Psychology Science Quarterly, Volume 50, 2008 (1), p. 44-63

The susceptibility of a mixed model measure of emotional intelligence to faking: A Solomon four-group design

DANIEL S. WHITMAN[1], DAVID L. VAN ROOY[2], CHOCKALINGAM VISWESVARAN[3] & ALEXANDER ALONSO[4]

Abstract

This study used a Solomon four-group design to investigate the fakability of a widely used measure of emotional intelligence (EI). Administration instructions (faking/honest-response) and testing effects (pre-test/no pre-test) were the two conditions (i.e., 2 x 2) examined; two different Solomon four-group designs – one for fake-good instructional treatments and one for fake-bad instructional treatments – were assessed. Participants (n = 300) were randomly assigned to one of the six conditions and results indicate a significant pre-testing effect for fake-bad but not fake-good instructions. However, the interaction of testing and treatment was only significant for fake-good but not fake-bad. As expected, within-subjects designs resulted in higher distortion than between-subjects designs for both fake-good (d-value of 1.08 compared to 0.10) and fake-bad instructions (4.07 vs. 3.56, respectively). Participants were able to fake-bad more than fake-good, irrespective of the design used and scaling effects. Implications for EI assessments are discussed.

Key words: emotional intelligence; test faking; pre-testing effects; assessment; selection

[1] Daniel S. Whitman, DM-266, Department of Psychology, University Park Campus, Florida International University, Miami, FL, 33199, USA; Phone: 404-213-1066, email: dwhitman@yahoo.com

[2] Marriott International

[3] Florida International University

[4] American Institutes for Research

The popular literature on emotional intelligence (EI) has at times seemed to outpace empirical research on the topic. This is partly due to journalists and public commentators making claims about the importance of EI that have not been based on rigorous research findings (Ashkanasy & Daus, 2005; Brody, 2004; Matthews, Zeidner, & Roberts, 2002; Murphy, 2006). In the past few years, scientists have begun to rectify this and report more empirical evidence (Murphy, 2006). For instance, EI has recently been shown to be a valid predictor of important outcomes such as leadership, teamwork, and job performance (Cote & Miners, 2006; Van Rooy & Viswesvaran, 2004; Wong & Law, 2002). However, as EI is increasingly being used in organizational settings (Daus & Ashkanasy, 2005), a number of practical concerns – in addition to criterion-related validity – still need to be addressed about this construct (Conte, 2005).

One such area that requires additional research is on the susceptibility of a measure of EI to overt faking (e.g., Barchard, 2003; Charbonneau & Nicol, 2001; Sjoberg, 2001). In the present study, we build on previous research and examine response distortion to EI by adopting a more rigorous experimental technique known as the Solomon four-group design. Our major goal in employing this sophisticated experimental design is to disentangle faking effects from instructional and pre-testing effects as they relate to scores on an EI test (e.g., Alliger & Dwight, 2000). We also examine how individual differences, namely general mental ability (GMA) and personality, influence the capability to fake, as prior research on response distortion has seldom investigated these important variables.

Models of emotional intelligence

Many tests of EI (e.g., Bar-On EQ-i, 1997; MSCEIT V2.0; Mayer, Salovey, Caruso & Sitarenios, 2003) have been marketed and are currently being used in personnel selection contexts (Murphy, 2006). It is therefore important to understand how responses on EI tests can be distorted in high-stakes testing (Conte, 2005). However, there is a controversy within the field of EI over proper conceptualizations of the construct. One conceptualization, known as the *ability model* (Salovey & Mayer, 1990), considers EI as a type of intelligence and, as such, it should be moderately related to GMA. This has been consistently demonstrated (e.g., Brackett & Mayer, 2003; Schulte, Ree & Caretta, 2004), and Van Rooy, Viswesvaran, and Pluta (2005) found a corrected correlation of .34 between the two constructs in a meta-analytic review. These findings suggest that ability-based EI shares the same positive manifold as other cognitive abilities do with GMA (Brody, 2004). Because of this overlap, ability-based measures of EI should not be overly susceptible to faking attempts (e.g., Jensen, 1998).

A competing conceptualization of EI has been commonly referred to as a *mixed model*. In addition to incorporating components of the ability model, the mixed model of EI includes other non-cognitive aspects such as personality, motivation and empathy. Van Rooy and colleagues (2005) reported a much smaller corrected correlation of .13 between mixed model EI and GMA, suggesting greater construct divergence than the ability model. Given that they are relatively orthogonal constructs, mixed model EI may have an appeal in selection contexts (Van Rooy, Dilchert, Viswesvaran, & Ones, 2006) as a predictor that could potentially supplement GMA as an assessment tool (Schmidt & Hunter, 1998).

Faking cognitive and non-cognitive measures

The two models highlight important issues to consider when determining if measures of EI can be faked. On the one hand, faking is extremely difficult for measures of GMA (e.g., Jensen, 1998). On the other hand, researchers have often found that non-cognitive measures can be faked to varying extents (e.g., Deller, Ones, Viswesvaran, & Dilchert, 2006; Frei, Griffith, McDaniel, Snell & Douglas, 1997). This suggests that the extent to which mixed model measures of EI (which do not claim to be a type of intelligence) can be faked is a concern that needs to be empirically reviewed. Prior research (Viswesvaran & Ones, 1999) has found that non-cognitive predictors like personality and biodata can be faked. In a similar vein, research on the fakability of EI (especially self-report measures) is needed to determine whether response patterns for these measures are similar to other non-cognitive constructs. This examination will serve as indirect evidence of the construct validity of EI. That is, if the fakability of a mixed model measure of EI is similar to the fakability found for Big Five factors of personality, such evidence will serve as another empirical node in the nomological net.

In the past decade there have been a number of examinations regarding the relative fakability of the Big Five personality dimensions. For example, Viswesvaran and Ones (2000) concluded that each of the dimensions is equally fakable in a recent meta-analysis. Another comparison of the fakability of non-cognitive measures (personality test, biodata inventory, and an integrity test) found that biodata and integrity tests were both more fakable than any of the Big Five (McFarland & Ryan, 2000). Given that different non-cognitive measures appear to vary in the degree that they can be faked, it is important to gauge the fakability of mixed model EI, which has been shown to overlap with aspects of the Big Five (e.g., Brackett & Mayer, 2003; Saklofske, Austin & Minski, 2003; Schutte et al., 1998). Thus, in the present study, we adopt a measure of EI based on the mixed model (or trait-based) conceptualization.

Previous research on faking EI

Although some researchers have probed response distortion in relation to emotional intelligence (e.g., Barchard, 2003; Charbonneau & Nicol, 2002; Janovics & Christiansen, 2004; Sjoberg, 2001), there are a number of limitations with the extant research that suggest more rigorous examinations are needed. For instance, Charbonneau and Nicol (2002) assessed response distortion in a sample of minors. While there is value in assessing EI for clinical purposes, the results from such studies are less relevant for organizational selection purposes where almost all applicants are over the age of 18. Although some measures of EI (e.g., EQ-i, Bar-On, 1997) have impression management scales built in, others have noted that the use of an impression management scale provides only *general* information about an individuals test taking propensities (McDaniel & Snell, 1998; McFarland & Ryan, 2000). Lie scales and impression management scales can assess typical faking in an assessment, but to assess the maximal level of faking possible in an instrument, experimental manipulations using faking instructions are necessary (Paulhus, 1984). Sjoberg (2001), for instance, examined fakability in a host of measures (e.g., EI, personality, locus of control, Machiavellianism), but the study was not able to isolate EI-specific faking. Indeed, Sjoberg acknowledges that a section involving experimentally manipulated faking instructions would be a valuable addition to the test battery employed in the study. To extend these past examinations of response sets and EI, we use an experimental study to measure active (i.e., conscious) attempts to fake responses.

The importance of addressing the fakability of a self-report (i.e., mixed model) measure of EI is underscored even more when one considers the proliferation of self-report EI measures that have been postulated and marketed as valuable predictors of performance (e.g., EIS; Schutte et al., 1998). Accordingly, Salovey, Mayer, and Caruso (2002) stated that simply asking someone if they were smart would not be a sufficient indicator of intelligence and that the same standard should be held for EI. Past studies using self-report measures of intelligence have typically demonstrated only moderate correlations with IQ tests (e.g., Paulhus, Lysy & Yik, 1998). However, this assumption will be tempered if a self-report measure of EI is shown to be difficult to fake. It is prudent to explore the fakability of a self-report, mixed model measure of EI now.

Mixed model EI and faking

If a mixed measure is shown to be resistant to faking, there will be evidence that EI is a unique construct. That is, such a finding would suggest that mixed model EI may be more cognitively loaded than previously thought. However, if it is fakable, as hypothesized, our main concern is the extent to which the measure can be faked. In this vein, we compare the degree of fakability of a mixed measure of EI with that of the fakability found for personality factors (Viswesvaran & Ones, 1999). Such a comparison is warranted given the rapid implementation of EI testing in high-stakes assessment such as personnel selection (e.g., Ashkanasy & Daus, 2005; Murphy, 2006). Furthermore, a rigorous experimental design is needed to determine the generalizability of the findings from a comparison of Big Five and EI assessments.

Influence of individual differences

Researchers have only recently recognized that individual difference variables may affect the degree to which an individual is able to alter their score on a selection test (McFarland & Ryan, 2000). Thus, we still know very little about how important constructs such as GMA and personality influence the ability to fake non-cognitive measures. Given that mixed EI is non-cognitive in nature, the present study assesses if test takers with higher levels of intelligence are better able to modify their EI scores in the direction instructed by test administrators. If so, such findings would undermine the construct validity of many non-cognitive measures, which claim to be distinct from GMA. Because general mental ability (GMA) is defined as “the ability to reason, plan, solve problems and learn from experience” (Gottfredson, 1997, p. 13), it would stand to reason that test takers with higher levels of intelligence would be better able to determine appropriate responses (i.e., reason and planning) and distort their responses accordingly (i.e., solving the “problem” or “puzzle” of faking EI in an appropriate manner). Thus, we expect that individuals with higher GMA will be able to change their scores more when instructed to fake good or bad.

Similarly, we examine the impact of Big Five personality variables on the ability to fake EI. While we believe that some of the Big Five factors (e.g., Extraversion, Emotional Stability, Openness) will not have an impact on the ability to fake EI, there is theoretical rationale for other factors. For instance, those who score high on the Conscientiousness factor of the Big Five are often described as dependable, cautious, and responsible (Goldberg, 1990). Thus, Conscientious subjects may see "faking" as frivolous or even irresponsible behavior, and may be less likely to fake even when instructed to do so. Agreeableness, which represents the tendency to be compliant, trusting, and helpful, may also affect a person's ability to fake EI. However, we believe that those who are high on Agreeableness will be more likely to fake when instructed to do so, considering that they are more likely to trust test administrators and comply with directions.

Faking good and faking bad

The current study examines the effects of both faking good and bad. Most of the past personnel selection research on faking has focused on faking good since that is the most likely scenario in selection contexts. Even though the likelihood of negative distortion in a selection setting is probably somewhat remote (Hough, Eaton, Dunnette, Kamp, & McCloy, 1990), there is still utility in assessing the influence of faking bad on a measure. Moreover, this appears to be especially critical during the early stages of construct development (Mook, 1983). Although ability to fake bad is a curious and unusual concept in personnel selection, it is useful to assess fake-good and fake-bad profiles in order to estimate cutoff points for individuals with unusually high or low scores on a test (Paulhus, Bruce & Trapnell, 1995). Becker and Martin (1995) suggest that some employees may intentionally attempt to look bad to avoid undesirable outcomes such as a promotion with greater responsibility or an unwanted transfer. Furthermore, the popularity of EI is extending from personnel selection to clinical, judicial and counseling settings where faking-bad is a clear issue (e.g., a cry for help).

Previous studies (Viswesvaran & Ones, 1999) have found that personality tests are more susceptible to fake-bad than fake-good instructions. Individuals may be able to better understand what "bad" is supposed to look like and may also be less certain about what would constitute the most desirable response in a given situation. It is also possible that a ceiling effect occurs in scoring non-cognitive assessments. If this is the case, on average, scores will tend to be closer to the maximum value of the scale than the minimum value. Conversely, the maximum value that a test-taker could fake-bad (i.e., lower their score to the scale minimum) would be greater than the maximum value a test-taker could fake-good (i.e., raise their score to the scale maximum). In either case, it will be of interest in the current experiment to determine if this finding holds true for a measure of mixed model EI, or if faking bad has the same magnitude of effect, albeit in the opposite direction, as faking good.

Experimental designs used in faking studies

The influence of faking has been evaluated for a number of non-cognitive measures including personality, biodata, and integrity tests (e.g., Ellingson, Smith & Sackett, 2001;

McFarland & Ryan, 2000; Ones & Viswesvaran, 1998; Ones, Viswesvaran, & Schmidt, 1993), and it is a widely held belief that faking can and does occur on non-cognitive tests and that the propensity for faking can depend on individual difference variables. However, reviews of the research on the fakability of non-cognitive measures suggest that studies have typically employed *either* a between-subjects design or a within-subjects design (cf. Viswesvaran & Ones, 1999).

In a between-subjects experimental design, two groups are formed and participants are randomly assigned to each group. One group takes the test under *faking* instructions whereas the other takes it under *respond honest* instructions. In a within-subjects design, the same set of respondents take the test twice, once with faking instructions and once with respond honest instructions. When a within-subjects design is employed, there is a potential for "testing" effects. That is, taking the test once could influence responses the second time. Hausknecht, Trevor and Farr (2002) report a significant pre-testing effect for ability tests, but such pre-testing effects are likely to be more pronounced for non-cognitive assessments such as EI. The testing effect can be assessed using a control group that responds to a measure of EI twice under respond honest instructions.

Furthermore, while both of these comparisons have relative merits, previous studies have found substantial differences in effect sizes between within-subjects and between-subjects comparisons (Viswesvaran & Ones, 1999). These differential findings may be attributable to the increased statistical power gained by using a within-subjects as compared to a between-subjects design (Hunter & Schmidt, 1990). Viswesvaran and Ones (1999), for instance, found that effect sizes of within-subjects estimates are larger than that of between-subjects design estimates and concluded that fakability may be an individual difference variable. Thus, another important research question concerns whether similar findings will hold for other non-cognitive predictors like mixed model EI.

Solomon four group design

Because faking is a function of both the ability and motivation to fake (McFarland & Ryan, 2000), a subject’s prior exposure to the test has the potential to interact with faking instructions. However, use of either a between-subjects design or a within-subjects design makes it impossible to estimate interactions between test instructions and previous test exposure. To address this concern, Solomon (1949) proposed a four-group research design in which two additional groups (one control, one experimental) were added to the typical two-group within-subjects or between-subjects designs. The four-group design is shown in Figure 1.

First, a group is designed where participants respond only once under faking instructions (i.e., they take the test only once). The second additional group is a control group that received no pre-test (and no treatment) and takes only the post-test (where they respond honestly). In addition to controlling for threats to internal validity, the design also allows tests to determine if there is a pre-test sensitization effect (Braver & Braver, 1988) and a pre-test-treatment interaction (Huck & Sandler, 1973). Still, even with the power of the design, it is often underused in research where it is appropriate because of (1) the number of participants required and (2) researchers’ uncertainty about the proper statistical analyses required due to

Solomon Four-Group Design (Fake-Good)			
Group	Pre-test	Treatment	Post-test
1	O_1	X	O_2
2^+	O_3		O_4
3		X	O_5
4*			O_6

Solomon Four-Group Design (Fake-Bad)			
Group	Pre-test	Treatment	Post-test
5	O_7	X	O_8
2^+	O_3		O_4
6		X	O_9
4*			O_6

Note. + and * indicate the same control group was used for both groups

Figure 1

the complexity of the design (Braver & Braver, 1988). Note that since we have two treatments (fake-good and fake-bad) we had a total of six groups, with the two controls being the same for the two treatments.

To summarize, we employ a Solomon four-group design in this paper in order to (1) assess the fakability of a mixed measure of EI, (2) determine if individual differences influence the extent to which EI can be faked, (3) test whether fake-bad instructions affected scores more than fake-good instructions, (4) examine whether within-subjects designs provide a higher effect size than between-subjects design, and, (5) disentangle pre-testing main effects, as well the interaction between pre-testing and faking instructions.

Method

Sample

The research sample consisted of 300 undergraduate psychology students at a large Southeastern university in the United States. Students participated in the experiment in order to obtain extra credit in their respective courses. The mean age was 23.23 and the sample was 77 % female. The sample was ethnically distributed as follows: Caucasian 14 %, African American 10 %, Hispanic 64 %, and all other demographic groups (e.g., Asian, Native American) combined for the remaining 12 %.

Measures

General Mental Ability. General mental ability (GMA) was measured using the 50-item timed Wonderlic Personnel Test published by Wonderlic Incorporated (WPT; Wonderlic & Associates, 1983). The WPT reports a single score that is interpreted as a person's overall level of GMA. The WPT does not measure or report separate components or facets of GMA (e.g., verbal ability, quantitative ability). The reported reliability estimates (Cronbach's alpha) for the Wonderlic range from .78 to .95 and it has shown good convergent validity with other intelligence measures (e.g., the Wechsler Adult Intelligence Scales).

Personality. Personality was assessed with the Individual Perceptions Inventory (IPI; Goldberg, 1990), which is a 50-item measure of the Big Five personality characteristics. The IPI has shown strong convergent validity with other established measures of personality such as the 16PF and the NEO-PI (Goldberg, 1999). In the present study, scale reliability coefficients ranged from .76 (Emotional Stability) to .89 (Openness).

Emotional Intelligence. The 33-item emotional intelligence scale (EIS; Schutte et al., 1998) was used in the current study. Schutte and colleagues adopt Mayer and Salovey's (1997, p.10) definition of EI, which states that EI is "the ability to perceive accurately, appraise, and express emotion; the ability to access and/or generate feelings when they facilitate thought; the ability to understand emotion and emotional knowledge; and the ability to regulate emotions to promote emotional and intellectual growth." Although the EIS is based on Mayer and Salovey's (1997) ability model of EI, the instrument uses a non-performance based format and is described as a unidimensional self-report measure of EI (i.e., omitting the performance-based items of an ability measure). Therefore, the EIS should provide an excellent measure from which to assess potential faking effects for mixed model emotional intelligence. In addition, we chose to use the EIS because it is one of the most widely used measures of EI (Van Rooy & Viswesvaran, 2004). In fact, Van Rooy and Viswesvaran report that of all predictive studies reported in the published literature, approximately 25 % had used the EIS. Furthermore, Zeidner, Roberts, and Matthews (2002) state that this type of measure may be vulnerable to response sets and social desirability factors (e.g., faking), noting that studies designed to examine these concerns are lacking and urgently needed in the field of EI.

The psychometric aspects of the EIS are strong. Previous studies using the EIS (e.g., Chan, 2004; Riley & Schutte, 2003) have reported good reliabilities (i.e., coefficient alpha scores above .80) and the measure has been shown to be predictive of various outcomes by Schutte and colleagues (1998), who provide the full scale and appropriate psychometric results in their original study. In the current study, reliability estimates for the EIS ranged from .82 to .96. Example questions include, "I find it hard to understand the non-verbal messages of other people", "When I experience a positive emotion, I know how to make it last", and "When I feel a change in emotions, I tend to come up with new ideas".

Design and procedure

Two Solomon four-group designs (Solomon, 1949) were used to assess faking on a measure of emotional intelligence: one for fake-good and one for fake-bad. We chose this design because it (1) provides us a way to compare individuals in a control group to those that have been exposed to a manipulation as well as (2) allows for within-subject and between-subject comparisons. All participants were randomly assigned to one of six experimental conditions. The 6 groups were comprised as follows (see Figure 1): Group 1 took the EIS once under respond honest instructions and then once under fake-good instructions. Group 2 took the EIS twice; both times under respond honest instructions. Group 3 took the test only once (there was no pre-test) and was asked to fake-good. Group 4 was administered the EIS only one time (no pre-test) with respond honest instructions. Group 5 took the EIS once under respond honest instructions and then for a second time under fake-bad instructions. Group 6 took the test once (no pre-test) and they were asked to fake-bad. The same control groups were used in this condition as described in the fake-good design (i.e., Groups 2 and 4).

Similar to McFarland and Ryan (2000), participants in the fake-good conditions were told to take the test as if they were applying for a job they really wanted and to make themselves look as good as possible (to get selected for the position) when taking the test. Instructions in the fake-bad condition were similar and participants were told to take the test in an attempt to portray themselves as negatively as possible (so as to avoid selection). For the groups that took the EIS twice, a break of approximately one hour was given between the two administrations; subjects completed a distracter task during this time. The Wonderlic and IPI were administered after all EI responses were given.

Analyses

Analyses were conducted separately for fake-good and fake-bad treatments. Consider the fake-good instructed groups and the control groups (Groups 1-4). We first conducted a 2 x 2 ANOVA on the four post-test scores (see Figure 1). The two factors are treatment (fake vs. respond honestly) and testing (pre-testing vs. no pre-testing). A significant interaction suggests that testing interacts with instructions.

We used hierarchical multiple regression analyses to test whether the individual difference variables (GMA, Big Five) influenced a subject's ability to fake. Post-treatment EI scores in each condition (fake-good, fake-bad) were regressed onto pre-treatment EI scores (i.e., honest response EI), which was entered into the first step of the regression model, and then on the block of individual difference variables in step two. Support for each hypothesis was determined by the magnitude and significance of the incremental variance explained by the addition of the individual difference variables at step two, along with the significance of the regression coefficients.

We computed two effect sizes (Cohen's *d*-values) using the means reported in cells (O2 and O1, as well as O5 and O6). The first effect size is a fake-good within-subjects estimate whereas the second is a between-subjects estimate of the same. We again computed two effect sizes using the means reported in cells (O7 and O8, as well as O9 and O6). The first effect size is a fake-bad within-subjects estimate whereas the second is a between-subjects estimate of the same.

We compared the effect size from O2 and O1 with the effect size from O7 and O8. This compares the within-subjects fake-good effect size to the within-subjects fake-bad effect size. Similarly, the effect size from O5 and O6 compared with the effect size from O9 and O6 estimates the strength of the between-subjects fake-good versus the between-subjects fake-bad conditions.

Contrasting the effect size from O2 and O1 with the effect size from O5 and O6 compares within-subjects fake-good with between-subjects fake-good effect sizes. In addition, we evaluated the effect size from O7 and O8 with the effect size from O9 and O6, which compared within-subjects fake-bad with between-subjects fake-bad effect sizes.

Results

Table 1 summarizes the means, standard deviations, reliabilities, and zero-order correlations among all the variables used in the study. We also report the means, standard deviations and reliabilities for all conditions in the fake-good design (Table 2) and the fake-bad design (Table 3). The nine different conditions are identified as O1 to O9 in Tables 2 and 3 and correspond to the same conditions depicted in Figure 1. As shown in the tables, the reliability of the EIS was .86 both times when it was taken under normal instructions and no other measures were given before it. As would be expected, the reliability increased in all conditions (.90 to .95) where the EIS was taken once before (i.e., pre-testing occurred), with the exception of condition O8 (.82).

To assess the adequacy of random assignments we tested the equivalence of participants' EI scores who were asked to respond honestly in cells O1, O3 and O7. Results of the ANOVA were non-significant, F (2, 147) = 0.75, indicating equivalence among control groups.

We started with an examination of the interaction between faking instructions and pre-testing. For the fake-good instructions, the interaction between instructions and pre-testing was significant, F (1, 196) = 14.40, $p < .01$. For the fake-bad instructions, the interaction was non-significant, F (1, 196) = 0.05. In the fake-good design, there was a main effect found for treatment (fake vs. honest), F (1, 196) = 20.71, $p <$. 01, but there was no main effect for pre-testing, F (1, 196) = 0.46. For the fake-bad design, there was a main effect for both treatment (fake vs. honest), F (1, 196) = 647.70, $p < .01$, and pre-testing, F (1, 196) = 8.03, $p < .01$.

Regression analyses (summarized in Table 4) were conducted to test our hypotheses regarding individual difference variables, which predicted that GMA and agreeableness are positively related to the ability to fake an EI measure and that conscientiousness is negatively related to the ability to fake EI. For the fake-good condition (represented in the upper-half of Table 4), pre-test (i.e., honest response) EI scores did not explain a significant amount of variance in the ability to positively inflate EI scores (R^2 = .02). That is, pre-test scores explained only 2 % of the variance in post-test EI scores. The addition of the individ-

Table 1:
Correlations among study variables

Variables	Mean (*SD*)	1	2	3	4	5	6	7	8	9
1. Honest EI	130.14 (15.55)	(.86)								
2. Fake-good EI	141.60 (29.36)	--	(.94)							
3. Fake-bad EI	54.27 (25.93)	--	--	(.89)						
4. GMA	21.19 (6.05)	.01	.06	-.22*	--					
5. Conscientiousness	3.52 (.62)	.41*	.21*	.09	-.02	(.85)				
6. Extraversion	3.31 (.73)	.32*	.13	-.12	.09	.09	(.83)			
7. Openness	3.58 (.62)	.33*	.29*	-.12	.26*	.26*	.40*	(.89)		
8. Agreeableness	3.92 (.65)	.39*	.29*	-.27*	.21*	.30*	.43*	.49*	(.88)	
9. Emotional Stability	2.98 (.76)	.10	-.02	-.26*	.02	.12*	.23*	.23*	.20*	(.76)

Note. n = 100 for variables 1-3; n = 300 for variables 4-9
* $p < .05$

Table 2:
Fake-good comparisons

		Time 1			Time 2		
Group	Condition	α	Mean	SD	α	Mean	SD
1	Respond Honest (O1); Fake-good (O2)	.86	130.20	13.76	.93	147.02	17.11
2	Respond Honest (O3); Respond Honest (O4)	.86	129.30	12.60	.92	126.26	16.87
3	Fake-good Only (O5)	---	---	---	.95	135.90	22.06
4	Respond Honest Only (O6)	---	---	---	.90	134.02	13.17

n = 50 in each condition

Table 3:
Fake-Bad comparisons

		Time 1			Time 2		
Group	Condition	α	Mean	SD	α	Mean	SD
5	Respond Honest (O7); Fake-good (O8)	.86	130.41	13.31	.82	49.65	24.44
2	Respond Honest (O3); Respond Honest (O4)	.86	129.30	12.60	.92	126.26	16.87
6	Fake-good Only (O9)	---	---	---	.96	58.80	26.79
4	Respond Honest Only (O6)	---	---	---	.90	134.02	13.17

n = 50 in each condition

ual difference variables at step two did not explain a significant amount of incremental variance ($\Delta R^2 = .04$). In other words, GMA and the Big Five variables only explained an additional 4 % of the variance in post-test EI scores. Further examination of the standardized regression coefficients (*β)* in step two indicated that neither GMA nor any of the Big Five variables were related to higher post-test EI scores.

However, it is also possible that our non-significant findings may have been partly attributable to ceiling effects. Therefore, we accounted for ceiling effects by conducting additional analyses in which post-test EI scores were transformed to represent percentages of the maximum possible deviation based upon pre-test scores[5]. We use the following example to illustrate this process. Consider that the maximum score on the EIS is 165. An individual who scores 135 on the pre-treatment (or pre-test) administration of the EI measure would be able to maximally inflate their score by 30 points (i.e., 165 - 135 = 30). If the same individual scores a 155 after being asked to fake-good (in the post-treatment condition), they would have improved their score by 20 points. Transforming this number to a percentage of the possible maximum deviation (30 points) gives a score of 20/30 = .66. Thus, .66 would be the

[5] We thank an anonymous reviewer for this suggestion.

dependent variable regressed onto the individual difference variables; in the previous regression analyses, a score of 155 would have been the dependent variable for this individual.

The analyses in which we attempted to control for ceiling effects are represented in the far right columns of Table 4 under headings %FG. The results suggest that transforming the dependent variable did not alter our findings. Pre-test (i.e., honest response) EI scores did not explain a significant amount of variance in the percentages of possible deviation (R^2 = .03), and the addition of the individual difference variables at step two did not explain a significant amount of incremental variance (ΔR^2 = .06). These findings suggest that a ceiling effect may not be present in these data.

For the fake-bad condition (represented in the lower-half of Table 4), pre-test (i.e., honest response) EI scores explained a significant amount of variance in the ability to lower EI scores (R^2 = .16, $p < .05$). Addition of the individual difference variables at step two explained a significant amount of incremental variance (ΔR^2 = .14, $p < .05$). Examination of the standardized regression coefficients in step two indicated that "honest response" EI scores (β = -.41, $p < .05$) and Conscientiousness (β = .29, $p < .01$) were significantly related to lower post-test EI scores. That is, for every one standard deviation unit that Conscientiousness increases, post-test EI scores are expected to go up .29 standard deviation units. Additionally, we conducted transformations (in a process similar to the one described above) to examine for the presence of floor-effects. The results of these analyses (on the right side of Table 4 under the heading %FB) did not alter our initial findings, suggesting the absence of a floor-effect.

The effect size for the within-subjects comparison (O2 and O1) was 1.08 (t = 5.84, p < .01), indicating that instructions to fake-good significantly increased EI scores. Next, we computed the between-subjects effect size comparing O5 (honest response) and O6 (fake-good). The effect size was non-significant (t = 0.10). In the fake-bad condition, we compared the means reported in cells (O7 and O8; within-subject estimate). The effect size was 4.07 (t = 17.55, $p < .01$), indicating that participants instructed to fake-bad could significantly lower their scores. We also compared the means from cells O9 and O6 (between-subjects estimate). The between-subjects d-value was 3.56 (t = 17.82, $p < .01$). Thus, participants were able to significantly lower their EI scores in both types of comparisons for the fake-bad design.

We also compared the fake-good effect size (O2 and O1) with the fake-bad effect size (O7 and O8). This compares two within-subjects effect sizes. The effect sizes were 1.08 and 4.07, respectively. We also compared the fake-good between-subjects effect size (O5 and O6) with fake-bad between-subjects effect sizes (O9 and O6). The d-values were 0.10 and 3.56, respectively. There was a clear difference in fake-good and fake-bad irrespective of the experimental design (within or between-subjects design) and participants were able to fake-bad to a greater extent than they could fake-good. As with the regression analyses reported above, we also wanted to rule out that these differences were not caused by ceiling- or floor-effects. Therefore, we transformed O2–O1 and O8–O7 score changes to represent percentages in maximum possible deviations (as described in the section above). These supplementary results did not alter our findings; participants in the fake-bad condition were able to fake to a significantly greater degree (84 % of maximum possible deviation) than participants in the fake-good condition (56 % of maximum possible deviation). The d-value of 1.02 (t = 5.31, $p < .01$) suggests that ceiling- and floor-effects do not account for our findings.

Table 4:
Individual differences regression results for EI faking score and percentage of possible deviation in faking score

Individual differences	β (FG)	β (FG)	*t-value*	β (%FG)	β (%FG)	*t-value*
Step 1						
Respond Honest EI Scores	.14		.99	-.18		-1.28
R^2	**.02**			**.03**		
Step 2						
Respond Honest EI Scores		.09	.47		-.13	-.68
Wonderlic (GMA)		-.14	-.82		-.18	-1.07
Agreeableness		.15	.73		.09	.45
Conscientiousness		.00	-.01		-.08	-.40
Emotional Stability		-.11	-.64		-.11	-.65
Extraversion		.05	.22		.07	.38
Openness to Experience		.07	.33		-.11	-.54
R^2		**.06**			**.09**	
ΔR^2		**.04**			**.06**	

Individual differences	β (FB)	β (FB)	*t-value*	β (%FB)	β (%FB)	*t-value*
Step 1						
Respond Honest EI Scores	-.40*		-3.02	-.29		-2.11
R^2	**.16***			**.09***		
Step 2						
Respond Honest EI Scores		-.41*	-2.46		-.30	-1.71
Wonderlic (GMA)		-.24	-1.63		-.18	-1.21
Agreeableness		-.06	-.39		-.02	-.09
Conscientiousness		.29*	1.97		.32*	2.13
Emotional Stability		-.14	-.94		-.21	-1.39
Extraversion		-.01	-.07		.08	.45
Openness to Experience		-.02	-.08		-.11	-.61
R^2		**.30**			**.25**	
ΔR^2		**.14***			**.13**	

Note. FG = total EI score after being instructed to fake good; %FG = percentage of possible maximum upper deviation that subjects inflated their scores; FB = total EI score after being instructed to fake bad; %FB = percentage of possible maximum lower deviation that subjects deflated their scores; R^2 = percentage of variance in the dependent variable explained by the individual difference variables; β = represents the expected change in the dependent variable that results from a change of one standard deviation in the specific individual difference variable.
n = 50 [respond honest (RH) and then fake good (FG)]
n = 50 [respond honest (RH) and then fake bad (FB)]
* $p < .05$

To compare the two designs, we first looked at the fake-good condition and compared the effect size from O2 and O1 (within-subjects) with the effect size from O5 and O6 (between-subjects). The *d*-values were 1.08 and 0.10, respectively. In the within-subjects design, participants improved their scores more when instructed to fake compared to improvements in the between-subjects design. These results are more pronounced given our earlier analysis showing successful random assignment to groups. For fake-bad instructions, the values were 4.07 and 3.56, respectively (compared the effect size from O7 and O8 with the effect size from O9 and O6). For fake-bad, it appears that the design does not influence fakability magnitudes.

Discussion

The current research was designed to evaluate how responses vary on a widely used measure of EI depending on the administrative instructions provided (Ellingson, Sackett, & Hough, 1999). The results clearly demonstrate that a self-report measure of emotional intelligence is susceptible to intentional faking attempts. Furthermore, the magnitude of faking is much more pronounced when participants are instructed to fake-bad rather than fake-good. Our results also indicate that previous exposure to a test influences a person's subsequent ability to fake. Specifically, individuals who first took the test under normal response instructions were better able to fake than subjects who were not previously exposed to the test. Lastly, we did not find that GMA or the Big Five played a significant role in influencing fakability.

Theoretical implications

The history of experimental design has seen rapid advances during the last century. In fact, Solomon (1949) notes that a control group was not utilized until the first decade of the 1900's. Since then the Solomon four-group design has been recognized as the most elaborate and comprehensive design available to researchers (Braver & Braver, 1988; Huck & Sandler, 1973, Shadish et al., 2002). Even so, few studies have used the design to assess intervention effectiveness and, as far as we are aware, none have used it to assess the fakability of any of the frequently used non-cognitive selection tools (cf. Viswesvaran & Ones, 1999). This study can therefore serve as an illustration of the design for use in future fakability studies.

With other non-cognitive predictors such as personality (e.g., Barrick & Mount, 1991; Hurtz & Donovan, 2000; Mount, Barrick, & Stewart, 1998), biodata (e.g., Russell, Mattson, Devlin, & Atwater), and integrity (Ones, Viswesvaran, & Schmidt, 1993), estimates of predictive validity were generated and research was subsequently conducted to determine if tests of the constructs could be faked and to what extent. Indeed, successful faking efforts have been shown for these measures (e.g., Hough, Eaton, Dunnette, Kamp & McCloy, 1990; Kluger, Reilly, & Russell, 1991; McFarland & Ryan, 2000; Viswesvaran & Ones, 1999) and EI can now be added to this list. To the extent that mixed model EI can be faked, it appears to function similarly to personality variables. Although our statement is based on one measure of EI, albeit a commonly used one (Van Rooy & Viswesvaran, 2004), the results are likely to generalize to other self-report measures of the construct.

Past research has shown that Big Five personality factors, namely Conscientiousness and Emotional Stability, are related to faking (McFarland & Ryan, 2000). Our results, however, did not extend this finding to EI. With the exception of Conscientiousness in the fake-bad condition, personality did not play a role in the ability to fake responses on a measure of EI.

Surprisingly, GMA was also unrelated to faking. Perhaps this provides further evidence that mixed model EI is orthogonal to GMA (e.g., Van Rooy et al., 2006).

The EIS used in this study, though originally described as based on the ability-based model of EI, does not use a performance-based scoring format (i.e., expert or consensus); future studies now need to experimentally replicate the current results with one of these measures (e.g., MSCEIT V2.0; Mayer et al., 2003). Briefly, consensus scoring utilizes the response frequency of another group to determine what the answer should be and expert scoring relies on the answers given be experts in the field, which are assumed to be correct. Not only will it be of interest to see the extent, if any, to which performance-based measures can be faked, but also if differences vary according to the scoring method employed.

Practical implications

Because EI is a new area of research, many applicants will encounter measures for the first time and may not know what type of scores are "desirable" for the job in question. Thus, even though applicants may be able to fake, they might be uncertain what the preferred profile actually looks like. For instance, in personality research, there is still disagreement about what openness to experience taps and what a desirable response is (McFarland & Ryan, 2000). As another example, certain organizations may not want employees who are too high on a scale of conformity because it may limit independent and creative thinking. Similarly, research may show that individuals who are too high on certain dimensions of EI might actually make decisions that are overly emotionally laden and not always in the best interest of the organization. If an applicant were to fake a measure of EI, they would still have to determine what EI profile fits the job of interest (e.g., salesperson vs. computer programmer). Of course, now that it has been shown that people can fake, the issue must be extended to high-stake (e.g., employment) situations (Hough & Schneider, 1996). Similarly, just because someone can fake a test and "inflate" their score, it does not necessarily mean that this inflation will be in the direction desired by the hiring organization (Ones, Viswesvaran, & Reiss, 1996).

Like past research (Viswesvaran & Ones, 1999), our results also show that within-subject comparisons produced larger faking differences than between-subject comparisons; this held true for faking good but was not as pronounced for faking bad. Within-subject comparisons involve having the same person take a test twice: once under each response set and then comparing score differences. Between-subject comparisons use participants from two different groups (i.e., with different response instructions) to examine score differences. This contradicts previous research that has not found order effects for Big Five personality dimensions, biodata, and integrity measures (McFarland & Ryan, 2000). It is important to look at both of these in experimental settings, but the within-subject comparison is more important for generalization to employment settings. If an employer utilizes a measure of EI in a selection system, our results clearly indicate that an applicant could score considerably higher if they were given the same measure of EI a second time and wanted to portray themselves in a more favorable light. Employers would therefore be prudent to prevent retakes unless a certain minimum amount of time has passed since the original administration.

Limitations and future research

The EIS employed in this study is a relatively short measure (i.e., 33 items) and it may be easier to fake than longer measures, in which response patterns become more difficult to maintain. The EIS test score is also scored based on a single overall score, whereas other measures of EI often assess subdimensions. Future research should look at how subdimension scores vary according to response instructions. As is typical for faking studies, a student sample was used. Although this is a limitation, it will remain one in this line of research, as most organizations are unlikely to allow researchers to experimentally manipulate or even access to candidate pools. Furthermore, a laboratory study is necessary in order to provide the control needed to ensure that individuals did attempt to fake (McFarland & Ryan, 2000). More importantly, differences between incumbents and applicants have also been reported (cf. Birkeland, Manson, Kisamore, Brannick, & Liu, 2003), which suggests that future research should try to replicate the four-group experimental design not only with employees but also with applicants (something that may not be possible). Even so, our results provide a solid base for the refinement of existing measures and construction of new measures of EI.

It is also important to note that the time-interval between the first and second administration was short (1 hour) and this could have increased the stability of the scores (lowering the effect sizes for faking instructions). However, this may not be a major concern. First, as seen in Tables 2 and 3, the within-subjects effect sizes were larger than the between-subjects effect sizes. Second, the test-retest reliability for Group 2 (respond honest twice) was only .74, which is a moderate estimate that compares with other stability coefficients reported for personality scales (cf. Viswesvaran & Ones, 2000). A counterargument can be made that within-subjects estimates were high because participants tried to contrast their scores with remembered responses from the first administration (so as to comply with experimenter expectancies). While possible, the moderate stability coefficient of .74 does not support this potential alternative hypothesis. Nonetheless, researchers should replicate this study using longer retest intervals.

We should also note that observed variability was higher in the fake conditions than in the respond honest conditions (fake-good = 17.11 vs. 13.76; fake-bad = 24.44 vs. 13.31). However, the reliability increased from the first to second administration suggesting that the increased variability is reflected in true score variability. Thus, while plausible, this effect does not drown out our main experimental manipulation. Moreover, it is important to stress that the current research investigated the general potential to fake good or bad on a measure of EI. Future research should explore these effects in specific occupations, jobs, organizations and situations.

In line with other research (McFarland & Ryan, 2002; Mersman & Shultz, 1998), we found that there was considerably less room for faking good than there was for faking bad. It is therefore possible for some to suggest that a self-report measure of EI can only be faked to a certain extent because there is not a lot of room to increase scores. However, in our analyses, we attempted to control for both ceiling- and floor-effects by transforming EI scores into percentages of possible maximum deviation. The fact that such transformations did not affect our findings or conclusions indicates that ceiling- and floor-effects may not be present. It is nevertheless important to note that pre-testing still facilitated increased scores when participants were instructed to fake-good.

Quantifying the extent of faking is the first step in gauging the problem (Viswesvaran & Ones, 1999). We did not report results for gender differences in the current study because the sample was predominantly female and the number of males in each condition was not sufficient. An examination of the two, although lacking adequate power, showed that mean scores were roughly comparable in the respond honest conditions. Interestingly, males were better than females in their efforts to fake when told to respond in as negative or positive a manner as possible. It could very well be that males are higher self-monitors and are more likely to engage in socially desirable responding (Day, Schleicher, Unckless, & Hiller, 2002) than females. Future research should examine gender differences in this area.

Conclusion

Emotional intelligence is now being widely used in applied settings (Murphy, 2006). Although psychologists may have some reservations about the uniqueness of this construct vis a vis more traditional predictors, it is imperative to empirically test whether such commonly used measures are fakable. Using a Solomon four-group design, we were able to address this question and estimate some of the confounding effects, such as pre-testing and individual differences. Hopefully, this research will spur further investigations on the applicability of EI.

References

Alliger, G. M., & Dwight, S. A. (2000). A meta-analytic investigation of the susceptibility of integrity tests to faking and coaching. *Educational and Psychological Measurement, 60*, 59-72.

Ashkanasy, N. M. & Daus, C. S. (2005). Rumors of the death of emotional intelligence in organizational behavior are vastly exaggerated. *Journal of Organizational Behavior, 26*, 441- 452.

Barchard, K. A. (2003). Does emotional intelligence assist in the prediction of academic success? *Educational & Psychological Measurement, 65*, 840-858.

Bar-On, R. (1997). *Bar-On emotional quotient inventory: Technical manual*. Toronto: Multi-health Systems.

Barrick, M. R., & Mount, M. K. (1991). The Big Five personality dimensions and job performance: A meta-analysis. *Personnel Psychology, 44*, 1-26.

Becker, T. E., & Martin, S. L. (1995). Trying to look bad at work: Methods and motives for managing poor impressions in organizations. *Academy of Management Journal, 38*, 174-199.

Birkeland, S., Manson, T., Kisamore, J. L., Brannick, M. T., & Liu, Y. (2003, April). *A meta-analytic investigation of job-applicant faking on personality measures*. Poster presented at the 18th annual conference of the Society for Industrial and Organizational Psychology, Orlando, FL.

Brackett, M. & Mayer, J. (2003). Convergent, discriminant, and incremental validity of competing measures of emotional intelligence. *Personality and Social Psychology Bulletin, 29*, 1147-1158.

Braver, M. C. W., & Braver, S. L. (1988). Statistical treatment of the Solomon Four-Group Design: A meta-analytic approach. *Psychological Bulletin, 104*, 150-154.

Brody, N. (2004). What cognitive intelligence is and what emotional intelligence is not. *Psychological Inquiry, 15,* 234-238.

Chan, D. (2004). Perceived emotional intelligence and self-efficacy among Chinese secondary school teachers in Hong Kong. *Personality and Individual Differences, 36,* 1781-1795.

Conte, J. M. (2005). A review and critique of emotional intelligence measures. *Journal of Organizational Behavior, 26*, 433-440.

Cote, S., & Miners, T. H. C. (2006). Emotional intelligence, cognitive intelligence, and job performance, *Administrative Science Quarterly, 51*, 1-28.

Daus, C. S. & Ashkanasy, N. M. (2005). The case for the ability-based model of emotional intelligence in organizational behavior. *Journal of Organizational Behavior, 26*, 453-466.

Day, D. V., Schleicher, D. J., Unckless, A. L., & Hiller. N. J. (2002). Self-monitoring personality at work: A meta-analytic investigation of construct validity. *Journal of Applied Psychology*, 87, 390-401.

Deller, J., Ones, D. E. Viswesvaran, C., & Dilchert, S. (2006). Considering response distortion in personality measurement for industrial, work and organizational psychology research and practice: Special Issue. *Psychology Science, 48(3*).

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.

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.

Frie, R. L., Griffith, R. L., McDaniel, M. A., Snell, A. F., & Douglas, E. F. (1997). *Faking of noncognitive measures: Factor invariance using multiple groups LISREL.* Poster presented at the 12th annual conference of the Society for Industrial and Organizational Psychology, St. Louis, MO.

Goldberg, L. R. (1990). An alternative 'description of personality': The big-five factor structure. *Journal of Personality and Social Psychology*, *59*, 1216-1229.

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.

Gottfredson, L. S. (1997). Mainstream science on intelligence: An editorial with 52 signatories, history and bibliography. *Intelligence, 24,* 13-23.

Hausknecht, J. P., Trevor, C. O., & Farr, J. L. (2002). Retaking ability tests in a selection setting: Implications for practice effects, training performance, and turnover. *Journal of Applied Psychology, 87*, 243-254.

Hough, L. M., Eaton, N. K., Dunnette, M. D., Kamp, J. D., McCloy, R. A. (1990). Criterion related validity of personality constructs and the effects of response distortion on those validities. *Journal of Applied Psychology, 75*, 581-595.

Hough, L. M., & Schneider, R. J. (1996). The frontiers of I/O personality research. In K. R. Murphy (Ed.), *Individual differences in behavior in organizations* (pp. 31-88). San Francisco: Jossey-Bass.

Huck, S. W., & Sandler, H. M. (1973). A note on the Solomon 4-Group Design: Appropriate statistical analyses. *The Journal of Experimental Education, 42*, 54-55.

Hunter, J. E., & Schmidt, F. L. (1990). *Methods of meta-analysis*. Newbury Park: Sage Publications.

Hurtz, G. M., & Donovan, J. J. (2000). Personality and job performance: The Big Five revisited. *Journal of Applied Psychology, 85,* 869-879.

Janovics, J., & Christiansen, N. D. (2004). *Emotional intelligence in a selection context: Criterion related validity and vulnerability to response distortion*. In K. E. Fox & R. P. Tett (Eds.), 19th Annual Conference for the Society for Industrial and Organizational Psychology. Chicago, IL.

Jensen. A. (1998). *The g factor: The science of mental ability*. Westport, Connecticut: Praege.

Kluger, A. N., Reilly, R. R., & Russell, C. J. (1991). Faking biodata tests: Are option-keyed instruments more resistant? *Journal of Applied Psychology, 76*, 889-896.

Matthews, G., Zeidner, M., & Roberts, R. D. (2002). *Emotional intelligence: Science and myth*. Boston, MA: MIT Press.

Mayer, J. D., & Salovey, P. (1997). What is emotional intelligence? In P. Salovey, & D. Sluyter (Eds.), *Emotional development and emotional intelligence: educational implications* (pp. 3–34). New York: Basic Books.

Mayer, J. D., Salovey, P., Caruso, D. R., & Sitarenios, G. (2003). Measuring emotional intelligence with the MSCEIT V2.0. *Emotion, 3*, 97-105.

McFarland, L. A., & Ryan, A. M. (2000). Variance in faking across noncognitive measures. *Journal of Applied Psychology, 85*, 812-821.

Mersman, J. L. & Shultz, K. S. (1998). Individual differences in the ability to fake on personality measures. *Personality and Individual Differences. 24*, 217-227.

Mook, D. G. (1983). In defense of external invalidity. *American Psychologist, 38*, 379-387.

Mount, M. K., Barrick, M. R., & Stewart, G. L. (1998). Five-factor model of personality and performance in jobs involving interpersonal interactions. *Human Performance, 11*, 145-165.

Mueller-Hanson, R., Heggestad, E. D., & Thornton, G. C., III (2003). Faking and selection: Considering the use of personality from select-in and select-out perspectives. *Journal of Applied Psychology, 88*, 348-355.

Murphy, K. R. (Ed.). (2006). The Case Against Emotional Intelligence: What are the Problems and How Can They be Fixed? Mahwah, NJ: Erlbaum.

Ones, D. S., & Viswesvaran, C. (1998). The effects of social desirability and faking on personality and integrity assessment for personnel selection. *Human Performance, 11*, 245-269.

Ones, D. S., Viswesvaran, C., & Reiss (1996). Role of social desirability in personality testing for personnel selection: The Red Herring. *Journal of Applied Psychology, 81*, 660-679.

Ones, D. S., Viswesvaran, C., & Schmidt, F. L. (1993). Comprehensive meta-analysis of integrity test validities: Findings and implications for personnel selection and theories of job performance. *Journal of Applied Psychology, 78*, 679-703.

Paulhus, D. L. (1984). Two-component models of social desirable responding. *Journal of Personality and Social Psychology, 46*, 598-609.

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., Lysy, D. C., & Yik, M. S. M. (1998). Self-report measures of intelligence: Are they useful as proxy IQ tests? *Journal of Personality, 66*, 525-554.

Petrides, K. V., & Furnham, A. (2000). On the dimensional structure of emotional intelligence. *Personality and Individual Differences, 29*, 313-320.

Riley, H. & Schutte, N. (2003). Low emotional intelligence as a predictor of substance-use problems. *Journal of Drug Education, 33,* 391-398.

Roberts, R. D., Zeidner, M., & Matthews, G. (2001). Does emotional intelligence meet traditional standards for an intelligence? Some new data and conclusions. *Emotion, 1*, 196-231.

Russell, C. J., Mattson, J., Devlin, S. E., & Atwater, D. (1990). Predictive validity of biodata items generated from retrospective life experience essays. *Journal of Applied Psychology, 75*, 569-580.

Saklofske, D. H., Austin, E. J., & Minski, P. S. (2003). Factor structure and validity of a trait emotional intelligence measure. *Personality and Individual Differences, 34,* 707-721.

Salovey, P., & Mayer, J. D. (1990). Emotional intelligence. *Imagination, Cognition, and Intelligence, 9*, 185-211.

Salovey, P., Mayer, J. D., & Caruso, D. (2002). The positive psychology of emotional intelligence. In C. R. Snyder & S. J. Lopez (Eds.), *The Handbook of Positive Psychology*. New York: Oxford University Press.

Schulte, Ree, & Carretta (2004). Emotional intelligence: not much more than g and personality. *Personality and Individual Differences, 37,* 1059-1068.

Schmit, M. J., & Ryan, A. M. (1993). The Big Five in personnel selection: Factor structure in applicant and nonapplicant populations. *Journal of Applied Psychology, 78*, 966-974.

Schutte, N. S., Malouff, J. M., Hall, L. E., Haggerty, D. J., Cooper, J. T., Golden, C. J., & Dornheim, L. (1998). Development and Validation of a measure of emotional intelligence. *Personality and Individual Differences, 25*, 167-177.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin.

Sjoberg, L. (2001). Emotional intelligence: A psychometric analysis. *European Psychologist, 6*, 79-95.

Smith, D. B., Hanges, P. J., & Dickson, M. W. (2001). Personnel selection and the Five-Factor Model: Reexamining the effects of applicant's frame of reference. *Journal of Applied Psychology, 86*, 304-315.

Snell, A. F., & McDaniel, M. A. (1998). *Faking: Getting data to answer the right questions*. Poster presented at the 13th annual conference of the Society for Industrial and Organizational Psychology, Dallas, TX.

Solomon, R. L. (1949). An extension of control group design. *Psychological Bulletin, 46*, 137-150.

Van Rooy, D. L., Dilchert, S., Viswesvaran, C., & Ones, D. S. (2006). Multiplying Intelligences: Are General, Emotional, and Practical Intelligences Equal? In K. R. Murphy (Ed.), *The Case Against Emotional Intelligence: What are the Problems and How Can They be Fixed?* Mahwah, NJ: Erlbaum.

Van Rooy, D. L., & Viswesvaran, C. (2004). Emotional Intelligence: A Meta-Analytic Investigation of predictive Validity and Nomological Net. *Journal of Vocational Behavior*, *65*, 71-95.

Van Rooy, D. L., Viswesvaran, C., & Pluta, P. (2005). An examination of construct validity: What is this thing called emotional intelligence? *Human Performance, 18,* 445-462.

Viswesvaran, C., & Ones, D. S. (1999). Meta-analysis 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*, 197-224.

Wonderlic & Associates (1983). Wonderlic Personnel Test manual. Northfield, IL: Wonderlic, Inc.

Wong, C., & Law, K. S. (2002). The effects of leader and follower emotional intelligence on performance and attitude: An exploratory study. *The Leadership Quarterly, 13,* 243–274

Zeidner, M., Roberts, R. D., & Matthews, G. (2002). Can emotional intelligence be schooled? A critical review. *Educational Psychologist, 37*, 215-231.