Faculty Assessment of Emergency Medicine Resident Grit: A Multicenter Study

Nathan Olson, MD, MAEd, Adriana Segura Olson, MD, MAEd, Kelly Williamson, MD, Nicholas Hartman, MD, MPH, Jeremy Branzetti, MD, and Patrick Lank, MD, MS, for the Emergency Medicine Education Research Alliance (EMERA)

ABSTRACT

Background: Assessment of trainees' competency is challenging; the predictive power of traditional evaluations is debatable especially in regard to noncognitive traits. New assessments need to be sought to better understand affective areas like personality. Grit, defined as "perseverance and passion for long-term goals," can assess aspects of personality. Grit predicts educational attainment and burnout rates in other populations and is accurate with an informant report version. Self-assessments, while useful, have inherent limitations. Faculty's ability to accurately assess trainees' grit could prove helpful in identifying learner needs and avenues for further development.

Objective: This study sought to determine the correlation between EM resident self-assessed and faculty-assessed Grit Scale (Grit-S) scores of that same resident.

Methods: Subjects were PGY-1 to -4 EM residents and resident-selected faculty as part of a larger multicenter trial involving 10 EM residencies during 2017. The Grit-S Scale was administered to participating EM residents; an informant version was completed by their self-selected faculty. Correlation coefficients were computed to assess the relationship between residents' self-assessed and the residents' faculty-assessed Grit-S score.

Results: A total of 281 of 303 residents completed the Grit-S, for a 93% response rate; 200 of 281 residents had at least one faculty-assessed Grit-S score. No correlation was found between residents' self-assessed and faculty-assessed Grit-S scores. There was a correlation between the two faculty-assessed Grit-S scores for the same resident.

Conclusion: There was no correlation between resident and faculty-assessed Grit-S scores; additionally, faculty-assessed Grit-S scores of residents were higher. This corroborates the challenges faculty face at accurately assessing aspects of residents they supervise. While faculty and resident Grit-S scores did not show significant concordance, grit may still be a useful predictive personality trait that could help shape future training.

Accurately assessing trainee fortitude and resolve is challenging for educators.[1,2] Traditionally, assessments of trainee characteristics have focused on test scores, grades, and evaluations. The value and predictive power of traditional evaluations remain debatable, although it does seem that certain standardized tests can predict later scores, grades, and achievement.[3] In the field of postgraduate medical education,

higher scores on Step 1 of the United States Medical Licensing Examination (USMLE) confer a greater likelihood of passing the specialty boards after training in general surgery, pediatrics, obstetrics and gynecology, and emergency medicine (EM).[4–7] However, test-taking aptitude and knowledge acquisition are not the only domains that determine the effectiveness of physicians. Noncognitive affective traits, including persistence, self-discipline, and teamwork skills, are essential to producing well-rounded competent physicians.[8] These traits are beginning to be explored through a variety of assessment tools.[8–10] As a more expansive view on competency develops, new assessments of skills and personality traits are needed. However, these are often less quantifiable than medical knowledge and it is unclear how well faculty, who evaluate and mentor residents, are at assessing noncognitive traits.

Grit, defined as "perseverance and passion for long-term goals" has emerged as a means to quantify an aspect of personality.[11,12] Grit is a noncognitive trait and is not correlated with IQ.[11] The Grit score, calculated through a self-reported questionnaire, has been found to help predict items such as educational attainment, grade point average among Ivy League undergraduates, and retention among cadets at the United States Military Academy.[13–15] Within medicine, the Grit score has been found to negatively correlate with surgery resident burnout rates and likelihood of leaving their training program and to positively correlate with pharmacy students' pursuit of residencies and fellowships.[16–18] The short Grit Scale (Grit-S) is a validated eight-question test scored on a 1–5 scale (5 is the highest score) with the average of the eight responses representing a person's grit. The Grit-S has been shown to predict the same outcomes as the original Grit Scale but in a more efficient manner.[19]

In other populations, the Grit-S score has also been shown to be accurate with an informant report version, which is filled out by someone who knows the subject well.[4] Informant report scores of noncognitive assessments are important as research has shown that personality traits are likely better assessed in a multimodal fashion as this may help address some aspects of response bias. Additionally, some personality traits have been shown to be better assessed by others rather than by oneself.[20,21] Therefore, as personality traits are being explored in medicine, it may be useful to gather both self-assessed and informant-assessed scores when possible. Since Grit has been shown to predict achievement in other populations as well as correlate with burnout and attrition rates in surgical residents, this type of brief quantifiable measurement of an important personality trait holds promise for medical educators in determining the current and future needs of their trainees.[16,17] Knowledge of Grit-S scores may be useful in helping faculty foster Grit in residents as well as to identify residents that may benefit from closer monitoring and training. However, these are often less quantifiable than medical knowledge and it is unclear how well faculty can assess noncognitive traits in their residents. In this study, we sought to assess the correlation between an EM resident-assessed Grit-S score and a faculty-assessed Grit-S score of that same resident.

METHODS

Study Setting

This study was a secondary analysis of a larger multi-center educational trial investigating the effectiveness of a wellness didactic curriculum and involved 10 allopathic EM residency programs in the United States during 2017. All sites obtained institutional review board approval. Eight sites were postgraduate year (PGY)-1 to -3 residencies and two sites were PGY-1 to -4 residencies.

Study Subjects

Study subjects were PGY-1 to -4 EM residents at 10 allopathic Accreditation Council for Graduate Medical Education (ACGME)-approved United States EM residencies. There were no further exclusion criteria for resident subjects. Additional study subjects were resident-selected faculty at the 10 sites who were asked to participate in person by each respective site investigator. Faculty were eligible to participate if they worked primarily at the main clinical hospital site of the residency program.

Study Protocol

Informed consent was obtained from all study participants. In February 2017, the Grit-S was administered to all EM residents participating in the study. After completing the Grit-S, resident study participants were asked to "list two faculty members that you feel know you well. These two faculty members will fill out a Grit-S about you." Site investigators then asked identified faculty participants to complete the Grit-S as pertaining to that particular resident. Each resident was assigned a unique identifier number known only to the individual participant and the respective site

investigator. Faculty were given the name of the resident to complete the form, but only the unique identifier number remained on the form when it was returned to the site investigator.

Statistical Analysis

Categorical outcomes were summarized with frequencies and percentages and continuously distributed outcomes were summarized with the sample size, mean, and standard deviation (SD. The resident-assessed Grit-S scores by PGY were assessed with a one-way analysis of variance (ANOVA). A Pearson's product-moment correlation coefficient was computed to assess the relationship between residents' self-assessed Grit-S score, the residents' faculty-assessed Grit-S score, and the relationship between the two faculty members who calculated a Grit-S score for the same resident.

RESULTS

A total of 281 of 303 residents completed the Grit-S as part of a larger study for a response rate of 93%.

Table 1
Demographics of participants

Participant Characteristics	$n = 281$*
Mean age[†]	30 ($\pm$3.1)
Gender[‡]	
Male	197 (70.4)
Female	83 (29.6)
Ethnicity[§]	
Caucasian	201 (73.9)
Mixed	15 (5.5)
Latino	14 (5.1)
East Asian	14 (5.1)
African American	11 (4)
Asian	10 (3.7)
Middle Eastern	6 (2.2)
Pacific Islander	1 (0.4)
EM PGY	
PGY-1	81 (28.8)
PGY-2	94 (33.5)
PGY-3	88 (31.3)
PGY-4	18 (6.4)
Resident self-assessed Grit-S score	3.58 ($\pm$0.54)
Faculty-assessed Grit-S score[¶]	4.22 ($\pm$0.54)

Data are reported as mean ($\pm$SD) or n (%).
*a 281/303 participants completed self-assessed Grit-S score.
[†]$n = 262$, 19 participants did not indicate age.
[‡]$n = 280$, 1 participant did not indicate
[§]$n = 272$, 9 participants did not indicate ethnicity
[¶]$n = 200$, number of participants with at least 1 faculty-assessed Grit-S Score

The mean ($\pm$SD) age of participants was 30 ($\pm$3.1) years. The participants were 70% male and 30% female (Table 1).

The mean ($\pm$SD) resident-assessed Grit-S score was 3.58 ($\pm$0.54; Table 1 and Figure 1). There was no statistically significant difference between the Grit-S scores of each PGY($p = 0.976$; Table 2).

Of the 281 residents with a self-assessed Grit-S score 81 residents did not have a faculty member assess their Grit-S score either because they did not select faculty or the faculty did not return the Grit-S; these residents were removed from further analysis. Of the 200 residents with faculty-assessed Grit-S scores, 174 had two faculty assessed Grit-S scores and 26 only had one faculty assessed Grit-S score. The mean ($\pm$SD) faculty-assessed Grit-S score for residents was 4.22 ($\pm$0.54; Table 1). A subgroup analysis by ANOVA of faculty-assessed Grit-S scores by PGY could not be performed secondary to a small sample size (Table 2). There was no correlation between the residents' self-assessed Grit-S score and the residents' faculty-assessed Grit-S score ($r = 0.13$, $n = 200$, $p = 0.06$; Figure 2). Of the residents with faculty-assessed Grit-S scores, 167 of 200 (84%) had lower self-assessed Grit-S scores than faculty-assessed Grit-S scores. There was a moderate correlation between the two faculty-assessed Grit-S scores for the same resident ($r = 0.47$, $n = 174$, $p < 0.0001$; Figure 3).

DISCUSSION

Graduate medical education has moved toward competency-based standards.[22–24] The goals of competency-based education encompass both cognitive knowledge-based skills and practice and affective noncognitive attributes such as professionalism, personality, communication, teamwork, empathy, and perseverance. The field of medicine does not fully understand what assessments, tools, or trainings help advance the noncognitive domains of competency-based education.[24] Multiple studies have indicated that these noncognitive skills and traits, including personality, can help predict career interest and satisfaction, academic performance, and health care clinical outcomes.[8,11,14,19,25,26] Different aspects of personality such as conscientiousness and openness have been shown to be important in gaining and applying medical knowledge.[8,27–29] Research indicates that people are relatively effective at providing accurate self-assessments about noncognitive traits; however, they have
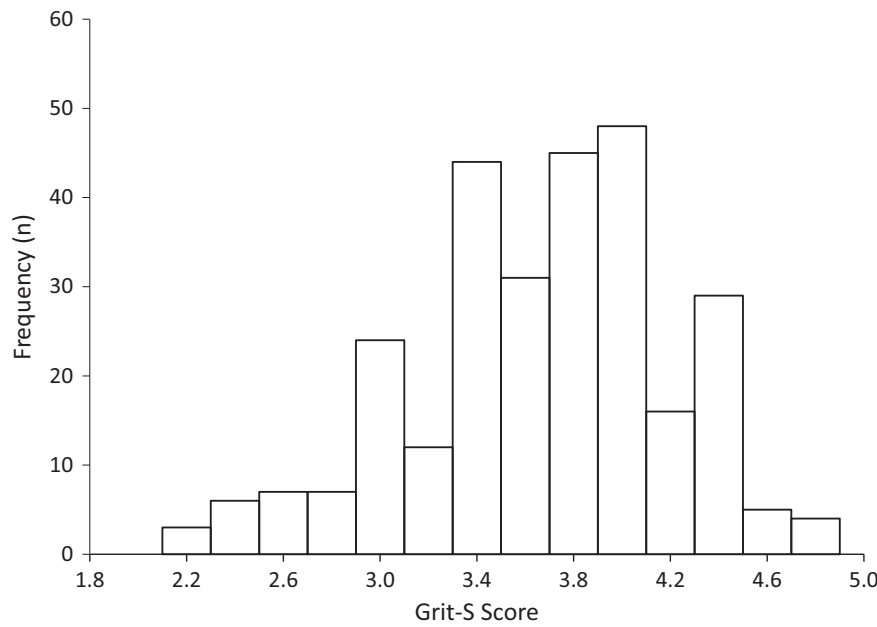
Figure 1. Resident self-assessed Grit-S score. Grit-S = Grit Scale.

Table 2
Grit-S Score by PGY

Grit-S Score	PGY-1 ($n = 81$)	PGY-2 ($n = 94$)	PGY-3 ($n = 88$)	PGY-4 ($n = 18$)	p-value
Self-assessed Grit-S	3.6 ± 0.54	3.58 ± 0.51	3.56 ± 0.55	3.62 ± 0.65	0.976
Faculty-assessed Grit-S	4.32 ± 0.97	3.45 ± 2.19	3.9 ± 1.56	4.12 ± 0.41	-

Data are reported as meant $\pm$ SD.

difficulty predicting their actual performance.[30–32] Furthermore, observer reports of noncognitive traits may be better predictors of actual performance than self-assessments.[32]

In 2007, Duckworth et al.[11] introduced Grit as a personality level trait that measures perseverance and passion for long-term goals. Grittier individuals, controlled for natural talent (IQ), have higher educational attainment, higher GPA, and fewer career changes and are less likely to drop out of a rigorous military training program.[11,19] Grit is also important in medicine as studies have shown that surgical residents with lower Grit-S scores have higher burnout rates and were more likely to leave their training program.[16,17]

While studies have examined personality and medical education, to the best of our knowledge, no study to date has compared faculty-assessed Grit-S scores to resident self-assessed Grit-S scores. While Grit-S scores for an individual can be predicted by family members and close peers,[19] it is not clear how well faculty can assess personality traits such as grit in their residents. The evidence as to whether supervising faculty can accurately assess different aspects of residents is mixed. Pediatric faculty demonstrated correlation between their assessment of trainee medical knowledge and the performance on the pediatrics in-training examination (ITE).[33] However, internal medicine and EM faculty were not able to accurately predict ITE scores for their trainees.[34,35] Similarly, a recent study showed that EM faculty were poor at predicting burnout in their trainees.[36] In addition to non-observable traits, faculty also have difficulty providing assessments of observed clinical skills for resident physicians.[1]

In our study, there was no correlation between the resident-assessed and faculty-assessed Grit-S scores. This corroborates the challenges that faculty have at accurately assessing aspects of residents they are supervising not only in cognitive, but also noncognitive, domains.[1,34,36] There are multiple reasons that we believe this may be occurring. First, there may be a lack of familiarity and decreasing observation time of faculty spent interacting with residents, secondary to faculty's nonclinical academic responsibilities, pressure for increased patient flow, and lack of dedicated
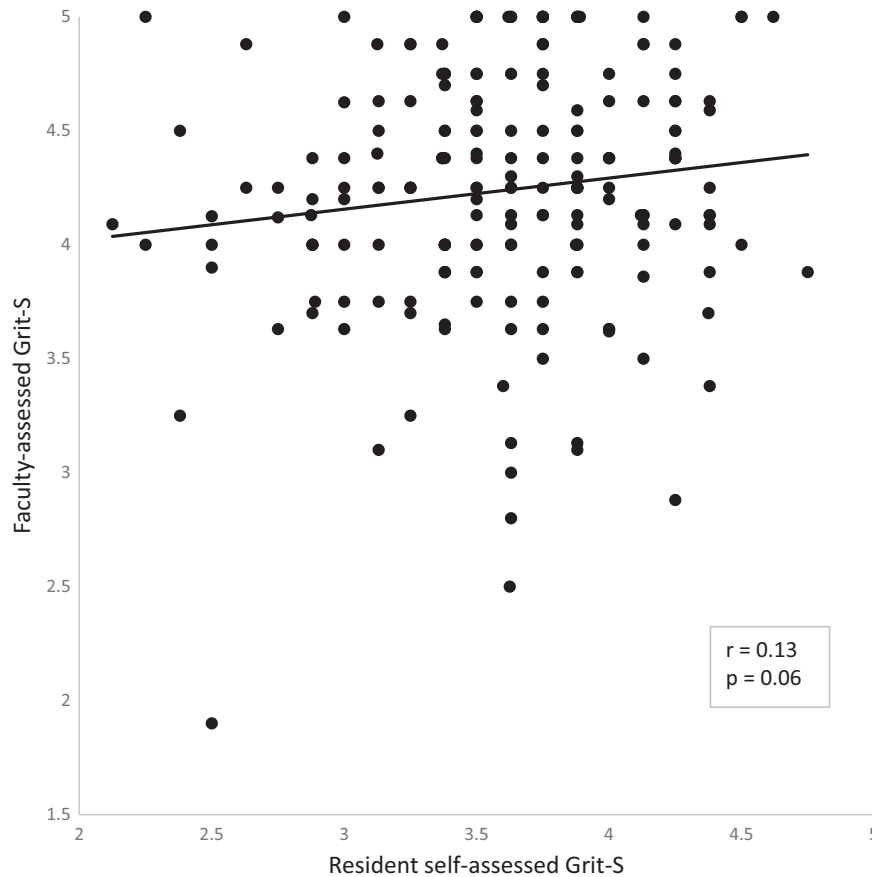
Figure 2. Correlation of residents' self-assessed Grit-S score and the residents' faculty-assessed Grit-S score. Grit-S = Grit Scale.

teaching time.[37,38] While the residents self-selected faculty evaluators, it is likely that faculty do not know residents nearly as well as the family members or close peers in the validation study of the Grit-S informant report. It is unclear from our study how knowing residents in different capacities may have influenced the faculty-assessed Grit score. We did not, for example, differentiate how faculty knew the resident, whether it was from clinical shifts, from completing research projects with them, from mentorship, as program leadership, or socially.

There was no correlation between the resident-assessed and the faculty-assessed Grit-S score; the mean Grit-S Score for residents was 3.58 and the mean faculty-assessed Grit-S score was 4.22. At all sites, faculty scored residents higher than residents scored themselves, with 84% of residents having lower self-assessed Grit-S scores. This is similar to previous findings of faculty prediction of resident attributes and achievements. Faculty overestimated the ITE scores of EM residents and underestimated the rate of burnout in EM residents.[34,36] This suggests that faculty may view their residents more favorably than their actual

performance or self-assessments would suggest, whether that is within cognitive or noncognitive domains. There may be a number of reasons for this trend of faculty overestimation of trainees. Faculty likely think very highly about their residents, either because they were involved in recruiting them to the program or because they are their mentors. Because of these relationships, faculty may be self-projecting onto their residents and assessing their Grit more highly because they want them to be successful. Another reason for higher faculty-assessed Grit-S scores could be a generosity error (or bias), which typically make ratings more favorable than other data would suggest they should be.[39] Residents, on the other hand, are often burned out and overwhelmed during residency so they may be more critical of themselves than their faculty are, accounting for the lower resident-assessed Grit-S scores.[36]

LIMITATIONS

There are several important limitations of this study. Seventy percent of EM residents who participated in
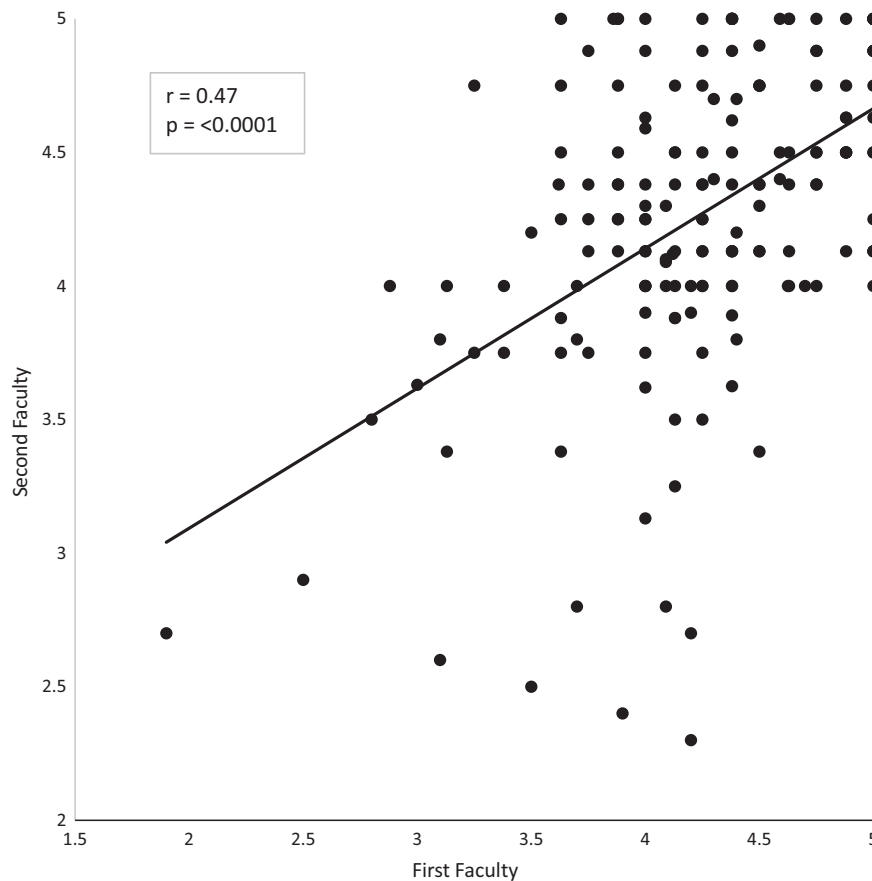
Figure 3. Correlation of faculty members' prediction of the same resident's Grit-S Score. Grit-S = Grit Scale.

the study were males and 30% were females. The proportion of females in our study was lower than the national average of 47% for female EM residents.[40] The majority of our participants were male (70%), whereas in the validation study by Duckworth and Patrick[19] of the informant report of the Grit-S, the majority of participants were female (89%). This difference in the sex of participants may have affected the results of our study. Twenty-six percent of residents self-identified as underrepresented minorities, which is higher than the 14% of U.S. EM residents self-identifying as underrepresented minorities.[41]

Residents selected the specific faculty whom they believed would best be able to assess their Grit. In the original Grit-S informant report study, the selected individuals were family and friends who likely knew the study subjects over many years and different types of experiences giving them more insight into their passion and perseverance for long-term goals. Program leadership (program directors, assistant/associate program directors, etc.) or more experienced faculty may be more accurate assessors of noncognitive traits, but this was not measured in this study. However, many

subjects did select at least one member of program leadership to assess them in this study. It is likely that a certain amount of interaction and observation time is necessary to accurately assess another person's Grit. Thus, it may be true that assessments between faculty and senior residents are more accurate than juniors; however, a subgroup analysis by ANOVA of faculty-assessed Grit-S scores of residents by PGY level was not possible due to the small sample size in each PGY. Due to the complexity of EM shifts, rotations, and faculty availability, we also lacked the means to explicitly track the amount of interaction a faculty participant had with a resident subject. Since approximately one-third of the residents did not have a faculty-assessed Grit-S score, it is also possible that there was a selection bias toward residents thought of more highly by the faculty.

The Grit-S score has been shown to have predictive value in other disciplines and correlated with burnout and attrition rates in surgical residents but has yet to be studied in depth in medicine. Thus, regardless of predictive accuracy of faculty, more work needs to be done to understand the value of Grit in physicians. If

the correlation with burnout seen in surgical residents is replicated in further assessments of Grit and other medical specialties, it may be a useful tool in physician training programs.[16]

CONCLUSION

Our study was unable to demonstrate a correlation between emergency medicine resident self-assessed and faculty-assessed Grit-S scores; however, we believe that noncognitive affective traits like Grit should be explored further. Information gleaned from Grit-S scores may help identify learner needs and future educational courses and career development opportunities. Since Grit has been correlated with surgery residents' burnout and attrition rate,[16,17] further study is needed to determine if a self-assessed Grit-S score is predictive or revealing of EM residents, as well as how to better utilize noncognitive assessments to improve resident training and education.

References

1. Kogan JR, Conforti LN, Iobst WF, Holmboe ES. Reconceptualizing variable rater assessments as both an educational and clinical care problem. Acad Med 2014;89:721–7.
2. Shih AF, Maroongroge S. The importance of grit in medical training. J Grad Med Educ 2017;9:399.
3. Bridgeman B, McCamley-Jenkins L, Ervin N. Predictions of freshmen grade-point average from the revised and recentered Sat I: Reasoning test. ETS Research Report Series 2000;2000(1):i–16.
4. Sutton E, Richardson JD, Ziegler C, Bond J, Burke-Poole M, McMasters KM. Is USMLE step 1 score a valid predictor of success in surgical residency? Am J Surg 2014;208:1029–34.
5. McCaskill QE, Kirk JJ, Barata DM, Wludyka PS, Zenni EA, Chiu TT. USMLE step 1 scores as a significant predictor of future board passage in pediatrics. Ambul Pediatr 2007;7:192–5.
6. Armstrong A, Alvero R, Nielsen P, Do US, et al. medical licensure examination step 1 scores correlate with council on resident education in obstetrics and gynecology in-training examination scores and American board of obstetrics and gynecology written examination performance? Mil Med 2007;172:640–3.
7. Harmouche E, Goyal N, Pinawin A, Nagarwala J, Bhat R. USMLE scores predict success in ABEM initial certification: a multi-center study. West J Emerg Med 2017;18:544–9.
8. Doherty EM, Nugent E. Personality factors and medical training: a review of the literature. Med Educ 2011;45:132–40.
9. Lalec JF, Torsher LC, Wiegmann DA, Arnold JJ, Brown DA, Phatak V. The Mayo high performance teamwork scale: reliability and validity for evaluating key resource management skills. Simul Healthc 2007;2:4–10.
10. Hojat M, Gonnella JS, Nasca TJ, Mangione S, Veloksi JJ, Magee M. The Jefferson Scale of Physician Empathy: further psychometric data and differences by gender and specialty at item level. Acad Med 2002;77(10 Suppl):S58–60.
11. Duckworth AL, Peterson C, Matthews MD, Kelly DR. Grit: perseverance and passion for long-term goals. J Pers Soc Psychol 2007;92:1087–101.
12. Duckworth A, Gross JJ. Self-control and grit: related but separable determinants of success. Curr Dir Psychol Sci 2014;23:319–25.
13. Eskreis-Winkler L, Shulman EP, Beal SA, Duckworth AL. The grit effect: predicting retention in the military, the workplace, school, and marriage. Front Psychol 2014;5:1–12.
14. Robertson-Kraft C, Duckworth AL. True grit: trait-level perseverance and passion for long-term goals predicts effectiveness and retention among novice teachers. Teach Coll Rec 2014;116:1–27.
15. Duckworth A, Kirby TA, Tsukayama E, Berstein H, Ericsson KA. Deliberate practice spells success: why grittier competitors triumph at the national spelling bee. Soc Psychol Pers Sci 2010;2:174–81.
16. Burkhart RA, Tholey RM, Guinto D, Yeo CJ, Chojnacki KA. Grit: a marker of residents at risk for attrition? Surgery 2014;155:1014–22.
17. Salles A, Cohen GL, Mueller CM. The relationship between grit and resident well-being. Am J Surg 2014;207:251–4.
18. Palisoc AJL, Matsumoto RR, Ho J, Perry PJ, Tang TT, Ip EJ. Relationship between grit with academic performance and attainment of postgraduate training in pharmacy students. Am J Pharm Educ 2017;81:67.
19. Duckworth AL, Patrick DQ. Development and validation of the short grit scale (Grit-S). J Pers Asses 2010;91:166–74.
20. Rosenman R, Tennekoon V, Hill LG. Measuring bias in self-reported data. Int J Behav Healthc Res 2011;2:320–22.
21. Balsis S, Cooper LD, Oltmanns TF. Are informant reports of personality more internally consistent than self reports of personality? Assessment 2016;22:399–404.
22. Capriccio C, Wolfsthal SD, Englander R, Ferentz K, Martin C. Shifting paradigms: from flexner to competencies. Acad Med 2002;77:361–7.
23. Iobst WF, Sherbino J, Cate OT, et al. Competency-based medical education in postgraduate medical education. Med Teach 2010;32:651–6.

24. Holmboe ES, Sherbino J, Long DM, Swing SR, Frank JR. The role of assessment in competency-based medical education. Med Teach 2010;32:676–82.

25. Del Canale S, Louis DZ, Maio V, et al. The relationship between physicians empathy and disease complications: an empirical study of primary care physicians and their diabetic patients in Parma. Italy. Acad Med 2012;87:1243–9.

26. Gonnela JS, Hojat M, Erdmann JB, Veloski JJ. What have we learned and where do we go from here? Acad Med 1993;68(2 Suppl):S79–87.

27. Ferguson E, James D, O'Hehir F, Sanders A, McManus IC. Pilot study of the roles of personality, references, and personal statements in relation to performance over the five years of a medical degree. BMJ 2003;326:429–32.

28. Lievens F, Coetsier P, De Fruyt F, De Maeseneer J. Medical students' personality, characteristics and academic performance: a five-factor model perspective. Med Educ 2002;36:1050–6.

29. Lievens F, Ones DS, Dilchert S. Personality scale validities increase throughout medical school. J Appl Psychol 2009;94:1514–35.

30. Oh IS, Wang G, Mount MK. Validity of observer ratings of five-factor model of personality traits: a meta-analysis. J Appl Psychol 2011;96:762–73.

31. Duckworth AL, Yeager DS. Measurement matters: assessing personal qualities other than cognitive ability for educational purposes. Educ Res 2015;44:237–51.

32. Dunning D, Heath C, Suls JM. Flawed self-assessment: implications for health, education, and the workplace. Psychol Sci Public Interest 2004;5:69–106.

33. David JK, Inamdar S, Stone RK. Interrater agreement and predictive validity of faculty ratings of pediatric residents. J Med Educ 1986;61:901–5.

34. Aldeen AZ, Quattromani EN, Williamson K, Hartman ND, Wheaton NB, Branzetti JB. Faculty prediction of in-training examination scores of emergency medicine residents: a multicenter study. J Emerg Med 2015;49:64–9.

35. Hawkins RE, Sumption KF, Gaglione MM, Holmboe ES. The in-training examination in internal medicine: resident perceptions and lack of correlation between resident scores and faculty prediction of resident performance. Am J Med 1999;106:206–10.

36. Lu DW, Lank PM, Branzetti JB. Emergency medicine faculty are poor at predicting burnout in individual trainees: an exploratory study. AEM Educ Train 2017;1:75–8.

37. Albanese MA. Challenges in using rater judgments in medical education. J Eval Clin Pract 2000;6:305–19.

38. Ramani S, Orlander JD, Strunin L, Barber TW. Whither bedside teaching? A focus-group study of clinical teachers. Acad Med 2003;78:384–90.

39. Mehrens W, Lehmann I. Measurement and Evaluation in Education and Psychology, 3rd ed. New York, NY: Holt, Rinehart and Winston, 1984.

40. American Association of Medical Colleges. Residency Applicants to ACGME-accredited Programs by Specialty and Sex, 2017-2018. Available at: https://www.aamc.org/download/321558/data/factstablec1.pdf. Accessed January 1, 2018.

41. Landry AM, Stevens J, Kelly SP, Sanchez LD, Fischer J. Under-represented minorities in emergency medicine. J Emerg Med 2013;45:100–4.

Supporting Information

The following supporting information is available in the online version of this paper available at http://onlinelibrary.wiley.com/doi/10.1002/aet2.10309/full

Data Supplement S1. Grit-S Scale.