

Reducing Bias in Interviews with Undergraduate Student Applicants for STEM Research (Work In Progress)

Dr. Jad El Harake, Vanderbilt University

Prof. Jonathan Ehrman, Department of Biomedical Engineering, Vanderbilt University, Nashville, TN, USA

Katrina L. Leaptrot, Vanderbilt University

Reducing Bias in Interviews with Undergraduate Students for STEM Research Involvement (Work In Progress)

Abstract

Undergraduate research can play a large role in diversifying STEM fields, giving many students from underrepresented groups the opportunity to explore research careers for the first time in their educational journey. However, research programs and opportunities may be biased against underrepresented students; this can occur due to implicit biases amongst interviewers, or due to the format of the selection process and the implemented scoring criteria or lack thereof. For example, unstructured interviews have been shown to favor white applicants over applicants of color, while structured interviews with predetermined questions are more equitable. Several studies have investigated these biases, typically within the context of medical and graduate school applications. Drawing on the recommendations of these studies, as well as the authors' prior experiences interviewing applicants, we have developed a framework which seeks to minimize bias for interviewing and admitting students to STEM undergraduate research opportunities. The proposed framework is intended to assess applicants holistically, evaluating academic potential as well as an applicant's ability to effectively and creatively collaborate with others. A written application is used as a screening tool, followed by an in-person interview. The interview process consists of a multiple mini-interview portion (MMI), a group activity portion, and a program overview. In the MMI segment, applicants are given 3 minutes to answer one question before rotating to the next station, with two interviewers assigned to each station. The group activity allows for applicant evaluation in a more relaxed setting as they collaborate with their peers. Finally, the program overview ensures all applicants have a baseline understanding of the program goals and values. This framework exposes applicants to several different interviewers, mitigating individual biases, and additionally presents applicants with various opportunities to showcase their strengths both individually and in group settings. This process has also been designed to be time- and resource-efficient, which is crucial due to the typical mismatch between the number of applicants and number of reviewers available. This work-in-progress describes the process and the rationale behind its design, and reports on preliminary metrics of fairness and efficiency such as the demographic distribution of accepted applicants and the required time to evaluate an applicant cohort.

Introduction

Studies consistently highlight the benefits of diversity within teams, such as improved creativity, problem-solving, and overall research outcomes [1], [2]. In contrast, scientific research can stagnate without the synergy of diverse perspectives and insights. Academia is particularly important to analyze, as disparities within academic settings propagate to affect the future scientific workforce [3]. Despite efforts to create more inclusive educational and work environments, implicit biases in recruitment and selection processes remain a significant barrier [4], [5], [6]. Reviewer biases can influence applicant evaluations in favor of candidates with certain demographic characteristics, background, or socioeconomic status; for example reviewers may subconsciously favor applicants who share similar traits with them, such as gender, race, or personalities [7]. Furthermore, it has been shown that certain selection processes which may seem benign can produce biased selection outcomes; examples include unstructured interviews [8] or overly emphasizing “cultural fit” as a selection criteria [9], [10]. These phenomena can reinforce disparities within academic fields by creating higher barriers for underrepresented groups and individuals with nontraditional backgrounds or merits.

A structured and evidence-based framework for the evaluation and selection of researchers is thus needed to address implicit biases. Several studies have investigated best practices for unbiased selection within the context of medical school admission [11], [12], [13], [14], [15], and to a lesser extent graduate school admission [4]. We hypothesize that the findings and recommendations from these studies are generalizable and can guide other academic selection processes. In this study we implement recommendations from the literature in the context of undergraduate applicants for STEM research funding and mentorship through a year-round undergraduate research program at Vanderbilt University, the SyBBURE Searle Undergraduate Research Program.

This specific research program seeks to foster a culture of creativity, innovation, and teamwork, set within an engaging and flexible setting [16], [17], [18], [19]. Drawing students from majors across the university’s colleges, the program emphasizes the value of diverse perspectives, promoting rich interdisciplinary collaboration. Students are offered a mix of individual research opportunities, team-driven design projects, technical training related to these projects, and various activities that promote community engagement and all-around development. The program prides itself on its holistic approach to learning, preparing students not only for the laboratory but for a balanced life and career after graduation. Fairness and diversity are thus necessary guiding values for the program to cultivate a culture of collaborative innovation and personal and professional growth.

This work-in-progress aims to disseminate the framework that has been developed such that it may be replicated at similar programs, and to solicit feedback from others in order to continually improve the fairness of the selection process. We have endeavored to develop an

equitable evaluation process to assess values and characteristics of interest, such as creativity, motivation, curiosity, openness to collaboration. The following sections are organized as follows; in the methods, the application framework will be described in detail along with the rationale behind specific design choices. Preliminary results will then be presented, primarily concerning the demographic distributions of accepted applicants over time. Finally, the discussion will elaborate on the authors' experiences implementing this process, as well as the limitations and future directions of this work.

Methods

A. General Considerations

To ensure an equitable and effective selection process, the following principles were prioritized throughout the application process:

- (1) **Diversity:** In particular, diversity not only of background but also of applicant personalities, strengths, and interests are highly valued in the program. On a related note, applicants' potential is valued over prior achievements in order to minimize the effects of socioeconomic background and their pre-existing network. This approach additionally ensures that individuals with unconventional prior experience are not unfairly disadvantaged compared to applicants with typical research experiences through coursework or internships.
- (2) **Fairness:** Reviewer biases are mitigated by redacting identifying information when possible, pooling evaluations from multiple reviewers for each applicant and ensuring a low standard deviation between scores.
- (3) **Efficiency:** The process is streamlined to accommodate the high volume of applications within the constraints of limited reviewers and time, enabling thorough evaluation without compromising assessment accuracy.
- (4) **Structure:** Establishing clear yet adaptable evaluation criteria, which are summarized in Table 1, ensuring that sufficient structure is provided to guide transparent decision-making while allowing enough flexibility to accommodate applicants' unique strengths and experiences.

In-person interviews are expected to yield the most accurate evaluations, but are logistically more challenging to organize particularly when considering the high applicant-to-reviewer ratio. We thus utilized a written application screening followed by an in-person interview, improving efficiency while providing all applicants with the opportunity to showcase their strengths in a rigorous and unbiased evaluation. Within the interview process, moderated group-based activities provide an opportunity to assess how candidates navigate interdisciplinary challenges and communicate with peers in different fields, while individual interviews can reveal an applicants' interests and critical thinking skills with more nuance and in greater depth than what can be gleaned from a group setting. In the case of the individual interviews, it has been shown that structured multiple mini-interviews (MMIs) improve evaluation reliability and

fairness [13], [20]. The inclusion of all these evaluation metrics and environments is expected to achieve a more holistic assessment.

B. Written application

Written questions, when designed thoughtfully, allow candidates to demonstrate their ability to reflect and grow from experience, articulate complex ideas, and present well-reasoned arguments, all of which are important skills in research. The inclusion of lighthearted creative writing prompts can additionally highlight applicants' curiosity, resourcefulness, and general approach to problem solving. However, written assessments alone may not capture the full range of a candidate's abilities, particularly in dynamic, collaborative environments. The written component is therefore more effectively used as a screening tool, while in-person interviews can be used to evaluate candidates' interpersonal skills, adaptability, and problem-solving approaches.

The written application consisted of a broad set of questions designed to be inclusive of diverse experiences and to capture various indicators of aptitude and potential. Care was taken to avoid reliance on metrics that could introduce bias, such as grade point average (GPA) or prior experiences, as these may disproportionately favor applicants from more advantaged backgrounds. Instead, the questions were crafted to assess critical thinking, creativity, motivation, and alignment with program goals. These questions were selected to evaluate intrinsic motivation, problem-solving skills, and resilience while remaining accessible to a wide range of applicants.

C. Interview Process

Applicants shortlisted from the written application stage were invited to participate in a three-part interview process as depicted in Figure 1. The process consisted of MMIs, a collaborative activity, and a program overview segment. In prior iterations of this framework, unstructured ranking was attempted in which reviewers were asked to rank applicants based on overall impression. However, this led to a high rate of discrepancies between reviewers. In addition, unstructured evaluation of applicants holistically by each reviewer is expected to be less effective at mitigating implicit bias compared to focusing evaluation on a specific characteristic. Thus, each set of 2-3 reviewers were assigned just one of the 9 characteristics listed in Table 1 to evaluate for all applicants.

C1: Multiple Mini-interviews (MMIs)

Applicants rotate between several interview stations; at each station, two reviewers were assigned to evaluate the criteria outlined in Table 1. The interviewers at each station select one question to be posed to all applicants with the same phrasing each time, and applicants have 3

minutes to answer before rotating to the next MMI station. Once all applicants have had the chance to interview at each station, they move on to the next evaluation activity.

Activity	Criteria	Description	Example Interview Questions and Prompts
MMIs	Character	Sense of self, individuality, internal motivation	<ul style="list-style-type: none"> • Describe a time when you had to make a tough decision. • How do you stay motivated during challenging or monotonous tasks? • What is an accomplishment you are proud of?
	Communication	Ability to interact, both verbally and nonverbally	<ul style="list-style-type: none"> • Tell us about yourself; what is your story?
	Curiosity	Desire to learn about new things, breadth and depth of interests	<ul style="list-style-type: none"> • Describe a time you questioned someone's understanding of an idea or concept. • What is something new you've tried to learn recently, and why did you try it?
	Commitment	Having time to fully engage with program	<ul style="list-style-type: none"> • What does a typical day look like for you? • What are your top priorities and how do you manage your time to balance them?
	Coherence	Ability to turn vague ideas and concepts into executable plans with little guidance	<ul style="list-style-type: none"> • Pretend you are in charge of planning a birthday party for an elephant. What do you do?
Group Activity	Creativity	Variety and originality of meaningful ideas, and the depth of description of those ideas	<ul style="list-style-type: none"> • Pretend your group is responsible for building an ecosystem for a mythical creature. Work together to develop a plan, then use craft materials to construct a model of your enclosure. Explain your strategies and reasonings as you work.
	Collaboration	Ability to listen to and synergize with others' ideas, knowing when to compromise	<ul style="list-style-type: none"> • Work together to envision a better world and devise a plan of action to achieve this.
	Courtesy	Respectful treatment of others and shared spaces	
Program Overview	Contribution	Potential impact on the program and on the other students	<ul style="list-style-type: none"> • What would you bring to the program? • What would you want to make, discover, invent, or learn if you were accepted into the program?

Table 1. List of criteria to be evaluated for each applicant within each of the three interview segments, as well as examples of questions that have been used at each station.

C2: Mediated Group Activity

To assess creativity, collaboration, and courtesy towards peers. Applicants participated in a facilitated group activity designed to simulate a collaborative task. The task is designed to be complex in the number of approaches that could be applied, without requiring specific prior knowledge in order to ensure accessibility to applicants from different academic disciplines. In this activity, applicants are evaluated for their creativity, courtesy, and collaboration.

C3: Program Overview

In the program overview, interviewers and current program participants informally talk to applicants about the main program components from a student's perspective. Interviewers introduce themselves and then describe aspects of the program, including the independent research project, team-based design project, and the community element of the program. Interviewers also ask questions to gain insights about the applicants' interests and potential contributions to the program. This provides applicants with the opportunity to ask any clarifying questions, as well as align their understanding of the program mission and values based on the description provided by current program participants.

D. Scoring

The written applications were each scored independently by two reviewers using predefined rubrics summarized in Table 1. If the standard deviation between reviewers' scores exceeded a set threshold, a third reviewer was assigned to mediate discrepancies. Demographic information, such as gender and ethnicity, was hidden during review and scoring. This information was reviewed after selection to ensure that the demographic distribution of the selected students was representative of the applicant pool and university's larger student body.

After the written application scores are finalized, the number of available spots in the cohort is determined based on the program resources. Approximately three times as many applicants are invited to interview, aiming for a ~33% acceptance rate following the interviews. The in-person interviews are assessed according to the criteria shown in Table 1; each applicant is ranked by at least two reviewers for each of the listed criteria, and these rankings are combined to yield the final applicant score. The reviewers then discuss the rankings, paying particular attention to any cases with large differences between the individually assigned rankings. This discussion aims to better understand and resolve any large disparities in applicant evaluations,

improving the robustness of the process to inter-reviewer discrepancies. Once rankings are finalized, the highest-scoring applicants are offered a spot in the program. Applicants who scored well but did not meet the cutoff based on available program spots are offered a position on the waitlist. Decisions are typically ready to be shared with the applicants approximately one hour following the conclusion of the interviews.

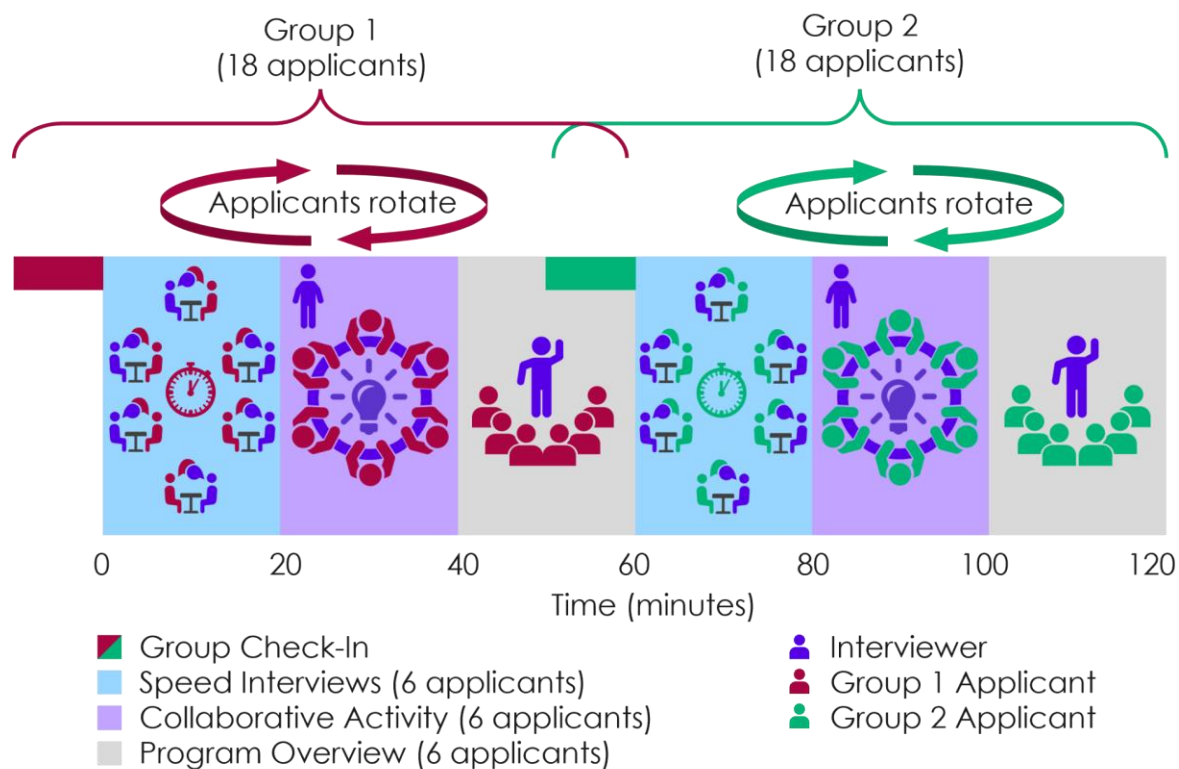


Figure 1. Schematic summarizing the proposed in-person interview process. 36 students were divided into two groups of 18 to be interviewed at time 0-60 min (Group 1, red) or time 60-120 min (Group 2, green). Each group was further divided into subgroups of 6 during check-in, and these subgroups rotated between the MMIs (blue), collaborative activity (purple), and program overview (grey) segments. Each segment lasted 20 minutes, totalling one hour for the interview process for each group of 18 applicants.

Results

The framework above was most recently applied to evaluate a pool of 134 applicants. After an estimated 8 hours of asynchronous application scoring by 9 reviewers each, 36

applicants were invited to the in-person interview. The 36 applicants were divided into two groups of 18 and both sets of interviews proceeded according to the schedule shown in Figure 1 with no major delays, taking two hours with 9 faculty and staff reviewers and 14 student reviewers. Reviewer scores were pooled immediately following the conclusion of the interviews, and one hour of discussion was necessary to finalize the lists of accepted and waitlisted applicants. This process in total takes 99 faculty and staff hours and 42 students hours. If each of the 134 applicants had been given a 30 minute interview with the faculty and staff, it would have taken 603 person-hours, highlighting the efficiency of prescreening and a group interview.

In preliminary evaluation of fairness, we considered the racial and ethnic distribution of program participants over time since the implementation of the interview framework, as well as the distribution of gender identity. The program began in 2007, and the interview framework was utilized beginning in 2016. Figure 2 shows an improvement over time in the consistency of program racial diversity since the framework has been implemented. In particular, Hispanic and Black students who were severely underrepresented at the outset of the program now make up a significant percentage of the program. Similarly, Figure 3 shows a more balanced distribution of gender identity which has persisted since the interview framework was adopted.

Although these preliminary findings are encouraging, there are several limitations to this analysis which must be considered. Firstly, it is challenging to prove a causal relationship between the implemented process and the observed improvements in diversity. This is in large part due to the large number of other factors which may have also affected the demographic distribution of the program; the overall diversity of the student body at the university has increased over time, which likely contributed to the increased representation of marginalized groups within the program. In addition, as was previously mentioned in the introduction, there are many more facets of diversity than racial and gender distribution. In the future, analysis of socioeconomic status, field of study, English fluency, neurodivergence, and personality may also be considered both when evaluating fairness of outcomes and when designing improvements to the interview framework. For example, interviewees for whom English is a second language may benefit from additional time accommodations, particularly during the MMI segment. Furthermore, a formal analysis of the specific implicit biases which may affect selection is necessary in order to understand how effective the described methods are at reducing bias, and to identify any potential shortcomings of this interview protocol. Finally, a statistical comparison between the distributions of accepted applicants before and after the implementation of this framework, normalized against the distribution of all applicants and of the larger university student body, could more definitively reveal the impact of the proposed framework.

Despite these limitations, the presented framework is believed to be an efficient method of selecting applicants based on program values, with several strengths backed by evidence from the literature. This framework relies on structured interviews and well-defined evaluation criteria

which has been defined such that applicants' potential is emphasized over prior experience, mitigating potential unfair advantages due to applicants' prior access to opportunities. In addition, multiple reviewers are assigned to score each component of a students' application, and any large deviations in scoring are addressed by additional reviewer assignments. Finally, applicants are exposed to a wide range of written questions and in-person interview environments, ensuring a holistic evaluation of applicants.

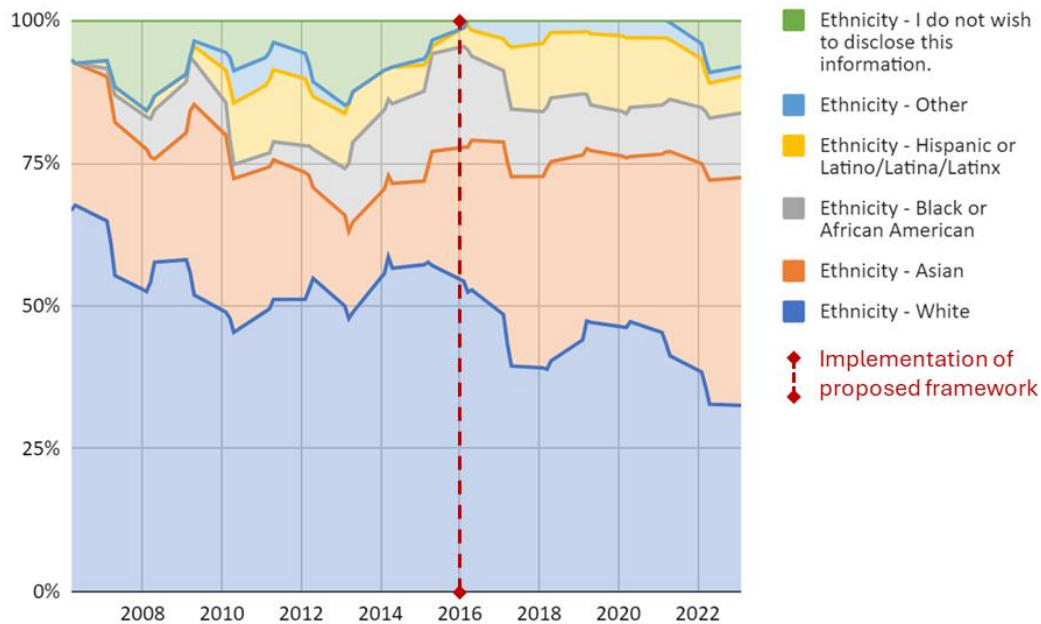


Figure 2. Distribution of program participants' ethnicities over time, beginning with the establishment of the program in 2007. The dashed red line indicates the implementation of the proposed framework.

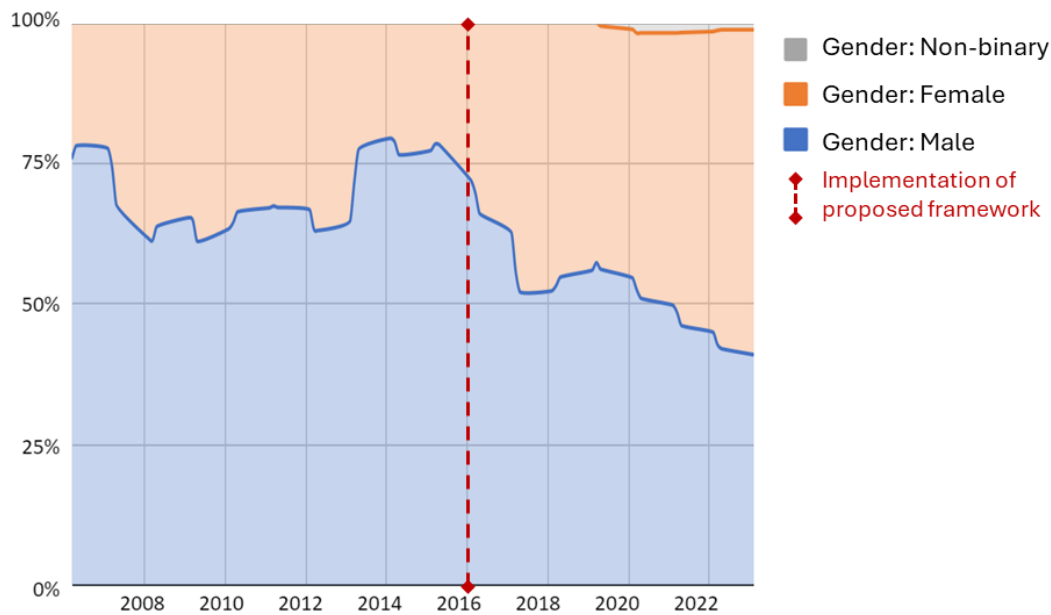


Figure 3. Distribution of program participants' gender identities over time, beginning with the establishment of the program in 2007. The dashed red line indicates the implementation of the proposed framework.

Conclusion

In this work-in-progress manuscript, we have presented an efficient research program application and evaluation framework which seeks to minimize biases during selection. A written application is followed by a structured in-person interview, which consists of three different activities used to assess various aspects of each applicant. The preliminary analysis of program demographic distribution shows promising results supporting the fairness of the proposed framework, but more rigorous investigation is needed to demonstrate causality. Future work will seek to shed more light on the impact of the interview framework on applicant outcomes, and to formally define the specific biases which are mitigated by each component of the proposed process.

References

- [1] D. G. Smith and N. B. Schonfeld, "The Benefits of Diversity what the Research Tells Us," *Campus*, vol. 5, no. 5, pp. 16–23, Nov. 2000, doi: 10.1177/108648220000500505.

- [2] D. van Knippenberg and J. N. Mell, "Past, present, and potential future of team diversity research: From compositional diversity to emergent diversity," *Organ. Behav. Hum. Decis. Process.*, vol. 136, pp. 135–145, Sep. 2016, doi: 10.1016/j.obhdp.2016.05.007.
- [3] P. R. Hernandez, A. Woodcock, M. Estrada, and P. W. Schultz, "Undergraduate Research Experiences Broaden Diversity in the Scientific Workforce," *BioScience*, vol. 68, no. 3, pp. 204–211, Mar. 2018, doi: 10.1093/biosci/bix163.
- [4] S. E. Woo, J. M. LeBreton, M. G. Keith, and L. Tay, "Bias, Fairness, and Validity in Graduate-School Admissions: A Psychometric Perspective," *Perspect. Psychol. Sci.*, Jun. 2022, doi: 10.1177/17456916211055374.
- [5] C. Houser and K. Lemmons, "Implicit bias in letters of recommendation for an undergraduate research internship," *J. Furth. High. Educ.*, vol. 42, no. 5, pp. 585–595, Jul. 2018, doi: 10.1080/0309877X.2017.1301410.
- [6] D. M. Easterly and C. S. Ricard, "Conscious Efforts to End Unconscious Bias: Why Women Leave Academic Research," *J. Res. Adm.*, vol. 42, no. 1, pp. 61–73, 2011.
- [7] K. L. Milkman, M. Akinola, and D. Chugh, "What happens before? A field experiment exploring how pay and representation differentially shape bias on the pathway into organizations," *J. Appl. Psychol.*, vol. 100, no. 6, pp. 1678–1712, Nov. 2015, doi: 10.1037/apl0000022.
- [8] I. Bergelson, C. Tracy, and E. Takacs, "Best Practices for Reducing Bias in the Interview Process," *Curr. Urol. Rep.*, vol. 23, no. 11, pp. 319–325, Nov. 2022, doi: 10.1007/s11934-022-01116-7.
- [9] L. A. Rivera, "Hiring as Cultural Matching: The Case of Elite Professional Service Firms," *Am. Sociol. Rev.*, vol. 77, no. 6, pp. 999–1022, Dec. 2012, doi: 10.1177/0003122412463213.
- [10] H. H. Bye, J. G. Horverak, G. M. Sandal, D. L. Sam, and F. J. van de Vijver, "Cultural fit and ethnic background in the job interview," *Int. J. Cross Cult. Manag.*, vol. 14, no. 1, pp. 7–26, Apr. 2014, doi: 10.1177/1470595813491237.
- [11] J. C. Lin, A. Lokhande, C. E. Margo, and P. B. Greenberg, "Best practices for interviewing applicants for medical school admissions: a systematic review," *Perspect. Med. Educ.*, vol. 11, no. 5, pp. 239–246, Oct. 2022, doi: 10.1007/s40037-022-00726-8.
- [12] S. Ali, M. S. Sadiq Hashmi, M. Umair, M. A. Beg, and N. Huda, "Multiple Mini-Interviews: Current Perspectives on Utility and Limitations," *Adv. Med. Educ. Pract.*, vol. 10, no. null, pp. 1031–1038, Dec. 2019, doi: 10.2147/AMEP.S181332.
- [13] E. L. Rees, A. W. Hawarden, G. Dent, R. Hays, J. Bates, and A. B. Hassell, "Evidence regarding the utility of multiple mini-interview (MMI) for selection to undergraduate health programs: A BEME systematic review: BEME Guide No. 37," *Med. Teach.*, vol. 38, no. 5, pp. 443–455, May 2016, doi: 10.3109/0142159X.2016.1158799.
- [14] K. Kumar, C. Roberts, I. Rothnie, C. Du Fresne, and M. Walton, "Experiences of the multiple mini-interview: a qualitative analysis," *Med. Educ.*, vol. 43, no. 4, pp. 360–367, 2009, doi: 10.1111/j.1365-2923.2009.03291.x.
- [15] J.-F. Lemay, J. M. Lockyer, V. T. Collin, and A. K. W. Brownell, "Assessment of non-cognitive traits through the admissions multiple mini-interview," *Med. Educ.*, vol. 41, no. 6, pp. 573–579, 2007, doi: 10.1111/j.1365-2923.2007.02767.x.
- [16] K. H. Oliver, J. D. Ehrman, and C. C. Marasco, "Vigilante Innovation (VIX): case study on the development of student skills through a team-based design process and

environment,” *Int. J. STEM Educ.*, vol. 6, no. 1, p. 36, Oct. 2019, doi: 10.1186/s40594-019-0190-3.

[17] C. O. Pax, K. Cyr, J. Ehrman, and C. C. Marasco, “Vigilante Innovation,” *Vigilante Innovation*. Accessed: Jan. 15, 2025. [Online]. Available:

<https://vigilanteinnovation.com>

[18] J. D. Ehrman, “Guide To Research,” *Guide To Research*. Accessed: Jan. 15, 2025. [Online]. Available: <https://www.guidetoresearch.com>

[19] J. D. Ehrman and C. C. Marasco, “Guide To Mentoring,” *Guide To Mentoring*. Accessed: Jan. 15, 2025. [Online]. Available: <https://www.guidetomentoring.com>

[20] J.-M. Leduc, R. Rioux, R. Gagnon, C. Bourdy, and A. Dennis, “Impact of sociodemographic characteristics of applicants in multiple mini-interviews,” *Med. Teach.*, vol. 39, no. 3, pp. 285–294, Mar. 2017, doi: 10.1080/0142159X.2017.1270431.