The relationship of Grit to faculty evaluations and standardized test scores of anesthesia residents: a pilot study

Ryan Fink[1], Catherine Kuhn[2], N. David Yanez[1], Jeffrey Taekman[2]

Corresponding author: Dr Ryan Fink finkr@ohsu.edu
Institution: 1. Oregon Health and Science University, 2. Duke University Hospital
Categories: Assessment, Learning Outcomes/Competency

Received: 05/10/2018
Published: 04/12/2018

Abstract

Background: Grit, defined as perseverance and passion for long-term goals, is associated with high levels of success in multiple domains but has only been minimally studied in the medical fields.

Aims: This study examined whether higher levels of grit are associated with measures of anesthesiology resident performance – standardized test scores and faculty evaluations.

Methods: 23 graduated anesthesiology residents volunteered to participate in the study. Participants completed the Grit Survey via an online interface. The program director provided performance data to a neutral data manager, who de-identified the data prior to analysis.

Results: Correlation estimates between standardized test scores and grit ranged between -0.12 and 0.31 (not statistically significant at the one percent level of significance; p values ranging from 0.16 to 0.85). Correlation estimates between PGY-4 faculty evaluations and grit scores ranged from 0.43 to 0.49, but were not statistically significant. There may be a trend towards significance, as p-values ranged from 0.021 to 0.044.

Conclusion: This pilot study showed a trend toward higher grit being associated with higher overall faculty evaluation of resident performance in their final year of training. This finding was non-significant and thus should be considered preliminary. There was no significant association between grit and performance on standardized tests.

Keywords: Grit; Performance; Recruitment; Personality tests
Introduction

The ability to predict which residency applicants will be successful resident trainees, and eventually consultant anesthesiologists, is an area of active interest and research, especially for program directors. Residency programs compete for top resident applicants not only to provide excellent clinical care to patients, but also to enhance a program's reputation, which can have a positive impact on a program's ability to recruit top resident physicians in the future. There are economic considerations as well, as departments may utilize significant resources in money, time, and energy (Metro et al. 2005) to recruit successful residents; not to mention the investment in training once residents arrive to a program. Consequently, there is interest within academic training programs to identify the personal attributes or other pre-residency characteristics that may be predictive of high-level performance and success during training. Unfortunately for program directors and selection committees, predicting which applicants will become high-performing residents has proven challenging.

Performance on standardized tests is used extensively for screening applicants to residency programs (Boyse et al. 2002; Dirschl et al. 2006; Brothers and Wetherholt 2007; Tolan et al. 2010; Results of the 2014 NRMP Program Director Survey 2014) and results from standardized tests such as the United States Medical Licensing Examination (USMLE, or Comprehensive Osteopathic Medical Licensing Examination [COMLEX]) provide some of the few pieces of residency application data that are standardized amongst applicants. Many studies have found that USMLE scores are predictive of performance on subsequent in-training exam (ITE) and board certification tests (Bell et al. 2002; Boyse et al. 2002; Dirschl et al. 2006; McCaskill et al. 2007; Shellito et al. 2010; Guffey et al. 2011). Additionally, a meta-analysis of eighty studies confirmed that a resident selection strategy based on standardized examinations is most strongly correlated with examination-based outcomes (Kenny et al. 2013). However, achievement on pre-residency standardized exams does not reliably correlate with clinical performance (Bell et al. 2002; Boyse et al. 2002; Dirschl et al. 2006; Brothers and Wetherholt 2007; Tolan et al. 2010). In response to this dichotomy, at least one group has submitted a “plea” to training programs to stop using the USMLE as a screening tool for residency (Prober et al. 2016).

A multitude of other factors have been studied in the quest to find a pre-residency variable (or combination of variables) that may be predictive of clinical and/or overall performance – not just performance on standardized exams. These factors often focus on “non-cognitive factors” (Farrington et al. 2012) (sometimes now called “soft skills,” metacognitive learning skills,” or “agency”) or “emotional intelligence” (Cherniss et al. 1998), especially empathy, integrity, consciousness, emotional stability, tolerance, and communication skills. Commonly used approaches to assess these personality characteristics include face-to-face interviews (George et al. 1989; Altmaier et al. 1992; Metro et al. 2005; Dubovsky et al. 2008; Alterman et al. 2011), the Medical School Performance Evaluation (MSPE), and letters of recommendation (LOR) (Harfmann and Zirwas 2011; Kenny et al. 2013), but studies on their ability to predict future performance is mixed. Personality testing/surveying has also been studied as a possible tool for selecting applicants with a high likelihood of success during training (Gough et al. 1991; McDonald et al. 1994; Merlo and Mateevskii 2009; Lubelski et al. 2016), and positive correlations have been found in certain sub-scales of personality.

Duckworth et al aimed to define a personality trait that would be specific to, and consistent with, high levels of success in any domain (Duckworth A. L. et al. 2007; Duckworth A. L. and Quinn 2009). This group conducted a series of investigations that suggest grit, defined as “perseverance and passion for long term goals” (Duckworth A. L. et al. 2007), is essential to success. Grit requires an individual to work “strenuously toward challenges, maintaining effort and interest over years despite failure, adversity, and plateaus in progress” (Duckworth A. L. et al. 2007). Duckworth's validated Grit Scale and Short Grit Scale have been used to understand the characteristics of success in multiple arenas including college students, West Point military cadets, Scripps National Spelling Bee finalists, and
teachers (Duckworth A. L. et al. 2007; Duckworth A. L. and Quinn 2009; Duckworth A.L. et al. 2009; Eskreis-Winkler et al. 2014). In each of these domains, individuals with higher grit scores had higher levels of success, even when compared to other common predictors such as intelligence or the Big Five personality traits (i.e., openness to experience, conscientiousness, extraversion, agreeableness, and neuroticism).

As noted, the scientific literature is mixed regarding the ability of a wide array of pre-residency variables to predict success during residency training. Grit appears to have the ability to predict high levels of success, even among already high-achieving individuals. To assess whether grit is associated with better performance in residency training, our group conducted this pilot study of previous anesthesiology residents. We hypothesized that higher grit scores would be positively correlated with two markers of success in residency: standardized test scores and faculty evaluations of clinical performance.

**Methods**

The institutional review board (IRB) approved this study. An invitation email was sent to anesthesiologists who graduated from the Duke anesthesiology residency program between the years 2009 and 2012. Recently graduated residents were studied because all performance outcome variables would be complete by nature of having finished the program. Potential subjects were directed to a website to consent and then complete the “Short Grit Survey” (Grit-S) tool (Duckworth A. L. and Quinn 2009). The short version (6 questions, Figure 1) of the Grit Survey was used because it is shorter and thus may be perceived as less burdensome to the subjects to participate. The Grit-S was shown to have internal consistency, test–retest consistency, and predictive validity, similar to the longer, original Grit Survey (Duckworth A. L. and Quinn 2009). In other studies of grit in residency (Burkhart et al. 2014; Halliday et al. 2017; Salles et al. 2017), the Short Grit Scale was used, but it is important to note that we used the original 6-item Grit-S, whereas these other studies used the revised 8-item Grit-S. De-identification of the data was performed such that the program director was not aware of the Grit Survey results.

**Figure 1: The Short Grit Scale**
Those who chose to participate consented to have specific data from their educational record included in the study. Data collected from consenting participants' files included results from the pre-residency USMLE Step 1, Step 2 Clinical Knowledge, and intra-residency Step 3, reported as the numerical score. Results were also collected from the Anesthesia Knowledge Test (AKT), an anesthesiology-specific, intra-residency exam administered at 0, 1, 6 and 24 months of training. AKT results were recorded as the percentile based on the national average for the year of the test. Reporting of AKT24 required some additional normalization. Results of AKT24 are reported to programs in 7 sub-sections. For each participant, an “AKT 24 Overall Percentile” was calculated by averaging the 7 sub-section percentiles based on the national average.

In addition to standardized test scores, we also examined faculty ratings of residents. Resident clinical performance was assessed by faculty evaluation of ACGME Core Competencies (on a scale from 0 to 9). The average score for each core competency was calculated for each Post-Graduate Year (PGY) 2 through 4 (the years of clinical anesthesia training).

To protect participant confidentiality of sensitive academic information, a neutral-party data manager completed matching of grit scores to performance data such that a de-identified data set (lacking participant name/identifying information) was sent to the study team for analysis. The primary study end-points were the correlation of grit score to standardized test results (USMLE and AKT), and grit score to faculty evaluation of clinical performance as assessed by the ACGME Core Competencies.

SAS (Cary, North Carolina) and R (ver. 3.3.3) were used for statistical tests. Descriptive statistics were calculated for grit score, standardized test scores, and faculty evaluations. Pearson correlation statistics were used to assess the relationship between the standardized test scores and grit scores, and faculty evaluations and grit scores. We used
Pearson correlations to summarize the strength of a linear association (or trend) between our variables. The p-values are not adjusted for multiple comparisons. In addition, we included locally weighted scatterplot smooth (LOWESS) curves in our figures. These curves do not assume a linear association between variables, but computes a smooth curve that allows for easier visual evaluation of the type of association between the variables.

**Results/Analysis**

Forty-eight previous residents were contacted and 23 elected to participate, for a response rate of 48%. One participant was missing key information, thus some measures only included data on 22 participants. Demographic information of study subjects is shown in Table 1.

**Table 1: Demographic information of study participants**

| Class of 2009 | n  | Male/Female, n |
|---------------|----|----------------|
|               | 2  | 2/0            |
| Class of 2010 | 5  | 2/3            |
| Class of 2011 | 8  | 7/1            |
| Class of 2012 | 8  | 4/4            |
| **Total sample** | **23** | **15/8** |

Sixty-one percent of participants were male. The mean grit score was 3.52 ± 0.65 (range, 2.0 – 4.33). As a comparison, a grit score of 3.5 is at the 40th percentile of a large sample of adults in America (Duckworth A. L. 2016). In other studies of resident trainees the mean grit scores were: 3.61 – 3.67 (Halliday et al. 2017), 3.64 (Salles et al. 2014), and 3.87 (Burkhart et al. 2014).

Descriptive statistics for the standardized test scores are shown in Table 2.

**Table 2: Standardized test scores and simple statistics**

| USMLE reported as numerical score. AKT results reported as percentile based on the national average. AKT 24 Overall Avg %ile is the mean of the average percentile score of all 7 sub-scores of the AKT 24. Std Dev = standard deviation. |
|-----------------|-----------------|-----------------|-----------------|-----------------|
| **Mean**        | **Std Dev**     | **Min**         | **Max**         |
| USMLE 1 (score) | 237.2           | 12.3            | 211             | 264             |
| USMLE 2 CK (score) | 239.0         | 12.0            | 214             | 262             |
| USMLE 3 (score) | 229.0           | 9.6             | 206             | 246             |
| AKT 0           | 71.5            | 21.8            | 26              | 99              |
| AKT 1           | 71.3            | 19.6            | 26              | 98              |
| AKT 6           | 70.8            | 18.3            | 37              | 98              |
| **AKT 24 – Overall Average %** | **65.2** | **20.5** | **41** | **83** |

AKT results are presented as a percentile based on the national average. Thus, the small decrease in mean AKT24 compared to the other AKT results could signify either a lower average score (signifying the learners performing worse on the exam) or could represent an increase in average scores of other national test-takers such that the subject’s mean percentile was lower for a similar exam performance. However, the results are not extreme and could also be the result of chance. Correlation estimates between the standardized test scores and grit scores ranged between -0.12 and 0.31 (Table 3 and Figure 2). The associations, however, were not statistically significant at the one percent level of significance (p values ranging from 0.16 to 0.85).
Table 3: Pearson correlation coefficients for grit and standardized test scores

| Test          | n  | Pearson   | p    |
|---------------|----|-----------|------|
| USMLE 1       | 23 | 0.28798   | 0.1827|
| USMLE 2 CK    | 23 | 0.14654   | 0.5047|
| USMLE 3       | 23 | 0.11535   | 0.6002|
| AKT 0         | 22 | -0.11632  | 0.6062|
| AKT 1         | 22 | -0.015833 | 0.4816|
| AKT 6         | 22 | 0.31173   | 0.1579|
| AKT 24        | 23 | 0.04240   | 0.8477|

Figure 2: Correlations of grit and standardized test scores. Each dot represents a study participant.

For each Core Competency in the PGY-4 year, the final year of anesthesia residency, there was a trend towards a positive correlation between PGY4 faculty evaluation score and grit score (Table 4 and Figure 3). Correlation coefficients ranged from 0.43 to 0.49, and the characteristics were marginally associated. The p-values for the correlation between PGY-4 faculty evaluations and grit ranged from 0.021 to 0.044. There were no statistically significant associations with PGY-2 and PGY-3 faculty evaluations and grit score (data not shown).
### Table 4: Pearson correlation coefficients for grit and PGY-4 faculty evaluations

| Faculty Evaluation               | Pearson | p    |
|----------------------------------|---------|------|
| Patient Care                     | 0.47    | 0.027|
| Medical Knowledge                | 0.45    | 0.037|
| Practice-based Learning          | 0.44    | 0.041|
| Interpersonal/Communication      | 0.43    | 0.044|
| Professionalism                  | 0.49    | 0.021|
| Systems-based Practice           | 0.48    | 0.023|

### Figure 3: Correlation of grit score with PGY-4 ACGME Core Competency, average faculty evaluation score. Each dot represents a study participant.

### Discussion

As previously noted, the scientific literature on success in residency training is quite clear – performance on standardized tests may predict performance on future standardized exams but does not reliably predict **clinical** performance. “Non-cognitive” characteristics, especially empathy, integrity, tolerance, and communication skills are essential to the successful practice of medicine (Allen et al. 1997; Marley and Carman 1999; Lambe and Bristow
2010). While these attributes are often assessed by subjective means and thus may be more difficult to standardize and study, the assessment of an applicant's non-cognitive attributes may hold promise as a predictive tool for success compared to cognitive factors (such as examination scores or medical school grades). Studies investigating “non-cognitive skill”-focused or “behavioral interview” techniques show correlations to future resident performance (Altmaier et al. 1992; Olawaiye et al. 2006; Brothers and Wetherholt 2007; Strand et al. 2011). For example, in a study of a surgical residency program, ratings of applicant's “Personal Characteristics,” including attitude, motivation, integrity, and interpersonal skills, were strongly correlated with resident clinical performance as assessed by faculty, while cognitive measures (medical school GPA and USMLE scores) were negatively correlated (Brothers and Wetherholt 2007). An interview selection process for an Obstetrics/Gynecology training program that was heavily weighted toward the interview and non-cognitive qualities showed a positive correlation with subsequent resident clinical performance (Olawaiye et al. 2006). The Multiple Mini Interview (MMI) (Eva et al. 2004) is probably the most well-known standardized technique for trying to assess the non-cognitive, or “soft,” skills of applicants, and is often used in medical school admission processes. In studies looking at predicting post-graduate performance, the MMI may hold promise, but the literature is still in its infancy (Hofmeister et al. 2009; Dore et al. 2010; Burkhardt et al. 2015; Sklar et al. 2015).

Personality testing is another area of applicant or trainee assessment that has been gaining popularity. In reviews of the academic performance literature, “conscientiousness” – often described as how a person controls, regulates, and directs their impulses, and focused on organization, thoroughness, and reliability - is a personality trait consistently found to be associated with academic performance, both in general (Poropat 2009) and in medical school, specifically (Doherty and Nugent 2011). For residency training, multiple investigators have found that extensive personality testing may yield positive correlations to performance. For example the group of Gough and McDonald (Gough et al. 1991; McDonald et al. 1994) used the 462-item California Psychological Inventory and the > 200 question Strong Interest Inventory to study anesthesiology residents and found that certain sub-scales of personality were associated with high-performing residents. Merlo et al. (Merlo and Matveevskii, 2009) used the 300-item International Personality Item Pool Representation (IPIP-NEO) and found that personality traits such as confidence and conscientiousness were associated with success in anesthesiology residents. Lubelski et al. (Lubelski et al. 2016) used the 574-item Hogan Personality assessments in neurosurgery residents, and while they did not specifically study markers of performance, they concluded that personality testing may help predict future resident behavior.

The Grit Scale is a personality survey that focuses on the propensity of an individual to doggedly pursue goals they find valuable over the long-term, despite setbacks; grit is an assessment of endurance and perseverance (Duckworth A. L. et al. 2007). In stark contrast to the previously mentioned personality assessments, the Grit Survey is short – only 8 questions in the current version. Notwithstanding the survey's short length, higher grit has been found to be associated with greater success in many domains (Duckworth A. L. et al. 2007; Duckworth A. L. and Quinn 2009; Eskreis-Winkler et al. 2014). For example, individuals with higher average grit scores obtain more education and tend to have higher GPAs (Duckworth A. L. and Quinn 2009). Of competitors in the National Spelling Bee, children with higher grit progressed further in the competition; military cadets with higher grit scores were more likely to finish summer training or complete a Special Operations course; grittier men stayed married longer; employees with more grit kept their jobs longer and had fewer career changes; and grittier teachers were more effective at their jobs (Duckworth A. L. et al. 2007; Duckworth A. L. and Quinn 2009; Duckworth A.L. et al. 2009; Eskreis-Winkler et al. 2014). While a variety of personality characteristics (and other factors) may interact to mediate behaviors associated with success – like self-control/self-discipline (Duckworth A. L. and Seligman 2005), need for achievement, and self-efficacy – the relationship between conscientiousness and success seems to have the most support in the medical literature (Merlo and Matveevskii 2009; Doherty and Nugent 2011). Conscientiousness is likely closely related to grit, but at least in some sub-groups of people studied - US Military Academy training retention (Duckworth A. L. et al. 2007) and children in the National Spelling Bee (Duckworth A. L. and Quinn 2009) - the Grit Survey shows
predictive validity over measures of conscientiousness.

Using the Grit Survey in a pilot study of anesthesiology residents, we found a trend toward a positive correlation between increasing grit and final-year (PGY-4) faculty evaluation of residents’ ACGME Core Competencies. There was no significant association between grit and PGY-2 and PGY-3 faculty evaluations of core competencies. Faculty evaluations tend to encompass a wide view of the individual resident and their characteristics and abilities, including technical skills, knowledge, judgment, empathy, and professionalism. In this context, it seems reasonable that grit, which encompasses many aspects of personality that relate to success (motivation, endurance, perseverance, and passion), may be likely to correlate with residency performance evaluations by faculty. While this is a preliminary finding and was not statistically significant in our study, future studies with more power to detect differences are ongoing to assess if grit is associated with resident performance. In addition, it may be possible to assess for grit in the pre-residency time period and use this as a residency selection tool to help predict which future trainees may perform well. This will have to be evaluated in future studies.

The fact that only the final year’s evaluations showed a trend toward correlation with grit could be explained in multiple ways. It could be that years of exposure to a trainee are needed for faculty evaluators to gain a comprehensive appraisal of resident performance. This finding could also mean that these gritty characteristics are most notable in the final year of training, when most senior residents are given greater responsibility and autonomy. Finally, this could also mean that residents became grittier as they progressed through training. While it does appear that grit increases with age (on the order of decades) (Duckworth A. L. et al. 2007; Duckworth A. L. 2016), over shorter periods of time (year-to-year) grit appears to be stable (Duckworth A. L. and Quinn 2009). In a study of surgical residents, Salles, et al. found grit to be stable over multiple years of training (Salles et al. 2014), but in our sample Grit Scores were collected one to four years after training was complete, thus we do not know if grit changed either during residency or after.

One of the ultimate goals of this line of study is to assess if grit may be able to help predict which residents will do well in a training program. In the current literature on grit and residency training, there is some evidence that the Grit Survey may be used for this purpose, and specifically that low grit may be a marker for attrition. Burkhart, et al. studied general surgery trainee grit in relation to attrition and found that those with below median grit were twice as likely to consider leaving a training program (Burkhart et al. 2014). Of the three trainees who did leave their training program, all of them had below median grit (but the difference was non-significant due to the low attrition rate) (Burkhart et al. 2014). Another study in general surgery residents had similar findings – those with higher grit were less likely to consider leaving a program, but the true attrition rate was low and there was no significant correlation between leaving a program and grit (Salles et al. 2017). While higher grit has been found to be associated with higher levels of success in other fields, low grit may also be a predictor of poor resident performance or attrition (although attrition does not always result from poor performance).

We found that grit did not correlate with standardized exam performance taken either during medical school (the USMLE) or during residency (AKT). This is consistent with another study of grit in the medical field that showed no correlation between the grit of surgical residents and performance on the American Board of Surgery In-Training Examination (ABSITE) (Burkhart et al. 2014). In addition, having more grit is not necessarily indicative of having stronger “cognitive” abilities – usually identified with intelligence and the ability to solve abstract problems (as often measured by the IQ test and standardized tests) (Brunello and Schlotter 2011) – and in fact grit predicts success more strongly than measures of intelligence (Duckworth A. L. et al. 2007; Duckworth A. L. and Quinn 2009). Given that overall resident performance encompasses many more aspects than merely cognitive skills like knowledge acquisition, it seems plausible that grit may correlate with measures of overall performance (such as faculty evaluations) and not test scores.
This study has several limitations. In this pilot study, a primary limitation was that we had a relatively small sample size and thus, a low level of power to detect significant associations. And given the multiple comparisons, we used the more rigorous p-value cut-off of 0.01 to indicate statistical significance. As participation in this study was voluntary, selection bias is of concern - grit scores and residency performance may be different between those who chose to participate versus those that did not. We did not collect performance data of those not participating, thus we cannot know if performance differed between participants and non-participants.

The self-report nature of the grit questionnaire makes it susceptible to social desirability bias (the desire to “look good” (Gnambs and Kaspar 2017)), and questions on the Grit Survey are transparent. We attempted to minimize this effect by emphasizing the de-identified nature of the study with multiple safeguards in place to protect participants’ privacy; however, it was up to each study participant to decide how honestly to answer each question. Despite these safeguards, other more objective measures of grit may be desirable. In one series of studies on grit it was found that family and friends could reliably assess a person’s grit (Duckworth A. L. and Quinn 2009), and grit as scored by a blind rater based on résumé data correlated with teacher effectiveness (Robertson-Kraft and Duckworth 2014). Therefore, in future studies it may be possible to assess grit in a manner independent of a self-report score, which would be essential if a grit score is to be used for applicant screening.

We cannot rule out the possibility that grit changed over time between when the study subjects were applicants versus when the in-residency exams, faculty evaluations, and assessment of grit occurred. While grit does increase with life experience and age, it seems to be relatively stable over short periods of time (Duckworth A. L. and Quinn 2009), including during residency training (Salles et al. 2014), and an ongoing study by our group will assess this same question.

Finally, faculty evaluations have well-known limitations (Holmboe 2004). We hoped that by averaging all faculty evaluations each year, multiple evaluations by many faculty members would be captured and represent an accurate picture of resident clinical performance. Most residency training programs have now transitioned to Milestone evaluations, and it’s unclear how these changes may affect future assessments of clinical performance and grit.

**Conclusion**

This pilot study showed a trend toward higher grit being associated with higher overall faculty evaluation of resident performance in their final year of training. This finding was non-significant and thus should be considered preliminary. There were no highly significant associations between grit and performance on standardized tests. Future larger scale studies are needed to confirm the associations of grit and overall performance in a larger cohort of trainees, to assess whether low grit can predict risk of poor performance, and to investigate if grit can be assessed by objectives means (i.e., not a self-report scale). Our ultimate goal is to find applicant characteristics predictive of future success in residency training and beyond.

**Take Home Messages**

- Predicting which residency applicants will be successful in a training program is challenging and often subjective, though personality testing may have value in predicting resident performance.
- Measuring Grit, which encompasses passion and perseverance, appears to be associated with high achievers in many domains, though studies in the medical field are few.
The results of this study did not show a correlation between grit and standardized test scores. While also not statistically significant, there may be a trend towards a relationship of grit and faculty evaluations.

Notes On Contributors

Ryan J. Fink, M.D., is Assistant Professor of Anesthesiology and Critical Care Medicine, and is the Associate Program Director for the Anesthesiology-Critical Care Fellowship Program at Oregon Health and Science University.

Catherine M. Kuhn, M.D. is Professor with Tenure in the Department of Anesthesiology, where she served as Program Director for the Core Residency program for nearly twenty years, and is now the Designated Institutional Official and Director of Graduate Medical Education for Duke University Hospital and Health System.

N. David Yanez, PhD, is Professor of Biostatistics in the OHSU-PSU School of Public Health, and is co-Director of the Biostatistics and Design Program. His research interests include measurement error models, generalized linear and quasi-likelihood models, longitudinal and correlated data methods, distribution-free tests, and cardiovascular disease research.

Jeffrey M. Taekman, M.D., is Professor of Anesthesiology at Duke University Hospital and Health System, and is Director of the Human Simulation and Patient Safety Center. His research focuses on the use of technology (primarily simulation) to improve learning and processes within healthcare systems.

Acknowledgements

We would like to thank Dr. Angela Duckworth and her team for their advice and the use of the Grit Scale. We would also like to thank the staff of the Duke Anesthesiology Education Office for their collection and input of data. Thank you to Marie Kane for assistance with reference formatting.

Bibliography/References

Allen, I., Brown, P. and Hughes, P. (1997) Choosing Tomorrow's Doctors, London: Policy Studies Institute.

Alterman, D. M., Jones, T. M., Heidel, R. E., Daley, B. J., et al. (2011) ‘The predictive value of general surgery application data for future resident performance’, J Surg Educ, 68(6), pp. 513-518. https://doi.org/10.1016/j.jsurg.2011.07.007

Altmaier, E. M., Smith, W. L., O'Halloran, C. M. and Franken, E. A., Jr. (1992) ‘The predictive utility of behavior-based interviewing compared with traditional interviewing in the selection of radiology residents’, Invest Radiol, 27(5), pp. 385-389. https://doi.org/10.1097/0000424-199205000-00013

Bell, J. G., Kanellitsas, I. and Shaffer, L. (2002) ‘Selection of obstetrics and gynecology residents on the basis of medical school performance’, Am J Obstet Gynecol, 186(5), pp. 1091-1094. https://doi.org/10.1067/mob.2002.121622

Boyse, T. D., Patterson, S. K., Cohan, R. H., Korobkin, M., et al. (2002) ‘Does medical school performance predict
radiology resident performance?', *Acad Radiol*, 9(4), pp. 437-445. [https://doi.org/10.1016/j.acra.2002.11.007](https://doi.org/10.1016/j.acra.2002.11.007)

Brothers, T. E. and Wetherholt, S. (2007) ‘Importance of the faculty interview during the resident application process’, *J Surg Educ*, 64(6), pp. 378-385. [https://doi.org/10.1016/j.jsurg.2007.05.003](https://doi.org/10.1016/j.jsurg.2007.05.003)

Brunello, G. and Schlotter, M. (2011) ‘Non Cognitive Skills and Personality Traits: Labour Market Relevance and their Development in Education & Training Systems’, Bonn, Germany: Institute for the Study of Labor.

Burkhardt, J. C., Stansfield, R. B., Vohra, T., Losman, E., *et al.* (2015) ‘Prognostic value of the Multiple Mini-Interview for emergency medicine residence performance’, *J Emer Med*, 49(2), pp. 196-202. [https://doi.org/10.1016/j.jemermed.2015.02.008](https://doi.org/10.1016/j.jemermed.2015.02.008)

Burkhart, R. A., Tholey, R. M., Guinto, D., Yeo, C. J., *et al.* (2014) ‘Grit: a marker of residents at risk for attrition?’, *Surgery*, 155(6), pp. 1014-1022. [https://doi.org/10.1016/j.surg.2014.01.015](https://doi.org/10.1016/j.surg.2014.01.015)

Cherniss, C., Goleman, D., Emmerling, R., Cowan, K., *et al.* (1998) *Bringing emotional intelligence to the workplace: A technical report issued by the Consortium for Research on Emotional Intelligence in Organizations*. Available at: [http://www.eiconsortium.org/pdf/technical_report](http://www.eiconsortium.org/pdf/technical_report)

Dirschl, D. R., Campion, E. R. and Gilliam K. (2006) ‘Resident selection and predictors of performance: can we be evidence based?’, *Clin Orthop Relat Res*, 449, pp. 44-49.

Doherty, E. M. and Nugent, E. (2011) ‘Personality factors and medical training: a review of the literature’, *Med Educ*, 45(2), pp. 132-140. [https://doi.org/10.1111/j.1365-2923.2010.03760.x](https://doi.org/10.1111/j.1365-2923.2010.03760.x)

Dore, K. L, Kreuger, S., Ladhani, M., Rolfson, D., *et al.* (2010) ‘The reliability and acceptability of the Multiple Mini-Interview as a selection instrument for postgraduate admissions’, *Acad Med*, 85(10 Suppl), S60-63. [https://doi.org/10.1097/ACM.0b013e3181ed442b](https://doi.org/10.1097/ACM.0b013e3181ed442b)

Dubovsky, S. L., Gendel, M. H., Dubovsky, A. N., Levin, R., *et al.* (2008) ‘Can admissions interviews predict performance in residency?’ *Acad Psychiatry*, 32(6), pp. 498-503. [https://doi.org/10.1176/appi.ap.32.6.498](https://doi.org/10.1176/appi.ap.32.6.498)

Duckworth, A. L. (2016) *Grit: The power of passion and perseverance*. New York, NY: Simon & Schuster, Inc.

Duckworth, A. L., Peterson, C., Matthews, M. D. and Kelly, D.R. (2007) ‘Grit: perseverance and passion for long-term goals’, *J Pers Soc Psychol*, 92(6), pp. 1087-1101. [https://doi.org/10.1037/0022-3514.92.6.1087](https://doi.org/10.1037/0022-3514.92.6.1087)

Duckworth, A. L. and Quinn, P. D. (2009) ‘Development and validation of the short grit scale (grit-s)’, *J Pers Assess*, 91(2), pp. 166-174. [https://doi.org/10.1080/00223890802634290](https://doi.org/10.1080/00223890802634290)

Duckworth, A. L., Quinn, P. D. and Seligman, M. E. P. (2009) ‘Positive predictors of teacher effectiveness’ *J Positive Psychol*, 4(6), pp. 540-547. [https://doi.org/10.1080/17439760903157232](https://doi.org/10.1080/17439760903157232)

Duckworth, A. L. and Seligman, M. E. (2005) ‘Self-discipline outdoes IQ in predicting academic performance of adolescents’, *Psychol Sci*, 16(12), pp. 939-944. [https://doi.org/10.1111/j.1467-9280.2005.01641.x](https://doi.org/10.1111/j.1467-9280.2005.01641.x)

Eskreis-Winkler, L., Shulman, E. P., Beal, S. A. and Duckworth, A. L. (2014) ‘The grit effect: predicting retention
in the military, the workplace, school and marriage’, *Front Psychol.*, 5, pp. 36. https://doi.org/10.3389/fpsyg.2014.00036

Eva, K. W., Rosenfeld, J., Reiter, H. I. and Norman, G. R. (2004) ‘An admissions OSCE: the Multiple Mini-Interview’, *Med Educ.*, 38(3), pp: 314-326. https://doi.org/10.1046/j.1365-2923.2004.01776.x

Farrington, C. A., Roderick, M., Allensworth, E., Nagaoka, J., et al. (2012) *Teaching adolescents to become learners. The role of noncognitive factors in shaping school performance: A critical literature review*, Chicago: University of Chicago Consortium on Chicago School Research. Available at: https://consortium.uchicago.edu/sites/default/files/publications/Noncognitive%20Report.pdf

George, J. M., Young, D. and Metz, E. N. (1989) ‘Evaluating selected internship candidates and their subsequent performances’, *Acad Med.*, 64(8), pp. 480-482. https://doi.org/10.1097/00001888-198908000-00013

Gnambs, T. and Kaspar, K. (2017) ‘Socially desirable responding in web-based questionnaires: a meta analytic review of the Candor hypothesis’, *Assessment*, 24(6), pp. 746-762. https://doi.org/10.1177/1073191115624547

Gough, H. G., Bradley, P. and McDonald, J. S. (1991) ‘Performance of residents in anesthesiology as related to measures of personality and interests’, *Psychol Rep.*, 68(3 Pt 1), pp. 979-994. https://doi.org/10.2466/pr0.1991.68.3.979

Guffey, R. C., Rusin, K., Chidiac, E. J. and Marsh, H. M. (2011) ‘The utility of pre-residency standardized tests for anesthesiology resident selection: the place of United States Medical Licensing Examination scores’, *Anesth Analg.*, 112(1), pp. 201-206. https://doi.org/10.1213/ANE.0b013e3181f9facd

Halliday, L., Walker, A., Vig, S., Hines, J., etc. (2017) ‘Grit and burnout in UK doctors: a cross-sectional study across specialties and stages of training’, *Postgrad Med J.*, 93(1101), pp. 389-394. https://doi.org/10.1136/postgradmedj-2015-133919

Harfmann, K. L. and Zirwas, M. J. (2011) ‘Can performance in medical school predict performance in residency? A compilation and review of correlative studies’, *J Am Acad Dermatol.*, 65(5), pp. 1010-1022.e1012. https://doi.org/10.1016/j.jaad.2010.07.034

Hofmeister, M., Lockyer, J. and Crutcher, R. (2009) ‘The Multiple Mini-Interview for selection of international medical graduates into family medicine residency education’, *Med Educ.*, 43(6), pp. 573-579. https://doi.org/10.1111/j.1365-2923.2009.03380.x

Holmboe, E.S. (2004) ‘Faculty and the observation of trainees’ clinical skills: problems and opportunities’, *Acad Med.*, 79(1), pp. 16-22. https://doi.org/10.1097/00001888-200401000-00006

Kenny, S., McInnes, M. and Singh, V. (2013) ‘Associations between residency selection strategies and doctor performance: a meta-analysis’, *Med Educ.*, 47(8), pp. 790-800. https://doi.org/10.1111/medu.12234

Lambe, P. and Bristow, D. (2010) ‘What are the most important non-academic attributes of good doctors? A Delphi survey of clinicians’, *Med Teach.*, 32(8), pp. e347-354. https://doi.org/10.3109/0142159X.2010.490603

Lubelski, D., Healy, A. T., Friedman, A., Ferraris, D., etc. (2016) ‘Correlation of personality assessments with
standard selection criteria for neurosurgical residency applicants’, *J Neurosurg*, 125(4), pp. 986-994. [https://doi.org/10.15694/mep.2018.0000273.1](https://doi.org/10.15694/mep.2018.0000273.1)

Marley, J. and Carman, I. (1999) ‘Selecting medical students: a case report of the need for change’, *Med Educ*, 33(6), pp. 455-459. [https://doi.org/10.1046/j.1365-2923.1999.00345.x](https://doi.org/10.1046/j.1365-2923.1999.00345.x)

McCaskill, Q. E., Kirk, J. J., Barata, D. M., Wludyka, P. S., *et al.* (2007) ‘USMLE step 1 scores as a significant predictor of future board passage in pediatrics’, *Ambul Pediatr*, 7(2), pp. 192-195. [https://doi.org/10.1016/j.ambp.2007.01.002](https://doi.org/10.1016/j.ambp.2007.01.002)

McDonald, J. S., Lingam, R. P., Gupta, B., Jacoby, J., *et al.* (1994) ‘Psychologic testing as an aid to selection of residents in anesthesiology’, *Anesth Analg*, 78(3), pp. 542-547. [https://doi.org/10.1213/00000539-199403000-00022](https://doi.org/10.1213/00000539-199403000-00022)

Merlo, L. J. and Matveevskii, A. S. (2009) ‘Personality testing may improve resident selection in anesthesiology programs’, *Med Teach*, 31(12), pp. e551-554. [https://doi.org/10.3109/01421590903390593](https://doi.org/10.3109/01421590903390593)

Metro, D. G., Talarico, J. F., Patel, R. M. and Wetmore, A. L. (2005) ‘The resident application process and its correlation to future performance as a resident’, *Anesth Analg*, 100(2), pp. 502-505. [https://doi.org/10.1213/01.ANE.0000154583.47236.5F](https://doi.org/10.1213/01.ANE.0000154583.47236.5F)

Olawaiye, A., Yeh, J. and Withiam-Leitch, M. (2006) ‘Resident selection process and prediction of clinical performance in an obstetrics and gynecology program’, *Teach Learn Med*, 18(4), pp. 310-315. [https://doi.org/10.1207/s15328015tlm1804_6](https://doi.org/10.1207/s15328015tlm1804_6)

Poropat, A. E. (2009) ‘A meta-analysis of the five-factor model of personality and academic performance’, *Psychol Bull*, 135(2), pp. 322-338. [https://doi.org/10.1037/a0014996](https://doi.org/10.1037/a0014996)

Prober, C. G., Kolars, J. C., First, L. R. and Melnick, D. E. (2016) ‘A plea to reassess the role of United States medical l examination step 1 scores in residency selection’, *Acad Med*, 91(1), pp. 12-15. [https://doi.org/10.1097/ACM.0000000000000855](https://doi.org/10.1097/ACM.0000000000000855)

Results of the 2014 NRMP Program Director Survey. 2014. Washington, DC: National Resident Matching Program. [https://www.nrmp.org/wp-content/uploads/2014/09/PD-Survey-Report-2014.pdf](https://www.nrmp.org/wp-content/uploads/2014/09/PD-Survey-Report-2014.pdf)

Robertson-Kraft, C. and Duckworth, A. L. (2014) ‘True grit: trait-level perseverance and passion for long-term goals predicts effectiveness and retention among novice teachers’, *Teach Coll Rec (1970)*, 116(3).

Salles, A., Cohen, G. L. and Mueller, C. M. (2014) ‘The relationship between grit and resident well-being’, *Am J Surg*, 207(2), pp. 251-254. [https://doi.org/10.1016/j.amjsurg.2013.09.006](https://doi.org/10.1016/j.amjsurg.2013.09.006)

Salles, A. Lin, D., Liebert, C., Esquivel, M., *et al.* (2017) ‘Grit as a predictor of risk of attrition in surgical residency’, *Am J Surg*, 213(2), pp. 288-291. [https://doi.org/10.1016/j.amjsurg.2016.10.012](https://doi.org/10.1016/j.amjsurg.2016.10.012)

Shellito, J. L., Osland, J. S., Helmer, S. D. and Chang, F. C. (2010) ‘American Board of Surgery examinations: can we identify surgery residency applicants and residents who will pass the examinations on the first attempt?’, *Am J Surg*, 199(2), pp. 216-222. [https://doi.org/10.1016/j.amjsurg.2009.03.006](https://doi.org/10.1016/j.amjsurg.2009.03.006)
Sklar, M. C., Eskander, A., Dore, K. and Witterick, I. J. (2015) ‘Comparing the traditional and Multiple Mini Interviews in the selection of post-graduate medical trainees’, Can Med Educ J, 6(2), pp. e6-e13. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4795078/pdf/cmej0606.pdf

Strand, E. A., Moore, E. and Laube, D. W. (2011) ‘Can a structured, behavior-based interview predict future resident success?’, Am J Obstet Gynecol, 204(5), pp. 446.e441-446.e413. https://doi.org/10.1016/j.ajog.2011.02.019

Tolan, A. M., Kaji, A. H., Quach, C., Hines, O. J., et al. (2010) ‘The electronic residency application service application can predict accreditation council for graduate medical education competency-based surgical resident performance’, J Surg Educ, 67(6), pp. 444-448. https://doi.org/10.1016/j.jsurg.2010.05.002

University of Washington Graduate Medical Education Office. Assessment and evaluation policy final. [accessed 2016 May 29].
http://www.uwmedicine.org/education/Documents/gme/Assessment%20and%20Evaluation%20Policy%20FINAL.pdf.

Appendices
None.

Declarations

The author has declared that there are no conflicts of interest.

This has been published under Creative Commons "CC BY 4.0" (https://creativecommons.org/licenses/by-sa/4.0/)

Ethics Statement

This study was approved by the Duke University Institutional Review Board, reference number: Pro00038507.

External Funding

This paper has not had any External Funding

AMEE MedEdPublish: rapid, post-publication, peer-reviewed papers on healthcare professions’ education. For more information please visit www.mededpublish.org or contact mededpublish@dundee.ac.uk.