



## Predicting creativity and academic success with a “Fake-Proof” measure of the Big Five

Jacob B. Hirsh, Jordan B. Peterson\*

Department of Psychology, University of Toronto, Sidney Smith Hall, 100 St. George Street, Toronto, Ont., Canada M5S 3G3

### ARTICLE INFO

*Article history:*  
Available online 27 April 2008

*Keywords:*

Performance prediction  
Big Five  
Personality  
Psychometrics  
Biased responding

### ABSTRACT

Self-report measures of personality appear susceptible to biased responding, especially when administered in competitive environments. Respondents can selectively enhance their positive traits while downplaying negative ones. Consequently, it can be difficult to achieve an accurate representation of personality when there is motivation for favourable self-presentation. In the current study, we developed a relative-scored Big Five measure in which respondents had to make repeated choices between equally desirable personality descriptors. This measure was contrasted with a traditional Big Five measure for its ability to predict GPA and creative achievement under both normal and “fake good” response conditions. While the relative-scored measure significantly predicted these outcomes in both conditions, the Likert questionnaire lost its predictive ability when faking was present. The relative-scored measure thus proved more robust against biased responding than the Likert measure of the Big Five.

© 2008 Elsevier Inc. All rights reserved.

### 1. Introduction

Prediction of real-world performance outcomes is one of the primary goals of psychometric assessment. In the study of personality, this goal has been significantly advanced by the emergence of the “Big Five” model of personality structure (Goldberg, 1992; McCrae & John, 1992). The Big Five model describes personality variation across five broad trait domains: Neuroticism, Extraversion, Openness, Agreeableness, and Conscientiousness (Costa & McCrae, 1992). These personality dimensions appear to be valid cross-culturally (McCrae & Costa, 1997), are relatively stable across the lifespan (Costa & McCrae, 1997) and can be reliably used to predict real-world outcomes (for a review, see Ozer & Benet-Martinez, 2006).

The broad domain trait of Conscientiousness in particular has emerged as a significant predictor of academic success, above and beyond differences in cognitive ability (Goff & Ackerman, 1992). Individuals who score highly on scales of Conscientiousness are hard working, organized, efficient, and self-disciplined. As might be expected, these individuals are more likely to succeed in the academic realm. Recent studies suggest that Conscientiousness accounts for 12–25% of the variance in academic performance (Gray & Watson, 2002; Higgins, Peterson, Pihl, & Lee, 2007). In one such study, composite measures of self-discipline, a construct apparently related to trait Conscientiousness, were twice as effective as IQ at predicting academic performance (Duckworth & Seligman, 2005). Conscientiousness is also the best single personality predictor of workplace performance across a variety of job categories (Barrick & Mount, 1991; Hertz & Donovan, 2000).

After Conscientiousness, Emotional Stability (the inverse of Neuroticism) is the best Big Five predictor of workplace performance (Salgado, 1997). Individuals high on Emotional Stability are secure, confident, and not easily disturbed. Such individuals may have an easier time accomplishing difficult tasks than those who score lower on this trait. Low scorers tend to be

\* Corresponding author. Fax: +1 416 978 4811.

E-mail address: [jordanbpeterson@yahoo.com](mailto:jordanbpeterson@yahoo.com) (J.B. Peterson).

anxious, depressed, and worrisome, which makes them more susceptible to emotional exhaustion. Emotional Stability is also a good predictor of job satisfaction and organizational commitment (Thoresen, Kaplan, Barsky, Warren, & de Chermont, 2003).

Extraverts, who are assertive, enthusiastic, and sociable, are good candidates for team-based activities (Barrick, Mount, & Judge, 2001). Their high levels of positive affect and enthusiasm also help make Extraverts effective performers in leadership positions (Judge, Bono, Ilies, & Gerhardt, 2002) and accounts for their comparatively high levels of job satisfaction and sense of personal accomplishment (Thoresen et al., 2003). Trait Agreeableness, like Extraversion, is also a good predictor of team-based work performance (Barrick et al., 2001). Highly agreeable individuals are warm, considerate, trusting, and empathic, in contrast to their tough-minded, selfish, and hostile counterparts, at the low end of the spectrum. When combined with Extraversion, high Agreeableness also predicts a transformational leadership style, which is associated with increased commitment, satisfaction, and motivation among group members (Judge & Bono, 2000).

Openness to Experience, finally, has been linked to higher levels of creative achievement (Carson, Peterson, & Higgins, 2005). Open people are curious, imaginative, and willing to entertain new ideas. People who score highly on this dimension have a greater tendency towards cognitive exploration and also manifest higher levels of cognitive flexibility and divergent thinking (DeYoung, Peterson, & Higgins, 2005; McCrae, 1987). Neuropsychological investigations suggest that individual differences in Openness are related to dopaminergic function in the prefrontal cortex (DeYoung et al., 2005). The increased cognitive flexibility afforded by dopaminergic activity is thought to underlie the generation of novel associations central to the creative process (Eysenck, 1995). Scores on personality questionnaires measuring Openness thus appear to be significant predictors of an individual's creative capacity.

Despite the frequently reported predictive utility of questionnaires assessing these Big Five traits, their implementation in real-world selection processes can be hindered, at least in some circumstances, by the presence of biased responding. When individuals are asked to rate themselves on a series of personality dimensions, they sometimes exaggerate their positive and downplay their negative qualities (Paulhus, 2002). This tendency presents a potentially serious problem in the domain of performance prediction, because respondents may be highly motivated to make a good impression. A large literature now shows that motivated individuals are able to fake their scores on a five factor personality scale when attempting to do so (e.g., Furnham, 1997; Viswesvaran & Ones, 1999).

Although there has been some debate in the literature as to whether response bias is a problem in real-world assessment contexts (e.g., Barrick & Mount, 1996; Ones, Viswesvaran, & Reiss, 1996), a recent meta-analysis of job applicant faking on personality questionnaires has demonstrated that applicants score significantly higher than non-applicants on Extraversion, Conscientiousness, emotional stability, and openness (Birkeland, Manson, Kisamore, Brannick, & Smith, 2006). Furthermore, these traits are differentially biased, with Conscientiousness and Emotional Stability, the two most important predictors of real-world success, being inflated more than the other dimensions. Higher scores on these traits may therefore indicate greater levels of self-presentation and biased responding rather than an accurate description of personality. Higgins et al. (2007) demonstrated, for example, that self-rated Conscientiousness predicted self but not manager rated job performance, indicating the presence of inflationary bias across outcome and predictor variables. It thus becomes difficult to distinguish individuals who are authentically high on positive traits from those who are simply trying to present themselves in a favourable light. Consequently, personality questionnaires can lose a substantial portion of their predictive validity when there is an incentive for respondents to make a good impression (Mueller-Hanson, Heggstad, & Thornton, 2003; Rosse, Stecher, Miller, & Levin, 1998).

One approach to resolving this issue has been to administer tests of socially desirable responding, assessing the extent to which respondents are willing to admit to undesirable traits or behaviours. These tests originated as "lie" or "response bias" scales, and were designed to detect individuals who fake good while completing personality questionnaires (Eysenck, 1994; Furnham, 1986; Paulhus, 1991). These scales include the K scale of the MMPI (Block, 1965), Edward's Social Desirability Scale (1953; 1957), Sackeim and Gur's Self-Deception Questionnaire (SDQ, 1978), the Marlowe-Crowne Social Desirability Scale (MCSD, Crowne & Marlowe, 1960; Reynolds, 1982), Byrne's Repression-Sensitization Scale (Byrne & Bounds, 1964), Allaman, Joyce, and Crandall's (1972) Censure-Avoidance questionnaire, the Lie Scale in Eysenck's Personality Questionnaire (EPQ, Eysenck, Eysenck, & Barrett, 1985), Paulhus' Balanced Inventory of Desirable Responding (BIDR, 1991), and the NEO Research Validity Scales (Schinka, Kinder, & Kremer, 1997).

Despite their purported function, these bias scales appear to be associated with more genuine personality variance than response bias, particularly when responses are anonymous (Borkenau & Amelang, 1985; DeYoung, Peterson, & Higgins, 2002; McCrae & Costa, 1983; Piedmont, McCrae, Riemann, & Angleitner, 2000). Although social desirability measures appear to be correlated with discrepancies between self-reports and observer ratings of personality (Paulhus & John, 1998), controlling for them statistically tends to decrease the correlation between self-reports and observer ratings (Borkenau & Amelang, 1985; Piedmont et al., 2000). Furthermore, controlling for socially desirable responding does not appear to improve criterion-related validities of personality predictors of job performance (Ellingson, Sackett, & Hough, 1999; Hough, Eaton, Dunnette, & Kamp, 1990; Ones et al., 1996). In a recent meta-analysis, neither measures of conscious or unconscious response bias were able to improve the predictive validity of their accompanying personality questionnaires (Li & Bagger, 2006). In fact, high scores on the NEO PI-R Positive Presentation Management scale actually correlate positively with workplace productivity, even though the former is also highly correlated with Self-Deceptive Enhancement as measured by the BIDR (Reid-Seiser & Fritzsche, 2001). Overall, the ability to fake good on personality questionnaires appears to be unrelated to scores on measures of social desirability and response bias, which themselves appear to reflect genuine variance in personality (Mersman

& Shultz, 1998). Although the failure of these scales to improve predictive validity has been cited as evidence for the lack of response bias in job applicant samples (e.g., Ones et al., 1996), an alternative explanation is that these scales are simply ineffective at predicting the degree to which one's self-reported personality is biased in a positive direction. Indeed, there is substantive evidence suggesting that motivated responding on personality questionnaires is a real problem that poses a significant threat to predictive validity (Birkeland et al., 2006).

To address the problem of biased responding and the lack of success in detecting and controlling for this tendency, we sought to improve the predictive validity of the personality assessment instruments themselves. Specifically, the current study involved the construction and validation of a Big Five personality questionnaire that could prove more resistant to biased responding. Personality measures were created using a variety of comparative scaling techniques, in which each trait domain was scored relative to all the others, rather than being scored separately. In the currently most common non-comparative test format, respondents rate their agreement with a variety of descriptions using a scale from 1 (*Strongly Disagree*) to 5 (*Strongly Agree*) (e.g., Costa & McCrae, 1992). In principle, this allows individuals to inflate their scores by selectively ranking themselves higher on all the positive dimensions and lower on all the negative ones. In the current study, we employed three questionnaire formats designed to prevent this type of self-enhancement by requiring respondents to choose between equally valued descriptors. Previous research suggests that these relative-scored, or ipsative, survey formats may be less susceptible to distortion than their Likert scored counterparts (Christiansen, Burns, & Montgomery, 2005; Jackson, Wroblewski, & Ashton, 2000). If such formats are indeed more effective at reducing response distortion, they may produce a better estimate of an individual's personality than traditional Likert format questionnaires (Baron, 1996). While previous studies have applied relative-scored techniques to single constructs (e.g., Conscientiousness, Integrity), the current study extends this research by creating and testing a relative-scored measure that assesses each of the Big Five dimensions.

The new questionnaire was compared with the traditional Likert format in its ability to predict performance in two independent domains. A "fake good" response condition was also included to test the relative-scored measures' effectiveness at reducing biased responding and maintaining predictive validity under explicit faking conditions. Analog faking designs such as this one appear appropriately comparable to real-world situations in which response distortion is likely (Bagby & Marshall, 2003). Indeed, analog faking studies tend to produce even more response distortion than is found in real-world selection procedures (Birkeland et al., 2006; Viswesvaran & Ones, 1999). Thus, the resistance of a questionnaire to explicit fake good instructions can be considered a strong indicator of the likely ecological validity of that questionnaire.

The first performance domain to be predicted was academic success, indexed by participants' grade point average (GPA). In order to obtain a good GPA, students must sustain high levels of academic performance over an extended period of time. This requires the continued demonstration of multiple abilities in a variety of domains, all within a rapidly changing environment. The diverse raters, contexts, and content areas that contribute to one's overall GPA make it a balanced measure of an individual's academic performance. Additionally, the established correlations between Conscientiousness and GPA (e.g., Goff & Ackerman, 1992; Higgins et al., 2007) make it suitable for a test of the predictive validity of the relative-scored personality questionnaires. The second targeted area of performance was the domain of creative achievement. As discussed above, creativity is most strongly related to the personality trait of Openness (Carson et al., 2005). If the relative-scored questionnaire is effective at eliminating biased responding, then the relative-scored Openness dimension should be a better predictor of creativity than the standard Openness scale.

We hypothesized that the relative-scored Big Five variant would be as valid as the traditional Big Five measure for predicting performance in both academic and creative domains and significantly better in the fake good condition. Specifically, we expected both BFI and Relative-Scored (RS) Conscientiousness to predict GPA in the normal response condition, but only RS Conscientiousness to predict GPA in the fake good condition. Similarly, BFI and RS Openness were both expected to predict creativity in the normal condition, but only RS Openness was expected to predict creativity in the fake good condition.

## 2. Methods

### 2.1. Participants

We tested 205 undergraduate students from the University of Toronto (59 male, 146 female) ranging in age from 18 to 35 ( $M = 21$ ,  $SD = 3.0$ ). Participants were recruited through campus flyers advertising the experiment, and were paid \$15 for their time. Because the experiment was conducted online, there were no limits to the number of simultaneous respondents. Nine participants did not complete all of the questionnaires, leaving us with partial data for these individuals. Removing their data completely did not affect our results, so the available responses were kept in the analyses.

### 2.2. Materials

The relative-scored personality questionnaire employed in this study is comprised of three different comparative scaling methods: paired comparisons, forced-choice, and rank order techniques. The questionnaire was constructed using items from the International Personality Item Pool (IPIP), a public-domain resource for obtaining questionnaire items validated against commonly used scales (Goldberg, 1999; International Personality Item Pool, 2005). The personality descriptors used in the current study were taken from the IPIP five factor questionnaires, including the IPIP NEO, BFI, and the Big Five items

from the Seven Factor questionnaire. Items from these scales were combined to create a pool of descriptors from each of the five dimensions, which were then used as factor markers in our relative-scored methods. Each of the relative-scored scales was constructed to have an equal number of positive and negative items from all five of the trait dimensions.

The first relative-scored method used in our questionnaire was [Thurstone's \(1927\)](#) paired comparisons technique. In this survey format, respondents have to make a series of choices between two personality descriptions. During each question, the participant is asked to choose the most appropriate self-description from two different trait categories (e.g., “Rarely get irritated” vs. “Am full of ideas” contrasts Emotional Stability with Openness, respectively). In a single comparison block, one item is taken from each of the five dimensions. The item from each of the dimensions is then compared to an item from each of the others, leading to 10 comparisons per block. After 100 of these comparisons are made, all five dimensions end up being compared to each of the other ones ten different times. Half of the blocks compare two positive items with each other, while the remaining blocks compare two negative items with each other. Altogether, ten unique items are presented from each of the five dimensions. Domain scores are calculated by summing the number of times that positive items from a given dimension are chosen and subtracting the number of times that negative items from that dimension are chosen. Raw scores can have a potential range of  $-20$  to  $20$ .

In the forced-choice method, the Big Five markers were split into five groups of positive items and five groups of negative items. In the positive groups, respondents had to select the 10 most appropriate personality descriptions from a list of 20 available options. Each group contained four items from each of the five trait dimensions. In the negative groups, only 5 choices were required from a list of 20 items. The difference between positive and negative item groups was intended to make it easier for the participants to choose negative self-descriptions. A total of 200 unique items were included in this section, balanced between each of the five trait dimensions. Domain scores were again calculated by summing the number of positive items selected from each dimension, and subtracting the number of negative items. The potential raw scores range from  $-20$  to  $20$ .

In the rank order method, participants were presented with five personality descriptions (one from each trait domain) and were asked to rank them with regards to how well they applied to their own personality. In total, twenty groups of five were presented, with ten groups of positive items and ten groups of negative items. Altogether, 100 unique descriptors were displayed. Items were reverse-scored for the order that they were chosen (i.e., items ranked as most applicable were given a 5, and items that were least applicable were given a 1). Domain totals were calculated by summing the positive scores within each dimension and subtracting the negative scores. The potential raw scores ranged from  $-40$  to  $40$ . Altogether, the combined administration time for the three relative-scored methods was approximately 35 min.

For a traditional Likert personality questionnaire, we administered the Big Five Inventory ([John, Donahue, & Kentle, 1991](#)). This questionnaire features 44 items across the five trait domains, and requires respondents to rate their agreement with a variety of personality descriptions on a 5-point scale (e.g., “I see myself as someone who is a reliable worker”).

As a measure of creative achievement, we employed the Creative Achievement Questionnaire (CAQ). The CAQ requires participants to indicate the extent to which their creative achievements have been recognized across a variety of domains (such as writing, science, or visual arts). It is a reliable measure of creative accomplishments, and is characterized by good convergent, discriminant, and predictive validity ([Carson et al., 2005](#)).

### 2.3. Design

The study employed a mixed within-subjects and between-groups design. The primary within-subjects independent variable was the questionnaire format being used to predict academic performance and creative achievement (Likert vs. relative-scored). The between-groups independent variable was the response condition of the participant (normal vs. fake good). The dependent variables for both analyses were students' university grades and their CAQ scores. To prevent order effects, we counterbalanced the presentation order of the Likert and relative-scored personality questionnaires.

### 2.4. Procedure

The experiment took approximately one hour to complete, and was administered entirely over the internet via online survey software ([Select Survey ASP Advanced, 2005](#)). Previous research suggests equivalence between online and paper administration of personality questionnaires ([Chuah, Drasgow, & Roberts, 2006](#)). Upon responding to the advertisements for the study, participants were sent a username and password to login to the survey site. Participants were free to complete the questionnaires from any computer with internet access. The initial web page presented the participants with links to all of the questionnaires that needed to be completed for the study. The participants were instructed to work through these questionnaires one at a time, taking short breaks between them. After completing each questionnaire, they were returned to the initial index page where they could continue to the next section. Before completing any of the questionnaires, however, participants were required to agree to the online informed consent form. This form also requested the consent of the participant to allow access to their academic transcripts for the purposes of the study. These were obtained directly from the office of the faculty registrar to ensure accuracy.

Once the participants agreed to participate in the study, they completed a brief demographics questionnaire, followed by the four methodological variants of the Big Five measure. The presentation order was counterbalanced across participants, with half receiving the Likert questionnaire first and the other half receiving the relative-scored variants first. The partici-

**Table 1**  
Intercorrelations for Big Five, CGPA, CAQ, and Years English in the normal condition

Variables	1	2	3	4	5	6	7	8	9	10	11	12
1 BFI Extra	—											
2 BFI Agree	-.04	—										
3 BFI Consc	.10	.39**	—									
4 BFI EmStab	.28**	.27**	.16	—								
5 BFI Open	.37**	.04	-.04	-.02	—							
6 RS Extra	.80**	-.25*	-.22*	.05	.32**	—						
7 RS Agree	-.12	.33**	-.16	-.17	-.10	-.12	—					
8 RS Consc	-.38**	.02	.62**	-.25*	-.36**	-.51**	-.29**	—				
9 RS EmStab	-.30**	.20	-.08	.61**	-.32**	-.42**	-.14	-.14	—			
10 RS Open	-.04	-.24*	-.34**	-.31**	.56**	.05	-.19	-.30**	-.37**	—		
11 CGPA	-.01	.04	.29**	-.05	-.09	-.14	-.24*	.32**	-.02	-.02	—	
12 CAQ	.02	.09	-.01	-.03	.35**	-.03	.02	-.10	-.11	.29**	.01	—
13 Years Eng	.05	.09	.25*	.15	.13	-.13	.00	.11	-.05	.09	-.08	.21*

\*  $p < .05$ , two-tailed.

\*\*  $p < .01$ , two-tailed.

pants were also randomly assigned to one of two response conditions. In the normal response condition, participants were asked to answer the questionnaires honestly and accurately. In the fake good condition, participants were told to answer the personality questionnaires as though they were applying for a job and wanted to make the best impression possible. Previous research has demonstrated that this type of faking instruction strongly elevates individual trait scores, especially in the domain of Conscientiousness (e.g., Paulhus, Bruce, & Trapnell, 1995; Viswesvaran & Ones, 1999). The fake good condition thus allowed us to provisionally test the ability of the relative-scored measure to attenuate the effects of intentionally biased responding. After completing the personality questionnaires, participants in the fake good condition were given a manipulation check to ensure that they had faked their responses in a positive manner. Participants in both groups were then told to complete the CAQ honestly and accurately, and to inform us when they had completed all of the surveys. After completing the surveys, participants were fully debriefed about the study and sent a \$15 Interac email money transfer. Student grades were collected through the university, with the written consent of the participants, in order to conduct the analyses. The study in total was approved by the University of Toronto's Institutional Research Board.

### 3. Results

#### 3.1. Deriving relative-scored values

Results from the three relative-scored methods were highly consistent with each other in the normal condition, with average within-trait correlations ranging from .78 (Agreeableness) to .88 (Extraversion) across the three measures. These within-trait correlations dropped in the fake good condition, with values ranging from .64 (Emotional Stability) to .75 (Openness). Across both conditions, the within-trait correlations were similar to the reliabilities that are commonly observed with Big Five scales (Viswesvaran & Ones, 2000). Because scores from the three relative-scored methods were highly correlated with each other, and no single technique proved significantly superior to the others across all of our comparisons, we combined all three into a single composite measure. Combining ipsative scores derived from a variety of measures also provides additional psychometric benefits, as discussed later. Each composite Big Five score was obtained by calculating the mean of the standardized domain values for the paired comparison, forced-choice, and rank order methods. In combining these methods, we hoped to develop a relative-scored measure with maximal breadth and robustness. Any reference to the relative-scored questionnaire in the following results refers to this composite measure.

#### 3.2. Characteristics of the relative-scored questionnaire

Tables 1 and 2 present the intercorrelations between dimensions of the Likert and relative-scored (RS) Big Five questionnaires. In the normal response condition, each of the relative-scored Big Five dimensions correlated significantly with their Likert counterparts. These correlations had an upper range of .80 for Extraversion and a lower range of .33 for Agreeableness. As expected, the correlations dropped in the fake condition, ranging from .48 (Openness) to .19 (Agreeableness).<sup>1</sup> A comparison of the average relative-scored/Likert scale correlations from the normal ( $r = .61$ ,  $n = 490$ ) and fake good ( $r = .36$ ,  $n = 470$ ) conditions revealed a significant difference between the two ( $z = 5.13$ ,  $p < .01$ ).

<sup>1</sup> The relatively low intercorrelations obtained for trait Agreeableness may be a result of using the IPIP items, as lower intercorrelations for this trait are also obtained with other IPIP-derived measures (DeYoung, Quilty, & Peterson, 2007). According to the IPIP website, the IPIP Agreeableness items have the lowest correlations with Goldberg's factor markers ( $r = .54$ ). When compared with longer Big Five measures such as the NEO-PI-R, however, the IPIP Agreeableness items demonstrate better convergent validity ( $r = .77$ ).

**Table 2**  
Intercorrelations for Big Five, CGPA, CAQ, and Years English in the fake good condition

Variables	1	2	3	4	5	6	7	8	9	10	11	12
1 BFI Extra	—											
2 BFI Agree	.38**	—										
3 BFI Consc	.51**	.67**	—									
4 BFI EmStab	.62**	.62**	.66**	—								
5 BFI Open	.36**	.35**	.47**	.38**	—							
6 RS Extra	.44**	-.14	-.07	.08	-.01	—						
7 RS Agree	-.27**	.19	-.19	-.24*	-.18	-.26**	—					
8 RS Consc	.03	.08	.31**	.08	-.16	-.32**	-.33**	—				
9 RS EmStab	.10	.09	.00	.38**	-.18	.00	-.25*	-.14	—			
10 RS Open	-.16	-.22*	-.01	-.15	.48**	-.25*	-.35**	-.18	-.32**	—		
11 CGPA	-.16	-.09	.09	-.06	.06	-.27**	-.16	.24*	.11	.13	—	
12 CAQ	-.01	-.10	-.07	-.09	.13	.13	-.11	-.21*	-.09	.25*	-.02	—
13 Years Eng	.12	.10	.18	.17	.26**	-.13	.08	-.10	-.09	.18	-.11	-.03

\*  $p < .05$ , two-tailed.

\*\*  $p < .01$ , two-tailed.

Because items from each dimension were continually being contrasted with items from the other dimensions, the relative-scored composite domains tended to be negatively correlated with each other to varying levels of significance, in both the normal response (weakest  $r = .05$ , strongest  $r = -.51$ ) and fake good conditions (weakest  $r = .00$ , strongest  $r = -.35$ ). In contrast, any significant relationships among the standard BFI traits tended to be positive in both the normal (weakest  $r = -.02$ , strongest  $r = .39$ ) and fake response conditions (weakest  $r = .35$ , strongest  $r = .67$ ).

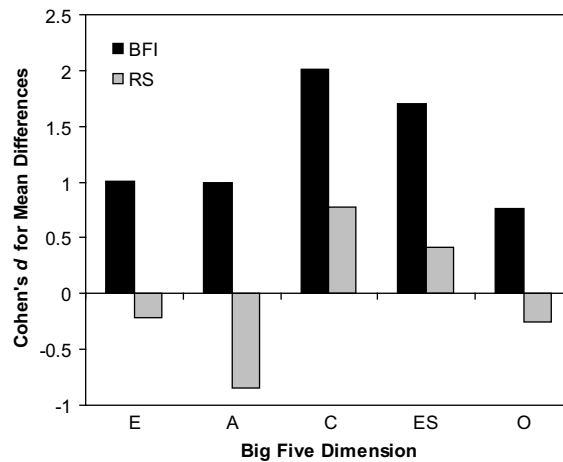
### 3.3. Differences between normal and fake response conditions

Descriptive statistics for both measures and both conditions are presented in Table 3. Mean responses for each of the five dimensions of the BFI were significantly higher in the fake condition, confirming the effectiveness of the fake good manipulation. In accordance with previous research investigating intentional response distortion on the Big Five (Paulhus et al., 1995; Viswesvaran & Ones, 1999), Conscientiousness was the most susceptible to faking ( $d = 2.01$ ), followed by Emotional Stability ( $d = 0.76$ ). While the fake good standard BFI produced significantly higher mean scores on all dimensions, the fake good relative-scored questionnaire had higher means on two dimensions (Conscientiousness and Emotional Stability) and lower means on the others. However, significant differences between conditions were observed only for Agreeableness, Conscientiousness, and Emotional Stability. Fig. 1 displays a graph of the effect size differences across the two response conditions. The relative-scored questionnaire had much smaller differences across conditions (average  $d = -0.03$ ) compared to the standard questionnaire (average  $d = 1.29$ ).

No significant differences were found between conditions for CGPA or CAQ (both  $ps > .05$ ), confirming baseline equivalence of the criterion variables. A small difference was found for Years English, with the fake condition having a

**Table 3**  
Descriptive statistics and  $t$ -tests (two-tailed) for the fake good and normal samples

	Fake Good			Normal			Difference		
	$n$	$M$	$SD$	$n$	$M$	$SD$	$T$	$df$	$p$
<i>Standard BFI</i>									
Extraversion	97	30.87	4.18	99	25.79	5.86	7.0	194	.00
Agreeableness	97	37.98	5.62	99	32.76	4.96	6.9	194	.00
Conscientiousness	97	40.48	4.61	99	29.82	6.00	14.9	194	.00
Emotional Stability	97	32.80	5.59	99	23.16	5.75	11.9	194	.00
Openness	97	40.98	4.94	99	37.13	5.22	5.3	194	.00
<i>Relative-scored</i>									
Extraversion	101	-0.11	0.73	102	0.09	1.11	-1.5	201	.14
Agreeableness	101	-0.37	0.90	102	0.34	0.78	-6.0	201	.00
Conscientiousness	101	0.35	0.70	102	-0.33	1.04	5.5	201	.00
Emotional Stability	101	0.20	0.65	102	-0.16	1.09	2.8	201	.01
Openness	101	-0.12	0.92	102	0.12	0.91	-1.9	201	.06
CGPA	95	2.81	0.63	98	2.85	0.69	-0.4	191	.72
CAQ	101	2.69	0.53	98	2.58	0.49	1.5	197	.13
Years English	103	17.51	5.54	102	15.78	5.93	2.2	203	.03



**Fig. 1.** Effect sizes for the mean differences of each Big Five dimension across normal and fake good conditions (E = Extraversion; A = Agreeableness; C = Conscientiousness; ES = Emotional Stability; O = Openness; BFI = Standard BFI measure; RS = Relative-scored Big Five measure).

slightly higher mean ( $M = 17.51$ ,  $SD = 5.54$ ) than the normal group ( $M = 15.78$ ,  $SD = 5.93$ ),  $t(203) = 2.15$ ,  $p < .05$  (two-tailed),  $d = 0.30$ . However, controlling for Years English did not affect any of the analyses.

### 3.4. Distortion of factor structure under fake responding

Factor analysis with Direct Oblimin rotation ( $\delta = 0$ ) demonstrated that the standard BFI yielded the familiar five factor structure in the normal response condition, with each factor accounting for between 8% and 33% of the total variance. In the fake-good condition, however, much of the standard BFI variance collapsed into a single factor, instead of decomposing into the usual five traits. This single factor accounted for over 60% of the total variance, and can be interpreted as the result of ranking oneself positively on all dimensions. A similar distortion of the Big Five factor structure is sometimes observed in job applicant samples, where faking is more likely (Birkeland et al., 2006; Higgins et al., 2007; Schmit & Ryan, 1993; but see Marshall, De Fruyt, Rolland, & Bagby, 2005). Table 4 presents the correlations of the extracted “positivity” factor with each of the standard and relative-scored Big Five dimensions.

While this single factor was highly correlated with each of the standard-scored dimensions, there were no significant correlations with any of the relative-scored factors. The single factor's correlations with the standard and relative-scored domains were compared using Fisher's  $r$ -to- $z'$  transformation. Significant differences emerged for each of the five factors: Extraversion ( $\Delta r = .66$ ,  $z' = 5.84$ ,  $p < .01$ ), Agreeableness ( $\Delta r = .96$ ,  $z' = 8.45$ ,  $p < .01$ ), Conscientiousness ( $\Delta r = .76$ ,  $z' = 8.31$ ,  $p < .01$ ), Emotional Stability ( $\Delta r = .74$ ,  $z' = 7.97$ ,  $p < .01$ ), Openness ( $\Delta r = .66$ ,  $z' = 5.20$ ,  $p < .01$ ). The lack of relationship of this factor with any of the relative-scored dimensions helps clearly demonstrate the usefulness of the new questionnaire in attenuating biased responding.

**Table 4**  
Correlations between Big Five and single BFI factor in the fake good condition

Variables	Single BFI Factor
BFI Extraversion	.73**
BFI Agreeableness	.79**
BFI Conscientiousness	.87**
BFI Emotional Stability	.86**
BFI Openness	.62**
RS Extraversion	.07
RS Agreeableness	-.17
RS Conscientiousness	.11
RS Emotional Stability	.12
RS Openness	-.04
Years English	.21*

\*  $p < .05$ , two-tailed.

\*\*  $p < .01$ , two-tailed.

### 3.5. Attenuation of relationship between personality responses and Years English

Familiarity with the English language also appears to help individuals present themselves in a more positive light, as may also be seen in Table 4. In both conditions, the number of years speaking English correlated significantly with the participants' summed standard BFI totals across domains (normal  $r = .24$ , fake  $r = .21$ ). Higher scores overall indicate that the respondents were rating themselves more positively on each dimension. In contrast, experience with English had no significant relationship to any of the relative-scored domains, in either response condition (average normal  $r = .00$ , average fake  $r = -.01$ ). This suggests that the relative-scored questionnaire helped to reduce the advantage that native English speakers have over non-native speakers when their personality is being assessed via self-report, and provides another piece of evidence that the relative-scored measures are resistant to biased responding.

### 3.6. Predictive validity across measures and conditions

As Table 1 reveals, the standard BFI was able to significantly predict both CGPA and CAQ scores in the normal condition (Conscientiousness and CGPA ( $r = .29$ ); Openness and CAQ scores ( $r = .35$ ), as hypothesized). However, the standard BFI produced no significant predictors in the fake good condition (Conscientiousness and CGPA ( $r = .09$ ); Openness and CAQ ( $r = .13$ ). To test whether the overall predictive validity of the BFI differed significantly between response conditions, we averaged the correlations of Openness with CAQ and Conscientiousness with CGPA and compared this value across response conditions as a measure of cross-domain predictive ability. As expected, the average correlation in the normal condition ( $r = .32$ ,  $n = 189$ ) was significantly greater than the average correlation in the fake condition ( $r = .11$ ,  $n = 183$ ),  $z' = -2.16$ ,  $p < .05$ . Our sample thus confirmed that the standard BFI loses predictive validity under at least some conditions promoting biased responding.

In contrast to the standard measure, the relative-scored Big Five questionnaire was robust against attempts to fake good. The relative-scored measure of Conscientiousness significantly predicted CGPA in both the normal ( $r = .32$ ) and fake ( $r = .24$ ) response conditions. Similarly, Openness was also able to maintain its predictive ability across conditions (normal  $r = .29$ , fake  $r = .25$ ). A comparison of the weighted average correlation (normal  $r = .31$ ,  $n = 194$ , fake  $r = .24$ ,  $n = 193$ ) showed that there was no significant difference overall in the predictive ability of the relative-scored questionnaire across conditions ( $z' = -0.64$ ,  $p = .26$ ). These are the final two—and the most powerful—of four pieces of evidence demonstrating the robustness of the relative-scored measures to distortion by response bias.

## 4. Discussion

Students exposed to relatively simple instructions to fake good, as if simulating a job assessment situation, appeared able (1) to distort the factor structure of a standard Big Five personality measure; (2) to successfully present themselves in an enhanced manner, particularly exaggerating Conscientiousness and Emotional Stability, the two best personality predictors of job performance; and (3) to reduce the relationship between standard Big Five measures and two measures of performance (CGPA and CAQ) to insignificance. However, students completing the novel relative-scored Big Five questionnaires were much less able to produce such positive distortion. The standard BFI lost its predictive validity in the fake response condition, but the relative-scored questionnaires were able to predict creative achievement and academic success in both conditions. This pattern of results mirrors previous studies in which relative-scored questionnaires were able to predict workplace delinquency (Jackson et al., 2000) and supervisor performance ratings (Christiansen et al., 2005) under instructions to fake good. Making repeated choices between equally socially desirable personality descriptors thus appears to be a process less sensitive to biased responding than rating individual items on a Likert scale. Even when trying to present themselves favourably, the choices that participants made revealed a great deal about their personality and could be used to predict behavioural outcomes. The fact that this assessment technique was resilient against explicit instructions to fake good demonstrates that it retains its predictive validity under laboratory conditions that significantly increase self-report bias. Given that real-world faking tends to be less pronounced than laboratory faking studies (Birkeland et al., 2006; Viswesvaran & Ones, 1999), the current study provides a strong test of the relative-scored format's resistance to biased responding.

Four main findings provide strong support for the robustness of the relative-scored personality questionnaires against faking. First, the relative-scored Big Five dimensions were not correlated significantly with the single factor extracted from the fake-good BFI and hypothetically indexing positive self-presentation. Second, the relative-scored questionnaire was not susceptible to the potential for producing positive bias that was apparently characteristic of speakers with increased English fluency. Third, the relative-scored questionnaire was able to successfully predict students' GPA under explicit instructions to fake good. Fourth, and finally, the new measure was also a valid predictor of creative achievement, a measure which bore no significant relationship to academic success. The sustained predictive validity of the relative-scored questionnaire across these two independent performance domains, both relying on separate personality traits, emphasizes the resilience and utility of the new measure. This suggests that much of the variance associated with self-enhancement on the Likert-style questionnaire has been eliminated through use of the new measure.

Support for the validity of the present study was provided by the fact that as in previous research, faking was most pronounced in the traits of Conscientiousness ( $d = 2.01$ ), and Emotional Stability ( $d = 0.76$ ) (Paulhus et al., 1995; Viswesvaran &



Ones, 1999). It should be noted that these factors are the two best personality predictors of workplace performance (Barrick & Mount, 1991), highlighting the potentially detrimental impact of biased responding on selection procedures. Self-reported personality among job applicants also tends to be inflated on these dimensions compared to other populations, further suggesting that biased responding is a significant issue to contend with (Birkeland et al., 2006; Stark, Chernyshenko, Chan, Lee, & Drasgow, 2001). Considering that the relative-scored questionnaire was able to attenuate the effects of biased responding in both of these domains, it may prove particularly useful in the prediction of workplace performance. The massive variability in productivity typically obtaining between individuals means that even the moderate improvements in predictive validity potentially gained from the new questionnaire could have large economic benefits when used in real world selection procedures (e.g., Hunter, Schmidt, & Judiesch, 1990).

It is worth noting that relative-scored, or ipsative, techniques have been severely criticized for some of their mathematical shortcomings, such as range restriction and reduced variance (Bartram, 1996; Hicks, 1970; Johnson, Wood, & Blinkhorn, 1988). There is thus some real cost to be paid for accruing the benefits of potentially increased validity. However, this manifests itself primarily in the effects of relative scoring on the independence of the Big Five traits and the consequences for certain statistical procedures. Whenever an increased score is observed in one dimension, a lower score is necessarily observed in another dimension. The resultant collinearity between domains makes the relative-scored survey format problematic for multiple regression and factor analyses (Cornwell & Dunlap, 1994). Consequently, such scales are most useful when a single domain can be used for predictive purposes, without attempting to combine it with any of the other domains. Such scales should also not be relied upon to assess the relationships between traits, because these are necessarily forced to be more negatively correlated with each other than would be the case for a non-ipsative measure.

However, such criticisms (1) are most pertinent when applied to fully ipsatized measures and (2) do not necessarily mean that ipsatized scores are by necessity invalid, or even less valid, under all conditions. The scoring procedure utilized in the present study reduces the problems of ipsatization appreciably by standardizing scores on the individual scales, derived using different methods, and then averaging across the standardized values to extract composite trait scores. This means that our scale is “normative ipsatized”, and is thus less affected by the mathematical problems of fully ipsatized measures (Hicks, 1970). In a fully ipsatized measure, for example, the sum of the rows and columns in the intercorrelation matrix should both equal zero. As can be seen from Table 1 and Table 2, however, this is not the case for the relative-scored dimensions used in this study.

On a theoretical note, each of the composite domain scores derived from the new measure represents the relative strength of a given trait, as compared to the relative strength of that trait in others. In other words, a high Conscientiousness score on this questionnaire indicates that such an individual places a greater within-person emphasis on Conscientiousness, compared to other individuals. The current study thus suggests that the relative strength of different personality traits within an individual can still be an effective performance predictor. The fact that such within-person ranking of personality traits is an effective predictor of performance deserves attention in future research. It is worth noting, however, that the high correlations between Likert and relative-scored Big Five dimensions also suggests that within-individual trait rankings converge considerably with absolute trait scores rated across individuals (cf. Jackson, Neill, & Bevan, 1973).

Use of the new scale is finally justified by our study's fulfillment of Hicks' (1970) requirements for properly employing ipsative measures. In his critique of ipsative measurement, Hicks concluded that such techniques should only be used when '(a) significant response bias exists; (b) this bias reduces validity and (c) an ipsative format successfully diminishes bias and increases validity to a greater extent than do non-ipsative controls for bias' (Hicks, 1970, p. 181). Each one of these criteria was met in the current study. Thus, there is good reason to assume that the relative-scored questionnaire described herein might be a useful instrument for enhancing the predictive validity of personality questionnaires under conditions of biased responding.

Under the specific experimental conditions detailed in this paper, the collinearity between traits characteristic of partially ipsatized scales appears to have had the strongest influence on Agreeableness, as scores on this dimension were significantly lowered in the fake good condition ( $d = -0.85$ ). This suggests that Agreeableness was “sacrificed” in order to raise scores on Conscientiousness and Emotional Stability, such that respondents were more likely to choose items from the latter two domains. One interpretation of this result is that the participants in our study considered high levels of Agreeableness to be less important in the eyes of potential employers. This supposition may in fact be justified, practically, given that relatively lower levels of Agreeableness may be predictive of enhanced workplace performance in high-autonomy jobs (Barrick & Mount, 1993). Support for such interpretation can be derived from research demonstrating that participants are aware of the most desirable traits for a given assessment purpose, implicitly or explicitly, and can modify their responses appropriately (Furnham, 1990; Martin, Bowen, & Hunt, 2002). However, even if such strategic response manipulation manifested itself to some degree in our study, the relative-scored questionnaire still maintained its predictive validity very well, in contrast to the Likert scored counterpart.

Overall, then, the present study provides evidence that the relative-scored measure of the Big Five can help to limit the effects of biased responding. Perhaps individuals motivated to employ Big Five trait questionnaires might choose between the Likert and relative-scored measures, according to their explicit purposes. The former may well prove more effective under two conditions: first, when the goal is to assess the statistical nature of the relationship between different traits, as the correlation between those traits is not exaggerated by the administration methodology and second, when the relationship between a criterion external to the test is to be measured under conditions when the test-takers are not motivated to look good. The relative-scored measures, by contrast, may be particularly useful when prediction under motivated conditions is

the aim. Such questionnaires are likely to be useful, for example, under competitive, zero-sum conditions where respondents will be motivated towards favourable impression management. It should also be noted that although some research suggests that relative-scored techniques may not provide improved assessment at the individual level (Heggestad, Morrison, Reeve, & McCloy, 2006), their ability to predict performance criteria under faking conditions makes them a potentially valuable tool for selection purposes. Because an individual score on any personality scale is a function of the true score plus measurement error or response bias, there is never any guarantee that any particular individual will be accurately assessed, even when using normative Likert questionnaires. The utility of personality questionnaires for selection purposes operates at the group level, such that repeated use of such measures will on average lead to benefits in line with the scale's predictive validity.

Finally, the study also suggests something somewhat unexpected and potentially interesting. Subjects in the fake good condition appeared willing to sacrifice their appearance on Agreeableness in order to enhance their scores for Conscientiousness and Emotional Stability, the two best personality predictors of job success. This suggests that the relative measures might be used to investigate the structure of implicit or explicit models of expected ideal behaviour, in different situations of administrator or target demand. In consequence, we are currently investigating self-presentation using relative measures in other stressful and competitive situations that are not specifically job performance related, hoping that we can derive some insight into what individuals consider specifically worth highlighting and denigrating about their personality, in relationship to their particular goals.

## References

- Allaman, J. D., Joyce, C. S., & Crandall, V. C. (1972). The antecedents of social desirability response tendencies of children and young adults. *Child Development, 43*, 1135–1160.
- Bagby, R. M., & Marshall, M. B. (2003). Positive impression management and its influence on the revised NEO personality inventory: A comparison of analog and differential prevalence group designs. *Psychological Assessment, 15*, 333–339.
- Baron, H. (1996). Strengths and limitations of ipsative measurement. *Journal of Occupational and Organizational Psychology, 69*, 49–56.
- Barrick, M. R., & Mount, M. K. (1991). The Big Five personality dimensions and job performance: A meta-analysis. *Personnel Psychology, 44*, 1–26.
- Barrick, M. R., & Mount, M. K. (1993). Autonomy as a moderator of the relationships between the Big Five personality dimensions and job performance. *Journal of Applied Psychology, 78*, 111–118.
- Barrick, M. R., & Mount, M. K. (1996). Effects of impression management and self-deception on the predictive validity of personality constructs. *Journal of Applied Psychology, 81*, 261–272.
- Barrick, M. R., Mount, M. K., & Judge, T. A. (2001). Personality and performance at the beginning of the new millennium: What do we know and where do we go next? *International Journal of Selection and Assessment, 9*, 9–30.
- Bartram, D. (1996). The relationship between ipsatized and normative measures of personality. *Journal of Occupational and Organizational Psychology, 69*, 25–39.
- 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*, 317–335.
- Block, J. (1965). *The challenge of response sets: Unconfounding meaning, acquiescence, and social desirability in the MMPI*. New York: Appleton-Century-Crofts.
- Borkenau, P., & Amelang, M. (1985). The control of social desirability in personality inventories: A study using the principal-factor deletion technique. *Journal of Research in Personality, 19*, 44–53.
- Byrne, D., & Bounds, C. (1964). The reversal of F Scale items. *Psychological Reports, 14*, 216.
- Carson, S., Peterson, J. B., & Higgins, D. (2005). Reliability, validity, and factor structure of the Creative Achievement Questionnaire. *Creativity Research Journal, 17*, 37–50.
- Christiansen, N. D., Burns, G. N., & Montgomery, G. E. (2005). Reconsidering forced-choice item formats for applicant personality assessment. *Human Performance, 18*, 267–307.
- Chuah, S. C., Drasgow, F., & Roberts, B. W. (2006). Personality assessment: Does the medium matter? No. *Journal of Research in Personality, 40*, 359–376.
- Cornwell, J. M., & Dunlap, W. P. (1994). On the questionable soundness of factoring ipsative data: A response to Saville and Willson (1992). *Journal of Occupational and Organizational Psychology, 67*, 89–100.
- Costa, P. T., Jr., & McCrae, R. R. (1992). *Revised NEO Personality Inventory and NEO Five-Factor Inventory professional manual*. Odessa, FL: Psychological Assessment Resources.
- Costa, P. T., Jr., & McCrae, R. R. (1997). Longitudinal stability of adult personality. In R. Hogan, J. A. Johnson, & S. R. Briggs (Eds.), *Handbook of personality psychology* (pp. 269–290). San Diego: Academic Press.
- Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology, 24*, 349–354.
- DeYoung, C. G., Peterson, J. B., & Higgins, D. (2002). Higher-order factors of the Big Five predict conformity: Are there neuroses of health? *Personality and Individual Differences, 33*, 533–553.
- DeYoung, C. G., Peterson, J. B., & Higgins, D. (2005). Sources of Openness/Intellect: Cognitive and neuropsychological correlates of the fifth factor of personality. *Journal of Personality, 73*, 825–858.
- DeYoung, C. G., Quilty, L. C., & Peterson, J. B. (2007). Between facets and domains: Ten aspects of the Big Five. *Journal of Personality and Social Psychology, 93*, 880–896.
- Edwards, A. L. (1953). The relationship between the judged desirability of a trait and the probability that the trait will be endorsed. *Journal of Applied Psychology, 37*, 90–99.
- Edwards, A. L. (1957). *The social desirability variable in personality assessment and research*. New York: Dryden.
- Ellingson, J. E., Sackett, P. R., & Hough, L. M. (1999). Social desirability corrections in personality measurement: Issues of applicant comparison and construct validity. *Journal of Applied Psychology, 84*, 155–166.
- Eysenck, H. J. (1994). Neuroticism and the illusion of mental health. *American Psychologist, 49*, 971–972.
- Eysenck, H. J. (1995). *Genius: The natural history of creativity*. New York: Cambridge University Press.
- Eysenck, S. B., Eysenck, H. J., & Barrett, P. (1985). A revised version of the Psychoticism Scale. *Personality and Individual Differences, 6*, 121–129.
- Furnham, A. (1986). Response bias, social desirability and dissimulation. *Personality & Individual Differences, 7*, 385–400.
- Furnham, A. (1990). Faking personality questionnaires: Fabricating different profiles for different purposes. *Current Psychology: Research & Reviews, 9*, 46–55.
- Furnham, A. (1997). Knowing and faking one's five-factor personality scale. *Journal of Personality Assessment, 69*, 229–243.
- Goff, M., & Ackerman, P. L. (1992). Personality-intelligence relations: Assessment of typical intellectual engagement. *Journal of Educational Psychology, 84*, 537–552.
- Goldberg, L. R. (1992). The development of markers for the Big-Five factor structure. *Psychological Assessment, 4*, 26–42.
- Goldberg, L. R. (1999). A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models. In I. Mervielde, I. Deary, F. De Fruyt, & F. Ostendorf (Eds.), *Personality psychology in Europe* (Vol. 7, pp. 7–28). Tilburg, The Netherlands: Tilburg University Press.
- Gray, E. K., & Watson, D. (2002). General and specific traits of personality and their relation to sleep and academic performance. *Journal of Personality, 70*, 177–206.

- Heggstad, E. D., Morrison, M., Reeve, C. L., & McCloy, R. A. (2006). Forced-choice assessments of personality for selection: Evaluating issues of normative assessment and faking resistance. *Journal of Applied Psychology*, 91, 9–24.
- Hicks, L. E. (1970). Some properties of ipsative, normative, and force-choice normative measures. *Psychological Bulletin*, 74, 167–184.
- Higgins, D. M., Peterson, J. B., Pihl, R. O., & Lee, A. G. M. (2007). Prefrontal cognitive ability, intelligence, Big Five personality, and the prediction of advanced academic and workplace performance. *Journal of Personality and Social Psychology*, 93, 298–319.
- Hough, L. M., Eaton, N. K., Dunnette, M. D., & Kamp, J. D. (1990). Criterion-related validities of personality constructs and the effect of response distortion on those validities. *Journal of Applied Psychology*, 75, 581–595.
- Hunter, J. E., Schmidt, F. L., & Judiesch, M. K. (1990). Individual differences in output variability as a function of job complexity. *Journal of Applied Psychology*, 75, 28–42.
- Hurtz, G. M., & Donovan, J. J. (2000). Personality and job performance: The Big Five revisited. *Journal of Applied Psychology*, 85, 869–879.
- International Personality Item Pool (2005). A scientific collaboratory for the development of advanced measures of personality traits and other individual differences. Available from <http://ipip.ori.org/>. Retrieved September, 2005.
- Jackson, D. N., Neill, J. A., & Bevan, A. R. (1973). An evaluation of forced-choice and true-false item formats in personality assessment. *Journal of Research in Personality*, 7, 21–30.
- Jackson, D. N., Wroblewski, V. R., & Ashton, M. C. (2000). The impact of faking on employment tests: Does forced choice offer a solution? *Human Performance*, 13, 371–388.
- John, O. P., Donahue, E. M., & Kentle, R. L. (1991). *The "Big Five" Inventory—Versions 4a and 54*. Berkeley: University of California, Berkeley, Institute of Personality and Social Research.
- Johnson, C. E., Wood, R., & Blinkhorn, S. F. (1988). Spuriouser and spuriouser: The use of ipsative personality tests. *Journal of Occupational Psychology*, 61, 153–162.
- Judge, T. A., & Bono, J. E. (2000). Five-factor model of personality and transformational leadership. *Journal of Applied Psychology*, 85, 751–765.
- Judge, T. A., Bono, J. E., Ilies, R., & Gerhardt, M. W. (2002). Personality and leadership: A qualitative and quantitative review. *Journal of Applied Psychology*, 87, 765–780.
- Li, A., & Bagger, J. (2006). Using the BIDR to distinguish the effects of impression management and self-deception on the criterion validity of personality measures: A meta-analysis. *International Journal of Selection and Assessment*, 14, 131–141.
- 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.
- Martin, B. A., Bowen, C. C., & Hunt, S. T. (2002). How effective are people at faking on personality questionnaires? *Personality and Individual Differences*, 32, 247–256.
- McCrae, R. R. (1987). Creativity, divergent thinking, and openness to experience. *Journal of Personality and Social Psychology*, 52, 1258–1265.
- McCrae, R. R., & Costa, P. T. Jr., (1983). Social desirability scales: More substance than style. *Journal of Consulting and Clinical Psychology*, 51, 882–888.
- McCrae, R. R., & Costa, P. T. Jr., (1997). Personality trait structure as a human universal. *American Psychologist*, 52, 509–516.
- McCrae, R. R., & John, O. P. (1992). An introduction to the five-factor model and its applications. *Journal of Personality*, 60, 175–215.
- Mersman, J. L., & Shultz, K. S. (1998). Individual differences in the ability to fake on personality measures. *Personality and Individual Differences*, 24, 217–227.
- Mueller-Hanson, R., Heggstad, E. D., & Thornton, G. C. (2003). Faking and selection: Considering the use of personality from select-in and select-out perspectives. *Journal of Applied Psychology*, 88, 348–355.
- 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.
- Ozer, D. J., & Benet-Martinez, V. (2006). Personality and the prediction of consequential outcomes. *Annual Review of Psychology*, 57, 401–421.
- Paulhus, D. L. (1991). Measurement and control of response bias. In J. P. Robinson & P. R. Shaver, et al. (Eds.), *Measures of personality and social psychological attitudes* (pp. 17–59). San Diego, CA: Academic Press, Inc.
- Paulhus, D. L. (2002). Socially desirable responding: The evolution of a construct. In H. I. Braun & D. N. Jackson (Eds.), *The role of constructs in psychological and educational measurement* (pp. 37–48). Mahwah, NJ: Lawrence Erlbaum.
- 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.
- Paulhus, D. L., & John, O. P. (1998). Egoistic and moralistic biases in self-perception: The interplay of self-deceptive styles with basic traits and motives. *Journal of Personality*, 66, 1025–1060.
- 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.
- 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.
- Reynolds, W. M. (1982). Development of reliable and valid short forms of the Marlow-Crowne Social Desirability Scale. *Journal of Clinical Psychology*, 38, 119–125.
- Rosse, J. G., Stecher, M. D., Miller, J. L., & Levin, R. A. (1998). The impact of response distortion on preemployment personality testing and hiring decisions. *Journal of Applied Psychology*, 83, 634–644.
- Sackeim, H. A., & Gur, R. C. (1978). Self-deception, self-confrontation, and consciousness. In G. E. Schwartz & D. Shapiro (Eds.), *Consciousness and self-regulation, advances in research and theory* (Vol. 2, pp. 139–197). New York: Plenum Press.
- Salgado, J. F. (1997). The five-factor model of personality and job performance in the European community. *Journal of Applied Psychology*, 82, 30–43.
- 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.
- 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.
- Select Survey ASP Advanced (Version 8.1.5) [Computer Software]. Clifton, NJ: ClassApps.
- Stark, S., Chernyshenko, O. S., Chan, K., Lee, W. C., & Drasgow, F. (2001). Effects of the testing situation on item responding: Cause for concern. *Journal of Applied Psychology*, 86, 943–953.
- Thoresen, C. J., Kaplan, S. A., Barsky, A. P., Warren, C. R., & de Chermont, K. (2003). The affective underpinnings of job perceptions and attitudes: A meta-analytic review and integration. *Psychological Bulletin*, 129, 914–945.
- Thurstone, L. L. (1927). The method of paired comparisons for social values. *Journal of Abnormal and Social Psychology*, 21, 384–400.
- Viswesvaran, C., & Ones, D. S. (1999). Meta-analyses of fakability estimates: Implications for personality measurement. *Educational and Psychological Measurement*, 59, 197–210.
- Viswesvaran, C., & Ones, D. S. (2000). Measurement error in "Big Five factors" personality assessment: Reliability generalization across studies and measures. *Educational and Psychological Measurement*, 60, 224–235.