Validity of the MMPI-2-RF (restructured form) L-r and K-r scales in detecting underreporting in clinical and nonclinical samples. — Source link ↗

Martin Sellbom, R. Michael Bagby

Institutions: Kent State University, Centre for Addiction and Mental Health

Related papers:

- MMPI-2-RF characteristics of custody evaluation litigants.

- Examination of the MMPI-2 Restructured Form (MMPI-2-RF) Validity Scales in Civil Forensic Settings: Findings from Simulation and Known Group Samples

- MMPI-2 : Minnesota Multiphasic Personality Inventory-2 : manual for administration and scoring

- Detection of Overreported Psychopathology with the MMPI-2 RF Form Validity Scales.

- Utility of the MMPI-2-RF (Restructured Form) validity scales in detecting malingering in a criminal forensic setting: a known-groups design.

Validity of the MMPI-2-RF (Restructured Form) L–r and K–r Scales in Detecting Underreporting in Clinical and Nonclinical Samples

Martin Sellbom
Kent State University

R. Michael Bagby
Centre for Addiction and Mental Health
and University of Toronto

In the current investigation, the authors examined the validity of the L–r and K–r scales on the recently developed Minnesota Multiphasic Personality Inventory–2–Restructured Form (MMPI-2-RF; Y. S. Ben-Porath & A. Tellegen, in press) in measuring underreported response bias. Three archival samples previously collected for examining MMPI-2 validity scales were reanalyzed in 2 studies. In Study 1 L–r and K–r significantly differentiated 2 groups of participants (patients with schizophrenia and university students) who had been instructed to underreport on the MMPI-2 from participants who took the test under standard instructions. L–r and K–r also added incremental predictive variance to one another in differentiating these groups. In Study 2 a similar set of outcomes emerged through the use of a differential prevalence design in which L–r and K–r significantly differentiated a group of child custody litigants who were administered the MMPI-2 from university students taking the test under standard instructions.

Keywords: MMPI-2-RF, MMPI-2, underreporting, social desirability, validity scales

Test takers in a variety of assessment contexts may be motivated to minimize or deny behavioral and emotional difficulties or pathological personality traits on measures of psychopathology. These may include parents undergoing a child custody evaluation, who may not endorse problematic symptoms or undesirable personality traits in order to improve their chances of being awarded custody; job applicants who wish to highlight their interpersonal strengths and minimize their psychological problems; repeat offenders who wish to be reintegrated into society; and psychiatric inpatients who deny or otherwise minimize their symptoms in order to secure release from a hospital. This phenomenon has been variously labeled *faking good, defensive responding, socially desirable responding*, and *positive impression management*, among other terms. In the current investigation, we use the descriptive term *underreporting* to characterize this response style, as it does not imply that such responding is necessarily intentional (Ben-Porath, 2003).

The potential for respondents to underreport, and its effects on the validity of self-report measures of psychopathology, has led to the development and continual evaluation of validity scales designed to detect this test-taking approach. The Minnesota Multiphasic Personality Inventory (MMPI; Hathaway & McKinley, 1940) and its revised version, the Minnesota Multiphasic Person-ality Inventory–2 (MMPI-2; Butcher et al., 2001), have a history of measuring underreporting responses styles spanning nearly 7 decades. The MMPI-2 currently includes three standard scales that assess the underreporting of psychopathology: Lie (L), Correction (K), and Superlative (S). Several additional scales, such as the Edwards Social Desirability scale (Edwards, 1957) and the Wiggins Social Desirability scale (Wiggins, 1959), have also been investigated.

In meta-analyses of MMPI (Baer, Wetter, & Berry, 1992) and MMPI-2 (Baer & Miller, 2002) underreporting validity scales, average effect sizes (Cohen's d) were 1.05 and 1.25, respectively, suggesting that the MMPI-2 underreporting scales are effective in detecting this type of response bias. More specifically, Baer and Miller (2002) conducted a meta-analysis of 22 studies that examined underreporting on the MMPI-2 and reported large effect sizes for L (mean $d = 1.19$) and K (mean $d = 1.13$), as well as good classification accuracy rates with reasonable false positive rates (12%–20%) at a low base rate (.30). These findings are important to consider when evaluating the utility of revised versions of the L (L–r) and K (K–r) scales that appear on the restructured form of the MMPI-2.

The MMPI-2-RF (Restructured Form)

The MMPI-2-RF (Ben-Porath & Tellegen, in press) is a substantially shortened version of the MMPI-2 and is linked conceptually and empirically to modern theories and models of psychopathology and personality. The measure includes five sets of scales: Validity, Higher-Order (H-O); Restructured Clinical (RC); Specific Problem (SP); Interest; and Personality Psychopathology Five–Revised (PSY–5R) scales.

The current investigation focused on the validity of the MMPI-2-RF scales designed to measure underreporting: Uncommon Virtues (L–r) and Adjustment Validity (K–r). Tellegen and Ben-Porath (in press) developed these scales by factor analyzing the

items derived from the L, K, S, and Wiggins Social Desirability scales in several samples, including personnel selection, underreporting simulation, and clinical samples. They consistently found two primary factors in these analyses. Items that loaded distinctly on the respective factors (without substantial cross-loadings on the opposite factor) were selected for inclusion in two nonoverlapping scales. L–r consists of 14 items, 11 of which appear on the original L scale and 3 new items that originally appeared on the Wiggins Social Desirability scale. K–r consists of 14 items, all of which appear on the original K scale. Tellegen and Ben-Porath reported that both scales are highly correlated with their original MMPI-2 counterparts in a variety of samples. For instance, L and L–r are correlated .91 and .92 in a personnel selection and underreporting simulation sample, respectively. K and K–r are significantly correlated, .84 and .88, in these samples as well. However, L–r and K–r are correlated only .32 and .31 in these two samples. The current investigation was the first in which these scales were evaluated in samples outside the MMPI-2-RF technical manual.

The Current Investigation

The current investigation was designed to examine the validity of L–r and K–r in two studies (Study 1 and Study 2) by reanalyzing archival data from previously published research on the MMPI-2 validity scales (Bagby & Marshall, 2004; Bagby, Nicholson, Buis, Radovanovic, & Fidler, 1999; Bagby, Rogers, Buis, & Kalemba, 1994; Bagby et al., 1997). Study 1 was designed to examine L–r and K–r in an analogue simulation research design in which both clinical (patients with schizophrenia) and nonclinical samples had been instructed either to underreport or to take the MMPI-2 under standard (i.e., honest) instructions. We hypothesized that both L–r and K–r would differentiate between the underreporting and standard instruction conditions. We further hypothesized that L–r and K–r would add significant incremental predictive utility over each other due to low intercorrelations between them and theoretical differences in measuring uncommon virtues versus unlikely psychological adjustment.

Several researchers have noted that a potential limitation of analogue simulation designs concerns the generalizability of findings to actual individuals who distort their responses and have therefore argued for the use of designs that include samples where misleading responding is clearly established (known-group designs) or suspected to have occurred (differential prevalence designs; see, e.g., Rogers, 1997). To address this potential limitation, in Study 2 we added a differential prevalence design with a sample of divorced or separated parents involved in a family custody evaluation. We compared this condition to analogue simulation conditions in which participants had been instructed either to underreport within a child custody scenario or to take the test honestly. We hypothesized that L–r and K–r would significantly differentiate both analogue underreporting and differential prevalence conditions from the analogue standard instruction condition. As in Study 1, we expected L–r and K–r to add incremental predictive validity over each other.

Study 1

Two archival samples originally used in Bagby et al. (1997) to examine the MMPI-2 validity scales were utilized in Study 1. The first sample was composed of undergraduate university students asked to complete the MMPI-2 either with instructions to underreport or with standard (honest) instructions. Although the ability of well-adjusted university students to manipulate responses to appear more psychologically adjusted may be relevant to certain kinds of assessment situations (e.g., personnel selection), such a finding does not directly address the issue of how patients with psychiatric disorders perform the same task, and this is a critical question in applied settings (e.g., release from hospital). Thus, the second sample consisted of psychiatric patients diagnosed with schizophrenia who completed the MMPI-2 under either underreporting or standard instructions.

Method

Participants

The university sample was composed of 98 students (30 men, 68 women) recruited from undergraduate psychology courses. Invalid MMPI-2-RF profiles due to either nonresponding or inconsistent responding (i.e., CNS–r > 18 raw, VRIN–r or TRIN–r > 79T; Ben-Porath & Tellegen, in press) were excluded, resulting in a final sample of 29 men and 65 women with a mean age of 23.82 years ($SD = 5.59$). Although race and ethnicity were not formally assessed, most of these participants were White and born in Canada. The patient sample consisted of 99 patients (57 men and 42 women) diagnosed with schizophrenia. Each patient had been formally diagnosed through the Structured Clinical Interview for *DSM–III–R* Axis I Disorders (SCID–I/P; Spitzer, Williams, Gibbon, & First, 1988). They were chronic outpatients in the residual phase of the disorder who were being maintained on psychotropic medication. After MMPI-2-RF exclusionary criteria had been applied (see above), the final sample was composed of 50 male and 37 female participants with an average age of 39.49 years ($SD = 9.01$). Although race and ethnicity was not formally assessed, most of these participants were White and born in Canada (see Bagby et al., 1997, for more information about the samples).

Measure

The MMPI-2-RF (Ben-Porath & Tellegen, in press) is a 338-item self-report inventory on which participants answer "true" or "false" to a variety of statements indicating personality and psychopathology. The entire MMPI-2-RF item pool can be derived from the original MMPI-2, and the same normative sample is used with a few modifications (see Ben-Porath & Tellegen, in press). The MMPI-2-RF technical manual provides extensive reliability and validity data for this instrument (Tellegen & Ben-Porath, in press). In addition, Tellegen and Ben-Porath (in press) presented data indicating that the MMPI-2-RF scale scores that were derived from administration of the 567-item MMPI-2 booklet (used in this study) are interchangeable with results obtained from administration of the 338-item MMPI-2-RF booklet. For this study, we used the L–r and K–r validity scales. (See above for a description of their development and initial validation.)

Procedures

The university students were solicited from psychology courses and paid $25 for their participation. Each participant was randomly

assigned to complete the MMPI-2 under either underreporting or standard instructions. Student participants with underreporting instructions were asked to "produce a profile that is relatively free of psychological problems and symptoms" (Bagby et al., 1997, p. 408). They were asked to respond in a believable manner so that they could avoid detection by the "various 'lie' scales built into the questionnaire" (Bagby et al., 1997, p. 408).

The patient sample was derived from the Schizophrenia Research Registry maintained at the Clarke Institute of Psychiatry, Toronto, Canada (now the Centre for Addiction and Mental Health). Members in the registry had been diagnosed previously through the SCID–I/P and had given consent to be contacted for further research. The patients were contacted by telephone and given a brief description of the study; if they expressed an interest, a testing session was scheduled. Approximately two thirds of those patients contacted by telephone agreed to participate in the study. Patients were paid $25 for their participation. All patients completed the MMPI-2 either in small groups or alone, under the supervision of a research assistant. They were randomly assigned to complete the MMPI-2 under either standard or underreporting instructions. Patients with underreporting instructions were asked not to reveal their psychiatric symptoms. More specifically, they received the following instruction: "we would like you to answer the questions as if you were applying for employment and wanted to keep your history of psychiatric problems and symptoms private. In other words, your goal in answering the questions of this psychological test is to appear as well-adjusted as possible" (Bagby et al., 1997, p. 408).

Results and Discussion

RC Scales

Overall model. We first conducted a 2×2 multivariate analysis of variance (MANOVA) to determine any differences in clinical presentation as a function of response condition (i.e., standard instructions vs. underreporting) and group (i.e., schizophrenic patients vs. undergraduate students) with four planned comparisons. The RC scales served as the dependent variables. The overall model was statistically significant, Wilks's $\lambda = .561$, $F(9, 169) = 14.70$, $p < .001$, $\eta^2 = .44$, for response condition, and Wilks's $\lambda = .721$, $F(9, 169) = 7.25$, $p < .001$, $\eta^2 = .28$, for group. These main effects were qualified by a Response Condition $\times$ Group interaction, Wilks's $\lambda = .859$, $F(9, 169) = 3.09$, $p = .002$, $\eta^2 = .14$.

Planned comparisons. We conducted four planned comparisons as a follow-up to the overall model. First, we sought to determine whether there were any differences in clinical presentation between the patients with schizophrenia and students who took the MMPI-2 under standard instructions. This comparison was statistically significant, Hotelling's $T^2 = 0.882$, $F(9, 79) = 7.79$, $p < .001$, $\eta^2 = .44$. Follow-up analyses of variance (ANOVAs) indicated that the schizophrenia patients scored significantly higher on all scales except RC3, RC4, and RC9, $Fs(1, 81) = 10.10–25.35$, $ps < .002$, $\eta^2s = .10–.23$.[1] In our second and third comparisons, we examined whether students and patients were able to alter their clinical presentations when underreporting versus responding to standard instructions. Students were able to alter their RC profiles significantly, Hotelling's $T^2 = 0.770$, $F(9, 84) = 7.18$,

$p < .001$, $\eta^2 = .44$, across response sets. Follow-up univariate ANOVAs indicated that students who took the MMPI-2 with underreporting instructions scored significantly lower on all scales except RC6 and RC9 relative to those who took the test under standard instructions, $Fs(1, 92) = 13.15–32.69$, $ps < .001$, $\eta^2s = .13–.26$. Patients with schizophrenia were also able to alter their clinical presentations when underreporting, Hotelling's $T^2 = 1.16$, $F(9, 77) = 9.92$, $p < .001$, $\eta^2 = .54$. Patients who took the test with instructions to underreport scored significantly lower on all scales except RC3 and RC9 relative to those with standard instructions, $Fs(1, 85) = 14.63–72.90$, $ps < .001$, $\eta^2s = .15–.46$. Our final planned comparison examined differences between underreporting patients and students who took the test with standard instructions. The purpose was to determine whether the patients who underreported were able to produce MMPI-2 profiles that were indistinguishable from a "normal" sample. There were no overall multivariate differences in their RC scale profiles, Hotelling's $T^2 = 0.219$, $F(9, 80) = 1.95$, $p > .05$, $\eta^2 = .18$.

In sum, these analyses suggest that (a) the patients with schizophrenia presented with more psychopathology than the students, (b) both patient and student groups had significantly lower mean scores on the RC scales when underreporting relative to standard instructions, and (c) underreporting patients were able to present themselves similarly to a nonclinical sample of students.

Validity Scale Group Differences

Overall model. We conducted a 2×2 MANOVA to identify differences on L–r and K–r as a function of response condition (i.e., standard instructions vs. underreporting) and group (i.e., patients with schizophrenia vs. undergraduate students) with three planned comparisons. The overall model was significant only for a response condition main effect, Wilks's $\lambda = .673$, $F(2, 176) = 42.74$, $p < .001$, $\eta^2 = .33$, indicating that underreporting participants scored significantly higher on validity scales relative to those taking the test under standard instructions.

Planned comparisons. We followed up with ANOVAs for three planned comparisons: (a) students with standard instructions (SI) versus students with underreporting instructions (UI), (b) schizophrenia patients with SI versus schizophrenia patients with UI, and (c) schizophrenia patients with UI versus students with SI. This latter comparison was conducted because the patients with schizophrenia were trying to appear normal, and thus should be compared to normal (i.e., nonclinical) individuals as well. As summarized in Table 1, the results indicated that underreporting patients scored significantly higher on both L–r and K–r relative to both patients and students who took the test under standard instructions. Moreover, underreporting students scored higher on both L–r and K–r relative to students under standard instructions. Despite underreporting patients being able to present clinically as normal individuals on the substantive scales, the validity scales were able to differentiate them from a nonclinical student sample with large effects.

---

[1] We used a Bonferroni corrected alpha of .006 (.05/9 comparisons) to determine statistical significance for all univariate comparisons.

Table 1

Underreporting Versus Standard Instructions in Patient and Undergraduate Samples

	Patients		Undergraduates					
Scale	SI ($n = 43$)	UI ($n = 44$)	SI ($n = 46$)	UI ($n = 48$)	F	d_1	d_2	d_3
L–r	51.67$_a$ (10.78)	63.66$_b$ (14.75)	49.71$_a$ (9.57)	57.92$_b$ (15.11)	10.88***	0.93	0.65	1.13
K–r	46.60$_a$ (8.79)	57.81$_b$ (9.97)	46.54$_a$ (9.68)	59.42$_b$ (8.23)	26.35***	1.19	1.44	1.15

Note. Means with different subscripts are significantly different at $p < .05$. Values in parentheses represent standard deviations. SI = standard instructions; UI = underreporting instructions; d_1 = schizophrenia patients SI vs. UI; d_2 = undergraduate SI vs. UI; d_3 = undergraduate SI vs. schizophrenia UI.
*** $p < .001$.

Incremental Validity

We next conducted hierarchical logistic regression analyses to determine whether L–r and K–r would add significant incremental predictive utility to each other in differentiating underreporting and standard instruction profiles. We examined this issue using the three planned comparisons described earlier. Table 2 reports these analyses. L–r added significant incremental utility in differentiating underreporting patients from both other patients and students with standard instructions. Unexpectedly, L–r did not add significantly to K–r in differentiating undergraduate underreporters from those under standard instructions. Conversely, K–r added a substantial increment to L–r in each comparison.

Study 2

The second study was designed to replicate and extend the findings of Study 1. To generalize across designs and settings, we examined the scales in a second analogue simulation design and included a differential prevalence condition (i.e., a sample where underreporting was suspected to have occurred). A unique feature of this study was the use of a child custody context—a setting in which substantial underreporting on psychological testing occurs frequently (e.g., Bagby et al., 1999)—for both simulation and differential prevalence groups. We examined the validity of the L–r and K–r in this study through reanalyzing previously collected data for other investigations that studied the MMPI-2 (Bagby et al., 1994, 1999; see also Bagby & Marshall, 2004).

Method

Participants

Analogue simulation design. This sample consisted of 140 university students who were asked either to underreport or to take the MMPI-2 under standard instructions. Invalid MMPI-2-RF

Table 2

Hierarchical Logistic Regression Analysis Results for Predicting Underreporting Status

	Parameters							
Step	Model fit χ^2(df)	w'	χ^2 chg(df)	w' chg	Odds ratio[a]	SE[a]	R^2	ΔR^2
			Patients with schizophrenia SI vs. UI					
Step 1: L–r	17.40 (1)**	.45			1.06	.024	.18	
Step 2: K–r	32.35 (2)**	.61	14.95 (1)**	.41	1.11	.031	.31	.13
Step 1: K–r	26.19 (1)**	.55			1.11	.031	.26	
Step 2: L–r	32.35 (2)**	.61	6.16 (1)*	.27	1.06	.024	.31	.05
			Undergraduate SI vs. UI					
Step 1: L–r	9.69 (1)**	.32			1.03	.023	.10	
Step 2: K–r	39.50 (2)**	.65	29.82 (1)**	.56	1.15	.032	.34	.24
Step 1: K–r	37.82 (1)**	.63			1.15	.032	.33	
Step 2: L–r	39.50 (2)**	.65	1.68 (1)	.13	1.03	.023	.34	.01
			Undergraduate SI vs. Patients with schizophrenia UI					
Step 1: L–r	25.36 (1)**	.53			1.07	.025	.25	
Step 2: K–r	35.13 (2)**	.62	9.77 (1)**	.33	1.09	.029	.32	.07
Step 1: K–r	25.24 (1)**	.52			1.09	.029	.25	
Step 2: L–r	35.13 (2)**	.62	9.89 (1)**	.33	1.07	.025	.32	.07

Note. Cox and Snell R^2 estimation was used for logistic regression. w's of .10, .30, and .50 are small, medium, and large effect sizes, respectively (see Cohen, 1988). w' = effect size for nonparametric χ^2 statistic; w' chg = effect size for χ^2 chg statistic; SI = standard instructions; UI = underreporting instructions.
[a] From final model.
* $p < .05$. ** $p < .01$.

Table 3
Underreporting Versus Standard Instructions in Undergraduate and Custody Differential Prevalence Samples

	Undergraduates		Custody			
Scale	SI ($n = 67$)	UI ($n = 65$)	DPG ($n = 109$)	F	d_1	d_2
L–r	49.60$_a$ (9.81)	64.57$_b$ (17.68)	59.69$_b$ (12.11)	22.09***	1.05	0.89
K–r	47.70$_a$ (11.66)	58.77$_b$ (9.87)	56.12$_b$ (10.66)	20.60***	1.02	0.76

Note. Means with different subscripts are significantly different at $p < .05$. Values in parentheses represent standard deviations. SI = standard instructions; UI = underreporting instructions; DPG = differential prevalence group; d_1 = effect size for undergraduates SI vs. UI; d_2 = effect size for undergraduate SI vs. custody DPG.
*** $p < .001$.

profiles due to either nonresponding or inconsistent responding (i.e., CNS–r > 18 raw, VRIN–r or TRIN–r > 79T; Ben-Porath & Tellegen, in press) were excluded, resulting in a final sample of 29 men and 65 women with a mean age of 23.47 years ($SD = 6.23$). Although race and ethnicity were not formally assessed, most of these participants were White and Canadian born.

Custody sample. This sample consisted of 117 (57 men, 60 women) individuals being assessed for the purposes of child custody and access evaluations, and was used as a differential prevalence group for comparisons. After invalid MMPI-2-RF profiles based on noncontent responsiveness (see above) were excluded, the sample consisted of 56 men and 53 women with a mean age of 36.40 years ($SD = 8.71$). Bagby et al. (1999) provide a detailed description of this sample.

Measure

See the description of the MMPI-2-RF in Study 1.

Procedures

Analogue simulation design. The university students were solicited from psychology courses and randomly assigned to completing the MMPI-2 under either underreporting or standard instructions. Approximately half ($n = 67$) of the undergraduate participants received instructions to fake good and were provided with specific scenarios as examples of an assessment situation where such a possibility might exist, including gaining custody or access to their child. The remaining half ($n = 65$) were asked to take the MMPI-2 under standard instructions. Those in the underreporting and standard conditions were awarded $5 and $10, respectively, for their participation. As an additional incentive, participants in the underreporting condition were told that if they were not detected as faking, they would qualify for a lottery ($100), which was awarded at the completion of the data collection. However, they were not informed of the existence of scales to detect underreporting.

Custody sample. These individuals were administered the MMPI-2 as part of a psychological evaluation for child custody court evaluation. Of the participants, 31 were evaluated at a family court clinic, whereas 86 were evaluated at a private clinic. Bagby et al. (1999) reported no differences in age or gender of participants across the two clinics.

Results and Discussion

RC Scales

Overall model. Consistent with Study 1, we first examined whether the study groups displayed any differential clinical presentation across the three conditions. We conducted a one-way MANOVA in which we used the RC scales as dependent variables. The overall model was statistically significant, Wilks's $\lambda = .577$, $F(18, 460) = 8.09$, $p < .001$, $\eta^2 = .24$.

Planned comparisons. We had two planned comparisons. First, we examined differences between the underreporting group and the standard instruction group, and found that the two groups had significantly different profiles, Hotelling's $T^2 = 0.610$, $F(9, 122) = 8.27$, $p < .001$, $\eta^2 = .38$. Follow-up univariate tests revealed that underreporting individuals scored significantly lower on all scales, except RC6, $Fs(1, 130) = 8.50–45.41$, $ps < .004$, $\eta^2s = .06–.26$.[2] In the second comparison, we tested the difference between the differential prevalence and standard instruction groups, which was significant, Hotelling's $T^2 = 0.385$, $F(9, 166) = 7.10$, $p < .001$, $\eta^2 = .28$. Follow-up one-way ANOVAs revealed that individuals in the differential prevalence group scored significantly lower on all scales, except RC1, RC2, and RC6, $Fs(1, 174) = 13.34–37.62$, $ps < .001$, $\eta^2s = .07–.18$.

In sum, these analyses indicate that both the differential prevalence group and the student group asked to underreport had significantly lower mean scores on the RC scales relative to students taking the test under standard instructions.

Validity Scale Group Comparisons

We conducted a one-way three-group MANOVA to determine whether there were overall differences in underreporting across the three conditions. The overall model was statistically significant, Wilks's $\lambda = .795$, $F(4, 472) = 15.19$, $p < .001$, $\eta^2 = .11$. We followed up with ANOVAs and Tukey's HSD post-hoc tests for each individual validity scale (i.e., L–r and K–r) to test for differences between the three groups. These results are displayed in Table 3. As expected, both the analogue underreporting and differential prev-

---

[2] As with Study 1, we used a conservative alpha level of .006 to determine statistical significance for the univariate tests involving the RC scales.

Table 4
Hierarchical Logistic Regression Analysis Results for Predicting Underreporting Status

Step	Model fit $\chi^2(df)$	w'	χ^2 chg(df)	w' chg	Odds ratio[a]	SE[a]	R^2	ΔR^2
				Parameters				
			Undergraduates SI vs. UI					
Step 1: L–r	33.44 (1)**	.50			1.05	.052	.22	
Step 2: K–r	41.45 (2)**	.56	8.01 (1)**	.25	1.06	.062	.27	.05
Step 1: K–r	32.58 (1)**	.50			1.06	.062	.22	
Step 2: L–r	41.45 (2)**	.56	8.86 (1)**	.26	1.05	.052	.27	.05
			Undergraduates SI vs. Custody DPG					
Step 1: L–r	32.06 (1)**	.43			1.07	.070	.17	
Step 2: K–r	39.55 (2)**	.47	7.49 (1)**	.21	1.05	.044	.20	.03
Step 1: K–r	22.22 (1)**	.36			1.05	.044	.12	
Step 2: L–r	39.55 (2)**	.47	17.33 (1)**	.31	1.07	.070	.20	.08

Note. Cox and Snell R^2 estimation was used for logistic regression. w's of .10, .30, and .50 are small, medium, and large effect sizes, respectively (see Cohen, 1988). w' = effect size for nonparametric χ^2 statistic; w' chg = effect size for χ^2 chg statistic; SI = standard instructions; UI = underreporting instructions; DPG = differential prevalence group.
[a] From final model.
** $p < .01$.

alence groups had significantly higher mean scores on L–r and K–r relative to the analogue standard instruction group. The effect sizes for these differences as estimated by Cohen's d were large. The underreporting and differential prevalence groups did not differ on these scales, as hypothesized. These findings replicated those from Study 1. They also indicate that these two validity scales perform as expected in a group where underreporting is very prevalent (estimated 74%; see Bagby et al., 1999).

Incremental Validity

We next conducted hierarchical logistic regression analyses to determine whether L–r and K–r would add significant incremental predictive utility to each other in differentiating underreporting and standard instruction profiles. We examined this issue using the two planned comparisons (i.e., underreporting vs. standard instructions and differential prevalence group vs. standard instructions) reported earlier. As evident from Table 4, L–r and K–r added significant incremental utility to each other in each comparison, suggesting that both scales can be used in conjunction when identifying underreported profiles.

General Discussion

In the current investigation, our primary goal was to examine the validity of the revised L and K scales for the MMPI-2-RF through two studies that used different designed and research participants. The data for these two studies were derived from archival data sets, which had been published previously. The results from both studies indicate that L–r and K–r are able to differentiate between individuals instructed to underreport from those who responded to standard (honest) instructions to the test. These differences were almost always associated with large effect sizes. The effect sizes derived from the current study were mostly similar to those reported in Baer and Miller's (2002) meta-analysis of the MMPI-2 validity scales (i.e., mean ds = 1.19 and 1.13 for L and K,

respectively), which suggests that L–r and K–r are as effective in differentiating between underreported and honest profiles as their original counterparts.

There was one instance among the various comparisons in which L–r performed comparatively poorer in the context of the other outcomes. In Study 1 the effect size for L–r was substantially smaller ($d = .65$) in differentiating university student underreporters from students who took the test under standard instructions; in contrast the mean effect sizes for the other comparisons was $d = 1.00$. Moreover, unlike all other comparisons, in this instance it also failed to add incremental variance to K–r. One potential explanation for this outcome is that Study 1's underreporting instructions explicitly focus on the individual being symptom free, whereas Study 2's instructions and differential prevalence group status emphasize not just the individual but also relationships with others (through child custody). Thus, when relationships with others is of lesser concern, the L–r scale may play a smaller role, at least in nonclinical samples.[3]

The current findings support the view that those who underreport on the MMPI-2-RF are likely to alter their clinical presentation substantially relative to those taking the test under standard instructions. Indeed, across the comparisons there was a substantial decrease in profile elevation on most RC scales. This bolsters the argument that it is necessary to assess for potential underreporting bias, particularly in settings where such bias is suspected (e.g., personnel selection, child custody), as the interpretation of an underreported profile would suggest significantly less psychopathology when in fact that might not be the case.

The current results also have implications on the relative utility of L–r and K–r. The scores on both scales were associated with significant incremental validity relative to each other, which indicates that these scales could be used in conjunction when making

_____

[3] We thank an anonymous reviewer for suggesting this explanation.

decisions regarding underreporting. L–r and K–r were developed with different strategies regarding underreporting in mind, which appear consistent with Bagby and Marshall's (2004) factor analysis of MMPI-2 underreporting scales. They established two major dimensions of underreporting—impression management and self-deception—which have been clearly supported in the general social desirability literature (e.g., Paulhus, 2002). Indeed, Bagby and Marshall found that the original L scale adhered to the impression management strategy, whereas the original K score loaded on self-deception. Given that the revised version of L and K for the MMPI-2-RF are quite similar to their original counterparts, it is likely that these scales continue to mark these empirically identified dimensions of underreporting. Linking these scales to such response styles would have theoretical implications regarding their construct validity and interpretation. However, further research is necessary to more closely tie L–r and K–r to these particular response styles before firm conclusions are warranted.

There are several limitations that should be acknowledged. First, our clinical sample consisted of relatively stable outpatients with schizophrenia. A sample with more acutely ill participants might have generated different results, particularly in that these individuals would have been less able to underreport symptoms of psychopathology. Second, there were substantial differences in the student versus patient comparison in Study 1 in terms of age, gender, and socioeconomic distribution, which may have accounted for some of these differences. Finally, neither study employed a posttest questionnaire to determine whether the participants followed the directions. Although we performed manipulation checks by examining the symptomatic presentation on the RC scales for the various planned comparisons, which confirmed that the participants followed the directions as a group, future research should include such forms to ensure that all participants included in the analyses followed the directions.

Future research should examine the utility of the MMPI-2-RF validity scales in detecting underreporting in more naturalistic settings by using alternative methodological designs. Although the present study introduced a differential prevalence group, these are generally considered weaker than known-group designs simply because the underreporting status cannot be verified among those individuals. Perhaps the optimal design is the known-group design in which underreporting versus valid response groups are established on the basis of some external criterion and, preferably, are examined in a setting where underreporting research has the most utility (e.g., personnel selection, child custody, civil commitment release). Unfortunately, such designs are very difficult to follow. Unlike overreporting (or malingering) research in which structured interviews are available to form such groups (e.g., the Structured Interview for Reported Symptoms; Rogers, Bagby, & Dickens, 1992), such criteria are not readily available in this research paradigm. Another future direction involves further examination of what elevations on L–r and K–r mean and, in particular, what differential elevation patterns portray. We propose such interpretations in the theoretical and empirical framework of socially desirability responding (e.g., Paulhus, 2002), but these hypotheses require further validation. Finally, future studies should examine the optimal classification accuracy statistics for various cut scores. This will require multiple large samples for cross-validation,

which is why we were unable to conduct these analyses in the present investigation.

References

Baer, R. A., & Miller, J. (2002). Underreporting of psychopathology on the MMPI-2: A meta-analytic review. *Psychological Assessment, 14,* 16–26.

Baer, R. A., Wetter, M. W., & Berry, D. T. R. (1992). Detection of underreporting of psychopathology on the MMPI: A meta-analysis. *Clinical Psychology Review, 12,* 509–525.

Bagby, R. M., & Marshall, M. B. (2004). Assessing underreporting response bias on the MMPI-2. *Assessment, 11,* 115–126.

Bagby, R. M., Nicholson, R. A., Buis, T., Radovanovic, H., & Fidler, B. J. (1999). Defensive responding on the MMPI-2 in family custody evaluations. *Psychological Assessment, 11,* 24–28.

Bagby, R. M., Rogers, R., Buis, T., & Kalemba, V. (1994). Malingered and defensive response styles on the MMPI-2: An examination of validity scales. *Assessment, 1,* 31–38.

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 nonclinical samples. *Psychological Assessment, 9,* 406–413.

Ben-Porath, Y. S. (2003). Assessing personality and psychopathology with self-report inventories. In J. R. Graham & J. A. Naglieri (Eds.), *Handbook of psychology* (Vol. 10, pp. 553–578). New York: Wiley.

Ben-Porath, Y. S., & Tellegen, A. (in press). *Minnesota Multiphasic Personality Inventory–2 Restructured Form: Manual for administration, scoring, and interpretation.* Minneapolis: University of Minnesota Press.

Butcher, J. N., Graham, J. R., Ben-Porath, Y. S., Tellegen, A., Dahlstrom, W. G., & Kaemmer, B. (2001). *Minnesota Multiphasic Personality Inventory–2: Manual for administration and scoring* (rev. ed.). Minneapolis: University of Minnesota Press.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). New York: Academic Press.

Edwards, A. L. (1957). *The social desirability variable in personality assessment and research.* New York: Dryden.

Hathaway, S. R., & McKinley, J. C. (1940). A multiphasic personality schedule (Minnesota): I. Construction of the schedule. *Journal of Psychology, 10,* 249–254.

Paulhus, D. L. (2002). Socially desirable responding: The evolution of a construct. In H. I. Braun, D. N. Jackson, & D. E. Wiley (Eds.), *The role of constructs in psychological and educational measurement* (pp. 49–69). Mahwah, NJ: Erlbaum.

Rogers, R. (Ed.). (1997). *Clinical assessment of malingering and deception* (2nd ed.). New York: Guilford Press.

Rogers, R., Bagby, R. M., & Dickens, S. E. (1992). *Structured Interview of Reported Symptoms (SIRS): Professional manual.* Odessa, FL: Psychological Assessment Resources.

Spitzer, R. L., Williams, J. B. W., Gibbon, M., & First, M. B. (1988). *Structured Clinical Interview for DSM–III–R Axis I Disorders.* New York: New York State Psychiatric Institute.

Tellegen, A., & Ben-Porath, Y. S. (in press). *Minnesota Multiphasic Personality Inventory–2 Restructured Form: Technical manual.* Minneapolis: University of Minnesota Press.

Wiggins, J. S. (1959). Interrelationships among the MMPI measures of dissimulation under standard and social desirability instructions. *Journal of Consulting Psychology, 23,* 419–427.