



Evaluating the Whole Applicant: Use of Situational Judgment Testing and Personality Testing to Address Disparities in Resident Selection

Elizabeth B. Takacs¹ · Chad R. Tracy¹

Accepted: 18 July 2022

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract

Purpose of Review Urology program directors are faced with increasing numbers of applications annually, making holistic review of each candidate progressively more difficult. Efforts to streamline evaluation using traditional cognitive metrics have fallen short as these do not predict overall resident performance. Situational judgment tests (SJTs) and personality assessment tools (PATs) have been used in business and industry for decades to evaluate candidates and measure non-cognitive attributes that better predict subsequent performance. The purpose of this review is to describe what these assessments are and the current literature on the use of these metrics in medical education.

Recent Findings SJTs relative to PATs have more original research. Data suggests that SJTs decrease bias, increase diversity, and may be predictive of performance in residency. PATs are also emerging with data to support use with ability to assess fit to program and certain traits identified more consistently among high-performing residents and correlation to performance on ACGME milestones. PATs may be more coachable than SJTs.

Summary SJTs and PATs are emerging as techniques to supplement the current resident application review process. Early evidence supports their use in undergraduate medical education as does some early preliminary results in graduate medical education.

Keywords Situational judgment test · Personality assessment tools · Resident selection · Urology · Recruitment · Diversity

Introduction

Resident selection has recently been challenged in ways that many could not have anticipated. In February 2020, the United States Medical Licensing Examination (USMLE) provided notification that the step 1 certification exam would be converted to a pass/fail format effective January 2022. This was quickly followed by the upheaval of COVID-19, leading to multiple medical schools transitioning to pass/fail clerkship grades and a shift toward virtual interviews. Simultaneously, the question of equity and diversity in medical honors societies led many schools to suspend or remove their involvement in the Alpha Omega Alpha (AOA) honor society [1]. Concurrently, there has been increasing pressure to implement a more holistic and transparent review process

for screening residency applications and recognition of the need to do more to create workforce diversity.

“The match” as it was designed by the National Resident Matching Program (NRMP) is created to be an applicant-centric model. Within this model, the goal is to ensure the applicant has the most favorable outcome, with programs having stable results over time [2]. Applicants and programs have evolved their approach to this process, placing increased value on USMLE step 1 performance, away rotations, and the number of programs to which applicants apply [3, 4]. In urology alone, the average number of applications submitted per applicant has increased from 63 in 2015 to 82 in 2022 [5]. During this time, however, programs have been slower to evolve, relying on traditional and often antiquated methods of applicant screening that are steeped in bias and provide little relevance to subsequent resident performance.

Historically, USMLE step 1 and 2 scores, grades in required clerkships and specialty electives, and letters of recommendation were the driving force for determination of who would be invited for interviews [6, 7]. Of these, many programs place the highest importance on USMLE step 1 score, despite repeated studies showing its lack of

This article is part of the Topical Collection on *Education*

✉ Elizabeth B. Takacs
elizabeth-takacs@uiowa.edu

¹ Department of Urology, University of Iowa, Iowa City, IA 52242-1089, USA

utility in predicting resident performance. Accordingly, for decades step 1 score has been used for counseling students interested in competitive specialties like urology, as a screening tool for application review, and as a criterion in the final rank list [6].

There is increasing evidence that these cognitive metrics are not sufficient in resident selection. Studies have demonstrated that USMLE scores may disadvantage females, underrepresented minorities (URM), and those of socioeconomic disadvantage [8–10]. Though predictive of performance on written board exams, USMLE step 1 scores do not correlate with clinical outcomes, professionalism, or performance on ACGME core competencies [6, 11]. Likewise, there is mixed utility in using class rank, AOA honor society membership, junior year clerkship grades, and the medical student performance evaluation (MSPE) in order to predict a successful resident. Even if effective, these metrics have become progressively less reliable given a move toward pass/fail grading within medical schools, removal of class rank and peer comparisons from the MSPE, and the fact that AOA is not available at all institutions [6].

Because of the lack of objective metrics in determining resident success during the selection process, many programs rely on evaluating fit through the application or during the interview. Optimal fit is difficult to determine and relies on attempting to select candidates who thrive in clinical and academic settings and who contribute to and benefit from those environments equally [12]. The effect of virtual interviews on evaluating an applicant's fit within a program is unknown, though many worry about a decrease in social interaction outside of the interviews during an in-person visit, which is critical in determining fit within a program. The movement toward virtual interviews has emphasized the need to better understand fit through more objective components [13]. Fit itself, however, is highly subjective and may lead to a less diverse resident pool, such that it is recommended that residency programs assess fit in terms of institutional mission, goals, and learning environment [14].

Holistic application review is recommended by the Association of American Medical Colleges (AAMC) to systematically evaluate applicants in an equitable fashion, with an emphasis on equity and diversity in alignment with institutional goals [15]. Unlike traditional application review where much of the focus is on cognitive measures, the holistic approach reviews the candidate with consideration of non-cognitive attributes, reflection on an applicant's experiences, and assessment of the value that applicant may provide to the institution [11, 15]. Studies in undergraduate education and graduate medical education have demonstrated that holistic review leads to an increase in female, URM, and first-generation applicants that are invited for interview [3, 16].

With a need to diversify our workforce, an increasing emphasis on holistic applicant review, ballooning numbers of applications per program, and recognition that our traditional methods of screening are inadequate, the residency application screening process appears to be at an impasse. As such, it may be time for residency programs to consider alternative methods for evaluation of candidates. Two such methods, which have been utilized in the business world but less so in resident selection, are situational judgment testing and personality assessment.

Situational Judgment Test (SJT)

SJTs are designed to measure important non-cognitive characteristics, such as conscientiousness, integrity, accountability, teamwork, stress tolerance, and adaptability. SJTs are program-specific, developed and administered to applicants during the screening process, and attempt to determine key applicant characteristics important to that particular program. These assessments present video-based or written hypothetical but common clinical scenarios likely to be encountered in residency, and candidates are asked to select a response to that scenario. Candidates may be asked how they would most likely respond, how they would most and least likely respond, or to rank the answers from most likely to least likely response. Scoring is pre-determined for each response based on the key qualities the question is intended to address, with scoring determined by subject matter experts. Since the importance of various competencies varies between jobs, each SJT should be individualized for each program [17]. As such, there may be extensive variability to the scenario content, response instructions, response formats, and scoring approach (Table 1) [18]. SJTs are commercially available and may be customizable or developed from scratch by an individual program.

Although various forms of SJT have been around since the 1940s, their use in screening of job applicants did not become widespread until the late 1990s, and their use in health science education did not begin until the early 2000s [19–21]. As such, although used in industry for some time for the initial screening of applicants, the use of SJTs in medical education is still evolving.

Can SJTs Predict Resident Performance?

The value of a screening tool is dependent on its ability to predict subsequent performance. Studies in medical students have shown SJTs can predict performance on ACGME patient care, interpersonal and communication skills, and professionalism competencies as well as grade point average, internship performance, and eventual job performance [22, 23]. Likewise, the SJT-based Computer-based Assessment

Table 1 Summary of SJT aspects that can be customized in testing development [18]

Scenario content	Response instructions	Response format	Scoring approach
Reflects work-, education-, or training-related situations that candidates will likely encounter in target role (e.g., medication error, patient not wanting to discharge)	<ul style="list-style-type: none"> • Knowledge-based: What is the best option? • Behavioral tendency: What would you be most likely to do? 	<ul style="list-style-type: none"> • Ranking actions in order • Rating all response options independently • Multiple choice: choose the 3 best options • Choose best and/or worst option 	Compare candidates' responses to pre-determined scoring key developed by in-depth review of subject matter experts

for Sampling Personal Characteristics (CASPer) exam, which is administered to students applying to Canadian medical schools, has been shown to predict personal/professional characteristics and can provide discriminant validity over traditional cognitive attributes [24]. Based on these studies as well as a decade of research, the AAMC recently established the AAMC PREview professional readiness exam, which is an SJT designed to measure non-cognitive pre-professional competencies [25]. This test will be widely available to medical schools starting in 2022–2023 for the selection of incoming students. The results and follow-up over the subsequent years of medical school and into residency could cause a profound shift in student selection from cognitive to non-cognitive attributes.

Information regarding the use of SJTs in the selection of residents is more limited. One study found that a higher score on the SJT positively correlated with faculty evaluations, medical student evaluations, and overall performance, and that SJT scores provided significant incremental validity over USMLE Step 1 alone with regards to overall performance [26]. Similarly, a multi-institutional study across 21 residency programs found that higher SJT scores were predictive of overall milestones performance and higher scores on multisource professional assessments, with SJTs offering incremental validity over USMLE Step 1 alone [10]. Interestingly, in addition to predicting success on traditional objective measures, SJTs are also capable of predicting overall difficulties in professionalism, such as remediation and probation [27].

What is the Impact of SJTs on the Applicant Screening Process by Programs?

Over the last decade, the number of applicants to urology programs has increased significantly, with the average applicant applying to > 80 programs and any one residency receiving 100 to 150 applications per residency position. For programs, a holistic review of each application may require nearly 100 h for the initial screening (equivalent to 4.5 h/day from application release until batched interviews in the 2022 match cycle). To that end, SJTs are able to screen large numbers of applicants in a more efficient and meaningful manner [28]. Decreasing the time for upfront review of applications would allow residency programs to spend more time focusing on the applications of students who best align with their departmental competencies as pre-determined by the SJT. Similarly, if a particular candidate aligns better with a specific program, candidates could spend less time on a large number of interviews and instead spend more meaningful time evaluating programs to which they are a better fit.

Importantly, studies on SJTs generally revolve around how residents fare during training at an institution based on

the institution's defined core characteristics. Equally interesting would be to determine whether a resident who was a better fit with a particular program's core characteristics would be more likely to thrive in that program compared to one where their core characteristics were less aligned. It stands to reason that if a program and resident share similar goals and characteristics, the resident and program would both benefit.

What is the Impact of SJTs on the Applicant Pool Invited for Interviews?

Because of their discriminant validity to weight the application process more toward non-cognitive attributes, SJTs produce a decidedly different applicant pool compared to traditional metrics. In one study, only 23% of applicants identified through an SJT would have been selected for interview based on traditional application review alone. Further strengthening their use, the authors noted that of all 7 of their matched PGY1 residents, none would have been offered an interview if traditional metrics had been applied in the selection process [29].

Improving trainee and physician diversity is critical for developing a diverse work force that serves the needs of all patients. Traditional screening methods disproportionately impact URM [14]. Non-white students are more likely to receive lower scores or fail the USMLE, achieve lower grades in all clerkships, and may be less likely to be elected to AOA [1, 9, 30]. Additionally, granular assessment of letters of recommendation reveals that URM applicants are less likely to be described as outstanding, excellent, very good, or good [31]. Here, too, SJTs may offer some benefit as they have been shown to increase the number of women and applicants from a lower socioeconomic class [32, 33].

The ability of SJTs to improve cultural diversity during the medical student application process is less clear, with both of the above studies showing that African American and Hispanic/Latino applicants scored lower on SJTs. Importantly, however, the difference in SJT score was significantly smaller than the difference when using traditional metrics, suggesting that SJTs may decrease, but not eliminate, the effect of bias introduced through more traditional cognitive measures. Similarly, although students of lower socioeconomic status did not fare as well on the CASPer, differences were less significant than differences observed with academic metrics. Taken together, as the weight of the SJT increases compared to the weight of cognitive factors, the number of females, African Americans, and Hispanic/Latinos also increases [33].

Studies in the postdoctoral selection process have also been shown to increase resident and fellow diversity. In a large multi-institutional study of surgical residents across

7 programs, use of a customized SJT and lowering the USMLE cutoff resulted in all but one program increasing the number of URM applicants for interview, ranging from a 1 to 17% increase [34]. Similarly, in screening fellowship applicants, the use of an SJT resulted in a 22% absolute increase in the percentage of URM applicants being invited for interviews compared to traditional methods [28]. While encouraging, these studies included women and other groups within their definition of URM that typically do not fall under the traditional definition of URM, such that the effect of SJTs to increase ethnic diversity in the postdoctoral setting remains unknown.

Faculty and Applicant Perception of the Use of SJTs

In considering the use of SJTs in the applicant screening process, it is important to consider the perspectives of both the faculty and the applicant. A single study that utilized an SJT in the screening process for a surgical fellowship found universal agreement among 5 faculty that there was value in the process of developing the SJT. The process itself helped them understand attributes important for fellows at their program, and they had greater confidence in identifying which candidates would be a good fit [28]. It is not surprising that the faculty investing significant time in the process of SJT development would reflect positively on its use. Further research is required to determine if interviewers can determine a difference in candidates during the interview when blinded to their program compatibility based on SJTs.

The use of additional testing or requirements beyond the standard application increases the burden to applicants as SJTs may take up to 75 min to complete [33]. If implemented by individual programs, the additional time would likely be overly burdensome and may dissuade applicants from applying to certain programs [26]. Despite these time demands, survey data of applicants to a variety of postgraduate programs demonstrates that a majority of applicants perceive SJT as relevant and easy to complete and would not deter them from applying [28, 35, 36]. Nevertheless, despite seeing some benefit to SJTs, most applicants believe that the traditional process (interviews, letter of recommendation, and past achievements) is more representative of them as an applicant [35, 36]. After spending years honing their academic achievements, applicants may be concerned about basing their future on (another) high-stakes test. As such, any attempt to institute SJTs broadly would likely be met with resistance and would require significant education of all parties on their merits and established validity and reliability [26, 35]. Importantly, SJTs should be seen only as an adjunct to traditional measures, as each portion of the application is predictive of separate performance metrics.

Limitations of SJTs

Although they have shown significant promise, SJTs also have several limitations [37], not the least of which is generalizability. Residency programs, more so than medical schools, have unique values, culture, and performance measures, which means each program may need a program-specific SJT [26]. Unfortunately, SJTs are complex and resource-heavy in their initial development, which requires identification of key applicant attributes, question generation, consensus scoring, tests for validity, and continued refinement (Fig. 1) [26]. Most often, this requires outside help from experts in organizational science and buy-in from key stakeholders. While there are consulting firms that will perform this work, the upfront cost for even a small program easily may exceed \$50,000.

Regarding the tests themselves, SJT scenarios tend to be brief, which may remove some of the intended realism and reduce the quality and depth of candidate assessment. Furthermore, SJTs that rely on multiple choice answers may lead a candidate to select a scenario that varies from their natural response. Finally, each individual SJT question is likely to be multidimensional, making it difficult to test any one specific attribute.

In general, SJTs have shown validity, increase diversity, and correlate with competency performance metrics such as the ACGME milestones. However, the long-term utility of SJTs may be questioned as they become more common, and more resources are spent on coaching for the exam. The ability to discriminate between non-cognitive and cognitive factors may be compromised if applicants begin to study ways to master SJT exams. As it stands, there are dozens of websites and companies ready to coach applicants in SJTs.

Importantly, coachability can be decreased, though not eliminated, by using a knowledge-based format and institution-specific questions and by increasing the complexity of the assessment [10].

Important to the baseline understanding of SJTs is that they only measure the constructs to which they are designed to test. As such, the evidence for determining performance can be difficult to measure as objective measurements of success do not always measure the attributes that are sought in development of an SJT. The metrics to which we should measure outcomes are ill-defined and determining the benefit of SJTs may well require entirely different performance metrics (e.g., an SJT on empathy should not be measured through an ABSITE score). Similarly, particular traits do not exist in a vacuum, and we may not fully understand the interactions between various “good” and “bad” attributes and how they affect resident and physician performance.

Finally, using SJTs as a method for initial screening suggests that characteristics that are important in residents and future physicians are fixed and can only be acquired prior to medical training. Like many things in training, it is possible that core characteristics can be learned if specific training is provided in the right context. Currently, it is not known to what extent these characteristics may be modifiable.

Personality Assessment Tool (PAT)

Personality is influenced by genetics and environment and cannot be directly observed. Although personality remains stable over time, self-awareness can enable a person to overcome undesirable traits [38–40]. PATs are tests that seek information about a person’s motivations, preferences,

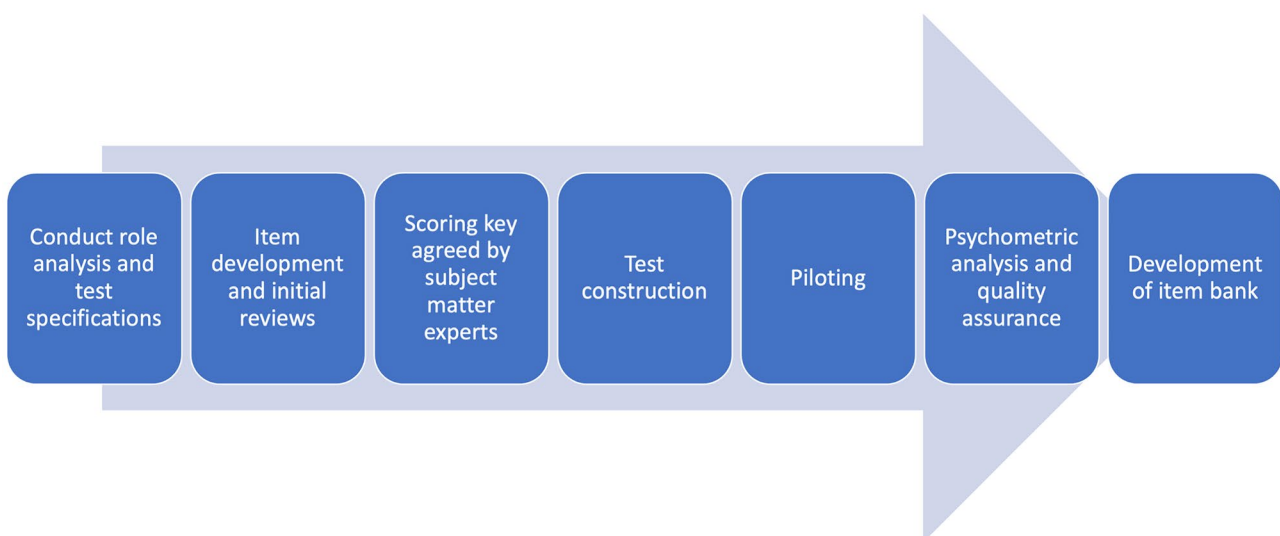


Fig. 1 Development of SJT [18]

interests, emotional makeup, and interaction with others and their environment to categorize their personality type. The five most commonly analyzed personality domains (“Big 5”) are agreeableness, conscientiousness, extroversion, neuroticism (emotional stability), and openness to experience (Table 2) [39, 41–43]. Beyond the individual personality characteristics, PATs can categorize individuals into discrete personality types (e.g., Myers-Briggs type indicator) or provide assessments along a spectrum (e.g., Big Five personality test).

PATs have been used in business and industry for decades as it is recognized that personality characteristics and job performance are related across a variety of occupations [44, 45]. Indeed, some studies have shown that personality is the third best predictor of job performance, behind cognitive ability and job-related knowledge [44]. As a result, many organizations have implemented personality testing in job applicant screening, leadership development, employee onboarding, coaching, and team building, demonstrating improvement in organization outcomes such as job satisfaction, decreased attrition, and work motivation [40, 46].

In medicine, PATs have been studied to define personality characteristics common to specific specialties, define generational differences, aid in mentorship, and characterize personality types of medical students, residents, and faculty. On personality assessment, surgeons score higher on conscientiousness and extroversion but lower on agreeableness and neuroticism relative to general practitioners [47–50]. Within urology, residents score higher for extroversion, openness, and conscientiousness relative to the general population [51].

To be a successful resident and physician, it takes more than just cognitive ability. Efforts to identify and define personality traits of successful residents are emerging. While PATs may provide reassurance in creation of final rank lists, ensuring that candidates have traits that are compatible with the program, their use as a screening tool and ability to predict future success as a resident or physician remains poorly defined [52].

Are There Specific Personality Traits that Perform Better in Medicine?

Results are mixed as to the ability of personality tests to predict overall resident performance. Studies supporting a link between personality and performance identified that high-performing residents have higher scores on cooperation, self-efficacy, adventurousness, extroversion, conscientiousness, agreeableness, and emotional stability and lower scores on neuroticism, anxiety, anger, and vulnerability [41, 53, 54]. Similarly, residents who score higher on independence have higher case volumes and completeness within their surgical case logs, and those who perform poorly in traits

linked to stress (excitable, skeptical, and imaginative) perform poorly on tasks related to communication compared to those with high scores in emotional stability, agreeableness, conscientiousness, and openness [55]. Other studies have shown that while residents who are rated higher in outgoingness and kindness have higher medical student evaluations, personality characteristics were not related to faculty evaluations or overall performance [26]. Taken together, these studies indicate that while there is some information emerging linking personality characteristics with high-performing residents, data is limited and future work needs to be done to understand the role of personality in resident performance.

Can PATs Help in Resident Selection?

Part of the challenge of the current resident selection process has been the dependence on objective data such as grades, USMLE scores, and publications. As we are seeing shifts in these measures, many are seeking alternative metrics that may provide surrogate correlation such as personality tests [56]. The Residency Select/J3Personica test is a validated instrument developed specifically to assess characteristics that are expected in residency and is based on the concept that applicants can be compared with individual program profiles and national benchmarks to determine personality fit [56]. There is a paucity of research regarding this instrument, though there is some evidence that a low score on the imaginative scale correlated with USMLE, and high adjustment scale correlated with greater number of publications [57].

Interviewers and interviewees often use the interview process as a means to judge personality fit within a program. From the program’s perspective, the interview is important for assessment of non-academic factors, including personality, and directly affects rank list [58, 59]. From the applicant’s perspective, interviews provide an opportunity to present desirable traits and fit [60, 61]. Contrary to perceptions, studies have demonstrated no correlation between formal applicant PAT results and rank on the match list, suggesting that personality testing evaluates traits and fit differently than what is measured in a traditional interview [41, 62].

Limitations of PATs

While they are more generic than SJTs and, therefore, may be used across a wide variety of programs without program-specific questions, PATs remain time-consuming for completion and can come at a cost upwards of \$1000. Because of the number of attributes that need to be measured and analyzed, interpretation of PATs may also be challenging and, therefore, it is critical to work with an organization psychologist or other subject matter expert to ensure proper selection, administration, and interpretation [40].

Table 2 Summary of the “Big Five” personality traits [41–43]

Trait	Description	Characteristics assessed	High score	Low score
Agreeableness	Degree to which applicant is oriented toward community	Cooperation, empathy, trustworthiness, sympathy	Optimistic view of human nature; greater levels of altruism, trust, and modesty	Confronts poor performers; tough; willing to take unpopular positions; may place own interests above others
Conscientiousness	Tendency toward self-control in order to achieve goals	Self-discipline, detail-orientation, organization, goal-directedness	Orderly, reliable, attentive to rules and details, prioritizes tasks	Comfortable with ambiguity, flexible about rules, capable of quickly changing direction
Extraversion	Tendency to seek out activity and company of others; comprises: Ambition Degree to which a person seems socially self-confident, leader-like, competitive, energetic Sociability Degree a person seems to need or enjoy social interaction	Talkativeness, energy, assertiveness, outgoingness	High energy, productive team leader, goal oriented	Laid back, flexible, good team player
Neuroticism	Tendency to experience negative emotions, emotional triggers	Confidence, self-acceptance, stability under pressure	Approachable and friendly, busy and full of energy, talks more than listens, builds relationships with a variety of people Less stress tolerant, anxious, or tense	Talk, task-oriented, works well independently, prefers structured meetings Handles pressure well and expects to succeed, solid self-esteem, positive outlook on life, higher emotional stability
Openness	Tendency to seek out activity and company of others, creative and multiple interests, learn new things; comprises: Inquisitive Evaluates imagination and problem-solving Learning approach Extent to which a person seems to enjoy learning	Creativity, abstract thinking, curiosity, imagination	Creative, adventurous, open-minded, thinks quickly and strategically in the moment about the big picture Goal-focused, values education for the sake of education, averse to working on uninteresting tasks	Levelheaded, pragmatic, process-focused Hands-on approach to learning, focus on applying existing skills, might be unconcerned with learning new skills

As with any assessment, validity is always a concern. Similar to SJTs, there is significant concern that a PAT as a high-stakes exam may be coachable and that applicants may lean toward characteristics they think will be more desirable to programs. Indeed, research has shown that in a high-stakes environment, applicants may engage in substantial response distortion in order to display characteristics that may be more socially desirable [63, 64]. Though response distortion adds noise to the assessment, it has less impact on rank ordering of applicants as applicants with lower scores tend to distort responses more [63]. As with SJTs, personality testing should not be used in isolation when measuring interpersonal constructs and should only be considered in the overall context of the remaining application.

Conclusions

SJTs and PATs have been successfully utilized in the screening of applicants across a wide range of industries. With rapid changes and poor validity of traditional metrics in resident selection, SJTs and PATs may be considered as adjunct measures of non-cognitive applicant attributes. Combined with traditional metrics, the non-cognitive measures contribute discriminant validity that gives a better all-around picture of each individual.

Compliance with Ethical Standards

Conflict of Interest The authors declare that they have no conflict of interest.

Human and Animal Rights and Informed Consent This article does not contain any studies with human or animal subjects performed by any of the authors.

References

- Byyny RL, Martinez D, Cleary L, Ballard B, Barth BE, Christensen S, et al. Alpha Omega Alpha Honor Medical Society: a commitment to inclusion, diversity, equity, and service in the profession of medicine. *Acad Med.* 2020;95(5):670–3. <https://doi.org/10.1097/ACM.0000000000003088>.
- Roth AE. The origins, history, and design of the resident match. *JAMA.* 2003;289(7):909–12. <https://doi.org/10.1001/jama.289.7.909>.
- Tidwell J, Yudien M, Rutledge H, Terhune KP, LaFemina J, Aarons CB. Reshaping residency recruitment: achieving alignment between applicants and programs in surgery. *J Surg Educ.* 2022;79(3):643–54. <https://doi.org/10.1016/j.jsurg.2022.01.004>.
- Winterton M, Ahn J, Bernstein J. The prevalence and cost of medical student visiting rotations. *BMC Med Educ.* 2016;16(1):291. <https://doi.org/10.1186/s12909-016-0805-z>.
- American Urological Association. Urology Residency Match Statistics. 2022. <https://www.auanet.org/meetings-and-education/for-residents/urology-and-specialty-matches>. Accessed May 2022.
- Pershing S, Co JPT, Katznelson L. The new USMLE Step 1 paradigm: an opportunity to cultivate diversity of excellence. *Acad Med.* 2020;95(9):1325–8. <https://doi.org/10.1097/ACM.0000000000003512>.
- Huang MM, Clifton MM. Evaluating urology residency applications: what matters most and what comes next? *Curr Urol Rep.* 2020;21(10):37. <https://doi.org/10.1007/s11934-020-00993-0>.
- 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(12):1253–6. <https://doi.org/10.1097/00001888-200112000-00021>.
- Rubright JD, Jodoin M, Barone MA. Examining demographics, prior academic performance, and United States medical licensing examination scores. *Acad Med.* 2019;94(3):364–70. <https://doi.org/10.1097/ACM.0000000000002366>.
- Cullen MJ, Zhang C, Marcus-Blank B, Braman JP, Tiryaki E, Konia M, et al. Improving our ability to predict resident applicant performance: validity evidence for a situational judgment test. *Teach Learn Med.* 2020;32(5):508–21. <https://doi.org/10.1080/10401334.2020.1760104>.
- Winkel AF, Morgan HK, Burk-Rafel J, Dalrymple JL, Chiang S, Marzano D, et al. A model for exploring compatibility between applicants and residency programs: right resident, right program. *Obstet Gynecol.* 2021;137(1):164–9. <https://doi.org/10.1097/AOG.0000000000004179>.
- Patel H, Yakkanti R, Bellam K, Agyeman K, Aiyer A. Innovation in resident selection: life without step 1. *J Med Educ Curric Dev.* 2022;9:23821205221084936. <https://doi.org/10.1177/23821205221084936>.
- Gabrielson AT, Kohn JR, Sparks HT, Clifton MM, Kohn TP. Proposed changes to the 2021 residency application process in the wake of COVID-19. *Acad Med.* 2020;95(9):1346–9. <https://doi.org/10.1097/ACM.0000000000003520>.
- Phillips MR, Charles A. Addressing implicit bias in the surgical residency application and interview process for underrepresented minorities. *Surgery.* 2021;169(6):1283–4. <https://doi.org/10.1016/j.surg.2021.01.018>.
- Association of American Medical Colleges. Holistic review. 2022. <https://www.aamc.org/services/member-capacity-building/holistic-review>. Accessed May 2022.
- Sungar WG, Angerhofer C, McCormick T, Zimmer S, Druck J, Kaplan B, et al. Implementation of holistic review into emergency medicine residency application screening to improve recruitment of underrepresented in medicine applicants. *AEM Educ Train.* 2021;5(Suppl 1):S10–8. <https://doi.org/10.1002/aet2.10662>.
- Koenig TW, Parrish SK, Terregino CA, Williams JP, Dunleavy DM, Volsch JM. Core personal competencies important to entering students' success in medical school: what are they and how could they be assessed early in the admission process? *Acad Med.* 2013;88(5):603–13. <https://doi.org/10.1097/ACM.0b013e31828b3389>.
- Patterson F, Zibarras L, Ashworth V. Situational judgement tests in medical education and training: research, theory and practice: AMEE Guide No. 100. *Med Teach.* 2016;38(1):3–17. <https://doi.org/10.3109/0142159X.2015.1072619>.
- Cardall AJ. Preliminary manual for the test of practical judgement. Chicago: Science Research Associates; 1942.
- Chan D, Schmitt N. Situational judgment and job performance. *Hum Perform.* 2002;15(3):233–54. https://doi.org/10.1207/s15327043hup1503_01.
- Koczwara A, Patterson F, Zibarras L, Kerrin M, Irish B, Wilkinson M. Evaluating cognitive ability, knowledge tests and situational judgement tests for postgraduate selection. *Med Educ.* 2012;46(4):399–408. <https://doi.org/10.1111/j.1365-2923.2011.04195.x>.

22. Marcus-Blank B, Dahlke JA, Braman JP, Borman-Shoap E, Tiryaki E, Chipman J, et al. Predicting performance of first-year residents: correlations between structured interview, licensure exam, and competency scores in a multi-institutional study. *Acad Med.* 2019;94(3):378–87. <https://doi.org/10.1097/ACM.0000000000002429>.
23. Lievens F, Sackett PR. The validity of interpersonal skills assessment via situational judgment tests for predicting academic success and job performance. *J Appl Psychol.* 2012;97(2):460–8. <https://doi.org/10.1037/a0025741>.
24. Dore KL, Reiter HI, Kreuger S, Norman GR. CASPer, an online pre-interview screen for personal/professional characteristics: prediction of national licensure scores. *Adv Health Sci Educ Theory Pract.* 2017;22(2):327–36. <https://doi.org/10.1007/s10459-016-9739-9>.
25. Association of American Medical Colleges. AAMC PREview professional readiness exam. 2022. <https://students-residents.aamc.org/aamc-preview/aamc-preview-professional-readiness-exam>. Accessed 19 May 2022.
26. Gardner AK, Dunkin BJ. Evaluation of validity evidence for personality, emotional intelligence, and situational judgment tests to identify successful residents. *JAMA Surg.* 2018;153(5):409–16. <https://doi.org/10.1001/jamasurg.2017.5013>.
27. Saxena A, Desanghere L, Dore K, Reiter H. Incorporating situational judgment tests into postgraduate medical education admissions: examining educational and organizational outcomes. *Acad Med.* 2021;96(11S):S203–4. <https://doi.org/10.1097/ACM.0000000000004280>.
28. Gardner AK, Dunkin BJ. Pursuing excellence: the power of selection science to provide meaningful data and enhance efficiency in selecting surgical trainees. *Ann Surg.* 2019;270(1):188–92. <https://doi.org/10.1097/SLA.0000000000002806>.
29. Lyons J, Bingmer K, Ammori J, Marks J. Utilization of a novel program-specific evaluation tool results in a decidedly different interview pool than traditional application review. *J Surg Educ.* 2019;76(6):e110–7. <https://doi.org/10.1016/j.jsurg.2019.10.007>.
30. Lee KB, Vaishnavi SN, Lau SK, Andriole DA, Jeffe DB. “Making the grade:” noncognitive predictors of medical students’ clinical clerkship grades. *J Natl Med Assoc.* 2007;99(10):1138–50.
31. Rinard JR, Mahabir RC. Successfully matching into surgical specialties: an analysis of national resident matching program data. *J Grad Med Educ.* 2010;2(3):316–21. <https://doi.org/10.4300/JGME-D-09-00020.1>.
32. Lievens F, Patterson F, Corstjens J, Martin S, Nicholson S. Widening access in selection using situational judgement tests: evidence from the UKCAT. *Med Educ.* 2016;50(6):624–36. <https://doi.org/10.1111/medu.13060>.
33. Juster FR, Baum RC, Zou C, Risucci D, Ly A, Reiter H, et al. Addressing the diversity-validity dilemma using situational judgment tests. *Acad Med.* 2019;94(8):1197–203. <https://doi.org/10.1097/ACM.0000000000002769>.
34. Gardner AK, Cavanaugh KJ, Willis RE, Dunkin BJ. Can better selection tools help us achieve our diversity goals in postgraduate medical education? Comparing use of USMLE Step 1 scores and situational judgment tests at 7 surgical residencies. *Acad Med.* 2020;95(5):751–7. <https://doi.org/10.1097/ACM.0000000000003092>.
35. Reed BN, Armahizer MJ, Devabhakthuni S, Lemens L, Yeung SYA. Candidate reactions to a postgraduate year 1 pharmacy residency supplemental application. *Am J Health Syst Pharm.* 2022. <https://doi.org/10.1093/ajhp/zxac007>.
36. Shipper ES, Mazer LM, Merrell SB, Lin DT, Lau JN, Melcher ML. Pilot evaluation of the computer-based assessment for sampling personal characteristics test. *J Surg Res.* 2017;215:211–8. <https://doi.org/10.1016/j.jss.2017.03.054>.
37. Benbassat J. Assessments of non-academic attributes in applicants for undergraduate medical education: an overview of advantages and limitations. *Med Sci Educ.* 2019;29(4):1129–34. <https://doi.org/10.1007/s40670-019-00791-5>.
38. Lewin K. A dynamic theory of personality: selected papers. *J Nerv Ment Dis.* 1936;84(5):612–3.
39. Costa PT, McCrae RR, Kay GG. Persons, places, and personality: career assessment using the revised NEO personality inventory. *J Career Assess.* 2016;3(2):123–39. <https://doi.org/10.1177/106907279500300202>.
40. Tornetta P 3rd, Jacobs JJ, Sterling RS, Kogan M, Fletcher KA, Friedman AM. Personality assessment in orthopaedic surgery: AOA critical issues. *J Bone Joint Surg Am.* 2019;101(4): e13. <https://doi.org/10.2106/JBJS.18.00578>.
41. Hughes BD, Perone JA, Cummins CB, Sommerhalder C, Tyler DS, Bowen-Jallow KA, et al. Personality testing may identify applicants who will become successful in general surgery residency. *J Surg Res.* 2019;233:240–8. <https://doi.org/10.1016/j.jss.2018.08.003>.
42. Srivastava S. Measuring the Big Five personality factors. 2022. <https://psdlab.uoregon.edu/bigfive.html>. Accessed 30 June 2022.
43. Hogan Assessments. The Big Five personality characteristics: a look behind the Hogan personality tests. 2021. <https://www.hoganassessments.com/blog/big-five-personality-characteristics-behind-hogan-personality-tests>. Accessed 30 June 2022.
44. Barrick MR, Mount MK. The Big Five personality dimensions and job performance: a meta-analysis. *Pers Psychol.* 1991;44(1):1–26. <https://doi.org/10.1111/j.1744-6570.1991.tb00688.x>.
45. Tett RP, Jackson DN, Rothstein M. Personality measures as predictors of job performance: a meta-analytic review. *Pers Psychol.* 2006;44(4):703–42. <https://doi.org/10.1111/j.1744-6570.1991.tb00696.x>.
46. Beagrie S. How to excel at psychometric assessments. *Personnel Today.* 2005;25:25.
47. Drosdeck JM, Osayi SN, Peterson LA, Yu L, Ellison EC, Muscarella P. Surgeon and nonsurgeon personalities at different career points. *J Surg Res.* 2015;196(1):60–6. <https://doi.org/10.1016/j.jss.2015.02.021>.
48. Borges NJ, Savickas ML. Personality and medical specialty choice: a literature review and integration. *J Career Assess.* 2016;10(3):362–80. <https://doi.org/10.1177/1067272010003006>.
49. McGreevy J, Wiebe D. A preliminary measurement of the surgical personality. *Am J Surg.* 2002;184(2):121–5. [https://doi.org/10.1016/s0002-9610\(02\)00919-4](https://doi.org/10.1016/s0002-9610(02)00919-4).
50. Thomas JH. The surgical personality: fact or fiction. *Am J Surg.* 1997;174(6):573–7. [https://doi.org/10.1016/s0002-9610\(97\)00208-0](https://doi.org/10.1016/s0002-9610(97)00208-0).
51. Eng MK, Macneily AE, Alden L. The urological personality: is it unique? *Can J Urol.* 2004;11(5):2401–6.
52. Bell RM, Fann SA, Morrison JE, Lisk JR. Determining personal talents and behavioral styles of applicants to surgical training: a new look at an old problem. part I. *J Surg Educ.* 2011;68(6):534–41. <https://doi.org/10.1016/j.jsurg.2011.05.016>.
53. Merlo LJ, Matveevskii AS. Personality testing may improve resident selection in anesthesiology programs. *Med Teach.* 2009;31(12):e551–4. <https://doi.org/10.3109/01421590903390593>.
54. Phillips D, Egol KA, Maculatis MC, Roloff KS, Friedman AM, Levine B, et al. Personality factors associated with resident performance: results from 12 accreditation council for graduate medical education accredited orthopaedic surgery programs. *J Surg Educ.* 2018;75(1):122–31. <https://doi.org/10.1016/j.jsurg.2017.06.023>.
55. Holmes KS, Zuckerman JD, Maculatis MC, Friedman AM, Lawrence E, Phillips DP. Personality predictors of communication skills among orthopedic surgery residents. *J Surg Educ.* 2020;77(1):202–12. <https://doi.org/10.1016/j.jsurg.2019.08.012>.
56. Horan DP, Baldwin K, Purtill JJ, Namdari S. Predictors of success in an orthopaedic residency. *JBJS Rev.* 2021. <https://doi.org/10.2106/JBJS.RVW.20.00180>.

57. Lubelski D, Healy AT, Friedman A, Ferraris D, Benzel EC, Schlenk R. Correlation of personality assessments with standard selection criteria for neurosurgical residency applicants. *J Neurosurg.* 2016;125(4):986–94. <https://doi.org/10.3171/2015.7.JNS15880>.
58. Johnson EK, Edwards JC. Current practices in admission interviews at U.S. medical schools. *Acad Med.* 1991;66(7):408–12. <https://doi.org/10.1097/00001888-199107000-00008>.
59. Gong H Jr, Parker NH, Apgar FA, Shank C. Influence of the interview on ranking in the residency selection process. *Med Educ.* 1984;18(5):366–9. <https://doi.org/10.1111/j.1365-2923.1984.tb01284.x>.
60. Marciani RD, Smith TA, Kohn MW. Applicants' opinions about the selection process for oral surgery programs. *J Oral Surg.* 1977;35(8):648–51.
61. Marciani RD, Smith TA, Heaton LJ. Applicants' opinions about the selection process for oral and maxillofacial surgery programs. *J Oral Maxillofac Surg.* 2003;61(5):608–14. <https://doi.org/10.1053/joms.2003.50091>.
62. Frantsve LM, Laskin DM, Auerbach SM. Personality and gender influences on faculty ratings and rankings of oral and maxillofacial surgery residency applicants. *J Dent Educ.* 2003;67(11):1252–9.
63. Anglim J, Bozic S, Little J, Lievens F. Response distortion on personality tests in applicants: comparing high-stakes to low-stakes medical settings. *Adv Health Sci Educ Theory Pract.* 2018;23(2):311–21. <https://doi.org/10.1007/s10459-017-9796-8>.
64. Wood JK, Anglim J, Horwood S. Effect of job applicant faking and cognitive ability on self-other agreement and criterion validity of personality assessments. *Int J Sel Assess.* 2022. <https://doi.org/10.1111/ijasa.12382>.

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

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.