Examining the necessity for and utility of the Psychopathic Personality Inventory-Revised (PPI-R) validity scales. — Source link ↗

Jaime L. Anderson, Martin Sellbom, Dustin B. Wygant, John F. Edens

Related papers:

- Factor structure of the psychopathic personality inventory: validity and implications for clinical assessment.

- Development and preliminary validation of a self-report measure of psychopathic personality traits in noncriminal populations.

- Evidence for response bias as a source of error variance in applied assessment.

- An examination of the Psychopathic Personality Inventory's nomological network: a meta-analytic review.

- Triarchic conceptualization of psychopathy: developmental origins of disinhibition, boldness, and meanness.

Examining the Necessity for and Utility of the Psychopathic Personality Inventory—Revised (PPI–R) Validity Scales

Jaime L. Anderson and Martin Sellbom
University of Alabama

Dustin B. Wygant
Eastern Kentucky University

John F. Edens
Texas A&M University

The present study aimed to investigate the need for and utility of the Psychopathic Personality Inventory—Revised (PPI–R) Deviant Responding (DR) and Virtuous Responding (VR) validity scales in identifying overreporting and underreporting, respectively. Since the PPI–R was published, there has not been an independent peer-reviewed examination of these scales. Participants were 384 undergraduate individuals asked to respond to the PPI–R under standard, underreporting, or overreporting instructions. A comparison group consisting of 200 forensic psychiatric patients was also used for the overreporting analyses. Effects of response bias on mean elevations on the PPI–R substantive scales were examined along with the effects on the PPI–R total, factor, and content scales' correlations with other relevant extratest measures of psychopathy. Mean elevations differed significantly, and correlations with extratest measures of psychopathy were significantly lower. Substantial decrement in psychometric validity of PPI–R scores was observed in the simulation conditions. In addition, the utility of the PPI–R validity scales in differentiating between groups was also determined. Both the VR and DR scales showed utility in differentiating between their respective dissimulation condition and the comparison groups, with acceptable rates of sensitivity and specificity.

Keywords: psychopathy, Psychopathic Personality Inventory- Revised, PPI-R, validity scales, malingering

Psychopathy is a personality syndrome characterized by guiltlessness, failure to form close attachments, callousness, and behavioral traits, such as impulsivity and antisocial actions (Cleckley, 1941/1976). Information about psychopathic traits can play a major role in risk evaluations, criminal sentencing, treatment considerations, and employment evaluations (DeMatteo & Edens, 2006; Wu & Lebreton, 2011). Individuals with psychopathy are more prone to violence (Hare & Neumann, 2008) and more likely to recidivate (Skeem, Polaschek, Patrick, & Lilienfeld, 2011), making this a disorder of substantial public health relevance.

Any forensic psychological evaluation comes with the potential of response bias, or misrepresentation of symptomatology and personality characteristics, because of the presence of external incentives. Although they do not appear disproportionately successful in their attempts at feigning (Marion et al., 2012; Poythress, Edens, & Watkins, 2001), individuals who score higher on psy-chopathy measures are more likely to attempt to misrepresent their symptoms on psychological tests (Kucharski, Falkenbach, Egan, & Duncan, 2006; Marion et al., 2012). Indeed, Kucharski et al. (2006) estimated that on the basis of the Structured Interview of Reported Symptoms (SIRS), 21.6% of individuals in the moderate psychopathy group and 33.3% of individuals in the high psychopathy group met the malingering cut-off, whereas only 8.1% of those in the low psychopathy group met this cut-off. Thus, assessing response bias is particularly important in evaluating psychopathy.

The Psychopathic Personality Inventory—Revised (PPI–R; Lilienfeld & Widows, 2005) is a 154-item self-report questionnaire designed for the assessment of psychopathy, particularly in nonincarcerated populations. However, it is also frequently used in forensic populations, and the PPI–R manual provides norms for both community and correctional settings. It provides a total score of psychopathy, three factor scores, and eight content scale scores. In addition, the PPI–R includes three validity scales designed to measure random responding, and the over- and underreporting of symptoms. Inconsistent Responding (INC) measures random responding, the Deviant Responding scale (DR) indexes overreporting, and the Virtuous Responding scale (VR) measures underreporting. The VR scale is designed to measure social desirability and positive impression management, whereas the DR scale is designed to measure the endorsement of odd or bizarre responses. Although these scales have been included on the PPI–R since the test was commercially released in 2005, they have not yet been formally evaluated in independent, peer-reviewed investigations. The present study was designed to fill this gap in the literature.

Whether personality inventories actually benefit from the inclusion of measures of response distortion has been a source of ongoing debate in the field. For instance, the developers of the NEO PI–R and NEO FFI have argued that validity scales are not needed for self-report personality and psychopathology measures, largely because these scales have not been shown to increase criterion validity (Costa & McCrae, 1992). More recently, McGrath and colleagues (2010) raised concerns about the need for validity scales and other response bias indicators. They asserted that response bias may not change protocol validity in most cases, that there is insufficient research to show the utility of validity measures, that validity scales may not adequately assess for response bias, and thus may not even be necessary in many testing circumstances. This introduces a new question in the case of the PPI–R; do these claims by McGrath et al. (2010) have some legitimacy with regard to the two validity scales included on this test? That is, can the PPI–R validity scales detect response bias and would such response bias reduce the test's psychometric validity?

The need for and utility of validity measures has been shown in research on some widely used personality measures such as the Minnesota Multiphasic Personality Inventory 2 (MMPI–2; e.g., Rogers, Sewell, Martin, & Vitacco, 2003), MMPI–2 Restructured Form (MMPI–2–RF; e.g., Sellbom & Bagby, 2010; Wygant et al., 2011), NEO Personality Inventory (NEO PI–R; e.g., Sellbom & Bagby, 2008b), and Personality Assessment Inventory (PAI; e.g., Edens & Ruiz, 2006; Hawes & Boccaccini, 2009; Sellbom & Bagby, 2008a). These studies and reviews show that the utility of these scales in detecting response bias is unquestionable, and that substantive scale profiles are altered when individuals are dissimulating. However, more research is needed on these well-established measures to show that the psychometric validity is moderated by response bias. Some preliminary evidence has been shown for the MMPI–2–RF (e.g., Burchett & Ben Porath, 2010) and the NEO PI–R (e.g., Caldwell-Andrews et al., 2001), but there is very little work on the validity scales of the PPI–R, and further research in general is needed to address McGrath and colleagues' (2010) claims.

There has only been one peer-reviewed article published on the PPI–R validity scales (Nikolova, Hedry, Douglas, Edens, & Lilienfeld, 2012), but this study only examined the INC scale. For the original version of the PPI (Lilienfeld & Andrews, 1996), Edens (2004) showed that mean elevations on the PPI subscales reflecting socially deviant traits are heavily influenced by both over- and underreporting in an undergraduate sample. When examining the validity scales, Edens, Buffington, and Tomicic (2000) found that the DR scale was associated with high sensitivity and specificity rates when differentiating between undergraduate students who responded to the PPI in an honest versus overreporting manner. Similarly, the VR scale was able to significantly differentiate between honest and underreporting simulation groups, albeit at somewhat lower sensitivity and specificity rates relative to DR (Edens, Buffington, Tomicic, & Riley, 2001). However, each of these studies relied only on samples of undergraduate students with no comparison groups from civil or psychiatric forensic patient samples. Rogers (2008) has argued that validity scales should not be used in the clinical evaluation of overreporting unless they can demonstrate utility in differentiating between overreporters and genuine psychiatric patients.

The PPI–R professional manual (Lilienfeld & Widows, 2005) reports research on the DR and VR scales using a very small ($n = 40$) sample of undergraduate students. When the revised version of this measure was developed, the DR scale was changed significantly from its original version, whereas the VR scale was completely recreated in its revised form. The research presented in the manual on these two scales did show that both VR and DR were able to differentiate undergraduates who were asked to over- or underreport psychopathy traits from those who took the test under standard instructions. Importantly, we included a forensic psychiatric inpatient sample as a comparison group to address concerns about excessive reliance on college student control samples who would exhibit little in the way of serious psychopathology. The research reported in the manual, however, has not been through the vetting process that is typical for a peer-reviewed journal and also has limitations due to the very small sample size and the lack of a patient comparison group (for overreporting). Therefore, more research is needed on these scales to ascertain their utility in the detection of response bias.

The current investigation had two general objectives. First, we examined the need for validity scales on the PPI–R by determining the effects that under- and overreporting have on mean elevations on the PPI–R substantive scales and also the effects that under- and overreporting have on the PPI–R total, factor, and content scales' correlations with other relevant extratest measures of psychopathy. Second, we sought to determine the psychometric validity of the DR and VR scales and examined the utility of these scales in differentiating between groups instructed to over- and underreport and those who completed the measure under standard instructions. On the basis of the very limited literature on these scales (including the work on their original PPI counterparts), we expected (a) significantly different mean elevation between the groups on the substantive scales; (b) correlations with conceptually relevant psychopathy scales to decrease significantly in magnitude when response bias was introduced, which would suggest that response bias moderates the validity of the PPI–R scales; and (c) the VR and DR scales to show utility in identifying individuals who were under- or overreporting symptoms, respectively.

Method

Participants

College sample. The first sample consisted of 384 potential participants from two public universities in the southeastern United States. Upon excluding 49 individuals found to engage in excessive random responding (per the PPI–R manual's cutoff of 44 on the Inconsistency scale), 111 male and 224 female undergraduate students comprised the sample. The individuals engaging in random responding were removed to reduce random measurement error. Since the PPI–R was the last administered instrument and the only measure that indexes inconsistent responding, there is no way to assure that participants took the earlier measures seriously. Therefore, we excluded these participants entirely. Participants completed the study in small groups under the supervision of a trained research assistant. Each participant was randomly assigned to one of three groups: honest responding, overreporting, and underreporting. Among these participants, 127 were assigned to the honest responding group, 94 to the overreporting group, and

114 to the underreporting group. Participants had a mean age of 19.9 years ($SD = 3.88$), ranging from 17 to 54. Participants were also predominantly Caucasian (81.3%), with the largest minority being African American (14.8%) and the remaining belonging to other racial or ethnic groups (3.9%). There were no statistically significant differences between the groups on the demographic variables, supporting the random assignment.

Forensic psychiatric sample. Undergraduate students do not constitute an appropriate comparison group for evaluating the utility of validity scales in detecting overreporting (e.g., Rogers, 2008), because they are part of the "normal" population, and thus assumed to be less pathological than a patient sample. The most appropriate benchmark to determine whether validity scales can differentiate between overreporters and patients with genuine psychopathology is to use a patient sample as a comparison group. Here, because the PPI–R is a measure of psychopathy, a forensic psychiatric sample was deemed most appropriate for comparative purposes. We used data from an archival forensic inpatient sample (see Edens & McDermott, 2010, for more detail) as a comparison group in examining the detection of overreporting on the PPI–R. This sample comprised 200 forensic psychiatric patients (166 after removing participants on the basis of the INC scale), who were primarily hospitalized after being found incompetent to stand trial, not guilty by reason of insanity, or classified as a mentally disordered offender. These individuals had volunteered to participate in a confidential research study, which had no bearing on their status as patients at the hospital, and had therefore no incentives to dissimulate. The sample was predominantly male (85.5%) and Caucasian (65%), with the largest minority being African American (20.5%), with the remaining belonging to other races/ethnicities (14.5%). The majority of this sample had a diagnosis of schizophrenia (67.5%), 60% had a co-occurring substance disorder, and 26.5% had a diagnosis of antisocial personality disorder.

Measures

Psychopathy Personality Inventory—Revised (PPI–R; Lilienfeld & Widows, 2005). The PPI–R is a 154-item self-report questionnaire designed to assess psychopathic personality traits.

As described earlier, it includes three validity scales: Virtuous Responding (VR), Deviant Responding (DR), and Inconsistent Responding. A total score, three factor scores, and eight content scale scores are also generated (see Table 1 for scale labels). It has been shown have acceptable construct validity with other measures of psychopathy (Lilienfeld, & Widows, 2005; Marcus, Fulton, & Edens, 2012; Poythress et al., 2010; Ray, Weir, Poythress, & Rickelm, 2011). Internal consistencies ranged from .79 (Coldheartedness) to .88 (Carefree Nonplanfulness) for the PPI–R content scales.

Boldness Inventory (Patrick, Vaidyanathan, Benning, Hicks, & Kramer, 2010). The Boldness Inventory is a 19-item self-report inventory of psychopathic traits within the context of the Triarchic conceptualization of psychopathy (Patrick, Fowles, & Krueger, 2009). Boldness represents the nexus of social dominance, thrill seeking/fearlessness, and low stress–reactivity (Patrick et al., 2009), and can be traced back to seminal descriptions of psychopathy such as those of Cleckley (1941/1976) and Lykken (1995). Internal consistency for the Boldness Inventory was .74.

Hare Self-Report Psychopathy Scale—II (SRP–II; Hare, 1991). The SRP–II is a 60-item self-report inventory designed to measure the same constructs assessed in the Psychopathy Checklist—Revised (PCL–R; Hare, 2003). The SRP–II has been found to be correlated with the PCL–R (Hare, 1991), along with several other measures of psychopathy, including the PPI (Benning, Patrick, Salekin, & Leistico, 2005; Lilienfeld & Andrews, 1996). Internal consistency was .87 in the current sample.

Levenson Self-Report Psychopathy Scale (LSRP; Levenson, Kiel, & Fitzpatrick, 1995). The LSRP is a 26-item self-report measure of psychopathy designed for use in nonincarcerated populations. Respondents rate themselves from 1 to 4 on each item, and a total score along with three factor scores (Egocentrism, Callous, Antisocial) are generated (Brinkley, Diamond, Magaletta, & Heigel, 2008; Sellbom, 2011). Convergent and discriminant validity of this measure has been shown in several studies (Brinkley, Schmitt, Smith, & Newman, 2001; Brinkley et al., 2008; Levenson et al., 1995; Poythress et al., 2010; Sellbom, 2011).

Table 1
Mean PPI–R Total, Factor, and Content Scale Score Differences Across Response Style Groups

Scale	Honest students ($n = 127$)	Underreporting ($n = 114$)	Overreporting ($n = 94$)	F	p	Cohen's d (95% CI) honest vs. underreporting	Cohen's d (95% CI) honest vs. overreporting
PPI total score	279.26$_a$ (39.32)	262.32$_a$ (37.68)	391.68$_b$ (64.45)	192.039	<.001	−0.44 (−0.69, −0.18)	−2.18 (−2.51, −1.84)
I. Fearless dominance	109.98$_a$ (18.92)	106.82$_a$ (16.20)	125.18$_b$ (23.65)	21.78	<.001	−0.18 (−0.43, 0.08)	−0.72 (−0.99, −0.44)
Social influence	47.20$_a$ (9.05)	45.36$_a$ (7.88)	49.64$_b$ (11.50)	7.09	<.001	−0.22 (−0.47, 0.04)	−0.24 (−0.51, 0.03)
Fearlessness	31.54$_a$ (9.89)	27.91$_b$ (8.43)	44.13$_b$ (10.71)	69.54	<.001	−0.39 (−0.65, −0.14)	−1.23 (−1.51, −0.93)
Stress immunity	31.23$_a$ (7.01)	33.55$_b$ (6.76)	30.98$_a$ (7.35)	9.64	<.001	−0.34 (−0.59, −0.08)	0.03 (−0.23, 0.30)
II. Self-centered impulsivity	139.09$_a$ (26.21)	128.19$_b$ (26.68)	215.67$_c$ (36.07)	208.20	<.001	0.41 (0.16, 0.67)	−2.49 (−2.83, −2.13)
Machiavellian egocentricity	40.73$_a$ (9.72)	35.20$_b$ (9.46)	62.67$_b$ (12.76)	156.00	<.001	0.13 (−0.13, 0.38)	−1.97 (−2.29, −1.64)
Rebellious nonconformity	31.46$_a$ (7.47)	28.52$_b$ (6.70)	48.23$_b$ (10.29)	115.71	<.001	0.09 (−0.17, 0.34)	−1.91 (−2.22, −1.58)
Blame externalization	31.69$_a$ (7.47)	33.79$_a$ (9.57)	47.12$_b$ (8.58)	62.93	<.001	−0.25 (−0.50, 0.001)	−1.94 (−2.25, −1.61)
Carefree nonplanfulness	35.21$_a$ (9.50)	30.68$_b$ (7.60)	57.65$_c$ (12.78)	187.88	<.001	0.52 (0.26, 0.78)	−2.04 (−2.36, −1.70)
III. Coldheartedness	30.19$_a$ (6.90)	27.30$_b$ (6.81)	50.83$_c$ (12.44)	164.94	<.001	0.42 (0.16, 0.68)	−2.14 (−2.46, −1.80)

Note. Mean scores are reported as raw scores. PPI–R = Psychopathic Personality Inventory—Revised; CI = confidence interval. Means with different subscripts are significantly different ($p < .05$).

Internal consistencies were .85 (Total), .84 (Egocentricity), .66 (Callous), and .72 (Antisocial) in the current sample.

Antisocial Process Screening Device—Youth Version (APSD; Frick & Hare, 2001). The APSD is a 20-item self-report scale on which respondents rate themselves from 0 to 2 on each item. This measure adheres to a three-factor model of psychopathy (Frick & Hare, 2001), with factors reflecting Callous–Unemotional Traits (7 items), Narcissism (5 items), and Impulsivity/Conduct Problems (6 items). The APSD has been shown to differentiate between groups of adolescents with differing levels of violence and severity of psychopathic traits (Kruh, Frick, & Clements, 2005). Internal consistencies were .85 (Total), .50 (Callous–Unemotional), .72 (Narcissism), and .58 (Impulsivity) in the current sample. The APSD subscales are very short, which affects Cronbach's alpha. Average interitem correlations were in the acceptable range for both Callous–Unemotional (.16) and Impulsivity (.22).

Inventory of Callous Unemotional Traits (ICU; Essau, Sasagawa, & Frick, 2006). The ICU is a 24-item self-report measure of callous and unemotional traits for adolescents. It was designed to improve upon the Callous–Unemotional scale found on the APSD by adding items, and changing the wording and rating system. Previous research on this measure supports its use in assessing for CU traits (Essau et al., 2006; Kimonis et al., 2008). The internal consistency for this measure was .70.

Procedures

College sample. Participants in the college sample were first administered a battery of tests under standard instructions (i.e., all measured described above except PPI–R). Next, participants were randomly assigned into one of three groups in terms of responding to the PPI–R: honest responding, underreporting, and overreporting. The investigator read a set of instructions to each group specifying how each participant should respond on the PPI–R. Participants in the underreporting group were given a scenario in which they were faced with criminal charges and were told to underreport their symptoms on the PPI–R in order to appear to be a better fit for a less secure prison. Participants in the overreporting group were also given a scenario in which they were charged for a criminal offense but were told to overreport their symptoms on the PPI–R and "fake mental illness" in order to evade prosecution. Specific instructions given to the undergraduate sample in each of the dissimulation groups can be found in the Appendix. Finally, the honest responding group was asked to continue to respond honestly to the PPI–R as they had for the previous measures administered.

Forensic psychiatric sample. Each patient in this sample was administered the PPI–R as a volunteer in a confidential research study. They were given this measure as part of a standard research battery, were given no special instructions, and none of the patients had an external incentive to response dishonestly. Prior to enrolling in the study, potential participants were administered the MacArthur Competence Assessment Tool for Clinical Research (MacCAT–CR; Appelbaum & Grisso, 2001) to ensure that they had the capacity to provide informed consent.

Results

Impact of Response Bias on Mean Elevations and Psychometric Validity

We first examined whether the mean elevations on PPI–R scale scores would be substantially different across the three response bias groups, because such differences would have substantial interpretative implications. A one-way analysis of variance (ANOVA) was conducted to determine group differences across total, factor, and content scale scores. These results are shown in Table 1 and Figure 1. For the underreporting group, mean scale elevations were significantly lower than the other groups for the total score, all three factor scores, and each of the content scales with the exception of Blame Externalization and Stress Immunity. This was not surprising given that underreporters are likely to report fewer problems in general and will therefore show greater immunity to stress. For the overreporting group, mean scale elevations were significantly higher (in relation to the other groups) for the total score, three factors scores, and each of the content scales with the only exception being Stress Immunity.

We next examined the impact of response bias on the psychometric validity of PPI–R total, factor, and content scale scores (see Table 2). For this purpose, we calculated correlations between PPI–R scores and scores on criterion psychopathy measures separately for each group (i.e., honest responding, overreporting, and underreporting). More specifically, PPI–R Total scores were correlated with other psychopathy total scores (e.g., APSD, SRP–II), whereas the PPI–R factor and content scale scores were correlated with conceptually relevant subscale scores derived from the other measures. Fischer's z test for independent correlations was used to determine whether the correlation magnitudes differed significantly across groups, and Cohen's q statistic was used to estimate the effect size associated with these differences (with qs of .10–.29, .30–.49, and .50+, indicating small, medium, and large effect sizes, respectively; see Cohen, 1992).

Overall, correlations between PPI–R scores and extratest psychopathy scale scores were significantly lower when participants were asked to either overreport or underreport their symptoms in relation to when the test was taken under standard instructions. The differences were most pronounced when correlations for the stan-
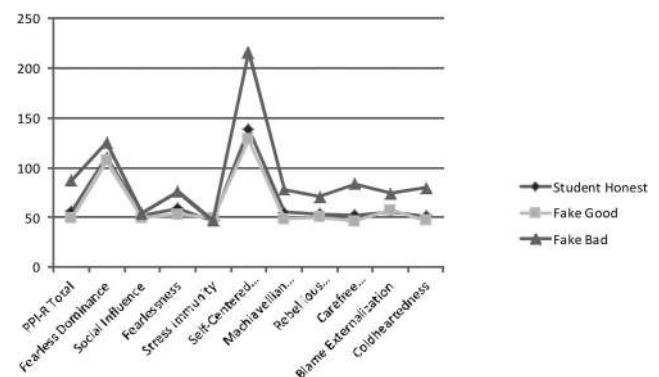

Figure 1. Mean Psychopathic Personality Inventory—Revised (PPI–R) total, factor, and content scale elevations across undergraduate student response style groups.

Table 2

Correlations Between PPI–R Total, Factor, and Subscale Scores and Extratest Criteria Across Response Style Groups

Scale	Honest	Underreporting	Z	Cohen's q (95% CI)	Overreporting	Z	Cohen's q (95% CI)
PPI Total							
SRP	.72	.42	3.69	.46 (.34, .60)	−.20	8.85	1.11 (1.00, 1.26)
LSRP	.68	.41	3.10	.39 (.26, .52)	−.16	7.90	.99 (.87, 1.13)
APSD	.62	.32	3.13	.39 (.26, .52)	−.13	6.84	.86 (.74, 1.00)
PPI Fearless Dominance							
BI	−.66	−.33	3.58	.45 (.33, .58)	.15	7.54	.94 (.82, 1.09)
LSRP–1	.18	.12	0.54	.07 (−.06, .19)	−.03	1.71	.21 (.08, .35)
PPI Self-Centered Impulsivity							
LSRP–3	.65	.38	2.96	.37 (.25, .51)	−.14	7.33	.92 (.79, 1.07)
APSD–N	.60	.33	2.82	.35 (.23, .49)	−.03	5.79	.72 (.60, .87)
APSD–I/C	.62	.39	2.55	.32 (.20, .45)	−.12	6.78	.85 (.73, 1.00)
PPI Machiavellian Egocentricity							
LSRP–1	.69	.37	3.63	.45 (.33, .59)	−.07	7.32	.86 (0.74, 1.00)
ICU	.50	.23	2.51	.31 (.19, .45)	−.13	5.41	.68 (.50, .83)
APSD–N	.65	.33	3.41	.43 (.31, .56)	.01	6.07	.76 (.63, .91)
PPI Social Influence							
BI	−.48	−.20	2.57	.32 (.20, .46)	.14	5.31	.66 (.54, .81)
PPI Coldheartedness							
LSRP–2	.42	.41	0.18	.02 (−.10, .15)	−.11	4.47	.56 (.44, .71)
APSD–C	.29	.18	0.97	.12 (.00, .26)	−.24	4.39	.55 (.42, .69)
ICU	.47	.22	2.31	.29 (.16, .42)	−.23	5.95	.74 (.62, .89)
PPI Fearlessness							
BI	−.54	−.29	2.46	.31 (.19, .44)	.17	6.23	.78 (.66, .93)
PPI Rebellious Noncomformity							
LSRP–3	.49	.25	2.22	.28 (.15, .41)	.00	4.31	.54 (.41, .68)
APSD–I/C	.51	.39	1.25	.16 (.03, .29)	−.10	5.37	.57 (.45, .71)
PPI Carefree Nonplanfulness							
LSRP–3	.52	.41	1.14	.14 (.02, .28)	−.22	6.38	.80 (.68, .95)
APSD–I/C	.45	.24	1.87	.23 (.11, .37)	−.13	4.89	.61 (.49, .76)
PPI Blame Externalization							
LSRP–3	.37	.19	1.53	.19 (.07, .32)	−.15	4.25	.53 (.41, .68)
APSD–I/C	.40	.32	0.76	.10 (−.03, .22)	−.04	3.66	.46 (.33, .60)
PPI Stress Immunity							
BI	−.39	−.19	1.72	.22 (.09, .34)	.03	3.49	.44 (.31, .58)

Note. $z > 1.96$ is significant at $p < .05$. PPI–R = Psychopathic Personality Inventory—Revised; CI = confidence interval; SRP = Self-Report Psychopathy Scale—II; APSD = Antisocial Processes Screening Device; APSD–I/C = Impulsivity/Conduct Problems; APSD–N = Narcissism; APSD–C = Callous–Unemotional; LSRP = Levenson Self-Report Psychopathy Scale; LSRP–1 = Egocentricity; LSRP–2 = Callous; LSRP–3 = Antisocial; BI = Boldness Inventory; ICU = Inventory of Callous Unemotional Traits.

dard instruction versus overreporting groups were compared. In these comparisons, numerous correlations went from being moderate to high for the honest condition to being negatively correlated in the overreporting condition. On the other hand, for the honest versus underreporting groups, the correlations with relevant criteria for some of the PPI–R subscales (Carefree Nonplanfulness, Blame Externalization, and Stress Immunity) were not significantly different. For most significant correlation comparisons, effect sizes were moderate to large. For instance, when differences between correlations of the PPI–R total score and SRP, LSRP, and APSD total scores were examined, Cohen's q statistics for the underreporting group ranged from 0.39 to 0.46 and qs for the overreporting group ranged from 0.86 to 1.11. Similar results were found for the correlations between the factor and subscale scores and the relevant total and factor scores of other measures for both the underreporting group (qs = 0.02-−0.45) and the overreporting group (qs = 0.21−0.99). Thus, these results portray a very clear pattern: the criterion-related validity of PPI–R substantive scale scores is substantially attenuated when individuals are either over- or underreporting.

Group Differences on Validity Scales

The next step involved examining the validity of the PPI–R validity scales in differentiating underreporting and overreporting groups from the honest responding group. For these analyses, we also included the forensic psychiatric sample as a comparison for the overreporting group. More specifically, one-way ANOVAs were used to determine if scores on the DR and VR validity scales differed across groups. We used Tukey's HSD post hoc test for bivariate comparisons and Cohen's d to characterize effect size magnitudes. These results are shown in Table 3. As expected, the overreporting group scored significantly higher on the DR scale than the forensic patient, undergraduate honest, underreporting groups, $F(3, 584) = 287.96, p < .001$, with large effect sizes when compared with both the undergraduate honest group and the forensic patient group. On the other hand, the underreporting group scored significantly higher on the VR scale than all other groups, $F(3, 584) = 49.10, p < .001$, with a large effect size when comparing the underreporting group with the undergraduate honest responding group.

Table 3
Mean PPI–R Validity Scale Score Differences Across Response Style Groups

Scale	Honest students	Honest patients	Under-reporting	Over-reporting	F	p	Cohen's d (95% CI) honest vs. under-reporting	Cohen's d (95% CI) honest vs. over-reporting	Cohen's d (95% CI) patient vs. over-reporting
DR	14.596_a (4.68)	13.625_a (3.48)	13.702_a (3.93)	27.779_b (6.52)	287.963	<.001	0.21 (−0.05, 0.46)	−2.38 (−2.72, −2.03)	−2.95 (−3.30, −2.58)
VR	29.114_a (5.13)	29.810_a (5.67)	34.182_b (7.20)	24.434_c (7.40)	49.097	<.001	−0.82 (−1.08, −0.55)	0.76 (0.48, 1.03)	0.85 (0.58, 1.11)

Note. DR = Deviant Responding; VR = Virtuous Responding; CI = Confidence interval; Honest = Honest Students; Patient = Honest Patients. Means with different subscripts are significantly different ($p < .05$).

We did not find significant differences between the undergraduate honest and forensic patient groups on the DR scale. However, this was expected given that the PPI–R manual states that the items in the DR scale were designed explicitly to measure symptoms of superficially plausible, but nonexistent psychological disorders. It is therefore a positive finding that patients with bona fide psychopathology do not endorse a greater number of items on this scale than do undergraduate students who take the PPI–R honestly, and speaks well of the scale's content validity and clinical utility.

Classification Accuracy

Finally, we examined the classification accuracy of both of these scales in differentiating between honest groups and groups that were asked to over- or underreport their symptoms on the basis of rates of sensitivity and specificity (see Table 4). Sensitivity refers to the percentage of individuals who were correctly identified as showing response bias, whereas specificity refers to the percentage of individuals who were correctly identified as answering truthfully. Additionally, positive and negative predictive powers (PPP and NPP) indicate the probability that an individual is (or is not) engaging in response bias given a certain cut-off value. Because predictive power is heavily influenced by base rates, and estimated base rates of response bias vary from setting to setting, we examined the PPP and NPP statistics using a series of hypothetical base rates. A series of cut scores for each scale was also examined to determine at what score the least amount of false positive predictions would be made, while also being able to effectively identify those attempting to response dishonestly. Bagby, Rogers, Buis, & Kalemba (1994) recommended that PPP rates should be .80 or higher, though Rogers (2008) argued that in high-stakes evaluations, PPP rates of .90 or higher may be necessary.

We first examined the classification accuracy of the DR scale in identifying overreporting when compared with the forensic patient group. On the basis of these results, we found that the DR scale had an optimal cut score of 23. At this cutoff, acceptable rates of sensitivity (.80) and specificity (.99) were determined with a 97% hit rate. Across base rates ranging from .10 to .50, PPP at this cut score ranged from .90 to .99, whereas NPP ranged from .83 to .98. When using the undergraduate honest responding group as a comparison, an optimal cut score of 25 was identified. At this cut score, rates of sensitivity (.72) and specificity (.96) were somewhat lower than when patients were used, with a hit rate of 93%. At this score, PPP ranged from .65 to .94 and NPP ranged from .77 to .97. Finally, we examined the VR scale in identifying underreporting when compared to the undergraduate group who took the PPI–R under standard instructions. In this case, an optimal cut off score of 38 was determined, which rendered high specificity (.98) but low sensitivity (.32) with a hit rate of 91%. PPP at this cut score ranged from .63 to .94 and NPP ranged from .59 to .93. Of course, lower cut scores would result in improved identification of underreporters, but at the expense of a higher rate of false positive prediction errors.

Discussion

The current investigation aimed to examine the need for and utility of the PPI–R validity scales. In general, our hypotheses were supported. When individuals were asked to respond dishon-

Table 4

Classification Accuracy Statistics for Overreporting, Underreporting, and Honest Group Comparisons

Subscale	Cutoff	SENS	SPEC	HR	Predictive power: positive/negative base rates				
					.10	.20	.30	.40	.50
DR (patient comparison)	20	.90	.92	.92	.55/.99	.73/.97	.82/.96	.88/.93	.92/.90
	21	.89	.94	.93	.61/.99	.78/.97	.86/.95	.90/.93	.93/.90
	22	.85	.96	.95	.68/.98	.83/.96	.89/.94	.93/.91	.95/.87
	23	.80	.99	.97	.90/.98	.95/.95	.97/.92	.98/.88	.99/.83
	24	.75	1.00	.97	.94/.97	.97/.94	.99/.90	.99/.86	.99/.80
	25	.72	1.00	.97	.94/.97	.97/.93	.98/.89	.99/.84	.99/.78
DR (student comparison)	20	.90	.82	.82	.35/.99	.55/.97	.68/.95	.77/.93	.83/.89
	21	.89	.83	.84	.37/.99	.57/.97	.69/.95	.78/.92	.84/.89
	22	.85	.87	.87	.43/.98	.63/.96	.74/.93	.82/.90	.87/.86
	23	.80	.91	.90	.51/.98	.70/.95	.80/.91	.86/.87	.90/.82
	24	.75	.94	.92	.60/.97	.77/.94	.85/.90	.90/.85	.93/.79
	25	.72	.96	.93	.65/.97	.81/.93	.88/.89	.92/.84	.94/.77
VR	35	.48	.81	.64	.22/.93	.38/.86	.52/.78	.62/.70	.71/.61
	36	.45	.89	.84	.30/.94	.49/.86	.63/.79	.72/.71	.80/.62
	37	.36	.93	.87	.36/.93	.56/.85	.69/.77	.77/.69	.84/.59
	38	.32	.98	.91	.63/.93	.79/.85	.87/.77	.91/.68	.94/.59
	39	.27	.98	.91	.59/.92	.76/.84	.85/.76	.89/.67	.93/.57
	40	.23	.99	.92	.78/.92	.89/.84	.93/.75	.96/.66	.97/.56

Note. DR = Deviant Responding; VR = Virtuous Responding; Deviant Responding was compared to both the patient group and the undergraduate honest groups; Virtuous Responding was compared to the undergraduate honest group. SENS = sensitivity; SPEC = specificity; and HR = hit rate.

estly, the resulting protocols were significantly affected. Scores on the PPI–R total, factor, and content scale scores in the overreporting group were significantly higher, and scores on these scales in the underreporting group were significantly lower, relative to those of the honest responding group. Moreover, the psychometric validity of PPI–R scale scores (and especially total and factor scores) were substantially attenuated (and in some cases for the overreporting group, opposite of theoretical expectations) when participants were asked to engage in dissimulated responding. Thus, if the validity scales are not used to rule out potential over- and underreporting, conclusions drawn from PPI–R protocols could be highly inaccurate. Such responding can have significant implications for diagnosis and treatment of psychopathy and other mental health conditions. Furthermore, in terms of the utility of the PPI–R validity scales in distinguishing between groups of honest, overreporting, and underreporting participants, we found that the scales performed well at distinguishing between groups with high rates of sensitivity and specificity.

Considering our findings in conjunction, clinicians need to be careful in attributing any meaning to PPI–R total, factor, and content scale scores when individuals are engaged in suspected dissimulated responding. Fortunately, the current findings also indicate that the PPI–R validity scales are useful in identifying over- and underreporting individuals at an acceptable error rate. Clinicians are therefore encouraged to use the PPI–R validity scales in detecting over- and underreporting, and when cross-validated, using the cut scores presented here.

In a broader context, the results of the current investigation are generally inconsistent with arguments made against the use of validity scales (Costa & McCrae, 1992; McGrath et al., 2010). These results add to the growing literature showing that validity scales are indeed needed, given that the substantive scale profiles and criterion-related validity are significantly affected by response bias. Furthermore, this study shows the effects of response bias on

protocol validity and that self-report measures can indeed be easily feigned. Thus, the advantage of self-report measures such as the PPI–R is that there are built-in scales used to detect these response styles. Therefore, it behooves any forensic examiner to select measures that have validity scales that evidence good utility in detecting such response styles (e.g., MMPI–2–RF, PAI). Specifically with regard to psychopathy, this is particularly important given that the PCL–R can also be feigned without having any response bias indicators (Rogers et al., 2002).

This research also further adds to the already extensive literature that validity scales are useful in detecting dissimulated responding. Nevertheless, one of the general arguments by McGrath and colleagues that the validity scales are not useful in increasing criterion-related validity (and thus acting as suppressor variables) is likely true, as evidenced by their results as well as the lack of utility for the MMPI K-correction (see, e.g., Barthlow, Graham, Ben-Porath, & McNulty, 2004), but this does not negate the need for or utility of these scales. As Ben-Porath and Waller (1992) argued, the goal of validity scales is to assess the utility of the self-report profile that is produced, not to increase criterion-related validity of substantive scales. Our results support the utility of such scales in achieving this goal.

There are some limitations associated with this study that must be acknowledged. One of these concerns the samples that were used. The use of a predominately female undergraduate sample is a limitation because such samples lack external validity and the psychopathological personality trait variability seen in general populations. Thus, the results may be difficult to generalize in male offender populations in which psychopathic traits are present at higher mean levels. Therefore, future research should replicate and extend this research to samples of individuals with legitimate incentives to overreport or underreport their symptoms, including forensic and correctional settings. In addition, it should be noted that although the forensic group did not have an incentive to underreport their symptoms, it is possible that some of those

individuals may have engaged in some minimal underreporting in order to appear more prepared to be discharged from the hospital, which could have caused slightly inflated effect sizes when the overreporting group was compared with this sample. Future research should use other forensic patient groups as well to replicate these findings.

In conclusion, the results of this study showed support for the need for and utility of the PPI–R VR and DR validity scales, and builds on the limited research base on these scales. Mean elevations and psychometric validity of PPI–R scale scores are grossly affected by response bias, consistent with literature on other self-report inventories, and validity scales are thus clearly needed to detect dissimulation. The PPI–R validity scales evidenced promising utility in differentiating between dissimulators and genuine test-takers.

References

Appelbaum, P. S., & Grisso, T. (2001). *MacArthur Competence Assessment Tool for Clinical Research (MacCAT–CR)*. Sarasota, FL: Professional Resource Press.

Barthlow, D. L., Graham, J. R., Ben-Porath, Y. S., & McNulty, J. L. (2004). Construct validity of the MMPI-2 College Maladjustment (Mt) Scale. *Assessment, 11,* 251–262. doi:10.1177/1073191104268317

Benning, S., Patrick, C. J., Salekin, R. T., & Leistico, A. R. (2005). Convergent and discriminant vailidty of psychopathy factors assessed via self-report: A comparison of three instruments. *Assessment, 12,* 270–289. doi:10.1177/1073191105277110

Ben-Porath, Y. S., & Waller, G. (1992). Five big issues in clinical personality assessment: A rejoinder to Costa and McCrae. *Psychological Assessment, 4,* 23–25. doi:10.1037/1040-3590.4.1.23

Brinkley, C. A., Diamond, P. M., Magaletta, P. R., & Heigel, C. P. (2008). Cross-validation of Levenson's Psychopathy Scale in a sample of federal female inmates. *Assessment, 15,* 464–482. doi:10.1177/1073191108319043

Brinkley, C., Schmitt, W., Smith, S., & Newman, J. (2001). Construct validation of a self-report psychopathy scale: Does Levenson's Self-Report Psychopathy Scale measure the same constructs as Hare's Psychopathy Checklist—Revised? *Personality and Individual Differences, 31,* 1021–1038. doi:10.1016/S0191-8869(00)00178-1

Burchett, D. L., & Ben-Porath, Y. S. (2010). The impact of overreporting on MMPI–2–RF substantive scale score validity. *Assessment, 17,* 497–516. doi:10.1177/1073191110392553

Caldwell-Andrews, A. A. (2001, January). Relationships between MMPI-2 validity scales and NEO PI-R experimental validity scales in police candidates. *Dissertation Abstracts International,* 61.

Cleckley, H. H. (1976). *The mask of sanity: An attempt to reinterpret the so-called psychopathic personality*. Oxford, England: Mosby. (Original work published 1941)

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112,* 155–159. doi:10.1037/0033-2909.112.1.155

Costa, P. T., & McCrae, R. R. (1992). "Normal" personality inventories in clinical assessment: General requirements and the potential for using the NEO Personality Inventory: Reply. *Psychological Assessment, 4,* 20–22. doi:10.1037/1040-3590.4.1.20

DeMatteo, D., & Edens, J. F. (2006). The role and relevance of the Psychopathy Checklist—Revised in court: A case law survey of U.S. courts (1991–2004). *Psychology, Public Policy, and Law, 12,* 214–241. doi:10.1037/1076-8971.12.2.214

Edens, J. F. (2004). Effect of response distortion on the assessment of divergent facets of psychopathy. *Assessment, 11,* 109–112. doi:10.1177/1073191103261474

Edens, J. F., Buffington, J. K., & Tomicic, T. L. (2000). An investigation of the relationship between psychopathic traits and malingering on the Psychopathic Personality Inventory. *Assessment, 7,* 281–296. doi:10.1177/10731910000700307

Edens, J. F., Buffington, J. K., Tomicic, T. L., & Riley, B. D. (2001). Effects of positive impression management on the Psychopathic Personality Inventory. *Law and Human Behavior, 25,* 235–256. doi:10.1023/A:1010793810896

Edens, J. F., & McDermott, B. E. (2010). Examining the construct validity of the Psychopathic Personality Inventory—Revised: Preferential correlates of fearless dominance and self-centered impulsivity. *Psychological Assessment, 22,* 32–42. doi:10.1037/a0018220

Essau, C. A., Sasagawa, S., & Frick, P. J. (2006). Callous-unemotional traits in a community sample of adolescents. *Assessment, 13,* 454–469. doi:10.1177/1073191106287354

Frick, P. J., & Hare, R. D. (2001). *The Antisocial Process Screening Device*. Toronto, Ontario, Canada: Multi-Health Systems.

Hare, R. D. (1991). *The Self-Report Psychopathy Scale—II. Unpublished test*. Vancouver, British Columbia, Canada: University of British Columbia.

Hare, R. D. (2003). *Manual for the Revised Psychopathy Checklist* (2nd ed.). Toronto, ON, Canada: Multi-Health Systems.

Hawes, S. W., & Boccaccini, M. T. (2009). Detection of overreporting of psychopathology on the Personality Assessment Inventory: A meta-analytic review. *Psychological Assessment, 21,* 112–124. doi:10.1037/a0015036

Kimonis, E. R., Frick, P. J., Skeem, J. L., Marsee, M. A., Cruise, K., Munoz, L. C., ... Morris, A. S. (2008). Assessing callous-unemotional traits in adolescent offenders: Validation of the inventory of callous-unemotional traits. *International Journal of Law and Psychiatry, 31,* 241–252. doi:10.1016/j.ijlp.2008.04.002

Kruh, I. P., Frick, P. J., & Clements, C. B. (2005). Historical and personality correlates to the violence patterns of juveniles tried as adults. *Criminal Justice and Behavior, 32,* 69–96. doi:10.1177/0093854804270629

Kucharski, L., Falkenbach, D. M., Egan, S. S., & Duncan, S. (2006). Antisocial personality disorder and the malingering of psychiatric disorder: A study of criminal defendants. *International Journal of Forensic Mental Health, 5,* 195–204. doi:10.1080/14999013.2006.10471243

Levenson, M. R., Kiel, K. A., & Fitzpatrick, C. M. (1995). Assessing psychopathic attributes in a noninstituionalized population. *Journal of Personality and Social Psychology, 68,* 151–158. doi:10.1037/0022-3514.68.1.151

Lilienfeld, S. O., & Andrews, B. P. (1996). Development and preliminary validation of a self-report measure of psychopathic personality traits in noncriminal populations. *Journal of Personality Assessment, 66,* 488–524. doi:10.1207/s15327752jpa6603_3

Lilienfeld, S. O., & Widows, M. R. (2005). *Psychopathic Personality Inventory—Revised: Professional manual*. Lutz, FL: Psychological Assessment Resources.

Lykken, D. T. (1995). *The antisocial personalities*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Marcus, D. K., Fulton, J. J., & Edens, J. F. (2012). The two-factor model of psychopathic personality: Evidence from the Psychopathic Personality Inventory. *Personality Disorders: Theory, Research, and Treatment, 3,* 140–154. doi:10.1037/a0025282

Marion, B. E., Sellbom, M., Salekin, R. T., Toomey, J. A., Kucharski, L. T., & Duncan, S. (2012). An examination of the association between psychopathy and dissimulation in forensic psychological evaluations. *Law and Human Behavior.* Advance online publication.10.1037/lhb0000008

McGrath, R. E., Mitchell, M., Kim, B. H., & Hough, L. (2010). Evidence for response bias as a source of error variance in applied assessment. *Psychological Bulletin, 136,* 450–470. doi:10.1037/a0019216

Nikolova, N., Hedry, M. C., Douglas, K. S., Edens, J. F., & Lilienfeld, S. O. (2012). The inconsistency of inconsistency scales: A comparison of two widely used measures. *Behavioral Sciences and the Law, 30,* 16–27. doi:10.1002/bsl/1996

Patrick, C. J. (2010). Conceptualizing the psychopathic personality: Dis-inhibited, bold, . . . Or just plain mean? In R. T. Salekin & D. R. Lynam (Eds.), *Handbook of child and adolescent psychopathy* (pp. 15–48). New York, NY: Guilford Press.

Patrick, C. J., Vaidyanathan, U., Benning, S. D., Hicks, B. M., & Kramer, M. D. (2010). *Refining measurement of the fearless-dominance component of psychopathy: Development and validation of the Boldness Inventory. Unpublished manual.*

Poythress, N. G., Edens, J. F., & Watkins, M. M. (2001). The relationship between psychopathic personality features and malingering symptoms of major mental illness. *Law and Human Behavior, 25,* 567–582. doi:10.1023/A:1012702223004

Poythress, N. G., Lilienfeld, S. O., Skeem, J. L., Douglas, K. S., Edens, J. F., Epstein, M., & Patrick, C. J. (2010). Using the PCL–R to help estimate the validity of two self-report measures of psychopathy with offenders. *Assessment, 17,* 206–219. doi:10.1177/1073191109351715

Ray, J. V., Weir, J. W., Poythress, N. G., & Rickelm, A. (2011). Correspondence between the Psychopathic Personality Inventory and the Psychopathic Personality Inventory—Revised: A look at self-reported personality traits. *Criminal Justice and Behavior, 38,* 375–385. doi:10.1177/0093854811398178

Rogers, R. (Ed.). (2008). *Clinical assessment of malingering and deception* (3rd ed.). New York, NY: Guilford Press.

Rogers, R., Sewell, K. W., Martin, M. A., & Vitacco, M. J. (2003). Detection of feigned mental disorders: A meta-analysis of the MMPI–2 and malingering. *Assessment, 10,* 160–177. doi:10.1177/1073191103010002007

Rogers, R., Vitacco, M. J., Jackson, R. L., Martin, M., Collins, M., & Sewell, K. W. (2002). Faking psychopathy? An examinations of re-sponse styles with antisocial youth. *Journal of Personality Assessment, 78,* 31–46. doi:10.1207/S15327752JPA7801_03

Sellbom, M. (2011). Elaborating on the construct validity of the Levenson Self-Report Psychopathy Scale in incarcerated and non-incarcerated samples. *Law and Human Behavior, 35,* 440–451. doi:10.1007/s10979-010-9249-x

Sellbom, M., & Bagby, R. M. (2008a). Response styles on multi-scale inventories. In R. Rogers (Ed.), *Clinical assessment of malingering and deception* (3rd ed.). New York: Guilford Press.

Sellbom, M., & Bagby, R. (2008b). The validity and utility of the Positive Presentation Management and Negative Presentation Management Scales for the revised NEO Personality Inventory. *Assessment, 15,* 165–176. doi:10.1177/1073191107310301

Sellbom, M., & Bagby, R. (2010). Detection of overreported psycho-pathology with the MMPI–2 RF form validity scales. *Psychological Assessment: A Journal of Consulting and Clinical Psychology, 22,* 757–767. doi:10.1037/a0020825

Skeem, J. L., Polaschek, D. L., Patrick, C. J., & Lilienfeld, S. O. (2011). Psychopathic personality: Bridging the gap between scientific evidence and public policy. *Psychological Science in the Public Interest, 12,* 95–162. doi:10.1177/1529100611426706

Wu, J., & Lebreton, J. M. (2011). Reconsidering the dispositional basis of counterproductive work behavior: The role of aberrant personality. *Personnel Psychology, 64,* 593–626.

Wygant, D. B., Anderson, J. L., Sellbom, M., Rapier, J. L., Allgeier, L. M., & Granacher, R. P. (2011). Association of the MMPI–2 restructured form (MMPI–2–RF) validity scales with structured malingering criteria. *Psychological Injury and Law, 4,* 13–23. doi:10.1007/s12207-011-9098-z

Appendix

Response Bias Groups PPI-R Instructions

Instructions: "Feign Good"

Please try to fill out the following questionnaire in order to give a good impression of yourself. Please pretend that you have been convicted of a crime, and are being evaluated by a psychologist. You want to make sure that you present yourself in a positive light, so that the psychologist does not classify you as "difficult," or "dangerous." Your answers could decide whether you are placed in a maximum or minimum security facility. However, keep in mind that this questionnaire has scales designed to detect dishonest individuals. Thus, try to be as believable as you can in presenting yourself in a positive manner, without problems, so that you are not detected by the psychologist.

Instructions: "Feign Bad"

Please try to fill out the following questionnaire in order to give a negative impression of yourself. Please pretend that you have committed a crime, and are being evaluated by a psychol-ogist before you go to trial. You decide to fake mental illness, hoping that it will get you a lighter sentence (or maybe even get you off the hook completely). However, keep in mind that this questionnaire has scales designed to detect dishonest individuals. Thus, try to be as believable as you can in presenting yourself negatively, so that you are not detected by the psychologist.