were interviewed from 2005 to 2007. The association between blinded interviewers for the categories was 87%, 63%, 68%, 73%, and 90% ($P = .0000$), respectively. Comparing blinded to nonblinded interviewers showed an association of 75% (63%), 71% (86%), 68% (58%), 66% (79%), and 72.7% (82%) ($P = .0000$), respectively. A strong degree of agreement (Cohen κ, 0.75) for the 2 ranking scores resulted in 90% agreement for Top 10 and Upper Third and 85% for Middle Third and Lower Third categories. No correlation was found between United States Medical Licensing Examination scores and final ranking; moderate agreement was found between ranking and deans' letters (Cohen κ, 0.59, $P = .0000$).

Conclusion Candidate rankings on nonacademic attributes were not affected by interview type.

Introduction

Selecting residents is a demanding yet critical part of the responsibilities of the program director and faculty in any residency program. It is a difficult process, as the future success of a medical student in residency has not been shown to be predicted by academic performance in medical school alone.[1–4]

It is common practice to use the components of the Electronic Residency Application Service (ERAS) application in an attempt to evaluate the prospective applicants' cognitive and noncognitive characteristics. For example, objective data, such as the candidates' United States Medical Licensing Examination (USMLE) scores and medical school transcripts, can be used to judge medical knowledge. However, the qualities of a person's character, such as interpersonal and communication skills, and his or her propensity to exhibit professionalism, are noncognitive and therefore not as well reflected in the ERAS application.[5] Recommendation letters, the personal statement, and the curriculum vitae are subjective information, for they are initiated by the candidate and have the potential to be biased. Therefore, since there is overall agreement that success in residency is dependent upon a candidate who excels in both academic and interpersonal skills, most residency programs incorporate a personal interview in the selection process. This is a very time-consuming part of the process, but it is a direct method to attempt to evaluate a residency candidate's interpersonal and communication skills and other nonacademic attributes.

Regardless of the type of interview used in the residency selection process, the interviewer has access to the candidate's application prior to or during the interview. It has been suggested that knowledge of an applicant's academic performance may bias an interviewer in regard to noncognitive traits.[6–8] This is thought to create a "halo effect," where interviewers are overly influenced by a favorable or unfavorable trait that affects their judgment.[6,7] Attributes such as interpersonal and communication skills, leadership, and motivation have been shown to be better evaluated at an interview if academic performance is de-emphasized. Therefore, it has been suggested that to best utilize the significance of the interview, interviewers should be blinded to all academic and other objective data at the time of the interview.

The purpose of this study was to evaluate whether a nonblinded interview biased the interviewer's evaluation of

All authors are in the Department of Obstetrics and Gynecology, St. Luke's-Roosevelt Hospital Center, Columbia University, New York, NY. **Lois E. Brustman, MD,** is Program Director; **Fern L. Williams,** is Program Coordinator; **Katherine Carroll, BA,** is MFM Administrative Secretary; **Heather Lurie, MD,** is Attending Physician; **Eric Ganz, MD,** is Attending Physician; and **Oded Langer, MD, PhD,** is Chairman.

Corresponding author: Lois Brustman, MD, Department of Obstetrics and Gynecology, St. Luke's-Roosevelt Hospital Center, 1000 Tenth Avenue, Suite 10C, New York, NY 10019, 212.523.8366, lbrustma@chpnet.org

Received: March 23, 2010; revision received June 18, 2010; accepted: July 9, 2010.

DOI: 10.4300/JGME-D-10-00051.1

the candidate's subjective, noncognitive performance when compared to the evaluation of a blinded interviewer.

Methods

Study Design

This was a prospective, descriptive study of the process of resident selection in an obstetric and gynecology program over a 3-year period (2005–2007).

Application Review

All applications from American, international, and osteopathic medical schools submitted through ERAS each year were reviewed by the program director. A select group of resident applicants were invited for an interview after their applications were screened on the basis of USMLE scores, minimal score on USMLE of ≥200, medical school transcripts, and successful completion of all course requirements, dean's letters, personal statements, curriculum vitae, and letters of recommendation. Each year, approximately 80 applicants were interviewed, and ultimately 65 to 70 were ranked.

Training of Interviewers

Each year, all interviewers participated in a 1-hour mandatory training session. A total of 35 interviewers were trained, consisting of 27 faculty members and 8 residents (postgraduate year-3 [PGY-3] and PGY-4). Selected faculty had been practicing at least 5 years or were prior residents in the program.

At the training session, the rationale behind the *blinded interview* (interviewer only had access to the candidate's personal statement prior to or during the interview) and the *nonblinded interview* (interviewer had full access to the application) was reviewed. The goals for the interview process were defined in relation to the type of individuals that were to be recruited. The goals were good interpersonal and communication skills, professionalism (ie, integrity, sensitivity to patients and family), and their potential to be a team player. Each interviewer was given a packet with guidelines for interviewing resident applicants. The packet included a description of the selection process, examples of questions that should and should not be asked, and a copy of the evaluation sheet. In addition, examples of open-ended questions that could be used to elicit the desired traits of the candidates were distributed and reviewed. It was stressed that USMLE scores and grades should not be discussed.

Interview Process

During November to January each year, there were 4 interview days when approximately 20 to 25 candidates were invited. All residency candidates had three 30-minute interviews; 2 were blinded and 1 was nonblinded.

The blinded interview was performed by faculty and selected PGY-3 and PGY-4 residents; these interviewers only had access to the applicant's personal statement. The nonblinded interview was conducted by the chairman, program director, or vice chair/assistant program director, who had full access to the candidate's application. All interviewers used the same evaluation sheet on which the applicants were rated by the interviewer in categories of selected traits on a 1 to 10 basis (1 = worst, 10 = best) and each candidate was ranked at the end of an interview session as Top 10, Upper Third, Middle Third, Lower Third, or Do Not Rank (DNR). Resident applicants were assigned by the program coordinator to interviewers; students who had done an elective at our institution were assigned to interviewers who did not know them by interviewers who knew them.

At the end of the interview day, there was a required meeting for all interviewers (both blinded and nonblinded) during which each prospective resident applicant was presented and discussed. The interviewers' evaluation sheets were collected prior to the start of this meeting. Each candidate's ERAS application was summarized in regard to USMLE scores, synopsis of the dean's letter, transcript, curriculum vitae, and letters of recommendation; the total application was also available for review by all present. Interviewers were given 5 minutes each to present their resident applicant and defend their ranking. A final ranking score was agreed on for each candidate (Top 10, Upper Third, Middle Third, Lower Third, and DNR) based on a consensus of the rank given by each interviewer at the time of their interview.

Statistical Analysis

The inter- and intra-observer variability between the interviewers and the type of interview was compared. For certain statistical analyses, candidates were regrouped into Group I (Top 10 and Upper Third) and Group II (Middle Third, Lower Third, and DNR). The deans' letters were categorized by the program director at the time the candidates' applications were reviewed for interview selection. They were categorized as excellent, average, or nonsupportive based on the following criteria: (1) perceived familiarity with the candidate, (2) academic record (ie, medical, surgical, and obstetric and gynecology clerkship), (3) association between the transcript information and the dean's letter, and (4) the summary paragraph (''key words,'' class rank). Categorical data were analyzed with the χ^2 and Fisher exact test, while numerical data were analyzed with the Student *t* test. Degree of agreement was determined using a Cohen κ test and Pearson correlation test.

Results

A total of 1496 ERAS applications were prescreened by the program director, resulting in a total of 234 candidates interviewed for 6 PGY-1 positions each year over the 3-year study. Eighty-eight percent of the candidates were from American medical schools and 12% were graduates of international or osteopathic medical schools. The mean

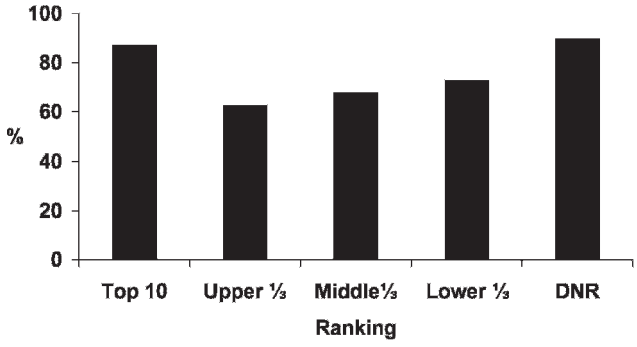

FIGURE 1 **Association Between the Blinded Interviewers to Individual Ranking of Candidates**

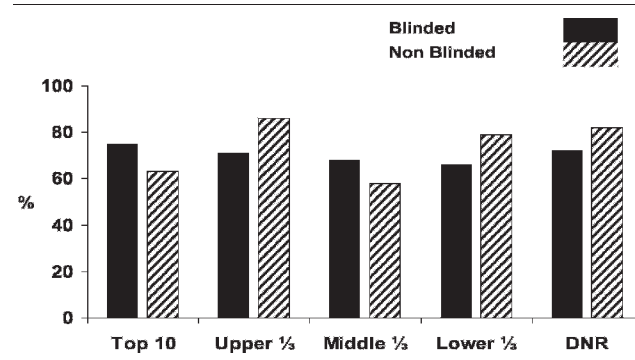

FIGURE 2 **Relationship of the Ranking of Candidates Between the Blinded and Nonblinded Interviewers**

USMLE Step 1 scores were similar between the American graduates when compared to the international and osteopathic school graduates (214 ± 15 vs 214 ± 17, respectively). In addition, in candidates for whom the USMLE Step 2 results were available prior to the interview, the scores were similar between the American medical school graduates and the combined scores of the international and osteopathic candidates (220 ± 16 vs 221 ± 22, respectively). There was a minimal increase in the average USMLE scores over the 3 years of the study, but this was not statistically significant (USMLE Step 1: 211 ± 14 in 2005 to 216 ± 12 in 2007; USMLE Step 2: 215 ± 17 in 2005 and in 2007, 225 ± 14).

The intra- and intervariability between the types of interviews and candidate ranking was studied. The association between the 2 blinded interviewers for the Top 10, Upper Third, Middle Third, Lower Third, and DNR, was 87%, 63%, 68%, 73%, and 90%, respectively ($P = .000$) (FIGURE 1). In comparison, there was less of an association in the ranking of the 5 categories consisting of Top 10, Upper Third, Middle Third, Lower Third, and DNR when blinded interviewers 1 and 2 were compared to nonblinded interviewers: 75% (63%), 71% (86%), 68% (58%), 66% (79%), and 72.7% (82%), respectively ($P = .000$) (FIGURE 2). Further analysis was done after the data were restratified into 2 groups. Group I consisted of the Top 10 and Upper Third and Group II consisted of the Middle Third, Lower Third, and DNR categories. This analysis revealed a strong degree of agreement (Cohen κ of 0.75, $P = .000$) for the ranking of Blinded Interview I to the nonblinded interview, resulting in an 89% agreement for the combined Top 10 and Upper Third categories and 85% agreement for the lower (Middle Third, Lower Third, and DNR) categories. Similar results were found when Blinded Interview II was compared to the nonblinded interviewer, with 87% agreement in Group I (Top 10 and Upper Third) and 90% for Group II (Middle Third, Lower Third, DNR; Cohen κ of 0.77, $P = .000$) In addition, when the ranking scores of all blinded interviewers were compared to those of the nonblinded interviewers, the strong degree of agreement was persistent, measured by a Cohen κ of 0.77 (Group I, 81% agreement; Group II, 95% agreement).

Twenty-five percent of the deans' letters were categorized as excellent, 43% were considered average, and 32% were considered nonsupportive. Analysis of the data revealed that only 38% and 35% of the excellent letters were for candidates in the Top 10 and Upper Third categories, respectively. The letters that were classified as average were only moderately associated with the ranking categories, ranging from 13% to 35%. In addition, 90% of the nonsupportive letters were in the Middle Third, Lower Third, and DNR categories (FIGURE 3).

Further analysis revealed only a moderate degree of agreement between a candidate's dean's letter (classified as

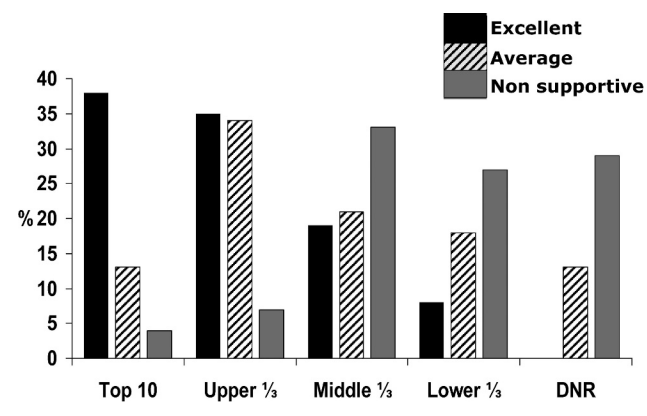

FIGURE 3 **The Association of the Assigned Deans' Letter Categories to the Ranking of the Resident Candidates**

Candidates' deans' letters were classified as excellent, average, or nonsupportive and then compared to the final ranking of the candidates, revealing a moderate degree of agreement, with a Cohen κ score of 0.36 ($p = .000$).

excellent, average, or nonsupportive) when compared to the ranking of the candidates into Group I (Top 10 or Upper Third), and Group 2 (Middle Third, Lower Third, and DNR) with a Cohen κ score of 0.36 ($P = .000$).

Finally, there was no association found between the candidate's final ranking and the scores on the USMLE Step 1 and Step 2 exams, the years of experience of the interviewers, and the type of interview (blinded or nonblinded). In contrast, the correlation coefficient between the final ranking score and the nonblinded interviewers was 0.90 ($P = .000$) and, for the blinded interviewer, the final score was 0.88 ($P = .000$).

Discussion

We found that there appears to be no effect on the evaluation of a resident's noncognitive skills when the interviewers are blinded or nonblinded to the applicant's cognitive achievements. In our selection process, a nonblinded interviewer does not seem to be biased by access to the candidate's application during the interview.

As early as 1979, program directors from different subspecialties revealed in a survey that the personal interview was the most important part of the residency selection process, allowing them to evaluate both cognitive and noncognitive skills.[9] Of note, this finding was reported in obstetrics and gynecology programs, where noncognitive skills are stressed. In general, the interview helps to identify the prospective residents with the greatest potential to successfully master residency, in addition to having the best fit with the program. Other purposes of the interview are for verification of information and recruitment, due to the competitive nature of programs trying to recruit the best candidates in areas that are either isolated or in cities where there are many programs from which to choose. The importance of the interview in the residency selection process is clear; however, there is a paucity of medical literature on attempts to standardize the reliability and reproducibility of the personal interview. Our residency, like others, has traditionally placed a strong emphasis on choosing candidates who have good interpersonal and communication skills, will be team players, and exhibit professional behavior. Therefore, the personal interview is overall the most important part of the residency selection process.[9,10]

The educational and psychology literature over the years has described various interview techniques.[11–14] Interviews can be either unstructured, structured, or semi-structured. An unstructured interview is commonly used in selection of residency candidates where 1 or more interviewers ask random questions based on the candidate's application. Another type of interview is the structured interview, which must meet all of the following criteria: (1) content is developed from job analysis; (2) questions are standardized, with the same questions asked of every applicant; (3) sample questions are provided to the interviewers; and (4) the interview is conducted by a board of interviewers. Semi-structured interviews meet only some of these criteria. Edwards et al[11] demonstrated that for medical school admission, first structured, then semi-structured interviews are the most valued and reliable interview techniques, while the unstructured interview falls behind in reproducibility and reliability.

Academic performance can bias interviewers' perception of noncognitive attributes of residency applicants.[6,7] Several studies have shown that favorable or unfavorable traits reported in a candidate's application can bias an interviewer, creating a "halo effect"[6–8] and thereby impair the judgment of the interviewer when evaluating the candidate for nonrelated characteristics. Further, negative aspects affect decisions more than favorable ones do.[11] To avoid this "halo effect," interviewers in these studies were "blinded" to the academic performance of the resident candidate prior to and during the interview. Robin et al[7] showed that academic performance favorably influenced the interview score from an interviewer not blinded to academic credentials when compared to the score given by the blinded interviewer in a surgical residency program. In 2001, Smilen et al[6] conducted a study to determine whether USMLE scores influenced interviewers in the process of selecting obstetric and gynecology resident candidates. The study was conducted over a 2-year period during which interviewers were privy to the applicants' scores in the first year of the study, but were blinded in the second year. The study revealed a statistically significant correlation between interview scores and USMLE Step 1 scores when the interviewers knew these grades and demonstrated no correlation when they did not, suggesting that to best utilize the significance of the interview, interviewers should be blinded to all academic and other objective data at the time of the interview.

In contrast to Smilen et al,[6] we found that no difference existed between blinded and nonblinded interviews. Miles et al[8] interviewed resident candidates at 2 surgical programs using a blinded and nonblinded interview technique and had similar results to our study. One of the institutions in that study indicated that academic credentials and medical knowledge were the most significant attributes in residency selection, while the other, like ours, considered noncognitive personal qualities to be more important. At the institution that deemed personal qualities to be the goal of the interview process, there was no difference in the interviewers' ranking, whether they were blinded or nonblinded. This result differed from the program that stressed academic qualities, where the nonblinded interviewer did appear to be biased by academic criteria. The authors attributed their finding to the fact that a program's philosophical approach to the residency selection process determined whether a blinded interview would influence candidate selection. In our study, we created an applicant screening procedure that attempted to evaluate

resident candidates based not only on their academic performance, but placed equal emphasis on evaluating their interpersonal communication skills and professionalism. As opposed to Miles et al, we used 2 teams of interviewers in the same institution, 1 that was blinded and 1 that was nonblinded. Our results revealed that no difference exists in candidate ranking between the blinded and the nonblinded interviewer, as well as between the blinded interviewers, resulting in minimal intra- and intervariability. In addition, there was a strong degree of agreement between the different interview types and the ranking of the candidates (Cohen κ score of 0.77, $P = .000$). The range of interviewers (residents to faculty, with different levels of experience) did not have any impact on candidate ranking, regardless of the type of interview conducted. It is possible that the structured, goal-oriented training process for all interviewers, emphasizing the importance of noncognitive traits, may be the explanation for our findings. In addition, the screening of the applications by the program director prior to the interview process may eliminate the need for the nonblinded interviewer to feel the necessity to scrutinize the academic component of the application. This would allow the interviewers to be more focused on the nonacademic traits of the applicants and could account for the fact that they were not biased by the knowledge of the components of the application. Finally, no association between USMLE Step 1 or Step 2 scores and deans' letters was noted in the final ranking of the candidates.

The importance of the residency selection process can not be underestimated. We are in search of resident candidates that not only have the potential to excel in academic competencies but also those who will master the more subjective educational goals such as interpersonal and communication skills, professionalism, and patient care. We believe that the best way to evaluate these qualities in a residency candidate is through the personal interview. If candidates are preselected and interviewers are well trained in the goals of the process, it is probable that the interviewers will not be biased by the components of the ERAS application.

References

1. Bell JG, Kanellitsas I, Shaffer L. Selection of obstetrics and gynecology residents on the basis of medical school performance. *Am J Obstet Gynecol*. 2002;186:1091–1094.
2. Dawkins K, Ekstrom, Maltbie A, Golden R. The relationship between psychiatry residency applicant evaluations and subsequent residency performance. *Acad Psychiatry*. 2005;29:69–75.
3. Borowitz SM, Saulsbury FT, Wilson WG. Information collected during the residency match process does not predict clinical performance. *Arch Pediatr Adolesc Med*. 2000;154(3):256–260.
4. Warrick SS, Cumrine RS. Predictors of success in an anesthesiology residency. *J Med Educ*. 1986;61(7):591–595.
5. Papp KK, Polk HC, Richardson JD. The relationship between criteria used to select residents and performance during residency. *Am J Surg*. 1997;173(4):326–329.
6. Smilen SW, Funai EF, Bianco AT. Residency selection: should interviewers be given applicants' board scores? *Am J Obstet Gynecol*. 2001;184(3):508–513.
7. Robin AP, Bombeck CT, Pollak R, Nyhus LM. Introduction of bias in residency—candidate interviews. *Surgery*. 1991;110(2):253–258.
8. Miles WS, Shaw V, Risucci D. The role of blinded interviews in the assessment of surgical residency candidates. *Am J Surg*. 2001;182:143–146.
9. Wagoner NE, Gray GT. Report on a survey of program directors regarding selection factors in graduate medical education. *J Med Educ*. 1979;54(6):445–452.
10. Scherl SA, Lively N, Simon MA. Initial review of Electronic Residency Application Service charts by orthopaedic residency faculty members. Does applicant gender matter? *J Bone Joint Surg Am*. 2001;83-A(1):65–70.
11. Edwards JC, Johnson EK, Molidor JB. The interview in the admission process. *Acad Med*. 1990;65(3):167–177.
12. Janz T. Initial comparisons of patterned behavior description interviews versus unstructured interviews. *J Appl Psychol*. 1982;67:577–580.
13. Pursell ED, Campion MA, Gaylord SR. Structured interviewing: avoiding selection problems. *Pers J*. 1980;59(11):907–912.
14. Powis DA, Naem RLB, Bristow T, Murphy LB. The objective structured interview for medical student selection. *Br Med J (Clin Res Ed)*. 1988;296:765–768.