

Response Distortion on the NEO PI-R among Candidates Taking the Entrance Examination to the National School of Civil Aviation (ENAC-France)

Jacques Juhel¹, Sophie Brunot¹, Gérard Zapata²

¹Centre de Recherche en Psychologie, Cognition et Communication, Université Européenne de
Bretagne-Rennes 2, Rennes, France

²Ecole Nationale d'Aviation Civile, Toulouse, France
Email: jacques.juhel@univ-rennes2.fr

Received February 2nd, 2012; revised March 2nd, 2012; accepted April 6th, 2012

This study examined the effect of a selection context on the responses to a five factor personality questionnaire (Revised NEO Personality Inventory) of 974 candidates taking the entrance examination to the ENAC (National School of Civil Aviation, France). A response distortion index was calculated using Schinka, Kinder and Kremer (1997)'s method. The results indicate: 1) lower neuroticism scores but higher conscientiousness, agreeableness and extraversion scores compared to standard conditions; 2) a high estimated effect of the response distortion index on all dimensions of the NEO PI-R except for Openness to experience; 3) substantial stability of scores at a one-year interval (N = 117). The paper discusses the conceptual and practical value of this response distortion index for measuring the dimensions of the NEO PI-R in candidates taking the entrance examination to the ENAC.

Keywords: Five Factor Model of Personality; Presentation Management; Social Desirability; Selection for Training

Introduction

Despite being heavily criticized, personality inventories based on the Five Factor Model are still widely used in psychological assessment in a variety of settings (Hogan, 2005), particularly when selecting personnel or prospective candidates for a training course. A major reason for the continued use of such measures is the higher predictive validity associated with the collected data (Hogan, De Fruyt, & Rolland, 2006; Mount & Barrick, 1998). Researchers have established that Conscientiousness (being dependable, organized, consistent) is both a positive predictor (compliance with rules and procedures) and a negative predictor (creativity in addressing new issues) of professional performance in various occupational groups. They have also shown that Neuroticism (vs. Emotional stability) is negatively (vs. positively) related to job proficiency and training proficiency (Mount, Barrick, & Stewart, 1998). It has been also found that Extraversion (sociable, energetic, active) and above all Openness to experience (intellectual, imaginative) are predictors of the ability to benefit from training (Barrick & Mount, 1991). Finally, Agreeableness (cooperative, considerate, trusting) has been shown to be a valid predictor of performance at work where the job involves interactions with a team (Judge, Higgins and Thoresen, 1999; Salgado, 1997).

However, the tendency of people to bias their responses to personality items have raised some questions about the validity of five factor personality tests (e.g., Ones, Viswesvaran, Dilchert, & Deller, 2006). These response biases can be defined as systematic tendencies to respond to items tests in some basis other than the specific items contents (i.e., what the items were designed to measure) (Paulhus, 1991). Several bias has been

identified (e.g., acquiescence bias, central tendency bias) but the most frequently studied refers to socially desirable responding (SDR) that is to say the tendency for participants to present a favorable image of themselves (Paulhus, 1991). There is indeed some research evidence available to suggest that in a context that has significant social implications, people generally respond according to what they perceive to be socially desirable and useful to them. Several studies have revealed that SDR contaminates the data collected with personality inventories. For example, Ones, Viswesvaran and Reiss (1996) have established, using meta-analysis methods, that SDR was mainly related to Neuroticism (-.37), Agreeableness (.20) and Conscientiousness (.14). It has also been found with a more experimental approach that subjects instructed to "fake good" in an experimental situation with no social implications were generally able to do so (up to a standard deviation of difference compared to the control group; Viswesvaran & Ones, 1999).

Many studies have been undertaken to assess whether SDR (i.e., social desirability) does compromise criterion-validity of personality scale scores and represents construct irrelevant variance. Based on current evidence, there is no obvious answer to this question. A number of studies have suggested that SDR adds noise and results in an increased mean score, a change in the rank order of traits (Rosse, Stecher, Miller, & Levin, 1998), a change in factor structure (Schmit & Ryan, 1993) and criterion-related validity (Caldwell-Andrews, Baer, & Berry, 2000). In addition, SDR could be a negative predictor of performance in stressful conditions (Sandal, Musson, Helmreich, & Gravdal, 2005). The interpretation of this result is that a higher SDR level is associated with a higher level of arousal in potentially threatening situations for self-esteem and a higher probability

of implementing defense mechanisms. However, other researchers have argued that the factor structure of scores remains unchanged in a selection context (Ellingson, Smith, & Sackett, 2001) and that the correction of the bias caused by an overly positive presentation of the self does not increase the criterion-related validity of personality inventories (see for example Piedmont, McCrae, Riemann, & Angleitner, 2000). In short, while there is strong evidence of the SDR phenomenon (particularly in a selection context), the effect of SDR on the construct validity of personality inventories is not unequivocal.

The implications of SDR in personality measurement clearly depend on whether the bias is seen as the involuntary consequence of positive self-image or a more intentional behavior (Wiggins, 1964). In the first case, socially desirable responding might reflect an unconscious and not deliberate bias in self regard. According to Paulhus and Reid (1991), such bias might refer to the extent to which individuals inadvertently exaggerate their desirable qualities (self enhancement) or inadvertently conceal or minimize undesirable qualities (self denial). This so-called self-deception might represent a substantial feature of personality (Costa & McCrae, 1997; Hogan, Hogan, & Roberts, 1996). In contrast, socially desirable responding might be a deliberate attempt to give a positive self-image. This “propagandistic” bias or Impression Management (Paulhus, 1984) might represent a circumstantial style of response bias that may interfere with the measurement of personality.

Despite the argument of some researchers against the development of response distortion scales for personality inventories (e.g., Costa & McCrae, 1997), specific scales have been proposed to detect various patterns of response distortion among applicants. Schinka, Kinder, & Kremer (1997) in particular have developed two research validity scales to be used with the Revised NEO Personality Inventory (Costa & McCrae, 1992): a Positive Presentation Management scale (PPM) that was intended to identify applicants who present in an overly positive fashion and a Negative Presentation Management scale (NPM) that was intended to identify applicants who present in an overly negative fashion. Rather than introducing new items, Schinka and associates exclusively selected NEO PI-R items deviating from the mean in an extreme positive or negative direction for inclusion in these scales. Several studies utilizing these PPM and NPM scales have found support that they were sensitive to impression management strategies (Ballenger, Caldwell-Andrews, & Baer, 2001; Berry et al., 2001; Caldwell-Andrews, Baer, & Berry, 2000; Morey, Quigley, Sanislow et al., 2002; Reid-Seizer & Fritzsche, 2001; Young & Schinka, 2001).

This paper aims to measure the SDR phenomenon and to explore the links between SDR and the personality dimensions of candidates taking the ENAC entrance exam. These were evaluated using the French version of the NEO PI-R (ECPA, 1998, 2003), and the two Presentation Management scales developed by Schinka et al. (1997). Since the general research hypothesis predicts a positive response bias among candidates, a decrease of the N scores (Bagby & Marshall, 2003) and an increase of the E, A and C scores (McFarland, Wiechmann, & Chandler, 2001; Ones, Viswesvaran, & Reiss, 1996; Reid-Seiser & Fritzsche, 2001) are predicted. Given the lack of consensus in the literature, no prediction can be made about O scores (Griffin, Hesketh, & Grayson, 2004; McFarland and Ryan, 2000; Paulhus, Bruce, & Trapnell, 1995; Topping & O’Gorman, 1997). The paper also examines individual trends and average

trends in NEO PI-R scores among candidates retaking the exam a year later. There are two possible hypothesis. The first hypothesis predicts that SDR will reach a maximum threshold in the first exam session and that the scores in the NEO PI-R facets and validity scales (PPM, NPM) will not vary significantly between the two exam sessions (“ceiling” effect hypothesis) (e.g., Hogan, Barrett, & Hogan, 2007). A need to be consistent across time could also explain the lack of difference between exam sessions. However, it is also plausible that the degree of response bias in applicants after having been rejected in the first exam session will be greater in the second exam session because of a more acute perception of the social implications of the exam (“forced” trait hypothesis). It is therefore expected a decrease of the N scores and an increase of the C, E and A scores from one session to another.

Method

Participants and Procedure

The NEO PI-R, an operationalization of the five-factor model, was administered to a sample of 974 candidates (including 862 men and 112 women aged on average in their twenties) taking the ENAC entrance exam and to a sub-sample of 117 candidates (including 111 men and 6 women) retaking the exam a year later.

Material

A French version of the NEO PI-R elaborated by Rolland (Costa, McCrae, Rolland, 1998-2004) which appears to be largely equivalent to the original language version (Rolland, Parker, & Stumpf, 1998) was used in that study. The inventory consisted of 240 items with 30 facet scales and 5 domain scores (N: Neuroticism, E: Extraversion, O: Openness to experience, A: Agreeableness, and C: Conscientiousness). Only the domain scores (48 items per domain) were taken into account in the present paper. Each validity scale consisted of 10 items of the NEO PI-R: 2 per domain for the PPM scale (e.g., “It’s easy to make me afraid”, “I try generally to be thoughtful and considerate”, “I try to conscientiously perform all the tasks entrusted to me”); and 2 for N, 3 for E, 3 for O, 1 for A and 1 for C for the NPM scale (e.g., “It’s difficult for me to make a decision”, “Over the years, I made a number of silly things”, “I do not always assert myself as much as I should”). The validity scales were calculated based on the method developed by Schinka et al. (1997). A Response Distortion experimental index was measured by subtracting the NPM score from the PPM score.

Results

Figure 1 presents T scores ($m = 50$; $\sigma = 10$) obtained by the 974 candidates sitting the ENAC entrance exam and by a French reference sample (norms level 3: baccalaureate up to 2 years of post-baccalaureate higher education studies, see Costa et al., 1998-2004). A comparison of these scores indicates as expected that the exam candidates were significantly ($p < .001$) less likely to describe themselves as being emotionally unstable [$t(973) = -53.20$], more likely to describe themselves as being more extravert [$t(973) = 23.37$], more agreeable [$t(973) = 14.98$], and more conscientious [$t(973) = 34.08$]. Moreover, they described themselves as less open to experience [$t(973) = -7.03$] than the reference group did. The effect size measured

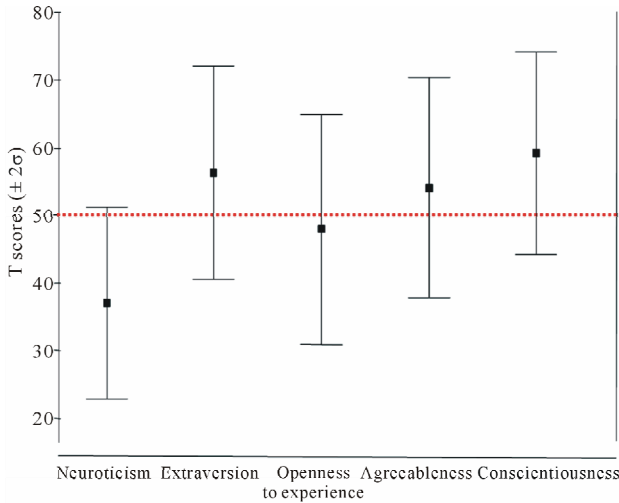


Figure 1. Mean T scores of the 974 candidates sitting the National School of Civil Aviation entrance exam (T = 50 for reference sample, norms level 3).

using Cohen’s d on adjusted means for independent samples was significant for N (d = -1.28), large for C (d = .86), moderate for E (d = .63) and low for A (d = .41) and O (d = -0.20).

Since norms adapted to the sample were not available for the validity scale scores, means for PPM and NPM were compared to the American norms for men and women of similar ages and educational levels (Schinka et al., 1997). The means of the American sample were weighted for gender to have a similar representation of men and women to that of the French sample. Whereas no difference was found between the ENAC and the reference samples for NPM scores (Ms = 8.96 and 8.90 and SDs = 3.31 and 4.10 respectively), the PPM score of the ENAC sample (M = 23.10, SD = 4.50) was significantly higher than the one obtained for the reference sample (M = 20.67, SD = 4.10, confidence interval [13,31] at .05; Cohen’s d around .50; score higher than the mean score of the reference sample for 74% of the experimental sample). As predicted, a bias toward positive self-presentation was observed in the ENAC sample.

The relationships between the five personality dimensions of the NEO PI-R and the PPM-NPM index were investigated in the total sample (N = 974) by structural equation modeling using Mplus version 6.1 (Muthén & Muthén, 1987, 2011). We chose maximum likelihood parameter estimation because scores were roughly normally distributed. Comparative Fit Index (CFI), Non-Normed Fit Index (TLI) and root mean square error of approximation (RMSEA) were used as goodness-of-fit indicators to assess the models. As suggested by Hu and Bentler (1999), cut-off levels for determining model fit for continuous data were CFI > .95, TLI > .95, RMSEA < .06 and standard root mean square residual (SRMR) < .08.

The correlations between the five dimensions of the NEO PI-R and the PPM and NPM scores are presented in **Table 1**. The importance of the correlations between the five dimensions of the NEO PI-R suggests considering first a one factor model measured by N, C, E, A and O ($\chi^2 = 240.704$; ddl = 5; CFI = .760; TLI = .520; RMSEA = .220). Since this last model did not fit well the data, we estimated the parameters of a full two factor model obtained by exploratory structural equation modeling (ESEM). Inspection of the item loadings of this saturated

model led us to specify a confirmatory factor analysis (CFA) model with only one factor measured by C, N and A and correlated with E. In addition, O was allowed to be correlated with E and to freely co-vary with residual variances of C, N and A. This CFA model yielded acceptably high goodness of fit indices ($\chi^2 = 2.513$; ddl = 1; CFI = .998; TLI = .985; RMSEA = .039). In the final model, E, O and the factor measured by C, N and A were regressed on the PPM-NPM index. Results of this last analysis revealed a sound model fit ($\chi^2 = 11.802$; ddl = 6; p = .067; CFI = .997; TLI = .993; RMSEA = .032) in which the proportion variance explained in each observed variable was: C (.632), N (.593), E (.359), A (.172) and O (.000).

The hypothesized model schematically represented in **Figure 2** points to an interpretation of the common variance of C, N and A in terms of a “social competence in a selection context” that is adjustment to what is believed to be the best expected psychological profile (i.e., efficient, conscientious, emotionally stable, able to come into contact with others, ...) when one is a candidate for the entrance examination at a School of Civil Aviation. Results also highlight the specific status of E and O

Table 1. Correlations between the five NEO PI-R domain scores and NPM and PPM scores (N = 974).

	N	E	O	A	C	NPM	PPM
N	1						
E	-.281	1					
O	.256	.300	1				
A	-.302	.192	.051	1			
C	-.611	.360	-.135	.333	1		
NPM	.359	-.546	-.270	-.389	-.524	1	
PPM	-.655	.423	-.209	.209	.577	-.246	1

Note: N = Neuroticism; E = Extraversion; O = Openness; A = Agreeableness; C = Conscientiousness; NPM = Negative Presentation Management; PPM = Positive Presentation Management. All correlations are significant (p < .001) except the one between O and A (p = .113).

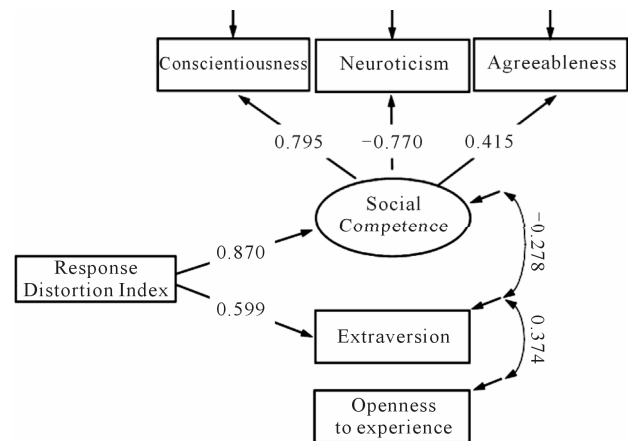


Figure 2. Final model (standardized estimations; all < .001) representing the relationships between Response Distortion (PPM-NPM) Index and the five personality dimensions of the NEO PI-R (Residual covariances between O and C, N and A were modeled but are not drawn in the figure; N = 974).

with respect to the dimension measured by C, N and A. Furthermore, examination of estimated parameters in the model reveals two significant structural regression paths: 1) the path from SDR index to “social competence” ($\beta = .870$, $SE = .014$) and 2) the path from SDR index to E ($\beta = .599$, $SE = .020$), the latter coefficient being significantly less important than the first one ($\Delta\chi^2 = 24.641$; $ddl = 1$; $p = .000$). Finally, no relationship was observed between SDR index and O.

Table 2 presents the findings relating to the 117 candidates observed on two separate occasions. The mean scores for N and C varied significantly between the two exam sessions. The mean N scores decreased while the mean C scores increased. However, based on Hopkins’s criteria ($<.60$), the effect size measured using Cohen’s d adjusted for intraclass correlation was low. All variables displayed very good test-retest reliability: N (.83), E (.87), O (.81), A (.83) and C (.89). An additional analysis aimed at testing the stability of N and C scores between the two exam sessions for each individual (Reliable Change Index; Maassen, 2004) shows that this hypothesis can only be rejected for 7.69% of candidates (1).

In accordance with the “ceiling” effect hypothesis, there was no significant mean difference between the two sessions: PPM: 23.02 (SD = 4.54) vs. 23.58 (SD = 4.09); NPM: 9.14 (SD = 2.74) vs. 9.33 (SD = 2.95). The two research validity scales also displayed good reliability: NPM (.67) and PPM (.75). Finally, the PPM-NPM index displayed high reliability (.73) but was found to be temporally slightly less stable than the N ($z = 2665$), E ($z = 3426$) and C ($z = 4766$) personality dimensions.

Discussion

The first objective of this study was to evaluate the effect of a selection context on NEO PI-R scores. As predicted, exam candidates in a selection context obtained lower scores for N compared to the reference sample but higher scores for C, E and A. As a result, such mean score changes in NEO PI-R render necessary the construction of norms specific to candidates taking the entrance examination to the French National School of Civil Aviation (2). These findings are consistent with the results of studies that compared a “motivating” context with a standard context (Ones et al., 1996; Sandal et al., 2005). As noted by Hogan (1991), the findings suggest that “well-adjusted people have positively biased self-images; consequently, well-adjusted people tend to ignore minor criticisms, discount their failures, avoid negative thoughts and expect to succeed in most of their undertakings”. More surprisingly, the exam can-

didates claimed to be less open to experience than individuals in the reference sample, although the significance of this finding is low given the very small effect size. However, the measurement of O appears once again (Griffin, Hesketh and Grayson, 2004) to be the most reliable measurement—whether intentionally or not—among the dimensions of the five-factor model.

Two sets of results suggest that the findings relating to candidates who took the NEO PI-R twice remained stable over time. Intraclass correlations were high (ranging from .66 to .81). The results of the analysis of mean change combined with the results of the analysis of individual change also point to temporal stability in the five domain scores. Thus, it is logical to think that construct validity of the personality scales remained intact across the 2 administrations of the NEO PI-R. Therefore, the hypothesis of a maximum-threshold SDR in the first exam session seems more likely to account for the results obtained for the experimental sample than the “forced” trait hypothesis of an increase of SDR between the two exam sessions.

The final objective of this study was to assess the relationships between the domains of the NEO PI-R and SDR measured by the PPM-NPM index. Two key points need to be emphasized. The first point concerns the significant correlations between 4 of the 5 NEO PI-R factors. While it is not altogether new (Reid-Seiser & Fritzsche, 2001), this finding challenges the hypothesis of an orthogonal and invariant factor structure of the NEO PI-R in a selection context (Schmit & Ryan, 1993; Bernard and Walsh, 2004; see also Marshall et al., 2005). Another important finding concerns the value of the estimated correlation between SDR measured by PPM-NPM and the latent variable expressing the common variance of C, N and A. This correlation (which was equally significant in both exam sessions) indicates that the validity scales developed by Schinka and colleagues measure a form of response bias that may be closely linked to candidate responses in certain contexts (for example, in this case, 63% of the variance of C and 59% of the variance of N on the whole sample). The presence of a high level of SDR in data collected in a selection context therefore merits attention (McFarland, 2003), particularly since a recent study conducted among airplane pilots indicated that it may be a negative predictor of professional performance in stressful situations (Sandal, Musson, Helmreich, & Gravidal, 2005).

Conclusion

This study demonstrates the importance of response distortion on the NEO PI-R in real-world selection settings. The

Table 2.

Mean and standard deviation of raw NEO PI-R domain scores in the 1st exam session (T1) and 2nd exam session (T2); mean change, intraclass correlation and effect size between T1 and T2; *: $p < .05$ ($N = 117$).

		Neuroticism	Extraversion	Openness to experience	Agreeableness	Conscientiousness
Mean (standard deviation)	T1	60.62 (18.09)	122.21 (17.17)	110.65 (15.49)	128.26 (13.93)	133.91 (19.45)
	T2	56.64 (16.31)	121.76 (15.43)	108.42 (16.49)	128.43 (14.35)	136.38 (17.82)
Mean change		-3.98*	-0.44	-2.23	.16	2.47*
Confidence interval (95%)		[-6.32; -1.65]	[-2.46; 1.57]	[-4.57; 0.11]	[-1.82; 2.14]	[.34; 4.60]
Intraclass correlation		.726	.773	.682	.708	.805
Effect size (adjusted Cohen’s d)		-.44	-.06	-.25	.02	.30

findings indicate that the tendency of candidates to give self-enhancing biased responses can be assessed using the validity scales developed by Schinka and colleagues. However, the issue of the conceptual status of SDR remains unresolved (see for example Morey et al., 2002). Viewed as a substantial personality trait without any effect on the criterion-related validity of the NEO PI-R (Hogan, Barrett, & Hogan, 2007), response distortion may imply a form of context adjustment that can be easily measured using the PPM-NPM index. If viewed as circumstantial, the PPM-NPM index may be useful for identifying invalid approaches provided the threshold at which response distortion biases NEO PI-R scores is determined.

Notes

1) N: significant decrease (.05) in 4 individuals but significant increase in 5 individuals. C: significant increase in 4 individuals but significant decrease in 4 individuals.

2) Norms available upon request from the first author.

REFERENCES

- Bagby, R. M., & Marshall, M. B. (2003). Positive Impression Management and its influence on the NEO PI-R: A comparison of analog and differential prevalence group designs. *Psychological Assessment, 15*, 333-339. doi:10.1037/1040-3590.15.3.333
- Ballenger, J. F., Caldwell-Andrews, A., & Baer, R. A. (2001). Effects of positive impression management on the NEO Personality Inventory-Revised in a clinical population. *Psychological Assessment, 13*, 254-260. doi:10.1037/1040-3590.13.2.254
- Barrick, M. R., & Mount, M. K. (1991). The Big Five personality dimensions and job performance: A meta-analysis. *Personnel Psychology, 44*, 1-26. doi:10.1111/j.1744-6570.1991.tb00688.x
- Bernard, L. C., & Walsh, R. P. (2004). Socially desirable and non-purposeful responding on the NEO PI-R. *Counseling and Clinical Psychology Journal, 1*, 4-16.
- Berry, D. T. R., Bagby, R. M., Smerz, J., Rinaldo, J. C., Caldwell-Andrews, A., & Baer, R. A. (2001). Effectiveness of NEO PI-R research validity scales for discriminating analog malingering and genuine psychopathology. *Journal of Personality Assessment, 76*, 496-516. doi:10.1207/S15327752JPA7603_10
- Caldwell-Andrews, A., Baer, R. A., & Berry, D. T. R. (2000). Effects of response sets on NEO-PI R scores and their relation to external criteria. *Journal of Personality Assessment, 74*, 472-488. doi:10.1207/S15327752JPA7403_10
- Costa, P. T., & McCrae, R. R. (1992). *NEO PI-R professional manual*. Florida: Psychological Assessment Resources Inc. (version française, ECPA, 1998, 2003).
- Costa, P. T., & McCrae, R. R. (1997). Stability and change in personality assessment: The NEO personality inventory in the year 2000. *Journal of Personality Assessment, 68*, 86-94. doi:10.1207/s15327752jpa6801_7
- Costa, P. T., McCrae, R. R., & Rolland, J. P. (1998-2004). *Manuel de l'inventaire de personnalité NEO PI-R*. Paris: ECPA.
- Ellingson, J. E., Smith, D. B., & Sackett, P. R. (2001). Investigating the influence of social desirability on personality factor structure. *Journal of Applied Psychology, 86*, 122-133. doi:10.1037/0021-9010.86.1.122
- Griffin, B., Hesketh, B., & Grayson, D. (2004). Applicants faking good: evidence of item bias in the NEO PI-R. *Personality and Individual Differences, 36*, 1545-1558. doi:10.1016/j.paid.2003.06.004
- Hogan, R. (1991). Personality and personality measurement. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology* (Vol. 2, pp. 873-919). Palo Alto, CA: Consulting Psychologists Press.
- Hogan, R. (2005). In defense of personality measurement: New wine for old whiners. *Human Performance, 18*, 331-341. doi:10.1207/s15327043hup1804_1
- Hogan, J., Barrett, P., & Hogan, R. (2007). Personality measurement, faking, and employment selection. *Journal of Applied Psychology, 92*, 1270-1285. doi:10.1037/0021-9010.92.5.1270
- Hogan, R., Hogan, J., & Robert, B. W. (1996). Personality measurement and employment decisions: Questions and answers. *American Psychologist, 51*, 469-477. doi:10.1037/0003-066X.51.5.469
- Hogan, R., De Fruyt, F., & Rolland, J.-P. (2006). Validité et intérêt des méthodes d'évaluation de la personnalité à des fins de sélection: une perspective de psychologie appliquée aux problématiques des entreprises. *Psychologie Française, 51*, 245-264. doi:10.1016/j.psf.2005.12.007
- Judge, T. J., Higgins, C. A., Thoresen, C. J., & Barrick, M. R. (1999). The big five personality traits, general mental ability, and career success across the life span. *Personnel Psychology, 52*, 621-652. doi:10.1111/j.1744-6570.1999.tb00174.x
- Maassen, G. H. (2004). The standard error in the Jacobson and truxel reliable change index: The classical approach to the assessment of reliable change. *Journal of the International Neuropsychological Society, 10*, 888-893. doi:10.1017/S1355617704106097
- 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*, 379-384. doi:10.1037/1040-3590.17.3.379
- McFarland, L. A. (2003). Warning against faking on a personality test: Effects on applicant reactions and personality test scores. *International Journal of Selection and Assessment, 11*, 265-276. doi:10.1111/j.0965-075X.2003.00250.x
- McFarland, L. A., & Ryan, A. M. (2000). Variance in faking across non-concomitant measures. *Journal of Applied Psychology, 85*, 812-821. doi:10.1037/0021-9010.85.5.812
- McFarland, L. A., Wiechmann, D., & Chandler, C. W. (2001). *Using appropriateness fit to identify faking on a personality test*. Paper presented at the Society for Industrial and Organizational Psychology conference, April 2001, San Diego, CA.
- Morey, L. C., Quigley, B. D., Sanislow, C. A., Skodol, A. E., McGlashan, T. H., Shea, M. T., Stout, R. L., Zanarini, M. C., & Gunderson, J. G. (2002). Substance or style? An investigation of the NEO-PI-R validity scales. *Journal of Personality Assessment, 79*, 583-599. doi:10.1207/S15327752JPA7903_11
- Mount, M. K., & Barrick, M. R. (1998). Five reasons why the "big five" article has been frequently cited. *Personnel Psychology, 51*, 849-858. doi:10.1111/j.1744-6570.1998.tb00743.x
- Mount, M. K., Barrick, M. R., & Stewart, G. L. (1998). Personality predictors of performance in jobs involving interaction with others. *Human Performance, 11*, 145-166. doi:10.1080/08959285.1998.9668029
- Muthén, L. K., & Muthén, B. O. (1987, 2012). *Mplus: Statistical analysis with latent variables* (Version 6.12) [Computer software]. Los Angeles: Author.
- Ones, D. S., Viswesvaran, C., & Reiss, A. D. (1996). Role of social desirability in personality testing for personnel selection: The red herring. *Journal of Applied Psychology, 81*, 660-679. doi:10.1037/0021-9010.81.6.660
- Ones, D. S., Viswesvaran, C., Dilchert, C., & Deller, J. (2006) (Eds.). Special issue: Considering response distortion in personality measurement for industrial, work and organizational psychology research and practice. *Psychology Science, 48*, 207-297.
- Paulhus, D. L. (1984). Two-component models of socially desirable responding. *Journal of Personality and Social Psychology, 46*, 598-609. doi:10.1037/0022-3514.46.3.598
- Paulhus, D. L. (1991). Measurement and control of response bias. In J. P. Robinson, J. R. Shaver, & L. S. Wrightsman (Eds.), *Measure of personality and social psychological attitudes* (pp. 17-59). San Diego, CA: Academic Press.
- Paulhus, D. L., & Reid, D. B. (1991). Enhancement and denial in socially desirable responding. *Journal of Personality and Social Psychology, 60*, 307-317. doi:10.1037/0022-3514.60.2.307
- Paulhus, D. L., Bruce, M. N., & Trapnell, P. D. (1995). Effects of

- self-presentation strategies on personality profiles and their structure. *Personality and Social Psychology Bulletin*, 21, 100-108. doi:10.1177/0146167295212001
- Piedmont, R. L., McCrae, R. R., Riemann, R., & Angleitner, A. (2000). On the invalidity of validity scales: Evidence from self-reports and observer ratings in volunteer samples. *Journal of Personality and Social Psychology*, 78, 582-593. doi:10.1037/0022-3514.78.3.582
- Reid-Seiser, H. L., & Fritzsche, B. A. (2001). The usefulness of the NEO PI-R positive presentation management scale for detecting response distortion in employment contexts. *Personality and Individual Differences*, 31, 639-650. doi:10.1016/S0191-8869(00)00168-9
- Rolland, J.-P., Parker, W. D., & Stumpf, H. (1998). A psychometric examination of the French translations of the NEO-PI-R and NEO-FFI. *Journal of Personality Assessment*, 7, 269-291. doi:10.1207/s15327752jpa7102_13
- Rosse, J. G., Stecher, M. D., Levin, R. A., & Miller, J. L. (1998). The impact of response distortion on preemployment personality testing and hiring decisions. *Journal of Applied Psychology*, 83, 634-644. doi:10.1037/0021-9010.83.4.634
- Salgado, J. F. (1997). The five factor model of personality and job performance in the European Community. *Journal of Applied Psychology*, 82, 30-43. doi:10.1037/0021-9010.82.1.30
- Sandal, G. M., Musson, D., Helmreich, R. L., & Gravdal, L. (2005). Social desirability bias in personality testing: Implications for astronaut selection. *Acta Astronautica*, 57, 634-641. doi:10.1016/j.actaastro.2005.03.011
- Schinka, J. A., Kinder, B. N., & Kremer, T. (1997). Research validity scales for the NEO PI-R: Development and initial validation. *Journal of Personality Assessment*, 68, 127-138. doi:10.1207/s15327752jpa6801_10
- Schmit, M. J., & Ryan, A. M. (1993). The Big Five in personnel selection: Factor structure in applicant and non applicant populations. *Journal of Applied Psychology*, 78, 966-974. doi:10.1037/0021-9010.78.6.966
- Topping, G. D., & O'Gorman, J. G. (1997). Effects of faking set on validity of the NEO-FFI. *Personality and Individual Differences*, 23, 117-124. doi:10.1016/S0191-8869(97)00006-8
- Viswesvaran, C., & Ones, D. S. (1999). Meta-analyses of fakability estimates: Implications for personality measurement. *Educational and Psychological Measurement*, 59, 197-210. doi:10.1177/00131649921969802
- Wiggins, J. S. (1964). Convergences among stylistic response measures from objective personality tests. *Educational and Psychological Measurement*, 24, 551-562. doi:10.1177/001316446402400310
- Young, M. S., & Schinka, J. A. (2001). Research validity scales for the NEO-PI-R: Additional evidence for reliability and validity. *Journal of Personality Assessment*, 76, 412-420. doi:10.1207/S15327752JPA7603_04