

# Perspectives in Radiologic Education

ROBERT A. NOVELLINE, MD, *Editor*

## The Predictive Utility of Behavior-Based Interviewing Compared with Traditional Interviewing in the Selection of Radiology Residents

ELIZABETH M. ALTMAIER, PhD,\* WILBUR L. SMITH, MD,† COLLEEN M. O'HALLORAN, BA,\* AND E.A. FRANKEN, JR., MD†

Altmaier EM, Smith WL, O'Halloran CM, Franken EA Jr. The predictive utility of behavior-based interviewing compared with traditional interviewing in the selection of radiology residents. *Invest Radiol* 1992;27:385-389.

**RATIONALE AND OBJECTIVES.** This study compares the predictive use of measures based on traditional faculty and resident interviews of residency applicants with measures obtained through behavior-based interviewing. A special emphasis was placed on predicting residents' noncognitive abilities.

**METHODS.** One hundred fifty-one resident applicants, over a 3-year period, were interviewed using standard interviews by faculty and residents. These residents also were interviewed with an experimental behavior-based accomplishment interview. Four years later, during their diagnostic radiology residency, evaluations of performance were gathered on these applicants from their residency director.

**RESULTS.** Results indicated that scores based on responses given during the accomplishment interviews added considerable predictive utility to the low prediction demonstrated by traditional interviews.

**CONCLUSIONS.** These findings imply that improving unstructured faculty and resident interviews to obtain, in a more rigorous manner, desired information about noncognitive abilities may be a key to successful resident selection.

**KEY WORDS:** predictive utility; residents, interviews.

---

From the \*College of Education and Center for Health Services Research, the †Department of Radiology, College of Medicine, The University of Iowa, Iowa City, IA.

Supported by a grant from the Iowa Measurement Research Foundation (Dr. Altmaier).

Reprint requests: Elizabeth M. Altmaier, PhD, N459 Lindquist Center, The University of Iowa, Iowa City, IA 52242.

Received June 17, 1991, and accepted for publication, after revision, November 27, 1991.

RESEARCHERS AND EDUCATORS have long attempted to define selection measures that would adequately predict residency performance in medical specialties. Most often, the same type of measures that predicted performance in medical school, such as objective measures of academic achievement, have been used. In fact, program directors in other specialties have reported an increased use of the academic predictors of National Board of Medical Examiners (NBME) Part I scores, grades in clerkships, and class rank in evaluating prospective residents.<sup>1</sup> Unfortunately, this increasing reliance on academic achievement predictors ignores data that consistently show a lack of predictive use for these measures. The failure of academic data to predict later performance has been extensively demonstrated, ranging from evaluations of grades<sup>2</sup> to test scores.<sup>3</sup> A recent study<sup>4</sup> which assessed performance 4 years into the residency showed that NBME scores and Alpha Omega Alpha (a medical honors society) membership failed to correlate with any of the performance measures studied.

One possible explanation for the lack of successful prediction demonstrated by purely academic variables is that the resident needs to have a variety of qualities in addition to an adequate knowledge base to be successful. As has been noted,<sup>5</sup> the resident needs the ability to use and apply knowledge and the personal qualities to enhance this application. Such "noncognitive" qualities as personality traits and work habits play a critical role in unsatisfactory resident performance evaluations.<sup>6</sup> In addition, faculty physicians from a variety of specialties report valuing certain qualities: willingness to admit error, professional maturity, enthusiasm and energy, and interpersonal rapport.<sup>7-9</sup>

The importance of these qualities has not been overlooked in radiology resident selection, and prospective residents usually participate in a series of unstructured interviews with faculty and current residents to assess qualities such as maturity, motivation, and communication skills. Unfortunately, the degree to which these interviews rigorously assess these qualities is unknown. What is known, however, is that ratings obtained from these interviews also are poor predictors of performance, often on the very dimensions that the interviews sought to predict. A recent study,<sup>4</sup> for example, showed that faculty and resident interviews failed to correlate with later performance measures on a variety of dimensions, including motivation and interpersonal skills.

Arguments against a traditional interview format, combined with an unwillingness to give up the opportunity to assess critical noncognitive abilities, led to the development of an experimental behavior-based accomplishment interview.<sup>10,11</sup> This interview is a standardized method of gathering information about residents' competencies in key noncognitive areas (conscientious, curiosity, confidence, recognition of limits, and interpersonal skills) through elicitation of incidents that the resident believes best demonstrate his or her abilities within a given area. Recently, within a single cohort of radiology residents, the accomplishment interview significantly predicted later performance, whereas traditional objective academic measures and standard faculty and resident interviews did not.<sup>4</sup>

The purpose of this study is to determine the degree to which the accomplishment interview predicted performance on a variety of dimensions for three successive cohorts of radiology resident applicants. Because three cohorts were studied, the number of participants allowed use of multiple regression analyses to examine the incremental predictive utility of the experimental interview above that gained by traditional interviewing.

## Materials and Methods

One hundred fifty-one applicants applied for a radiology residency at The University of Iowa in 1983, 1984, or 1985 and participated in a traditional on-site, day-long interview conducted by various faculty members. This interview was given a mean score based on the subjective evaluations of at least four faculty interviewers. The applicant also was interviewed by at least one current radiology resident who used the same interview format. The interviewers had access to the applicants' files, and thus were not blind to the applicants' objective data (eg, NBME scores, class rank, etc).

The applicants also participated in an accomplishment interview during their on-campus visit. (Approximately 6 residents did not agree to participate in this interview over the 3 years.) Before the visit, the applicants had been sent a questionnaire that requested descriptions of past situations in which the applicant had best demonstrated the following qualities: conscientiousness, curiosity, confidence, recognition of limits, and interpersonal skills. A definition of each of the qualities was included. During the accom-

plishment interview, trained interviewers, who used standard probes to obtain additional information, asked the applicants to elaborate on the incidents that had been described in the questionnaire. The interviewers were blind to the applicants' other application data.

The responses to the accomplishment interviews were scored independently by two raters. For the 1983 responses, two psychology raters were used; two physician raters scored the 1984 and 1985 data. The raters used behavioral benchmarks that had been developed by two radiology physicians to reflect poor versus outstanding performance in each area. Scores ranged from 1 to 5; disagreements were resolved consensually.

Four years after the interviews, attempts were made to recontact the residency applicants who were by then in the middle of their third or fourth year. The residents who agreed to participate consented to allowing the investigators to obtain an assessment of their current performance from their residency director. The directors were sent two evaluation forms. One was of the type currently used by The University of Iowa's Department of Radiology to assess skills in four areas (motivation, interaction with others, manual dexterity, and interpretive skills) on a five-point scale of unacceptable (1) to outstanding (5).

The second evaluation form was a behavioral observation scale developed to assess actual behaviors that the evaluator had observed. On this scale, the evaluator was asked to rate the resident on the frequency with which he or she has been observed doing certain behaviors. Thirty items were included, six for each area (conscientiousness, curiosity, confidence, recognition of limits, or interpersonal skills). (The recognition of limits subscale was omitted from the 1983 follow-up.) Items reflect either desirable or undesirable behaviors, and are scored such that a high score indicates more desirable behaviors in a given area. An example of an item from the conscientiousness area is "Pulls previous films in addition to current ones to consider the whole sequence." The scale demonstrated acceptable reliability, with area alpha coefficients ranging from .57 to .77.

Of the 151 residents who interviewed for a residency position, complete data were obtained on 72 (48%). Over the 3 years, 22 residents could not be located, 42 refused further participation, 12 did not respond to our letters, and 3 failed to be evaluated by their residency director. Student's *t* tests examining the applicants' interview data as a function of follow-up showed no significant differences between those residents later followed up and those residents not included (all Student's *t* tests < 1.00).

## Results

Table 1 displays the correlation coefficients among the interview ratings and objective data. Because of occasional missing data, the number of responses for each variable range from 135 to 151. Clearly, with a few exceptions, the accomplishment interview scores are unrelated to the objective predictors. A combined score reflecting both research experience and actual publications correlates with both curiosity and the total accomplishment interview score. In addition, accomplishment interview scores are negatively correlated with NBME scores. Among the traditional predictors, faculty interview ratings are highly correlated with Alpha Omega Alpha membership and NBME scores as well as with resident interview ratings.

Our research question in this study concerned the comparative utility of traditional variables and behavior-based

TABLE 1. Intercorrelations among Predictor Variables

	Resident interview	Faculty interview	AOA	Research publications	NBME
Faculty interaction	0.42*	—	—		
AOA	0.15	0.17†	—		
Research publications	-0.09	0.08	-0.04	—	
NBME	0.06	0.27*	0.44*	-0.14	—
Conscientiousness	-0.05	0.05	-0.08	0.11	-0.18
Curiosity	0.10	0.12	-0.14	0.21*	-0.38*
Confidence	-0.12	-0.01	0.18	0.07	-0.04
Recognition of limits	0.05	-0.03	0.12	0.01	-0.02
Interpersonal skills	0.04	0.05	0.06	0.07	-0.03
Noncognitive total	0.01	0.06	0.04	0.16†	-0.26†

AOA: Alpha Omega Alpha; NBME: National Board of Medical Examiners.

\* $P < .01$ .

† $P < .05$ .

interviews in predicting later performance. The data were first examined with bivariate Pearson product-moment correlations. The degree to which faculty and resident interviews predicted later performance was considered, and these correlation coefficients are presented in Table 2. The only significant predictors were among the resident interview ratings, and these were in a negative direction. Resident interview scores negatively predicted later conscientiousness, curiosity, and confidence. No faculty interview ratings achieved significance as a predictor.

The five components of the accomplishment interview also were considered as individual predictors of later performance. (Because the recognition of limits items were omitted from the 1983 follow-up, the number of responses available for the correlations for those interview scores was 39; other responses range from 64 to 72.) As shown in Table 3, conscientiousness and recognition of limits did not positively predict later performance. However, interview-based ratings of curiosity successfully predicted later curiosity, confidence predicted later interpersonal skills and interpretive skills, and interpersonal skills predicted later interpretive skills and overall evaluation by physicians.

Our primary interest was in considering the increment in prediction achieved by the behavior-based accomplishment interview compared with traditional faculty and resident in-

terviewing. Thus, the data were analyzed with a series of hierarchical multiple regressions. In each regression, the faculty interview and resident interview scores were entered first as a block; the five accomplishment interview scores were then entered as a second block. Since in previous research,<sup>4</sup> a resident applicant's research and publications predicted later performance, we also included this score in the first block of predictors.

The results of the multiple regression analysis for the criterion variable of total score on the behavioral observation scale showed that ratings from faculty and resident interviews accounted for 2.4% of the variability in performance ( $R^2 = .024$ ); when the accomplishment interview scores were added, the  $R^2$  value rose to .26. Individual significant predictors were faculty interview scores, with a beta of  $-.737$ , and recognition of limits, with a beta of  $.699$ , both significant at  $P < .05$ . Thus, when the criterion variable was overall performance in five key noncognitive areas, faculty interview scores were significantly negatively associated with performance, whereas recognition of limits, assessed during the accomplishment interview, was a significant positive predictor.

A second regression analysis considered overall resident evaluation at follow-up assessed by the total score on the departmental form. The first block of predictors (faculty and resident interview, research and publications) accounted for 4% of the variance,  $R^2 = .04$ . The second block of predictors (the accomplishment interview scales) increased the  $R^2$  to  $.18$ . There were no individual significant predictors at the  $.05$  level.

Regression analyses also were carried out using the individual scales of the traditional evaluation measure (motivation, interaction, manual dexterity, and interpretive skills) as criterion measures, because this measure considered resident performance components not in the behavioral observation scale. The hierarchical regression on motivation showed an increment in  $R^2$  of  $.06$  for the accomplishment interview scores; there were no individual significant predictors. The residents' interactions with others (eg, patients, staff), when analyzed within the multiple regression,

TABLE 2. Correlations between Faculty and Resident Interviews and Performance Data

	Faculty	Resident
Conscientiousness	-0.10	-0.24*
Confidence	-0.13	-0.27*
Curiosity	0.02	-0.22*
Interpersonal skills	-0.01	-0.15
Recognition of limits	-0.05	-0.17
Total behavior observation scale	-0.11	-0.19
Motivation	-0.12	-0.02
Interaction with others	-0.10	-0.03
Manual dexterity	0.14	0.16
Interpretive skills	-0.04	-0.10
Total resident evaluation form	-0.04	-0.06

\* $P < .05$ .

TABLE 3. Correlations between Accomplishment Interview Scores and Performance Measures

	Conscientiousness	Curiosity	Confidence	Recognition of limits	Interpersonal
Conscientiousness	0.15	0.06	0.16	-0.07	0.16
Confidence	0.17	0.03	0.02	-0.04	0.03
Curiosity	0.15	0.21*	-0.11	0.10	0.08
Interpersonal skills	0.06	0.03	0.20*	-0.16	0.14
Recognition of limits	-0.09	0.05	0.08	0.17	0.11
Total behavior					
observation scale	0.12	0.04	0.01	0.21	0.06
Motivation	-0.05	0.09	-0.02	-0.10	0.10
Interaction with others	-0.14	0.15	0.00	-0.27*	0.18
Manual dexterity	-0.05	0.06	0.04	-0.23*	0.14
Interpretive skills	0.06	0.09	0.21*	-0.09	0.21*
Total resident					
evaluation form	-0.06	0.14	0.08	-0.19	0.21*

\* $P < .05$ .

showed an increment in  $R^2$  from .04 to .26. Specific predictors were interpersonal skills (beta = .29) and curiosity (beta = .28), both significant at  $P < .05$ . The multiple regression on manual dexterity scores showed an increment in  $R^2$  from .04 to .14, with no individual significant predictors. Finally, the multiple regression on interpretive skills showed an  $R^2$  increase from .02 to .14 and a significant contribution by confidence (beta = .29).

### Discussion

The most noteworthy finding of this study was that, considered in the aggregate, the behavior-based interview scores significantly added predictive utility to ratings obtained from unstructured faculty and resident interviews. Although individual components of the behavior-based accomplishment interview have demonstrated predictive utility in previous research, the accomplishment interview's overall contribution to selection should be considered. The data from three cohorts of resident applicants studied here suggest that unstructured faculty and resident interviews do not predict later performance, particularly in key noncognitive areas. The prediction increment obtained by the behavior-based accomplishment interview scores underlines the importance of obtaining accurate data on noncognitive variables during the interview process.

Why might this increase in prediction have occurred? First, as noted earlier, the accomplishment interview personnel were blind to applicants' other application data. Our previous research<sup>4</sup> and the data of this study indicate significant correlations between faculty interviews and objective data available on the application. This pattern suggests that faculty and residents may be partially biased by applicants' records, and use "confirmatory" strategies of interviewing where information that appears contradictory to an initial favorable or unfavorable impression is discounted. In contrast, the accomplishment interview focused on specific incidents, used standard probes, and was objectively scored by two raters. Thus, there may be more rigor in the mea-

asures obtained by behavioral-based accomplishment interviewing than in traditional interviews, where faculty and residents pursue lines of inquiry somewhat idiosyncratically. In addition, the lack of correlation between accomplishment interview scores and ratings based on faculty and resident interviews suggests that the accomplishment interview allows an assessment of information not obtained during traditional interviews.

Two important points should be noted. First, key noncognitive aspects of resident performance, such as conscientiousness, were more successfully predicted by structured behavior-based interviews than by traditional faculty and resident interviews. Second, the noncognitive predictor of confidence was highly related to interpretive skills during residency, suggesting that some noncognitive aspects may be critical components of the learning process during residency, and thus may influence performance in cognitive areas. It is interesting that confidence has been shown to be a critical feature of success in residency in obstetrics-gynecology<sup>12</sup> and pediatrics<sup>13</sup> as well as in the internship year before the doctorate in psychology.<sup>14</sup>

There are a few limitations to this study that should be noted. First, the reliability of the traditional evaluation measure may be suspect. Our data showed a significant positive skew to the ratings, and thus the lack of positive correlations between this measure and predictor variables may be partially a function of measurement error. In addition, these evaluations were completed by only one rater per subject; thus, the inter-rater reliability is unknown. Second, although three cohorts of residents were studied, the data may represent an unknown sampling bias, in that all subjects were applicants to a single midwestern residency.

Overall, our results suggest that improving unstructured faculty and resident interviews to obtain, in a more rigorous manner, desired information about noncognitive abilities may be a key to successful resident selection. The failure to account for later performance with traditionally conducted interviews, we argue, reflects a lack of rigor in collecting

information through the interview process, as opposed to a lack of importance in later performance for the noncognitive variables that such an interview is intended to predict.

### Acknowledgments

The authors thank Valerie Tarico, Michelle Greiner, and Randall Ross for their assistance in interviewing resident applicants.

### References

1. Wagoner NE, Suriano JR, Stoner JA. Factors used by program directors to select residents. *J Med Educ* 1986;61:10-21.
2. Gardner B. A multivariate computer analysis of students' performances as a predictor of performance as a surgical intern. *J Surg Res* 1972;12:216-219.
3. McCollister RJ. The use of Part 1 National Board Scores in the selection of residents in ophthalmology and otolaryngology. *JAMA* 1988;259:240-242.
4. Wood PS, Smith WL, Altmaier EM, Tarico VS, Franken EA. A prospective study of cognitive and noncognitive selection criteria as predictors of resident performance. *Invest Radiol* 1990;25:855-859.
5. Lloyd JS, Langsley DB, eds. *How to Evaluate Residents*. Chicago: American Board of Medical Specialties. 1986.
6. Anwar RAH, Bosk C, Greenberg AG. Resident evaluation: Is it, can it, should it be objective? *J Surg Res* 1981;30:27-41.
7. Funk DI, Ronai AK, Kinzer JB, Barrett MJ. Personality traits in anesthesia resident evaluations. *Anesthesiology* 1984;61:A462.
8. Greganti MA, McGaghie WC, Finn WF. Personal characteristics of house staff candidates: a quantitative analysis of relative weights. *Ann Intern Med* 1982;97:108-111.
9. Ross CA, Leichner P. Criteria for selecting residents. *Can J Psychiatr* 1984;29:681-686.
10. Hough LM. Development and evaluation of the "Accomplishment Record" method of selecting and promoting professionals. *J Appl Psychol* 1984;69:135-146.
11. Tarico VS, Altmaier EM, Smith WL, Franken EA, Berbaum K. An accomplishment interview for selection of radiology residents. *J Med Educ* 1986;61:845-847.
12. Altmaier EM, Johnson SR, Tarico VS, Laube D. An empirical specification of residency performance dimensions. *Obstet Gynecol* 1988;72:126-130.
13. Altmaier EM, McGuinness G, Wood PS, Ross RR, Bartley J, Smith WL. Defining successful performance among pediatric residents. *Pediatrics* 1990;85:139-143.
14. Ross RR, Altmaier EM. A job analysis of psychology internships in counseling center settings. *J Consult Clin Psychol* 1990;37:459-464.