JOURNAL OF PERSONALITY ASSESSMENT, *88*(3), 276–283

ARTICLES

The Utility of the NEO–PI–R Validity Scales to Detect Response Distortion: A Comparison With the MMPI–2

Benjamin J. Morasco

Behavioral Health and Clinical Neurosciences Division
Portland VA Medical Center
Portland, Oregon
Department of Psychiatry
Oregon Health & Science University

Jeffrey D. Gfeller

Departments of Psychology and Psychiatry
Saint Louis University

Katherine A. Elder

Departments of Psychiatry and Surgery
Oregon Health & Science University

In this psychometric study, we compared the recently developed Validity Scales from the Revised NEO Personality Inventory (NEO PI–R; Costa & McCrae, 1992b) with the MMPI–2 (Butcher, Dahstrom, Graham, Tellegen, & Kaemmer, 1989) Validity Scales. We collected data from clients ($n = 74$) who completed comprehensive psychological evaluations at a university-based outpatient mental health clinic. Correlations between the Validity Scales of the NEO–PI-R and MMPI–2 were significant and in the expected directions. The relationships provide support for convergent and discriminant validity of the NEO–PI–R Validity Scales. The percent agreement of invalid responding on the two measures was high, although the diagnostic agreement was modest ($\kappa = .22$–$.33$). Finally, clients who responded in an invalid manner on the NEO–PI–R Validity Scales produced significantly different clinical profiles on the NEO–PI–R and MMPI–2 than clients with valid protocols. These results provide additional support for the clinical utility of the NEO–PI–R Validity Scales as indicators of response bias.

The Revised NEO Personality Inventory (NEO–PI–R; Costa & McCrae, 1992b) is a commonly used measure designed to assess the Five-factor model of personality (McCrae & Costa, 1997). Although the NEO–PI–R was not developed to assess disordered personality functioning or psychopathology, the measure has been applied in this fashion (Costa & McCrae, 1992a; Quirk, Christiansen, Wagner, & McNulty, 2003; Reynolds & Clark, 2001). For example Reynolds and Clark (2001) determined there was substantial overlap between the NEO–PI–R Facet scales and the scales of the Schedule for Nonadaptive and Adaptive Personality (SNAP; Clark, 1993), which is a self-report instrument that assesses traits relevant to personality disorders. Additionally, Quirk et al. (2003) reported that NEO–PI–R scale scores provided incremental validity beyond selected Minnesota Multiphasic Personality Inventory–2 (MMPI–2; Butcher, Dahstrom, Graham, Tellegen, & Kaemmer, 1989) scales in the prediction of Axis I and Axis II diagnoses, which were determined by structured interviews.

As the NEO–PI–R is increasingly utilized in clinical contexts, concerns have been raised regarding the lack of validity scales to detect response bias, distortion, or random responding (Ben-Porath & Waller, 1992a, 1992b). This concern has received support, as subsequent studies have found that the

NEO–PI–R is susceptible to faking (Ballenger, Caldwell-Andrews, & Baer, 2001; Caldwell-Andrews, Baer, & Berry, 2000; Drury, 2001; Rosse, Stecher, Miller, & Levin, 1998). However, the authors (Costa & McCrae, 1992b) of the NEO–PI–R purposely omitted validity scales from the measure because of the belief that patient self-report is generally trustworthy. Costa and McCrae (1992b) argued that validity scales can be counterproductive, attempts to improve cooperation are more likely to improve test validity than attempts to correct protocol invalidity, and clinicians should consider test results in the context of a comprehensive assessment and not rely on single measures or indicators of response bias (Costa & McCrae, 1992a, 1992c, 1997). Others have suggested that validity scales, in general, may not be effective at detecting the test-taking attitudes they intend to measure (Borkenau & Ostendorf, 1992; McCrae & Costa, 1997; Nicholson & Hogan, 1990).

Although this issue remains unsettled, validity scales for the NEO–PI–R were developed (Schinka, Kinder, & Kremer, 1997). These scales were intended to detect random responding (Inconsistency Scale; INC), efforts to present in an overly positive fashion (Positive Presentation Management; PPM), and efforts to present in an overly negative fashion (Negative Presentation Management; NPM). Rather than adding additional items to the NEO–PI–R, Schinka et al. (1997) identified items within the measure to assess response validity. Several studies utilizing these NEO–PI–R Validity Scales revealed that they have been effective at identifying participants instructed to present positively (Ballenger et al., 2001; Caldwell-Andrews et al., 2000) or negatively (Berry et al., 2001; Caldwell-Andrews et al., 2000). In contrast, Piedmont, McCrae, Riemann, and Angleitner (2000) did not find support for the use of the NEO–PI–R Validity Scales, as they did not mediate validity scores or content scores on another psychological measure.

Although providing mixed results, the previously mentioned studies were limited by their analog design and/or use of nontreatment-seeking participants. To account for these limitations, Young and Schinka (2001) examined the utility of the NEO–PI–R Validity Scales to detect NPM in a clinical sample of 118 male veterans seeking treatment for alcohol dependence. In this study, the NEO–PI–R Validity Scales had satisfactory internal consistency (expressed as coefficient alphas of .70 for the PPM and .75 for the NPM; internal consistencies greater than .60 can be considered adequate; Nunnally & Bernstein, 1994; Streiner & Norman, 2003) and were significantly correlated with Personality Assessment Inventory (PAI; Morey, 1991) validity scales. Furthermore, individuals who attempted to present in a negative fashion on the NEO–PI–R had significantly different PAI profiles than individuals with valid NEO–PI–R profiles (Young & Schinka, 2001). This study provided additional support for the use of the NEO–PI–R Validity Scales to detect attempts to present oneself in a negative fashion.

A subsequent study (Morey et al., 2002) with 668 participants diagnosed with personality disorders or major depression utilized a multimethod-multitrait approach to evaluate the effectiveness of the NEO–PI–R Validity Scales. Results indicated that the NEO–PI–R Validity Scales corresponded with indicators of global functioning and response validity. Specifically, NEO–PI-R validity indicators correlated with observer ratings of the Global Assessment of Functioning, total number of Axis I and Axis II diagnoses, and validity indexes from the SNAP (Clark, 1993). In addition, four confirmatory factor analyses were conducted. The best fitting model was one in which NEO–PI–R Validity Scales were considered stylistic variables that were distinct from but correlated with substantive trait variables. This indicates that the scales may actually measure a presentational style that might relate to aspects of mental health. Moreover, based on convergent data with measures of psychopathology as well as symptom presentation, Morey et al. cautioned against interpreting NEO–PI–R positive or NPM indicators solely as reflecting purposeful response distortion.

To expand on prior research, in this study, we examined the utility of the NEO–PI–R Validity Scales to detect response distortion among clients seeking comprehensive psychological evaluations at a university-based psychological services center. Clients' scores on the NEO–PI–R Validity Scales were compared with the Validity Scales from the MMPI–2 (Butcher et al., 1989). This provides an important comparison, as the MMPI–2 Validity Scales have been shown to reliably detect random responding, malingering, and positive impression management (PIM; Bagby et al., 1997; Berry, Baer, & Harris, 1991; Gallen & Berry, 1996; Graham, Watts, & Timbrook, 1991; Rogers, Sewell, & Salekin, 1994). We conducted this study in a clinical setting with diverse clients and referral questions. The most frequent reason for assessment was to identify psychological factors that may be impacting educational functioning to determine whether clients merited academic accommodations. Other clients were referred for neuropsychological evaluations, psychiatric differential diagnoses, or psychotherapy. Thus, we drew clients from a clinical setting that included individuals who may have presented themselves in a negative manner as well as clients who may have attempted to present positively. Because these factors can negatively affect test validity, the sample provided an excellent opportunity to study characteristics of the NEO–PI–R Validity Scales.

METHOD

We drew data for this study from the records of clients completing comprehensive psychological evaluations at a university-based psychological services center. The clinic is located in an urban area and provides services to members of the university as well as the community at large. All clients completed evaluations between 1999 and 2002. Participants were eligible for inclusion in this study if they were at least 18 years old at the time of the evaluation and spoke English as a first language. A total of 74 clients completed the NEO–PI–R (Costa & McCrae, 1992b) and the MMPI–2 (Butcher et al.,

1989) in the course of their evaluations and met the preceding inclusion criteria. We did not exclude anyone from this study based on other demographic, psychological, or medical variables. Clinical psychology graduate students, under the supervision of licensed psychologists, conducted all evaluations.

PARTICIPANTS

The mean age of the sample ($n = 74$) was 32.7 years ($SD = 11.5$, range = 18–65). A majority of clients were female ($n = 40$, 54.1%) and single ($n = 45$, 60.8%). Regarding ethnicity, 60 (81.1%) clients indicated that they were White, 10 (13.5%) African American, and the remaining 4 (5.5%) reported another ethnicity. A total of 33 (44.6%) clients were full-time students, 21 (28.4%) were working full-time, 3 (4.1%) were working part-time, 10 (13.5%) were unemployed, and 6 (8.2%) had another employment status. A total of 45 (60.8%) clients were referred for psychological evaluations, and the remaining 29 (39.2%) were referred for therapy. Of the clients referred for psychological evaluations, 38 of the 45 (84.4%) were evaluated for a possible learning disorder, attention deficit hyperactivity disorder, or for some other factor possibly affecting educational functioning. The remaining clients who were referred for a psychological evaluation completed a comprehensive neuropsychological evaluation ($n = 3$, 4.1%), personality evaluation ($n = 3$, 4.1%), or custody evaluation ($n = 1$, 1.4%). Of the 29 clients referred for therapy, 23 of 29 (79.3%) sought individual therapy, 4 (13.8%) sought couple's therapy, and the remaining 2 (6.9%) sought family therapy. A majority of the clients (75.7%) reported a history of mental health treatment: 49 clients (66.2%) had previously received outpatient therapy, 20 (27.0%) had previously taken psychotropic medication, 11 (14.9%) had undergone a psychological assessment, and 7 (9.5%) had received inpatient psychiatric care.

Procedures

We reviewed the clinical charts of individuals who met the preceding inclusion/exclusion criteria. One of three clinical psychology doctoral students recorded all pertinent data onto a data collection sheet. We never documented identifying information, and we sorted data collection sheets by unique identification numbers. The Saint Louis University Institutional Review Board provided approval for this study.

Measures

We drew demographic data from the clinical reports. Data collected included age, gender, race/ethnicity, marital status, reason for referral, and history of psychiatric treatment.

The NEO–PI–R (Costa & McCrae, 1992b) is a self-report measure that provides an assessment of the Five-factor model of personality. The measure is comprised of 240 items answered on a 5-point scale ranging from 1 (*Strongly Disagree*) to 5 (*Strongly Agree*). The domains assessed by the NEO–PI–R include Neuroticism, Extraversion, Openness, Agreeableness, and Conscientiousness. Each of the five domains includes six facet scales. The NEO–PI–R has demonstrated excellent psychometric characteristics. Internal consistency has ranged from .86 to .92 for the five domains. Test–retest reliability has ranged from .63 to .83 (Costa & McCrae, 1992b). Standard scoring of the NEO–PI–R does not produce validity scale scores; therefore, we calculated the PPM, NPM, and INC Validity Scales for the NEO–PI–R using the procedures outlined by Schinka et al. (1997). Internal consistencies in this sample were .43 for PPM and .60 for NPM. We did not calculate internal consistency data for the INC scale, as the items do not share a similar content (Young & Schinka, 2001). In addition, we evaluated a validity index developed by Berry et al. (2001). This scale is derived by subtracting PPM from NPM, which may be analogous to the MMPI–2 F – K Index and suggests an attempt to present in an overly negative fashion. Consistent with prior research (Caldwell-Andrews et al., 2000; Yang, Bagby, & Ryder, 2000), cutoff scores for detecting invalid responding were greater than or equal to 22 on PPM and greater than or equal to 16 on NPM. For the NPM – PPM index, scores greater than or equal to –1 were considered invalid (Berry et al., 2001).

The MMPI–2 (Butcher et al., 1989) is a broad measure of personality and psychopathology. The test contains 567 true or false items that comprise the Validity Scales, 10 Clinical Scales, and Content and Supplementary Scales. Psychometric properties of the MMPI–2 have been adequate, with internal consistencies that have ranged from .34 to .87. One-week test–retest reliabilities have ranged from .58 to .92 (Butcher et al., 1989; Butcher, Graham, Dahlstrom, & Bowman, 1990). The MMPI–2 Validity Scales have been well-researched and studies have indicated they can detect invalid responding (Bagby et al., 1997; Berry et al., 1991; Gallen & Berry, 1996; Graham, 2000; Graham et al., 1991; Rogers et al., 1994). When collecting data for this study, we recorded MMPI-2 Clinical and Validity Scale scores onto a data collection sheet. We did not obtain individual responses. Thus, it is not possible to compute internal consistency coefficients for the MMPI–2 Validity and Clinical Scales. The specific validity indexes examined in this study were the L, F, K, F – K, F(b), and Variable Response Inconsistency (VRIN) scales. We used the L and K scales as measures of PIM that identified individuals who attempted to present in an overly favorable manner. The F scale is a measure of deviant response taking and indicates attempts to exaggerate psychopathology. The F – K scale is the raw score difference between the F scale and K scale, and we used this as an index for detecting fake-bad profiles. The F(b) scale measures invalid test taking on the second half of the MMPI–2 items and may be used to identify exaggerated symptoms. The VRIN scale is a measure of random responding.

We made a series of comparisons to evaluate the diagnostic overlap of invalid responding between the NEO–PI–R Validity Scales using MMPI–2 Validity Scales as the criterion measures. For these analyses, we compared a T score ≥ 65 on the MMPI–2 K scale with the PPM as indicative of attempting to present in an overly positive fashion. We used a T score ≥ 80 on the MMPI–2 F scale to compare with the NPM scale to indicate malingered pathology. A raw score ≥ 11 on the MMPI–2 F – K scale was the criterion for the NPM – PPM scale, which was an indication of exaggerated psychopathology.

RESULTS

We analyzed the proportion of invalid profiles using MMPI–2 F and K scales as the validity criterion by type of referral question. For the F scale, 18.4% (7/38) of clients who completed an assessment of educational functioning 11.1% (3/27) of clients referred for therapy, and none (0/9) of the clients who completed other assessments (referral questions included neuropsychological evaluation, personality assessments, and custody evaluations) had elevated T scores. A chi-square analysis of the proportions for these three groups was nonsignificant ($p = .31$). For the K scale, 7.9% (3/38) of clients referred for educational assessments, 3.7% (1/27) of clients referred for therapy, and 22.2% (2/9) of the "other" assessments had excessively elevated MMPI–2 K scores. A chi-square analysis indicated the proportion of K scale elevated profiles was not significantly different between the groups ($p = .21$).

Correlations between the NEO–PI–R Validity Scales and MMPI–2 Validity Scales are presented in Table 1. The NPM scale and MMPI–2 scales assessing exaggerated symptom reporting were positively and significantly correlated (correlations with F, F–K, and F(b) were all greater than or equal to .30, $p < .01$). In addition, NPM was significantly negatively correlated with the MMPI–2 L scale but not the K scale. The correlations between the PPM scale and MMPI–2 scales assessing PIM were positive and statistically significant (correlations with the L and K scales were greater than or equal to .39, $p < .001$). The PPM scale was also significantly negatively correlated with MMPI–2 scales of symptom exaggeration (correlations with F, F – K, and F(b) were stronger than or equal to –.46, $p < .001$). The NPM – PPM scale significantly ($p < .001$) correlated with the F – K Index. Scores on the NEO–PI–R INC were not significantly ($p > .05$) correlated with the MMPI–2 VRIN scale.

We examined the extent of agreement between the recommended cutoff scores for NEO–PI–R Validity Scales and MMPI–2 Validity Scales by overall classification accuracy and kappa statistics. Table 2 presents results from measures of PIM and NIM. We used the MMPI–2 K scale as the criterion for comparison with the PPM scale. The scales yielded the same classification in 82.4% of the cases, which indicated that a similar proportion of patients were attempting to present in a positive fashion. The diagnostic agreement for the PPM scale and K scale (as measured by kappa) was .30. A similar comparison between the NPM scale and the MMPI-2 F scale yielded 82.4% classification agreement between the two scales. The diagnostic agreement of invalid responding by faking bad was .22. Finally, to evaluate the new NPM – PPM scale, we compared its diagnostic accuracy with the MMPI–2 F-K index. The NPM – PPM scale and F – K Index had 81.1% classification accuracy, and the kappa was .33. We did not analyze the diagnostic agreement between INC and VRIN, as the correlation between the two scales was not significant.

We examined receiver operator characteristic curves to detect the best cut point for profile validity on the

TABLE 1
Zero-Order Correlations Between NEO–PI–R Validity Scales and MMPI–2 Validity Scales

Scale	*NPM*	*PPM*	*INC*	*NPM–PPM*	*L*	*F*	*K*	*F–K*	*F(b)*	*VRIN*
NEO–PI–R										
NPM	1.00									
PPM	−.48***	1.00								
INC	−.08	.30**	1.00							
NPM – PPM	.85***	−.87***	.22	1.00						
MMPI–2										
L	−.25*	.39***	.29*	−.38***	1.00					
F	.31**	−.52***	−.14	.49***	−.24*	1.00				
K	−.19	.41***	.23	−.35**	.43***	−.62***	1.00			
F – K	.30**	−.51***	−.20	.48***	−.35**	.90***	−.89***	1.00		
F(b)	.38***	−.46***	−.09	.49***	−.23*	.76***	−.63***	.79***	1.00	
VRIN	.26	−.09	.11	.20	−.08	.15	−.34**	.23*	.20	1.00

Note. $N = 74$. NEO–PI–R = Revised NEO Personality Inventory; MMP–2 = Minnesota Multiphasic Personality Inventory–2; NPM = Negative Presentation Management scale; PPM = Positive Presentation Management scale; INC = Inconsistency scale; NPM – PPM = NPM scale minus PPM scale; L = L scale; F = F scale; K = K scale; F – K = F minus K scale; F(b) = F(b) scale; VRIN = Variable Response Inconsistency scale. $*p < .05$, two-tailed. ** $p < .01$, two-tailed. ***$p < 001$, two-tailed.

TABLE 2
Concordance Between the NEO–PI–R Validity Scales and MMPI–2 Validity Scales

		MMPI-2 K Scale		*MMPI-2 F Scale*		*MMPI-2 F-K Index*			
Scale	*Validity*	*Valid (T score ≤ 64)*	*Invalid (T score ≥ 65)*	*Valid (T score ≤ 79)*	*Invalid (T score ≥ 80)*	*Valid (Raw score ≤ 10)*	*Invalid (Raw score ≥ 11)*	*Overall Agreement*	*Cohen's Kappa*
PPM	Valid (≤ 21)	57	2					82.4%	.30
	Invalid (≥ 22)	11	4						
NPM	Valid (≤ 15)			58	7			82.4%	.27
	Invalid (≥ 16)			6	3				
NPM – PPM	Valid (≤ –2)					56	1	81.1%	.33
	Invalid (≥ –1)					13	4		

Note. NEO–PI–K = Revised NEO Personality Inventory; MMPI–2 = Minnesota Multiphasic Personality Inventory–2; PPM = Positive Presentation Management; NPM = Negative Presentation Management; NPM–PPM = Negative Presentation Management scale minus the Positive Presentation Management scale.

TABLE 3
Index Scores of Participants With Invalid (≥ 22) Versus Valid (< 22) Scores on the Positive Presentation Management (PPM) Scale

	Invalid PPM[a]		*Valid PPM*[b]				
Scale	*M*	*SD*	*M*	*SD*	*Statistic (df)*	*p value*	*Cohen's d*
NEO–PI–R							
Neuroticism	49.2	5.3	59.5	12.7	$F(1, 73) = 9.43$	.003[c]	.84
Extraversion	56.1	9.5	48.1	12.1	$F(1, 73) = 5.66$	.02	.67
Openness	57.8	6.6	56.6	9.3	$F(1, 73) = 0.22$	.64	.14
Agreeableness	52.7	10.2	49.0	9.6	$F(1, 73) = 1.74$	.19	.38
Conscientiousness	50.5	12.0	40.0	13.0	$F(1, 73) = 8.12$	.006[c]	.84
MMPI–2							
1	57.7	9.3	57.8	10.6	$F(1, 73) = 0.02$	.97	.01
2	53.9	8.8	64.2	12.7	$F(1, 73) = 8.72$	.004[c]	.81
3	64.0	12.4	58.7	10.1	$F(1, 73) = 3.00$	.09	−.50
4	58.6	8.4	62.8	12.0	$F(1, 73) = 1.62$	.21	.37
5	52.5	11.6	51.3	10.7	$F(1, 73) = 0.15$	.70	.11
6	60.0	9.4	61.1	13.5	$F(1, 73) = 0.09$	.76	.09
7	54.9	10.1	66.0	13.9	$F(1, 73) = 8.34$	.005[c]	.80
8	56.6	6.8	66.3	13.1	$F(1, 73) = 7.52$	.008	.76
9	57.1	11.0	53.6	12.2	$F(1, 73) = 0.99$	.32	−.29
0	43.1	8.3	56.1	11.0	$F(1, 73) = 18.11$	.001[c]	1.11

Note. NEO–PI–R = Revised NEO Personality Inventory; MMP–2 = Minnesota Multiphasic Inventory–2. After controlling for multiple comparisons, the threshold for statistical significance was $p \leq 01$ on the NEO–PI–R and $p \leq .005$ on the MMPI–2.
[a] $n = 15$. [b] $n = 59$. [c] Denotes a statistically significant difference between the two groups.

NEO–PI–R Validity Scales using MMPI–2 Validity scores as the criterion. For PIM, a cutoff of 21 or greater on the PPM scale had the best balance for sensitivity (sensitivity = .83, specificity = .75), whereas a score of 22 or greater improved specificity (sensitivity = .67, specificity = .84; area under the curve [AUC] = .82). For negative impression management, a cutoff score of 12 or greater maximized sensitivity on the NPM scale (sensitivity = .70, specificity = .63), whereas a score of 15 or greater on NPM improved specificity (sensitivity = .50, specificity = .84; AUC = .69). Finally, on the NPM – PPM scale, a score greater than or equal to 0 was the best choice for a balance between sensitivity and specificity (sensitivity = .80, specificity = .84; AUC = .84).

Table 3 summarizes NEO–PI–R and MMPI–2 index scores of clients who responded invalidly on the PPM scale compared with valid PPM responders. We modified the alpha level to detect significance on these comparisons using the Bonferroni correction applied at the individual test (i.e., .05/5 for the NEO–PI–R, .05/10 for the MMPI–2) to control for multiple comparisons. Clients who presented in an overly positive fashion had significantly different NEO–PI–R profiles and MMPI–2 profiles than clients with valid PPM scores. Univariate tests indicate that invalid PPM responders had lower scores on the NEO–PI–R Neuroticism scale and higher scores on Conscientiousness. On the MMPI–2, invalid PPM respondents had lower scores on Scales 2, 7, and 0 compared with valid PPM responders. A review of the effect sizes in these comparisons, expressed as Cohen's d, indicated that the magnitude of differences were large.

TABLE 4
Index Scores of Participants With Invalid (≥ 16) Versus Valid (< 16) Scores on the Negative Presentation Management (NPM) scale

	Invalid PPM[a]		*Valid PPM*[b]				
	M	*SD*	*M*	*SD*	*Statistic (df)*	*p Value*	*Cohen's d*
NEO–PI–R							
Neuroticism	64.0	10.2	56.3	12.3	$F(1, 73) = 3.01$	.09	.63
Extraversion	40.4	6.8	51.1	12.0	$F(1, 73) = 6.67$	.01[c]	.89
Openness	56.4	8.8	56.9	8.8	$F(1, 73) = 0.02$	.88	.06
Aggreeableness	47.6	13.5	50.9	9.2	$F(1, 73) = 0.53$	.47	.34
Conscientiousness	33.8	10.7	43.3	13.5	$F(1, 73) = 4.16$	.05	.71
MMPI–2							
1	59.9	7.0	57.5	10.7	$F(1, 73) = 0.43$	.51	.23
2	65.3	11.4	61.7	12.8	$F(1, 73) = 0.66$	.42	.28
3	57.1	8.0	60.1	11.1	$F(1, 73) = 0.63$	.43	−.28
4	68.1	12.5	61.1	11.1	$F(1, 73) = 3.10$	.08	.61
5	54.6	12.2	51.2	10.6	$F(1, 73) = 0.78$	.38	.31
6	62.1	14.3	60.7	12.7	$F(1, 73) = 0.09$	.77	.11
7	67.9	11.7	63.2	14.2	$F(1, 73) = 0.90$	.35	.34
8	68.2	11.8	63.8	12.8	$F(1, 73) = 0.98$	.33	.35
9	48.1	8.4	55.2	12.2	$F(1, 73) = 2.82$	.10	−.59
0	63.7	7.7	52.1	11.5	$F(1, 73) = 8.62$	.004[c]	.99

Note. NEO–PI–R = Revised NEO Personality Inventory; MMP–2 = Minnesota Multiphasic Personality Inventory–2. After controlling for multiple comparisons, the threshold for statistical significance was $p \leq .01$ on the NEO–PI–R and $p \leq .005$ on the MMPI–2.
[a] $n = 9$. [b] $n = 65$. [c] Denotes a statistically significant difference between two groups.

A comparison of invalid and valid NPM responders is presented in Table 4. Univariate comparisons revealed that individuals with a NPM style had significantly lower scores on the NEO–PI–R Extraversion scale. On the MMPI–2, the only clinical scale that differed between clients with valid and invalid scores on the NPM was Scale 0, which was higher for the invalid NPM responders. The magnitude of this effect was large.

DISCUSSION

Findings from this study provide additional evidence that the NEO–PI–R Validity Scales have potential utility for detecting response bias. The scales correlated significantly with MMPI–2 validity indicators in the expected directions, and they demonstrated reasonably high overall percent agreement with the MMPI–2 Validity Scales. In addition, participants who responded in an invalid manner on the NEO–PI–R Validity Scales yielded significantly different results on the NEO–PI–R and MMPI–2 Clinical Scales relative to participants who responded in a valid manner.

Although this report provides support for the NEO–PI–R Validity Scales, caution is warranted. Using the MMPI–2 Validity Scales as the comparison, the NEO–PI–R Validity Scales yielded relatively high rates of false negative classification. In addition, the internal consistency scores for the NEO–PI–R validity indicators were low (.43 for PPM and .60 for NPM). These alpha coefficients were lower than previously reported (Young & Schinka, 2001) and may reflect the heterogeneity of this sample. Therefore, the reliability of these indicators may be questionable in some samples. Additional research is needed to explore the psychometric properties of the NEO–PI–R Validity Scales in clinical settings and to compare the scales with other reliable and valid measures to better understand the diagnostic accuracy of these indexes.

Several characteristics of this study limit the results. The cutoff rates for test validity on the MMPI–2 were based on general guidelines. In clinical practice, the validity index T scores may be modified to the population under examination (Graham, 2000). Different rates of diagnostic agreement might be identified if the cutoff level for profile validity were altered. A second limitation was the omission of the F(p) scale from the MMPI–2. The MMPI–2 protocols were scored via computer, which did not report the F(p) scale when these clinical batteries were completed. Research has suggested the F(p) scale may be superior to other MMPI–2 indicators in detecting malingering (Arbisi & Ben-Porath, 1995, 1998; Bagby, Nicholson, Bacchiochi, Ryder, & Bury, 2002; Rogers, Sewell, Martin, & Vitacco, 2003). Studies investigated the diagnostic agreement of the NEO–PI–R NPM scale with the MMPI–2 F(p) scale are needed to corroborate the results presented here. Third, the comparisons of NEO–PI–R Clinical scales between those with valid and invalid protocols may have been affected by unequal and small cell sizes. However, the reduced power did not preclude observation of significant differences. Another limitation relates to nature of the sample utilized for this investigation. Participants in this report were

seeking psychological services at a university-based clinic and were not recruited for a research study. There may have been certain clinical factors that led to the administration of these particular psychological measures, which in turn could have impacted the findings.

In conclusion, the results reported in this study combined with prior research (Ballenger et al., 2001; Berry et al., 2001; Caldwell-Andrews et al., 2000; Yang et al., 2000; Young & Schinka, 2001) suggest that the NEO–PI–R Validity Scales have promise as indictors for the detection of biased responding. As indicated by Morey et al. (2002), responses on the NEO–PI–R Validity scale items may also reflect a presentational style that is distinct from, yet similar to, response distortion. It remains to be seen if modifications to the scales, the use of different cut points for validity, or empirical investigation in diverse populations results in improved diagnostic accuracy. Based on the results from this and prior investigations combined with the general clinical utility of the NEO–PI–R, future research of the measure, including the Validity Scales, appears highly warranted.

ACKNOWLEDGMENTS

We thank Lauren Schwarz for help with data collection, Jeremiah Weinstock for assistance with data analysis, and the comments from three reviewers on a prior draft of this article.

REFERENCES

Arbisi, P. A., & Ben-Porath, Y. S. (1995). An MMPI–2 infrequent response scale for use with psychopathological populations: The Infrequency-Psychopathology scale, F(p). *Psychological Assessment, 7,* 424–431.

Arbisi, P. A., & Ben-Porath, Y. S. (1998). The ability of Minnesota Multiphasic Personality Inventory–2 Validity Scales to detect fake-bad responses in psychiatric inpatients. *Psychological Assessment, 10,* 221–228.

Bagby, R. M., Nicholson, R. A., Bacchiochi, J. R., Ryder, A. G., & Bury, A. S. (2002). The predictive capacity of the MMPI–2 and PAI validity scales and indexes to detect coached and uncoached feigning. *Journal of Personality Assessment, 78,* 69–86.

Bagby, R. M., Rogers, R., Nicholson, R. A., Buis, T., Seeman, M. V., & Rector, N. A. (1997). Effectiveness of the MMPI–2 validity indicators in the detection of defensive responding in clinical and non-clinical samples. *Journal of Personality Assessment, 68,* 650–664.

Ballenger, J. F., Caldwell-Andrews, A., & Baer, R. A. (2001). Effects of positive impression management on the NEO Personality Inventory–Revised in a clinical population. *Psychological Assessment, 13,* 254–260.

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

Ben-Porath, Y. S., & Waller, N. G. (1992b). "Normal" personality inventories in clinical assessment: General requirements and the potential for using the NEO Personality Inventory. *Psychological Assessment, 4,* 14–19.

Berry, D. T., Baer, R. A., & Harris, M. J. (1991). Detection of malingering on the MMPI: A meta-analysis. *Clinical Psychology Review, 11,* 585–598.

Berry, D. T. R., Bagby, R. M., Smerz, J., Rinaldo, J. C., Caldwell-Andrews, A., & Baer, R. A. (2001). Effectiveness of NEO–PI–R research Validity Scales for discriminating analog malingering and genuine psychopathology. *Journal of Personality Assessment, 76,* 496–516.

Borkenau, P., & Ostendorf, F. (1992). Social desirability scales as moderator and suppressor variables. *European Journal of Personality, 6,* 199–214.

Butcher, J. N., Dahlstrom, W. G., Graham, J. R., Tellegen, A., & Kaemmer, B. (1989). *MMPI–2: Minnesota Multiphasic Personality Inventory–2: Manual for administration and scoring*. Minneapolis: University of Minnesota Press.

Butcher, J. N., Graham, J. R., Dahlstrom, W. G., & Bowman, E. (1990). The MMPI–2 with college students. *Journal of Personality Assessment, 54,* 1–15.

Caldwell-Andrews, A., Baer, R. A., & Berry, D. T. R. (2000). Effects of response sets on NEO–PI–R scores and their relations to external criteria. *Journal of Personality Assessment, 74,* 472–488.

Clark, L. A. (1993). *Manual for the Schedule for Nonadaptive and Adaptive Personality.* Minneapolis: University of Minnesota Press.

Costa, P. T., Jr., & McCrae, R. R. (1992a). Normal personality in clinical practice: The NEO Personality Inventory. *Psychological Assessment, 4,* 5–13.

Costa, P. T., Jr., & McCrae, R. R. (1992b). *Professional manual for the NEO Personality Inventory.* Odessa, FL: Psychological Assessment Resources.

Costa, P. T., Jr., & McCrae, R. R. (1992c). Reply to Ben-Porath and Waller. *Psychological Assessment, 4,* 20–22.

Costa, P. T., Jr., & McCrae, R. R. (1997). Stability and change in personality assessment: The Revised NEO Personality Inventory in the year 2000. *Journal of Personality Assessment, 68,* 86–94.

Drury, J. A. (2001). *The detection of malingering and positive impression management profiles on the NEO–PI–R: A comparison of methods.* Unpublished doctoral dissertation, Saint Louis University, Saint Louis, Missouri.

Gallen, R. T., & Berry, D. T. R. (1996). Detection of random responding in MMPI–2 protocols. *Assessment, 3,* 171–178.

Graham, J. R. (2000). *MMPI–2: Assessing personality and psychopathology* (3rd ed.). New York: Oxford University Press.

Graham, J. R., Watts, D., & Timbrook, R. E. (1991). Detecting fake-good and fake-bad MMPI–2 profiles. *Journal of Personality Assessment, 57,* 264–277.

McCrae, R. R., & Costa, P. T., Jr. (1997). Personality trait structure as a human universal. *American Psychologist, 52,* 509–516.

Morey, L. C. (1991). *Personality Assessment Inventory: Professional manual*. Odessa, FL: Psychological Assessment Resources.

Morey, L. C., Quigley, B. D., Sanislow, C. A., Skodol, A. E., McGlashan, T. H., Shea, M. T., et al. (2002). Substance or style? An investigation of the NEO–PI–R Validity Scales. *Journal of Personality Assessment, 79,* 583–599.

Nicholson, R. A., & Hogan, R. (1990). The construct validity of social desirability. *American Psychologist, 45,* 290–292.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.

Piedmont, R. L., McCrae, R. R., Riemann, R., & Angleitner, A. (2000). On the invalidity of validity scales: Evidence from self-reports and observer ratings in volunteer samples. *Journal of Personality and Social Psychology, 78,* 582–593.

Quirk, S. W., Christiansen, N. D., Wagner, S. H., & McNulty, J. L. (2003). On the usefulness of measures of normal personality for clinical assessment: Evidence of the incremental validity of the Revised NEO Personality Inventory. *Psychological Assessment, 15,* 311–325.

Reynolds, S. K., & Clark, L. A. (2001). Predicting dimensions of personality disorder from domains and facets of the five-factor model. *Journal of Personality Assessment, 69,* 199–222.

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.

Rogers, R., Sewell, K. W., & Salekin, R. T. (1994). A meta-analysis of malingering on the MMPI–2. *Assessment, 1,* 227–237.

Rosse, J. G., Stecher, M. D., Miller, J. L., & Levin, R. A. (1998). The impact of response distortion on preemployment personality testing and hiring decisions. *Journal of Applied Psychology, 83,* 634– 644.

Schinka, J. A., Kinder, B. N., & Kremer, T. (1997). Research Validity Scales for the NEO–PI–R: Development and initial validation. *Journal of Personality Assessment, 68,* 127–138.

Streiner, D., & Norman, G. (2003). *Health measurement scales* (3rd ed.). Oxford, England: Oxford University Press.

Yang, J., Bagby, R. M., & Ryder, A. G. (2000). Response style and the Revised NEO Personality Inventory: Validity scales and spousal ratings in a Chinese psychiatric sample. *Assessment, 7,* 389–402.

Young, M. S., & Schinka, J. A. (2001). Research Validity Scales for the NEO–PI–R: Additional evidence for reliability and validity. *Journal of Personality Assessment, 76,* 412–420.

Benjamin Morasco, Ph.D.
Portland VA Medical Center
3710 SW US Veterans Hospital Road (P3MHDC)
Portland, OR 97239
Email: benjamin.morasco@va.gov

Received November 13, 2005
Revised August 3, 2006