

EVALUATION APPREHENSION, SOCIAL DESIRABILITY, AND THE INTERPRETATION OF TEST CORRELATIONS

CHARLES B. THOMAS AND JUDITH A. HALL

The John Hopkins University

FREDERICK D. MILLER

New York University

JOSEPH R. DEWHIRST

Harvard University

GARY ALAN FINE

University of Minnesota

MARYLEE TAYLOR

Pennsylvania State University

RALPH L. ROSNOW

Temple University

In this paper we investigate the relationship between the concepts of social desirability and evaluation apprehension. The Marlowe-Crowne Social Desirability Scale and the Taylor Manifest Anxiety Scale were administered to 63 Harvard and Radcliffe students. As predicted, there was a moderate negative correlation between social desirability and manifest anxiety in the (anonymous) high evaluation apprehension condition; and a substantially reduced correlation in the (anonymous) low evaluation apprehension condition. Nonanonymous participants also had a lower mean score on the Taylor Manifest Anxiety Scale than did anonymous participants. The results demonstrate a link between evaluation apprehension and social desirability, and indicate the importance of the nature of the testing situation in clinical or applied settings.

Keywords: evaluation apprehension, social desirability, interpretation of test correlations.

Weber and Cook (1972, p. 287) have noted the connection between the concept of evaluation apprehension in psychological experiments (Rosenberk, 1965, 1969) and the concept of social desirability in personality testing. "The apprehensive subject role," they wrote, "is obviously related to the literature on social desirability...though explicit links still need to be made" (see also Rosenberg, 1969, p. 280, footnote.) The present study demonstrates just such an empirical link.

Rosenberg (1965, 1969) introduced the notion that evaluation apprehension can affect participants' behavior in psychological experiments, and can lead to invalid causal

This study was supported by an allocation from the Department of Psychology and Social Relations at Harvard University, where the last author was a visiting professor. The first author thanks Linda Gottfredson, John Hollifield, James McPartland, and especially Robert Slavin at the John Hopkins University for their comments on an earlier draft of this paper.

Correspondence and reprint requests should be addressed to: Charles B. Thomas, Center for Social Organization of Schools, Johns Hopkins University, Baltimore, MD 21218, USA.

inference. Rosenberg (1969, p. 281) defined evaluation apprehension as “an active, anxiety-toned concern that he (the participant) win a positive evaluation from the experimenter, or at least that he provide no grounds for a negative one”. As a result of evaluation apprehension, participants have conformed less in conformity studies and exhibited quicker conditioning in conditioning studies as part of a positive self-presentation (Weber & Cook, 1972). Other evaluation apprehension researchers have shown that when they must make a choice, participants are more concerned with presenting themselves in a favorable light (this has been called the apprehensive participant role) than with helping the experimenter to validate the experimental hypothesis (the good participant role) (Rosnow et al., 1973).

Concern with giving a positive self-presentation is also implicit in the social desirability concept. This concept refers to a tendency to give the socially desired response – e.g., a response that would typically be considered well-adjusted – in answering items on personality measures. This response set is important for personality researchers because it threatens valid interpretation of test results. Crowne and Marlowe (1964), for example, found a moderate negative correlation (-0.25) between scores on the Taylor Manifest Anxiety Scale (TMAS) and the Marlowe-Crowne Social Desirability Scale (MCSDS): participants with a high need to manifest socially desirable traits and behaviors reported themselves to be less anxious (the socially desirable response) than participants with lower MCSDS scores.

Rosenberg (1969) demonstrated an empirical link between evaluation apprehension and social desirability. In his studies (Rosenberg, 1969, pp. 295, 306), participants with high versus low desirability scores on the Marlowe-Crowne Social Desirability Scale were found to be differentially responsive to the experimental manipulation of evaluation apprehension. In a related way, we hoped to demonstrate in this study that the impact of social desirability as a response set affecting personality test scores is mediated by evaluation apprehension. We hypothesized that in a high evaluation apprehension condition there would be a moderately negative correlation (approximating the -0.25 found by Marlowe & Crowne, 1964) between scores on the MCSDS and scores on another personality measure, namely, the Taylor Manifest Anxiety Scale; and this correlation between TMAS and MCSDS would be substantially reduced in the low evaluation apprehension condition. High evaluation apprehension was operationalized as informing participants that they were to be tested nonanonymously; low evaluation apprehension participants were tested anonymously. This was done on the assumption that participants who believe they can be identified should be more concerned with making a positive self-presentation than participants who believe they are anonymous. Silverman (1968) and Rosnow et al. (1973) indicated that anonymity-nonanonymity can be used to effectively manipulate evaluation apprehension.

It is important to note two points about the nature of the hypothesis. The prediction is made for the correlational results in the two conditions (high vs. low evaluation apprehension). No prediction is made regarding possible mean differences. The predicted correlational result can be obtained for any of seven patterns of mean differences: the experimental group (nonanonymous condition) mean may exceed the control group (anonymous condition) mean for (1) both MCSDS and TMAS, (2) neither MCSDS nor TMAS, (3) TMAS alone, or (4) MCSDS alone. The control group mean may

exceed the experimental group mean for (5) both MCSDS and TMAS, (6) TMAS alone, or (7) MCSDS alone. The second important feature of the hypothesis to note is that the maximum difference between the correlation (between MCSDS and TMAS) in the experimental condition and the correlation in the control condition is expected to be about $0.25 - 0.00 = 0.25$. Crowne and Marlowe (1964, Chapter 21) give theoretical reasons why correlations found in this type of research would not be expected to be much more strongly negative than $r = -0.25$.

METHOD

All 33 items of the MCSDS (Crowne & Marlowe, 1964) and all 50 items of the TMAS (Taylor, 1953) were administered to 63 men and women who were students at Harvard and Radcliffe Colleges, enrolled in an intermediate-level undergraduate course. The experimental manipulation consisted of varying instructions concerning anonymity. Half of the participants received booklets in which they were instructed to write their name, sex, and graduating class. Booklets for the other half of the participants contained instructions to write sex and graduating class only. The cover pages of the booklets were identical in both manipulations. To randomize the sample, test booklets were initially stacked so that nonanonymous and anonymous test booklets alternated; booklets were then distributed in this order so that every other student was in the nonanonymous condition. Items from the two scales were randomly interspersed in the booklet.

The experiment was carried out during the lecture hour, without advance notification to the students. While students were not explicitly required to participate in the experiment, no students were seen to leave the room after the announcement of the experiment. However, five participants (four experimental, one control) were later deleted from the sample because they left more than a predetermined number of items blank (more than eight items, or roughly 10%). The final sample therefore consisted of 38 men, 18 women, and 2 participants who did not state their sex.

RESULTS AND DISCUSSION

The hypothesis entertained in this study dealt with the relationship between the dependent variables. For the group as a whole the Pearson product moment correlation between TMAS scores and MCSDS scores was -0.19 . For the nonanonymous group ($n = 27$) this correlation was significant ($r = -0.35$, $p = 0.054$, one-tailed), while for the anonymous group ($n = 31$) the correlation was negligible ($r = -0.04$). Thus, as hypothesized, participants in the high evaluation apprehension condition replicated the relationship originally reported in Crowne and Marlowe (1964), whereas the correlation vanished in the low evaluation apprehension condition.^{2,3}

²The correlation for anonymity accounts for 12.1% less variance than the correlation for nonanonymity. The difference between the two correlations (after transformation of each by Fisher's z), or $q = 0.325$, corresponds to what Cohen (1969) has called a "medium" effect size.

³Reliabilities for the four conditions of the experiment were computed using Cronbach's alpha. For anonymity, MCSDS reliability = 0.76; and for nonanonymity, MCSDS reliability = 0.76. For anonymity, TMAS reliability 0.86; and for nonanonymity, TMAS reliability = 0.90.

We also explored for possible effects on the scale means attributable to the experimental manipulation. Table 1 presents these means; there was no main effect of sex, nor an interaction effect of sex with either variable in Table 1. The cell means were all within the normal range reported for the two scales. By way of comparison, the mean score of a large sample of Iowa students on the TMAS was 14.56 (Taylor, 1953); with regard to the MCSDS, Crowne and Marlowe (1964) reported the range for college students to be 10.06 (Dartmouth men) to 16.04 (University of North Dakota women). Examining the cell means separately for each test revealed no differences on the MCSDS ($t < 1$), but there were fewer anxious symptoms reported by nonanonymous than by anonymous participants on the TMAS ($t = 1.69$, $p < 0.10$, two-tailed test). This noteworthy finding suggests that evaluation apprehension can affect the rate at which anxiety symptoms are self-reported.

TABLE 1
MEANS AND STANDARD DEVIATIONS ON THE TMAS AND MCSDS FOR PARTICIPANTS TESTED
ANONYMOUSLY OR NONANONYMOUSLY

Experimental conditions	TMAS		MCSDS	
	M	SD	MS	D
Anonymity (n = 31)	17.19	7.89	9.55	4.79
Nonanonymity (n = 27)	13.63	8.22	10.18	4.66

Although prediction in this experiment was limited to the correlational result, the pattern of mean differences found is understandable. The correlational results indicate that social desirability operated in the experimental condition (where a moderate negative correlation between MCSDS and TMAS was found) but not the control condition (where a zero correlation was found) - Thus, it is not surprising that participants in the experimental condition reported fewer symptoms of anxiety (the socially desired response) than did participants in the control condition. The finding of no mean difference between experimental and control participants on MCSDS is also what might be expected, since randomization procedures would be expected to produce roughly equal distributions of MCSDS scores (a trait, as opposed to state, variable) for participants in each condition of the experiment.

Some additional issues merit consideration. First, it should be noted that, with a relatively small sample such as used in this research, it is more difficult to obtain a statistically significant result than with a larger sample size (Cohen, 1969). A further question involves the appropriate statistical test to employ. It is inappropriate to use a test for the significance of the difference between two correlation coefficients, because of the constraints on the correlations. The correlation between the tests for the anonymous condition ($r = -0.04$) is as small as it can be; and the correlation for the non-anonymous condition is about as high as it can reasonably be expected to be in this area of research (Crowne & Marlowe, 1964). Our correlation of $r = -0.35$ is actually stronger than the -0.25 correlation found by Marlowe-Crowne. Further, there is a sizable difference in variance explained by the correlation in the experimental condition compared to that explained by the correlation in the control condition.

There is also a matter of interpretation to be considered. Perhaps the results of this experiment are due to a "carelessness factor" rather than to the experimental manipu-

lation. That is, perhaps the MCSDS or TMAS was less valid or internally consistent when taken anonymously than when taken nonanonymously: perhaps participants in the anonymous condition took the test less seriously and responded in a more slipshod fashion. The fact that there were no differences in reliability between experimental and control participants on MCSDS (0.76 vs. 0.76 reliability) or on TMAS (0.90 vs. 0.86) (see footnote 2) counts as evidence against this rival interpretation. Furthermore, the findings gained by Silverman (1968) and Rosnow et al. (1973) and the fact that more experimental than control participants had their data deleted for incomplete responding (see Method section above) support an evaluation apprehension rather than "carelessness factor" explanation.

CONCLUSION

In this study one empirical link between evaluation apprehension and social desirability was demonstrated. It is hoped that the work reported here will serve as a useful point of departure for future work of both an empirical and a theoretical nature. The result reported here also has implications for personality researchers and for those working in clinical or applied settings. In this experiment it was demonstrated that an observed relationship between scores on a specimen personality test and scores on a test of social desirability may be moderated in an important way by the nature of the testing situation (specifically, whether the testing situation was anonymous or nonanonymous). It was also found that evaluation apprehension can affect the rate at which anxiety symptoms are self-reported. The nature of the testing situations should be carefully considered when: (1) interpreting the correlation between a personality test and a measure of social desirability; and (2) interpreting the actual scores on personality measures (such as the TMAS) which seem strongly susceptible to the influence of social desirability.

REFERENCES

- Cohen, J. (1969). *Statistical power analysis for the behavioral sciences*. New York: Academic Press.
- Crowne, D., & Marlowe, D. (1964). *The approval motive: Studies in evaluative dependence*. New York: Wiley.
- Rosenberg, M. J. (1965). When dissonance fails: On eliminating evaluation apprehension from attitude measurement. *Journal of Personality and Social Psychology*, 1, 28-42.
- Rosenberg, M. J. (1969). The conditions and consequences of evaluation apprehension. In R. Rosenthal & R. L. Rosnow (Eds.), *Artifact in behavioral research*. New York: Academic Press.
- Rosnow, R., Goodstadt, B., Suls, J., & Gitter, A. (1973). More on the social psychology of the experiment: When compliance turns to self-defense. *Journal of Personality and Social Psychology*, 27, 337-343.
- Silverman, I. (1968). Role-related behavior of subjects in laboratory studies of attitude change. *Journal of Personality and Social Psychology*, 8, 343-348.
- Taylor, J. (1953). A personality scale of manifest anxiety. *Journal of Abnormal and Social Psychology*, 48, 285-290.
- Weber, S. J., & Cook, T. D. (1972). Subject effects in laboratory research: An examination of subject roles, demand characteristics, and valid inference. *Psychological Bulletin*, 77, 273-295.