Singapore Management University

Institutional Knowledge at Singapore Management University

Research Collection Lee Kong Chian School Of Business

Lee Kong Chian School of Business

5-2018

Response distortion on personality tests in applicants: Comparing high-stakes to low-stakes medical settings

Jeromy ANGLIM
Deakin University

Stefan BOZIC
Deakin University

Jonathon LITTLE

Filip LIEVENS
Singapore Management University, filiplievens@smu.edu.sg
DOI: https://doi.org/10.1007/s10459-017-9796-8

Follow this and additional works at: https://ink.library.smu.edu.sg/lkcsb_research

Part of the Human Resources Management Commons, Industrial and Organizational Psychology Commons, and the Organizational Behavior and Theory Commons

Citation

Response distortion on personality tests in applicants: Comparing high-stakes to low-stakes medical settings

Jeromy Anglim, School of Psychology, Deakin University, Locked Bag 20000, Geelong 3220, Australia

Corresponding author, jeromy.anglim@deakin.edu.au

Stefan Bozic, School of Psychology, Deakin University

Jonathon Little, Abbotsford, Melbourne, VA, Australia

Filip Lievens, Department of Personnel Management, Work and Organizational Psychology, Ghent University, Ghent, Belgium

Abstract

The current study examined the degree to which applicants applying for medical internships distort their responses to personality tests and assessed whether this response distortion led to reduced predictive validity. The applicant sample (n = 530) completed the NEO Personality Inventory whilst applying for one of 60 positions as first-year post-graduate medical interns. Predictive validity was assessed using university grades, averaged over the entire medical degree. Applicant responses for the Big Five (i.e., neuroticism, extraversion, openness, conscientiousness, and agreeableness) and 30 facets of personality were compared to a range of normative samples where personality was measured in standard research settings including medical students, role model physicians, current interns, and standard young-adult test norms. Applicants had substantially higher scores on conscientiousness, openness, agreeableness, and extraversion and lower scores on neuroticism with an average absolute standardized difference of 1.03, when averaged over the normative samples. While current interns, medical students, and especially role model physicians do show a more socially desirable personality profile than standard test norms, applicants provided responses that were substantially more socially desirable. Of the Big Five, conscientiousness was the strongest predictor of academic performance in both applicants (r = .11) and medical students (r = .21). Findings suggest that applicants engage in substantial response distortion, and that the predictive validity of personality is modest and may be reduced in an applicant setting.

Keywords

Personality traits, Academic performance, Medical students, Five Factor Model, Medical student selection

Introduction

Given the importance of integrity, motivation, and interpersonal skills to being an effective doctor, it is important to incorporate measures of non-cognitive characteristics into academic and employment admissions procedures (Patterson et al. 2016). Despite their importance, it remains unclear how best to measure these characteristics, particularly in the context of high-stakes testing where applicants may be motivated to distort their responses (Albanese et al. 2003; Bore et al. 2009; Musson 2009; Patterson et al. 2016). In recent years, a variety of selection tools have been developed that aim to assess these non-cognitive characteristics, including situational judgment tests (Bore et al. 2009; De Leng et al. 2017; Lievens 2013; Patterson et al. 2009, 2012), multiple mini-interviews (Eva and Macala 2014; Eva et al. 2004, 2009; Griffin and Wilson 2012a; Kulasegaram et al. 2010), emotional intelligence tests (Libbrecht et al. 2014), and personality tests (Griffin and Wilson 2012b; Lievens et al. 2002; MacKenzie et al. 2017; Rothstein and Goffin 2006). The current study focuses on personality testing in the context of the high-stakes selection of medical interns assessing the degree to which response distortion might limit the utility of personality tests.

A recurring criticism of personality testing is that applicants might intentionally distort their responses when they know that their responses will be used to inform admission decisions (Morgeson et al. 2007). However, most of what is known about personality testing in selection contexts comes from the large literature on employee selection (Oswald and Hough 2008; Rothstein and Goffin 2006). Meta-analytic research comparing job applicants with non-applicants suggests that applicants distort responses (Birkeland et al. 2006), with applicants typically scoring around half a standard deviation higher on measures of conscientiousness (Birkeland et al. 2006). In contrast, the vast majority of research on medical students and interns has administered personality tests in contexts where scores are used only for low-stakes (research) purposes (McLarnon et al. 2017). The few studies that have examined personality testing in high-stakes (selection) contexts generally have methodological limitations such as small sample sizes and lack of comparison groups (Hobfoll et al. 1982; Shen and Comrey 1997). Arguably, the best current estimate of applicant response distortion in a medical context comes from a small-sample repeated-measures study (n = 63) (Griffin and Wilson 2012b). Responses in the selection context were approximately two-thirds of a standard deviation higher on the Big Five (i.e., extraversion, openness, conscientiousness, agreeableness, and reversed neuroticism) compared to non-applicant responses (Griffin et al. 2008).

A related important issue is whether response distortion reduces the predictive validity of personality test scores (Griffin and Wilson 2012b; MacKenzie et al. 2017; Rothstein and Goffin 2006). The Big Five personality traits have provided a useful organizing framework, with meta-analytic research, largely in non-medical student contexts, finding modest correlations with academic grades for conscientiousness (r = .19) and openness, r = .10 (Poropat 2009).

Although research has examined correlations of medical student personality with academic performance (Doherty and Nugent 2011; Ferguson et al. 2003; Haight et al. 2012; Knights and Kennedy 2007; Lievens et al. 2002, 2009; McLarnon et al. 2017; Peng et al. 1995) and other outcomes (Hojat et al. 2015; Jerant et al. 2012; McManus et al. 2004; Pohl et al. 2011; Song and Shi 2017; Tyssen et al. 2007), there appears to be no research that has systematically compared responses to personality in a large sample of medical

students with applicants to medical programs. So, the "jury is still out" regarding the extent to which applicants to medical programs distort their responses. Thus, this study examined the degree to which graduated medical students distort their responses on personality tests when applying for one of several prestigious medical internships. To assess the degree of response distortion in a high-stakes medical selection context we rely on an established paradigm in social desirability research and compare results in the applicant context to several normative samples where responses were collected in a standard low-stakes research context (Mesmer-Magnus and Viswesvaran 2006). To assess whether this response distortion leads to reduced predictive validity in a high-stakes medical selection context, we used university grades that were available both for the applicants and for a large non-applicant sample of medical students.

Methods

Supplementary materials and analyses are available on the OSF at https://osf.io/bjxwu.

Participants and procedure

The applicant sample consisted of medical graduates applying for one of 60 medical internship positions at a major health provider in Australia. The internship represents the first-year of post-graduate education and consists of a period of supervised clinical experience that is typically completed immediately after a graduate degree in medicine. Satisfactory completion is a requirement for general registration with the Medical Board of Australia. Applying for an intern position is a competitive process. Students in the final year of their medical degree record their preferences for where they wish to complete their intern year. Each health provider administers their own selection process. Importantly, the intern program that formed the basis for this research was especially prestigious and applications outstripped positions by approximately ten to one suggesting that applicants would have an incentive to distort responses on the personality test. As only limited places were offered for applicants outside the state, and because of issues with grade standardization, only within-state applicants were retained.

The application process required applicants to first complete an initial application form. From this, 15% of applicants were not asked to continue with the selection process on the basis that they were clearly not competitive for a position. Remaining applicants were then required to complete the personality test online. Applicants were informed that the personality test would influence selection decisions. Participants were encouraged to answer questions honestly and informed that applicants found to have faked would significantly diminish their chances of being selected. The final applicant sample used for present analyses consisted of 530 participants (55% female; age at time of personality testing, M = 24.9, SD = 3.0, range 21–46).

Materials

Personality

Applicants completed versions of the NEO Personality Inventory-3 (NEO PI-3) measuring five domains (i.e., neuroticism, extraversion, openness, agreeableness, and conscientiousness) along with six facets per

domain. The test consists of 240 items with 8 items per facet. Items were measured on a 1–5 scale, and scales were scored as the item-mean after reversing reversed-items. The test is one of the most established and well-validated measures of the Big Five with Cronbach's alpha reliabilities for domains of around .90 (Costa and Mccrae 1992). The NEO PI-3 (McCrae et al. 2005) involved minor revisions to 37 of the 240 items from the earlier NEO PI-R (Costa and Mccrae 1992) in order to improve readability. A large cross-cultural analysis comparing NEO PI-R and NEO PI-3 found that norms were not substantially different between the two versions (De Fruyt et al. 2009).

Grade point average

Academic performance was measured using grade point average (GPA): i.e., the mean student grade over the entire medical degree. Grades were obtained directly from universities. Grades were averaged over years and then z-score standardized within universities to remove any systematic differences in grading practices between universities.

Comparison samples

To assess the degree to which the applicant context led to response distortion, we calculated mean standardized differences on personality scores between the applicants and several normative samples. Each normative sample completed the personality test in a standard research context where there was no obvious incentive to make a positive impression. In addition, the medical student sample also provided measures of GPA from which comparative predictive validities of personality on GPA could be derived.

Medical students

Individual-level data was obtained for a sample of medical students drawn from research in Flemish universities (n = 539, 62% female, age at time of completion of personality test, M = 18.25, SD = 0.83). Students completed the official Dutch translation of the NEO-PI-R in the first year of their medical degree. Student grades were then obtained throughout the degree and z-score standardized in the same way as was done for the applicant sample. Students who completed the study were drawn from the 1997 and 1998 cohorts of a larger longitudinal study (Lievens et al. 2009). Present analyses differ from previous uses of the data in that they (a) combine the 1997 and 1998 cohorts to maximize sample size, and (b) make the sample more comparable to the applicant sample, who all finished their medical degree, by only including students who provided at least 6 years of academic grades. Students completed the personality test during the first year of their medical degree as part of a longitudinal research project. Thus, the average time difference between personality measurement and grades was similar for the student and the applicant sample, and both samples excluded students who did not complete their medical degrees.

Current interns and physician role models

Scale means and standard deviations for the NEO PI-R were obtained from a study by Hojat et al. (1999). This included a current intern sample of 104 physicians in internal medicine residency (33% female). A

second physician role model sample of 188 physicians was used. Participants in this sample were invited to participate if their managers deemed them to be positive role models (13% female).

NEO PI-3 and NEO-PI-R norms

Combined gender norms for NEO PI-3 Form S young adults (21–30 years, n = 218) (McCrae et al. 2005) and combined college-aged norms for NEO PI-R (n = 389) (Costa and Mccrae 1992) were obtained from the NEO PI test manuals. Age and gender of these two samples are very similar to the applicant sample. The correspondence of the NEO-PI-3 norms and the NEO-PI-R norms illustrates how the small changes between versions of the test do not substantively influence conclusions about applicant response distortion.

Data analytic approach

The degree of applicant response distortion was quantified using the standardized differences in means between applicants and non-applicants. These standardized differences (i.e., Cohen's d) involved first subtracting the non-applicant mean from the applicant mean and then dividing this difference by the applicant standard deviation. The applicant standard deviation was used in order to have a consistent denominator across normative samples. A common rule of thumb is to interpret d values of 0.2, 0.5, and 0.8 as indicating small, medium, and large effects respectively (Cohen 1992). Correlations were used to examine the bivariate relationship between personality and GPA, and regression models were used to examine the overall prediction of GPA by personality. Facet-level correlations are presented and discussed in relation to existing literature (Anglim and Grant 2016; Anglim et al. 2017; de Vries et al. 2011; Gray and Watson 2002; Griffin et al. 2004; Horwood et al. 2015; Lievens et al. 2002; Marshall et al. 2005; Paunonen and Jackson 2000; Woo et al. 2015) in the online supplement. Item-level descriptive statistics, which are relevant to quantifying item-level social desirability on the NEO-PI3, are also provided in the online supplement.

Results

Means and standard deviations for applicants along with standardized differences between applicants and norm groups are presented for the Big Five and 30 personality facets in Table 1 (confidence intervals are provided in the online supplement). Applicant responses were more socially desirable than all comparison groups, although these differences were slightly reduced in the interns and medical students samples compared to standard age norms. After reversing neuroticism, the average Cohen's d effect size for the five samples was d = 1.03, and the average for each sample was d = 0.48 (physician role models), d = 0.95 (interns), d = 1.07 (medical students), d = 1.31 (NEO-PI-3 young adult norms), and d = 1.32 (NEO-PI-R young adult norms). Overall, substantial differences were present on all of the Big Five factors, but were largest for agreeableness, neuroticism, and conscientiousness. Specifically, when averaged over the five normative samples, standardized differences were −1.14 for neuroticism, 0.64 for extraversion, 0.84 for openness, 1.38 for agreeableness, and 1.14 for conscientiousness. Differences at the facet-level varied substantially within a given Big Five factor. For example, scores for openness to actions, ideas, and values were much higher in applicants, but scores for openness to aesthetics and feelings were about the same for applicants and non-applicants.

Table 1: Differences in personality factor and facet scores between applicants and normative samples

Trait	Intern applicants (530)		Effect sizes (applicants versus norm group)					Effect size
	M	SD	Resident role (188)	Current interns (104)	Medical students (539)	NEO-PI-3 (218)	NEO-PI-R (389)	Mean
Neuroticism	2.30	0.41	−0.18	−1.09	−1.49	−1.23	−1.73	−1.14
Extraversion	3.69	0.34	0.49	0.78	0.72	0.72	0.48	0.64
Openness	3.75	0.33	0.72	0.72	0.57	1.24	0.94	0.84
Agreeableness	3.95	0.32	0.60	1.30	1.24	1.95	1.81	1.38
Conscientiousness	3.96	0.35	0.43	0.86	1.31	1.42	1.66	1.14
N1. Anxiety	2.71	0.59	0.27	−0.49	−0.67	−0.64	−0.81	−0.47
N2. Angry hostility	1.98	0.48	−0.49	−1.48	−1.43	−1.71	−2.13	−1.45
N3. Depression	2.19	0.55	−0.16	−0.93	−1.49	−0.93	−1.31	−0.96
N4. Self-consciousness	2.46	0.55	−0.29	−0.81	−1.04	−0.40	−1.08	−0.72
N5. Impulsiveness	2.56	0.53	−0.46	−0.71	−1.13	−0.88	−1.39	−0.91
N6. Vulnerability	1.93	0.40	0.27	−0.78	−1.40	−1.41	−1.65	−0.99
E1. Warmth	4.33	0.41	0.76	1.06	1.62	1.37	1.06	1.17
E2. Gregariousness	3.68	0.51	1.09	0.77	0.29	0.72	0.55	0.68
E3. Assertiveness	3.34	0.48	−0.33	0.71	0.86	0.43	0.45	0.42
E4. Activity	3.37	0.45	−0.76	0.02	0.28	0.24	0.05	−0.03
E5. Excitement seeking	3.41	0.53	0.80	0.09	−0.03	−0.23	−0.52	0.02
E6. Positive emotions	3.98	0.52	0.41	0.69	0.25	0.65	0.55	0.51
O1. Fantasy	3.18	0.57	−0.19	−0.32	−0.60	−0.23	−0.58	−0.38
O2. Aesthetics	3.65	0.65	0.31	0.46	−0.04	0.83	0.50	0.41
O3. Feelings	3.75	0.46	−0.11	−0.03	−0.01	0.22	−0.11	−0.01
O4. Actions	3.66	0.43	1.45	1.42	1.37	1.53	1.59	1.47
O5. Ideas	4.10	0.54	0.64	0.73	1.04	1.40	1.33	1.03
O6. Values	4.13	0.42	1.04	0.83	0.88	1.37	1.28	1.08
A1. Trust	3.98	0.49	−0.05	0.99	0.91	1.47	1.29	0.92
A2. Straightforwardness	3.89	0.51	0.18	0.71	0.77	1.20	1.18	0.81
A3. Altruism	4.40	0.37	0.74	1.18	1.76	1.52	1.35	1.31
A4. Compliance	3.66	0.49	0.29	0.72	1.18	1.61	1.43	1.05
A5. Modesty	3.73	0.52	0.72	0.60	0.20	1.03	0.91	0.69
A6. Tender-mindedness	4.04	0.39	0.76	1.40	0.52	1.40	1.43	1.10
C1. Competence	4.10	0.40	−0.16	0.79	1.38	1.23	1.17	0.88
C2. Order	3.52	0.49	0.40	0.32	0.62	0.24	0.60	0.44
C3. Dutifulness	4.20	0.39	−0.10	0.38	0.99	1.46	1.39	0.82
C4. Achievement striving	4.15	0.44	0.54	1.02	1.07	1.40	1.80	1.17
C5. Self-discipline	4.18	0.46	0.71	1.15	1.38	1.53	1.78	1.31
C6. Deliberation	3.63	0.52	0.45	0.42	0.77	0.86	1.08	0.72

Effect size estimates are calculated as: (applicant mean − norm mean)/applicant standard deviation. Thus, positive values indicate that applicants scored higher than the norm group. Sample sizes are shown in parentheses below group names. Further supplementary materials are available on the OSF at https://osf.io/bjxwu

The correlations between the Big Five personality and GPA for both the applicant and medical student samples are shown in Table 2. In general, the correlations between personality and GPA were fairly small. Conscientiousness was significantly correlated with GPA in both medical students (r = .21, p < .001), and applicants (r = .11, p < .05). While this correlation was smaller in the applicant sample, this difference was not statistically significant, $\Delta r = -.11$, p = .07. Openness was a significant predictor of GPA in non-applicants (r = .14, p < .01), but not in applicants (r = .04, ns), but this difference was not statistically significant, $\Delta r = -.10$, p = .11. Correlations with GPA for neuroticism, extraversion, and agreeableness were close to zero in both samples. Overall regression models predicting GPA from the Big Five appeared lower for applicants (adjusted multiple R = .11) compared to medical students (adjusted multiple R = .25). Thus, overall there was modest evidence for a reduction in predictive validity of the Big Five domain scores in the applicant context.

Table 2: Comparison of bivariate correlations between big five domain scores and GPA in non-applicants and applicants

	Medical students	Intern applicants	Difference		
	GPA r	GPA r	Δr	z	p
Neuroticism	.04	.01	−.02	0.40	.69
Extraversion	.02	.02	.00	0.04	.96
Openness	**.14**	.04	−.10	1.62	.11
Agreeableness	.03	.03	.00	0.05	.96
Conscientiousness	**.21**	.11	−.11	1.80	.07

$r \geq .09$, $p < .05$; $r \geq .12$, $p < .01$; $r \geq .15$, $p < .001$. Correlations significant at .01 level are bolded. Δr represents differences in trait–GPA correlations between non-applicants and applicants. Statistical significance of differences is based on z test for independent correlations (Olkin and Finn 1995)

Discussion

Overall, the present study fills an important gap in the literature by providing the first nuanced assessment in a large medical sample of the degree to which a selection context influences responses on personality tests and the degree to which this might alter the predictive validity of academic medical performance. The key conclusions are twofold and can be summarized as follows: When used in a high-stakes medical selection context, applicants appear to respond in more socially desirable ways. In fact, applicant response distortion in this sample was somewhat larger than is commonly seen in the literature. Differences with the interns and medical students were around one standard deviation, and these two samples seem to share the most with the applicant sample. This is larger than the estimates of around two-thirds of a standard deviation from the small-sample repeated-measure study by Griffin et al. (2008), and it is also larger than meta-analytic estimates comparing applicants and non-applicants (Birkeland et al. 2006).

Second, response distortion in the applicant context may reduce but not remove the predictive validity of personality test scores. Correlations for openness and conscientiousness with GPA were slightly lower in the applicant context, albeit this difference in correlations was not significant. A small reduction in validity is consistent with a model of response distortion where response distortion adds a small amount of noise to personality measurement, but where the negative effect on predictive validity is offset because

applicants with lower scores tend to distort responses more (Anglim et al. 2017). Thus, the change in rank ordering is not as extreme as it would be were the size of response distortion to be unrelated to true personality scores. Whether the modest predictive validities obtained by personality testing are sufficient to overcome other concerns about their use is a matter of judgment. However, the present data are highly relevant to informing such decisions. At the very least, personality testing should not be used in isolation when measuring interpersonal constructs, and should be combined with other admissions procedures such as multiple mini-interviews and situational judgment tests.

Some limitations should be acknowledged. While comparing applicants to a range of different normative groups, other potential differences besides the selection context may partially explain differences between applicants and non-applicants. While ancillary analyses suggest that such effects are likely to be small relative to the observed differences, future research could seek to obtain samples of applicants and non-applicants that are matched on more variables. Applicants were older (mean age 25 years) than medical students (mean age 18 years); applicants had completed more years of education; applicants completed the English versions of the personality test, and medical students completed the Dutch version; and Flemish and Australian medical students may have different personality profiles. Nonetheless, comparisons with other normative samples suggested that the effects of age, years of education, test format, and cultural differences are likely to be small, relative to the large differences in personality that were obtained. In particular, the student and intern samples did present a slightly more socially desirable response profile than implied by standard young adult test norms.

Although the current research used academic performance as the outcome variable, the broader research goal is to predict what makes an effective doctor. Thus, the assessment of reductions of predictive validity in the applicant sample were mainly intended to assess the broader question of whether predictive validity might decline in general. In particular, many employers perceive the benefits of personality testing to be related more to predicting discretionary behavior rather than standard task performance. The aim is to avoid hiring people who might engage in acts such as bullying, fraud, and unsafe work practices, and to hire people who are more likely to create a positive work climate in the organization. Thus, future research should seek to obtain measures of actual performance in medical practice.

References

Albanese, M. A., Snow, M. H., Skochelak, S. E., Huggett, K. N., & Farrell, P. M. (2003). Assessing personal qualities in medical school admissions. Academic Medicine, 78(3), 313–321.

Anglim, J., & Grant, S. L. (2016). Predicting psychological and subjective well-being from personality: Incremental prediction from 30 facets over the big 5. Journal of Happiness Studies, 17, 59–80.

Anglim, J., Morse, G., De Vries, R. E., Maccann, C., & Marty, A. (2017). Comparing job applicants to nonapplicants using an item-level bifactor model on the HEXACO Personality Inventory. European Journal of Personality. doi:10.1002/per.2120.

Birkeland, S. A., Manson, T. M., Kisamore, J. L., Brannick, M. T., & Smith, M. A. (2006). A meta-analytic investigation of job applicant faking on personality measures. International Journal of Selection and Assessment, 14(4), 317–335.

Bore, M., Munro, D., & Powis, D. (2009). A comprehensive model for the selection of medical students. Medical Teacher, 31(12), 1066–1072.

Cohen, J. (1992). A power primer. Psychological Bulletin, 112(1), 155–159.

Costa, P. T., & Mccrae, R. R. (1992). Neo PI-R professional manual: Odessa. FL: Psychological Assessment Resources.

De Fruyt, F., De Bolle, M., Mccrae, R. R., Terracciano, A., & Costa, P. T. (2009). Assessing the universal structure of personality in early adolescence: The NEO-PI-R and NEO-PI-3 in 24 cultures. Assessment, 16(3), 301–311.

De Leng, W. E., Stegers-Jager, K. M., Husbands, A., Dowell, J. S., Born, M. P., & Themmen, A. P. N. (2017). Scoring method of a situational judgment test: Influence on internal consistency reliability, adverse impact and correlation with personality? Advances in Health Sciences Education, 22, 243–265.

De Vries, A., De Vries, R. E., & Born, M. P. (2011). Broad versus narrow traits: Conscientiousness and Honesty–Humility as predictors of academic criteria. European Journal of Personality, 25(5), 336–348.

Doherty, E. M., & Nugent, E. (2011). Personality factors and medical training: A review of the literature. Medical Education, 45(2), 132–140.

Eva, K. W., & Macala, C. (2014). Multiple mini-interview test characteristics: 'Tis better to ask candidates to recall than to imagine. Medical Education, 48(6), 604–613.

Eva, K. W., Reiter, H. I., Trinh, K., Wasi, P., Rosenfeld, J., & Norman, G. R. (2009). Predictive validity of the multiple mini-interview for selecting medical trainees. Medical Education, 43(8), 767–775.

Eva, K. W., Rosenfeld, J., Reiter, H. I., & Norman, G. R. (2004). An admissions OSCE: The multiple miniinterview. Medical Education, 38(3), 314–326.

Ferguson, E., Mcmanus, I., James, D., O'hehir, F., & Sanders, A. (2003). Pilot study of the roles of personality, references, and personal statements in relation to performance over the five years of a medical degree Commentary: How to derive causes from correlations in educational studies. BMJ, 326(7386), 429–432.

Gray, E. K., & Watson, D. (2002). General and specific traits of personality and their relation to sleep and academic performance. Journal of Personality, 70(2), 177–206.

Griffin, B., Harding, D. W., Wilson, I. G., & Yeomans, N. D. (2008). Does practice make perfect? The effect of coaching and retesting on selection tests used for admission to an Australian medical school. Medical Journal of Australia, 189(5), 270–273.

Griffin, B., Hesketh, B., & Grayson, D. (2004). Applicants faking good: Evidence of item bias in the NEO PI-R. Personality and Individual Differences, 36(7), 1545–1558.

Griffin, B., & Wilson, I. G. (2012a). Associations between the big five personality factors and multiple miniinterviews. Advances in Health Sciences Education, 17(3), 377–388.

Griffin, B., & Wilson, I. G. (2012b). Faking good: Self-enhancement in medical school applicants. Medical Education, 46(5), 485–490.

Haight, S. J., Chibnall, J. T., Schindler, D. L., & Slavin, S. J. (2012). Associations of medical student personality and health/wellness characteristics with their medical school performance across the curriculum. Academic Medicine, 87(4), 476–485.

Hobfoll, S., Anson, O., & Antonovsky, A. (1982). Personality factors as predictors of medical student performance. Medical Education, 16(5), 251–258.

Hojat, M., Michalec, B., Jon Veloski, J., & Tykocinski, M. L. (2015). Can empathy, other personality attributes, and level of positive social influence in medical school identify potential leaders in medicine? Academic Medicine, 90(4), 505–510.

Hojat, M., Nasca, T. J., Magee, M., Feeney, K., Pascual, R., Urbano, F., et al. (1999). A comparison of the personality profiles of internal medicine residents, physician role models, and the general population. Academic Medicine, 74(12), 1327–1333.

Horwood, S., Anglim, J., & Tooley, G. (2015). Type D personality and the Five-Factor Model: A facet-level analysis. Personality and Individual Differences, 83, 50–54.

Jerant, A., Griffin, E., Rainwater, J., Henderson, M., Sousa, F., Bertakis, K. D., et al. (2012). Does applicant personality influence multiple mini-interview performance and medical school acceptance offers? Academic Medicine, 87(9), 1250–1259.

Knights, J. A., & Kennedy, B. J. (2007). Medical school selection: Impact of dysfunctional tendencies on academic performance. Medical Education, 41(4), 362–368.

Kulasegaram, K., Reiter, H. I., Wiesner, W., Hackett, R. D., & Norman, G. R. (2010). Non-association between NEO-5 personality tests and multiple mini-interview. Advances in Health Sciences Education, 15(3), 415–423.

Libbrecht, N., Lievens, F., Carette, B., & Coˆteˊ, S. (2014). Emotional intelligence predicts success in medical school. Emotion, 14(1), 64–73.

Lievens, F. (2013). Adjusting medical school admission: Assessing interpersonal skills using situational judgement tests. Medical Education, 47(2), 182–189.

Lievens, F., Coetsier, P., De Fruyt, F., & De Maeseneer, J. (2002). Medical students' personality characteristics and academic performance: A five-factor model perspective. Medical Education, 36(11), 1050–1056.

Lievens, F., Ones, D. S., & Dilchert, S. (2009). Personality scale validities increase throughout medical school. Journal of Applied Psychology, 94(6), 1514–1535.

Mackenzie, R. K., Dowell, J., Ayansina, D., & Cleland, J. A. (2017). Do personality traits assessed on medical school admission predict exit performance? A UK-wide longitudinal cohort study. Advances in Health Sciences Education, 22, 365–385.

Marshall, M. B., De Fruyt, F., Rolland, J.-P., & Bagby, R. M. (2005). Socially desirable responding and the factorial stability of the NEO PI-R. Psychological Assessment, 17(3), 379–384.

Mccrae, R. R., Costa, J., Paul, T., & Martin, T. A. (2005). The NEO-PI-3: A more readable revised NEO personality inventory. Journal of Personality Assessment, 84(3), 261–270.

Mclarnon, M. J., Rothstein, M. G., Goffin, R. D., Rieder, M. J., Poole, A., Krajewski, H. T., et al. (2017). How important is personality in the selection of medical school students? Personality and Individual Differences, 104, 442–447.

Mcmanus, I., Keeling, A., & Paice, E. (2004). Stress, burnout and doctors' attitudes to work are determined by personality and learning style: A twelve year longitudinal study of UK medical graduates. BMC Medicine, 2, 29.

Mesmer-Magnus, J., & Viswesvaran, C. (2006). Assessing response distortion in personality tests: A review of research designs and analytic strategies. In M. H. Peterson & R. L. Griffith (Eds.), A closer examination of applicant faking behavior (pp. 85–113). Charlotte: Information Age Publishing.

Morgeson, F. P., Campion, M. A., Dipboye, R. L., Hollenbeck, J. R., Murphy, K., & Schmitt, N. (2007). Reconsidering the use of personality tests in personnel selection contexts. Personnel Psychology, 60(3), 683–729.

Musson, D. M. (2009). Personality and medical education. Medical Education, 43(5), 395–397.

Olkin, I., & Finn, J. D. (1995). Correlations redux. Psychological Bulletin, 118(1), 155–164.

Oswald, F. L., & Hough, L. M. (2008). Personality testing and industrial–organizational psychology: A productive exchange and some future directions. Industrial and Organizational Psychology: Perspectives on Science and Practice, 1(3), 323–332.

Patterson, F., Ashworth, V., Zibarras, L., Coan, P., Kerrin, M., & O'neill, P. (2012). Evaluations of situational judgement tests to assess non-academic attributes in selection. Medical Education, 46(9), 850–868.

Patterson, F., Baron, H., Carr, V., Plint, S., & Lane, P. (2009). Evaluation of three short-listing methodologies for selection into postgraduate training in general practice. Medical Education, 43(1), 50–57.

Patterson, F., Knight, A., Dowell, J., Nicholson, S., Cousans, F., & Cleland, J. (2016). How effective are selection methods in medical education? A systematic review. Medical Education, 50(1), 36–60.

Paunonen, S. V., & Jackson, D. N. (2000). What is beyond the big five? Plenty! Journal of Personality, 68(5), 821–835.

Peng, R., Khaw, H., & Edariah, A. (1995). Personality and performance of preclinical medical students. Medical Education, 29(4), 283–288.

Pohl, C. A., Hojat, M., & Arnold, L. (2011). Peer nominations as related to academic attainment, empathy, personality, and specialty interest. Academic Medicine, 86(6), 747–751.

Poropat, A. E. (2009). A meta-analysis of the five-factor model of personality and academic performance. Psychological Bulletin, 135(2), 322–338.

Rothstein, M. G., & Goffin, R. D. (2006). The use of personality measures in personnel selection: What does current research support? Human Resource Management Review, 16(2), 155–180.

Shen, H., & Comrey, A. L. (1997). Predicting medical students' academic performances by their cognitive abilities and personality characteristics. Academic Medicine, 72(9), 781–786.

Song, Y., & Shi, M. (2017). Associations between empathy and big five personality traits among Chinese undergraduate medical students. PLoS ONE, 12(2), e0171665.

Tyssen, R., Dolatowski, F. C., Røvik, J. O., Thorkildsen, R. F., Ekeberg, Ø., Hem, E., et al. (2007). Personality traits and types predict medical school stress: A six-year longitudinal and nationwide study. Medical Education, 41(8), 781–787.

Woo, S. E., Jin, J., & Lebreton, J. M. (2015). Specificity matters: Criterion-related validity of contextualized and facet measures of conscientiousness in predicting college student performance. Journal of Personality Assessment, 97(3), 301–309.