Prioritizing the Interview in Selecting Resident Applicants: Behavioral Interviews to Determine Goodness of Fit

Michael B. Prystowsky, MD, PhD[1], Evan Cadoff, MD[1],
Yungtai Lo, PhD[1], Tiffany M. Hebert, MD[1], and
Jacob J. Steinberg, MD[1]

Abstract

From our initial screening of applications, we assess that the 10% to 15% of applicants whom we will interview are all academically qualified to complete our residency training program. This initial screening to select applicants to interview includes a personality assessment provided by the personal statement, Dean's letter, and letters of recommendation that, taken together, begin our evaluation of the applicant's cultural fit for our program. While the numerical scoring ranks applicants preinterview, the final ranking into best fit categories is determined solely on the interview day at a consensus conference by faculty and residents. We analyzed data of 819 applicants from 2005 to 2017. Most candidates were US medical graduates (62.5%) with 23.7% international medical graduates, 11.7% Doctors of Osteopathic Medicine (DO), and 2.1% Caribbean medical graduates. Given that personality assessment began with application review, there was excellent correlation between the preinterview composite score and the final categorical ranking in all 4 categories. For most comparisons, higher scores and categorical rankings were associated with applicants subsequently working in academia versus private practice. We found no problem in using our 3-step process employing virtual interviews during the COVID pandemic.

Introduction

For decades, residency training programs have relied on numerical rankings to evaluate applicants for academic study and clinical training.[1-10] Certainly, some aspects of emotional intelligence are considered in examining personal statements, letters of recommendation ([LoR] including the recommender), and Dean's letters; usually, these considerations result in a tempering of a numerical ranking. During the past 2 to 3 decades, our pathology residency training program has utilized a process to override numerical ranking. This was likely driven by experience; that is, selecting a high scoring disruptive applicant and, alternatively, working with a lower scoring empathic

applicant who became a top of class performer. We needed to define ourselves and our training program culture, as each program needs to do. By knowing ourselves, our faculty and resident trainees could put applicants into categories based on

[1] Albert Einstein College of Medicine/Montefiore Medical Center, Bronx, NY, USA

Corresponding Author:
Michael B. Prystowsky, Department of Pathology, Albert Einstein College of Medicine/Montefiore Medical Center, 1300 Morris Park Avenue, Bronx, NY 10461, USA.
Email: michael.prystowsky@einsteinmed.org

goodness of fit for our program. Thus, the final ranking would be a categorical ranking, and the numerical ranking would be used within each category.

How did we achieve our goal to identify highly qualified candidates who will fit our culture[11-15] (eg, geography, ethos, practice norms, social mores...) and benefit most from our unique training program?[11-16] We did not know if a particular metric or characteristic would identify candidates who would perform optimally in our program.[16] Thus, we undertook a retrospective review of 5 years (2005-2009) to tabulate personality assessments and metrics (metadata) of those selected for an interview. We then reflected upon our process for residency candidate selection as generated by our algorithms and metrics that included academic records, LoRs, the applicant's narrative, interests, hobbies, publications, volunteer and research experiences. Ultimately the candidate's multifocal, multidimensional interviews with us, with the subsequent collaborative discussion, yielded our final impression and ranking. The composite features of our optimal applicant included sufficient intelligence, pertinent performance, optimal emotional intelligence, adequate pathology experience with a strong commitment to pathology, yielding the best-fit candidate for our culture and training program. This evaluation was then used prospectively. We now report on the outcomes of this recruitment process, up through the applicants evaluated in 2017 who have now completed their categorical training.

We used established evaluation methods for constructing our algorithms and metrics. We used personality assessment inventories (PAI, with the *Big-Five Factor*[17-20] as core). We particularly prioritized aspects of Highly Reliable Teamwork and their successful personality traits of "agreeableness and conscientiousness" with primarily additive features of detectable honesty, humility, and humor (the *HEXACO* addition to *OCEAN*—the *Big-Five*—[openness, conscientiousness, extraversion, agreeableness, and neuroticism]). *HEXACO*[21] is a 6-dimensional model of human personality that was created by Ashton and Lee and explained in their book, *The H Factor of Personality;* the 6 factors include Honesty-Humility (H), Emotionality (E), Extraversion (X), Agreeableness (A), Conscientiousness (C), and Openness to Experience (O) with each factor indicating high and low levels. Further, in our group discussions, we sought the "right-fit" to our program: the right personality, including resilience (Duckworth's Resilience and "GRIT"[22-24]), and potential for success in a busy city, and culturally complex, economically underserved academic hospital environment. In essence, we sought individuals who would successfully assimilate the ethos of Montefiore (eg, sense of "social justice").[25]

Worrisome indicators of potential future suboptimal performance included evidence of deficient communication skills. Some doctors have suboptimal communication skills and fail to communicate adequately or appropriately with peers, mentors, patients, and families. The Australian New South Wales Health Care Complaints Commission (who follow these matters closely) noted the number of doctors' complaints has been increasing annually, medical practitioners more than dental,

nurses, midwives, pharmacists, psychologists, and other health practitioners.[14,17] A significant percentage attract complaints and reported cases to the UK General Medical Council (GMC) had the highest ever number of complaints against doctors. This is a significant area of concern—and the ability to communicate with patients and peers is critical for success, under the rubric of "Can you talk to me?"[26,27]

Medical training can give a doctor the basic knowledge required and foster their skills, including updating that knowledge to ensure continued academic competence. It can also teach or nurture some of the other skills and attitudes in the competency list. However, it is unrealistic to expect that medical education can do it all, mainly if the student is attitudinally unsuited or otherwise ill-equipped in their psychological makeup to meet the profession's expectations and the community outlined above. Acceptance of this line of thought must lead us to acknowledge that we should take particular care in selecting medical practitioners, basing our choice on a range of criteria that reflect the excellent generic doctor's picture. Dr Powis describes techniques and methods used to measure some nonacademic and noncognitive qualities and provides empirical data on their reliability, construct validity, and, most importantly, their predictive validity that supports their adoption, suitable health professionals.[16]

Ultimately, the PAI advocates for a checklist[28] and dashboard of the most highlighted aspects of character and traits that one would like to include within a training program candidate and those that may be disruptive individually and likely to the group at large.

The attributes of the "better doctor" professional that can work within the present socially complex health care system include[29]: academic ability and cognitive skills; ability to communicate appropriately; good interpersonal skills; and demonstrated teamwork and empathy. Attributes that could detract from more optimal performance include: psychological vulnerability (inability to handle stress appropriately, low resilience); high levels of neuroticism; low levels of conscientiousness; extreme detachment or extreme emotional involvement; and high levels of impulsiveness and permissiveness. The PAI seeks to concretize and communalize these concepts by creating a common touchstone that evaluators could visibly see and discuss, to yield our final categorical ranking of the applicants who are likely to do best and gain most from our training program.

Materials and Methods

We used the following 3 types of metrics in the review process: numerical scores, a semi-quantitative score assigned to text documents (eg, LoR), and a rank priority rating based on interactions during the interview day (Figure 1). Our process has evolved during the past 2 decades primarily in how we weight non numerical data; the specific measures used today are presented below.
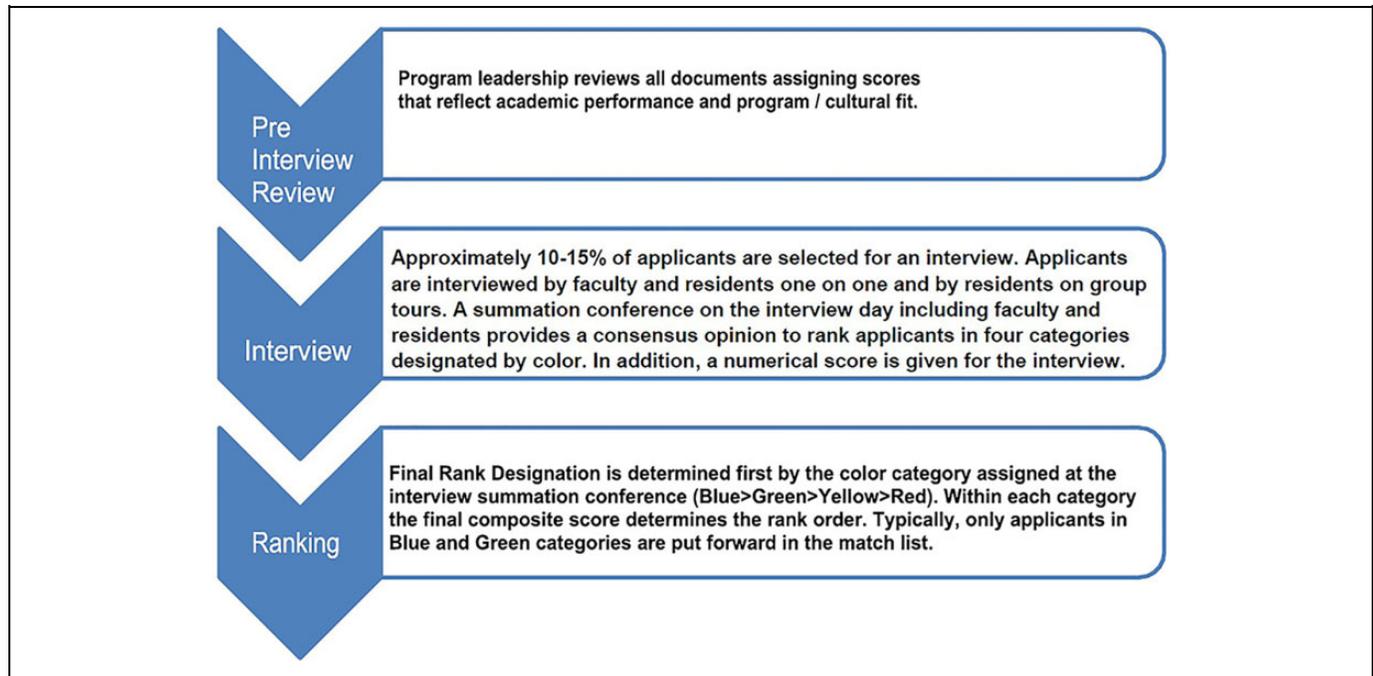
Figure 1. Three step application review. Applications are screened by the program director and associate directors. Approximately 10% to 15% of applicants are invited for interviews based on academic performance and potential programmatic fit. Programmatic fit is assessed during the interview. A consensus conference is held immediately after interviews to rank order applicants.

Initial Screen by Program Director

We have generally received 400 to 600 applications per year over the past 3 decades. The program director screens these applications with input from faculty and selects 60 to 70 candidates to interview for 5 positions. Elements in the ERAS application including Medical Student Performance Evaluation (MSPE),[30] LoR, Personal Statement, and Medical School (MS) transcript are given 1 to 5 points and multiplied by a weighting factor provided by ERAS. USMLE[31-37] scores are divided by 100. A Preliminary Composite Score is derived from the sum of the 5 previous elements; a hypothetical preinterview ranking sheet is shown in Figure 2.

Criteria used for exclusion from consideration include the following:

- Partial completion of residency training in another specialty, looking to transfer into pathology
- Dismissed from pathology residency training program elsewhere
- >10 years from clinical exposure
- MSPE or LoR with flags for professionalism
- No pathology LoRs
- No pathology clinical experience or observerships
- Unexplained leave of absence or gaps in curriculum vitae
- Inarticulate writing in the personal statement.

Additional points for outstanding performance. An additional score of up to 1 point could be added for exceptional achievements in the early days of developing our ranking algorithm, for example, *Alpha Omega Alpha,*[34,38] extensive research,[19,36,39,40] PhD, selective postdoctoral positions, exceptional meritorious publications or patents, Peace Corps, veterans, or government service. The additional point was added to the final ranking.

Social media profile[41-45]. A review of propriety content is now unavoidable. We begin with a general Google search, an examination of images from Google search, a Twitter review, and a Facebook review, if applicable.

Interview preparation-[46,47]. Because we rank order the candidates into discrete categories, we in-service the Faculty, House staff, and Associates for interviewing goals, methods, and management to obtain consistency and optimal results. We meet to review the interview process enabling us to discover more about each candidate, to assess the candidate's ability to communicate interests, and to learn and deliver the following:

- Candidate's passion for pathology, intellectual, and social qualities
- Positive impression of us and the ethos of Montefiore and Einstein
- Combination of self-motivation and successful teamwork
- Our training culture: A kind (empathy[19,48-54]), nurturing, visionary residency

Applicant	Pre-Interview Composite	Document	LoR	MSPE	USMLE
1	22.56	4.9	5.0	7.5	5.2
2	22.08	4.9	5.0	7.5	4.7
3	21.47	4.6	4.8	7.5	4.6
4	20.69	4.4	4.5	6.8	5.1
5	20.65	4.6	4.8	6.8	4.5
6	20.52	4.5	4.5	6.8	4.7
7	20.36	4.4	4.8	6.0	5.1
8	19.82	4.0	3.9	6.8	5.2
9	19.73	4.4	4.5	6.0	4.8
10	19.67	4.2	4.4	6.0	5.1
11	19.54	4.3	4.8	5.3	5.2
12	19.49	4.0	3.8	6.8	4.9
13	19.06	4.0	3.7	6.8	4.7
14	18.97	4.2	4.4	6.0	4.4
15	18.95	3.9	4.0	6.0	5.0
16	18.93	4.2	4.6	5.3	4.9
17	18.87	4.1	4.1	6.0	4.7
18	18.78	4.1	4.1	6.0	4.5
19	18.73	3.8	3.8	6.0	5.2
20	18.73	4.0	4.4	5.3	5.1
21	18.26	3.9	4.2	5.3	4.9
22	18.21	3.8	4.0	5.3	5.1
23	17.96	4.0	4.5	4.5	5.0
24	17.74	3.8	4.0	5.3	4.7
25	17.50	4.0	4.5	4.5	4.5
26	17.40	3.3	3.7	5.3	5.1
27	17.02	3.6	3.8	5.3	4.3
28	17.02	3.9	4.4	4.5	4.2
29	16.67	3.5	3.5	5.3	4.4
30	16.47	3.5	3.8	4.5	4.6
31	16.22	3.5	3.8	4.5	4.3
32	16.12	3.7	4.3	3.8	4.4
33	15.44	3.5	4.0	3.8	4.2
34	14.76	3.1	3.5	3.8	4.4
35	14.51	3.3	4.0	3.0	4.2

Figure 2. Representative preinterview rank order of applicants. Academic performance and academic/life experience as determined by information in the ERAS application yield a preinterview composite score that determines which applicants will be invited for interviews.

- Our goal to grow and develop the pathologist for the 21st century
- Forward-looking view for defining the competencies/milestones/outcomes

We outline and underscore our guidelines[55-57] for significant interview events. These guidelines (barring the handshakes and refreshments) are applicable in the virtual interview environment. We utilized a guideline and checklist of what to do—and what not to do. Ultimately, we would underscore the programs and hospital's collegiality, residency quality-of-life, teaching, research, national organizational membership, citizenship and leadership, teamwork[58-60] and communal involvement and activities, and success in fellowships and beyond.

Interview process. The interview process includes face-to-face interactions in multiple settings including, individual (Dyadic Mini Interviews; 3 per applicant), panels, and groups in multiple social settings. The 3 face-to-face interviews are scored 0 to 3 for each interview with the average multiplied by 2 and added to the preinterview composite score to yield a final composite score. The personal interactions of the formal interviews and informal interactions with current residents (on tours, at lunch) enable us to identify a candidate's weaknesses, and to address factual record concerns, for example, great interviews versus poor USMLE's. At the end of each interview day, all faculty and residents who participated in the interview process confer and assess each applicant with respect to their ability to adapt to our culture and take advantage of our program. Ultimately, this results in a categorical rank of applicants, usually focused primarily on emotional interactions and subjective assessment of how the candidate will fit in our program. Color coding includes 4 categories as follows from best to worst: Blue (4) > Green (3) > Yellow (2) > Red (1). Each candidate is presented by the 3 interviewers and the group (approximately 20 reviewers) assigns a color rank to the candidate. The most crucial decision at the summation conference is to delete candidates who do not fit in our program (Red).

In the program, we further understand that we attract and recruit young trainee candidates who will enter a particular culture. We'd like to know that they value equality, that they can deal with uncertainty and not avoid it but when presented, are not too overwhelmed by the complexities of the uncertain and unknown will pause, ask for help, reflect, be comfortable in the inevitable hierarchies of medicine. It is best if the candidate exhibits some sense of agency, executive function, planning, and organization; can demonstrate significant restraint, and is resourceful and has reliable resilience. These attributes will give them the best chance of success in survival within the excessive stresses that hammer them during residency training. These interactions we observe and try to verbalize and semi-quantify can only become more and more complex. Still, to avoid an over-obsessive or a sophistication that could impair our committee's reproducibility to decide to the point of non-utility, we try to adhere to our basics and algorithms. This is nicely presented in the "PPIK" theory of process, personality, interests, and knowledge that inform us, pique our selection process, hone our skills for common ground to choose the better candidate (hopefully!).

Some Questions that help us unravel a working-PAI:

- Any arts you enjoy? Examples?
- Memorable volunteer experience? Example?
- Extracurricular activity? Example?
- Memorable medical case example?
- Memorable pathology (clinical) case example?
- Group/team activity example?
- Why Montefiore?
- How do you unwind?
- Any measures of stress?

Telemedicine, telehealth, telepathology—interviews from afar[61-63]. Given the present COVID-19 pandemic and a sea-change in how we practice medicine, the ubiquitous presence of technology in our candidates' lives and our own, as well as in our clinical work through the increasing use of technologically mediated psychologically and medically informed health care and screening interviews, we needed to develop a heightened awareness of both the gains and the losses provided using available technologies. We in-serviced both our faculty and residents to optimize their Telemedicine ability for both care and the interview process in a human resources strategy and mode. We also are attentive to how heavy use of the internet, social media, texting, and so on may be affecting our candidates, concerning cognition, development, memory, attention span, future vulnerability, character tendencies, reality-testing, capacity for identity formation, and other consequences. We also reviewed the ethical and legal aspects of providing technologically mediated screening, interviews, and evaluations.

Final rank order. The last rank order is set initially by the color category. All candidates in the blue category are ranked higher than candidates in the green category regardless of the composite score as shown in a representative final ranking sheet (Figure 3). Candidates within a color category are ranked by the final combined score, which includes the PAI from interview day. Finally, the bottom of the rank list is trimmed by answering the following question—would we want this candidate rather than selecting[64] a candidate from the Supplemental Offer and Acceptance Program (SOAP) lists? All candidates in the red category are deleted, and most in the yellow category are deleted. Because the initial screening successfully selects candidates with acceptable academic achievement who are likely to fit in our program, we always have sufficient candidates in the blue and green categories to produce a match list with about 10 times the applicants required to fill our 5 slots.

Statistics. Differences in gender and medical school training among 4 color groups and between faculty position and private practice were compared using χ^2 tests or Fisher exact tests. Differences in composite scores, interview scores, recommendation letter scores, Dean's letter scores, and USMLE#1 scores among 4 color groups were compared using Kruskal-Wallis tests, and between faculty position and private practice were compared using Wilcoxon rank-sum tests. The Bonferroni correction was used to adjust *P* values for post hoc pairwise comparisons between color groups. Spearman correlation coefficients were used to assess the correlation of interview scores with preinterview composite scores, Dean's letter scores, recommendation letter scores, and USMLE #1 scores. All statistical analyses were performed using SAS version 9.4 (SAS Institute Inc.). Values of *P* .05 or less were considered statistically significant.

Results

The outcome data presented below shows that essentially all candidates interviewed and accepted into our program have had successful outcomes of both their residency training and ensuant careers.

Screening and Ranking

Our preinterview screen selected for academic performance and cultural fit for our program; thus, there was an intentional screening bias up front. The interview process enabled us either to confirm our initial bias or adjust our ranking based on personal interactions that probed for programmatic fit. In reviewing the hypothetical data in Figure 3, we present 2 possible scenarios for applicant 35 and applicant 4.

Applicant 35 has the lowest final composite score in the Blue category. This is likely a foreign applicant with a challenging life story where the applicant demonstrated empathy and strong teamwork. The applicant likely had a strong advocate from at least one interview who strongly recommended the Blue category convincing the review group. Since interview scores are given individually and averaged, the color category does not always reflect the average interview score.

Applicant 4 has one of the highest scores in the Green category. This scenario is typically a male with a perceived somewhat aggressive personality. Perhaps a reason for academic success but in an interview if perceived as too aggressive, residents will be concerned about teamwork and attendings are concerned about resistance to training. There was likely mixed opinion which is why the applicant remained acceptable and in the Green category. If there was universal concern, the candidate would have been placed in the Red category and not considered further.

We present these 2 examples to demonstrate how dependent our final selection is on the values that we consider highly and the personal interactions during the interview process. We believe that all of the candidates whom we interview are academically acceptable and will be successful in their careers. We focus our selection on the candidates who will fit our culture and training program.

Applicant Data

We analyzed data of 819 applicants from 2005 to 2017. Most candidates were US medical graduates (USG, 62.5%) with 23.7% international medical graduates (IMG), 11.7% DO, and 2.1% Caribbean medical graduates (CRB). When we compared our final ranking using the color categories (Table 1), there was predominantly gender equality in all color categories except for Red, where there were near twice as many males; the red category accounted for only 10% of the total candidate pool (*P* = .087). Considering undergraduate medical training, we have higher percentages of Blue in USG and IMG, whereas a proportionately higher percentage of DOs were in the RED and Yellow category (*P* = .001). This may have resulted from

Applicant	Final Composite	Interview	Pre-interview Composite	Document	LoR	MSPE	USMLE
1	28.19	5.6	22.56	4.9	5.0	7.5	5.2
2	27.66	5.6	22.08	4.9	5.0	7.5	4.7
3	26.22	4.8	21.47	4.6	4.8	7.5	4.6
6	25.67	5.2	20.52	4.5	4.5	6.8	4.7
7	25.37	5.0	20.36	4.4	4.8	6.0	5.1
5	25.33	4.7	20.65	4.6	4.8	6.8	4.5
12	24.72	5.2	19.49	4.0	3.8	6.8	4.9
8	24.35	4.5	19.82	4.0	3.9	6.8	5.2
11	24.10	4.6	19.54	4.3	4.8	5.3	5.2
14	23.74	4.8	18.97	4.2	4.4	6.0	4.4
17	23.69	4.8	18.87	4.1	4.1	6.0	4.7
19	23.21	4.5	18.73	3.8	3.8	6.0	5.2
20	22.83	4.1	18.73	4.0	4.4	5.3	5.1
28	21.82	4.8	17.02	3.9	4.4	4.5	4.2
26	20.67	3.3	17.40	3.3	3.7	5.3	5.1
30	20.35	3.9	16.47	3.5	3.8	4.5	4.6
35	18.01	3.5	14.51	3.3	4.0	3.0	4.2
9	24.59	4.9	19.73	4.4	4.5	6.0	4.8
13	24.30	5.2	19.06	4.0	3.7	6.8	4.7
4	24.23	3.5	20.69	4.4	4.5	6.8	5.1
10	23.87	4.2	19.67	4.2	4.4	6.0	5.1
18	23.09	4.3	18.78	4.1	4.1	6.0	4.5
15	22.93	4.0	18.95	3.9	4.0	6.0	5.0
16	22.18	3.3	18.93	4.2	4.6	5.3	4.9
21	22.03	3.8	18.26	3.9	4.2	5.3	4.9
23	21.90	3.9	17.96	4.0	4.5	4.5	5.0
22	21.68	3.5	18.21	3.8	4.0	5.3	5.1
25	21.66	4.2	17.50	4.0	4.5	4.5	4.5
24	20.83	3.1	17.74	3.8	4.0	5.3	4.7
31	20.40	4.2	16.22	3.5	3.8	4.5	4.3
29	19.75	3.1	16.67	3.5	3.5	5.3	4.4
33	18.80	3.4	15.44	3.5	4.0	3.8	4.2
32	18.74	2.6	16.12	3.7	4.3	3.8	4.4
27	20.28	3.3	17.02	3.6	3.8	5.3	4.3
34	16.40	1.6	14.76	3.1	3.5	3.8	4.4

Figure 3. Representative final rank order of applicants. Goodness of fit defines the color category of each applicant and academic scoring defines the rank order within each color category. The final rank order usually only containing the blue and green categories becomes the match list yielding approximately a 10:1 ratio of applicants to positions available. To demonstrate the impact of the interview within each color category, the preinterview ranking is given in the applicant column. Blue represents our highest priority applicants with Green as the next priority level. Yellow applicants are seldom included in the match and Red applicants are excluded from the match list.

earlier years when DO's lack of "Match" NRMP obligation enabled the better or best candidates to drop out of the NRMP/Match—presently not allowed. Ultimately, we sought the best bona fide candidates to attract and recruit from any UN/WHO-approved medical school.

It is important to note that the color category is determined solely by the group discussion after the interviews. All parameters are considered, but the interview goodness of

fit in our program determines the color. The composite (numerical) score determines the rank within a color category (Figure 3).

Candidate gender—males and females. There was an essential parity of males and females chosen—over a historical era that ranged from more male candidates in the past to the present decade, where more female candidates predominate (Table 1).

Table 1. Gender, Medical School Training, Composite Scores, Interview Scores, Recommendation Letter Scores, Dean's Letter Scores, USMLE#1 Scores, and Faculty Position by Four Color Groups.

	Blue N = 165	Green N = 468	Yellow N = 104	Red N = 82	P-value
Gender, n (%)					.087
Female	87 (52.7%)	238 (51.0%)	51 (49.0%)	30 (36.6%)	
Male	78 (47.3%)	229 (49.0%)	53 (51.0%)	52 (63.4%)	
Medical school training, n (%)					.001
CRB	2 (1.2%)	13 (2.8%)	1 (1.0%)	1 (1.2%)	
DO	6 (3.6%)	53 (11.4%)	24 (23.1%)	13 (15.9%)	
IMG	46 (27.9%)	108 (23.1%)	21 (20.2%)	18 (21.9%)	
USG	111 (67.3%)	293 (62.7%)	58 (55.8%)	50 (61.0%)	
Final composite score, median (IQR)	19.9 (16.4, 24.6)	20.6 (16.7, 22.9)	17.7 (16.2, 19.6)	15.3 (7.9, 18.5)	< .001
Preinterview composite score*, median (IQR)	17.1 (13.3, 20.5)	15.1 (12.9, 18.9)	12.7 (9.5, 15.2)	11.0 (7.3, 13.2)	< .001
Interview score, median (IQR)	3.8 (3.1, 4.6)	3.6 (3.4, 4.3)	3.4 (3.2, 3.8)	3.1 (0.6, 3.6)	< .001
Recommendation letter score, median (IQR)	3.8 (3.0, 4.5)	3.6 (3.0, 4.5)	3.2 (3.0, 4.0)	3.2 (2.5, 4.3)	< .001
Dean's letter score, median (IQR)	3.7 (3.4, 4.3)	3.5 (3.3, 4.1)	3.4 (3.0, 4.0)	3.3 (2.8, 3.8)	< .001
USMLE #1 score, median (IQR)	236 (222, 247)	220 (204, 236)	214 (196, 229)	212.5 (199, 230)	< .001
Faculty position	32 (59.3%)	109 (59.9%)	13 (37.1%)	12 (54.5%)	.094
Private practice	22 (40.7%)	73 (40.1%)	22 (62.9%)	10 (45.5%)	

Abbreviations: CRB, Caribbean schools; DO, osteopathic schools; IMG, international graduates; IQR, interquartile range; USG, US graduates.
*Preinterview composite score = final composite score − interview score.

Table 2. Correlation of Interview Scores With Preinterview Composite Scores, Dean's Letter Scores, Recommendation Letter Scores, and USMLE #1 Score in All and by 4 Color Groups.*

	Preinterview composite score	Dean's letter score	Recommendation letter score	USMLE #1 score
Interview score (for all 4 color groups)	r = 0.58 P < .001	r = 0.36 P < .001	r = 0.56 P < .001	r = 0.07 P = .120
Interview score (for Blue only)	r = 0.68 P < .001	r = 0.29 P = .024	r = 0.57 P < .001	r = −0.03 P = .782
Interview score (for Green only)	r = 0.41 P < .001	r = 0.37 P < .001	r = 0.59 P < .001	r = −0.06 P = .30
Interview score (for Yellow only)	r = 0.17 P = .234	r = 0.45 P = .022	r = 0.15 P = .378	r = 0.27 P = .092
Interview score (for Red only)	r = 0.77 P < .001	r = 0.39 P = .018	r = 0.44 P = .002	r = 0.03 P = .834

*r: Spearman correlation coefficient.

Composites (metric algorithm). The final composite score includes numerical ratings for academic performance, step exams, Dean's letter, LoR, and a weighted score for on-site interviews (Figure 3, Table 1). The median (interquartile range [IQR]) final composite scores were 19.9 (16.4, 24.6) in Blue, 20.6 (16.7, 22.9) in Green, 17.7 (16.2, 19.6) in Yellow, and 15.3 (7.9, 18.5) in Red (P < .001). The preinterview composite scores obtained from all numerical ratings except on-site interviews mostly mirrored the color categories with a median (IQR) of 17.1 (13.3, 20.5) in Blue, 15.1 (12.9, 18.9) for Green, 12.7 (9.5, 15.2) in Yellow, and 11.0 (7.3, 13.2) for Red (P < .001). This tandem parallel reality reinforced the probability of a functional process and diminished the extreme bias of decisions. That said, a numerically strong candidate was occasionally downgraded on the color rating for suboptimal interviews, and a numerically average candidate was periodically upgraded for strong interpersonal skills and cultural fit.

Interview score, including the numerous face-to-face encounters. Median quantitative interview score again mirrored the color category with ample separation between Blue (3.8 [3.1, 4.6]) and Red (3.1 [0.6, 3.6]), but much closer assessments in Green and Yellow of 3.6 (3.4, 4.3) versus 3.4 (3.2, 3.8)—indicating a more difficult decision to make (Table 2).

Recommendation letters. Scores were significantly different among 4 color groups (P < .001) though Yellow and Red were mainly the same number 3.2 (3.0, 4.0) and 3.2 (2.5, 4.3; P = .999). Recommendation scores were quite close in Blue (3.8 [3.0, 4.5]) and Green (3.6 [3.0, 4.5]; P = .999). We viewed many recommenders as simply advocates for the applicants.

Table 3. Gender, Medical School Training, Composite Scores, Interview Scores, Recommendation Letter Scores, Dean's Letter Scores, and USMLE#1 Scores by Academia Faculty Position and Private Practice.

	Faculty position, N = 166	Private practice, N = 127	P value
Gender, n (%)			.416
Female	89 (53.6%)	62 (48.8%)	
Male	77 (46.4%)	65 (51.2%)	
Medical school training, n (%)			.413
CRB	2 (1.2%)	3 (2.4%)	
DO	15 (9.0%)	18 (14.2%)	
IMG	32 (19.3%)	20 (15.7%)	
USG	117 (70.5%)	86 (67.7%)	
Final composite score, median (IQR)	21.7 (17.5, 23.8)	19.9 (17.3, 22.9)	.168
Preinterview composite score,* median (IQR)	17.5 (13.2, 19.6)	15.5 (11.1, 18.8)	.027
Dean's letter score, median (IQR)	4.4 (4.2, 4.6)	4.2 (4.1, 4.5)	.018
Interview score, median (IQR)	3.8 (3.5, 4.3)	3.6 (3.3, 4.0)	.005
Recommendation letter score, median (IQR)	4.5 (4.4, 4.6)	4.5 (4.4, 4.6)	.293
USMLE #1 score, median (IQR)	219.5 (203, 235.5)	212.5 (198, 234)	.039

Abbreviations: CRB, Caribbean schools; DO, osteopathic schools; IMG, international graduates; IQR, interquartile range; USG, US graduates.
*Preinterview composite score = final composite score − interview score.

Deans' Letters had been more evaluative in the past—but our opinion on this matter has changed over the last decade.

USMLE#1. Blue had the highest median score of 236 (222, 247), followed by Green 220 (204, 236), Yellow 214 (196, 229), and Red 212.5 (199, 230; $P < .001$). Red and Yellow median scores were equivalent ($P = .999$).

Correlation of interview scores with preinterview composite scores (including deans' scores, recommendation scores, and USMLE#1). For all 4 groups, the interview score was moderately to strongly correlated with the preinterview composite score ($r = 0.58$, $P < .001$) and the recommendation letter score ($r = 0.56$, $P < .001$), and weakly to moderately correlated with Dean's letter score ($r = 0.36$, $P < .001$) whereas the interview score was not significantly correlated with USMLE#1 score ($r = 0.07$, $P = .120$; Table 2). Among the Blue category, the interview score remained moderately to strongly correlated with the preinterview composite score ($r = 0.68$, $P < .001$) and recommendation letter score ($r = 0.57$, $P < .001$). Among the Greens, the interview score was moderate to strongly correlated with the recommendation letter score ($r = 0.59$, $P < .001$). The interview score was strongly correlated with the preinterview composite score ($r = 0.77$, $P < .001$) for the Red category, possibly implying the strength of dislike versus like (and supports most decision sciences theory that we may like something, but we are surer of what we do not care for). Unfortunately, the strength of the indicators for the Yellow's appeared somewhat weak with the possibility that the Dean's letter score would be the most helpful ($r = 0.45$).

Careers in academia faculty positions versus private practice. At the time of this report, for those graduates who had completed their fellowship training and were in practice, almost 60% of Blue and Green candidates were discovered in academic faculty practices (Table 1). This was only 5% greater than the Red category. The highest percentage of private practitioners occurred in the Yellow category. Higher preinterview composite scores ($P = .027$), Dean's letters ($P = .018$), interview scores ($P = .005$), and USMLE#1 score ($P = .039$) were associated with the graduate being in an academic faculty position (Table 3).

Candidate questionnaires' responses of interview day experience. Though not quantified, our anonymous, a brief questionnaire has generated very positive response (~ 4.5 positive on a scale of 5.0), with few comments of not being able to engage long enough or at all with select faculty (data not shown). For those applicants seeking more information. We would attempt to arrange a call or rarely, a second visit (usually for an applicant interested in working in a particular research laboratory).

Overview of our ranked candidates' outcome. The top of our lists were predominantly outstanding candidates who achieved impressive residencies and are now in positions typically at leading academic and medical research centers as faculty. A smaller group entered private practice in large multistate practice corporations, for example, dermatopathology. Surprisingly, the less "right-fit" selective to us, the bottom of our lists also did reasonably well and better than we would have expected. A higher percentage of candidates in the Yellow and Red groups went into smaller community private practice, and surprisingly some of them are harder to find on the internet. To seem to disappear is most unusual, but this seems to be occurring with a small cohort, particularly from the bottom of the list.

Adapting to COVID19—Virtual Interviews[65]

As a result of restrictions on travel and physical distancing, we needed to implement virtual tours and interviews for resident candidates.[65] Our process shown in Figure 1 remained

unchanged. We screened 711 applications and invited 60 candidates to interview, splitting them among 4 half-day interview sessions. In preparing for the virtual environment, we reviewed materials from the AAMC, ACGME, and various specialties (many surgical specialties had already been using this technology to some degree in their recruitment strategies) to determine best practices. Faculty, residents, and coordinators all underwent a thorough in-service on these best practices and potential pitfalls in the virtual environment. We also included practice sessions to facilitate comfort with the technology. The schedules were meticulously planned and practiced, minimizing the stress on both the interviewers and the candidates. Interviewing with virtual technology introduces additional opportunities for potential influence by unconscious bias. We added to our typical training on unconscious bias, emphasis on the need to recognize that candidates may have variable access to technology, access to a pleasant interview environment, and even comfortability with the virtual environment.[65] Interviewers were explicitly instructed not to allow these factors, largely out of the control of the interviewee, to negatively impact their evaluation of the candidate.[65] On the interview day, each candidate was interviewed by 2 faculty members and 1 resident. They also had small breakout groups with only residents present where they could have questions answered without any faculty present. The interviews took place in the morning of the interview day. We followed our same protocol for same-day preliminary discussion and ranking of candidates, albeit now in a virtual Zoom room.

A key component of attracting candidates of "good fit"[22,65] was our ability to give the candidates a sense of what it would be like to train as a resident in the Bronx as part of our Montefiore team. Within the virtual setting, providing a virtual tour was pivotal. We worked closely with the hospital's public relations department, developing a Montefiore general and Pathology-specific tour and overview of our program. The public relations team photographed all our laboratories and key educational staff. Our residents provided quotes and vignettes and recorded videos expressing what they felt makes Montefiore a special place to train. We invited all candidates, regardless of expressed affinity group, to attend the institution-wide House Staff Diversity Event. Over one-third of our interviewed candidates attended this event. Finally, our chief residents organized an informal, optional Zoom Q&A chat with only the residents and candidates. We do not typically offer a "second look" visit, but this year it felt necessary to offer candidates more opportunities to engage with our residents given the more impersonal, virtual setting that was required due to the pandemic.

In the end, we did not substantially change the number of ranked candidates over previous years. This year's outcome resembled previous years, with 5 slots comfortably filled.

Discussion

How do we find the best method to determine future performance of an applicant in an individual program? It is common to rely on numerical scores including test scores, grades, and the soft scores that we assign to letters of evaluation and dean's letters. If we allow these numbers to be our final ranking even with weighting, have we really evaluated applicants on the characteristics that we value most? For Montefiore/Einstein, our program focuses on training individuals to practice pathology as an integral physician member of a health care team. This requires empathy, teamwork, and communication skills. Our struggle over the past few decades was to find a way to use both the numerical score and a personality score to rank academically qualified applicants with the best personality fit for our residency training.

Our initial screening process yielded a small fraction of the academically qualified candidates whom we considered had the potential for a good cultural fit. The interpersonal interactions observed during the interview day constituted an essential factor for determining ranking on our match list (color category) and, of most importance, if the candidate was to be eliminated from the match list.

With this level of importance placed on the interview, we prioritize "the culture" at Montefiore/Einstein, in support of the institutional mission to *care for the underserved*.[2,66,67] Interviewers needed to identify candidates with clear evidence of empathy,[19,68-72] humility, humor, cultural skills, a definitive sense of teamwork—and of necessity, scholarship. We also were attentive to equality in our evaluation of the total applicant pool. We kept these principles visible before, during, and after the encounters of an interview.[73-78] This process made sure that we all had the common ground of what we sought in our ranked candidates, and undoubtedly, what we wanted to avoid.[79-86]

More detailed comment is in order regarding the dean's letter. There is forever difficulty and confusion in the various formats of the MSPE as it evolved over the last 2 decades.[30,87-92] The MSPE was originally an evaluative assessment that highlighted the young doctor's strengths and weaknesses in training and allowed programs to visualize how that young trainee would fit into their programs' culture and context. The literature strongly supports a realistic view of what the Dean's letter has become. In 2010, the Journal of *Academic Medicine* noted that the "good" description related to the bottom 50% of the class. In another subspecialty, a paper on the utility of the MSPE in Anesthesiology felt that the MSPE at that time remained a weak contributor in helping to determine and assess the quality of a resident.[93] In a 2014 paper from the *Journal of the American College of Radiology*, a graph showed that the term "excellence" appeared anywhere from the class ranking of the bottom 20% to upward of 90%, with apparently the 60th percentile being the determiner of "excellence".[94,95]

In 2017, a group of surgeons in Philadelphia[30,35] also looked at the revised MSPE at over 100 institutions; they noted that some 30+ percent of schools do not report summative comparative performance data. This lack of standardization among the institutions remains an ongoing issue. Indeed, even the highest assessment of honors, and a serious example of grade-inflation, appear worrisome, with some 40+ percent of

the medical students receiving honors. The MSPE has evolved with routine severe omissions from the Deans' offices for the 6 ACGME competencies to be assessed.

Some have found[94] difficulty with the MSPE, the Dean's letter, and the evaluative process; there is a marked difference between men's and women's medical students' assessment. Moreover, that they differed significantly depending on both the gender of the student and the authors' gender. This insightful analysis was carried out by Dr Carol Isaac[94] and her Team at the University of Wisconsin. In addition, there is emerging literature confirming the presence of racial bias in conferring Alpha Omega Alpha Honor Medical Society (AOA) honors, further muddying the waters in the assessment of medical student performance.[95-101]

To improve this process, a group headed by Dean Linda Hedrick[102] demonstrated that review and rereview and continuous review of prior years' Dean's letter was an ongoing process that was carried out by all the Dean's letter writers together and continuously. This process, which was structured along with the Institute for Healthcare Improvement (IHI) requirement of improvement of health care, achieves better success and uniformity and equivalence among student assessment and the repetition of this assessment every year.

We have an ongoing evolution related to the MSPE, the Dean's letter, how we describe, assess, evaluate, and recommend our medical students. How this information can be made more uniform and meaningful is a challenge. We are indeed entering a new era of the Flexnerian revolution. We hope that we can achieve a modicum of success for the deans' committees' evaluations but, more importantly improve our evaluation process with professionals in psychology and education for our graduating and trainee physician.[30,73,87,90-92,94,95,102-106]

Conclusion

We came to this present interview format based on faculty and associated desire to concentrate the more regular, long-drawn-out interview season into an efficient, compressed process. We also wanted to maximize involvement of our faculty and residents in the evaluation process, where the most "eyes" could interact with candidates.[87,107] Our postinterview summation promoted collaborative interactions with broad exchange of opinions, impressions, and ideas. Factual information about a candidate revealed through the interview process could be discussed in the broadest forum possible, with the attending faculty who would ultimately work with the candidate if accepted into the program.

We believe that our evaluation process has worked superbly for our program goals and mission. The strategy is highly rewarding to our faculty and, upon arrival in the program, to our trainees. This evaluation process establishes the foundation of community and promotes the ethos that we all carry together and with each other and underscoring of our communal social-capital. We also consider that this report constitutes opportunity to examine the anthropological process of evaluating, training, and graduating individuals who will carry this concept of "Bronx-Care" forward as they proceed in their careers.

Pragmatically, we believe the process, methodically presented in this article, has defined the "best-fit" candidate to work with us and succeed in a bustling medical center with highly challenging health care issues and the extraordinary diversity that comprises the citizens, patients, and neighbors we at Montefiore and Einstein care for, including city schools, social centers, nursing homes, and all aspects of acute and chronic care. The Montefiore enterprise has won the famous Foreman Award for community and population health care, named after a recent CEO and President of Montefiore, Dr. Spencer ("Spike") Foreman, who outlined his vision of future health care in 1995 in Academic Medicine, that still defines many of the challenges we face a generation later.

Our "PAI"[1,71,108,109] is the core guidance, our pragmatic and aspirational "compass" of how we choose the next generation of medical practitioners. We are all engaged in this better physician adventure. This paper outlines what we are doing—as flawed as it is—to select and create a better practitioner, a better trainee and colleague who will practice competently and with joy, avoid and diminish burnout[110-117] and medical error.[58,118-125] We need ongoing refinement, comments and improvement to improve and be better doctors.

References

1. Bore M, Munro D, Powis D. A comprehensive model for the selection of medical students. *Med Teach*. 2009;31:1066-1072. doi:10.3109/01421590903095510

2. Dowell J, Lumsden MA, Powis D, et al. Predictive validity of the personal qualities assessment for selection of medical students in Scotland. *Med Teach*. 2011;33:e485-e488. doi:10.3109/0142159x.2011.599448

3. Bernardi RA. Associations between Hofstede's cultural constructs and social desirability response bias. *J Bus Ethics*. 2006;65:43-53. doi:10.1007/s10551-005-5353-0

4. Hofstede G. What did GLOBE really measure? Researchers' minds versus respondents' minds. *J Int Bus Stud*. 2006;37:882-896. doi:10.1057/palgrave.jibs.8400233

5. Hofstede G. Asian management in the 21st century. *Asia Pac J Manag*. 2007;24:411-420. doi:10.1007/s10490-007-9049-0

6. Hofstede G, Garibaldi de Hilal AV, Malvezzi S, Tanure B, Vinken H. Comparing regional cultures within a country: lessons from Brazil. *J Cross Cult Psychol*. 2010;41:336-352. doi:10.1177/0022022109359696

7. Kirkman BL, Lowe KB, Gibson CB. A quarter century of culture's consequences: a review of empirical research incorporating Hofstede's cultural values framework. *J Int Bus Stud*. 2006;37:285-320. doi:10.1057/palgrave.jibs.8400202

8. Malach-Pines A, Kaspi-Baruch O. The role of culture and gender in the choice of a career in management. *Career Dev Int*. 2008;13:306-319. Malach-Pines A, ed. doi:10.1108/13620430810880808

9. McSweeney B. Hofstede's model of national cultural differences and their consequences: a triumph of faith—a failure of analysis. *Hum Relat*. 2002;55:89-118. doi:10.1177/0018726702551004

10. Sarah Powell I. Geert Hofstede: challenges of cultural diversity. *Hum Resour Manag Int Dig*. 2006;14:12-15. doi:10.1108/09670730610663187

11. Conrad P. Eliot Freidson's revolution in Medical Sociology. *Health*. 2007;11:141-144. doi:10.1177/1363459307074688

12. Cottingham AH, Suchman AL, Litzelman DK, et al. Enhancing the informal curriculum of a medical school: a case study in organizational culture change. *J Gen Intern Med*. 2008;23:715-722. doi:10.1007/s11606-008-0543-y

13. Dixon-Saxon S, Buckley M. Student selection, development and retention: a commentary on supporting student success in distance counselor education. *Prof Couns*. 2020;10:57-77. doi:10.15241/sds.l0.1.57

14. Mooney M. Professional socialization: the key to survival as a newly qualified nurse. *Int J Nurs Pract*. 2007;13:75-80. doi:10.1111/j.1440-172X.2007.00617

15. Wiedman D, Martinez IL. Organizational culture theme theory and analysis of strategic planning for a new medical school. *Hum Organ*. 2017;76:264-274. doi:10.17730/0018-7259.76.3.264

16. Powis DA, Neame RLB, Bristow T, Murphy LB. Interview for medical student selection: authors' reply. *BMJ*. 1988;296:1261-1261. doi:10.1136/bmj.296.6631.1261-b

17. Costa PT, McCrae RR. Personality disorders and the five-factor model of personality. *J Personal Disord*. 1990;4:362-371. doi:10.1521/pedi.1990.4.4.362

18. Haidet P, Stein HF. The role of the student-teacher relationship in the formation of physicians. The hidden curriculum as process. *J Gen Intern Med*. 2006;21:S16-S20. doi:10.1111/j.1525-1497.2006.00304.x

19. Hurwitz S, Kelly B, Powis D, Smyth R, Lewin T. The desirable qualities of future doctors—a study of medical student perceptions. *Med Teach*. 2013;35:e1332-e1339. doi:10.3109/0142159x.2013.770130

20. Murinson BB, Klick B, Haythornthwaite JA, Shochet R, Levine RB, Wright SM. Formative experiences of emerging physicians: gauging the impact of events that occur during medical school. *Acad Med*. 2010;85:1331-1337. doi:10.1097/acm.0b013e3181e5d52a

21. Ashton MC, Lee K. Empirical, theoretical, and practical advantages of the HEXACO model of personality structure. *Pers Soc Psychol Rev*. 2007;11:150-166. doi:10.1177/1088868306294907

22. Bulathwatta ADN, Witruk E, Reschke K. Effect of emotional intelligence and resilience on trauma coping among university students. *Health Psychol Rep*. 2017;1:12-19. doi:10.5114/hpr.2017.61786

23. Gardner AK, Steffes CP, Nepomnayshy D, et al. Selection bias: examining the feasibility, utility, and participant receptivity to incorporating simulation into the general surgery residency selection process. *Am J Surg*. 2017;213:1171-1177. doi:10.1016/j.amjsurg.2016.09.029

24. Lindskold S. Trust development, the GRIT proposal, and the effects of conciliatory acts on conflict and cooperation. *Psychol Bull*. 1978;85:772-793. doi:10.1037/0033-2909.85.4.772

25. Safyer SM. LEADERS Interview with Steven M. Safyer, M.D., President and Chief Executive Officer, Montefiore Medicine. Leadersmag.com. Published 2019. Accessed August 17, 2021. http://www.leadersmag.com/issues/2019.4_Oct/New20York%20City/LEADERS-Steven-Safyer-Montefiore.html

26. Ikkos G, Barbenel D. Complaints against psychiatrists: potential abuses. *Psychoanal Psychother*. 2000;14:49-62. doi:10.1080/02668730000700051

27. Murff HJ. Relationship between patient complaints and surgical complications. *Qual Safety Health Care*. 2006;15:13-16. doi:10.1136/qshc.2005.013847

28. Spreng RN, McKinnon MC, Mar RA, Levine B. The Toronto Empathy Questionnaire: scale development and initial validation of a factor-analytic solution to multiple empathy measures. *J Pers Assess*. 2009;91:62-71. doi:10.1080/00223890802484381

29. Powis D, Hamilton J, McManus IC. Widening access by changing the criteria for selecting medical students. *Teach Teach Educ*. 2007;23:1235-1245. doi:10.1016/j.tate.2007.06.001

30. Hook L, Salami AC, Diaz T, Friend KE, Fathalizadeh A, Joshi ART. The revised 2017 MSPE: better, but not "outstanding." *J Surg Educ*. 2018;75:e107-e111. doi:10.1016/j.jsurg.2018.06.014

31. Pohl CA, Hojat M, Arnold L. Peer nominations as related to academic attainment, empathy, personality, and specialty interest. *Acad Med*. 2011;86:747-751. doi:10.1097/acm.0b013e318217e464

32. Harrison LE. Using holistic review to form a diverse interview pool for selection to medical school. *Bayl Univ Med Cent Proc*. 2019;32:218-221. doi:10.1080/08998280.2019.1576575

33. Kenny S, McInnes M, Singh V. Associations between residency selection strategies and doctor performance: a meta-analysis. *Med Educ*. 2013;47:790-800. doi:10.1111/medu.12234

34. Rosenthal S, Howard B, Schlussel YR, et al. Does medical student membership in the Gold Humanism Honor Society influence selection for residency? *J Surg Educ*. 2009;66:308-313. doi:10.1016/j.jsurg.2009.08.002

35. Schaverien MV. Selection for surgical training: an evidence-based review. *J Surg Educ*. 2016;73:721-729. doi:10.1016/j.jsurg.2016.02.007

36. Vasan CM, Gentile M, Huff S, Vasan N. A study of test construct and gender differences in medical student anatomy examination performance. *Int J Anat Res*. 2018;6:5250-5255. doi:10.16965/ijar.2018.171

37. Max BA, Gelfand B, Brooks MR, Beckerly R, Segal S. Have personal statements become impersonal? An evaluation of

personal statements in anesthesiology residency applications. *J Clin Anesth*. 2010;22:346-351. doi:10.1016/j.jclinane.2009.10. 007

38. Low D, Pollack SW, Liao ZC, et al. Racial/ethnic disparities in clinical grading in medical school. *Teach Learn Med*. 2019;31: 487-496. doi:10.1080/10401334.2019.1597724

39. Adam J, Bore M, Childs R, et al. Predictors of professional behaviour and academic outcomes in a UK medical school: a longitudinal cohort study. *Med Teach*. 2015;37:868-880. doi:10.3109/ 0142159x.2015.1009023

40. Celebi JM, Nguyen CT, Sattler AL, Stevens MB, Lin SY. Impact of a scholarly track on quality of residency program applicants. *Educ Prim Care*. 2016;27:478-481. doi:10.1080/14739879.2016. 1197049

41. Bernstein SA, Gu A, Chretien KC, Gold JA. Graduate medical education virtual interviews and recruitment in the era of COVID-19. *J Grad Med Educ*. 2020;12:557-560. doi:10.4300/ JGME-D-20-00541.1

42. Bhayani RK, Fick L, Dillman D, Jardine DA, Oxentenko AS, O'Glasser A. Twelve tips for utilizing residency program social media accounts for modified residency recruitment. *MedEdPublish*. 2020;9. doi:10.15694/mep.2020.000178.1

43. Fick L, Palmisano K, Solik M. Residency program social media accounts and recruitment—a qualitative quality improvement project. *MedEdPublish*. 2020;9. doi:10.15694/mep.2020.000203.1

44. McHugh SM, Shaffer EG, Cormican DS, Beaman ST, Forte PJ, Metro DG. Use of social media resources by applicants during the residency selection process. *J Educ Perioper Med*. 2014;16:E071. doi:10.46374/volxvi-issue5-mchugh

45. Patel TY, Bedi HS, Deitte LA, Lewis PJ, Marx MV, Jordan SG. Brave new world: challenges and opportunities in the COVID-19 virtual interview season. *Acad Radiol*. 2020;27:1456-1460. doi:10.1016/j.acra.2020.07.001

46. Hörz S, Stern B, Caligor E, et al. A prototypical profile of borderline personality organization using the structured interview of personality organization (STIPO). *J Am Psychoanal Assoc*. 2009;57:1464-1468. doi:10.1177/000306510905700060802

47. Bryan CS. "Aequanimitas" Redux: William Osler on detached concern versus humanistic empathy. *Perspect Biol Med*. 2006; 49:384-392. doi:10.1353/pbm.2006.0038

48. Davis KL, Panksepp J, Normansell L. The affective neuroscience personality scales: normative data and implications. *Neuropsychoanalysis*. 2003;5:57-69. doi:10.1080/15294145.2003.10773410

49. Panksepp J. The cradle of consciousness: a periconscious emotional homunculus?: Commentary by Jaak Panksepp (Bowling Green). *Neuropsychoanalysis*. 2000;2:24-32. doi:10.1080/1529 4145.2000.10773278

50. Kernberg OF. Affective neuroscience: the foundations of human and animal emotions. *Neuropsychoanalysis*. 2000;2:270-272. doi:10.1080/15294145.2000.10773321

51. Decety J, Moriguchi Y. The empathic brain and its dysfunction in psychiatric populations: implications for intervention across different clinical conditions. *BioPsychoSocial Med*. 2007;1:22. doi:10.1186/1751-0759-1-22

52. Garden RE. The problem of empathy: medicine and the humanities. *New Lit Hist*. 2007;38:551-567. doi:10.1353/nlh.2007.0037

53. Hojat M, Louis DZ, Markham FW, Wender R, Rabinowitz C, Gonnella JS. Physicians? empathy and clinical outcomes for diabetic patients. *Acad Med*. 2011;86:359-364. doi:10.1097/acm. 0b013e3182086fe1

54. Stepien KA, Baernstein A. Educating for empathy. *J Gen Intern Med*. 2006;21:524-530. doi:10.1111/j.1525-1497.2006.00443.x

55. Al-Assaf AF, Wilson CN. Recruiting physicians for small and rural hospitals. *Hosp Top*. 1991;69:15-18. doi:10.1080/00185868.1991. 9948449

56. Birk C, Breland BD, Boyer ST, et al. ASHP guidelines on the recruitment, selection, and retention of pharmacy personnel. *Am J Health-Syst Pharm*. 2003;60:587-593. doi:10.1093/ajhp/60.6.587

57. Cohn KH, Harlow DC. Field-tested strategies for physician recruitment and contracting. *J Healthc Manag*. 2009;54: 151-158. doi:10.1097/00115514-200905000-00003

58. Baker DP, Day R, Salas E. Teamwork as an essential component of high-reliability organizations. *Health Serv Res*. 2006;41: 1576-1598. doi:10.1111/j.1475-6773.2006.00566.x

59. Castro MG, Dicks M, Fallin-Bennett K, et al. Teach students, empower patients, act collaboratively and meet health goals: an early interprofessional clinical experience in transformed care. *Adv Med Educ Pract*. 2019; 10:47-53. doi:10.2147/amep.s175413

60. Sasagawa M, Amieux PS. Dispositional humility of clinicians in an interprofessional primary care environment: a mixed methods study. *J Multidiscip Healthc*. 2019; 12:925-934. doi:10.2147/ jmdh.s226631

61. Knorr M, Hissbach J. Multiple mini-interviews: same concept, different approaches. *Med Educ*. 2014;48:1157-1175. doi:10.1111/ medu.12535

62. Nield LS, Saggio RB, Nease EK, et al. One program's experience with the use of Skype for SOAP. *J Grad Med Educ*. 2013;5: 707-707. doi:10.4300/jgme-d-13-00319.1

63. Taylor M, Wallen T, Mehaffey JH, et al. Interviews during the pandemic: a thoracic education cooperative group and surgery residents project [Published online March 25, 2021]. *Ann Thorac Surg*. 2021. doi:10.1016/j.athoracsur.2021.02.089

64. Powis D. How to do it. Select medical students. *BMJ*. 1998;317: 1149-1150. doi:10.1136/bmj.317.7166.1149

65. Fraser R, Dunleavy D. *Prep for Success in Your Virtual Interview*; 2020. Accessed August 17, 2021. https://www.aamc.org/system/ files/2020-07/AAMCPrepforSuccessinYourVirtualInterview

66. Farmer P. Challenging orthodoxies: the road ahead for health and human rights. *Health Hum Rights*. 2008;10:5. doi:10.2307/ 20460084

67. Farmer P. The profession in perspective: on teaching and learning. *Engl J*. 1978;67:15. doi:10.2307/815619

68. Halpern J. Empathy and patient–physician conflicts. *J Gen Intern Med*. 2007;22:696-700. doi:10.1007/s11606-006-0102-3

69. Iacoboni M. Imitation, empathy, and mirror neurons. *Annu Rev Psychol*. 2009;60:653-670. doi:10.1146/annurev.psych.60. 110707.163604

70. Kohut H. Introspection, empathy, and psychoanalysis an examination of the relationship between mode of observation and theory. *J Am Psychoanal Assoc*. 1959;7:459-483. doi:10.1177/0003065159 00700304

71. Powis D, Bore M, Munro D, Lumsden MA. Development of the personal qualities assessment as a tool for selecting medical students. *J Adult Contin Educ*. 2005;11:3-14. doi:10.7227/jace.11.1.2

72. Warren S. Poland. The patient's empathy. *Am Imago*. 2009;66: 495-500. doi:10.1353/aim.0.0066

73. Osborn M, Yanuck J, Mattson J, et al. Who to interview? Low adherence by U.S. medical schools to medical student performance evaluation format makes resident selection difficult. *West J Emerg Med*. 2017;18:50-55. doi:10.5811/westjem.2016. 10.32233

74. 232049 (E10) Amos Tversky and the ascent of behavioral economics. *Insur: Math Econ*. 1998;23:194. doi:10.1016/0167-6687(98) 90068-3

75. Kahneman D. A perspective on judgment and choice: mapping bounded rationality. *Am Psychol*. 2003;58:697-720. doi:10.1037/ 0003-066x.58.9.697

76. Thaler RH, Tversky A, Kahneman D, Schwartz A. The effect of myopia and loss aversion on risk taking: an experimental test. *Q J Econ*. 1997;112:647-661. doi:10.1162/003355397555226

77. Tversky A, Kahneman D. Availability: a heuristic for judging frequency and probability. *Cogn Psychol*. 1973;5:207-232. doi:10.1016/0010-0285(73)90033-9

78. Tversky A, Kahneman D. Advances in prospect theory: cumulative representation of uncertainty. *J Risk Uncertai*. 1992;5: 297-323. doi:10.1007/bf00122574

79. Bell BS, Kozlowski SWJ. A typology of virtual teams. *Group Organ Manage*. 2002;27:14-49. doi:10.1177/1059601102027 001003

80. Fernandez R, Vozenilek JA, Hegarty CB, et al. Developing expert medical teams: toward an evidence-based approach. *Acad Emerg Med*. 2008;15:1025-1036. doi:10.1111/j.1553-2712.2008.00232.x

81. Page DW. Professionalism and team care in the clinical setting. *Clin Anat*. 2006;19:468-472. doi:10.1002/ca.20304

82. Wilson KA. Promoting health care safety through training high reliability teams. *Qual Safety Health Care*. 2005;14:303-309. doi:10.1136/qshc.2004.010090

83. Powis DA, Bore M, Munro D. Selecting medical students: evidence based admissions procedures for medical students are being tested. *BMJ*. 2006;332:1156.2. doi:10.1136/bmj.332.7550. 1156-a

84. McManus IC, Powis DA, Wakeford R, Ferguson E, James D, Richards P. Intellectual aptitude tests and a levels for selecting UK school leaver entrants for medical school. *BMJ*. 2005;331: 555-559. doi:10.1136/bmj.331.7516.555

85. McManus IC, Ferguson E, Wakeford R, Powis D, James D. Predictive validity of the biomedical admissions test: an evaluation and case study. *Med Teach*. 2010;33:53-57. doi:10.3109/0142159x. 2010.525267

86. Chantrell S. Growth in emotional intelligence. Psychotherapy with a learning disabled girl. *J Child Psychother*. 2009;35: 157-174. doi:10.1080/00754170902996080

87. Andolsek KM. Improving the medical student performance evaluation to facilitate resident selection. *Acad Med*. 2016;91: 1475-1479. doi:10.1097/acm.0000000000001386

88. Edmond M, Roberson M, Hasan N. The dishonest dean?s letter: an analysis of 532 dean's letters from 99 U.S. medical schools. *Acad Med*. 1999;74:1033-1035. doi:10.1097/00001888-199909000-00019

89. Leiden LI, Miller GD. National survey of writers of dean?s letters for residency applications. *Acad Med*. 1986;61:943-953. doi:10. 1097/00001888-198612000-00001

90. Lurie SJ, Lambert DR, Nofziger AC, Epstein RM, Grady-Weliky TA. Relationship between peer assessment during medical school, dean's letter rankings, and ratings by internship directors. *J Gen Intern Med*. 2007;22:13-16. doi:10.1007/ s11606-007-0117-4

91. Naidich JB, Lee JY, Hansen EC, Smith LG. The meaning of excellence. *Acad Radiol*. 2007;14:1121-1126. doi:10.1016/j.acra.2007. 05.022

92. Kiefer CS, Colletti JE, Bellolio MF, et al. The "Good" dean?s letter. *Acad Med*. 2010;85:1705-1708. doi:10.1097/acm.0b013 e3181f55a10

93. Swide C, Lasater K, Dillman D. Perceived predictive value of the Medical Student Performance Evaluation (MSPE) in anesthesiology resident selection. *J Clin Anesth*. 2009;21: 38-43. doi:10.1016/j.jclinane.2008.06.019

94. Isaac C, Chertoff J, Lee B, Carnes M. Do students? and authors? genders affect evaluations? A linguistic analysis of medical student performance evaluations. *Acad Med*. 2011;86:59-66. doi:10.1097/acm.0b013e318200561d

95. Naidich JB, Grimaldi GM, Lombardi P, Davis LP, Naidich JJ. A program director's guide to the medical student performance evaluation (former dean's letter) with a database. *J Am Coll Radiol*. 2014;11:611-615. doi:10.1016/j.jacr.2013.11.012

96. Joshi ART, Vargo D, Mathis A, Love JN, Dhir T, Termuhlen PM. Surgical residency recruitment—opportunities for improvement. *J Surg Educ*. 2016;73:e104-e110. doi:10.1016/j.jsurg. 2016.09.005

97. Boatright D, Simon J, Jarou Z, et al. 167 Factors important to underrepresented minority applicants when selecting an emergency medicine program. *Ann Emerg Med*. 2015;66:S59-S60. doi:10.1016/j.annemergmed.2015.07.199

98. Boatright D, Branzetti J, Duong D, et al. Racial and ethnic diversity in academic emergency medicine: how far have we come? Next steps for the future. *AEM Educ Train*. 2018;2: S31-S39. Runde DP, ed. doi:10.1002/aet2.10204

99. Bramson R, Sadoski M, Sanders CW, van Walsum K, Wiprud R. A reliable and valid instrument to assess competency in basic surgical skills in second-year medical students. *South Med J*. 2007;100:985-990. doi:10.1097/smj.0b013e3181514a29

100. Stern DT, Papadakis M. The developing physician—becoming a professional. *N Engl J Med*. 2006;355:1794-1799. Cox M, Irby DM, eds. doi:10.1056/nejmra054783

101. Neumann M, Edelhäuser F, Tauschel D, et al. Empathy decline and its reasons: a systematic review of studies with medical students and residents. *Acad Med*. 2011;86:996-1009. doi:10.1097/ acm.0b013e318221e615

102. Hoffman KG, Brown RMA, Gay JW, Headrick LA. How an educational improvement project improved the summative evaluation of medical students. *Qual Safety Health Care*. 2009;18: 283-287. doi:10.1136/qshc.2008.027698

103. Westerman ME, Boe C, Bole R, et al. Evaluation of medical school grading variability in the United States. *Acad Med*. 2019; 94:1939-1945. doi:10.1097/acm.0000000000002843

104. Leydesdorff L, Ward J. Science shops: a kaleidoscope of science–society collaborations in Europe. *Public Underst Sci*. 2005;14:353-372. doi:10.1177/0963662505056612

105. Blaskiewicz R, Whitcomb W, Hoffstetter S. Medical student peer review: a unique evaluation tool. *Obstet Gynecol*. 1999; 93:S54. doi:10.1016/s0029-7844(99)90119-9

106. Shea JA, O?Grady E, Morrison G, Wagner BR, Morris JB. Medical student performance evaluations in 2005: an improvement over the former dean?s letter? *Acad Med*. 2008;83:284-291. doi:10.1097/acm.0b013e3181637bdd

107. Khaja MS, Jo A, Sherk WM, et al. Perspective on the new IR residency selection process: 4-year experience at a large, collaborative training program. *Acad Radiol*. Published online February 2021. doi:10.1016/j.acra.2021.01.028

108. Prideaux D, Roberts C, Eva K, et al. Assessment for selection for the health care professions and specialty training: consensus statement and recommendations from the Ottawa 2010 Conference. *Med Teach*. 2011;33:215-223. doi:10.3109/0142159x.2011.551560

109. Tsou K-I, Lin C-S, Cho S-L, et al. Using personal qualities assessment to measure the moral orientation and personal qualities of medical students in a Non-Western culture. *Eval Health Prof*. 2012;36:174-190. doi:10.1177/0163278712454138

110. Kjeldmand D, Holmström I. Balint groups as a means to increase job satisfaction and prevent burnout among general practitioners. *Ann Fam Med*. 2008;6:138-145. doi:10.1370/afm.813

111. Madrid PA, Schacher SJ. A critical concern: pediatrician self-care after disasters. *Pediatrics*. 2006;117:S454-S457. doi:10.1542/peds.2006-0099v

112. Maslach C, Leiter MP. Early predictors of job burnout and engagement. *J Applied Psychol*. 2008;93:498-512. doi:10.1037/0021-9010.93.3.498

113. Prins JT, Hoekstra-Weebers JEHM, Gazendam-Donofrio SM, et al. Burnout and engagement among resident doctors in the Netherlands: a national study. *Med Edu*. 2010;44:236-247. doi:10.1111/j.1365-2923.2009.03590.x

114. Schaufeli WB, Bakker AB, van der Heijden FMMA, Prins JT. Workaholism, burnout and well-being among junior doctors: the mediating role of role conflict. *Work Stress*. 2009;23:155-172. doi:10.1080/02678370902834021

115. Stafford L, Judd F. Mental health and occupational wellbeing of Australian gynaecologic oncologists. *Gynecol Oncol*. 2010;116: 526-532. doi:10.1016/j.ygyno.2009.10.080

116. West CP, Dyrbye LN, Shanafelt TD. Burnout in medical school deans. *Acad Med*. 2009;84:6. doi:10.1097/acm.0b013e318190147a

117. West CP, Halvorsen AJ, Swenson SL, McDonald FS. Burnout and distress among internal medicine program directors: results of a national survey. *J Gen Intern Med*. 2013;28:1056-1063. doi:10.1007/s11606-013-2349-9

118. Becich MJ, Gilbertson JR, Gupta D, Patel A, Grzybicki DM, Raab SS. Pathology and patient safety: the critical role of pathology informatics in error reduction and quality initiatives. *Clin Lab Med*. 2004;24:913-943. doi:10.1016/j.cll.2004.05.019

119. Croskerry P. Clinical cognition and diagnostic error: applications of a dual process model of reasoning. *Adv Health Sci Educ*. 2009;14:27-35. doi:10.1007/s10459-009-9182-2

120. Graber ML. Educational strategies to reduce diagnostic error: can you teach this stuff? *Adv Health Sci Educ*. 2009;14:63-69. doi:10.1007/s10459-009-9178-y

121. Jones BA, Meier FA. Patient safety in point-of-care testing. *Clin Lab Med*. 2004;24:997-1022. doi:10.1016/j.cll.2004.06.001

122. Murphy JG, Stee L, McEvoy MT, Oshiro J. Journal reporting of medical errors. *Chest*. 2007;131:890-896. doi:10.1378/chest.06-2420

123. Raab SS. Improving patient safety by examining pathology errors. *Clin Lab Med*. 2004;24:849-863. doi:10.1016/j.cll.2004.05.014

124. Tamuz M. Defining and classifying medical error: lessons for patient safety reporting systems. *Qual Safety Health Care*. 2004; 13:13-20. doi:10.1136/qshc.2002.003376

125. Weingart SN. Epidemiology of medical error. *BMJ*. 2000;320: 774-777. doi:10.1136/bmj.320.7237.774