

Selection Criteria for Internal Medicine Residency Applicants and Professionalism Ratings During Internship

MICHAEL W. CULLEN, MD; DARCY A. REED, MD, MPH; ANDREW J. HALVORSEN, MS;
CHRISTOPHER M. WITTICH, MD, PHARM.D; LISA M. BAUMANN KREUZIGER, MD; MIRA T. KEDDIS, MD;
FURMAN S. McDONALD, MD, MPH; AND THOMAS J. BECKMAN, MD

OBJECTIVE: To determine whether standardized admissions data in residents' Electronic Residency Application Service (ERAS) submissions were associated with multisource assessments of professionalism during internship.

PARTICIPANTS AND METHODS: ERAS applications for all internal medicine interns (N=191) at Mayo Clinic entering training between July 1, 2005, and July 1, 2008, were reviewed by 6 raters. Extracted data included United States Medical Licensing Examination scores, medicine clerkship grades, class rank, Alpha Omega Alpha membership, advanced degrees, awards, volunteer activities, research experiences, first author publications, career choice, and red flags in performance evaluations. Medical school reputation was quantified using *U.S. News & World Report* rankings. Strength of comparative statements in recommendation letters (0 = no comparative statement, 1 = equal to peers, 2 = top 20%, 3 = top 10% or "best") were also recorded. Validated multisource professionalism scores (5-point scales) were obtained for each intern. Associations between application variables and professionalism scores were examined using linear regression.

RESULTS: The mean \pm SD (minimum-maximum) professionalism score was 4.09 ± 0.31 (2.13-4.56). In multivariate analysis, professionalism scores were positively associated with mean strength of comparative statements in recommendation letters ($\beta=0.13$; $P=.002$). No other associations between ERAS application variables and professionalism scores were found.

CONCLUSION: Comparative statements in recommendation letters for internal medicine residency applicants were associated with professionalism scores during internship. Other variables traditionally examined when selecting residents were not associated with professionalism. These findings suggest that faculty physicians' direct observations, as reflected in letters of recommendation, are useful indicators of what constitutes a best student. Residency selection committees should scrutinize applicants' letters for strongly favorable comparative statements.

Mayo Clin Proc. 2011;86(3):197-202

AOA = Alpha Omega Alpha; ERAS = Electronic Residency Application Service; USMLE = US Medical Licensing Examination

Professionalism is an expected attribute of all physicians and a core competency of the Accreditation Council for Graduate Medical Education.¹ However, identifying residency applicants who will perform professionally is challenging. In 1995, the Association of American Medical Colleges created the Electronic Residency Application Service (ERAS) to facilitate medical students' applications to US residency programs.² Medical school deans and medical students place application information into ERAS for electronic distribution to residency programs. The ERAS

system contains data regarding applicants' medical school performance, including class rank and clerkship grades, research and volunteer experiences, recommendation letters, gaps in training, adverse actions, and academic remediation. Residency selection committees rely on these data for important decisions about selecting applicants. Unfortunately, despite the widespread use of ERAS data for resident selection, limited research exists regarding whether these data are associated with residency applicants' professional behaviors during subsequent years of training.

Lack of professionalism among medical learners carries negative consequences. For example, delinquent behaviors during preclinical years have been associated with lower professionalism scores on third-year clerkship evaluations.³ Unprofessional behaviors in both medical school and residency have been shown to predict disciplinary action by state medical licensing boards,^{4,6} and disruptive residents often have low professionalism scores.⁷ Conversely, highly professional behaviors in residency have been linked to conscientiousness, medical knowledge, and clinical skills.⁸

Much research has examined best practices for measuring professionalism.^{9,10} Experts suggest that professionalism assessments should reflect numerous observations from multiple individuals in realistic contexts over time.¹⁰ The purpose of the assessments should be transparent, and assessment standards should apply equally to all.¹⁰ We used these principles to assess residents' professionalism at

From the Division of Cardiovascular Diseases (M.W.C.), Internal Medicine Residency Program, Associate Program Director, and Division of Primary Care Internal Medicine (D.A.R.), Internal Medicine Residency Office of Educational Innovations (A.J.H.), Division of General Internal Medicine (C.M.W.), Division of Nephrology and Hypertension (M.T.K.), Internal Medicine Residency Program, Program Director, and Division of General Internal Medicine and Division of Hospital Internal Medicine (F.S.M.), Internal Medicine Residency Program, Associate Program Director, and Division of General Internal Medicine (T.J.B.), Mayo Clinic, Rochester, MN; and Division of Hematology, Oncology, and Transplantation, Department of Medicine, University of Minnesota, Minneapolis (L.M.B.K.).

This study was supported in part by the Mayo Clinic Internal Medicine Residency Office of Educational Innovations as part of the Accreditation Council for Graduate Medical Education Educational Innovation Project. One author (A.J.H.) performed the statistical analysis for this study.

Individual reprints of this article are not available. Address correspondence to Thomas J. Beckman, MD, Division of General Internal Medicine, Mayo Clinic, 200 First St SW, Rochester, MN 55905 (beckman.thomas@mayo.edu).

© 2011 Mayo Foundation for Medical Education and Research

Mayo Clinic via multisource ratings completed by faculty, peers, and nonphysician professionals.⁸ This professionalism assessment method has been shown to be reliable and valid.⁸

Few studies have explored the relationship between admission criteria and professional behavior among medical learners. Stern et al³ found no association between medical school admission packet materials and professionalism in medical school. We are unaware of previous studies examining associations between admission criteria for internal medicine residency applicants and subsequent professional behaviors during residency. Therefore, we determined whether selection criteria obtained from ERAS for national and international applicants to a large internal medicine residency program were associated with validated, observation-based ratings of professionalism during subsequent training.

PARTICIPANTS AND METHODS

We conducted a retrospective cohort study of all first-year categorical internal medicine residents (N=191) at Mayo Clinic who began training from July 1, 2005, through July 1, 2008 (4 consecutive classes of residents). We sought to examine associations between standard residency selection and professionalism ratings during internship. This study was deemed exempt from review by the Mayo Clinic Institutional Review Board.

Six raters (M.W.C., D.A.R., C.M.W., L.M.B.K., M.T.K., and T.J.B.) extracted internal medicine residency applicant characteristics from ERAS on a number of independent variables (Table 1). To assess the reliability of the data collection process, 15 randomly selected applications were scored by one of the 15 possible unique pairings of the 6 raters.

ASSESSMENT OF PROFESSIONALISM

Residents' professionalism was assessed using a multisource professionalism score consisting of observation-based assessments. Scores were collected from peers who worked with the interns for at least 1 month, third-year residents who provided supervision for at least 1 month, and faculty physicians who provided supervision for at least 2 weeks. The professionalism assessment included 16 items structured on a 5-point scale (1 = needs improvement, 3 = average resident, 5 = top 10%; Table 2). Validity for this professionalism assessment has been previously demonstrated. Specifically, content validity was established by items that represent known professionalism domains and best practices.⁹⁻¹³ Internal structure validity was supported by items representing a multidimensional assessment of professionalism with demonstrated internal consistency reliability.⁸ Criterion validity was evidenced by associations

between resident professionalism scores and in-training examination knowledge scores, Mini-Clinical Evaluation Exercise scores, and conscientious behaviors.⁸

STATISTICAL ANALYSES

Scores from professionalism assessments were averaged by resident to form a continuous overall mean professionalism score ranging from 1 to 5. Interrater reliability for abstraction of ERAS variables was determined using the Cohen κ for nominally scaled variables and intraclass correlation coefficients for ordinally scaled variables.

Associations between data abstracted from ERAS applications (the independent variables) and mean professionalism score (the dependent variable) were examined using simple and multiple linear regression. Given the large number of study participants assigned to the 50th percentile because of unreported data, a multiple linear regression model ignoring class rank was also fit. Medical school rankings from *U.S. News & World Report* were available for US medical school graduates only. They were assessed in a secondary analysis restricted to this subset of residents. To account for multiple comparisons, an α level of .01 was used to determine statistical significance. This sample of 191 residents provided 84% power to detect a Cohen f^2 effect size of 0.15 in the multiple linear regression model associating each resident's mean professionalism score with his or her 13 ERAS variable values. All calculations were performed using SAS statistical software, version 9.1 (SAS Institute, Cary, NC).

RESULTS

This study included 191 interns and 36,512 professionalism assessments from the 2005-2006 through 2008-2009 academic years. The sample included all eligible residents from the 4 classes studied. The mean \pm SD age of the study population was 28.4 \pm 3.0 years. Each resident's mean professionalism score was based on a mean \pm SD (range) of 191.2 \pm 25.1 (30-248) evaluations. For the 191 residents, the mean \pm SD (minimum-maximum) professionalism score was 4.09 \pm 0.31 (2.13-4.56). The sample included 68 female residents (35.6%) and 35 international medical school graduates (18.3%).

Interrater agreement for the nominally scaled variables,¹⁴ including Alpha Omega Alpha (AOA) membership, presence of advanced degree, indication of career choice, and presence of red flags was good (mean Cohen κ , 0.66).¹⁵ Interrater agreement for the ordinally scaled variables, including number of awards, volunteer experiences, research experiences, first author publications, medicine clerkship percentile, class rank, and average strength of comparative statements in letters, was excellent (mean intraclass correlation coefficient, 0.92).¹⁵

TABLE 1. Independent Variables Used in Abstracting Data From ERAS Applications^a

USMLE Step 1 and Step 2 Clinical Knowledge scores
Presence of an additional advanced degree (ie, PhD, MPH, MBA, PharmD, JD, MS)
Counts of awards, volunteer activities, research experiences, and first author publications
Indication of career intent in the personal statement (eg, cardiology, geriatrics), categorized as present or absent
Class rank on a percentile (0-100) scale ^b
Medicine clerkship grade on a percentile (0-100) scale ^c
Presence of red flags in the MSPE ^d
Comparative statements in the letters of recommendation ^e
AOA membership ^f
Medical school reputation on a percentile (0-100) scale ^g

^a AOA = Alpha Omega Alpha; ERAS = Electronic Residency Application Service; MSPE = Medical Student Performance Evaluation; USMLE = US Medical Licensing Examination.

^b Higher percentile indicated a higher class rank. Class rank was determined using either actual numeric rank (if available) or the midpoint of the most discriminating quantitative grouping provided. For example, if students were in the top third of their class, they were assigned to the 83.5th percentile (midpoint between the 100th and 67th percentiles). If no indication of class rank was available, applicants were assigned to the 50th percentile, which allowed us to retain all applicants in the analysis. A total of 76 applicants (39.8%) were assigned to the 50th percentile.

^c A higher percentile indicated a better performance. Clerkship grade was determined from the midpoint of the most discriminating range obtained from the medical school transcript or MSPE. For example, if an applicant received honors, and the top 20% of other students also received honors, then the applicant was assigned to the 90th percentile (ie, the midpoint between 100th and 80th percentiles). If an informative grading scale was unavailable, the applicant was assigned to the 50th percentile, which affected 14 applicants (7.3%). Clerkship grades were converted into percentile scores rather than categorical variables because individual medical schools used scales with different numbers of categories and reported different proportions of students in each category.

^d A red flag was defined as indication of remediation, a gap in training, or adverse action against the applicant. A concerning comment in the MSPE (ie, the applicant is argumentative or uninterested, lacks knowledge, or communicates poorly) was also considered a red flag.

^e Each letter of recommendation was evaluated for the strongest comparative statement according to the following scale: 0 = *no comparative statement in letter*; 1 = *neutral enthusiasm* ("performed at the level of his peers" or "performed as expected for students at his/her level of training"); 2 = *moderate enthusiasm* ("top group of medical students with whom I have worked," 11%-20% decile, top 20%); 3 = *most enthusiasm* ("best medical student with whom I have worked," top 10%). Scores were averaged across all letters as a measure of the cumulative letter strength for the application. The academic rank of the letter writer (full professor, associate professor, assistant professor, instructor, community physician, or unknown) was also recorded. We excluded quotations from evaluations embedded within letters from clerkship directors or department chairs when scoring comparative statements.

^f Alpha Omega Alpha membership (yes or no) was obtained from the ERAS common application form. For the "no" responses, the applicant's medical school was cross-referenced with the AOA Web site (<http://www.alphaomegaalpha.org/chapters.htm>; accessed November 20, 2009) to determine if the applicant's school sponsored an AOA chapter. Applicants from schools that did not offer AOA chapters were placed into a "not offered" category. Applicants from schools that offered AOA chapters in the 4th year of medical school, after submission of ERAS applications, were also considered in the "not offered" category.

^g A higher percentile indicated a better rank. Rankings were derived from 2009 *U.S. News & World Report* research rankings for 63 of 146 US medical schools (available at <http://grad-schools.usnews.rankingsandreviews.com/best-graduate-schools/top-medical-schools/research-rankings>; accessed on November 9, 2009). If multiple schools were tied for a rank, then the average of their percentiles was used. In accordance with this approach, the 83 US medical schools that were not ranked in the report were assigned the 28th percentile by averaging their possible percentiles (ie, 82/146, 81/146... 0/146). This approach affected 81 applicants (51.9%) from US medical schools.

ERAS APPLICATION VARIABLES

Table 3 provides the mean \pm SD for each of the ERAS application variables. The ERAS applications of the 191 residents in our cohort contained 725 distinct letters of recommendation. Of those 725 letters, 115 (15.9%) received a score of "3" for most enthusiastic comparative statement, 97 (13.4%) received a score of "2" for moderately enthusiastic comparative statement, and 10 (1.4%) received a score of "1" for a neutral comparative statement. The remainder of

the letters of recommendation (503, 69.4%) contained no comparative statements. Of the 191 applicants, 37 (19.4%) received a score of "3" as their strongest comparative statement, 23 (12.0%) received a score of "2," and 25 (13.1%) received a score of "1." No comparative statement was available for the remaining 106 residents (55.5%).

Information on the academic rank of the letter writer was available for 573 of the 725 recommendation letters: 264 (46.1%) were authored by full professors, 120 (20.9%) by

TABLE 2. Professionalism Assessment for First-Year Internal Medicine Residents

Type of assessment	Question
Resident of resident (peer)	Effectiveness and completeness of sign-outs
	Helpfulness in the completion of tasks
	Coverage of cross-cover issues and completion of necessary tasks when on call
	Effective communication with families, patients, and all members of team
	Level of integrity
	Dedication to learning environment
	Expressed respect for learners
	Desirability as a physician for one of your family members
Senior medical resident of intern	Desirability as a future coworker or team member
	Level of integrity
	Commitment to his/her education
	Desirability as future coworker or team member
	Desirability as a physician for one of your family members
Faculty of resident	Recognition of his/her own limitations and willingness to ask for help appropriately
	Commitment to his/her own education
	Humanistic qualities/professionalism

associate professors, 149 (26.0%) by assistant professors, 13 (2.3%) by instructors, and 27 (4.7%) by community physicians. Using generalized estimating equations to account for the correlation of letters within each application, no association between the strongest comparative statement and academic rank of the letter writer was found (all, $P > .13$).

ASSOCIATIONS BETWEEN ERAS APPLICATION VARIABLES AND MEAN PROFESSIONALISM SCORES

In the unadjusted model, volunteer activities ($P = .01$) and the average strongest comparative statement ($P < .001$) were significantly associated with mean professionalism score. In multivariate analysis, only the strongest comparative

TABLE 3. Association Between ERAS Variables and Mean Professionalism Scores During First Year of Internal Medicine Residency^a

ERAS variables	No. (%) (N=191)	Unadjusted ^b		Adjusted ^c	
		$\beta \pm SE$	<i>P</i> value	$\beta \pm SE$	<i>P</i> value
Elected to AOA					
Yes	52 (27.2)	-0.005±0.06	.94	-0.09±0.07	.20
No	88 (46.1)	0.003±0.05	.95	0.02±0.06	.68
Not offered	51 (26.7)				
Indicated career choice					
Yes	78 (40.8)	0.07±0.05	.13	0.06±0.05	.16
No	113 (59.2)				
Presence of advanced degree					
Yes	40 (20.9)	0.05±0.05	.36	0.06±0.06	.36
No	151 (79.1)				
Presence of red flags					
Yes	24 (12.6)	-0.06±0.07	.38	-0.10±0.07	.14
No	167 (87.4)				
	Mean ± SD				
USMLE Step 1 score	230.8±15.9	0.0007±0.001	.63	-0.001±0.002	.59
USMLE Step 2 score	236.6±19.1	0.002±0.001	.06	0.001±0.002	.41
Awards	1.6±1.7	0.02±0.01	.11	0.02±0.01	.22
Volunteer activities	5.7±5.0	0.01±0.004	.01	0.01±0.005	.04
Research experiences	2.4±1.8	0.004±0.01	.72	-0.002±0.01	.91
First author publications	0.66±1.97	-0.008±0.01	.46	-0.01±0.01	.33
Medicine clerkship grade, percentile	72.9±15.9	0.003±0.001	.02	0.001±0.002	.42
Class rank, percentile	65.6±19.0	-0.00007±0.001	.96	0.00006±0.001	.97
Strongest comparative statements score	0.75±0.65	0.13±0.03	<.001	0.13±0.04	.002

^a AOA = Alpha Omega Alpha; ERAS = Electronic Residency Application Service; USMLE = US Medical Licensing Examination.

^b Simple linear regression; *t* test of $\beta = 0$.

^c Multiple linear regression (accounting for all variables simultaneously); *t* test of $\beta = 0$.

statement in letters of recommendation remained significant ($P=.002$). A 1-point increase in the average strongest comparative statement was associated with a 0.13-point increase in mean professionalism score ($\beta=0.13$; 95% confidence interval, 0.05-0.20). Thus, a 1-unit increase in the rating of comparative statements was associated with a 0.13 increase in the mean professionalism score during internship. This regression estimate approximates one-half the SD of the mean professionalism scores in the sample. This finding remained unchanged in the multiple linear regression model excluding class rank.

Results were unchanged after incorporating the *U.S. News and World Report* medical school rankings into an analysis of the 156 US medical school graduates. Medical school ranking was not significantly associated with mean professionalism scores in either unadjusted ($P=.91$) or adjusted ($P=.65$) models. In the adjusted model, the average strongest comparative statement remained the only independent variable significantly associated with mean professionalism scores ($P=.001$; $\beta=0.14$; 95% confidence interval, 0.06-0.23).

DISCUSSION

This study demonstrated that the strength of comparative statements in applicants' letters of recommendation were positively associated with professionalism during internship, whereas other variables traditionally examined among applicants were not. To our knowledge, this is the first study to assess the association between comprehensive ERAS application variables and validated professionalism scores during subsequent training in internal medicine.

It is noteworthy that most of the application variables that residency programs typically consider when choosing residents, such as medical school reputation, AOA status, US Medical Licensing Examination (USMLE) scores, and clerkship grades, were not associated with professionalism during internship. This finding challenges the role of these variables in the selection of professional residents. Standard application variables (eg, USMLE scores and clerkship grades) predict medical knowledge and clinical performance¹⁶⁻¹⁹ and should thus be appraised during the selection process. However, such variables may not be sufficiently robust for identifying residents who will consistently uphold professional values. Residency programs are best served to identify professionalism during the application process because patients highly value professionalism among physicians^{1,20} and residents who cause disruptions often demonstrate deficiencies in professionalism.⁷

Although most ERAS variables were not associated with professionalism, this study identified an association between comparative statements in letters of recommen-

dation and multisource professionalism ratings during internship. This association was independent of writer experience because no relationship was found between letter writers' academic ranks and their strongest comparative statements. Our findings underscore the strength of observation-based assessment.²¹ Previous work has found a low frequency of comparative statements in letters of recommendation^{22,23} and little correlation between letters of recommendation and subsequent clinical performance.²⁴⁻²⁷ This lack of correlation is likely because these studies did not specifically identify comparative statements within letters²⁵⁻²⁷ or did not examine associations between letter content and professionalism scores that reflected observations of learners in realistic settings.²⁴ Our study specifically analyzed statements in letters that compared students to their peers. Furthermore, both letters of recommendation and our multisource professionalism ratings are based on first-hand observations of learners in clinical contexts over time. Our findings suggest that letters of recommendation can be strong markers of observation-based assessment in a residency application. Conversely, most other ERAS variables, such as USMLE scores, AOA status, and medical school reputation, do not involve direct observations of learners and thus may be less useful indicators of subsequent professional behavior. Therefore, residency program selection committees should consider scrutinizing letters of recommendation for observation-based comparative statements.

Our study extends the literature on application variables and trainee performance. Most prior research has involved medical students.^{3,28-30} Studies of residents have focused on medical knowledge and clinical skills.^{16,19,31} In the domain of professionalism, negative comments in the Medical Student Performance Evaluation have been associated with poor professionalism among psychiatry residents.³² Additionally, third-year medicine clerkship grades have been associated with professionalism ratings during internship.¹⁸ However, this study incorporated fewer independent variables and measured professionalism using a one-time survey of program directors.¹⁸ In contrast, our study examined a wide range of application variables and used a multi-source mean professionalism score consisting of numerous observation-based assessments of residents over the course of the internship. Furthermore, our validated professionalism assessment contains authoritative item content, reliable scores, and demonstrated associations between professionalism ratings and medical knowledge, clinical competence, and dutifulness.⁸

This study has some limitations. First, it was conducted at a single institution, and so additional studies are needed to further generalize the findings. However, the independent variables in this study, obtained from ERAS, are widely used by US residency programs. Second, some

independent variables were not available for all candidates (ie, AOA status). Third, *U.S. News & World Report* rankings have been criticized for lacking objective quality measurements.³³ Nonetheless, these rankings are commonly used to judge medical school quality. Fourth, grading the strength of comparative statements in candidates' letters of recommendation may be labor intensive. However, given the meaningful association with subsequent professionalism, we suggest that inspecting letters of recommendation for strongly positive comparative statements will enhance the ability to select highly professional residents. Fifth, comparative statements in clerkship directors' and department chairs' letters may not have involved direct observation. However, these comparative statements did reflect direct observations from the writers' colleagues, and the letters by clerkship directors and department chairs represented only a minority of the sample. Sixth, although we assumed that faculty members routinely observed the students assigned to them on clinical rotations, it is possible that, in rare instances, only limited amounts of observation occurred. Seventh, few applicants had information regarding "red flags" and specific career intent in their applications, which limited our ability to detect associations for these variables. Finally, although the professionalism outcome in this study was based on multisource assessments, which is a single method of measuring professionalism, our multisource assessment represents a best practice¹⁰ and was previously validated.⁸

CONCLUSION

We found that strongly favorable comparative statements in the recommendation letters for internal medicine residency applicants were associated with multisource assessments of professionalism during internship. This finding suggests that faculty members' observation-based assessments of students are powerful indicators of what constitutes a best student. Furthermore, this finding offers residency selection committees a useful tool for selecting applicants who may perform professionally during internship. Future research should examine associations between variables in ERAS and professionalism in later years of training and practice. Future work should also determine the relationships between these variables and other outcomes, including medical knowledge and clinical skills.

REFERENCES

1. Accreditation Council for Graduate Medical Education (ACGME). ACGME Outcome project. <http://www.acgme.org/outcome/implementation/rsvp.asp>. Accessed January 18, 2011.
2. Association of American Medical Colleges (AAMC). Electronic Residency Application Service. <https://www.aamc.org/services/eras/>. Accessed January 18, 2011.
3. Stern DT, Frohna AZ, Gruppen LD. The prediction of professional behaviour. *Med Educ*. 2005;39(1):75-82.
4. Papadakis MA, Teherani A, Banach MA, et al. Disciplinary action by medical boards and prior behavior in medical school. *N Engl J Med*. 2005;353(25):2673-2682.
5. Papadakis MA, Hodgson CS, Teherani A, Kohatsu ND. Unprofessional behavior in medical school is associated with subsequent disciplinary action by a state medical board. *Acad Med*. 2004;79(3):244-249.
6. Papadakis MA, Arnold GK, Blank LL, Holmboe ES, Lipner RS. Performance during internal medicine residency training and subsequent disciplinary action by state licensing boards. *Ann Intern Med*. 2008;148(11):869-876.
7. Yao DC, Wright SM. National survey of internal medicine residency program directors regarding problem residents. *JAMA*. 2000;284(9):1099-1104.
8. Reed DA, West CP, Mueller PS, Ficalora RD, Engstler GJ, Beckman TJ. Behaviors of highly professional resident physicians. *JAMA*. 2008;300(11):1326-1333.
9. Arnold L. Assessing professional behavior: yesterday, today, and tomorrow. *Acad Med*. 2002;77(6):502-515.
10. Stern DT. *Measuring Medical Professionalism*. New York, NY: Oxford University Press Inc; 2006.
11. American Board of Internal Medicine (ABIM). *Project Professionalism*. Philadelphia, PA: American Board of Internal Medicine; 2002.
12. Arnold L, Stern DT. *What is Medical Professionalism?* New York, NY: Oxford University Press; 2006.
13. Norcini J. Faculty observations of student professional behavior. In: Stern DT, ed. *Measuring Medical Professionalism*. New York, NY: Oxford University Press; 2006:147-157.
14. Kundel HL, Polansky M. Measurement of observer agreement. *Radiology*. 2003;228(2):303-308.
15. Landis JR, Koch GG. The measure of observer agreement for categorical data. *Biometrics*. 1977;33(1):159-174.
16. Andriole DA, Jeffe DB, Whelan AJ. What predicts surgical internship performance? *Am J Surg*. 2004;188(2):161-164.
17. Durning SJ, Pangaro LN, Lawrence LL, Waechter D, McManigle J, Jackson JL. The feasibility, reliability, and validity of a program director's (supervisor's) evaluation form for medical school graduates. *Acad Med*. 2005;80(10):964-968.
18. Greenburg DL, Durning SJ, Cohen DL, Cruess D, Jackson JL. Identifying medical students likely to exhibit poor professionalism and knowledge during internship. *J Gen Intern Med*. 2007;22(12):1711-1717.
19. Hamdy H, Prasad K, Anderson MB, et al. BEME systematic review: predictive values of measurements obtained in medical schools and future performance in medical practice. *Med Teach*. 2006;28(2):103-116.
20. Bendapudi NM, Berry LL, Frey KA, Parish JT, Rayburn WL. Patients' perspectives on ideal physician behaviors. *Mayo Clin Proc*. 2006;81(3):338-344.
21. Durning SJ, Cation LJ, Markert RJ, Pangaro LN. Assessing the reliability and validity of the mini-clinical evaluation exercise for internal medicine residency training. *Acad Med*. 2002;77(9):900-904.
22. Fortune JB. The content and value of letters of recommendation in the resident candidate evaluative process. *Curr Surg*. 2002;59(1):79-83.
23. O'Halloran CM, Altmaier EM, Smith WL, Edmund A, Franken J. Evaluation of resident applicants by letters of recommendation: a comparison of traditional and behavior-based formats. *Invest Radiol*. 1993;28(3):274-277.
24. Boyse TD, Patterson SK, Cohan RH, et al. Does medical school performance predict radiology resident performance? *Acad Radiol*. 2002;9(4):437-445.
25. DeZee KJ, Thomas MR, Mintz M, Durning SJ. Letters of recommendation: rating, writing, and reading by clerkship directors of internal medicine. *Teach Learn Med*. 2009;21(2):153-158.
26. Greenburg AG, Doyle J, McClure DK. Letters of recommendation for surgical residencies: what they say and what they mean. *J Surg Res*. 1994;56(2):192-198.
27. Leichner P, Eusebio-Torres E, Harper D. The validity of reference letters in predicting resident performance. *J Med Educ*. 1981;56(12):1019-1021.
28. Huff KL, Koenig JA, Treptau MM, Sireci SG. Validity of MCAT scores for predicting clerkship performance of medical students grouped by sex and ethnicity. *Acad Med*. 1999;74(10):S41-S44.
29. Julian ERP. Validity of the medical college admission test for predicting medical school performance. *Acad Med*. 2005;80(10):910-917.
30. Silver B, Hodgson CS. Evaluating GPAs and MCAT scores as predictors of NBME I and clerkship performances based on students' data from one undergraduate institution. *Acad Med*. 1997;72(5):394-396.
31. Markert RJ. The relationship of academic measures in medical school to performance after graduation. *Acad Med*. 1993;68(2):S31-S34.
32. Brenner AM, Mathai S, Jain S, Mohl PC. Can we predict "problem residents"? *Acad Med*. 2010;85(7):1147-1151.
33. Sehgal AR. The role of reputation in *U.S. News & World Report's* rankings of the top 50 American hospitals. *Ann Intern Med*. 2010;152(8):521-525.