



Reimagining Merit and Representation: Promoting Equity and Reducing Bias in GME Through Holistic Review

Nicolás E. Barceló¹ · Sonya Shadravan¹ · Christine R. Wells¹ · Nichole Goodsmith² · Brittany Tarrant¹ · Trevor Shaddox¹ · Yvonne Yang¹ · Eraka Bath¹ · Katrina DeBonis¹

Received: 8 December 2019 / Accepted: 30 September 2020 / Published online: 27 October 2020
© Academic Psychiatry 2020

Abstract

Objective This study aims to evaluate the capacity of a holistic review process in comparison with non-holistic approaches to facilitate mission-driven recruitment in residency interview screening and selection, with particular attention to the promotion of race equity for applicants underrepresented in medicine (URM).

Methods Five hundred forty-seven applicants to a psychiatry residency program from US allopathic medical schools were evaluated for interview selection via three distinct screening rubrics—one holistic approach (Holistic Review; HR) and two non-holistic processes: Traditional (TR) and Traditional Modified (TM). Each applicant was assigned a composite score corresponding to each rubric, and the top 100 applicants in each rubric were identified as selected for interview. Odds ratios (OR) of selection for interview according to URM status and secondary outcomes, including clinical performance and lived experience, were measured by analysis of group composition via univariate logistic regression.

Results Relative to Traditional, Holistic Review significantly increased the odds of URM applicant selection for interview (TR-OR: 0.35 vs HR-OR: 0.84, $p < 0.01$). Assigning value to lived experience and de-emphasizing USMLE STEP1 scores contributed to the significant changes in odds ratio of interview selection for URM applicants.

Conclusions Traditional interview selection methods systematically exclude URM applicants from consideration without due attention to applicant strengths or potential contribution to clinical care. Conversely, holistic screening represents a structural intervention capable of critically examining measures of merit, reducing bias, and increasing URM representation in residency recruitment, screening, and selection.

Keywords Equity · Admissions · Holistic review · Diversity · GME

Despite over 30 years of reported institutional intent to diversify the medical workforce, racial disparity in medical training persists [1, 2]. Attempts to address overt discrimination toward individual applicants have fallen short in impact: only prior to 1978 were there fewer Black men in medical training than today [3]. Similarly, interventions to promote workforce diversity for the potential health benefit to marginalized groups [4, 5] have not achieved their goal: at current rates of graduation, five centuries will pass before Latinos are

proportionally represented in California's physician workforce [6]. These findings call into question the efficacy of existing frameworks intended to promote diversity as well as the admissions procedures that account for trends in racial/ethnic representation of medical trainees.

A growing body of literature focused on trainee evaluation and promotion in medical education argues that existing definitions of “excellence” [7], as well as the metrics that serve as their evidence, reflect presumed racial hierarchies, obscure an uneven playing field, and do not reliably predict long term clinical contribution [8–13]. These processes thereby foster an ongoing segregation of marginalized groups to the detriment of medicine [14, 15]. In an effort to correct the systemic bias that propagates disparity, the Association of American Medical Colleges (AAMC) encourages the use of Holistic Review (HR)—“a flexible, individualized way of assessing an applicant's capabilities by which balanced consideration

✉ Nicolás E. Barceló
nbarcelo@mednet.ucla.edu

¹ University of California, Los Angeles, Los Angeles, CA, USA

² Greater Los Angeles Veterans Affairs Healthcare System, Los Angeles, CA, USA

is given to experiences, attributes, and academic metrics...” [16]. Despite its widespread implementation in medical schools nationwide, there are few studies examining the impact of HR on residency admissions [12].

This manuscript examines the potential of a holistic review screening process to dismantle bias and promote equitable representation relative to non-holistic processes in graduate medical education (GME) interview selection at one residency program [17]. In turn, it examines and quantifies how bias and privilege, both embedded in traditional GME selection procedures, influence patterns of representation.

Methods

Study Design and Population

The studied psychiatry residency program received a total of 806 applications in the 2018–2019 cycle. To maximize the internal validity of the study and increase consistency across application review, inclusion criteria consisted of graduates from allopathic US medical schools and excluded applicants with felony convictions or unexplained misdemeanors and graduates from schools without comparative data for overall performance or clinical rotations. Due to restrictions from the sponsoring institution, non-US citizens with Green Card and students applying for J-1 visa were excluded. Representation in medicine (RM) for study participants, including both underrepresented in medicine (URM) and not underrepresented in medicine (nURM), was based on self-identified race/ethnicity in the Electronic Residency Application Service (ERAS) application and using frameworks provided by the AAMC and in-state partners [16, 18].

This study qualified for exemption by the Institutional Review Boards at the University of California, Los Angeles.

Interventions

The Holistic Review (HR) tool was subjected to comparative analysis relative to two non-holistic models: the “Traditional” (TR) approach, which approximated rubrics previously used at the study site and reflects standards of practice in GME screening nationwide, and the “Traditional Modified” (TM), which expanded the Traditional approach by leveraging ERAS filters to attempt a more equity-enhancing approach without full holistic review.

Holistic Review

Residency program stakeholders developed a mission-driven and domains-based approach to applicant evaluation. This process included (1) identifying and devaluing metrics with known bias and limited predictive value for long-term clinical

strength (AOA induction, USMLE scores) [8, 9, 19], (2) reimagining and prioritizing personal qualities and professional characteristics that reflect program values, and (3) actively considering applicants in a broader social context—including acknowledgment of how institutional racism, poverty, and family educational achievement can impact applicant trajectory through medical school.

The Holistic Review approach ultimately resulted in a strengths-based rubric composed of eight domains. These domains were assessed according to criteria gleaned from multiple application elements—including curriculum vitae, personal statement, and the Medical Student Performance Evaluation (MPSE). Six categorical domains included (i) Leadership, (ii) Community Service, (iii) Clinical Performance, (iv) Research, (v) Reference Letters, and (vi) Professionalism and were scored according to duration, intensity, and degree of achievement (e.g. for leadership—Which position was held? For how long? What was the impact of experience?). With regard to professionalism, instances of unprofessional behavior generated a negative point score. Two measures of lived experience were included as binary outcomes, each with its own rubric: Resilience (achievement in enduring adversity—e.g. personal setback, illness, discrimination) and Distance Travelled (trajectory relative to family or community-level barriers reflecting marginalization at a population or structural level—e.g. first-generation college graduate, raised in community with high poverty/low educational resources). Self-identified race/ethnicity was not provided to reviewers as part of ERAS application material, and race/ethnicity was not explicitly considered in any Holistic Review domain nor was it sufficient to earn *de facto* recognition for Resilience or Distance Travelled. The pool of Holistic reviewers included faculty and residents trained in application of the Holistic Review rubric.

To generate a preliminary Holistic Review score, categorical domains were summed, with a weighted emphasis given to the Clinical domain. Where applicants demonstrated notable resilience and/or distance travelled, the preliminary score was multiplied by 1.1 (or 1.2 if both were noted) for the final Holistic Review composite score (see Table 1).

In light of mounting evidence of racial bias in STEP1 scores and their limited utility in predicting clinical performance, standardized scores were not included in the Holistic Review rubric [9, 20–22]. Additional elements, including “yellow” and “red” flags (e.g. significant difficulty in or repeated failure of clinical rotations), were included in the Holistic Review evaluation form but were not included in the composite score nor were they essential to this analysis.

Traditional Review

The Traditional Review rubric was based on screening rubrics previously used at the study site and parallels rubrics currently in use at residency programs nationwide. It relied on

Table 1 Holistic review rubric overview

Major Domains:	1. Each Major domain divided across 4 tiers of increasing excellence.
- Leadership	2. Tiers defined according to multiple considerations including the duration, intensity, role, and impact of the experience.
- Community	
- Research	
- Clinical	
- Reference Letters	
Professionalism	1. Professionalism was divided across 3 tiers of performance/behavior - consideration given to instances of “unprofessional behavior” as well as to professionalism “above the expected level” of medical student trainee.
Resilience and Distance Traveled	1. Divided across 2 tiers 2. Consideration to applicants life experience including achievement despite personal hardship or structural barriers.
Holistic Review (HR) Composite Score	1. Preliminary composite score with weighted preference to Clinical domain and “unprofessional” behavior with negative value. 2. Resilience and Distance Traveled with multiplicative (1.1–1.2 or 110–120%) contribution = HR Composite Score.

manipulation of discrete elements of the ERAS application, including graduating medical school, additional graduate degrees, STEP 1 scores, Alpha Omega Alpha Honors Society (AOA), Gold Humanism Honors Society (GHHS), peer-reviewed publications, and poster presentations. Following their transformation into either binary variables (e.g. AOA—yes/no) or, for continuous outcomes, into discreet categories

(e.g. STEP1 score), these metrics were combined to provide one composite Traditional Review score (see Table 2).

Traditional Modified Review

The Traditional Modified rubric represented an intermediary step between Traditional Review and Holistic Review and

Table 2 Traditional Review and Traditional Modified rubric overview

Traditional Rubric (TR)	
Continuous outcomes STEP1, posters, peer-reviewed articles, school rank ^a	1. Range of values broken into discreet categories. 2. Categories assigned increasing point value.
Binary outcomes AOA, GHHS, PhD, Masters	1. Yes/No and assigned point value reflecting significance of accomplishment.
Traditional (TR) Composite Score	1. Individual component scores summed into single composite score = TR Composite Score.
Traditional Modified (TM)	
Keywords ^b	1. Keywords were selected by multiple stakeholders in the recruitment process and would reflect program-specific values/mission—e.g. promoting health equity, increasing social justice. 2. ERAS platform was used to filter applicants according to use of identified keywords—describing past experience in research, volunteer, and work history. 3. Program recruitment stakeholders reviewed keyword use to ensure appropriate contextual meaning. e.g.: “diverse populations” vs “astrocyte physiological diversity” e.g.: “promotion of health equity” vs “work in a private equity group” 1 Point granted to applicants with experience or interest in program identified matters of importance—2-point maximum added score.
Traditional (TM) Composite Score	1. TR Composite score supplemented by Keywords score = TM Composite Score.

^a “School rank” generated via *US News and World Report* list of US Medical School rankings 2017–2018—including both Research and Primary Care criteria

was designed to examine whether more elaborate ERAS filters can be used to increase URM applicant representation through piecemeal modification of the Traditional Review approach without full holistic review. Through this method, “keywords” identified by the working group (e.g. “equity”) could be used as search filters within ERAS and potentially identify applicants with experience in areas related to equity, diversity, and/or inclusion. This would allow for enhancement of the Traditional Review score with only brief review of application materials to ensure relevant utilization of keywords (see Table 2).

Measures and Outcomes

Baseline applicant measures were obtained via AAMC ERAS. Each applicant was assigned three scores, one score reflecting their performance according to each of the three rubrics. Applicants selected for interview were identified by ranking in the top 100 applicants through a given screening approach—Traditional Review, Traditional Modified, or Holistic Review. The primary outcome was odds of selection for interview by URM status according to each screening tool. Secondary outcomes included odds of interview selection by individual elements of each rubric. Predicted probabilities of interview selection according to RM status were examined as alternative means to compare the relative contribution of each rubric to interview selection in light of differences in the number of applicants in each group.

Statistical Analysis

Statistical difference in the distribution of selection criteria by RM status was evaluated by chi-squared testing. This approach has been previously supported in the literature despite a value of zero for some categories of analysis [23].

We used logistic regression to model the binary outcome of “interview selection” according to exposure by review approach and as a function of each applicant dimension (e.g. URM, STEP1). In this manner, review paradigm served as the independent variable and was operationalized as a three-level nominal value: Traditional Review, Traditional Modified, and Holistic Review. In each regression, we included the interaction between review paradigm (e.g. Holistic Review) and a specified applicant dimension (e.g. URM) to quantify the association between review approach, applicant dimension, and interview selection. This approach generated an interaction term reflecting the relative impact of each review paradigm in determining interview selection according to applicant criteria. For example, a positive interaction term between Holistic Review and URM implies that Holistic Review strengthens the association between URM status and interview selection. To more intuitively display the impact of review approach to interview selection by applicant

dimension, we show the odds ratio and confidence interval for each dimension separately under each review paradigm.

We generated marginal predicted probabilities of selection for interview. These communicate the extent to which review paradigm influences the subsequent chance an applicant would have to be selected for interview according to a given dimension of their application. Because nURM applicants grossly outnumber URM applicants in our data, odds ratios obscure the absolute probability of interview selection. We therefore included predicted probabilities to convey a more intuitive measure of absolute effects, reengaging with the overwhelming predominance of nURM applicants in our model.

Analysis were conducted in, Stata 15.1 (Stata Corp, College Station, TX) and R Statistical Programming software 3.5.2 (R Core Team, Vienna, Austria).

Results

Representation of URM and Non-URM Applicants by Scoring Domain

Of 806 total applications, 574 applicants met inclusion criteria, including 154 (27%) URM applicants. Baseline characteristics, including distribution according to variables of interest by RM status, are shown in Table 3.

Regarding Traditional Review domains, nURM applicants were more likely than URM applicants to have STEP 1 scores above 240 (33% vs 12%; $p < 0.001$) and be AOA inductees (9.5% vs 2%; $p = 0.004$). There were no significant differences between groups in relation to GHHS membership, school in the highest ranked tier, distribution of additional graduate degrees (PhD or master’s), or number of applicants in the highest rank tier of either posters or peer reviewed publications.

Under the Traditional Modified rubric, URM applicants were more likely than nURM to have 2 or more relevant keywords in their applications (27% vs 15%; $p = 0.002$).

Using the Holistic rubric, URM applicants were more likely to be scored in the highest tier for resilience (37% vs 20%; $p < 0.01$) and distance travelled (31% vs 12%; $p < 0.01$) and less likely to be scored in the highest tier for clinical (8% vs 21%; $p < 0.01$) or research (13% vs 23%; $p < 0.01$) domains. There were no significant differences between groups in likelihood of being scored in the highest tiers for community service, leadership, reference letters, or professionalism.

Relative Impact of Screening Rubric on Interview Selection

As shown in Table 4, relative to Traditional Review, Holistic Review significantly increased the odds that a URM applicant would be selected for interview (0.35 vs 0.84, $p < 0.05$), while no statistically significant change was noted under Traditional

Table 3 Summary statistics of residency applicants by representation in medicine (RM) status ($n = 574$)

		Underrepresented in medicine (URM) ($N = 154$)	Non-underrepresented in medicine (non-URM), ($N = 420$)	<i>p</i> value
Traditional				
STEP 1	< 220, <i>n</i> (%)	71 (46%)	111 (26%)	**
	220–240, <i>n</i> (%)	64 (42%)	168 (40%)	
	240–260, <i>n</i> (%)	19 (12%)	123 (29%)	
	> 260, <i>n</i> (%)	0 (0%)	18 (4%)	
AOA, <i>n</i> (%)	3 (2%)	40 (9.5%)	**	
GHHS, <i>n</i> (%)	24 (16%)	47 (11%)		
PhD, <i>n</i> (%)	6 (4%)	31 (7%)		
Masters, <i>n</i> (%)	17 (11%)	46 (11%)		
School Rank ^a , <i>n</i> (%)	27 (18%)	80 (19%)		
Posters ^a , <i>n</i> (%)	18 (12%)	66 (16%)		
Peer Reviewed Articles ^a , <i>n</i> (%)	13 (8%)	57 (14%)		
Modified				
Keywords ^a , <i>n</i> (%)		42 (27%)	65 (15%)	**
Holistic				
Community ^a , <i>n</i> (%)		35 (23%)	68 (16%)	
Leadership ^a , <i>n</i> (%)		28 (18%)	74 (18%)	
Clinical ^a , <i>n</i> (%)		13 (8%)	88 (21%)	**
Research ^a , <i>n</i> (%)		20 (13%)	97 (23%)	**
Reference Letters ^a , <i>n</i> (%)		45 (29%)	150 (36%)	
Professionalism ^a , <i>n</i> (%)		29 (19%)	75 (18%)	
Resilience ^b , <i>n</i> (%)		57 (37%)	86 (20%)	**
Distance Travelled ^b , <i>n</i> (%)		47 (31%)	52 (12%)	**

^a Presented as the proportion of applications scoring within the highest (categorical) tier as defined by corresponding rubric criteria

^b Binary (Yes/No) outcomes, and defined by HR rubric as (i) resilience: achievement in enduring adversity at an individual level, (ii) distance travelled: trajectory relative to population-level barriers reflecting marginalization at a structural level

* $p < 0.05$

** $p < 0.01$

Modified (OR 0.54). Under the Holistic Review rubric, high STEP1 scores had a less pronounced impact on the odds of receiving an interview relative to Traditional Review (for scores > 260, OR 2.03 vs 24, respectively, $p < 0.01$), noting that only 18 applicants met criteria for highest tier of STEP1 scores. Membership in GHHS increased the odds for interview selection under Holistic Review relative to Traditional Review (OR 6.5 vs 1.59, $p < 0.001$), while applicants with a PhD or scoring in the highest tier for posters and peer-reviewed publications had significantly decreased odds of interview selection in Holistic Review. Use of the Traditional Modified rubric, compared with Traditional Review, did not significantly change the odds of selection for interview for applicants scoring in the highest tier of any Traditional Review domain. Applicants ranked in the highest tiers of the Holistic Review domains of Community Service, Leadership, Research, Reference Letters, and Professionalism or screening positive for Resilience had a greater odds ratio of selection for interview under Holistic Review than Traditional Review.

As shown in Fig. 1, relative to Traditional Review, the predicted probability of interview selection for URM applicants doubled via the Holistic Review approach (0.08 vs 0.16), with little change for nURM applicants across rubrics, (0.21 vs 0.18, Traditional Review and Holistic Review, respectively).

Discussion

This manuscript describes the development and implementation of an equity-minded and mission-driven Holistic Review process that critically examines and reconstructs the standards defining residency selection processes. In this single-site analysis, we note the efficacy of a Holistic Review approach in significantly increasing the odds of URM interview selection. Given the baseline distribution of residency applicants by RM status, the increased odds of interview selection for URM applicants comes without meaningful change in the predicted probability of interview selection for nURM applicants.

Table 4 Interaction effect of traditional rubric relative to traditional modified and holistic review including odds ratios (95% confidence interval) of interview selection

	Traditional Review (TR)	Traditional Modified (TM) ^a	Holistic Review (HR)	Interaction TR:HR
Under-Represented in Medicine (URM)	0.35 (0.19–0.65)	0.54 (0.32–0.94)	0.84 (0.51–1.38)	*
Traditional				
STEP 1	<220	–	–	**
	220–240	3.65 (1.7–7.8)	2.47 (1.27–4.79)	
	240–260	8.63 (4.04–18.4)	5.46 (2.79–10.66)	
	>260	24 (7.6–75)	10.4 (3.5–31)	
AOA	4.90 (2.57–9.34)	3.53 (1.83–6.79)	2.49 (1.26–4.92)	
GHHS	1.59 (0.88–2.88)	2.25 (1.28–3.96)	6.55 (3.84–11)	**
PhD	266 (35.9–1973)	80.9 (24.2–270)	0.4 (0.12–1.32)	**
Masters	2.09 (1.15–3.80)	2.51 (1.40–4.48)	3.24 (1.84–5.72)	
School Rank ^b	5.77 (3.58–9.29)	6.49 (4.03–10.5)	3.00 (1.86–4.87)	
Peer Reviewed Articles ^b	20.7 (11.5–37.3)	22.7 (12.5–41.1)	1.63 (0.9–2.94)	**
Posters ^b	7.45 (4.48–12.4)	7.44 (4.48–12.4)	2.90 (1.73–4.87)	*
Modified				
Keywords ^b	0.80 (0.45–1.43)	2.21 (1.35–3.62)**	3.0 (1.86–4.87)	**
Holistic				
Community ^b	1.37 (0.81–2.33)	2.07 (1.25–3.42)	19.8 (11.7–33.5)	**
Leadership ^b	1.49 (0.88–2.52)	1.84 (1.11–3.06)	17.7 (10.5–29.7)	**
Clinical ^b	5.09 (3.15–8.25)	3.54 (2.17–5.75)	4.79 (2.96–7.76)	
Research ^b	14.12 (8.58–23.2)	10.97 (6.75–17.8)	3.85 (2.41–6.15)	**
Reference Letters ^b	4.18 (2.67–6.57)	3.77 (2.41–5.89)	18.75 (10.4–33.8)	**
Professionalism ^b	1.78 (1.07–2.97)	2.46 (1.51–4.04)	10.6 (6.48–17.3)	**
Resilience ^c	0.82 (0.49–1.38)	1.00 (0.61–1.65)	2.53 (1.61–3.99)	**
Distance Traveled ^c	0.82 (0.45–1.49)	1.07 (0.61–1.87)	1.68 (1.0–2.83)	

^aIncluding also the interaction of TR:TM with significant outcomes noted at level of variable

^bPresented as the proportion of applications scoring within the highest (categorical) tier as defined by corresponding rubric criteria

^cBinary (Yes/No) outcomes, and defined by HR rubric as i) resilience: achievement in enduring adversity at an individual-level, ii) distance traveled: trajectory relative to population-level barriers reflecting marginalization at a structural-level

* $p < 0.05$, ** $p < 0.01$

The reduction in disparity demonstrated in this study, without direct consideration of applicant race/ethnicity, reflects the extent to which bias-producing metrics have themselves perpetuated racialized notions of excellence [9, 24]. For example, the distribution of STEP1 scores by RM status and the highly significant variation in the pattern of odds for interview selection by STEP1 score across rubrics underscore the role of USMLE as a potent barrier to equitable representation.

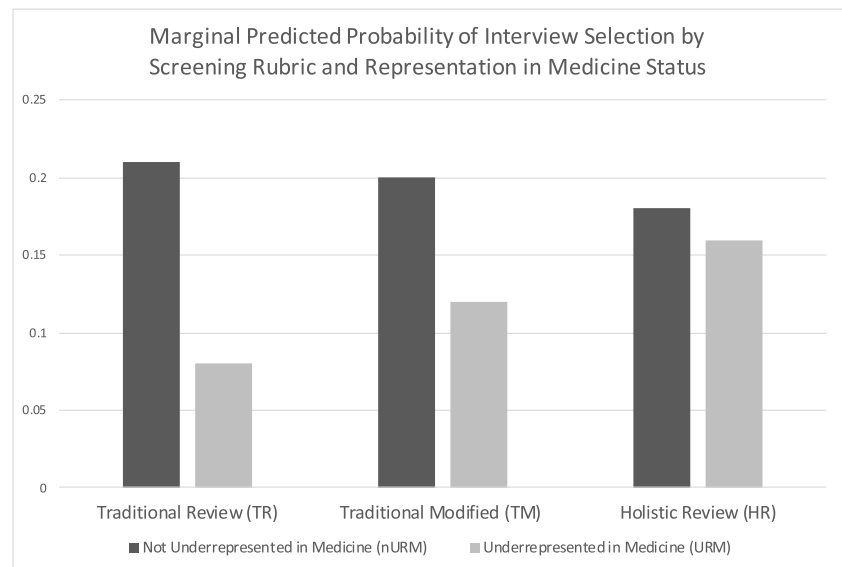
This study replicated known trends in honors society representation, including the exclusion of URM trainees from AOA selection and more equitable representation in GHHS [11]. While Holistic Review scoring did not prohibit consideration of AOA, its decreasing association to interview selection relative to Traditional Review (while not significant) likely reflects the repositioning of AOA election as but one example of achievement as opposed to the ultimate proxy for clinical excellence. Similarly, the defining of Holistic Review domains to allow multiple forms of justifying evidence likely

explains the equity-enhancing impact of community service, leadership, and professionalism in conjunction with GHHS.

In combination with a divestment from traditional screening elements, study outcomes are additionally explained by associations between URM status and lived experience. Rather than giving additive point contribution in recognition of resilience, the magnifying effect defined in this approach ($\geq 110\%$ of the preliminary score) gives due consideration to the experience of applicants who face structural barriers in a pervasive way, as opposed to overcoming discrete challenges without lasting repercussion. While distance travelled was also more frequently observed in the life of URM applicants and while its relative contribution to interview selection increased in Holistic Review relative to Traditional Review, its ultimate impact fell short of significance.

The structure of the Traditional Modified rubric, in which designated activities or experiences were given a minor additive point value in an otherwise largely “traditional” approach, likely

Fig. 1 Contrasts of the marginal predicted probabilities of interview selection (Y-axis) for applicants underrepresented in medicine (URM) relative to applicants who are not underrepresented in medicine (nURM) according to each screening and selection rubric examined (X-axis)—Traditional Review (TR), Traditional Modified (TM), and Holistic Review (HR)



represents a prevalent compromise to Holistic Review in modern GME. This compromise is presumably justified by time constraints—residency applications in psychiatry and across most specialties have increased substantially over the last decade [25, 26], and many programs nonetheless rely on the same small groups of administrators and faculty to carry out burdensome and time-sensitive screening and selection procedures. While not formally assessed, in this study, Holistic Review screening typically took ~10–15 min, while Traditional Modified required only ~2–3 min per applicant, and the Traditional Review processes was entirely automated. In this study, despite the fact that URM applicants were more likely to describe meaningful participation in endeavors aligning with program priorities, the additional point contribution defined by Traditional Modified was not sufficient to result in significant change in the odds of interview selection for URM applicants.

Several study outcomes require consideration of baseline sample characteristics. In the case of interview selection by RM status, the minimal change in predicted probability observed for nURM applicants is explained by the notably larger group of nURM applicants to GME. Similarly, while the magnitude of change in interview selection associated with PhD was unanticipated (from OR = 266 to OR = 0.4 in Traditional Review and Holistic Review, respectively), the trend can be understood when considering the relatively large point contribution defined by the Traditional Review rubric relative to the number of applicants with a PhD (~6% of total applicant pool).

In contemporary GME, diversity initiatives characterized by expedited and piecemeal interventions to “improve the numbers” by mitigating downstream bias are commonplace (e.g. unconscious/implicit bias trainings for individual reviewers or strategic advocacy for individual URM applicants) [27]. Despite their prevalence, these efforts have failed to shift population-level representation in medical training even

despite corresponding demographic shifts in the broader population [28]. By limiting their scope to downstream outcomes, these interventions leave intact the frameworks and selection processes that overlook the circumstances, accomplishments, and value-added of applicants from marginalized groups. Despite purported intention to seek out and support diversity, by continuing to rely on processes devoid of social and structural contextualization, academic medicine perpetuates institutional racism and thereby *manufactures* the continued underrepresentation of marginalized groups [29–32].

The method of evaluation described in this study moves upstream and represents a shift in focus away from quota-like recruitment of individual URM applicants (“how many were selected?”) and toward a realignment of values that promotes holistic excellence across groups by asking “under what conditions should applicants be selected?”

Study limitations include its small sample size, location at one study institution, lack of demonstrated inter-rater reliability for holistic review scoring, and the inability to account for the potentially compounding impacts of race- and gender-based bias in the applicant review process due to insufficient power to create subgroup analysis. The study sample was limited to US allopathic graduates, and future exploration may shed additional light on the contribution of Holistic Review to the consideration of osteopathic graduates as well as applicants trained internationally. Further, while metrics with extensive literature demonstrating embedded racial bias were excluded in the Holistic Review tool (e.g. STEP1), emerging literature is implicating some of the materials included, such as clinical grades [33] and MPSE letters [10] as sources of race- and gender-based disparity, and future iterations of the Holistic Review tool therefore must reflect this dynamic evidence base.

To the best of our knowledge, this manuscript represents the first comparative analysis of a systematic holistic approach to

applicant screening relative to non-holistic or traditional methods in addressing racial disparity in GME applicant selection. By identifying a representative cadre of trainees who embody diverse forms of excellence using a single, standardized procedure for all applicants, the study results affirm the efficacy of holistic tools to build a medical workforce with an increased capacity to promote equity within a residency training program [34].

In the context of new ACGME requirements to “engage in practices that focus on mission-driven, ongoing, systematic recruitment and retention of a diverse and inclusive workforce” [35], we share this intervention in order to advance existent knowledge regarding the viability and effectiveness of applied holistic review in residency recruitment [12, 36, 37]. In their application, we additionally hope that tools for holistic recruitment be accompanied by meaningful interventions to address the toxic climate of daily experiences with racism that characterize the lived experience of many GME trainees [38, 39]. Beyond effective recruitment, meaningful interventions to foster the retention and promotion of URM trainees and faculty will be essential to eradicate the inequity that remains ever-present in medical training and clinical care [40, 41].

Acknowledgments Thanks go to Evelyn Carter, PhD, for her contribution to conception of this Holistic Review rubric. Additional thanks to Dr. Theodore Hall and Lingqi Tan, PhD, for their insightful feedback. We additionally thank members of the “2018-19 Recruitment Task Force” including Drs. Erin Samplin, T.C. Scotton, Maria Gabriela Aguilera Nunez, and Evelyn Nelson. Finally, we thank each of the many residents who volunteered their time and energy to participate in Holistic Review.

Funding N.B. is supported by the APA SAMHSA Minority Fellowship, Grant 1H79SM080388-01. E.B. is supported by funding from the Los Angeles County Department of Probation and from the National Institute on Drug Abuse of the National Institutes of Health under the AACAP NIDA K12 program, Grant # K12DA000357, the California Community Foundation, and the UCLA Pritzker Center for Strengthening Children and Families. Dr. Goodsmith was supported by the VA Office of Academic Affiliations through the UCLA National Clinician Scholars Program. The contents do not represent the views of the U.S. Department of Veterans Affairs, the United States Government, or affiliated institutions.

Compliance with Ethical Standards This study qualified for exemption by the Institutional Review Boards at UCLA Medical Center.

Disclosures On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

- Talamantes E, Henderson M, Fancher T, Mullan F. Closing the gap - making medical school admissions more equitable. *N Engl J Med*. 2019;380:803–5.
- Deville C, Hwang W-T, Burgos R, Chapman C, Both S, Thomas CR. Diversity in graduate medical education in the United States by race, ethnicity, and sex, 2012. *JAMA Intern Med*. 2015;175:1706–9.
- An American Crisis. The growing absence of black men in medicine and science. Washington, DC: The National Academies Press; 2018.
- Komaromy M, Grumbach K, Drake M, Vranizan K, Lurie N, Keane D, et al. The role of Black and Hispanic physicians in providing health care for underserved populations. *N Engl J Med*. 1996;334:1305–10.
- Walker KO, Moreno G, Grumbach K. The association among specialty, race, ethnicity, and practice location among California physicians in diverse specialties. *J Natl Med Assoc*. Elsevier Masson SAS. 2012;104:46–52. [https://doi.org/10.1016/S0027-9684\(15\)30126-7](https://doi.org/10.1016/S0027-9684(15)30126-7).
- Hsu P, Balderas-Medina Anaya Y, Hayes Bautista D. 5 Centuries to reach parity: an analysis of how long it will take to address California’s Latino physician shortage. UCLA Lat Policy Polit Initiat. 2018; Available from: <https://latino.ucla.edu/wp-content/uploads/2018/10/5-Centuries-to-Reach-Parity-UCLA-LPPI.pdf>. Accessed 8 December 2019.
- Nivet MA. Diversity 3.0: a necessary systems upgrade. *Acad Med*. 2011;86:1487–9.
- Edmond MB, Deschenes JL, Eckler M, Wenzel RP. Racial bias in using USMLE Step 1 scores to grant internal medicine residency interviews. *Acad Med*. 2001;76:1253–6.
- Boatright D, Ross D, O’Connor P, Moore E, Nunez-Smith M. Racial disparities in medical student membership in the alpha omega alpha honor society. *JAMA Intern Med*. 2017;177:659–65.
- Ross DA, Boatright D, Nunez-Smith M, Jordan A, Chekroud A, Moore EZ. Differences in words used to describe racial and gender groups in Medical Student Performance Evaluations. *PLoS One*. 2017;12:e0181659.
- Wijesekera T, Kim M, Moore EZ, Sorenson O, Ross DA. All other things being equal: exploring racial and gender disparities in medical school honor society induction. *Acad Med*. 2019;94:562–9.
- Aibana O, Swails JL, Flores RJ, Love LT. Bridging the gap: holistic review to increase diversity in graduate medical education. *Acad Med*. 2019;94:1137–41.
- Chen DR, Priest KC, Batten JN, Frago LE, Reinfeld BI, Laitman BM. Student perspectives on the “step 1 climate” in preclinical medical education. *Acad Med*. 2019;94:302–4.
- Freeman AD, Freeman AD. Legitimizing racial discrimination through antidiscrimination law: a critical review of supreme court doctrine. *Minn Law Rev*. 1978;62.
- Boyd RW. The case for desegregation. *Lancet Elsevier Ltd*. 2019;393:2484–5.
- Coleman AL, Lipper KE, Taylor TE, Palmer SR. Roadmap to diversity and educational excellence: key legal and educational policy foundations for medical schools. A Report for the Association of American Medical Colleges. Washington, DC; 2014. Available from: https://store.aamc.org/downloadable/download/sample/sample_id/192
- Dort JM, Trickey AW, Kallies KJ, Joshi ART, Sidwell RA, Jarman BT. Applicant characteristics associated with selection for ranking at independent surgery residency programs. *J Surg Educ Elsevier*. 2015;72:e123–9.
- UCSF. URM Definition [Internet]. 2019. p. 2019. Available from: <https://diversity.ucsf.edu/URM-definition>. Accessed 8 December 2019.
- Cuddy MM, Swanson DB, Clauser BE. A multilevel analysis of examinee gender and USMLE step 1 performance. *Acad Med*. 2008;83:58–62.
- Veloski JJ, Callahan CA, Xu G, Hojat M, Nash DB. Prediction of students’ performances on licensing examinations using age, race, sex, undergraduate GPAs, and MCAT scores. *Acad Med*. 2000;75: S28–30.
- Girotti JA, Chanatry JA, Clinchot DM, McClure SC, Sein AS, Walker IW, et al. Investigating group differences in examinees’

- preparation for and performance on the new MCAT exam. *Acad Med.* 2019.
22. Moynahan KF. The current use of United States medical licensing examination step 1 scores: holistic admissions and student well-being are in the balance. *Acad Med.* 2018;93:963–5.
 23. Bewick V, Cheek L, Ball J. Statistics review 8: qualitative data - tests of association. *Crit Care.* 2004;8:46–53.
 24. Bonilla-Silva E. Rethinking racism: toward a structural interpretation. *Am Sociol Rev.* 2006;62:465.
 25. Agapoff JR, Tonai C, Eckert DM, Gavero G, Goebert DA. Challenges and perspectives to the rise in general psychiatry residency applications. *Acad Psychiatry.* 2018;42:674–6.
 26. AAMC. 2019 FACTS: Electronic Residency Application Service (ERAS) [Internet]. 2020 [cited 2020 Feb 23]. Available from: <https://www.aamc.org/data-reports/students-residents/interactive-data/2019-facts-electronic-residency-application-service-eras-data>. Accessed 23 February 2020.
 27. Johnson TJ, Ellison AM, Dalembert G, Fowler J, Dhingra M, Shaw K, et al. Implicit bias in pediatric academic medicine. *J Natl Med Assoc.* Elsevier Inc. 2017;109:156–63. <https://doi.org/10.1016/j.jnma.2017.03.003>.
 28. Iglehart JK. Diversity dynamics - challenges to a representative U.S. medical workforce. *N Engl J Med.* 2014;371:1471–4.
 29. Ford CL, Airhihenbuwa CO. The public health critical race methodology: praxis for antiracism research. *Soc Sci Med.* 2010;71:1390–8.
 30. Krieger N, Alegria M, Almeida-Filho N, Barbosa da Silva J, Barreto ML, Beckfield J, et al. Who, and what, causes health inequities? Reflections on emerging debates from an exploratory Latin American/North American workshop. *J Epidemiol Community Health.* 2010;64:747–9.
 31. Sklar DP. Racial violence, academic medicine, and academic medicine. *Acad Med.* 2015;90:1577–80.
 32. Mgbako O. The Unicorn. *JAMA.* 2019;321:149–50. Available from: <http://www.nber.org/papers/w24787>.
 33. Teherani A, Hauer KE, Fernandez A, King TE, Lucey C. How small differences in assessed clinical performance amplify to large differences in grades and awards: a cascade with serious consequences for students underrepresented in medicine. *Acad Med.* 2018;93:1286–92.
 34. Smith DG. Building institutional capacity for diversity and inclusion in academic medicine. *Acad Med.* 2012;87:1511–5.
 35. ACGME. Common program requirements (residency) [Internet]. Acgme. 2019. Available from: <https://www.acgme.org/Portals/0/PFAssets/ProgramRequirements/CPRResidency2019.pdf>. Accessed 23 February 2020.
 36. Bandiera G, Abrahams C, Ruetalo M, Hanson MD, Nickell L, Spadafora S. Identifying and promoting best practices in residency application and selection in a complex academic health network. *Acad Med.* 2015;90:1594–601.
 37. Fernandez A. Further incorporating diversity, equity, and inclusion into medical education research. *Acad Med.* 2019;94:S5–6.
 38. Cyrus KD. Medical education and the minority tax. *JAMA.* 2017;317:1833–4.
 39. Osseo-asare A, Balasuriya L, Huot SJ, Keene D, Berg D, Nunez-smith M, et al. Minority resident physicians' views on the role of race/ethnicity in their training experiences in the workplace. *JAMA Netw Open.* 2018;1:1–11.
 40. Palepu A, Carr PL, Friedman RH, Amos H, Ash AS, Moskowitz MA. Minority faculty and academic rank in medicine. *J Am Med Assoc.* 1998;280:767–71.
 41. Diaz T, Navarro JR, Chen EH. An institutional approach to fostering inclusion and addressing racial bias: implications for diversity in academic medicine. *Teach Learn Med Routledge;* 2019;0:1–7.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.