A VALIDATION OF THE CROWNE-MARLOWE SOCIAL DESIRABILITY SCALE

Timothy P. Johnson, Survey Research Laboratory, Univ. of Illinois at Chicago; and
Michael Fendrich, Dept. of Psychiatry, Univ. of Illinois at Chicago
Timothy P. Johnson, Survey Research Laboratory, Univ. of Illinois at Chicago,
412 S. Peoria St., Chicago, IL 60607

KEY WORDS: Social desirability, Crowne-Marlowe scale, drug testing.

1. Introduction

Social desirability is commonly thought of as the tendency of individuals to project favorable images of themselves during social interaction. Numerous measures of the tendency to respond in a socially desirable manner have been developed since World War II (Paulhus, 1991). One of the most commonly employed scales has been the Crowne-Marlowe (CM) Social Desirability, or Need for Approval, Scale (Crowne and Marlowe, 1960). As originally developed, this measure contains 33 true-false items that describe both acceptable but improbable behaviors, as well as those deemed unacceptable but probable. Perhaps as a consequence of the attention it has received, questions have been raised about the nature of the CM and how it functions. What we refer to as the *classic social desirability interpretation* suggests that the tendency to report information that is colored by social desirability concerns is best conceptualized as a personality trait which can be measured via the CM scale. A contrasting perspective, which we label the *true-behavior interpretation*, suggests instead that the CM reliably measures actual respondent behaviors and attitudes, rather than a propensity to edit self-reports. These competing interpretations are reviewed below.

1.1 The Classic Social Desirability Interpretation

Belief that the CM scale is able to identify persons with a propensity to supply survey interviewers with self-serving information comes from numerous empirical studies that have documented consistent relations between these measures and a range of sociological and psychological variables of substantive interest. For example, the CM has been found to be significantly correlated with a variety of measures of psychiatric and psychological health. Research consistently demonstrates a negative association between symptoms of poor mental health and the CM measure, a finding that has been interpreted as evidence of underreporting of undesirable qualities by persons with a high need for approval (Gove et al., 1976; Klassen et al., 1975). The CM scale has also been found to be inversely correlated with substance use reports (Bradburn and Sudman, 1979; Welte and Russell, 1993).

Significant positive correlations have been reported between the CM scale and favorable self-evaluations, including happiness, life satisfaction and morale (Carstensen and Cone, 1983; Kozma and Stones, 1987). These findings have been cites as evidence that persons scoring highly on the CM scale tend to over-report socially desirable and under-report socially undesirable information about themselves.

1.2 The True-Behavior Interpretation

Other investigators have challenged interpretations of negative correlations between the CM scale and various negative symptoms and behaviors (and positive correlations with measures of positive characteristics) as evidence of response artifacts. Bradburn and Sudman (1979), for example, interpreted these findings as evidence that persons who score highly on tests of social desirability do in fact behave in an altruistic manner consistent with the underlying personality trait represented by these measures. Welte and Russell (1992) have put forth a similar argument. Empirical support for this position comes from a study that employed an external criterion. McCrae and Costa (1983) demonstrated that persons with high CM and Eysenck Lie scores were in fact rated more positively by their spouse across a variety of psychosocial measures.

Given that the CM measure has been used in numerous studies as an indicator of socially desirable responding, it is somewhat surprising to observe that there have been very few efforts

This research was supported by National Institute on Drug Abuse Grant #R01DA12425. An earlier version of this paper was presented at the annual meeting of the American Association for Public Opinion Research, St. Petersburg, FL, May 2002.

to validate the scale. Doing so would seem to be an important priority given its continued use and the fact that a reasonable alternative explanation for the empirical evidence relevant to this measure has been available for decades. The purpose of this report is to use external evidence to examine the predictive power of alternative interpretations of the CM scale represented by the social desirability and true-behavior hypotheses.

2. Methods

2.1 Data Sources

The present study employs two sources of data collected as part of an epidemiological study of drug use among adults in Chicago, Illinois. From June, 2001 through January, 2002, over 600 Chicago residents aged 18 to 40 were randomly selected to participate in a survey on drug abuse. This survey used an area probability sample with households selected PPS (probability proportionate to size). The Trodahl-Carter-Bryant method was employed to select a single eligible adult respondent at the household level. This study was reviewed and approved by the University of Illinois at Chicago Institutional Review Board.

The drug survey portion of the study assessed lifetime and recent drug abuse using a format similar to that employed by the National Household Survey on Drug Abuse. The survey instrument included the 10-item version of the CM scale (Strahan and Gerbasi, 1972), which was administered following the drug use module. In an effort to maintain confidentiality of drug data collection, surveys were administered in the home by trained interviewers using audio computer-assisted self-interview (ACASI) procedures on laptop computers. Prior to the screening process, interviewers distributed letters to selected households introducing the study and the survey organization.

Immediately following the drug assessment portion of the survey, separate consent to participate in three drug testing procedures was requested from subjects: hair testing, saliva testing, and urine testing. Note that respondents were not informed of the drug testing procedures until after the drug survey portion of the study had been completed. The order of presentation for testing was random, and subjects were not asked to participate in the next procedure until they had decided upon participation on the previous procedure. Hair samples were taken by the interviewers who were instructed to cut approximately sixty strands of hair from the head of potential subjects. Subjects judged by interviewers as having insufficient hair length (approximately less than one half inch in length) were deemed ineligible to participate and were not recruited into the procedure. For saliva testing, subjects were asked to spit into a plastic tube presented to the subject by the interviewer. For urine testing, subjects were asked to privately urinate in a plastic cup provided to the subject by the interviewer. All biological specimens were assayed for the indications of cocaine and several other substances by US Drug Testing Laboratories in Des Plaines, Illinois. Of 627 subjects participating in the study, 614 (90.9%) answered all 10 CM items and a useable biological assay was available for 571 (91.1%). Both a test specimen and the CM scale were available 559 cases (89.2%).

2.2 Measures

In this analysis, we focus on self-reporting of last year cocaine use. Cocaine is believed to be one of the most stigmatizing of all drugs (Magura and Kang, 1996) and is thus an appropriate item for evaluating the ability of the CM measure to distinguish between respondents who are and are not willing to report highly sensitive information. Self-reports of cocaine use are compared to results from drug test assays. Reported cocaine use includes affirmative answers to questions regarding either crack cocaine of "any form of cocaine." Those respondents who tested positive for cocaine metabolites in hair, saliva or urine samples were classified as recent cocaine users. Those not testing positive using any of these procedures were classified as non-recent users. These two measures were used to estimate concordance, which is defined as the percentage of cases in which respondent self-reports of past year cocaine use match their drug test results. Concordance was estimated as:

$$RTC = [(a + b)/(a + b + c + d)],$$

where: RTC = Report-Test Concordance,
a = report is negative and test is negative,
b = report is positive and test is positive,
c = report is negative and test is positive, and
d = report is positive and test is negative.

The 10-item short form of the CM scale was employed in this analysis (Strahan and Gerbasi, 1972). In keeping with traditional practice, the scale was scored such that higher values represented larger numbers of socially desirable

responses endorsed. In this sample, the alpha reliability coefficient for the CM scale was 0.61.

Several sociodemographic variables were also examined, including age, education, gender and race/ethnicity. Each of these indicators has been previously shown to be associated with drug use reporting (Office of Applied Studies, 2000), the validity of drug use reporting (Fendrich et al., 1999; Hser, 1997) and the CM measure (Bradburn and Sudman (1979). For purposes of these analyses, age and education were collapsed into ordinal measures.

2.3 Analysis

Initial analyses included estimation of the level of concordance between self-reported cocaine use and drug test results. The level of agreement between these two data sources was also evaluated via the kappa statistic (Fleiss, 1981).

We first evaluate the classic social desirability hypothesis, which assumes that lower CM scores are a consequence of the tendency to provide socially desirable information, by comparing the average CM score for those who report past year cocaine use that is concordant versus disconcordant with their drug test assays. We examine the overall rate of concordance and also separately by type of concordance (concordant non-cocaine users vs. concordant cocaine users) and discordance (over-reports vs. under-reporters). Lower CM scores among persons providing concordant responses would be considered evidence supportive of the classic interpretation of the CM measure.

The true-behavior hypothesis is evaluated by comparing average CM scores for persons testing positive versus negative for recent cocaine use. Higher CM scores among those testing negative for cocaine would be considered evidence supportive of the true-behavior interpretation of the CM scale. Both sets of analyses are initially examined via bivariate comparisons of mean CM scores using t-tests and oneway analysis of variance. Logistic regression analyses are subsequently employed to examine associations between the CM scale and both concordance and positive drug testing after controlling for relevant sociodemographic items.

3. Results

Community subjects provided socially desirable answers, on average, to 5.35 of the 10 questions (SD = 2.15; median = 5, mode = 5). Consistent with previous findings, the CM scale was associated with several demographic measures, including race, education and age. Whites (mean = 4.56) and persons of other races (mean = 4.96) averaged lower CM scores than did African American (mean = 5.99) and Latino (mean = 5.55) respondents (oneway ANOVA F-test = 17.91, df = 3, 576, $p < 0.001$). CM scores also covaried with education, ranging from mean values of 6.02 for those with less than a high school education, to 5.59 among persons with a high school degree, to 5.51 among those with some college experience, to 4.74 among those having graduated from college (oneway ANOVA F-test = 10.19, df = 3, 610, $p < 0.001$). Age exhibited a nonlinear relationship with CM scores. Mean scores among persons aged 18-25 was 5.50. Among those aged 26-30, CM scores averaged 4.88, and among persons 31-40, the average CM score was 5.55 (oneway ANOVA F-test = 5.34, df = 2, 611, $p < 0.01$). CM scores did not vary by gender. The mean score was 5.20 among males and 5.45 among females (t-test = 1.41, df = 612, ns).

In our sample, 6.4% of all respondents reported cocaine use during the past year. Among those providing one or more biological samples, a total of 12.6% tested positive for cocaine. The concordance between self-reports of cocaine use during the past year and drug test results was 0.874, indicating that approximately 9 out of 10 respondents provided accurate information regarding recent cocaine use behaviors. The kappa statistic for assessing the overall level of agreement between these two data sources ($k = 0.297$), however, suggests poor agreement beyond chance (Fleiss, 1981: 218). As indicated in Table 1, the poor level of agreement is largely a consequence of respondents who test positive failing to report cocaine use during the past year. Approximately three-quarters of those testing positive for cocaine did not report use.

Mean CM scores by type of concordance are presented in Table 2. These data indicate that respondents who under-reported cocaine use (i.e., tested positive but reported no use in the past year) scored higher, on average, on the CM scale. No differences in mean CM scores were found between those who provided concordant information, regardless of whether they were or were not users. Those who over-reported cocaine use (i.e., tested negative but reported use in the past year) were also found to have CM scores lower than those who under-reported cocaine use. We believe that many of the over-reporters (n = 16) were so classified because they

indicated cocaine use at some time during the past year but not recently enough to have been detected using the biological assays. Also, the average CM scores of persons not providing a biological sample was above average, but not significantly different from any of the other groups examined. A oneway analysis (Table 2) of variance model suggested overall differences. Post hoc comparisons found the mean CM scores of concordant non-users and under-reporters to significantly vary at 0.05. These findings are consistent with the classic social desirability interpretation of the CM scale suggesting that persons who tend to provide socially desirable, if inaccurate, information will score higher on this measure.

The true-behavior hypothesis was next examined by comparing the mean CM scores of respondents testing positive versus negative for cocaine use. Contrary to expectations of this hypothesis, persons *not* having used cocaine, as assessed by drug assays, had lower CM scores (mean = 5.23; SD = 2.19) than those with positive drug test assays (mean = 5.86, SD = 1.76; t = 2.31, df = 557, p = 0.021).

Both the classic social desirability and true-behavior interpretations of the CM scale were next examined after controlling for relevant sociodemographic variables known to be associated with substance use behavior, including age, gender, education and race/ethnicity. These results are presented in Table 3. The first two columns in this table present odds ratios and 95% confidence intervals (CI) for a logistic regression model predicting cocaine concordance (coded '1') vs. under-reporting (coded '0'). In this analysis, the CM scale was found not to be associated with the accuracy of self-reported cocaine use after adjusting for age, education and race/ethnicity, all of which were found to be independently predictive of accurate reporting. The third and fourth columns present a second logistic regression model in which the same set of independent variables were employed to predict which respondents tested positive (coded '1') versus negative (coded '0') for cocaine use. Consistent with the bivariate results, CM scores were also not associated with the cocaine use. Education and race/ethnicity, in contrast, were highly predictive of testing positive for cocaine.

4. Discussion

In general, we find only weak evidence that the CM scale successfully discriminates between respondents who are and are not willing to report socially undesirable information. When examined within a bivariate context, the CM scale appears to perform in accordance with its original conceptualization. That is, persons who decline to report one socially stigmatizing behavior – last year cocaine use – score more highly on the CM measure of trait social desirability. However, the CM is unable to predict cocaine under-reporting once other measures also known to be associated with self-reported drug use and the CM scale are controlled in multivariate analyses. In particular, less educated respondents, males and members of some minority groups were more likely to fail to report recent cocaine use. These findings thus indicate that the CM scale is unable to predict socially desirable behavior once adjustments are made for common sources of measurement error. As such, the CM scale may not contribute sufficient new information to warrant its inclusion in many social investigations.

Contrary to the true-behavior interpretation of the CM measure, persons scoring highly were not less likely to use cocaine. The opposite was in fact observed: persons testing positive for cocaine, on average, scored higher on this measure. Reinterpretations of the CM scale that posit it to be a direct measure of a set of personal behaviors rather than an indicator of propensity to distort answers to socially desirable questions, while provocative, may also not be correct (Bradburn and Sudman, 1979).

These conclusions are presented with the acknowledgment that this study cannot be considered definitive. Most obviously, the CM scale was validated against a single objective. Although cocaine use does represent a clearly socially undesirable behavior, validating the CM against self-reports of other verifiable behaviors, such as criminal justice experiences, and desirable activities, such as civic participation and positive health behaviors, would provide additional evidence regarding the practical utility of this measure. In addition, the full CM scale was not employed in this study. Although the subset of items used has a strong correlation with the complete 33-item measure (Strahan and Gerbasi, 1972), the possibility that the use of the abbreviated measure may have influenced results cannot be ruled-out. Acknowledging these limitations, it is also important to note that our findings are based on a representative community sample for which self-reported data were collected prior to the knowledge that confirmatory biological specimens would be requested. As such, this represents one of the

few rigorous attempts to validate the widely used CM scale using external sources of evidence.

In summary, when contrasted with a confirmatory source of information, the CM scale is found to behave in a manner consistent with the expectations originally specified by Crowne and Marlowe (1960) some 40 years ago. This measure is also strongly associated with other respondent characteristics, including education, race/ethnicity and age. These demographic variables largely account for the observed variability in the accuracy of self-reported cocaine use. Consequently, these analyses suggest that the CM is not an independent predictor of the tendency to provide self-enhancing, or socially desirable, answers in social surveys.

5. References

Bradburn, N. M. and Sudman, S. (1979) *Improving Interview Method and Questionnaire Design*, San Francisco: Jossey-Bass.

Carstensen, L. L. and Cone, J. D. (1983), "Social Desirability and the Measurement of Psychological Well-Being in Elderly Persons," *Journal of Personality*, 38, 713-715.

Crowne, D. P. and Marlowe, D. (1960), "A New Scale of Social Desirability Independent of Psychopathology," *Journal of Consulting Psychology*, 24, 349-354.

Crowne, D. P. and Marlowe, D. (1964), *The Approval Motive*, New York: John Wiley & Sons.

Eysenck, H. J. and Eysenck, S. B. G. (1964), *Manual of the Eysenck Personality Inventory*, London: University Press.

Fendrich, M., Johnson, T. P., Sudman, S., Wislar, J. and Spiehler, V. (1999), "Validity of Drug Use Reporting in a High-Risk Community Sample: A Comparison of Cocaine and Heroin Survey Reports with Hair Tests," *American Journal of Epidemiology*, 149, 955-962.

Fleiss, J. L. (1981), *Statistical Methods for Rates and Proportions, Second Edition.* New York: John Wiley & Sons.

Gove, W. R., McCorkel, J., Fain, T. and Hughes, M. D. (1976), "Response Bias in Community Surveys of Mental Health: Systematic Bias or Random Noise?" *Social Science & Medicine*, 10, 497-502.

Hser, Y. I. (1997), "Self-Reported Drug Use: Results of Selected Empirical Investigations of Validity," in *The Validity of Self-Reported Drug Use: Improving the Accuracy of Survey Estimates*, eds. L. Harrison and A. Hughes, Rockville, MD: National Institute on Drug Abuse, pp. 320-343.

Klassen, D., Hornstra, R. K., and Anderson, P. B. (1975), "Influence of Social Desirability on Symptom and Mood Reporting in a Community Survey," *Journal of Consulting and Clinical Psychology*, 43, 448-452.

Kozma, A. and Stones, M. J. (1987), "Social Desirability in Measures of Subjective Well-Being: A Systematic Evaluation," *Journal of Gerontology*, 42, 56-59.

McCrae, R. R. and Costa, P. T. (1983), "Social Desirability Scales: More Substance than Style," *Journal of Consulting and Clinical Psychology*, 51, 882-888.

Maguara, S. and Kang, S. (1996), "The Validity of Self-Reported Drug Use in High Risk Populations: A Metaanalytic Review," *Substance Use and Misuse*, 31, 1131-1153.

Office of Applied Studies (2000), *Summary of Findings from the 1999 National Household Survey on Drug Abuse*, Rockville, MD: Substance Abuse and Mental Health Services Administration.

Paulhus, D. L. (1991), "Measurement and Control of Response Bias," in Measures of Personality and Social Psychological Attitudes, eds. J. P. Robinson and P. R. Shaver, San Diego: Academic Press, pp. 17-59.

Strahan, R. and Gerbasi, K. C. (1972), "Short, Homogeneous Versions of the Marlowe-Crowne Social Desirability Scale," *Journal of Clinical Psychology*, 28, 191-193.

Welte, J. W. and Russell, M. (1993), "Influence of Socially Desirable Responding in a Study of Stress and Substance Abuse," *Alcoholism: Clinical and Experimental Research*, 17, 758-761.

Table 1. Concordance of Past Year Reporting of Cocaine Use with Drug Test Results

	Tested Negative for Cocaine		Tested Positive for Cocaine	
Reported Cocaine Use	Percent	(n)	Percent	(*n*)
No	96.8	(483)	73.6	(53)
Yes	3.2	(16)	26.4	(19)
Total	100.0	(499)	100.0	(72)

Overall concordance rate = 0.879. Kappa coefficient = 0.297.

Table 2. Mean Crowne-Marlowe Score by Cocaine Self-Report/Drug Test Concordance

	Mean	Standard Deviation	(*n*)
Concordant – no use	5.24[a]	2.20	(472)
Concordant – used	5.21[a,b]	1.69	(19)
Under-reported	6.10[b]	1.74	(52)
Over-reported	4.94[a,b]	1.98	(16)
No test provided	5.82[a,b]	2.14	(55)
Total	5.36	2.15	(614)

Oneway ANOVA F-value = 2.72, df = 4, 613, $p = 0.029$.
NOTE: Percentages with different superscripts differ significantly (P < .05) on the basis of the least significant difference test.

Table 3. Logistic Regression Analyses of Cocaine Concordance and Testing Positive for Cocaine Use

	Cocaine Concordance		Cocaine Test Positive	
	Odds Ratio	95% CI	Odds Ratio	95% CI
Crowne-Marlowe Scale	0.90	0.77-1.05	1.04	0.91-1.19
Age 18-25 years	2.22*	1.03-4.77	0.50*	0.25-1.00
Age 26-30 years	1.00	---	1.00	---
Age 31-40 years	1.58	0.71-3.48	0.95	0.48-1.88
Female	1.00	---	1.00	---
Male	1.88	0.90-3.94	0.74	0.41-1.34
< high school degree	0.04**	0.004-0.29	25.33***	5.44-118.00
High school graduate	0.06**	0.007-0.48	13.99***	3.02-64.90
Some college	0.10*	0.01-0.81	9.49**	2.09-43.09
College graduate	1.00	---	1.00	---
White	1.00	---	1.00	---
African American	0.20*	0.04-0.88	2.68*	1.03-6.99
Latino	0.45	0.09-2.39	1.53	0.51-4.55
Other race group	0.07**	0.01-0.43	4.56*	1.11-18.78
Model X^2	66.89***		68.68***	
Degrees of freedom	10		10	
Hosmer-Lemeshow X^2 Test	5.27 (ns)		3.72 (ns)	
Nagelkerke R^2	0.25		0.22	
(n)	(537)		(553)	

* p < 0.05 ** p < 0.01 *** p < 0.001