JOURNAL OF PERSONALITY ASSESSMENT, *79*(2), 286–305

The Marlowe–Crowne Affair: Short Forms, Psychometric Structure, and Social Desirability

Steven D. Barger
Department of Psychology
Northern Arizona University

The Marlowe–Crowne Social Desirability scale (Crowne & Marlowe, 1960) is widely used to assess and control for response bias in self-report research. Several abbreviated versions of the Marlowe–Crowne scale have been proposed and adopted in psychology and medicine. In this article I evaluate the adequacy of 9 short forms using confirmatory factor analysis across 2 samples (combined $N = 867$). There was some evidence for the adequacy of different short forms, but model adequacy was not consistent across samples. Supplementary analyses revealed a multidimensional structure to the full Marlowe–Crowne scale and indicated that the apparent adequacy of model fit for some short forms might be a statistical artifact. Using the Marlowe–Crowne scale or its various short forms as a control for response bias is discouraged on empirical and conceptual grounds.

Personality assessment has long dealt with the possibility that individuals will respond to self-report instruments in a way that misrepresents their behavior. This distortion could be favorable or unfavorable, but among nonclinical populations, exaggerated virtues have received the most attention. Such social desirability in test behavior has led to the development of items and/or instruments that can measure such responses to discount or statistically adjust scores on measures of primary interest. Social desirability measures are also used for construct validation of personality instruments, in which small correlations with social desirability scales are provided as evidence for discriminant validity. Numerous scales have been developed, but the most enduring and popular of these is the 33-item Marlowe–Crowne Social Desirability (MC) scale (Crowne & Marlowe, 1960). Because of the putative importance of response bias, researchers have developed abbreviated subscales of this instrument. Such short forms reduce respondent burden and survey administration time but may or may not adequately capture the dimensions reflected in the full version of the scale. Using confirmatory factor analysis (CFA), this study evaluates the adequacy of nine Marlowe–Crowne scale short forms, and cross-validates these models in a second sample. The dimensional structure of the full scale is also explored.

DEVELOPMENT AND APPLICATION OF THE MARLOWE–CROWNE SCALE AND SHORT FORMS

The Marlowe–Crowne scale was developed to measure *social desirability*, defined as "the need of Ss to obtain approval by responding in a culturally appropriate and acceptable manner" (Crowne & Marlowe, 1960, p. 353). Crowne and Marlowe selected a pool of items from a number of personality inventories thought to represent this domain, with the caveat that item responses not reflect psychopathology or abnormality. This conceptualization distinguished the new scale from others whose content included maladjustment items and whose validation encompassed clinical diagnostic criteria (i.e., Social Desirability scale, Edwards, 1957). Items reflected approved behaviors but behaviors believed unlikely to occur (e.g., I never make a long trip without checking the safety of my car; Before voting I thoroughly investigate the qualifications of all the candidates). Thus, the primary rationale for this scale was to generate a social desirability measure that did not contain pathology-related content (Crowne & Marlowe, 1960).

An initial item pool was evaluated using respondents who were directed to answer questions in a socially desirable direction "from the point of view of college students" (Crowne & Marlowe, 1960, p. 350). Items with high agreement (90% or better) were retained and were subsequently judged to have less maladjustment content relative to the Edwards Social Desirability scale. These initial 47 items were administered to a sample (n = 76) of college students, and the 33 questions discriminating between high and low scorers at the .05 level were retained for the final version, named the Marlowe–Crowne Social Desirability scale. This set of items appeared to have good psychometric properties (internal consistency = 0.88, test–retest r = 0.89).

Convergent and discriminant validity for the new scale was evaluated by comparing correlations among the MC scale and Edwards's Social Desirability scale with various Minnesota Multiphasic Personality Inventory (MMPI; Hathaway & McKinley, 1943) scales. Edwards's scale appeared to tap neuroticism, as its correlation with anxiety and psychasthenia mirrored the average correlation among the scale's items (Crowne & Marlowe, 1960). On the other hand, the MC scale correlated less strongly with the clinical MMPI scales and more strongly with the MMPI validity scales. Crowne and Marlowe had apparently succeeded in their efforts to generate a measure free from psychopathological content.

The scale became widely adopted, and it continues to be popular today. A search of the Social Science Citation Index for the 1990s revealed 729 references to the original MC scale article. The scale was used predominantly to evaluate discriminant validity in personality assessment (Arroyo & Zigler, 1995; Campbell et al., 1996; Heatherton & Polivy, 1991; Helgeson, 1993; Zuckerman, Kuhlman, Joireman, Teta, & Kraft, 1993) or as a covariate to adjust for the potential influence of response bias (Hewitt & Flett, 1991; Pierce, Sarason, & Sarason, 1992; C. R. Snyder et al., 1991; Tangney et al., 1996). The MC scale has also been characterized as a measure of psychological defensiveness (Weinberger, 1990; Weinberger, Schwartz, & Davidson, 1979), but this review is limited to the social desirability applications of the MC scale.

Because of the scale's length, several abbreviated forms of the MC scale have been published (Hays, Hayashi, & Stewart,1989; Reynolds, 1982; Strahan & Gerbasi, 1972). In the 1990s, the Citation Index showed Reynolds's short forms cited in 128 studies, whereas Strahan and Gerbasi's forms were cited in 145 studies. Like the full scale, these short forms have been used as discriminant indexes when validating numerous constructs, including body image (Cash & Szymanski, 1995), interpersonal sense of control (Cook, 1993), measures of sexual motivation (Cooper, Shapiro, & Powers, 1998), and rumination (Scott & McIntosh, 1999); quality of life scales for breast cancer (Brady et al., 1997), lung cancer (Cella et al., 1995), brain cancer (Weitzner et al., 1995), and multiple sclerosis (Cella et al., 1996); and sexual behavior self-efficacy (Cecil & Pinkerton, 1998) and behavioral skill in negotiating condom use (Gordon, Carey, & Carey, 1997). These short forms are also used as self-report control variables in such diverse areas as psychiatric epidemiology (Dohrenwend et al., 1992), recovery from myocardial infarction (Frasure-Smith, Lespérance, Juneau, Talajic, & Bourassa, 1999), HIV progression (S. W. Cole, Kemeny, Taylor, Visscher, & Fahey, 1996), preventive behaviors for osteoporosis (Blalock et al., 1996), adjustment to chronic illness (Hatchett, Friend, Symister, & Wadhwa, 1997; Leake, Friend, & Wadhwa, 1999), health status among Persian Gulf War veterans (The Iowa Persian Gulf Study Group, 1997), and medical students' attitudes towards serving needy patients (Crandall, Volk, & Loemker, 1993). Strahan and Gerbasi's 10-item version has been evaluated by the Behavior Change Consortium of the National Institutes of Health (1997) and recommended for use in their health behavior change research. Thus, these scales are used to identify artifactual response bias in a wide range of fields, and their importance is underscored by their visibility in respected publications and institutions (e.g., *Health Psychology, Journal of the American Medical Association, Journal of Personality and Social Psychology, National Institutes of Health, Psychosomatic Medicine, Science*).

In sum, the MC scale has a historical and contemporary track record as a social desirability measure, and numerous short forms of the MC scale have been described and adopted in diverse literatures. Because of this continued popularity, it is especially important to evaluate the measurement properties of these scales. Several important issues regarding the measurement adequacy of the MC scale and its derivatives remain unresolved. The most important psychometric question is whether the full and short forms of the MC scale represent unitary constructs appropriate for their de facto applications. Previous factor analytic work suggests a multidimensional structure to the MC scale (Ballard, 1992; Crino, Svoboda, Rubenfeld, & White, 1983; Paulhus, 1984), which may render short versions inadequate representations of the full scale. In a similar vein, short forms of the MC scale may not be unidimensional. Internal consistency reliability and CFA provide two representations of scale dimensionality. Regarding the correlations among items, initial reports implied very good internal consistency for the full scale (α = .88; Crowne & Marlowe, 1960), but later reports were much lower (in the .70s; Ballard, 1992; Crino et al., 1983; Loo & Thorpe, 2000), a pattern also noted for the short forms (Ballard, 1992; Loo & Thorpe, 2000). Second, the adequacy of short forms can be more precisely evaluated with confirmatory, rather than exploratory, factor analytic techniques. CFA can test the adequacy of hypothesized models, rather than simply extracting latent factors from a data set. The full and short forms of the MC scale can be conceptualized as unitary subforms testable by CFA.

In general, CFA tests scale adequacy by analyzing item sets included in the full and abbreviated MC scale forms. Scale items are used as indicator variables and relations among the variables are hypothesized and evaluated. Generally this process includes first testing the null model, which hypothesizes no relations among the items, and then a model denoting that the item set represents a single latent factor (presumably social desirability). In addition, one can also compare model fit across different forms. Two studies using this strategy have been reported.

One analysis using 360 undergraduate volunteers (Fischer & Fick, 1993) found that all of the short forms had better fit than the full version and that Strahan and Gerbasi's (1972) Version X1 had the best fit. However, this study was unusual in that the internal consistencies of all scale versions were well above values reported previously (by .10 to .20), and the pattern of model fit was not replicated in a second report (Loo & Thorpe, 2000), albeit with a smaller sample (N = 232). In addition, both of these reports omitted a short form (i.e., Hays et al., 1989) utilized in the medical literature (e.g., Gritz et al., 1991). This study attempts to provide a more definitive psychometric evaluation of the various MC scale forms. Widely adopted versions of the MC scale will be evaluated using CFA, and model fit will be cross-validated with a second sample. The purpose of this study is to determine whether these popular scales represent the single dimension demanded by their application and by implication whether continued use of the MC scale and its short forms to measure social desirability is warranted.

METHOD

Participants

Undergraduate volunteers participated for course credit. In the first sample, 497 undergraduates participated in a mass testing session. Subsequent analyses are based on the 466 (232 women, 231 men, 3 did not report gender) who had complete MC data. Ages ranged from 17 to 45 years (*M* = 21 years, *Mdn* = 20 years). In the second sample, 431 participated in the mass testing session.[1] Subsequent analyses are based on the 401 (189 women, 207 men, 5 did not report gender) who completed all 33 MC items. Ages ranged from 17 to 46 years (*M* = 20 years, *Mdn* = 19 years).

Measurement Models

All of the models were tested using EQS (Multivariate Software, 2001). Modeling with structural equations involves several assumptions, multivariate normality being the most often violated assumption (Bentler & Dudgeon, 1996). Although the distribution for the full MC scale appears to be normal (see next), the degree of deviation from multivariate normality may vary as a function of the item subsets in the various short forms. Violating this assumption can render the maximum likelihood statistics inadequate for model evaluation (Hu, Bentler, & Kano, 1992). Therefore, to be conservative, corrected statistics (i.e., the Satorra–Bentler Scaled Statistic, known as the Robust Comparative Fit Index; RCFI; Satorra & Bentler, 1988) are reported. These statistics can identify maximum likelihood models accurately, even with distributional misspecification (Bentler & Dudgeon, 1996; Hu et al., 1992). They perform well in Monte Carlo experiments across various kinds of assumptive violations (Hu et al., 1992), and these corrected statistics are preferred to other distribution-free model estimation procedures (i.e., asymptotic distribution free; Chou, Bentler, & Satorra, 1991). The standardized root mean squared residual (RMR) is a complementary index of model fit and describes the average deviation

[1]The mass testing sessions took place the first week of the respective terms. The MC scale was included among other personality instruments (7 and 5 scales were included for Samples 1 and 2, respectively). The MC was the third scale in both sessions. Participants were told that the measures were to allow researchers to recruit individuals with certain psychological characteristics (e.g., those high in shyness) and that they could leave any questions blank they did not wish to answer. Given this context, an anonymous reviewer was concerned that the number of individuals with missing MC items might reflect a lack of motivation or care among respondents. This potential error variance was examined by comparing statistical information within the samples and across earlier work. First, the internal consistency reliabilities across the two samples are in line with other published work. Second, the scale structure revealed by the principal components analyses are also consistent, both across the two samples and with previously published work (see Results section). Thus, the psychometric consistency observed suggests that response error in this work is in line with expectations.

of the obtained matrix from the reproduced matrix. Average values less than 0.05 are considered desirable, although several individual residuals greater than 0.10 may suggest poor fit (D. A. Cole, 1987).

Model fit was evaluated using chi-square statistics, the RCFI, and the standardized RMR. Models are considered adequate with nonsignificant chi-squares, RCFIs > 0.90 and standardized RMRs less than 0.05. To allow comparison with other work, the Comparative Fit Index (CFI) is provided (cf. Bollen, 1990). The CFI evaluates model fit by comparing the hypothesized model to a more restrictive baseline model that specifies independence among indicators (items). Improvement over the baseline model is quantified from 0 to 1 with values greater than 0.90 indicating adequate fit (Bentler, 1990, 1992).

The full MC scale was administered, and short forms (see next) were extracted from the 33-item set. For seven of the nine short forms, as well as for the full scale, models were created with items as dependent variables and one latent factor as an independent variable (errors were also independent variables). For the version hypothesizing attribution and denial components of the MC scale (Ramanaiah, Schill, & Leung, 1977), two latent factors were modeled. Scaling for the latent factors was fixed by setting the variance to 1.0. Each of these hypothesized models was contrasted with its corresponding null model, which specifies no covariation among the variables (Bentler & Bonett, 1980). Raw data were analyzed, and parameters were modeled using maximum likelihood estimates.[2] No attempt was made to alter models to improve fit indexes.

Short Forms of the MC Scale

The most popular short forms (Reynolds, 1982; Strahan & Gerbasi, 1972) were originally created by identifying items that loaded strongly on the first latent factor of a principal components analysis. These first components accounted for 16% and 13% of the variance in total scores, respectively. Strahan and Gerbasi created two 10-item versions (designated X1 and X2) and combined these for a 20-item version (designated XX). The shorter forms had low internal consistency reliabilities (*M*s across 4 samples = 0.63 and 0.57, respectively), whereas the 20-item version was somewhat better (0.76). Uncorrected correlations between the full and abbreviated versions were "all in the .90's" (Strahan & Gerbasi, 1972, p. 192). Reynolds identified 3 new short forms (named A, B, and C) that contained 11, 12, and 13 items, respectively. Internal consistency reliabilities were better for these versions (from 0.74 to 0.76), and he concluded that these were reliable and valid short forms that could help researchers consider response biases in work using self-reports (Reynolds, 1982).

[2]Covariance matrixes are available on request.

In a slightly different approach, the MC scale was hypothesized to tap into attribution and denial dimensions (Millham as cited in Ramanaiah et al., 1977). These dimensions were denoted by the valence of the socially desirable response—socially desirable responses scored by a "true" response were classified as attribution, whereas socially desirable responses denoted by a "false" response were classified as denial. The scale was divided into 18 attribution and 15 denial items. No internal consistency reliabilities were reported for these subscales (Ramanaiah et al., 1977).

The last short form was developed by Hays et al. (1989) to meet the need of clinicians, especially physicians, for a social desirability measure that could be completed in under 1 min. They began with 10 of 11 items in Reynolds's (1982) Form A and altered the response format to a 5-point Likert-type metric ranging from 1 (*definitely true*) to 5 (*definitely false*). Responses were scored dichotomously, with extreme responses (i.e., 1 or 5) counted as socially desirable. They retained the 5 items with the highest item-total correlation with the 10-item subscale. The psychometric properties of this very short form appeared promising (internal consistencies from 0.66 to 0.68 across two samples; test–retest $r = 0.75$). In this study, however, items corresponding to Hays's short form were scored in the typical true–false format.

RESULTS

Descriptive Statistics for the Full MC Scale

Sample 1. The MC scale had adequate internal consistency ($\alpha = .73$), and the scores were normally distributed (skewness = –0.043, $p > .30$) with a mean of 17.2 (SD = 4.98). A significant gender difference in MC scale scores was observed, $t(461) = 1.97$, $p < .05$, but the size of this difference was small (women, $M = 17.7$, vs. men, $M = 16.7$; $r = 0.09$). Because the distribution of scores for each gender was similar, and because there were no differences observed in the second sample (see later), gender was collapsed in subsequent analyses.

Sample 2. Again, the MC scale had adequate internal consistency ($\alpha = .74$) and normality (skewness = 0.103, $p > .20$), with a mean of 16.2 ($SD = 4.96$). No gender difference in MC scale scores was observed, $t\,(394) = 0.61$, $p > .50$, (women, $M = 16.4$; men, $M = 16.1$).

Homogeneity of Short Forms

The internal consistency of the short forms was typically in the .60s ($Mdn = .62$) in both samples, with very poor reliability (αs =.44 to .45) observed for the five-item version (see Table 1). These statistics are generally consistent with previous research (Ballard, 1992; Loo & Thorpe, 2000; see Table 1).

CFAs

For all nine forms of the MC scale, null models were inadequate (chi-square *p*s < .001) indicating some degree of covariation among the items in each item set (see Table 2). Because the RMRs for all forms met or just exceeded the cutoff for model adequacy (all RMRs < .055), this index was not diagnostic and will not be emphasized. Similarly, the 20-item version (XX) and the attribution and denial model evaluated by Ramanaiah et al. (1977) performed poorly across both samples.

For Sample 1, most of the short forms performed poorly, having significant chi-square values and unacceptably low fit indexes (< .84). Only the 10-item Versions X1 and X2 approached an acceptable fit (RCFIs > .90), even though Version X1 had a significant chi-square. Thus, the two 10-item short forms may generously be described as having adequate (but not good) fit for a single latent factor model.

In Sample 2, Version X1 no longer had an adequate fit (significant chi-square and RCFI < .90), whereas its sibling X2 now had a significant chi-square despite an adequate fit (RCFI = .91). Form X2 could be called adequate, but given the low internal consistencies for these two versions (*Mdn* = .56, see Table 1), these scales are psychometrically undesirable. Similarly, Forms A and B (Reynolds, 1982) had adequate fit indexes (> .91) and RMR values in Sample 2, but their corresponding chi-squares were significant. Perhaps more important, these favorable indications of model fit were not consistent across samples. In contrast, the five-item version (Socially Desirable Response Set–5 [SDRS–5]; Hays et al., 1989) appeared to fit very well in Sample 2 (a nonsignificant chi-square, a low RMR, and a high fit index, RCFI = .989), but again, this model's adequacy is tempered by the poor fit observed in Sample 1 (significant chi-square, RCFI < .84).

Can any short forms be judged appropriate given these data? Only Hays et al.'s (1989) five-item version appeared worthy, but this apparent adequacy stands in sharp contrast to the poor fit observed in Sample 1. This inconsistency, coupled with very low internal consistency, suggests this short form is also inadequate. Similar inconsistencies among the other short forms requires caution in accepting any particular subscale as appropriate.

On the other hand, one could argue that for some short forms the previous classification as inadequate is equivocal. Given the limitations of CFA (e.g., McCrae, Zonderman, Costa, Bond, & Paunonen, 1996), how strongly do these data impugn the adequacy of MC scale short forms? For example, is it appropriate to discount the excellent fit of the five-item scale in Sample 2 on the basis of the correspondingly poor fit observed in Sample 1?

TABLE 1
Internal Consistency Reliabilities for the Marlowe–Crowne Scale and Short Forms

Scale Version	*No. of Items*	*Crowne & Marlowe*	*Strahan & Gerbasi*	*Reynolds*	*Crino et al.*	*Zook & Sipps*	*Hays et al.*	*Ballard*	*Fischer & Fick*	*Loo & Thorpe*	*Sample 1*	*Sample 2*	*Mdn*
Full MC scale	33	.88	.78	.82	.73			.75	.96	.72	.73	.74	.75
X1[a]	10		.64	.63				.50	.88	.52	.53	.56	.56
X2[a]	10		.62	.66				.54	.88	.42	.50	.56	.56
XX[a]	20		.81	.79				.71	.94	.68	.69	.72	.72
A[b]	11			.74				.64	.86	.59	.60	.63	.64
B[b]	12			.75				.67	.88	.61	.62	.64	.66
C[b]	13			.76		.74		.68	.89	.62	.62	.64	.68
Attribution[c]	18								.88	.56	.61	.59	.60
Denial[c]	15								.88	.63	.65	.64	.65
SDRS–5[d]	5						.68				.45	.44	.45

Note. Full references for the studies are in the Appendix. Where authors had multiple data sets in one publication, weighted means or medians were calculated. *Mdn* = median; MC = Marlowe–Crowne Social Desirability scale; SDRS–5 = Socially Desirable Response Set.

[a]Strahan and Gerbasi (1972). [b]Reynolds (1982). [c]Ramanaiah, Schill, and Leung (1977). [d]Hays, Hyaashi, and Stewart (1989) used a Likert-type response option for their scale, but scored it dichotomously.

TABLE 2
Confirmatory Factor Analytic Fit Indexes for the Marlowe–Crowne Scale and Its Short Forms

	Goodness of Fit Indexes													
	Sample 1[a]							*Sample 2*[b]						
	Null		*Model*					*Null*		*Model*				
Model	χ^2	*df*	*S–B* χ^2	*df*	*CFI*	*RCFI*	*RMR*	χ^2	*df*	*S–B* χ^2	*df*	*CFI*	*RCFI*	*RMR*
Full 33-item scale	1553**	528	845**	495	.650	.661	.054	1351**	528	667**	495	0.786	0.791	.051
X1[c]	245**	45	54.5*	35	.895	.904	.041	222**	45	54.4*	35	0.873	0.887	.046
X2[c]	187**	45	49.4	35	.900	.900	.039	219**	45	50.8*	35	0.911	0.911	.043
XX[c]	820**	190	321**	170	.750	.760	.052	773**	190	277**	170	0.808	0.815	.053
A[d]	352**	55	105**	44	.792	.800	.052	317**	55	65*	44	0.914	0.923	.044
B[d]	397**	66	117**	54	.806	.815	.051	362**	66	80*	54	0.906	0.915	.045
C[d]	447**	78	153**	65	.757	.767	.055	395**	78	107**	65	0.862	0.871	.049
Attribution and denial[e]	1553**	528	785**	494	.707	.718	.052	1351**	528	652**	494	0.804	0.808	.051
SDRS–5[f]	103**	10	20.6**	5	.819	.836	.051	69**	10	5.7	5	0.986	0.989	.027
Random 5-item measure[g]	57.6**	10	7.6	5	.947	.948	.030	50.2**	10	0.99	5	1.000	1.000	.011

Note. Adjusted and unadjusted chi-squares had similar p values. The two-factor model (attribution and denial; Ramanaiah, Schill, & Leung , 1977) allowed the factors to correlate. Null = no hypothesized relation among the items; model = one factor hypothesized for the items; S–B χ^2 = Satorra–Bentler scaled chi-square statistic; CFI = comparative fit index; RCFI = robust comparative fit index (adjusted for potential violations of distributional assumptions); RMR = standardized root mean squared residual; *df* = degrees of freedom.

[a]n = 466. [b]n = 401. [c]Strahan and Gerbasi (1972). [d]Reynolds (1982). [e]Ramanaiah et al. (1977). [f]Hays, Hyaashi, and Stewart (1989). [g]Created using items 9, 19, 20, 21, and 29.

$*p < .05$. $**p < .001$.

The apparent good fit of this scale may be a result of the fewer number of items. A five-item scale is inherently easier to model (vs. a model with 10 or 20 indicators) because of fewer parameters to estimate (Raykov, 1998; cf. Contrada & Jussim, 1992). Thus, the apparently adequate fit of any of these short forms (and their apparent improvement over the full item set, i.e., Fischer & Fick, 1993) may be an artifact of the fewer number of observed variables in the model. This rival hypothesis was tested in two ways. First, the number of items in a short form and its fit index were correlated across the two samples. There was a strong inverse relationship between the fit index and the number of items used in the form, $r(17) = -0.74$, $p = .001$. Second, if model fit is an artifact of a small number of observed variables, then *any* form that is shorter should perform better relative to longer versions. Therefore, a random five items of the MC scale were selected and evaluated as a short form. As shown in the last row of Table 2, this random subset convincingly outperformed all of the other versions, both in the fit criteria (RCFIs > .95) and, perhaps more importantly, in fit consistency across samples. The apparent adequacy of shorter forms, especially the five-item version, probably is an artifact of the less demanding covariance structure to be estimated in a model with fewer indicators.

Exploratory Analyses for Latent Variables in the MC Scale

Given the inadequacy of extant short forms and the potential for structural models to favor forms with fewer items, does the MC scale have an internal structure that would facilitate short form development? That is, are there large, cohesive latent factors defined by item sets that can be extracted? One approach to generating such forms would be to modify models using CFA, but such models should follow from theory or previous research. Because models hypothesized by existing literature were not adequate in the CFA, model modifications would necessarily be atheoretical. Therefore exploratory factor analyses were conducted to determine if any reasonable latent factor structure could be identified.

Sample 1. MC scale items were entered into a principal components analysis with varimax rotation. The first component accounted for 11% of the variance whereas the second and subsequent components accounted for 5% or less variance. Twelve components had eigenvalues greater than 1.0. These 12 components accounted for 53.8% of the total variability in scores (see Table 3).

Data from Sample 1 failed numerous criteria to extract component solutions (Hatcher, 1994): (a) eigenvalues descended smoothly (the change in value across components was graded; eigenvalues of 1.06, 1.03, 0.99, 0.96 were observed for the 11th through 14th components, respectively), making the Kaiser criterion misleading; (b) the “scree” test would suggest a one-component solution, but such a

Table 3
Principal Components Analyses of the Marlowe–Crowne Scale

	Sample 1[a]			*Sample 2*[b]		
Component	*Eigenvalue*	*VAF*	*Cumulative VAF*	*Eigenvalue*	*VAF*	*Cumulative VAF*
1	3.63	11.0	11.0	3.86	11.7	11.7
2	1.77	5.4	16.4	1.73	5.2	17.0
3	1.64	5.0	21.3	1.50	4.5	21.5
4	1.39	4.2	25.6	1.44	4.4	25.8
5	1.36	4.1	29.7	1.34	4.1	29.9
6	1.29	3.9	33.6	1.26	3.8	33.7
7	1.18	3.6	37.2	1.22	3.7	37.4
8	1.16	3.5	40.7	1.12	3.4	40.8
9	1.13	3.4	44.1	1.11	3.4	44.2
10	1.10	3.3	47.5	1.10	3.3	47.5
11	1.06	3.2	50.7	1.03	3.1	50.6
12	1.03	3.1	53.8	1.02	3.1	53.7

Note. VAF = percentage of variance accounted for by the component.
[a]$n = 466$. [b]$n = 401$.

solution would not explain 89% of the variance; and (c) a "variance accounted for" cutoff (such as 70%; Gorsuch, 1983) would require retaining 18 components for a 33-item scale. Finally, 41% of the residuals in the reproduced correlation matrix were greater than |.05|, which suggests the presence of additional components (Tabachnick & Fidell, 1989; cf. Crino et al., 1983). Therefore the 12-component solution would appear to be neither complete nor parsimonious.

Sample 2. The components analysis for Sample 2 was virtually identical to Sample 1. The first component accounted for 11.7% of the variance whereas the second and subsequent components accounted for 5% or less variance (see Table 3). Twelve components had eigenvalues > 1.0, but the change in magnitude across components was again graded (i.e., eigenvalues of 1.03, 1.02, 0.99, 0.94 for the 11th through 14th components, respectively). The 12 components accounted for 53.7% of the total variability in scores.[3]

As in Sample 1, these data failed various criteria to extract reasonable component solutions. Any of these solutions would fail the most important conceptual criteria, parsimony and interpretability. The heterogeneity in MC scale items across these two samples closely mirrors analyses reported previously (Ballard, 1992; Reynolds, 1982; Strahan & Gerbasi, 1972), strongly suggesting that the MC

[3]The same pattern of latent variables was observed using principal axis factoring, except that this method was more successful reproducing the correlation matrix (1% of the residuals greater than |.05|).

scale is multidimensional and that its structure precludes generating subscales that can characterize full scale scores. Because of the structural complexity of the MC scale, further short forms were not attempted.

DISCUSSION

The adequacy of several MC scale short forms was evaluated using CFA. All nine versions performed poorly, as evidenced by small fit indexes, significant chi-squares, inconsistency across samples, or all of these. Although several short forms could generously be called adequate, ancillary analyses suggest that model fit is a function of the number of items in the scale, not the true representation of one latent variable by the item subsets. Exploratory factor analysis revealed a heterogeneous structure for the MC scale, with large numbers of small-item clusters. This heterogeneity provided convergent evidence that any subset of items, regardless of structural model adequacy, will omit the bulk of variance in full scale scores. Therefore short forms of the MC scale appear to be inadequate as unidimensional proxies for the full MC scale.

To date, this is the largest confirmatory factor analytic study evaluating the structural adequacy of MC scale short forms, including a newer five-item scale. In contrast to previous reports (Fischer & Fick, 1993; Loo & Thorpe, 2000), this study did not consistently observe good model fit for the short forms. However, Versions X1 and X2 (both 10 items) were judged best in one study (Fischer & Fick, 1993), whereas Versions A and B (11 and 12 items) were judged best in the other (Loo & Thorpe, 2000). Thus, inconsistency across the two smaller studies was replicated in this research. However, it appears that acceptable model fit for MC scale short forms can be attributed to the less demanding estimation inherent with fewer indicator variables. Although those reports did not examine this question, correlations between the number of items and the fit index (AGFI) in the two prior studies are consistent with this hypothesis (rs = –0.97 and –0.85, respectively). Therefore, adequate fit reported for any MC scale short form is likely to be a statistical artifact.

It is acknowledged that the poor performance of the short forms may not generalize to populations other than college students. However, the MC scale was targeted toward and tested on college students, so these samples are consistent with those used in the scale's development. In addition, the five-item version (Hays et al., 1989) was not scored using the Likert-type response options used by the scale's authors. However, because this version was an abbreviated version of Reynolds's Form A, the structural limitations noted previously are unlikely to be overcome by a different response metric.

In practice, researchers continue to use the MC scale and its short forms to tap responses that may contaminate behavioral self-reports (Abreu & Gabarain, 2000;

Blalock et al., 1996; S. W. Cole et al., 1996; Cooper et al., 1998; Frasure-Smith et al., 1999; Gordon et al., 1997; Hatchett et al., 1997; Neighbors, O'Leary, & Labouvie, 1999; Scott & McIntosh, 1999; for an exception, see C. R. Snyder et al., 1991). The structural inadequacy of the short forms argues against this practice. Although the confirmatory models reject a unitary structure for the short forms, the following question is critical as to the appropriateness of this practice: Do MC scale scores actually represent contamination? Showing that the MC scale or short forms act as suppressor variables (Pedhazur, 1982) would provide such criterion validity. In the suppressor role, correlations between self-reports and objective (non-self-report) measures should be strengthened after statistically controlling for MC scale scores. To my knowledge, no such evidence exists for the short forms, and the full MC scale has seen few tests of this critical exercise. However, those studies that have included an objective, external (non-self-report) criterion fail to support an artifactual response conceptualization; controlling for MC scale scores either has no effect on or dilutes the relation between self-report and objective criteria.

For example, of 21 personality facets of the Neuroticism–Extraversion–Openness Inventory, partialing respondent MC scale scores from self-reports decreased correspondence between the respondent and spouse ratings on 16 facets, whereas the other 5 were unchanged (McCrae & Costa, 1983). In another study, partialing MC scale scores from self-reports decreased correspondence between raters and targets on 8 of 13 personality dimensions. In that study, only one correlation increased (by 0.01) whereas the other four were unchanged (Borkenau & Ostendorf, 1992). In a third study using both clinical and community samples, controlling for MC scale scores consistently reduced the correlation between self-reported and externally rated well-being (Kozma & Stones, 1987). Finally, using questions judged as high in social desirability, statistical adjustment provided no appreciable gain in predictive validity for behavioral observations across 50 comparisons using another personality inventory (the California Psychological Inventory; Dicken, 1963). Perhaps as important, social desirability was substantively related to the objective behavioral ratings in all but one of the six samples (Dicken, 1963). Thus, no evidence for suppression effects were found using ratings from spouses (Kozma & Stones, 1987; McCrae & Costa, 1983), acquaintances or relatives (Borkenau & Ostendorf, 1992), teachers and trained personality raters (Dicken, 1963), and psychiatric staff (Kozma & Stones, 1987; for a similar criticism of other validity scales, see Piedmont, McCrae, Riemann, & Angleitner, 2000; for an excellent evaluation of the MC scale in personnel selection, see Ones, Viswesvaran, & Reiss, 1996).

Directly manipulating the social demands of a situation provides another test of the substance versus style characterization. Weinberger and Davidson (1994) evaluated whether high MC individuals (who were also low in trait anxiety) were a "subgroup of impression managers" (p. 590) by comparing them to individuals who reported altering their self-presentations to avoid social disapproval (identified by a scale that included other-directed items from the Self-Monitoring scale; M. Snyder, 1974). Participants were instructed to be expressive or inhibited about personal limitations while completing a phrase association task. Although impression managers were able to meet the experimental demands, high MC individuals were inhibited regardless of condition. Thus, high MC individuals did not change their behavior to meet the social demands, whereas the impression managers did (see also Arkin & Lake, 1983). This supports a substantive rather than artifactual interpretation of MC scale scores, and it implies that statistical overlap between the MC scale and impression management scales (i.e., Paulhus, 1993) is unlikely to represent important response bias contamination (see Moorman & Podsakoff, 1992).

Other descriptive evidence supports a substantive interpretation of MC scale scores. For example, individuals scoring high on the MC scale are described by friends as precisely the rational, self-controlled people who would do the kinds of behaviors reflected in the MC items (Strahan & Strahan as cited in Weinberger, 1990). In addition, MC scale scores, alone or in combination with other measures, predict outcomes such as hypertension (Mann & James, 1998), cortisol levels (Brody et al., 2000), cholesterol (Niaura, Herbert, McMahon, & Sommerville, 1992), autonomic nervous system reactivity (Barger, Kircher, & Croyle, 1997), lifetime psychiatric morbidity (Lane, Merikangas, Schwartz, Huang, & Prusoff, 1990), and mortality following a cardiac event (Denollet, 1999). Such relations would not be expected from a response bias scale.[4]

Thus, research with the methodological leverage to determine whether the MC measures a response bias supports a substantive interpretation of the scale. Although this research also lacks this type of leverage, previous work, combined with the psychometric performance of the MC forms noted in this study, lead to two conclusions. First, short form or long, the MC scale does not measure one dimension of personality. Second, available evidence suggests that the dimensions captured by the MC scale are not response biases in need of adjustment (see also Weinberger, 1990).

Crowne and Marlowe (1964) recognized the substantive nature of their scale nearly 40 years ago, and rigorous tests of the scale's appellation have shown "social desirability" to be a misnomer (Borkenau & Ostendorf, 1992; Kozma & Stones, 1987; McCrae & Costa, 1983; Weinberger & Davidson, 1994). However, these observations have been largely ignored, as evidenced by the continued use of MC scale forms as proxies for socially desirable response sets. McCrae and Costa sagely observed that "conceptual habits die hard, and it is difficult to view the cor-

[4]In light of the limitations discussed here, researchers interested in the substantive dimensions captured by the Marlowe–Crowne scale (Crowne & Marlowe, 1964) might also consider identifying or developing alternate measures with better psychometric properties.

relation of a SD measure with a trait scale as anything but an indictment of that scale" (p. 887). It is hoped that this work will bring further scrutiny to this practice for the MC scale.

ACKNOWLEDGMENTS

This research was supported in part by a Northern Arizona University Organized Research Grant. I thank Virginia Blankenship and Heidi Wayment for their helpful feedback on earlier versions of this manuscript.

REFERENCES

Abreu, J. M., & Gabarain, G. (2000). Social desirability and Mexican American counselor preferences: Statistical control for a potential confound. *Journal of Counseling Psychology, 47,* 165–176.

Arkin, R. M., & Lake, E. A. (1983). Plumbing the depths of the bogus pipeline: A reprise. *Journal of Research in Personality, 17,* 81–88.

Arroyo, C. G., & Zigler, E. (1995). Racial identity, academic achievement, and the psychological well-being of economically disadvantaged adolescents. *Journal of Personality and Social Psychology, 69,* 903–914.

Ballard, R. (1992). Short forms of the Marlowe–Crowne social desirability scale. *Psychological Reports, 71,* 1155–1160.

Barger, S. D., Kircher, J. C., & Croyle, R. T. (1997). The effects of social context and defensiveness on the physiological responses of repressive copers. *Journal of Personality and Social Psychology, 73,* 1118–1128.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107,* 238–246.

Bentler, P. M. (1992). On the fit of models to covariances and methodology to the *Bulletin*. *Psychological Bulletin, 112,* 400–404.

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin, 88,* 588–606.

Bentler, P. M., & Dudgeon, P. (1996). Covariance structure analysis: Statistical practice, theory, and directions. *Annual Review of Psychology, 47,* 563–592.

Blalock, S. J., Devellis, R. F., Giorgino, K. B., Devellis, B. M., Gold, D. T., Dooley, M. A. F., et al. (1996). Osteoporosis prevention in premenopausal women: Using a stage model approach to examine the predictors of behavior. *Health Psychology, 15,* 84–93.

Bollen, K. A. (1990). Overall fit in covariance structure models: Two types of sample size effects. *Psychological Bulletin, 107,* 256–259.

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

Brady, M. J., Cella, D. F., Mo, F., Bonomi, A. E., Tulsky, D. S., Lloyd, S. R., et al. (1997). Reliability and validity of the Functional Assessment of Cancer Therapy–Breast quality-of-life instrument. *Journal of Clinical Oncology, 15,* 974–986.

Brody, S., Wagner, D., Heinrichs, M., James, A., Hellhammer, D., & Ehlert, U. (2000). Social desirability scores are associated with higher morning cortisol levels in firefighters. *Journal of Psychosomatic Research, 49,* 227–228.

Campbell, J. D., Trapnell, P. D., Heine, S. J., Katz, I. M., Lavallee, L. F., & Lehman, D. R. (1996). Self-concept clarity: Measurement, personality correlates, and cultural boundaries. *Journal of Personality and Social Psychology, 70,* 141–156.

Cash, T. F., & Szymanski, M. L. (1995). The development and validation of the body-image ideals questionnaire. *Journal of Personality Assessment, 64,* 466–477.

Cecil, H., & Pinkerton, S. D. (1998). Reliability and validity of a self-efficacy instrument for protective sexual behaviors. *Journal of American College Health, 47,* 113–121.

Cella, D. F., Bonomi, A. E., Lloyd, S. R., Tulsky, D. S., Kaplan, E., & Bonomi, P. (1995). Reliability and validity of the Functional Assessment of Cancer Therapy–Lung (FACT–L) quality of life instrument. *Lung Cancer, 12,* 199–220.

Cella, D. F., Dineen, K., Arnason, B., Reder, A., Webster, K. A., Karabatsos, G., et al. (1996). Validation of the functional assessment of multiple sclerosis quality of life instrument. *Neurology, 47,* 129–139.

Chou, C. P., Bentler, P. M., & Satorra, A. (1991). Scaled test statistics and robust standard errors for data in covariance structure analysis: A Monte Carlo study. *British Journal of Mathematical and Statistical Psychology, 44,* 347–357.

Cole, D. A. (1987). Utility of confirmatory factor analysis in test validation research. *Journal of Consulting and Clinical Psychology, 55,* 584–594.

Cole, S. W., Kemeny, M. E., Taylor, S. E., Visscher, B. R., & Fahey, J. (1996). Accelerated course of human immunodeficiency virus infection in gay men who conceal their homosexual identity. *Psychosomatic Medicine, 58,* 219–231.

Contrada, R. J., & Jussim, L. (1992). What does the Cook–Medley Hostility Scale measure? In search of an adequate measurement model. *Journal of Applied Social Psychology, 22,* 615–627.

Cook, W. L. (1993). Interdependence and the interpersonal sense of control: An analysis of family relationships. *Journal of Personality and Social Psychology, 64,* 587–601.

Cooper, M. L., Shapiro, C. M., & Powers, A. M. (1998). Motivations for sex and risky sexual behavior among adolescents and young adults: A functional perspective. *Journal of Personality and Social Psychology, 75,* 1528–1558.

Crandall, S. J. S., Volk, R. J., & Loemker, V. (1993). Medical students' attitudes toward providing care for the underserved. Are we training socially responsible physicians? *Journal of the American Medical Association, 269,* 2519–2523.

Crino, M. D., Svoboda, M., Rubenfeld, S., & White, M. C. (1983). Data on the Marlowe–Crowne and Edwards Social Desirability scales. *Psychological Reports, 53,* 963–968.

Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology, 24,* 349–354.

Crowne, D. P., & Marlowe, D. (1964). *The approval motive: Studies in evaluative dependence*. New York: Wiley.

Denollet, J. (1999, March). *Repressive coping predicts mortality in patients with coronary heart disease*. Paper presented at the 57th meeting of the American Psychosomatic Society, Vancouver, British Columbia, Canada.

Dicken, C. (1963). Good impression, social desirability, and acquiescence as suppressor variables. *Journal of Educational and Psychological Measurement, 4,* 699–720.

Dohrenwend, B. P., Levav, I., Shrout, P. E., Schwartz, S., Naveh, G., Link, B. G., et al. (1992). Socioeconomic status and psychiatric disorders: The causation-selection issue. *Science, 255,* 946–952.

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

Fischer, D. G., & Fick, C. (1993). Measuring social desirability: Short forms of the Marlowe–Crowne Social Desirability scale. *Educational and Psychological Measurement, 53,* 417–424.

Frasure-Smith, N., Lespérance, F., Juneau, M., Talajic, M., & Bourassa, M. G. (1999). Gender, depression, and one-year prognosis after myocardial infarction. *Psychosomatic Medicine, 61,* 26–37

Gordon, C. M., Carey, M. P., & Carey, K. B. (1997). Effects of a drinking event on behavioral skills and condom attitudes in men: Implications for HIV risk from a controlled experiment. *Health Psychology, 16,* 490–495.

Gorsuch, R. L. (1983). *Factor analysis* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Gritz, E. R., Carr, C. R., Rapkin, D. A., Chang, C., Beumer, J., & Ward, P. H. (1991). A smoking cessation intervention for head and neck cancer patients: Trial design, patient accrual, and characteristics. *Cancer Epidemiology, Biomarkers, and Prevention, 1,* 67–73.

Hatcher, L. (1994). *A step-by-step approach to using the SAS system for factor analysis and structural equation modeling*. Cary, NC: SAS.

Hatchett, L., Friend, R., Symister, P., & Wadhwa, N. (1997). Interpersonal expectations, social support, and adjustment to chronic illness. *Journal of Personality and Social Psychology, 73,* 560–573.

Hathaway, S. R., & McKinley, J. C. (1943). *The Minnesota Multiphasic Personality Inventory.* Minneapolis: University of Minnesota Press.

Hays, R. D., Hayashi, T., & Stewart, A. L. (1989). A five-item measure of socially desirable response set. *Educational and Psychological Measurement, 49,* 629–636.

Heatherton, T. F., & Polivy, J. (1991). Development and validation of a scale for measuring state self-esteem. *Journal of Personality and Social Psychology, 60,* 895–910.

Helgeson, V. S. (1993). Implications of agency and communion for patient and spouse adjustment to a first coronary event. *Journal of Personality and Social Psychology, 64,* 807–816.

Hewitt, P. L., & Flett, G. L. (1991). Perfectionism in the self and social contexts: Conceptualization, assessment, and association with psychopathology. *Journal of Personality and Social Psychology, 60,* 456–470.

Hu, L.-T., Bentler, P., & Kano, Y. (1992). Can test statistics in covariance structure analysis be trusted? *Psychological Bulletin, 112,* 351–362.

The Iowa Persian Gulf Study Group. (1997). Self-reported illness and health status among Gulf War Veterans. *Journal of the American Medical Association, 277,* 238–245.

Kozma, A., & Stones, M. J. (1987). Social desirability in measures of subjective well-being: A systematic evaluation. *Journal of Gerontology, 42,* 56–59.

Lane, R. D., Merikangas, K. R., Schwartz, G. E., Huang, S. S., & Prusoff, B. A. (1990). Inverse relationship between defensiveness and lifetime prevalence of psychiatric disorder. *American Journal of Psychiatry, 147,* 573–578.

Leake, R., Friend, R., & Wadhwa, N. (1999). Improving adjustment to chronic illness through strategic self-presentation: An experimental study on a renal dialysis unit. *Health Psychology, 18,* 54–62.

Loo, R., & Thorpe, K. (2000). Confirmatory factor analyses of the full and short versions of the Marlowe–Crowne Social Desirability scale. *Journal of Social Psychology, 140,* 628–635.

Mann, S. J., & James, G. D. (1998). Defensiveness and essential hypertension. *Journal of Psychosomatic Research, 45,* 139–148.

McCrae, R. R., & Costa, P. T. (1983). Social desirability scales: More substance than style. *Journal of Consulting and Clinical Psychology, 51,* 882–888.

McCrae, R. R., Zonderman, A. B., Costa, P. T., Bond, M. H., & Paunonen, S. V. (1996). Evaluating replicability of factors in the revised NEO Personality Inventory: Confirmatory factor analysis versus Procrustes rotation. *Journal of Personality and Social Psychology, 70,* 552–566.

Moorman, R. H., & Podsakoff, P. M. (1992). A meta-analytic review and empirical test of the potential confounding effects of social desirability response sets in organizational behavior research. *Journal of Occupational and Organizational Psychology, 65,* 131–149.

Multivariate Software. (2001). EQS (Version 5.8 for the Power Macintosh) [Computer software]. Encino, CA: Author.

National Institutes of Health. (1997). *Behavior Change Consortium-Recommended Nutrition Measures* [online]. Retrieved November 20, 2001, from http://www1.od.nih.gov/behaviorchange/measures/nutrition.htm

Neighbors, C. J., O'Leary, A., & Labouvie, E. (1999). Domestically violent and nonviolent male inmates' responses to their partners' requests for condom use: Testing a social-information processing model. *Health Psychology, 18,* 427–431.

Niaura, R., Herbert, P. N., McMahon, N., & Sommerville, L. (1992). Repressive coping and blood lipids in men and women. *Psychosomatic Medicine, 54,* 698–706.

Ones, D. S., Viswesvaran, C., & Reiss, A. D. (1996). Role of social desirability in personality testing: The red herring. *Journal of Applied Psychology, 81,* 660–679.

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

Paulhus, D. L. (1993). *The Balanced Inventory of Desirable Responding* [Reference manual, Version 6]. Vancouver, British Columbia, Canada: University of British Columbia.

Pedhazur, E. J. (1982). *Multiple regression in behavioral research* (2nd ed.). Chicago: Holt, Reinhart, & Winston.

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.

Pierce, G. R., Sarason, B. G., & Sarason, I. G. (1992). General and specific support expectations and stress as predictors of perceived supportiveness: An experimental study. *Journal of Personality and Social Psychology, 63,* 297–307.

Ramanaiah, N. V., Schill, T., & Leung, L. S. (1977). A test of the hypothesis about the two-dimensional nature of the Marlowe–Crowne Social Desirability scale. *Journal of Research in Personality, 11,* 251–259.

Raykov, T. (1998). On the use of confirmatory factor analysis in personality research. *Personality and Individual Differences, 24,* 291–293.

Reynolds, W. M. (1982). Development of reliable and valid short forms of the Marlowe–Crowne Social Desirability scale. *Journal of Clinical Psychology, 38,* 119–125.

Satorra, A., & Bentler, P. M. (1988). Scaling corrections for chi-square statistics in covariance structure analysis. *Proceedings of the Business and Economics Statistics Section, American Statistical Association,* 308–313.

Scott, V. B., & McIntosh, W. E. (1999). The development of a trait measure of ruminative thought. *Personality and Individual Differences,15,* 1045–1056.

Snyder, M. (1974). Self-monitoring of expressive behavior. *Journal of Personality and Social Psychology, 30,* 526–537.

Snyder, C. R., Harris, C., Anderson, J. R., Holleran, S. A., Irving, L. M., Sigmon, S. T., et al. (1991). The will and the ways: Development and validation of an individual-differences measure of hope. *Journal of Personality and Social Psychology, 60,* 570–585.

Strahan, R., & Gerbasi, K. C. (1972). Short, homogenous versions of the Marlow [*sic*]-Crowne Social Desirability scale. *Journal of Clinical Psychology, 28,* 191–193.

Tabachnick, B. G., & Fidell, L. S. (1989). *Using multivariate statistics* (2nd ed.). New York: HarperCollins.

Tangney, J. P., Hill-Barlow, D., Wagner, P. E., Marschall, D. E., Borenstein, J. K., Sanftner, J., et al. (1996). Assessing individual differences in constructive versus destructive responses to anger across the lifespan. *Journal of Personality and Social Psychology, 70,* 780–796.

Weinberger, D. A. (1990). The construct validity of the repressive coping style. In J. L. Singer (Ed.), *Repression and dissociation* (pp. 337–386). Chicago: University of Chicago Press.

Weinberger, D. A., & Davidson, M. N. (1994). Styles of inhibiting emotional expression: Distinguishing repressive coping from impression management. *Journal of Personality, 62,* 587–613.

Weinberger, D. A., Schwartz, G. E., & Davidson, R. J. (1979). Low anxious, high anxious, and repressive coping styles: Psychometric patterns and behavioral and physiological responses to stress. *Journal of Abnormal Psychology, 58,* 369–380.

Weitzner, M. A., Meyers, C. A., Gelke, C. K., Byrne, K. S., Cella, D. F., & Levin, V. A. (1995). The Functional Assessment of Cancer Therapy (FACT) scale. Development of a brain subscale and revalidation of the general version (FACT–G) in patients with primary brain tumors. *Cancer, 75,* 1151–1161.

Zuckerman, M., Kuhlman, D. M., Joireman, J., Teta, P., & Kraft, M. (1993). A comparison of three structural models for personality: The big three, the big five, and the alternative five. *Journal of Personality and Social Psychology, 65,* 757–768.

APPENDIX

Ballard, R. (1992). Short forms of the Marlowe–Crowne social desirability scale. *Psychological Reports, 71,* 1155–1160.

Crino, M. D., Svoboda, M., Rubenfeld, S., & White, M. C. (1983). Data on the Marlowe–Crowne and Edwards Social Desirability scales. *Psychological Reports, 53,* 963–968.

Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology, 24,* 349–354.

Fischer, D. G., & Fick, C. (1993). Measuring social desirability: Short forms of the Marlowe–Crowne Social Desirability scale. *Educational and Psychological Measurement, 53,* 417–424.

Hays, R. D., Hayashi, T., & Stewart, A. L. (1989). A five-item measure of socially desirable response set. *Educational and Psychological Measurement, 49,* 629–636.

Loo, R., & Thorpe, K. (2000). Confirmatory factor analyses of the full and short versions of the Marlowe–Crowne Social Desirability scale. *Journal of Social Psychology, 140,* 628–635.

Reynolds, W. M. (1982). Development of reliable and valid short forms of the Marlowe–Crowne Social Desirability scale. *Journal of Clinical Psychology, 38,* 119–125.

Strahan, R., & Gerbasi, K. C. (1972). Short, homogenous versions of the Marlow [*sic*]-Crowne Social Desirability scale. *Journal of Clinical Psychology, 28,* 191–193.

Zook, A., & Sipps, G. J. (1985). Cross-validation of a short form of the Marlowe–Crowne Social Desirability scale. *Journal of Clinical Psychology, 41,* 236–238.

Steven D. Barger
Department of Psychology
Northern Arizona University
P.O. Box 15106
Flagstaff, AZ 86011
E-mail: steven.barger@nau.edu

Received September 10, 2001
Revised November 26, 2001