Qual Quant (2012) 46:1883–1896
DOI 10.1007/s11135-011-9563-5

An evaluation of the validity of the Crowne–Marlowe need for approval scale

Timothy P. Johnson · Michael Fendrich · Mary Ellen Mackesy-Amiti

Published online: 8 August 2011

Abstract Over the past four decades, research has consistently documented negative correlations between the Crowne–Marlowe (CM) social desirability trait scale and numerous measures of sensitive behaviors, conditions, and opinions. These findings have been interpreted as evidence that persons with self-presentation concerns tend to under-report negative information. In contrast to this classic social desirability interpretation, a second perspective, labeled the true-behavior hypothesis, maintains that these correlations in fact reflect accurate reporting of both sets of variables, suggesting that the CM scale is not a sensitive indicator of the social desirability trait. We test these alternative interpretations by examining data from a community survey that collected both self-reports of cocaine use and the biological specimens necessary to validate the self-reports. In bivariate analyses, the CM scale was found to be associated with the concordance of cocaine use reporting and biological assays in a manner consistent with the classic social desirability hypothesis. The CM scale was not found to be associated with actual cocaine use, as measured by drug test assays, a finding inconsistent with the true-behavior hypothesis. After adjusting for other known correlates of substance use in logistic regression models, the CM scale was not associated with cocaine use under-reporting, nor with actual cocaine use behavior. Until further evidence is available, we conclude that the CM may be a questionable indicator of socially desirable reporting behavior in social surveys.

Keywords Social desirability · Crowne–Marlowe scale · Validation

T. P. Johnson (✉)
Survey Research Laboratory, Department of Public Administration,
University of Illinois at Chicago, 412 S. Peoria St., Chicago, IL 60607, USA
e-mail: tjohnson@srl.uic.edu

M. Fendrich
Center for Addiction and Behavioral Health Research, Helen Bader School of Social Welfare,
University of Wisconsin-Milwaukee, Milwaukee, WI, USA

M. E. Mackesy-Amiti
Division of Epidemiology and Biostatistics, School of Public Health,
University of Illinois at Chicago, Chicago, IL, USA

1 Introduction

For more than 50 years researchers have expressed concerns that the quality of survey self-reports may be influenced or biased by social desirability effects (DeMaio 1984). Social desirability is commonly thought of as the tendency of individuals to project favorable images of themselves during social interaction. For example, Edwards (1957, p. 35) defined this concept as "the tendency to give socially desirable responses in self-description." Crowne and Marlowe (1964) suggested the motivation for this tendency was "the need of subjects to respond in culturally sanctioned ways" in order to obtain social approval. Phillips and Clancy (1970, p. 505) observed that social desirability "reflects the tendency to say (or admit to) good rather than bad things about oneself." Ross and Mirowsky (1983, pp. 529–530) viewed "the tendency to give socially approved responses as an interaction strategy characterized by responding in normatively correct and conformist ways and generally trying to present a good face." Each of these conceptualizations conforms to Cicourel's (1964) theoretical expectation that survey respondents will tailor their answers in a manner consistent with the perceived values of the interviewer.

Numerous measures of the tendency to respond in a socially desirable manner have been developed (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. A considerable body of research employing the CM has accumulated, and it is believed to be the most commonly used measure of social desirability now available (Barger 2002; Beretvas et al. 2002; Moorman and Podsakoff 1992). 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 that can be measured via the CM scale. A contrasting perspective, which we label the *true-behavior interpretation,* challenges the classic hypothesis, suggesting 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

The general rationale behind the CM scale is that some respondents will answer survey questions based on the perceived social acceptability of the questions rather than item content (Phillips and Clancy 1972), and the CM scale is designed to detect this respondent behavior. Belief that the CM scale is able to identify persons with a propensity to report 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. Psychological correlates of the CM scale include the following: emotional problems (Klassen et al. 1975); depressive symptoms (Klassen et al. 1975); the Langner scale of psychiatric symptoms (Carr and Krause 1978; Gove et al. 1976; Klassen et al. 1975); the Beck hopelessness inventory (Linehan and Nielson 1983);

psychological distress (Ross and Mirowsky 1984); past suicidal behavior, recent suicide ideation, and likelihood of future suicide (Strosahl et al. 1984); and hostility (Gick et al. 1997). The CM scale has also been found to be inversely correlated with self-reported substance use, including alcohol consumption, intoxication, and marijuana use (Bradburn et al. 1979; see also Pleck et al. 1996; Watten 1996) and lifetime drug use (Welte and Russell 1993). Significant positive correlations have been reported between the CM scale and favorable self-evaluations. Kozma and Stones (1987) report correlations of 0.30 with a happiness scale and 0.21 with a measure of life satisfaction. Using an elderly sample, Carstensen and Cone (1983) report correlations of 0.58 and 0.70 with measures of life satisfaction and morale, respectively.

These findings are often cited as evidence that persons scoring highly on the CM scale tend to over-report socially desirable information about themselves and under-report socially undesirable information, presumably to manage other's impressions of them and more easily obtain social acceptance. It is important to note the body of research upon which this conclusion is based involves self-reported information only and that the strong assumption is made that the quality of these self-reports varies across CM scores. In fact, few direct attempts to examine the associations between the CM and validated self-report information exist. One indirect attempt to predict response bias using the CM has been unsuccessful. Bishop et al. (1986) reported no correlation between the CM and respondent willingness to offer opinions on fictitious issues. We are aware of only two direct attempts to evaluate the association between CM and self-reports of socially desirable behavior, neither of which support the classic social desirability hypothesis (McCrae and Costa 1983; Yoshino and Kato 1995). These will be reviewed shortly.

1.2 Alternative explanations

Other investigators have challenged conventional interpretations that elevated CM scores are evidence of deception in survey reports. Bradburn et al. (1979), for example, interpreted CM negative associations with survey reports of undesirable behavior as evidence that persons who score highly on tests of social desirability also behave in a manner consistent with the underlying personality trait represented by these measures. Welte and Russell (1993) have put forth a similar argument, suggesting that persons with high CM scores are very concerned with social approval and might reflect this concern via both opinions and behaviors that conform to social norms, rather than deliberate attempts at deception.

Thus, an alternative explanation for CM/survey behavior correlations emerges: Those scoring high on the CM and low on survey measures of undesirable behavior are providing honest responses about behaviors that they are unlikely to perform (or symptoms they are unlikely to experience). According to this alternative explanation, the CM is not a marker for deception (and can not be used to adjust for this in survey analyses). Rather, this alternative explanation suggests that CM is a construct reflecting high levels of conformity to socially desirable behavior.

Empirically, the only way that the claim that CM is a measure of deception can be verified is by examining its association with some other validated index gauging the underreporting of undesirable behavior. For example, if a survey inquires about substance abuse, one would need a test specifically confirming the extent to which a respondent underreported that behavior. If the CM is positively associated with underreporting among those testing positive for drug use, then support for the notion that the CM is a construct reflecting deception would be provided. A null or even a negative association between CM and this measure

of underreporting would fail to verify the classic social desirability, or deception, hypothesis and provide support for alternative explanations.

Empirical refutation of the deception hypothesis comes from the two available studies that have employed an external criterion. McCrae and Costa (1983) demonstrated that persons with high CM scores were in fact rated more positively by their spouse across a variety of psychosocial measures, including friendliness, good adjustment, and openness to experience. This report may be limited to the extent that validation information was based on collateral informants whose information may not be completely valid. For example deceptive informants may be consistently deceptive in their relationships with their spouses, thus causing their spouses to misjudge the nature of their actual behavior.

In a Japanese study, Yoshino and Kato (1995) examined the relationship between CM scores and the accuracy of self-reported alcohol abstinence among alcoholics receiving outpatient treatment in Tokyo. In their study, blood serum glutamyl γ-transpeptidase (GGT) alcohol abstinence was employed to verify self-reported alcohol abstinence. These researchers found that in general, reports of abstinence showed an association with biochemical markers, which reflected a consistent tendency toward honest self-disclosure. They found no consistent association between CM scores and deceptive reporting on drinking behavior among their study subjects. While this study is unique in its use of an independent biochemical marker for validation purposes, the sample itself is problematic, as it focused on patients in treatment for alcohol disorder who may have been highly motivated to honestly disclose their behavior.

Given that, consistent with classic interpretations of this measure, the CM scale is commonly employed as an indicator of socially desirable responding for adjustment purposes (Gove et al. 1976; Kozma and Stones 1987; McCrae 1986), it is somewhat surprising to observe how limited the efforts have been to validate the measure. Doing so would seem to be an important priority given the continued use of the CM measure in survey research (Couper et al. 2003). The purpose of this report is to examine the validity of the CM as a measure of respondent deception in survey reporting of sensitive behavior. We employ a validated measure of under-reporting of drug use that is based on biological assessment of this behavior. Furthermore, our analyses employ regression modeling that controls for potential socio-demographic confounders that have been shown to be correlated with CM in previous research (Warnecke et al. 1997; Klassen et al. 1975; Ross and Mirowsky 1983, 1984).

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 age 18–40 were randomly selected to participate in a survey on drug abuse. A multi-stage probability sample design was employed (Levy and Lemeshow 1991). At stage 1, census tracts in Chicago were selected randomly. At stage 2, one block was selected randomly from within each sampled tract. At stage 3, every household on the sampled block was screened for eligibility. At stage 4, one 18–40 year-old adult was selected from within each eligible household using the Trodahl-Carter-Bryant respondent selection procedure (Bryant 1975). This study was reviewed and approved by the University of Illinois at Chicago Institutional Review Board.

Using the American Association for Public Opinion Research (2000) response rate formula #3, the overall response and cooperation rates for the survey were 40 and 74%, respectively. These rates reflect the challenges of conducting in-person survey interviews in urban environments where household response rates tend to be lower for many reasons (Groves and Couper 1998). When restricted access, high-rise apartment buildings are excluded from consideration, the comparable response and cooperation rates were 51 and 75%, respectively. Restricted access apartment buildings are particularly problematic in urban surveys, as a single gatekeeper can successfully decline survey participation on behalf of dozens, and sometimes hundreds of potential respondents.

The drug survey portion of the study assessed lifetime and recent drug abuse using a format similar to that employed by the National Survey of Drug Use and Health. 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 60 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 U.S. Drug Testing Laboratories in Des Plaines, Illinois.

Of 627 subjects participating in the study, 614 (97.9%) responded to all CM items; biological specimens sufficient to be assayed were available for 565 (90.1%) respondents. Both the CM and biological assays were available for 553 (88.2%) of the respondents.

2.2 Measures

In this analysis, we focus on self-reporting of past 90-day 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. The self-report indicator of past 90-day cocaine use was based upon responses to questions concerned with "crack" and "any form of cocaine" use. 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. As a check on the psychometric viability of this combined measure, we computed a Cronbach's alpha reliability coefficient for the three individual biological sample results. This analysis yielded an alpha value of 0.66.

Table 1 Items included in the Crowne–Marlowe short form

1.	I'm always willing to admit it when I make a mistake (true)[a]
2.	I always try to practice what I preach (true)
3.	I never resent being asked to return a favor (true)
4.	I have never been irked when people expressed idea very different from my own (true)
5.	I have never deliberately said something that hurt someone's feelings (true)
6.	I like to gossip at times (false)
7.	There have been occasions when I took advantage of someone (false)
8.	I sometimes try to get even rather than forgive and forget (false)
9.	At times I have really insisted on having things my own way (false)
10.	There have been occasions when I felt like smashing things (false)

[a] Direction for coding socially desirable response

These two measures were used to estimate concordance, which is defined as the percentage of cases in which respondent self-reports of past 90-day cocaine use match their drug test results. Ninety days corresponds approximately to the longest window for cocaine detection using any of the three biological samples examined (Cone 1997). Concordance was estimated as:

$$\mathrm{RTC} = [(a + b) / (a + b + c + d)], \quad (1)$$

where RTC is report-test Concordance, a represents cases in which the report is negative and test is negative, b represents cases in which the report is positive and test is positive, c represents cases in which the report is negative and test is positive, and d represents cases in which the report is positive and test is negative.

In keeping with traditional practice, the CM scale was scored such that higher values represented larger numbers of socially desirable responses endorsed. The specific items in this scale are presented in Table 1. In this sample, the alpha reliability coefficient for the CM scale was 0.61.

Several socio-demographic variables also were examined, including age, education, gender and race/ethnicity. Age was measured in years. Education was measured using 8 ordinal categories that ranged from no formal education (coded as '1') through graduate or professional degree (coded as '8'). Gender and race/ethnicity were coded as dummy variables. 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 et al. 1979; Warnecke et al. 1997).

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 next employed one-way ANOVA and t-tests to examine mean CM scores by socio-demographic measures and by drug use reporting status. Higher CM scores in group c (the under-reporters), compared to those in other groups, would validate the classical interpretation of the CM as a measure of social desirability. Following these bivariate analyses, logistic regression was used to investigate associations between the CM scale and group c membership status, after controlling for relevant socio-demographic

items. OLS regression was also employed to examine the independent associations between socio-demographic items and CM scores.

3 Results

The final sample was 61.4% female. The average age was 27.9 years (SD = 6.30; age range = 18–40 years). African Americans constituted 40.6% of the sample. Non-Hispanic whites represented another 32.0%, and Latinos represented 18.2% of the sample. The remaining 9.2% included Asian Americans, Native Americans, and persons of mixed race/ethnicity. Persons with less than a high school education constituted 17.4%, and persons with a high school degree or equivalency, 21.9% of the sample. Persons with some college, and those with a college degree or higher, represented 27.3 and 33.5% of the sample, respectively.

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. The one-way ANOVA test for race was significant (F-test = 17.91, df = 3, 576, $p < 0.001$). Post hoc comparisons, using the least significant differences test, found that Whites (mean = 4.56) and persons of other races (mean = 4.96) averaged lower CM scores than did African American respondents (mean = 5.99). Whites also scored lower than Latino (mean = 5.55) respondents. CM scores also co-varied with education (one-way ANOVA F-test = 10.19, df = 3, 610, $p < 0.001$). Post hoc comparisons found that college graduates (mean = 4.74) had significantly lower CM scores than all other education groups (mean = 6.02 for those with less than a high school education; mean = 5.59 for persons with a high school degree; mean = 5.51 among those with some college experience); persons with some college also were found to have lower scores than persons with less than a high school degree. Age differences in mean CM scores were also noted (one-way ANOVA F-test = 5.34, df = 2, 611, $p < 0.01$). The mean score among persons 18–24 years old was 5.54. Among those 25–29 years old, CM scores averaged 4.88; among persons 30–34 years old, the average CM score was 5.44; and among persons aged 35–40, the average score was 5.60. The scores for persons aged 25–29 were found to be significantly lower than those of all other age groups. CM scores did not vary by gender. The mean score was 5.20 among males and 5.45 among females ($t = 1.41$, df = 612, ns).

In our sample, 4.0% of all respondents reported cocaine use during the past 90 days. Among those providing one or more biological samples, 12.7% tested positive for cocaine. The concordance between self-reports of cocaine use during the past 90 days and drug test results was 0.887, 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 ($k = 0.276$), however, suggests poor agreement beyond chance (Fleiss 1981, p. 218). As indicated in Table 2, the poor level of agreement is largely a consequence of respondents who test positive failing to report cocaine use during the past 90 days. Four out of five of those testing positive for cocaine did not report use.

The bivariate comparison between group c (under-reporters) and all others is summarized in Table 3. These data indicate that respondents who under-reported cocaine use (i.e., tested positive but reported no use in the past 90 days) scored higher than other respondents, on average, but not significantly different from any of the other groups examined. A one-way analysis of variance model yielded differences that were not significant (F-value = 1.87, df = 4, 609, ns). Subsequently, a t-test (not shown) compared under-reporters (group c) with respondents who provided concordant answers (groups a & b). In this analysis, under-reporters were

Table 2 Concordance of past 90-day reporting of cocaine use with drug test results

	Reported past 90-day cocaine use			
	% Yes	*n*	% No	*n*
Tested positive for cocaine	20.8	15	79.2	57
Tested negative for cocaine	1.4	7	98.6	486
No test available	4.8	3	95.2	59

Overall concordance rate (among persons providing test data and self-report) = 0.887
Kappa coefficient = 0.276

Table 3 Mean Crowne–Marlowe score by past 90-day cocaine self-report/drug test concordance

	Mean	Standard deviation	*n*
Concordant—no use	5.24	2.20	475
Concordant—used	5.40	1.84	15
Under-reported	5.98	1.73	56
Over-reported	5.29	2.43	7
No test provided	5.67	2.15	61
Total	5.36	2.15	614

One-way ANOVA
F-value = 1.87, df = 4, 609, ns
Levine statistic for test of homogeneity of variance = 1.73, df1 = 1, df2 = 609, ns

found to have significantly higher CM scores ($\bar{X} = 5.98$, SD= 1.73) than did concordant reporters ($\bar{X} = 5.25$, SD=2.18; $t = 2.93$, unequal variances df=76.5, $p = 0.004$).

This association between underreporting status and CM scale scores was next examined after controlling for relevant socio-demographic variables known to be associated with substance use behavior, including age, gender, education, and race/ethnicity. Results of a stepwise logistic regression model, used to examine the independent effects of the CM measure on cocaine under-reporting (coded as "1") vs. concordance (coded "0"), are presented in Table 4. The first two columns in this table present odds ratios and 95% confidence intervals (CI) for the initial step of this model, in which only the CM measure is entered into the equation. Consistent with the pattern of bivariate results revealed earlier (in Table 3), the CM measure was found to be positively associated with under-reporting such that under-reporters scored higher on the CM scale, relative to those respondents who provided concordant answers. When demographic measures are added during the second step of this model (columns 3 & 4 of Table 4), however, the effects of the CM scale are reduced to non-significance. In this model, persons with less education were more likely to under-report cocaine use, as were African American and other race group respondents. It should be noted that this analysis excluded the seven respondents who reported past year cocaine use but who tested negative (i.e., the "over-reporters"). An additional logistic regression model that included these cases produced identical conclusions.

An additional analysis was conducted of the classic social desirability hypothesis in which only persons testing positive for cocaine use were examined (results not shown). Comparing these two subgroups provides possibly the strongest test of the hypothesis as it most clearly contrasts individuals who were and were not willing to report socially undesirable information. For this analysis, logistic regression was again employed to determine if the CM scale was associated with concordant vs. under-reporting of cocaine use, albeit only among recent cocaine users. As Table 3 suggests, there were only 71 cases in these data that tested positive for cocaine use ($n = 15$ concordant reporters; $n = 56$ under-reporters). This model failed to

Table 4 Logistic regression analysis of 90-day cocaine use under-reporting

	Model 1		Model 2	
	Odds ratio	(95% CI)	Odds ratio	(95% CI)
Age (in years)			1.02	(0.98–1.07)
Female			1.00	–
Male			0.61	(0.31–1.21)
Education			0.54***	0.41–0.71)
White			1.00	–
African American			5.99**	(1.73–20.72)
Latino			2.04	(0.48–8.57)
Other race group			12.91**	(2.54–65.63)
Crowne–Marlowe scale	1.18**	(1.04–1.35)	1.06	(0.91–1.23)
Intercept	0.05***	–	0.19	–
Step χ^2	6.24**		56.45***	
Degrees-of-freedom	1		6	
Nagelkerke R^2	0.02		0.22	
n	546		546	

** $p < 0.01$
*** $p < 0.001$

produce reliable parameter estimates, most likely as a consequence of the small available sample size.

While it would be tempting to suggest that an exploration of the association between CM scores and test result status would identify whether or not the CM is an indicator of "true" socially desirable behavior among respondents, our data do not permit such a test since the majority of those who were detected as drug users underreported their drug use. Only 15 of the 71 subjects testing positive for cocaine actually disclosed use of this drug.

Given this lack of support for both the social desirability and true-behavior hypotheses, we explored other possible associations between CM and cocaine under- reporting. Our analyses indicated that those respondents with less education were also more likely to under-report answers and more likely to score highly on the CM measure. We thus explored the post hoc hypothesis that the association between CM scores and under-reporting is not a direct relationship but rather a consequence of associations each has with education. As education is highly correlated with cognitive skills (Ceci 1991) and is sometimes employed as an indicator of cognitive sophistication (Krosnick 1992), these findings might be interpreted as evidence that respondents low in cognitive sophistication are more likely to score highly on the CM and also misreport their drug use status (Dijkstra et al. 2001). This pattern of evidence would seem to be consistent with the general proposition that response effects and reporting errors are most common among those respondents who are able to invest the fewest cognitive resources in the construction of answers to survey questions and/or who are most likely to satisfice (Narayan and Krosnick 1996). If correct, higher scores on the CM might not represent deliberate attempts at impression management as much as less careful consideration of responses to vaguely worded questions that are now more than 40 years old. While attractive, this explanation also appears insufficient as additional multivariate analyses revealed that education does not predict CM scores independent of race/ethnicity in our

Table 5 Multiple linear regression analysis of Crowne–Marlowe scores

	Unstandardized coefficients	Standard errors
Age	0.01	0.01
Male (1=yes)	−0.06	0.18
Education	−0.12	0.07
African American (1=yes)	1.20	0.22***
Latino (1=yes)	0.80	0.26**
Other race group (1=yes)	0.29	0.42
Intercept	4.98	0.52***
R^2	0.08	
n	608	

** $p < 0.01$
*** $p < 0.001$

sample (see Table 5). A lack of association between education and CM scores has also been previously reported in other multivariate analyses (Warnecke et al. 1997).

Respondent race/ethnicity, in contrast, does independently influence both CM scores and cocaine use under-reporting. Specifically, Table 5 indicates that minority respondents, African American and other race group respondents in particular, appear to have higher CM scores on average and are also more likely to under-report cocaine use (Table 4). Ross and Mirowsky (1983) have interpreted similar findings as evidence that a relative lack of social power may lead minority groups to emphasize impression management and conformity. The endorsement of socially desirable survey items may be one method of doing so, and previous research has found significantly higher CM scores among minority groups in the U.S. (Klassen et al. 1975; Ross and Mirowsky 1984; Warnecke et al. 1997). In addition, the independent effects of race/ethnicity on cocaine under-reporting depicted in Table 4 are consistent with prior research suggesting lower quality reporting of substance use behaviors among minority respondents (Johnson and Bowman 2003). These findings are not unexpected given other research confirming the widespread belief that members of minority communities are more likely to be prosecuted for drug-related offenses and receive harsher sentences than white offenders convicted of similar crimes (Bonczar and Beck 1997; Roscoe and Morton 1994; Stone 1999). The spurious relationship between the CM measure and cocaine use under-reporting in this study may thus be a consequence of cultural variations in social power, perceptions of discrimination and sensitivity to drug use questions, which are together responsible for variations in both CM scores and drug use under-reporting in this sample.

4 Discussion

In general, we found some evidence that the CM scale successfully discriminates between respondents who under-report socially undesirable information and those who do not. 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 90 days cocaine use—score higher on the CM measure of trait social desirability than do respondents who provide accurate reports of non-use. 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. Unfortunately, more fine-grained analyses designed to investigate the CM scale's ability to discriminate between cocaine users who do and do not report their drug use were unsuccessful due to sample limitations. Consequently, our findings are inconclusive, although they tend to suggest the CM scale is unable to predict socially desirable behavior once adjustments are made for common sources of measurement error. As such, it is unclear whether the CM scale contributes sufficient new information to warrant its inclusion in many social investigations. Of course, it may be that the general absence of a CM-concordance association is a consequence of the nature of the dependent variable being examined: drug use report-test agreement. Indeed, only a small portion of the empirical evidence regarding CM correlations involves behaviors that are illegal. It may well be that the motivation for under-reporting behaviors with socially desirable answers may vary from those that in addition carry legal implications. It would thus be advisable to conduct additional validation work that examines a range of other self-report measures.

We note that having found even a weak association between under-reporting and CM scores necessarily undercuts the alternative "true behavior" hypothesis. The association between positive drug test outcomes and CM scores is further evidence for the lack of support of this alternative explanation. It should be further noted that since most positive tests were under-reported, test result outcomes and our index of under-reporting are closely related constructs. Thus, while we were unable to carry out a powerful test of the "true behavior" hypothesis, it can be argued that our findings obviated the need for such analyses.

These conclusions are presented with the acknowledgment that this study cannot be considered definitive. As acknowledged above, the CM scale was validated using biological measures of a single behavior. Although cocaine use does represent a clearly socially undesirable activity, 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, we should acknowledge that biological measures are not perfect indicators of use, and there is always the potential, however slight, for both false positives and false negatives to occur (for a specific critique of testing techniques, see Wolff et al. 1999).

Another consideration in evaluating these findings is the use of ACASI technology to collect the survey data. ACASI is a valuable method for the collection of high-quality survey data (Tourangeau and Smith 1996; Turner et al. 1998). One of its main advantages is the establishment of a more confidential reporting environment by eliminating the need for a respondent to verbally report answers to an interviewer. Hence, the social desirability demands associated with reporting potentially stigmatizing or incriminating questions, such as recent drug use behavior, is believed to be reduced. Our failure to validate the CM scale might thus be a consequence of the ACASI reporting method if it was in fact successful in improving drug reporting validity. Future research might consider this possibility via experimental manipulation of interview mode. Nevertheless, the utility of further comparisons with other data collection methods may be limited given that this mode of administration is generally considered "state of the art" for purpose of drug abuse research, including the National Survey on Drug Use & Health (Office of Applied Studies 2001).

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. As mentioned earlier, the CM scale was examined in this study because it has historically been the most commonly employed measure of social desirability employed in the literature. It is important to acknowledge that more recent measures of social desirability have also been developed that recognize the multidimensional character of this construct.

Paulhus (1998), for example, has developed a two dimensional instrument designed to measure other-deception (i.e., impression management) vs. self-deception. Future work should certainly be directed at evaluating the validity and usefulness of these measures.

Acknowledging these limitations, it is also important to note several advantages of this study design. Unlike much of the research evaluating the psychometric properties of CM, this research was drawn from a random sample of the general population. This design feature, along with a protocol design in which survey reports were completed prior to the introduction of biological testing, enhances the overall validity of our results. 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. Multivariate adjustments for demographic measures, race/ethnicity in particular, nonetheless suggest that the association between CM scores and reporting accuracy may be either spurious or very weak. Additional analyses reveal that several demographic variables, and the social processes they represent, largely account for the relationship between CM scores and actual cocaine use behavior. Consequently, the evidence presented in this study suggests that the CM measure may not be an independent predictor of the tendency to provide socially desirable answers in social surveys, nor does it appear to be a reliable indicator of actual social behavior. Given the limitations of this research, however, additional investigations will be necessary to resolve the status and utility of the CM measure in survey research.

Acknowledgment This research was supported by National Institute on Drug Abuse Grant #R01DA12425.

References

American Association for Public Opinion Research.: Standard Definitions: Final Dispositions of Case Codes and Outcomes for Surveys. Ann Arbor, Michigan: American Association for Public Opinion Research (2000)

Barger, S. D.: The Marlowe–Crowne affair: short forms, psychometric structure, and social desirability. J. Pers. Assess. **79**, 286–305 (2002)

Beretvas, S. N., Meyers, J. L., Leite, W. L.: A reliability generalization study of the Marlowe–Crowne social desirability scale. Educ. Psychol. Meas. **62**, 570–589 (2002)

Bishop, G. F., Tuchfarber, A. J., Oldendick, R. W.: Opinions on fictitious issues: the pressure to answer survey questions. Public Opin. Q. **50**, 240–250 (1986)

Bonczar, T. P., Beck, A. J.: Lifetime Likelihood of Going to State or Federal Prison. Special Report, U.S. Department of Justice, Office of Justice Programs, Bureau of Justice Statistics, Washington, DC (1997). Retrieved October 6, 2003 from http://www.ojp.usdoj.gov/bjs/pub/pdf/llgsfp.pdf

Bradburn N. M., Sudman Seymour and Associates: Reinterpreting the Marlowe-Crown scale. In: Bradburn, N. M. (ed.) Improving Interview Method and Questionnaire Design. Seymour Sudman and Associates, pp. 85–106. Jossey-Bass, San Francisco (1979)

Bryant, B. E.: Respondent selection in a time of changing household composition. J. Mark. Res. **12**, 129–135 (1975)

Carr, L. G., Krause, N.: Social status, psychiatric symptomatology, and response bias. J. Health Soc. Behav. **19**, 86–91 (1978)

Carstensen, L. L., Cone, J. D.: Social desirability and the measurement of psychological well-being in elderly persons. J. Pers. **38**, 713–715 (1983)

Ceci, S. J.: How much does schooling influence general intelligence and its cognitive components? A reassessment of the evidence. Dev. Psychol. **27**, 703–722 (1991)

Cicourel, A. V.: Method and Measurement in Sociology. Free Press, New York (1964)

Cone, E. J.: New developments in biological measures of drug prevalence. In: Harrison, L., Hughes, A. (eds.) The Validity of Self-Reported Drug Use: Improving the Accuracy of Survey Estimates, pp. 108–129. National Institute on Drug Abuse, Rockville (1997)
Couper, M. P., Singer, E., Tourangeau, R.: Understanding the effects of audio-CASI on self-reports of sensitive behavior. Public Opin. Q. **67**, 396–406 (2003)
Crowne, D. P., Marlowe, D.: A new scale of social desirability independent of psychopathology. J. Consult. Psychol. **24**, 349–354 (1960)
Crowne, D. P., Marlowe, D.: The Approval Motive. Wiley, New York (1964)
DeMaio, T. J.: Social Desirability and Survey Measurement: A Review. In: Turner, C., Martin, E. (eds.) Surveying Subjective Phenomena, pp. 257–282. Russell Sage, New York (1984)
Dijkstra, W., Smit, J. H., Comijs, H. C.: Using social desirability scales in research among the elderly. Qual. Quan. **35**, 107–115 (2001)
Edwards, A. L.: The Social Desirability Variable in Personality Assessment and Research. Vol. 2, Dryden Press, New York (1957)
Fendrich, M., Johnson, T. P., Sudman, S., Wislar, J., Spiehler, V.: Validity of drug use reporting in a high-risk community sample: a comparison of cocaine and heroin survey reports with hair tests. Am. J. Epidemiol. **149**, 955–962 (1999)
Fleiss, J. L.: Statistical Methods for Rates and Proportions. 2nd edn. Wiley, New York (1981)
Gick, M., McLeod, C., Hulihan, D.: Absorption, social desirability, and symptoms in a behavioral medicine population. J. Nerv. Ment. Dis. **185**, 454–458 (1997)
Gove, W. R., McCorkel, J., Fain, T., Hughes, M. D.: Response bias in community surveys of mental health: systematic bias or random noise?. Soc. Sci. Med. **10**, 497–502 (1976)
Groves, R. M., Couper, M. P.: Nonresponse in Household Interview Surveys. Wiley, New York (1998)
Hser, Y.-I.: Self-reported drug use: results of selected empirical investigations of validity. In: Harrison, L., Hughes, A. (eds.) The Validity of Self-Reported Drug Use: Improving the Accuracy of Survey Estimates, pp. 320–343. National Institute on Drug Abuse, Rockville (1997)
Johnson, T. P., Bowman, P. J.: Cross-cultural sources of measurement error in substance use surveys. Subst. Use Misuse **38**, 1447–1490 (2003)
Klassen, D., Hornstra, R. K., Anderson, P. B.: Influence of social desirability on symptom and mood reporting in a community survey. J. Consult. Clin. Psychol. **43**, 448–452 (1975)
Kozma, A., Stones, M. J.: Social desirability in measures of subjective well-being: a systematic evaluation. J. Gerontol. **42**, 56–59 (1987)
Krosnick, J. A: The impact of cognitive sophistication and attitude importance on response-order and question-order effects. In: Schwarz, N., Sudman, S. (eds.) Context Effects in Social and Psychological Research, pp. 203–218. Springer-Verlag, New York (1992)
Levy, P. S., Lemeshow, S.: Sampling of Populations. Methods and Applications. Wiley, New York (1991)
Linehan, M. M., Nielson, S. L.: Social desirability: its relevance to the measurement of hopelessness and suicidal behavior. J. Consult. Clin. Psychol. **51**, 141–143 (1983)
McCrae, R. R.: Well-being scales do not measure social desirability. J. Gerontol. **41**, 390–392 (1986)
McCrae, R. R., Costa, P. T. Jr.: Social desirability scales: more substance than style. J. Consult. Clin. Psychol. **51**, 882–888 (1983)
Magura, S., Kang, S.-Y.: The validity of self-reported drug use in high risk populations: a meta-analytic review. Subst. Use Misuse **31**, 1131–1153 (1996)
Moorman, R. H., Podsakoff, P. M.: A meta-analytic review and empirical test of the potential confounding effects of social desirability response sets in organizational behaviour research. J. Occup. Organ. Psychol. **65**, 131–149 (1992)
Narayan, S., Krosnick, J. A.: Education moderates some response effects in attitude measurement. Public Opin. Q. **60**, 58–88 (1996)
Office of Applied Studies: Summary of Findings from the 1999 National Household Survey on Drug Abuse. DHHS Publication No. SMA 00-3466, NHSDA Series H-12. Rockville, MD: Substance Abuse and Mental Health Services Administration (2000)
Office of Applied Studies: Development of Computer-Assisted Interviewing Procedures for the National Household Survey on Drug Abuse. DHHS Publication No. SMA 01-3514, Methodology Series: M-3. Rockville, MD: Substance Abuse and Mental Health Services Administration (2001)
Paulhus, D. L: Measurement and control of response bias. In: Robinson, J. P., Shaver, P. R. (eds.) Measures of Personality and Social Psychological Attitudes, pp. 17–59. Academic Press, San Diego (1991)
Paulhus, D. L.: Paulhus Deception Scales (PDS): The Balanced Inventory of Desirable Responding-7, User's Manual. Multi-Health Systems, Inc., North Tonawanda, NY (1998)
Phillips, D. L., Clancy, K. J.: Response biases in field studies of mental illness. Am. Sociol. Rev. **35**, 503–515 (1970)

Phillips, D. L., Clancy, K. J.: Some effects of 'Social desirability' in survey studies. AJS **77**, 921–940 (1972)
Pleck, J. H., Sonenstein, F. L., Ku, L.: Black-white differences in adolescent males' substance use: are they explained by underreporting by Blacks?. J. Gend. Cult. Health **1**, 247–265 (1996)
Roscoe M., Morton R.: "Disproportionate Minority Confinement." Fact Sheet #11. U.S. Department of Justice, Office of Justice Programs, Office of Juvenile Justice and Delinquency Prevention, Washington, DC (1994). Retrieved from October 6, 2003 from http://www.ncjrs.org/txtfiles/fs-9411.txt
Ross, C. E., Mirowsky, J.: The worst place and the best face. Soc. Forces **62**, 529–536 (1983)
Ross, C. E., Mirowsky, J.: Socially-desirable response and acquiescence in a cross-cultural survey of mental health. J. Health Soc. Behav. **25**, 189–197 (1984)
Stone, C.: Race, crime, and the administration of justice. Nat. Inst. Justice J. **239**, 26–32 (1999)
Strahan, R., Gerbasi, K. C.: Short, homogeneous versions of the Marlowe–Crowne social desirability scale. J. Clin. Psychol. **28**, 191–193 (1972)
Strosahl, K. D., Chiles, J. A., Linehan, M. M.: Will the real social desirability please stand up? Hopelessness, depression, social desirability, and the prediction of suicidal behavior. J. Consult. Clin. Psychol. **52**, 449–457 (1984)
Tourangeau, R., Smith, T. W.: Asking sensitive questions: the impact of data collection mode, question format, and question context. Public Opin. Q. **60**, 275–304 (1996)
Turner, C. F., Ku, L., Rogers, S. M., Lindberg, L. D., Pleck, J. H., Sonenstein, F. L.: Adolescent sexual behavior, drug use, and violence: increased reporting with computer survey technology. Science **280**, 867–873 (1998)
Warnecke, R. B., Johnson, T. P., Chavez, N., Sudman, S., O'Rourke, D. P., Lacey, L., Horm, J.: Improving question wording in surveys of culturally diverse populations. Ann. Epidemiol. **7**, 334–342 (1997)
Watten, R. G.: Coping styles in abstainers from alcohol. Psychopathology **29**, 340–346 (1996)
Welte, J. W., Russell, M.: Influence of socially desirable responding in a study of stress and substance abuse. Alcohol. Clin. Exp. Res. **17**, 758–761 (1993)
Wolff, K., Farrell, M., Marsden, J., Montiero, M. G., Ali, R., Welch, S., John, S.: A review of biological indicators of illicit drug use, practical considerations and clinical usefulness. Addiction **94**, 1279–1298 (1999)
Yoshino, A., Kato, M.: Influence of social desirability response set on self-report for assessing the outcome of treated alcoholics. Alcohol. Clin. Exp. Res. **19**, 1517–1519 (1995)