Can Behavior-Based Interviews Reduce Bias in Fellowship Applicant Assessment?

Melissa L. Langhan, MD, MHS; Michael P. Goldman, MD; Gunjan Tiyyagura, MD, MHS

From the Departments of Pediatrics and Emergency Medicine, Section of Pediatric Emergency Medicine, Yale University School of Medicine, New Haven, Conn
The authors have no conflict of interest to disclose.
Address correspondence to Melissa Langhan, MD, MHS, Section of Pediatric Emergency Medicine, Yale University, 100 York St, Suite 1D, New Haven, CT 06510 (e-mail: Melissa.Langhan@yale.edu).
Received for publication March 29, 2021; accepted December 13, 2021.

ABSTRACT

PURPOSE: Components of trainee applications may introduce bias based on race or gender. Behavior-based interviews (BBIs) rely on structured questions to elicit applicants' past experiences to predict future behavior. Our objective was to implement BBIs in one fellowship program and compare applicant assessment by race and gender when using a standardized assessment tool versus a BBI-based tool.

METHODS: In 2019 and 2020, we developed BBIs and BBI-specific assessments; 6 of 15 faculty were trained in this interview method. Applicants completed 6 interviews with either a BBI or unstructured format. All faculty completed a standardized assessment on applicants. BBI faculty also completed a BBI-specific assessment. Normalized average scores were calculated and used to rank applicants into quartiles. Race was categorized into White, underrepresented minorities (URMs; Black and Hispanics), and non-URMs (all others). Faculty and applicants were surveyed about BBIs.

RESULTS: Seventy-five applicants were interviewed. Significant differences were found in standardized assessment scores (White 1.01 +/- 0.09, non-URM 1.02 +/- 0.08, URM .94 +/- 0.07; $P = .02$) and quartiles by race ($P = .05$), but not for BBI scores (White 0.98 +/- 0.09, non-URM 1.03 +/- 0.09, URM 1.02 +/- 0.1; $P = .18$) or quartiles by race ($P = .17$). There were no significant differences in score or quartile by gender for either tool. The majority of faculty and applicant survey respondents commented positively about BBIs.

CONCLUSION: BBIs were successfully implemented and generally reviewed positively by faculty and applicants. BBIs reduced racial differences in applicant assessments. Applicant assessment may benefit from structured tools to mitigate potential biases.

KEYWORDS: applicants; bias; interviews

ACADEMIC PEDIATRICS 2022;22:478−485

WHAT'S NEW

Behavior-based interviews (BBIs) are one structured interview technique. BBIs were successfully developed and implemented at a fellowship program. While applicant assessment using a standardized tool exhibited significant differences by race, the BBIs did not show this difference.

WHILE THERE ARE increasing numbers of women and underrepresented minorities (URMs) entering medical school, both gender and racial inequalities continue to exist throughout medicine.[1−3] Residency and fellowship selection committees seek diverse pools of trainees to enter their programs who are likely to succeed in training. Trainees submit an array of application materials when applying to residency or fellowship programs including a personal statement, letters of recommendation, test scores (USMLE or COMLEX-USA), medical school transcript and a medical school performance letter (MSPE) or Dean's letter. These components contribute varying weight during the applicant selection process and ranking.[4−6] Prior reports revealed that all of these components are subject to gender[7−10] and/or racial biases.[8,10−13]

While written application materials are important, many programs report that interviews are considered of equal or greater weight when making ranking decisions.[5,6,14,15] Traditional interviews are often unstructured, or informal and conversational in nature. There is little evidence that these unstructured interviews (UIs) have a strong correlation with future behavior or success in training.[4,16−18] Further, UIs may be subject to unconscious biases of the interviewer and have poor reliability.[19] While interviewers likely have some degree of unconscious bias,[20] the use of structured interviews and multiple interviewers may reduce its overall effect.[21,22] Behavior-based interviews (BBIs) are one form of structured interviews.[23] They aim to assess noncognitive traits such as integrity and motivation that may not be gleaned from other sources. BBI questions rely on the applicant's

Copyright © 2021 by Academic Pediatric Association

responses about past experiences and behaviors in contrast to multiple mini-interviews, which often rely on applicants' responses to hypothetical situations.[24] BBIs have shown increased reliability and validity as compared to UIs, including stronger correlations to future performance; however, data regarding the impact of BBIs on interviewer biases are lacking.[16,25] This reliability is further improved when rating scales incorporate behavioral anchors for scoring instead of less structured tools such as a visual analog scale.[26] Given the benefits of BBIs and other forms of structured interviews, these methods have been recommended as a way to improve the selection process.[19,22,27,28]

To reduce bias in our fellowship candidate selection process, we implemented BBIs with an anchored scoring rubric. Our primary objective was to compare applicant assessment by race and gender when using a standardized assessment tool versus a BBI-based tool. We also aimed to assess faculty and applicant perceptions about the BBI.

METHODS

SETTING AND PARTICIPANTS

This was a prospective observational study at a single Pediatric Emergency Medicine (PEM) fellowship program conducted over 2 consecutive applicant seasons from 2019 to 2020. On average, 38 applicants are interviewed annually. Fifteen PEM attendings interviewed applicants for our program.

DEVELOPMENT AND VALIDATION

We gathered validity evidence for the BBI following Messick's framework.[29]

DEVELOPMENT AND VALIDATION OF THE BBI

Content Validity: The development and implementation of BBIs is visually depicted in Figure 1. Review of the literature and discussions with institutional experts who had implemented BBIs for medical school applicants informed the development of our BBIs.[23,28,30,31] First, faculty and fellows participated in a group discussion to determine characteristics deemed important to success at our program. The group then ranked these characteristics in order of perceived importance. After introduction of the BBI concept, interested faculty self-nominated to participate in this process and were asked to rate which characteristics they were most interested in exploring with applicants during an interview.

The program director (M.L.L.) developed 1 or 2 behavior-based questions to assess each characteristic and a corresponding anchored scoring rubric. The rubric was on a 6-point scale and included concrete examples of behaviors or traits along the continuum.[32] For example, in a behavior-based question focused on ethics, the rubric would denote that clear unethical actions are rated as low, a response in where the applicant describes their positive ethical actions would be average, and a response in which the applicant facilitated a discussion of differing opinions to promote ethical actions among others would be scored highly. Sixteen questions and rubrics were distributed to participating faculty for critical revision.

Response Process: Faculty were trained on the use of BBIs prior to implementation. Training included written materials and a mock interview with a residency graduate who was employed in our department and applying for fellowship in a different field.[33] BBI faculty performed the mock interview together, practiced asking BBI questions and discussed the mock applicant's responses to reach consensus on the scoring rubric, after which final

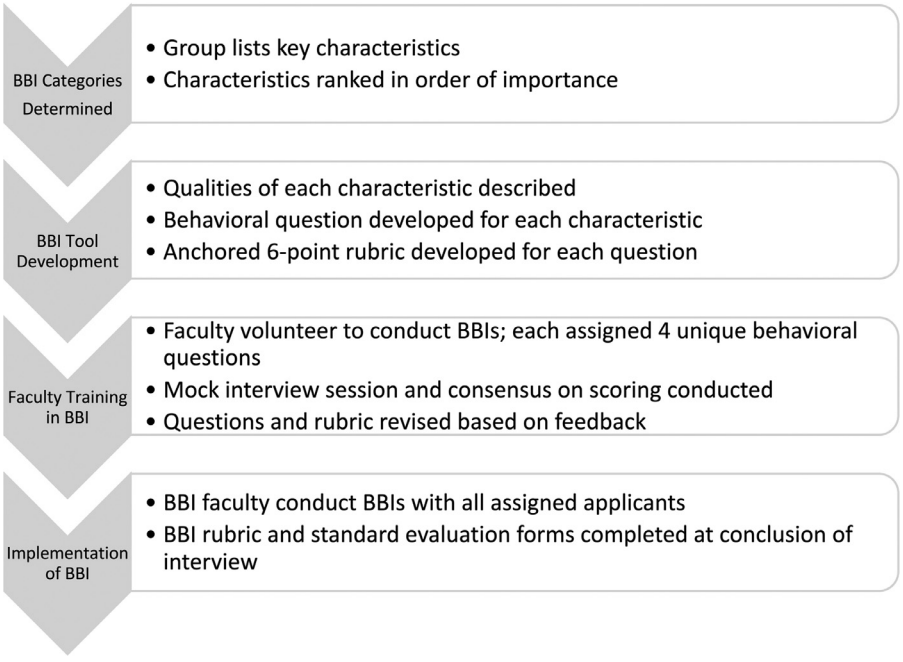


Figure 1. *Flow diagram of implementation methodology detailing development of BBIs.*

revisions to the BBI tools were made. Faculty performing BBIs were each assigned 4 unique questions. Each question was accompanied by an anchored scoring rubric specific to the trait being assessed. BBI faculty asked each applicant they interviewed the same 4 questions.

Applicant Procedure: Each applicant had one-on-one interviews with 5 PEM attendings and the Program Director. Interviews in 2019 were all conducted in person and in 2020 were all conducted virtually. Interviews were all 30 minutes in length. Given varying shift schedules and the desire of many faculty members to be involved in fellow selection, attendings assigned themselves to interview dates based on availability. Since 2019, 5 attendings were trained to perform BBIs. Four faculty (including the Program Director) performed BBIs in 2019; one left the institution and 2 additional faculty were recruited and trained in 2020. These faculty members performed BBIs with all applicants they interviewed, whereas remaining faculty performed UIs. The score from each of the 4 questions was added together for a summative BBI score.

Through the Electronic Residency Application Service (ERAS), all PEM fellowship applicants provided an ERAS application, curriculum vitae, personal statement, letters of recommendation, Board transcript (USMLE or COMLEX-USA), medical school transcript and MSPE. Of these materials, the ERAS application, personal statement, letters of recommendation and Board transcript were shared with the 6 faculty who interviewed the applicant. Each interviewer, including those completing BBIs, completed a standardized assessment tool for each applicant. This tool was composed of 7 categories derived from both the written application and interview; categories were scored on a 6-point scale with objective anchors and added together for a summative score. The 7 categories include: medical knowledge, academic scholarship, teaching, communication skills, initiative, letters of recommendation, and an overall assessment. Anchors for the standardized assessment tool were guided by objective data from the National Resident Match Program on characteristics of matched fellows in PEM. This report includes mean number of volunteer experiences, work experiences, research experiences, publications and mean USMLE scores, with 25th, 50th, and 75th percentiles for 3 consecutive years. The National Resident Match Program provides this information to program directors and designated institutional officials. Further details of our tools can be made available on request to the author.

Outcomes

Primary Outcome: Since individual faculty members interviewed a varying number of applicants, the summative scores from each faculty member were normalized. That is, the summative scores for each attending were averaged and the individual summative score for each applicant was divided by this average. Thus, scores were converted to a scale with 1.0 being average, scores >1 being above average and scores <1 being below average. The mean of the 6 normalized scores from the standardized assessment tool was calculated for each applicant. The mean of the normalized scores from the BBI tool was similarly calculated. Applicants had 2 to 4 BBI scores, depending upon the interviewers assigned that day. Mean scores from each tool were used to rank applicants into quartiles with the first quartile encompassing applicants at the top of the rank list and fourth quartile applicants at the bottom of the list. Applicant's self-identified gender and race were abstracted from the ERAS application. Race was categorized into White, URMs consisting of Black and Hispanic applicants, and non-URMs which comprised all other categories.

Secondary Outcome: Prior to the development of the BBI, faculty were sent an anonymous, web-based survey. Questions gauged perceptions about applicant interviews in general and BBIs using a 5-point Likert scale. BBI faculty were asked about their experiences at the end of the interview season through both an anonymous survey and informal discussion. After each interview season, applicants were sent an anonymous web-based survey after submission of their rank lists and asked to complete the survey prior to match day. The survey included one question about the BBI in a free text format, "How did you find the more structured interviews as compared to standard interviews?"

Analysis

Descriptive analysis of demographic variables and scores included frequencies with percentages and means with standard deviations (SD). Chi-square tests compared categorical variables. T-tests and analysis of variance, respectively, were used to analyze average scores by gender and race with 95% confidence intervals (CIs). Rank quartiles were compared using Mann-Whitney and Kruskal-Wallis tests by gender and race, respectively. A subanalysis compared standard scores and BBI scores among only BBI-trained faculty. Free text responses from the applicant survey were deductively coded by the authors into 3 categories (positive, neutral, and negative) and consensus achieved through open discussion. To achieve 80% power with a 2-tailed alpha of .05, a sample size of 21 applicants with 7 per racial group is needed to detect a difference of 0.1 in average scores between 3 groups with a standard deviation of .08. This study received exemption from the Yale University Human Research Protection Program.

Results

Seventy-five applicants were interviewed. Demographic variables of the applicants and faculty are presented in Table 1. BBI faculty were similar in race, but there was a higher proportion of female faculty compared to UI faculty. When comparing our applicants in 2019 and 2020, there were no differences in the gender or racial proportions of applicants. Similarly, there were no differences in the average assessment scores by race or gender on either tool when interviews were conducted in person in 2019 or virtually in 2020. The mean range of BBI

Table 1. Demographic Data for Applicants (N = 75) and Faculty (N = 15)

	N (%)
Applicants	
Gender	
Female	53 (71)
Male	22 (29)
Race/ethnicity	
White	42 (56)
Non-URM	21 (28)
URM	12 (16)
Faculty	
Gender	
Female	6 (36%)
Male	9 (64%)
Race/ethnicity	
White	13 (87%)
Non-URM	2 (13%)
URM	1 (7%)
Faculty rank	
Assistant professor	6 (40%)
Associate professor	7 (47%)
Professor	2 (13%)
Average years interviewing,	
1–5	3 (20%)
6–11	6 (40%)
12–17	2 (13%)
>18	4 (27%)

URM indicates underrepresented minorities.

scores was 2.8 points (SD 0.77) and the mean range of UI scores was 2.4 points (0.66). Of the 6 interviews, applicants had a median of 3 BBIs. Although there were no differences in the number of BBIs performed by race, males had a higher number of BBIs as compared to females (4 vs 3 respectively, $P = .016$). There was no significant difference in an applicant's mean BBI score based on the number of BBI interviews conducted (4 BBI mean score 0.988 vs 3 BBI mean score 1.01 vs 2 BBI mean score 0.989, $P = .54$).

There were statistically significant differences by race in standardized assessment scores, but not in the BBI scores (Table 2). Standardized assessment scores were above average for White and non-URM applicants, whereas the mean score for URM applicant was below average ($P = .024$). BBI scores were not significantly different between racial groups ($P = .18$). When analyzing only the standardized assessment scores of BBI interviewers, there remained a similar significant difference by race (White 1.01 [SD 0.09], non-URM 1.02 [SD 0.09], URM 0.94 [SD 0.07]; $P = .02$). There were significant differences among rank quartiles by race based on the standardized assessment, but not the BBI tool (Fig. 2).

There were no significant differences in average score by gender for either tool (Table 2). The mean difference in standardized assessment scores between females and males was -0.015 (95% CI -0.06, 0.03, $P = .98$). The mean difference in BBI scores between females and males was 0.004 (95% CI -0.04, 0.05, $P = .69$). There were no significant differences in rank quartiles by gender for either tool (Fig. 3).

Fourteen of 15 (93%) interviewing faculty completed the preseason survey (Table 3). All agreed that interviews were important in their ranking decision. Equal numbers agreed and disagreed about receiving training on performing interviews, and whether they believed a BBI would change their applicant rankings. A third of faculty were unfamiliar with BBIs. All BBI faculty had positive feedback about the experience with the new format and noted how they obtained useful and potentially unique information from the questions. For most interviews, BBI faculty were able to ask all 4 questions and had time for open discussion.

Fifty applicants (67%) responded to the post-season survey; 36 (72%) answered the open-ended BBI question. Positive comments about BBI were provided by 15 (42%) respondents. One noted, "I really liked the preset questions. I feel like I was able to express myself and my desire for pediatric emergency medicine much better than the standard 'tell me about yourself.'" Neutral comments were reported by 10 (28%) respondents, such as "It was nice to have both. I liked that some were free style and pick what to talk about but then liked that some interviews offered specific questions that helped guide the interview." Negative comments were provided by 11 (31%) respondents. One such comment was "I preferred the standard interviews as they gave me the opportunity to pursue whatever topics flowed naturally between the interviewer and myself." There was no significant difference in proportion of positive (45% vs 38%), neutral (25% vs 31%) and negative (30% vs 31%) comments between years 1 and 2, respectively.

DISCUSSION

After determining characteristics believed to be associated with successful training at our PEM fellowship program, we successfully developed BBI questions and

Table 2. Mean Scores of Applicants for Each Tool

Applicant Variable	Mean Standardized Assessment Score (SD)	Mean BBI-Based Score (SD)
Gender		
Female	0.997 (0.092)	1.002 (0.096)
Male	1.012 (0.092)	0.998 (0.089)
Race		
White	1.008 (0.094)*	0.983 (0.092)
Non-URM	1.024 (0.084)*	1.028 (0.089)
URM	0.939 (0.072)*	1.001 (0.094)

URM indicates underrepresented minorities; SD, standard deviation; and BBI, behavior-based interview.
*Statistically significant difference.

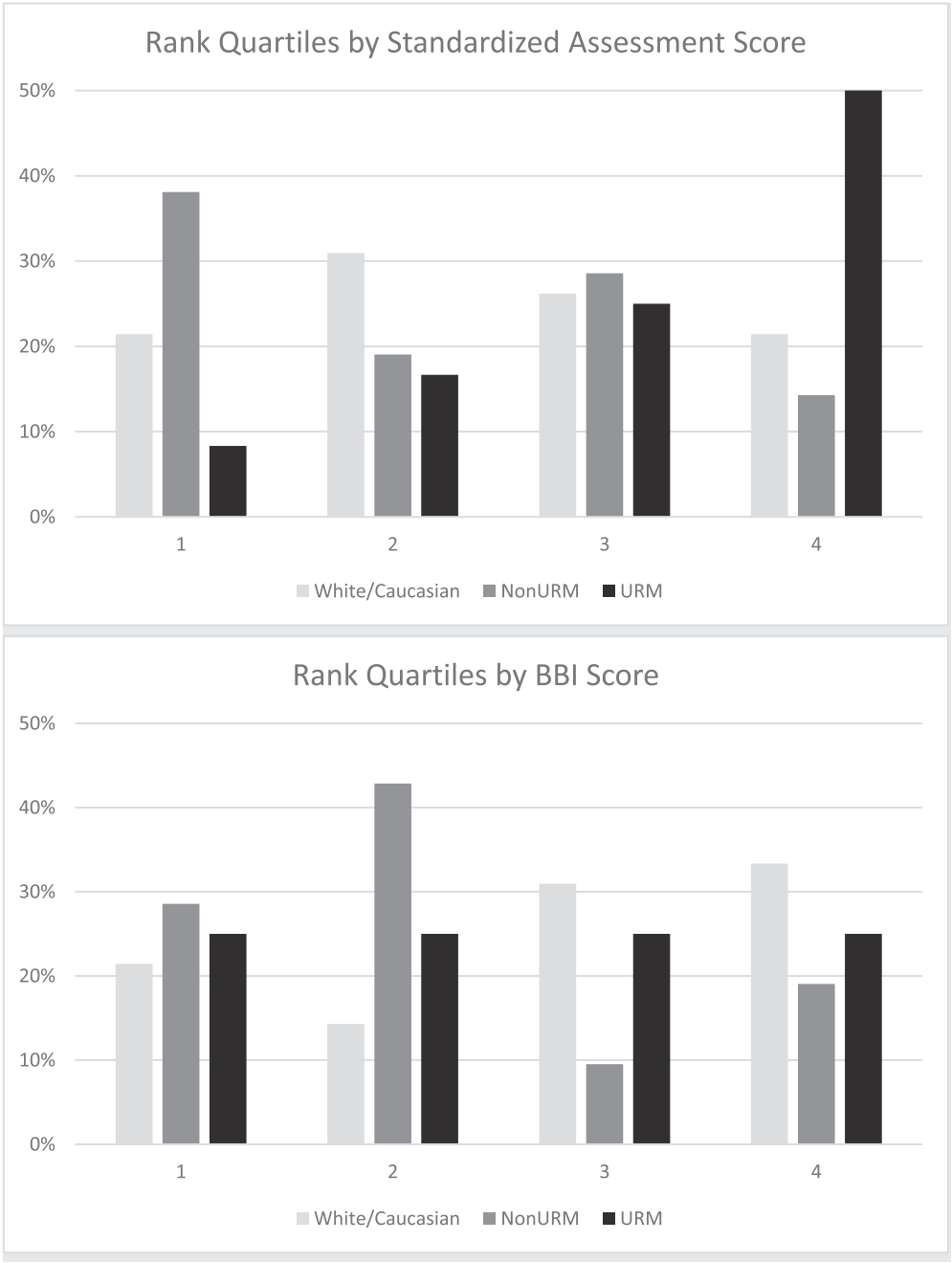


Figure 2. *Rank quartiles by self-identified race for each scoring tool (N = 75).*

corresponding anchored rubrics which were implemented both in person and virtually across 2 interview seasons. Our study demonstrated that while there were no differences in applicant assessments by gender, significant differences in standardized assessment scores were found by race. These differences were mitigated when using a BBI-based assessment tool. Both faculty and applicants generally perceived BBIs in a positive light. When examining the scores of only BBI faculty, differences in ranking by race remained significant with the standardized assessment tool, implying that the differences likely arose from components within the tool itself used to evaluate the applicants and not necessarily related to a variance in biases among the faculty using the different interview techniques.

As diversity, or lack thereof, is being increasingly recognized in medicine, it is paramount that we critically evaluate the systems in place to assess and advance students and trainees. Inherent biases found in application materials needs to be actively counteracted by admissions committees, program directors and those in leadership positions as few if any factors derived from application materials reliably predict future success.[17,18,34] Unlike the standardized assessment, the BBI-based assessment used only information gleaned from the interview. It is unclear to what degree the written application materials (which factored into scoring of the standardized assessment tool)

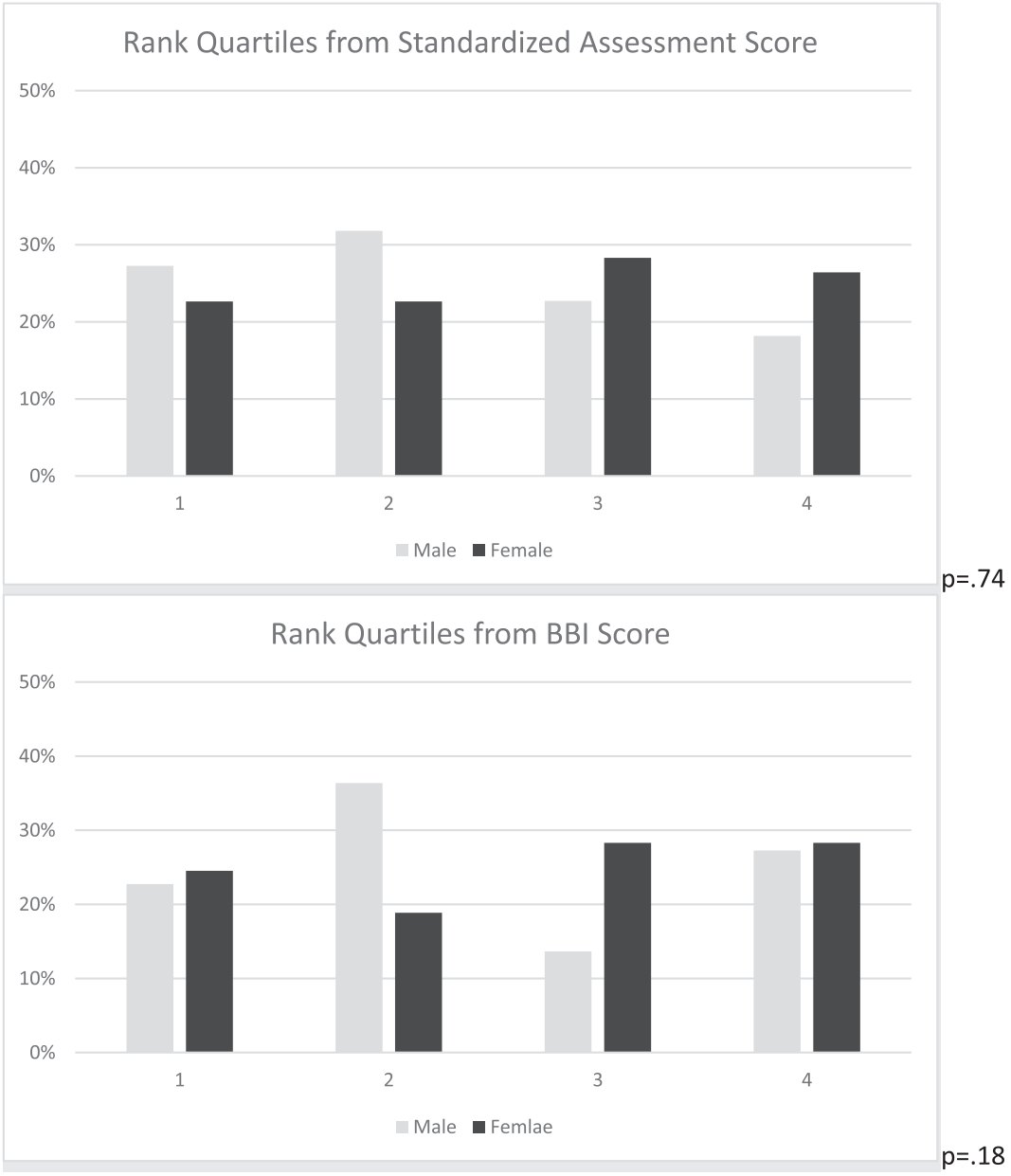


Figure 3. *Rank quartiles by self-identified gender for each scoring tool (N = 75).*

versus the UI format effected scores and led to the racial differences seen here. To assess further, we analyzed the subcomponents of our standardized evaluation tool and found statistically significant differences by race to exist in only one category, which relies on USMLE scores and accolades such as membership in Alpha Omega Alpha. When this component was removed, URM applicants continued to have lower than average mean scores, however there was no longer a statistically significant difference between racial groups. We plan on removing this category from the standardized assessment tool in the next applicant cycle and further revising the tool to try to reduce bias. Even though faculty were exposed to application information prior to their interview, it did not seem to negatively affect their BBI-based scores.

Many specialties have reported on the importance of the interview in recruitment, with some stating that their rank list was primarily determined by the applicant interview.[4,5,14,15,34,35] Despite the perception both by our faculty and in the literature that interviews are important in ranking applicants, it is interesting to note that most faculty reported no training in interviewing. Given this lack of training and evidence supporting the presence of implicit biases in trainee selection,[20,21] it is reasonable that faculty may unconsciously evaluate applicants on demographic factors or overall personality when performing UIs despite the belief that UIs lead to objective ranking decisions. In a literature review, Lee discussed the lack of validity and reliability of UIs including concerns that interviewers may interpret similar information differently depending upon the applicant and could be more heavily influenced by unfavorable perceptions.[19]

Different types of structured interviews have been successfully implemented, yet the extent of their use is

Table 3. Faculty Perceptions About Interviews for Fellowship Applicants

	Strongly Agree	Somewhat Agree	Neither Agree nor Disagree	Somewhat Disagree	Strongly Disagree
The interview is an important part of my ranking decision	10 (71%)	4 (29%)	0	0	0
I received training on how to interview others for a job position	1 (7%)	5 (36%)	2 (14%)	5 (36%)	1 (7%)
I am familiar with BBIs	0	9 (64%)	0	2 (14%)	3 (21%)
Compared to UIs, BBIs better allow the interviewer to assess those qualities which are important to be a physician.	1 (7%)	5 (36%)	3 (21%)	3 (21%)	2 (14%)
Compared to UIs, BBIs are difficult to perform	0	0	9 (64%)	5 (36%)	0
I am confident in my ability to provide an objective and reliable rank-ordering of fellowship applicants.	2 (14%)	9 (64%)	2 (14%)	1 (7%)	0
I think that BBI will change the way I rank applicants.	2 (14%)	2 (14%)	6 (43%)	2 (14%)	2 (14%)

UI indicates unstructured interview; BBI, behavior-based interview.

unclear.[16] While the BBI faculty and a portion of the applicants described positive experiences with the BBI, several faculty members were reluctant to relinquish the flexibility of the UIs and some applicants also reported a preference for the conversational format that UIs offered. One Canadian Emergency Medicine program examined the process of BBIs in residency selection after establishing performance criteria for applicants.[36,37] While they did not find improved reliability of BBIs compared to UIs, this may be due to the use of a visual analog scale with simple anchors of best, average and worst.[37,38] The use of increasing structure in the interview and scoring tools have been associated with improved reliability in scoring.[25,26] Although BBIs may lead to decreased bias and improve reliability in the interview process, further study to examine the acceptability of the BBI for faculty and applicants and to study the impact on predicting fellow success/outcomes is necessary prior to wide dissemination.

This study is limited to a single fellowship program and may not be generalizable to other programs. The BBI questions and scoring rubric were specifically designed to address traits that our faculty and fellows felt were important to success in training, such as initiative, accountability and professionalism. When implementing BBIs, a program must determine which characteristics they value in their trainees.[23] While a minority of our applicants self-identified as URMs, a significant difference in scores and ranking with 2 different tools was still demonstrated. However, we were unable to assess if there were differences by individual race categories due to our limited numbers. Currently, there is little literature assessing biases associated with the interview process and assessment of applicant interviews. Our program has a large faculty selection committee with limited diversity and applicants are interviewed by different faculty members. Both increased diversity and variation in interviewers has been shown to have positive effects on an interview's "fairness" (AAMC). Whereas faculty volunteered to conduct BBIs and may have inherently different characteristics than those who did not volunteer, these individuals appear to have the same unconscious biases as those performing UIs based on our analyses. Similarly, while there were no differences in average scores when comparing the years with in-person interviews and virtual interview, the change in modality may have unrecognized impact. Last, while we did find statistically significant differences among our scoring tools by race, the clinical significance of these differences is more difficult to assess. Future multi-institutional efforts may provide better data for comparing the impact of unstructured versus behavior-based tools on racial and gender-based biases. Whether interviewers should rate applicants solely on their interview or should also incorporate application materials is a question that a program's leadership should consider.

CONCLUSIONS

After program specific data gathering and brief training, a BBI with a structured scoring rubric was successfully implemented to evaluate PEM fellowship applicants on noncognitive characteristics. While no gender differences were revealed, significantly different applicant scores and subsequent rankings by race occurred when using different tools to assess fellowship candidates. A BBI with a specific scoring rubric was not prone to the same racial

biases found when using a standardized assessment. BBIs were generally viewed positively by faculty and fellow applicants. Improvements to address applicant concerns and evaluating the validity of BBIs may further enhance acceptability.

ACKNOWLEDGMENT

Financial statement: None.

REFERENCES

1. Lett LA, Murdock HM, Orji WU, et al. Trends in racial/ethnic representation among US medical students. *JAMA Netw Open*. 2019;2:e1910490.
2. Fassiotto M, Flores B, Victor R, et al. Rank equity index: measuring parity in the advancement of underrepresented populations in academic medicine. *Acad Med*. 2020;95:1844–1852.
3. Rotbart HA, McMillen D, Taussig H, et al. Assessing gender equity in a large academic department of pediatrics. *Acad Med*. 2012;87:98–104.
4. Shiroma PR, Alarcon RD. Selection factors among international medical graduates and psychiatric residency performance. *Acad Psychiatry*. 2010;34:128–131.
5. Hale MKP, Frank JR, Cheung WJ. Resident selection for emergency medicine specialty training in Canada: a survey of existing practice with recommendations for programs, applicants, and references. *Can J Emerg Med*. 2020;22:829–835.
6. Beheshtian E, Jalilianhasanpour R, Sahraian S, et al. Fellowship candidate factors considered by program directors. *J Am Coll Radiol*. 2020;17:284–288.
7. Demzik A, Filippou P, Chew C, et al. Gender-based differences in urology residency applicant personal statements. *Urology*. 2020 [epub ahead of print]. https://doi.org/10.1016/j.urology.2020.08.066.
8. Powers A, Gerull KM, Rothman R, et al. Race- and gender-based differences in descriptions of applicants in the letters of recommendation for orthopaedic surgery residency. *JB JS Open Access*. 2020;5:e20.00023. https://doi.org/10.2106/JBJS.OA.20.00023.
9. Li S, Fant AL, McCarthy DM, et al. Gender differences in language of standardized letter of evaluation narratives for emergency medicine residency applicants. *AEM Educ Train*. 2017;1:334–339.
10. Ross DA, Boatright D, Nunez-Smith M, et al. Differences in words used to describe racial and gender groups in Medical Student Performance Evaluations. *PLoS One*. 2017;12:e0181659.
11. Low D, Pollack SW, Liao ZC, et al. Racial/ethnic disparities in clinical grading in medical school. *Teach Learn Med*. 2019;31:487–496.
12. Williams M, Kim EJ, Pappas K, et al. The impact of United States Medical Licensing Exam (USMLE) step 1 cutoff scores on recruitment of underrepresented minorities in medicine: a retrospective cross-sectional study. *Health Sci Rep*. 2020;3:e2161.
13. Boatright D, Ross D, O'Connor P, et al. Racial disparities in medical student membership in the Alpha Omega Alpha Honor Society. *JAMA Intern Med*. 2017;177:659–665.
14. Swanson WS, Harris MC, Master C, et al. The impact of the interview in pediatric residency selection. *Ambul Pediatr*. 2005;5:216–220.
15. Nallasamy S, Uhler T, Nallasamy N, et al. Ophthalmology resident selection: current trends in selection criteria and improving the process. *Ophthalmology*. 2010;117:1041–1047.
16. Stephenson-Famy A, Houmard BS, Oberoi S, et al. Use of the interview in resident candidate selection: a review of the literature. *J Grad Med Educ*. 2015;7:539–548.
17. Dubovsky SL, Gendel MH, Dubovsky AN, et al. Can admissions interviews predict performance in residency? *Acad Psychiatry*. 2008;32:498–503.
18. Burkhardt JC, Parekh KP, Gallahue FE, et al. A critical disconnect: residency selection factors lack correlation with intern performance. *J Grad Med Educ*. 2020;12:696–704.
19. Lee AG, Golnik KC, Oetting TA, et al. Re-engineering the resident applicant selection process in ophthalmology: a literature review and recommendations for improvement. *Surv Ophthalmol*. 2008;53:164–176.
20. Capers IV Q, Clinchot D, McDougle L, et al. Implicit racial bias in medical school admissions. *Acad Med*. 2017;92:365–369.
21. Griffin BN, Wilson IG. Interviewer bias in medical student selection. *Med J Aust*. 2010;193:343–346.
22. Association of American Medical Colleges. *Best Practices for Conducting Residency Program Interviews*. Washington, DC: Association of American Medical Colleges; 2016.
23. Tatem G, Kokas M, Smith CL, et al. A feasibility assessment of behavioral-based interviewing to improve candidate selection for a pulmonary and critical care medicine fellowship program. *Ann Am Thorac Soc*. 2017;14:576–583.
24. Jones PE, Forister JG. A comparison of behavioral and multiple mini-interview formats in physician assistant program admissions. *J Physician Assist Educ*. 2011;22:36–40.
25. Bandiera G, Regehr G. Reliability of a structured interview scoring instrument for a Canadian postgraduate emergency medicine training program. *Acad Emerg Med*. 2004;11:27–32.
26. Taylor PJ, Small B. Asking applicants what they *would do* versus what the *did do:* A meta-analytic comparison of situational behaviour and past behaviour employment interview questions. *J Occ Organiz Psychol*. 2002;75:277–294.
27. Fuchs JW, Youmans QR. Mitigating bias in the era of virtual residency and fellowship interviews. *J Grad Med Educ*. 2020;12:674–677.
28. Prager JD, Myer CMt, Hayes KM, et al. Improving methods of resident selection. *Laryngoscope*. 2010;120:2391–2398.
29. Cook DA, Beckman TJ. Current concepts in validity and reliability for psychometric instruments: theory and application. *Am J Med*. 2006;119. 166.e167-116.
30. McArthur TA, Flug JA, Restauri N. Behavioral interviewing: integrating ACGME competency-based questions into the radiology resident selection process. *Curr Probl Diagn Radiol*. 2017;46:91–94.
31. Easdown LJ, Castro PL, Shinkle EP, et al. The behavioral interview, a method to evaluate ACGME competencies in resident selection: a pilot project. *J Educ Perioper Med*. 2005;7:E032.
32. Kell HJ, Martin-Raugh MP, Carney LM, et al. Exploring methods for developing behaviorally anchored rating scales for evaluating structured interview performance. *ETS Res Rep Ser*. 2017;1:1–26.
33. Gill BC, Myers E, Colpo D, et al. Structured interviewer training for the implementation of standardized behavioral questions in medical school admissions. *Simul Healthc*. 2020;15:128–132.
34. Brothers TE, Wetherholt S. Importance of the faculty interview during the resident application process. *J Surg Educ*. 2007;64:378–385.
35. Poirier MP, Pruitt CW. Factors used by pediatric emergency medicine program directors to select their fellows. *Pediatr Emerg Care*. 2003;19:157–161.
36. Blouin D, Dagnone JD. Performance criteria for emergency medicine residents: a job analysis. *Can J Emerg Med*. 2008;10:539–544.
37. Blouin D. Reliability of a structured interview for admission to an emergency medicine residency program. *Teach Learn Med*. 2010;22:246–250.
38. Blouin D, Day AG, Pavlov A. Comparative reliability of structured versus unstructured interviews in the admission process of a residency program. *J Grad Med Educ*. 2011;3:517–523.