

Internal and External Validity of Scores on the Balanced Inventory of Desirable Responding and the Paulhus Deception Scales

Richard I. Lanyon
Adam C. Carle
Arizona State University

The internal and external validity of scores on the two-scale Balanced Inventory of Desirable Responding (BIDR) and its recent revision, the Paulhus Deception Scales (PDS), developed to measure two facets of social desirability, were studied with three groups of forensic clients and two groups of college undergraduates (total $N = 519$). The two scales were statistically significantly related in all groups and for both versions of the inventory. A two-factor congeneric, orthogonal measurement model was rejected for all groups. However, a two-factor model that allowed cross-loadings among the items and correlation between the factors provided adequate fit. Concurrent validity data showed scores on both the Impression Management and Self-Deceptive Enhancement (SDE) scales to be satisfactory measures of their respective constructs and also of general social desirability, for both forensic clients and undergraduates. An exception was found in lower validity correlates for scores on the SDE scale in the PDS form.

Keywords: *desirability; deception; BIDR; PDS; forensic*

This article concerns the equivalence of scores on the Balanced Inventory of Desirable Responding (BIDR); its revision, the Paulhus Deception Scales (PDS); and the applicability of each to forensic populations. The BIDR is a 40-item, two-scale inventory developed by Delroy Paulhus (1984, 1986, 1994) based on college student responses to assess the dimensions of a two-factor model of social desirability. This work was based on the earlier work of Sackeim and Gur (1978), who had proposed the division of the traditional social desirability concept into “other” (or conscious) deception and “self” (or unconscious) deception.

Authors' Note: Adam C. Carle is now with the Department of Psychology, University of North Florida. Parts of this article are based on honors theses completed by Jamie Cottle in 2001 and Lauren J. Garby in 2002. Please address correspondence to Richard I. Lanyon, Department of Psychology, Arizona State University, Tempe, AZ 85287-1104; e-mail: Rlanyon@asu.edu.

Paulhus (1986) recognized the need for refined measures of these concepts and presented empirical support for his two-scale BIDR based on a model of two separate dimensions. Thus, scores on the Impression Management (IM) scale were demonstrated to assess conscious deception, as shown by correlations with other measures such as the MMPI Lie (L) scale and the L scale of the Eysenck Personality Inventory. A recent review of the general area of favorable self-presentation has supported the IM scale as a central measure of a relatively narrow concept that is appropriately referred to as exaggerated claims of extreme virtue (Lanyon, 2004). In regard to the Self-Deceptive Enhancement (SDE) scale, originally proposed to represent unconscious deception, Paulhus's (1986) data showed that scores were strongly correlated with Block's (1965) Ego-Resiliency (ERS) scale, Edwards' (1957) Social Desirability (So-r) scale, and (in reverse) the Taylor Manifest Anxiety Scale (At; Taylor, 1953). Most of the content of the latter two scales represents the simple affirmation or denial of psychopathology.

Paulhus subsequently made significant changes to the scoring system of the BIDR, replaced an item, and published the revision as the PDS (Paulhus, 1998). Data presented at that time assigned a further set of meanings to the SDE scale scores, involving such concepts as ego-inflation, narcissism, and arrogance and the general concept of self-deceptive positivity (e.g., Nichols & Greene, 1997). A review by Salekin (2000) supported these expanded meanings. The PDS also provides a single summary score, the sum of the two scales, as a single representation of favorable self-presentation or socially desirable responding.

Studies reviewed by Paulhus (1998) showed a small correlation between the SDE and IM scales, although a confirmatory factor analysis based on the PDS "sustained the two-factor structure previously obtained in a series of exploratory factor analyses" (p. 23), and found that an orthogonal model was superior to an oblique model. Other researchers were listed as having replicated the same structure (Borkeneau & Ostendorf, 1989; Nichols & Greene, 1989).

For the BIDR, the 20 items on each scale are each rated 1 through 7 by the respondent. Alternate items are reversed in content direction. Items are scored a point if an extreme rating is given (1 or 2; 6 or 7 for alternate items). Thus, the maximum score is 20 for each scale. For the PDS, items are rated 1 through 5. Ratings of 1 or 2 (4 or 5 for reversed items) count a point for the IM scale, but for the SDE scale, only ratings of 1 (5 for reversed items) are counted a point. The extent to which these changes have affected the comparability of the PDS and the BIDR is not known. Such knowledge is necessary in order to determine whether the existing validity research on the BIDR is applicable to the PDS. The mean score for the SDE scale (as listed in the *PDS Manual*; Paulhus, 1998) is now only 2.2, compared with 7.6, 7.3, 7.5, and 6.8 reported for the BIDR (Paulhus, 1991). For IM, the reported mean is 6.7, compared with 4.6 and 4.9 for the BIDR. These large discrepancies raise the question that scores resulting from these two versions may not be comparable either in internal structure or validity.

Norm groups for the PDS reported by Paulhus (1998) include the general population, prison inmates, and military recruits, indicating that it was intended for application beyond college students. And, in fact, the PDS has found increasing use in clinical-forensic assessment and research involving forensic populations (e.g., Hemphill & Howell, 2000; Kroner & Weekes, 1996a, 1996b; Looman, Abracen, Maillet, & DiFazio, 1998; Nugent & Kroner, 1996; Seto & Barbaree, 1993). The scales were developed in the context of social psychological research involving college undergraduates, and a wide variety of validity data were reported within that context. However, the appropriateness of using their scores in forensic contexts has not been adequately evaluated (see Salekin, 2000) for either the BIDR or the PDS. In particular, supportable use in a forensic context requires the availability of validity data specific to that context, in which there is often a great deal more at stake than in an undergraduate research setting. The question of the validity of scores on the BIDR/PDS in forensic work was the primary reason for conducting the present study. Early informal analysis of some of the PDS forensic data utilized in the present research showed a substantial correlation between the SDE and IM scales, well beyond the values (.14 and .20) reported in Paulhus's (1998) work.

The primary concerns and issues listed above can be summarized under two general headings: the extent to which the PDS is comparable to the BIDR in internal and external validity, and the validity of scores resulting from both tests in forensic settings (in addition to undergraduate settings). The answers to these questions would also inform conclusions about whether the existing validity research on the BIDR is applicable to the PDS.

The present research was designed to address these issues. Utilizing both forensic and undergraduate participants, the following tasks were set using both tests: (a) to examine the relationship of the SDE and IM scales, (b) to examine the fit of a two-factor model (Paulhus, 1998) for the individual items, and (c) to determine validity correlates. Specifically, it was hypothesized that previous findings for the BIDR and PDS with college undergraduates would be shown to apply in forensic assessment situations and would be confirmed for undergraduates. There have been several forms of the BIDR; the version utilized in the present research was the BIDR-6 Form 40 (Paulhus, 1991). This appears to have been the form used in most of the published work, although studies have often not included such information, and adequate psychometric information is often not reported by authors (see e.g., Vacha-Haase, Henson, & Caruso, 2002).

Method

Participants

Seven different sets of data were employed in this study, involving a total of 519 participants in five different groups, as follows.

First forensic group. These participants (Group 1, $n = 128$) were clients who had been evaluated by the first author within a forensic context for one of a variety of purposes, including child custody issues, personal injury litigation, sex offenses, and other criminal offenses. There were 86 men and 42 women, mean age was 33 years (range 18-72), and mean education level was 13 years (range 7-20 years). All had completed the MMPI-2, the BIDR, and the Psychological Screening Inventory (PSI; Lanyon, 1978, 1993) at a single sitting.

Second forensic group. These participants (Group 2, $n = 72$) were patients with personal injury, workers' compensation, or personal injury claims. Most had a claimed physical disorder or injury. All had been formally examined in a psychiatric clinic in order to determine the presence and extent of psychiatric and psychological difficulties and their relationship to the claimed disability. There were 23 men and 49 women, mean age was 42 (range 20-64), and mean education level was 13 years (range 6-20). All had completed the MMPI-2 and the BIDR, and 50 had completed the SCT-75 Sentence Completion Test (SCT-75; Lanyon, Almer, & Maxwell, 2002; Timmons, Lanyon, Almer, & Curran, 1993).

Third forensic group. These participants (Group 3, $n = 101$) came from the same source as Group 2, but the data were gathered in a different year and in the context of a different project. There were 57 men and 44 women, mean age was 44 (range 23-59), and mean education level was 14 years (range 10-20). Most claimed a disabling psychological condition. All had completed the MMPI-2, the PDS, and the SCT-75.

First undergraduate group. These participants (Group 4, $n = 103$) took part in the study to fulfill a course requirement. They completed the BIDR, the PDS, and the PSI. They also completed the MMPI-2 validity scales L, F (Infrequency), and K (Correction) as well as the So-r and At based on the MMPI.

Second undergraduate group. These participants (Group 5, $n = 115$) also took part as a way of fulfilling a course requirement but were tested in a different year than Group 4 and in a different context. They completed the BIDR, the PDS, the Good Impression (Gi) scale of the California Psychological Inventory (CPI; Gough, 1975), the Motivational Distortion (MD) and Faking-Bad (FB) scales of the Sixteen Personality Factors Questionnaire (16PF; Winder, O'Dell, & Karson, 1975), and the Desirability (Dy) scale of the Personality Research Form (PRF; Jackson, 1974). Group 4 participants completed the BIDR and the PDS 1 week apart, whereas Group 5 participants completed them at the beginning and end of a 1-hour testing period.

Measures

Measures selected for the assessment of concurrent validity were as follows. Because the data for the two undergraduate groups (Groups 4 and 5) were collected for research purposes, there was the opportunity to utilize different validation measures for each group. On the other hand, measures employed with Groups 1, 2, and 3 were constrained by what had been clinically appropriate to use in these real-life assessment situations.

Exaggeration of virtue. Because the concept of exaggerated virtue is known to be a narrow one (e.g., Lanyon, 2004; Strong, Greene, & Kordinak, 2002), selecting appropriate measures for assessing the concurrent validity of the IM scale in the present study was straightforward. Based on a recent review and analysis of the empirical literature in this area (Lanyon, 2004), the following measures were selected. For the forensic groups, measures included the MMPI-2 L scale, the sum of the L and K scales (here termed the L+K index; see Baer, Wetter, Nichols, Greene, & Berry, 1995), the Endorsement of Excessive Virtue (EEV) scale of the PSI (Lanyon, 1993), and the Excessive Virtue (EXCVIR) scale of the SCT-75 (Lanyon et al., 2002). For the undergraduate groups, measures included the Gi scale of the CPI (Gough, 1975), the MD scale of the 16PF (see Lanyon, Dannenbaum, Wolf, & Brown, 1989), and the MMPI-2 L + K index.

Self-deceptive enhancement. The concept assessed by the SDE scale is much more heterogeneous, mainly encompassing claims of superior mental health adjustment but also including claims of personal superiority and other concepts. To make the task of selecting validation measures for SDE even more difficult, analysis of item content has shown that scales purporting to assess these concepts often in fact assess a different dimension or more than one dimension (see Lanyon, 2004). For example, on the MMPI-2 Superlative (S; Butcher & Han, 1995) scale, presented as a measure of exaggerated positive characteristics, more than half of the items involve the denial of psychopathology. The same is true of Edwards' (1957) Social Desirability (So and So-r) scale. The item content of Wiggins' (1959) Social Desirability (Wsd) scale and the Positive Malingering (Mp) scale of Cofer, Chance, and Judson (1949) involve about 50% exaggerated-virtue items, with the remainder representing exaggeration of adjustment, personal superiority, and other concepts.

Taking these difficulties into consideration, the following scheme was adopted in selecting concurrent validation measures for the SDE scale. Based on the findings that the exaggeration of good adjustment and poor adjustment are probably opposite extremes of the same concept (Lanyon, 2004) and that SDE is highly correlated with this concept (Paulhus, 1986), several established measures of each "pole" of this concept—exaggeration of good/poor adjustment—were utilized. Also included were two scales of the original MMPI whose item content represents good/poor adjustment and that correlated highly with SDE in Paulhus's (1986) work: At and So-r.

For forensic groups, the PSI Endorsement of Superior Adjustment (ESA) scale (Lanyon, 1993) was selected to assess exaggerated claims of good adjustment. Scales selected for assessing exaggerated claims of poor adjustment included the Erroneous Psychiatric Stereotype (EPS) and Symptom Overendorsement (SOE) scales of the PSI; the F, Fb (Back-Page Infrequency), and Dsr2 (Dissimulation-Revised) scales and F-K index of the MMPI-2; and the Angry Negativity (ANGNEG) scales of the SCT-75. For undergraduate groups, scales selected with established validity for assessing exaggeration of good adjustment included the PSI ESA scale and the MMPI-2 So-r scale. Scales for assessing the exaggeration of poor adjustment included the MMPI-2 F and MMPI At scales and the Faking Bad (FB) scale of the 16PF.

General social desirability. Scales were also utilized that are held by the literature to represent social desirability or favorable self-presentation in general, including the MMPI-2 K and S scales, the PRF Dy scale, and the PSI Defensiveness (De) scale. These scales are presumably considered to be composites of both types of favorable self-presentation. However, their anticipated correlates were uncertain because their item content is not entirely consistent with the way they are described in the literature. For example, the content of the K scale (like the S scale) mostly involves the denial of psychopathology (Lanyon, 2004). Because of the secondary nature of the reported correlates of the SDE scale in the areas of ego inflation and narcissism, and because measures of these characteristics are not readily available, this aspect of SDE was not studied directly. For forensic groups, measures for assessing socially desirable responding in general included the MMPI-2 K and S scales, PSI De, and MMPI-2 adaptations of Mp (called Mp2) and Wsd (called Wsd2) scales. For undergraduate groups, scales with established validity for assessing socially desirable responding in general included PSI De, the MMPI-2 K scale, and PRF Dy. The MMPI-2 scales were available for all groups, whereas the PSI scales were available for Group 1, and ANGNEG and EXCVIR were available for Groups 2 and 3.

Results

Relationship of SDE and IM

In this and subsequent sections, the BIDR scales are designated SDE(B) and IM(B), and the PDS scales are designated SDE(P) and IM(P). The relationships between SDE(B) and IM(B) and between SDE(P) and IM(P) were evaluated by the simple correlation between them. For the BIDR in the forensic groups (1 and 2), the correlations between SDE and IM were .48, $p(.33 \leq \rho \leq .60) = .95$, and .58, $p(.40 \leq \rho \leq .72) = .95$, respectively. For the PDS (forensic Group 3), the correlation was .47, $p(.30 \leq \rho \leq .61) = .95$. For the undergraduate groups (4 and 5),

the correlations were .53, $p(.37 \leq \rho \leq .66) = .95$, and .30, $p(.12 \leq \rho \leq .46) = .95$, for the BIDR scales and .34, $p(.16 \leq \rho \leq .50) = .95$, and .15, $p(-.03 \leq \rho \leq .32) = .95$, for the PDS scales. For the combined undergraduate groups ($n = 218$), the correlations were .42, $p(.30 \leq \rho \leq .53) = .95$, and .25, $p(.12 \leq \rho \leq .37) = .95$, respectively.

The weighted average of the three correlations involving the forensic groups (McNemar, 1955; hereafter referred to as the “average”) was .50 and for the three correlations involving undergraduates was .34. The difference is statistically significant ($p < .01$), showing a stronger relationship for the SDE and IM scales in forensic than undergraduate participants. The average of all SDE/IM correlations involving the BIDR was .46, and the PDS was .33. This difference is statistically significant ($p < .05$), supporting Paulhus’s intention for greater independence of the PDS compared to the BIDR scales.

Goodness of Fit of the Two-Factor Model

Confirmatory factor analyses (CFA) examined the fit of several measurement models. To achieve adequate sample sizes, the BIDR was investigated for a combined forensic group (Groups 1 and 2), a combined undergraduate group (Groups 4 and 5), and a pooled forensic and undergraduate group (Groups 1, 2, 4, and 5). The PDS was explored in an undergraduate group (Groups 4 and 5) and a combined forensic and undergraduate group (Groups 3, 4, and 5).

Analyses first examined the fit of a two-factor, orthogonal, congeneric model for the BIDR and for the PDS as proposed by Paulhus (1998). The orthogonality hypothesis specified uncorrelated factors, and the congeneric hypothesis specified that the items loaded only on the construct they were intended to measure. The factors and items corresponded to the scales (IM and SDE) as identified by Paulhus. Mplus (Muthén & Muthén, 2004) was used for all analyses. Given the χ^2 's sensitivity to small deviations of fit and the fact that this test was statistically significant for each of the models considered below, fit analyses focused on close-fit indices and followed guidelines suggested by Hu and Bentler (1999) and Muthén and Muthén (2004).

CFA for ordered-categorical measures was used (Muthén, 1984) in the Mplus framework. This uses probit regression to estimate item correlations via the marginal and bivariate item distributions (Muthén & Muthén, 2004). These correlations are then used in a manner similar to CFA for continuous measures. In all analyses, the WLSMV (weighted least squares mean and variance adjusted) estimator, a robust estimator appropriate for categorical data (Muthén & Muthén, 2004), was used.

For each of the sample combinations, the two-factor, orthogonal, congeneric model fit poorly. Comparative fit index (CFI) values ranged from .468 to .739, weighted root-mean-square residual (WRMR) values ranged from 1.142 to 2.204, and none of the root-mean-square error of approximation (RMSEA) values were less than .05.

The next analysis, relaxing the orthogonality constraint, continued to demonstrate poor fit, evidencing a similar range of fit values. Thus, the fit of the two-factor, congeneric model allowing the factors to correlate was also rejected.

Two additional models were considered: a single-factor model and a two-factor model that allowed each item to cross-load on the factors and also allowed correlation between the factors. Across the groups, the set of fit indices suggested that the single-factor model was inadequate. CFI values ranged from .624 to .836, WRMR values ranged from 1.159 to 1.400, and none of the RMSEA values were less than .05. Given the indices, this model was rejected.

The final model examined the fit of a two-factor model allowing each item to cross-load on both factors and allowing the factors to correlate. For statistical identification (Millsap, 2001), the regression coefficient for Item 17 was constrained to one on the first factor and zero on the second, and the coefficient for Item 32 was fixed at one on the second factor and constrained to zero on the first. The remaining items were allowed to cross-load on both factors. The set of fit indices (see Table 1) demonstrated improved fit as compared to the previously considered models. CFI values ranged from .838 to .939, and WRMR values ranged from .904 to .995. None of the RMSEA values exceeded .044. Taken together, the fit indices suggested that the two-factor model allowing factors to correlate and the items to cross-load on the factors was tenable. For the models, the factor correlations ranged from $-.27$ to $-.43$, and although the models allowed cross-loadings, for each item across all models the highest regression coefficient was on the intended factor. Although space constraints limit presentation, the full set of model results is available by request.

Validity Correlates

As a general hypothesis, it was anticipated that the correlations of IM with the validity scales selected to assess exaggeration of virtue would be numerically greater than the correlations of SDE with these scales. Conversely, SDE was expected to show numerically higher correlations than IM with the validity scales selected to assess exaggeration of good/poor adjustment. In regard to criterion measures of general social desirability, it was predicted that SDE and IM would show comparable correlates with these measures. For each of the five participant groups, correlations were computed between scores on the BIDR/PDS scales and the particular concurrent validation scales as listed above. Data for the three forensic groups are shown in Tables 2, 3, and 4 and for the two undergraduate groups in Tables 5 and 6.

Overall, the vast majority of the confidence intervals (CIs) did not include zero (93/116). This finding allows the general conclusion that the concepts measured by both the IM and SDE scales of the BIDR and PDS indeed involve favorable self-presentation, broadly conceived. There are several ways to make a more detailed analysis of the concurrent validity correlations. Two methods were selected here.

Table 1
Final Two-Factor Model Allowing Correlated Factors and Item Cross-Loadings

Scale	Sample	χ^2	df^a	p	CFI	RMSEA	WRMR
BIDR	Forensic	176.694	127	.002	.939	.044	.904
	Undergraduate	151.530	110	.005	.848	.042	.968
	Pooled	339.890	209	< .001	.926	.039	.995
PDS	Undergraduate	297.011	90	< .001	.838	.041	.952
	Pooled	658.901	107	< .001	.888	.038	.951

Note: CFI = comparative fit index; RMSEA = root mean square of approximation; WRMR = weighted root-mean-square residual; BIDR = Balanced Inventory of Desirable Responding; PDS = Paulhus Deception Scales.

^aThe df for these models are a function of the WLSMV (weighted least squares mean and variance adjusted) estimator used in Mplus rather than the usual df .

First, comparisons were made involving weighted averages of subsets of the correlations; second, where available, comparisons were made between individual pairs of correlations.

Exaggeration of virtue. All but 1 of the 19 correlations of IM(B) and IM(P) with other measures of extreme virtue were positive. The average was .52; .50 for forensic groups, .48 for undergraduates, .54 for IM(B), and .49 for IM(P). The average correlation of SDE(B) and SDE(P) with the concurrent validity measures of exaggerated virtue was .37, statistically significantly less than the average for IM correlations ($p < .001$). In addition, the IM correlations were numerically higher than those for SDE for all but one correlate (involving the L+K index for disability claimants; see Table 3). These data indicate that both IM scales are satisfactory measures of exaggerated virtue (and are better than SDE) for forensic clients and also for undergraduates, but they also reflect the fact that SDE overlaps somewhat with IM in assessing exaggerated virtue.

Exaggeration of good/poor adjustment. Of the 24 correlations of SDE(B) and SDE(P) with measures of exaggerated good/poor adjustment, 17 of the CIs did not include zero. The average was .39: .42 for forensic groups, .35 for undergraduates, .46 for SDE(B), and .24 for SDE(P) ($p < .01$). The average correlation of the IM scales with the concurrent validity measures of exaggerated good/poor adjustment was .30, which does not differ statistically from the average for SDE correlations. The SDE correlations were nearly all numerically higher than those for IM for undergraduates but not for forensic clients. These data indicate that SDE(B) is a satisfactory measure of the exaggeration of good/poor adjustment, in particular for forensic clients, but raise a question about SDE(P), given that six of the seven CIs including zero were with the SDE(P) and one was with SDE(B).

Table 2
Means, Standard Deviations, and Concurrent Validity Correlates
of the BIDR Scales for 128 General Forensic Clients (Group 1)

Concurrent Validity Scales			BIDR Scales			
			SDE(B)		IM(B)	
<i>M</i>	<i>SD</i>	<i>r</i>	CI	<i>r</i>	CI	
<i>Exaggeration of virtue</i>						
L	4.34	2.27	.33	(.16 ≤ ρ ≤ .48)	.55	(.42 ≤ ρ ≤ .66)
L + K	19.80	6.64	.44	(.29 ≤ ρ ≤ .57)	.54	(.49 ≤ ρ ≤ .65)
EEV	15.76	6.99	.55	(.42 ≤ ρ ≤ .66)	.68	(.57 ≤ ρ ≤ .76)
<i>Exaggeration of good/poor adjustment</i>						
ESA	12.97	5.89	.64	(.53 ≤ ρ ≤ .73)	.43	(.27 ≤ ρ ≤ .56)
FK	42.14	11.16	-.52	(-.63 ≤ ρ ≤ -.38)	-.43	(-.56 ≤ ρ ≤ -.28)
Dsr2	8.48	5.97	-.49	(-.61 ≤ ρ ≤ -.34)	-.39	(-.53 ≤ ρ ≤ -.24)
EPS	9.17	5.04	-.50	(-.62 ≤ ρ ≤ -.36)	-.54	(-.65 ≤ ρ ≤ -.41)
SOE	6.40	5.27	-.51	(-.63 ≤ ρ ≤ -.37)	-.38	(-.52 ≤ ρ ≤ -.23)
F	7.18	5.87	-.49	(-.61 ≤ ρ ≤ -.34)	-.35	(-.49 ≤ ρ ≤ -.18)
<i>General social desirability</i>						
K	15.52	5.41	.43	(.28 ≤ ρ ≤ .56)	.48	(.33 ≤ ρ ≤ .60)
S	24.57	11.40	.44	(.39 ≤ ρ ≤ .57)	.54	(.40 ≤ ρ ≤ .65)
De	12.02	3.38	.47	(.32 ≤ ρ ≤ .60)	.49	(.34 ≤ ρ ≤ .61)

Note: BIDR = Balanced Inventory of Desirable Responding; SDE(B) = BIDR Self-Deceptive Enhancement; IM(B) = BIDR Impression Management; CI = confidence interval; L = MMPI-2 Lie; K = MMPI-2 Defensiveness; EEV = Psychological Screening Inventory (PSI) Endorsement of Excessive Virtue; ESA = PSI Endorsement of Superior Adjustment; F = MMPI-2 Infrequency; Dsr2 = MMPI-2 Dissimulation-Revised; EPS = PSI Erroneous Psychiatric Stereotype; SOE = PSI Symptom Overendorsement; Ra = PSI Random Responding; S = MMPI-2 Superlative; De = PSI Defensiveness.

General social desirability. Only 2 of the 30 CIs addressing the correlation of the SDE and IM with the measures of general social desirability included zero. The average correlation was .42: .40 for forensic groups and .45 for undergraduates, .41 for the SDE and .43 for IM, and .48 for the BIDR scales and .36 for the PDS scales ($p < .05$). Thus, both scales are satisfactory measures of general social desirability, but less so for the PDS than the BIDR.

Comparing individual correlations. The undergraduate data (Tables 5 and 6) allow direct comparison of concurrent validity correlations for the BIDR with those for the PDS. Specifically, there are 13 pairs of correlations for the SDE and 13 for IM. All of the correlation CI comparisons involving IM included zero, whereas five

Table 3
Concurrent Validity Correlates of the BIDR Scales
for 72 Disability Claimants (Group 2)

Concurrent Validity Scales			BIDR Scales			
			SDE(B)		IM(B)	
<i>M</i>	<i>SD</i>	<i>r</i>	CI	<i>r</i>	CI	
Exaggeration of virtue						
L	5.24	2.39	.40	(.19 ≤ ρ ≤ .58)	.53	(.35 ≤ ρ ≤ .68)
L + K	20.95	6.78	.52	(.32 ≤ ρ ≤ .67)	.50	(.31 ≤ ρ ≤ .66)
EXCVIR ^a	16.79	3.78	.26	(.03 ≤ ρ ≤ .46)	.30	(.08 ≤ ρ ≤ .50)
Exaggeration of good/poor adjustment						
F	8.09	6.29	-.51	(-.66 ≤ ρ ≤ -.32)	-.24	(-.44 ≤ ρ ≤ -.01)
Dsr2	9.94	6.57	-.58	(-.72 ≤ ρ ≤ -.40)	-.39	(-.57 ≤ ρ ≤ -.17)
F-K	-7.59	10.51	-.54	(-.69 ≤ ρ ≤ -.35)	-.34	(-.53 ≤ ρ ≤ -.12)
ANGNEG ^a	8.38	4.01	-.43	(-.60 ≤ ρ ≤ -.22)	-.25	(-.46 ≤ ρ ≤ -.02)
General social desirability						
K	15.72	5.38	.48	(.28 ≤ ρ ≤ .64)	.40	(.19 ≤ ρ ≤ .58)
S	26.35	9.87	.57	(.39 ≤ ρ ≤ .71)	.48	(.28 ≤ ρ ≤ .64)

Note: BIDR = Balanced Inventory of Desirable Responding; SDE(B) = BIDR Self-Deceptive Enhancement; IM(B) = BIDR Impression Management; CI = confidence interval; L = MMPI-2 Lie; K = MMPI-2 Defensiveness; EXCVIR = Sentence Completion Test-75 (SCT-75) Excessive Virtue; F = MMPI-2 Infrequency; Dsr2 = MMPI-2 Dissimulation-Revised; ANGNEG = SCT-75 Angry Negativity; S = MMPI-2 Superlative.

^a*N* = 50 for these correlations.

did not include zero for SDE: in Table 5, the correlations for L and At; and in Table 6, the correlations for Gi, FB, and Dy. These variables represent all three types of concurrent validity, indicating that the superiority of SDE(B) over SDE(P) has some generality, at least for the undergraduate participants.

Equivalence of SDE and IM on the BIDR and PDS. Correlations between SDE(B) and SDE(P), and between IM(B) and IM(P), were computed for the combined 218 undergraduate participants of Groups 4 and 5. The correlation was .69 for the SDE scales and .78 for the IM scales. The SDE correlation is lower than usual for short-term test-retest reliability.

Discussion

The main findings of the study were as follows: (a) Scores on the SDE and IM scales were positively related for both the BIDR and the PDS, more so for forensic

Table 4
Concurrent Validity Correlates of the PDS Scales for 101 Disability Claimants (Group 3)

	Concurrent Validity Scales						BIDR Scales					
	SDE(B)			IM(P)			SDE(B)			IM(P)		
	<i>M</i>	<i>SD</i>	<i>r</i>	CI	<i>r</i>	CI	<i>r</i>	CI	<i>r</i>	CI	<i>r</i>	CI
Exaggeration of virtue												
L	4.92	2.68	.48	(.32 ≤ <i>p</i> ≤ .62)	.49	(.32 ≤ <i>p</i> ≤ .62)	.56	(.42 ≤ <i>p</i> ≤ .68)				
L+K	20.83	6.80	.38	(.20 ≤ <i>p</i> ≤ .53)	.42	(.25 ≤ <i>p</i> ≤ .57)	.47	(.30 ≤ <i>p</i> ≤ .61)				
EXCVIR	14.73	3.46	.24	(.04 ≤ <i>p</i> ≤ .41)	.30	(.11 ≤ <i>p</i> ≤ .46)	.31	(.12 ≤ <i>p</i> ≤ .48)				
Exaggeration of good/poor adjustment												
F	8.50	6.14	-.14	(.32 ≤ <i>p</i> ≤ .06)	-.24	(-.41 ≤ <i>p</i> ≤ -.05)	-.22	(-.40 ≤ <i>p</i> ≤ -.03)				
Fb	5.35	5.65	-.16	(-.34 ≤ <i>p</i> ≤ .04)	-.17	(-.35 ≤ <i>p</i> ≤ .03)	-.19	(-.37 ≤ <i>p</i> ≤ .00)				
ANGNEG	9.44	4.30	-.26	(-.43 ≤ <i>p</i> ≤ -.06)	-.26	(-.44 ≤ <i>p</i> ≤ -.07)	-.30	(-.47 ≤ <i>p</i> ≤ -.11)				
General social desirability												
K	15.91	5.04	.25	(.06 ≤ <i>p</i> ≤ .43)	.31	(.12 ≤ <i>p</i> ≤ .48)	.33	(.14 ≤ <i>p</i> ≤ .49)				
S	25.80	10.10	.33	(.15 ≤ <i>p</i> ≤ .50)	.49	(.33 ≤ <i>p</i> ≤ .63)	.48	(.32 ≤ <i>p</i> ≤ .62)				
Mp2	10.32	3.85	.32	(.14 ≤ <i>p</i> ≤ .49)	.28	(.09 ≤ <i>p</i> ≤ .45)	.35	(.35 ≤ <i>p</i> ≤ .51)				
Wsd2	12.58	4.07	.27	(.08 ≤ <i>p</i> ≤ .44)	.22	(.02 ≤ <i>p</i> ≤ .39)	.28	(.28 ≤ <i>p</i> ≤ .45)				

Note: PDS = Paulhus Deception Scales; SDE(P) = PDS Self-Deceptive Enhancement; IM(P) = PDS Impression Management; CI = confidence interval; L = MMPI-2 Lie; K = MMPI-2 Defensiveness; EXCVIR = Sentence Completion Test-75 (SCT-75) Excessive Virtue; F = MMPI-2 Infrequency; Fb = MMPI-2 Back-Page Infrequency; ANGNEG = SCT-75 Angry Negativity; S = MMPI-2 Superlative; Mp2 = MMPI-2 Positive Malingering; Wsd2 = MMPI-2 Social Desirability.

Table 5
Concurrent Validity Correlates of the Self-Deceptive Enhancement (SDE) and the Impression Management (IM) Scales of the BIDR and the PDS for 103 Undergraduates (Group 4)

Concurrent Validity Scales	BIDR scales						PDS scales									
	SDE(B)			IM(B)			SDE(P)			IM(P)			Total (P)			
	<i>M</i>	<i>SD</i>	<i>r</i>	CI	<i>r</i>	CI	<i>r</i>	CI	<i>r</i>	CI	<i>r</i>	CI	<i>r</i>	CI	<i>r</i>	CI
Exaggeration of virtue																
L	3.52	2.2	.43	(.26 ≤ <i>p</i> ≤ .57)	.61	(.47 ≤ <i>p</i> ≤ .72)	.27	(.08 ≤ <i>p</i> ≤ .44)	.60	(.45 ≤ <i>p</i> ≤ .71)	.56	(.41 ≤ <i>p</i> ≤ .68)				
L+K	17.10	5.63	.54	(.38 ≤ <i>p</i> ≤ .66)	.63	(.50 ≤ <i>p</i> ≤ .74)	.41	(.24 ≤ <i>p</i> ≤ .56)	.58	(.44 ≤ <i>p</i> ≤ .70)	.62	(.49 ≤ <i>p</i> ≤ .73)				
EEV	15.18	5.32	.50	(.34 ≤ <i>p</i> ≤ .63)	.60	(.46 ≤ <i>p</i> ≤ .71)	.40	(.23 ≤ <i>p</i> ≤ .55)	.59	(.45 ≤ <i>p</i> ≤ .70)	.62	(.49 ≤ <i>p</i> ≤ .73)				
Exaggeration of good/poor adjustment																
So-r	28.41	6.99	.49	(.32 ≤ <i>p</i> ≤ .62)	.29	(.10 ≤ <i>p</i> ≤ .46)	.37	(.19 ≤ <i>p</i> ≤ .53)	.28	(.09 ≤ <i>p</i> ≤ .45)	.38	(.20 ≤ <i>p</i> ≤ .54)				
ESA	11.78	4.75	.54	(.38 ≤ <i>p</i> ≤ .66)	.38	(.21 ≤ <i>p</i> ≤ .54)	.43	(.26 ≤ <i>p</i> ≤ .57)	.40	(.22 ≤ <i>p</i> ≤ .55)	.50	(.34 ≤ <i>p</i> ≤ .63)				
F	6.86	5.43	-.24	(-.42 ≤ <i>p</i> ≤ (.05)	-.27	(-.44 ≤ <i>p</i> ≤ (.08)	-.24	(-.41 ≤ <i>p</i> ≤ (.05)	-.21	(-.39 ≤ <i>p</i> ≤ (.02)	-.27	(-.44 ≤ <i>p</i> ≤ (.08)				
At	18.53	9.84	-.56	(-.68 ≤ <i>p</i> ≤ (.41)	-.32	(-.49 ≤ <i>p</i> ≤ (.14)	-.37	(-.53 ≤ <i>p</i> ≤ (.19)	-.28	(-.45 ≤ <i>p</i> ≤ (.09)	-.38	(-.54 ≤ <i>p</i> ≤ (.20)				
General social desirability																
K	13.57	4.23	.48	(.32 ≤ <i>p</i> ≤ .62)	.51	(.35 ≤ <i>p</i> ≤ .64)	.41	(.23 ≤ <i>p</i> ≤ .56)	.45	(.29 ≤ <i>p</i> ≤ .60)	.53	(.37 ≤ <i>p</i> ≤ .65)				
De	10.83	2.61	.44	(.26 ≤ <i>p</i> ≤ .58)	.54	(.39 ≤ <i>p</i> ≤ .67)	.40	(.22 ≤ <i>p</i> ≤ .55)	.49	(.32 ≤ <i>p</i> ≤ .62)	.54	(.39 ≤ <i>p</i> ≤ .67)				

Note: BIDR = Balanced Inventory of Desirable Responding; PDS = Paulhus Deception Scales; SDE(B) = BIDR Self-Deceptive Enhancement; IM(B) = BIDR Impression Management; SDE(P) = PDS Self-Deceptive Enhancement; IM(P) = PDS Impression Management; CI = confidence interval; L = MMPI-2 Lie; K = MMPI-2 Defensiveness; EEV = PSI (Psychological Screening Inventory) Endorsement of Excessive Virtue; So-r = MMPI-2 Social Desirability-Revised; ESA = PSI Endorsement of Superior Adjustment; F = MMPI-2 Infrequency; At = MMPI-2 Manifest Anxiety; De = PSI Defensiveness.

Table 6
Concurrent Validity Correlates of the Self-Deceptive Enhancement (SDE) and Impression Management (IM) Scales of the BIDR and the PDS for 115 Undergraduates (Group 5)

Concurrent Validity Scales	BIDR scales						PDS scales								
	SDE(B)			IM(B)			SDE(P)			IM(P)			Total (P)		
	<i>M</i>	<i>SD</i>	<i>r</i>	CI	<i>r</i>	CI	<i>r</i>	CI	<i>r</i>	CI	<i>r</i>	CI	<i>r</i>	CI	
Exaggeration of virtue															
Gi	15.27	5.28	.42	(.25 ≤ <i>p</i> ≤ .56)	.56	(.43 ≤ <i>p</i> ≤ .68)	.27	(.10 ≤ <i>p</i> ≤ .44)	.53	(.38 ≤ <i>p</i> ≤ .65)	.55	(.41 ≤ <i>p</i> ≤ .66)			
MD	4.45	2.13	.37	(.20 ≤ <i>p</i> ≤ .51)	.42	(.26 ≤ <i>p</i> ≤ .56)	.26	(.08 ≤ <i>p</i> ≤ .43)	.43	(.27 ≤ <i>p</i> ≤ .57)	.47	(.31 ≤ <i>p</i> ≤ .60)			
Exaggeration of good/poor adjustment															
FB	2.34	1.71	-.10	(-.28 ≤ <i>p</i> ≤ .08)	-.19	(-.36 ≤ <i>p</i> ≤ -.01)	-.04	(-.22 ≤ <i>p</i> ≤ .15)	-.10	(-.28 ≤ <i>p</i> ≤ .08)	-.10	(-.28 ≤ <i>p</i> ≤ .08)			
General social desirability															
DY	11.83	2.75	.41	(.25 ≤ <i>p</i> ≤ .55)	.36	(.19 ≤ <i>p</i> ≤ .51)	.33	(.15 ≤ <i>p</i> ≤ .48)	.37	(.20 ≤ <i>p</i> ≤ .52)	.47	(.31 ≤ <i>p</i> ≤ .60)			

Note: BIDR = Balanced Inventory of Desirable Responding; PDS = Paulhus Deception Scales; SDE(B) = BIDR Self-Deceptive Enhancement; IM(B) = BIDR Impression Management; SDE(P) = PDS Self-Deceptive Enhancement; IM(P) = PDS Impression Management; CI = confidence interval; Gi = California Psychological Inventory (CPI) Good Impression; MD = Sixteen Personality Factors Questionnaire (16PF) Motivational Distortion; FB = 16PF Faking Bad; Dy = Personality Research Form (PRF) Desirability.

clients than undergraduates and more so for the BIDR than the PDS. (b) The two-factor, congeneric, orthogonal model utilized by Paulhus (1998) did not fit the data for either the BIDR or the PDS, for either forensic clients or undergraduates, even when the orthogonal constraint was relaxed. A tenable fit was obtained for both the BIDR and the PDS with a model that allowed all the items to cross-load and correlation between the factors. Results were similar for forensic clients and undergraduates, and despite relaxed constraints, the factor models provided evidence that scores on the SDE and IM scales provide sufficient discriminant validity. (c) Validity correlates showed good support for scores on the IM scale as a measure of exaggerated virtue for both the BIDR and the PDS and both types of participants. Scores on the SDE were shown to be satisfactory measures of exaggerated good/poor adjustment for both forensic clients and undergraduates, and more strongly for the BIDR than the PDS. As would be expected in view of correlations between SDE and IM, there was substantial overlap in the validity correlations for both exaggerated virtue and exaggerated good/poor adjustment. Validity correlates showed good support for both IM and SDE as measures of general social desirability for both forensic clients and undergraduates, though not as strongly for the PDS as for the BIDR. Comparisons of BIDR and PDS correlations showed higher validity correlates for SDE(B) than for SDE(P). The IM scores of the BIDR and PDS were reasonably equivalent; the SDE scales were less so.

The findings can be considered to confirm that scores on the IM scale are a satisfactory measure of the tendency to respond with exaggerated virtue, for both the BIDR and the PDS, both with forensic clients and with undergraduates. This finding is not surprising in view of the fact that the concept underlying the IM scale is relatively narrow and is readily defined (e.g., Strong et al., 2002). However, it must also be recognized that there is considerable overlap between the IM and SDE scales, and both are respectable measures of the traditional concepts of social desirability and general defensiveness. Scores on the SDE scale are shown to be a respectable measure of the exaggeration of good/poor adjustment, more so for the BIDR.

The findings also confirm that SDE scores are a stronger measure of the tendency to exaggerate good/poor adjustment, but weaker for SDE(P) than SDE(B). In regard to the generally lower validity correlates for SDE(P) than SDE(B), a likely reason involves the fact that the scoring system for SDE(P) results in a mean score of only 2.2 ($SD = 2.3$) for the general population (Paulhus, 1998). The mean and standard deviation for IM(P) are 6.7 and 4.0, respectively. This discrepancy likely results from the scoring practice of using only the extreme items rated 1 or 5 for SDE(P). Thus, the reliability of scores on the SDE(P) could be open to question. This concern is consistent with the data in the PDS *Manual* (Paulhus, 1998), which reported four internal consistency coefficients for the SDE(P) with a median of .72 and for the IM(P) with a median of .85. Stability coefficients are not reported, however.

The present data continued to reflect previous findings of statistically significant and noteworthy correlations between the SDE and IM scales for the BIDR and/or the PDS. Confirmatory factor analyses demonstrated some evidence for divergent validity with a minimally constrained two-factor model. There was substantial overlap in the validity correlates of both scales. Although this result can be explained in part by the difficulty of finding suitable concurrent validity measures for the SDE construct, it also suggests the presence of a general social desirability or favorable self-presentation factor underlying both constructs as assessed by scores on these two scales. Nevertheless, the ultimate issue is whether scores show divergent validity in the validity correlates. These data show overlap, although each is definitely a stronger measure of its own concept than is the other scale. Whether the two constructs could be more satisfactorily separated with different items is a question for future research.

In conclusion, the findings are supportive of the use of the BIDR/PDS in forensic contexts and allow the conclusion that research findings gathered with undergraduate populations can probably be generalized to forensic use. However, the greater overlap between the scales for the forensic data suggests that a single concept of favorable self-presentation is also salient here.

References

- Baer, R. A., Wetter, M. W., Nichols, D. S., Greene, R., & Berry, D. T. R. (1995). Sensitivity of MMPI-2 validity scales to under-reporting of symptoms. *Psychological Assessment*, *7*, 419-423.
- Block, J. (1965). *The challenge of response sets*. New York: Appleton-Century-Crofts.
- Borkenau, P., & Ostendorf, F. (1989). Descriptive consistency and social desirability in self- and peer reports. *European Journal of Personality*, *3*, 31-45.
- Butcher, J. N., & Han, K. (1995). Development of an MMPI-2 scale to assess the presentation of self in a superlative manner: The S scale. In J. N. Butcher & C. D. Spielberger (Eds.), *Advances in personality assessment* (Vol. 10, pp. 25-50). Mahwah, NJ: Lawrence Erlbaum.
- Cofer, C. N., Chance, J., & Judson, A. J. (1949). A study of malingering on the MMPI. *Journal of Psychology*, *27*, 491-499.
- Edwards, A. L. (1957). *The social desirability variable in personality assessment and research*. New York: Dryden.
- Gough, H. G. (1975). Some common misconceptions about neuroticism. *Journal of Consulting Psychology*, *18*, 287-292.
- Hemphill, J., & Howell, A. (2000). Adolescent offenders and stages of change. *Psychological Assessment*, *12*, 371-381.
- Hu, L., & Bentler, P. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*, 1-55.
- Jackson, D. N. (1974). *Personality Research Form: Manual* (Rev. ed.). Port Huron, MI: Research Psychologists Press.
- Kroner, D., & Weekes, J. (1996a). Balanced Inventory of Desirable Responding: Factor structure, reliability, and validity with an offender sample. *Personality and Individual Differences*, *21*, 323-333.
- Kroner, D., & Weekes, J. (1996b). Socially desirable responding and offense characteristics among rapists. *Violence and Victims*, *11*, 263-270.

- Lanyon, R. I. (1978). *Psychological Screening Inventory: Manual* (2nd ed.). Port Huron, MI: Research Psychologists Press.
- Lanyon, R. I. (1993). Development of scales to assess specific deception strategies on the Psychological Screening Inventory. *Psychological Assessment, 5*, 324-329.
- Lanyon, R. I. (2004). Favorable self-presentation on psychological inventories: An analysis. *American Journal of Forensic Psychology, 22*, 53-65.
- Lanyon, R. I., Almer, E. R., & Maxwell, B. M. (2002). Validation and norms for sentence completion task scales to assess misrepresentation during disability assessment. *American Journal of Forensic Psychology, 20*, 39-52.
- Lanyon, R. I., Dannenbaum, S. E., Wolf, L. L., & Brown, A. (1989). Dimensions of deceptive responding in criminal offenders. *Psychological Assessment, 1*, 300-304.
- Looman, J., Abracen, J., Maillet, G., & DiFazio, R. (1998). Phallometric nonresponding in sexual offenders. *Sexual Abuse: A Journal of Research and Treatment, 10*, 325-336.
- McNemar, Q. (1955). *Psychological statistics*. New York: John Wiley.
- Millsap, R. E. (2001). When trivial constraints are not trivial: The choice of uniqueness constraints in confirmatory factor analysis. *Structural Equation Modeling, 8*, 1-17.
- Muthén, B. O. (1984). A general structure equation model with dichotomous, ordered categorical and continuous latent variable indicators. *Psychometrika, 49*, 115-132.
- Muthén, L. K., & Muthén, B. O. (2004). *Mplus user's guide*. Los Angeles: Authors.
- Nichols, D. S., & Greene, R. L. (1989). *Adaptive or defensive: An evaluation of Paulhus's two-factor model of social desirability responding in the MMPI*. Unpublished manuscript, Dammasch State Hospital, Wilsonville, OR.
- Nichols, D. S., & Greene, R. L. (1997). Dimensions of deception in personality assessment: The example of the MMPI-2. *Journal of Personality Assessment, 67*, 251-266.
- Nugent, P., & Kroner, D. (1996). Denial, response styles, and admittance of offenses among child molesters and rapists. *Journal of Interpersonal Violence, 11*, 475-486.
- Paulhus, D. (1998). *Paulhus Deception Scales (PDS): The Balanced Inventory of Desirable Responding—7: User's manual*. North Tanawanda, NY: Multi-Health Systems.
- Paulhus, D. L. (1984). Two-component models of socially desirable responding. *Journal of Personality and Social Psychology, 46*, 598-609.
- Paulhus, D. L. (1986). Self-deception and impression management in tests responses. In A. Angleitner & J. S. Wiggins (Eds.), *Personality assessment via questionnaire* (pp. 143-165). New York: Springer.
- Paulhus, D. L. (1991). Balanced Inventory of Desirable Responding (BIDR). In J. P. Robinson, P. R. Shaver, & L. S. Wrightsman (Eds.), *Measures of personality and social psychological attitudes* (pp. 37-41). San Diego: Academic Press.
- Paulhus, D. L. (1994). *Balanced Inventory of Desirable Responding: Reference manual for the BIDR Version 6*. Unpublished manuscript, Department of Psychology, University of British Columbia at Vancouver.
- Sackeim, H., & Gur, R. (1978). Self-deception, self confrontation, and consciousness. In G. E. Schwartz & D. Shapiro (Eds.), *Consciousness and self-regulation: Advances in research* (Vol. 2, pp. 139-197). New York: Plenum.
- Salekin, R. (2000). Test review: The Paulhus Deception Scales. *American Psychology-Law Society News, 20*, 8-11.
- Seto, M., & Barbaree, H. (1993). Victim blame and sexual arousal to rape cues in rapists and nonoffenders. *Annals of Sex Research, 6*, 167-183.
- Strong, D. R., Greene, R. L., & Kordinak, S. T. (2002). Taxometric analysis of impression management and self-deception in college student and personnel evaluation settings. *Journal of Personality Assessment, 78*, 161-175.
- Taylor, J. (1953). A personality scale of manifest anxiety. *Journal of Abnormal and Social Psychology, 48*, 285-290.

- Timmons, L. A., Lanyon, R. I., Almer, E. R., & Curran, P. J. (1993). Development and validation of sentence completion indices of malingering during examination for disability. *American Journal of Forensic Psychology, 11*, 23-38.
- Vacha-Haase, T., Henson, R. K., & Caruso, J. (2002). Reliability generalization: Moving toward improved understanding and use of score reliability. *Educational and Psychological Measurement, 62*, 562-569.
- Wiggins, J. S. (1959). Interrelationships among MMPI measures of dissimulation under standard and social desirability instructions. *Journal of Consulting Psychology, 23*, 419-427.
- Winder, P., O'Dell, J., & Karson, S. (1975). New motivational distortion scales for the 16PF. *Journal of Personality Assessment, 39*, 532-537.