

## Assessing Marriage via Telephone Interviews and Written Questionnaires: A Methodological Note

*Telephone interviews are becoming an increasingly popular method of data collection in marital research. The present study examines whether the assessment of marital adjustment and perceived responsibility for marital events yields comparable results across telephone interviews and written questionnaires. Although the rank ordering of scores on both measures was consistent across these two administrative conditions, mean marital adjustment scores obtained on questionnaires were significantly lower than those obtained in the telephone interviews. Mean scores for the responsibility judgments did not differ significantly across response formats. The implications of these findings for different types of studies are discussed.*

The use of the telephone interview has become more frequent in marital research (e.g., Fincham & Bradbury, 1989; Krokoff, 1989). Telephone interviews are used for many purposes, including the collection of data on marriage and the recruitment of samples for laboratory research (e.g., Booth & Johnson, 1985; Krokoff, 1987). Despite the modification of paper and pencil measures for use in telephone interviews, there has been little research investigating the comparability of inter-

view and questionnaire data. Scores across these two administrative formats can be compared in terms of both their relative position in the sample and their absolute values. *Consistency in relative position* across administrative formats is important when correlational analyses are performed, since relative position in a sample influences obtained correlations. *Consistency in actual scores* across administrative formats is important when cutoff points are used to categorize individuals into groups. The present study therefore examines two measures which were used in recent telephone interviews by comparing relative and absolute scores obtained from telephone interviews and written questionnaires.

Several factors may differentially influence responses to the same or equivalent questions administered in interview and questionnaire formats. First, demand characteristics can be greatly increased when individuals respond orally to an interviewer rather than privately complete a questionnaire (Campbell & Stanley, 1963). For example, respondents may attempt to please the experimenter during a telephone interview because they see the experimenter as an authority figure. Respondents may also be influenced to a greater extent by experimenter expectancies during interviews than when completing a written questionnaire (Rosenthal, 1966).

Second, an interview format may rush the respondent to make a judgment more quickly than

if he or she were completing a written questionnaire. Respondents completing questionnaires often have time to look over their responses and make sure that they are presenting a consistent picture of themselves, while those completing interviews do not have the luxury of reviewing their responses.

Third, the tendency to respond in a socially desirable manner may increase when data are gathered over the telephone as compared to the completion of a written questionnaire. A decreased sense of anonymity when completing a telephone interview may lead respondents to feel greater evaluation apprehension and therefore present themselves in a psychologically healthy manner (Rosenberg, 1969). This impression management is not necessarily deliberate deception on the part of the respondent (Edwards, 1957) but it may nevertheless influence responses.

Fourth, because of the nature of some written questionnaires, some questions and response options must be reworded when administered in telephone interviews. Groves and Kahn (1979) noted that differing response formats in face to face interviews yielded significantly different results. The effects of altering the response formats of questionnaires for use in telephone interviews is largely unknown.

Finally, individuals differ in terms of their comfort with interpersonal interactions and this level of comfort is likely to influence the results obtained. For example, participants who are uncomfortable with an interview may be motivated to end it as soon as possible and may therefore be less attentive and more hurried in their responses.

In view of these observations, telephone and written questionnaire responses to the Marital Adjustment Test (MAT; Locke & Wallace, 1959) and a perceived responsibility questionnaire (Fincham & Bradbury, 1989) were investigated. These two measures were chosen because they recently have been adapted to allow data to be collected over the telephone. Although both telephone interview and questionnaire formats have been used for each measure, direct comparisons across these administrative conditions have been quite limited.

Krokoff (1989) was able to demonstrate that a telephone version of the MAT administered to wives successfully predicted their own and their husbands' marital adjustment 6 and 12 months later. The moderately high correlations between MAT scores measured first by telephone and later

in writing suggest that the *relative* position of individuals within the sample was consistent across the two administrative formats. Krokoff's study, however, did not address questions about possible *absolute* differences in scores on the MAT measured under the two procedures. In addition, Krokoff did not interview husbands and it is therefore not known whether gender differences exist in responses obtained across interview and questionnaire responses. The present study therefore examines possible gender differences.

As regards perceived responsibility, Fincham and Bradbury (1989) adapted a paper and pencil measure for telephone administration and found that the results from the telephone interviews yielded findings which were consistent with those documented for written questionnaires (e.g., Ross & Sicoly, 1979; Thompson & Kelley, 1981). However, no direct comparison was made between telephone interview and written questionnaire assessments of perceived responsibility. Because earlier research had shown that small changes in the wording of questions can significantly influence the allocation of responsibility (Ross & Sicoly, 1979; Study 5), Fincham and Bradbury (1989) recommended investigation of the relationship between interview and questionnaire ratings of responsibility for marital activities.

In sum, the present study compares scores on two measures across telephone interview and written questionnaire administration formats. The goal of the study is to examine the comparability of scores obtained via these two methods. In pursuing this goal we extend previous research by (a) investigating absolute as well as relative changes in scores and (b) examining gender differences. Finally, the inclusion of two different measures allows us to examine whether any differences obtained result from a general method effect or are measure specific.

## METHOD

### *Participants*

Fifty-six married couples were selected from 258 couples recruited for a larger project. The couples were recruited by means of advertisements in local newspapers. The sample used in this paper were selected to represent a wide range of marital quality. Marital quality was assessed by responses on a 10-point scale to a question asking respondents to

indicate "the degree of happiness, everything considered, in your marriage."

The 56 couples who participated in this study were married an average of 10.7 years ( $SD = 9.8$ ), had 1.0 children ( $SD = 1.2$ ), and earned a median family income of \$25,000 to \$30,000. The sample consisted of couples living in rural areas and small cities in a midwestern state. Wives averaged 33.6 ( $SD = 10.4$ ) years of age, 14.6 ( $SD = 2.3$ ) years of education, and scored 97.5 ( $SD = 29.0$ ) on the written MAT. Corresponding figures for husbands were 34.8 ( $SD = 11.5$ ) years of age, 14.6 ( $SD = 2.5$ ) years of education, and 103.2 ( $SD = 24.7$ ) for the MAT.

#### *Procedure*

Each spouse independently completed a telephone interview following which they were mailed questionnaires. The time between completion of the interview and the questionnaires averaged 5.26 days ( $SD = 4.25$ ), a period over which the stability of MAT and perceived responsibility scores was expected to be high. Each member of the couple was interviewed separately on the telephone. Care was taken to ensure that the partner could not hear the spouse being interviewed on the telephone. Similarly, couples were instructed to complete the questionnaires independently of one another and were given separate envelopes in which they sealed their responses. In the interview several demographic questions were asked first, followed by questions from the MAT and then the perceived responsibility measure. In the questionnaire packet, the MAT also preceded the perceived responsibility ratings.

#### *Measures*

**Marital quality.** The Marital Adjustment Test (Locke & Wallace, 1959) is a widely used measure of marital satisfaction that yields a score ranging from 2 to 158. It has adequate reliability (split half = .90) and discriminates between nondistressed spouses and spouses who have documented marital problems (Locke & Wallace, 1959). Scores on this instrument also correlate with clinicians' judgments of marital discord (Crowther, 1985). A score of 100 on the MAT is widely accepted as a cutoff for separating distressed from nondistressed spouses and is often used to form distressed and nondistressed subsamples.

The telephone administration of the MAT followed the procedures outlined by Krokoff (1989) which entailed rearrangement of the order of questions and minor rewording of some questions. The order of questions replicated that used by Krokoff (1989) and was deemed necessary so that the most threatening (personal) questions were asked last (those regarding demonstrations of affection and sexual relations). The minor rewording consisted of rephrasing open-ended statements into conversational forms (e.g., Written form: "When disagreements arise, they usually result in: husband giving in, wife giving in, agreement by mutual give and take"; Interview form: "When disagreements come up between you and your husband/wife, who usually gives in? Is it . . . your husband/wife, is it you, or do you reach agreement by mutual give and take?").

**Responsibility judgments.** Judgments of responsibility for marital activities were assessed with a questionnaire consisting of 12 commonly occurring events which had been used by Fincham and Bradbury (1989). Six items were positive (e.g., "How much do you contribute to planning joint activities for the two of you?") while the other six were negative (e.g., "How much do you contribute to causing arguments that occur between the two of you?"). Respondents described their perceptions of responsibility for each behavior on a 9-point Likert scale ranging from "primarily me" to "primarily partner." The administration of these questions over the telephone followed the procedure outlined by Fincham and Bradbury (1989), in which respondents were told to write down the numbers 1 through 9 on a piece of paper and put the words "primarily partner" next to the 1 and "primarily me" next to the 9. Respondents then referred to the scale when answering each of the 12 questions regarding their perceived contribution.

#### RESULTS

Scores on the MAT and the responsibility measure are compared across the two methods of data collection in terms of the consistency of relative score position and in terms of absolute score differences. Where significant differences are observed, further analyses are conducted to examine the nature of the differences between the two administrative formats.

TABLE 1. MARITAL ADJUSTMENT TEST SCORES AND RESPONSIBILITY RATINGS (IN PARENTHESES) AS A FUNCTION OF GENDER AND ADMINISTRATIVE CONDITION

Administrative Condition	Males		Females	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Telephone interview	103.2 (5.20)	24.7 (.63)	97.5 (5.49)	29.0 (.57)
Written questionnaire	99.9 (5.20)	24.0 (.57)	93.3 (5.44)	28.0 (.69)

To determine the consistency of respondents' relative positions (rank order) across the two administrative conditions, correlation coefficients were computed. As expected, the correlation between total MAT scores from the interview and written administrations was high for both husbands and wives ( $r = .86$  and  $.93$ , respectively), indicating that spouses, in large part, maintained their relative position within the sample. For responsibility judgments, scores from the interview and written responses correlated moderately for both husbands and wives ( $r = .44$  and  $.73$ , respectively). However, a significant gender difference emerged in the rank ordering of scores across methods of data collection ( $z = 2.12$ ,  $p < .05$ ).

To determine whether absolute differences in scores emerged across the two types of administrations, 2 (telephone  $\times$  questionnaire)  $\times$  2 (husband  $\times$  wife) analyses of variance were conducted for both the MAT and the perceived responsibility measures. Because the responses of each spouse in a couple are not considered independent in the marital literature, both factors were treated as repeated measures.

For marital quality, significant main effects were obtained for gender,  $F(1,55) = 4.02$ ,  $p < .05$ , and for administrative condition,  $F(1,55) = 10.34$ ,  $p < .002$ ; on average, husbands were more satisfied than wives and greater satisfaction was expressed over the telephone than in written responses (see Table 1). The interaction of gender by administrative condition was not significant,  $F(1,55) = 0.17$ ,  $p < .69$ .

As regards responsibility judgments, a significant gender main effect was found,  $F(1,55) = 2.60$ ,  $p < .03$ ; on average, wives saw themselves as making greater contributions to marital events than husbands (see Table 1). No significant difference was obtained between responses to interview and written questionnaire formats.

The significant main effect of administrative format on average MAT scores led us to examine whether differences in scores across the two formats changed individuals' classification as dis-

tressed versus nondistressed spouses. Using an MAT score of 100 as the cutoff point, individual score differences were sufficient to change 9.8% of the sample's classification, that is, 11 of 112 participants. Five spouses who had been classified as nondistressed during the telephone interview were classified as distressed from their completion of the written form of the MAT (average change score = 17.6; range = 9-28). The other 6 individuals had been classified as distressed during the telephone interview and scored in the nondistressed range on the written questionnaire (average change score = 12.5; range = 6-25).

In view of the significant difference found for administrative condition on mean MAT scores, an attempt was made to determine the source of this difference by examining each of the 15 items on the MAT. After family-wise error rates had been corrected ( $p < .05/15$  items  $< .003$ ), individual item analyses using 2 (telephone  $\times$  questionnaire)  $\times$  2 (husband  $\times$  wife) repeated measures analyses of variance indicated that scores were significantly different across interview and questionnaire formats on the global evaluative item, "How happy are you with your marriage, everything considered?",  $F(1,50) = 9.44$ ,  $p < .003$ , and on the question, "Do you confide in your mate . . . almost never, in some things, in most things, in everything?",  $F(1,55) = 15.42$ ,  $p < .001$ . Scores on all other items of the MAT did not result in significant differences across administrative formats. No significant interactions between administrative format and gender were found in the individual item analyses.

#### DISCUSSION

The present study indicates that the rank order of MAT scores is highly consistent across telephone interview and written questionnaire formats. This consistency in relative scores is comparable to that reported by Krokoff (1989) and was expected, given the trait-like connotations of marital quality discussed in the literature. The present study also

extends Krokoff's (1989) findings by showing that they apply to husbands as well as wives. These results suggest that scores from the two administrative formats are likely to yield comparable results in correlational research on marital satisfaction.

Despite the high correlations between MAT scores across the two administrative formats, a significant difference in absolute scores was found in that the average MAT score obtained over the telephone was significantly greater than that obtained from written responses. Although the data do not indicate why higher average scores were obtained on the administration of the telephone interview, it is possible that they result from factors such as respondents' attempts to please the experimenter, increased awareness of experimenter expectancies, and lower sense of anonymity during the interview. However, the nature of the differences in mean scores across administrative conditions was not found consistently across individuals. Our analysis of changes in individual MAT scores indicates that some participants' scores increased from telephone interview to questionnaire while others decreased. The varying nature of the changes in individual scores may result from individual differences in the tendency to answer in a socially desirable manner and in levels of comfort with interpersonal interactions that respondents experienced.

Absolute score differences across administrative formats, regardless of their direction, are particularly important because the MAT is frequently used to categorize individuals into distressed and nondistressed subsamples. In the present study, nearly 10% of the sample is classified differently by the data from the two methods of MAT administration. Absolute score differences across administrative formats will influence all types of classification into subsamples, even those which are based upon sample distributions rather than absolute cutoff points (e.g., median or tertile splits). Because of this, all procedures which use MAT scores to classify respondents into separate categories should be considered with caution, as the subsamples may not be comparable to those formed using an alternative method of data collection.

For the responsibility judgments, correlations obtained between telephone and written administrations showed moderate consistency in relative scores. However, unlike MAT scores, husbands' and wives' responsibility judgments differed signi-

ficantly in the extent to which their rank order correlates across the two administration formats. This finding is consistent with the discovery of numerous sex differences in marital research (Baum, Notarius, Burnett, & Haefner, 1990). Although intriguing, the reason for this difference is unclear. One possible explanation is that women place greater value on relationships than men and therefore attend to relationship events more carefully, leading to greater consistency in reports about such events. Finally, no absolute difference was observed across the two administrative modes in average responsibility scores. In sum, it appears that correlational analyses of responsibility judgments may be influenced by the manner in which data are collected more for husbands than for wives.

Because differences in absolute scores across types of administration occurred for the MAT but not for the responsibility judgments, the findings are unlikely to reflect a general method effect. One factor that may account for the differing results obtained across measures is the nature of the questions used in each measure. It is possible that the average MAT score differs across administrative conditions because the MAT includes questions that are more global and have less clear referents. Owing to their complexity, such questions may be more strongly influenced by the time demands of the interview situation.

In summary, the present study shows that scores from telephone and written administrations of the MAT are highly correlated. Although a mean difference in absolute scores was found, the high correlation between scores from the two administrative conditions suggests that data obtained through telephone and written administrations of the MAT can be used in place of questionnaire data when classification of respondents into subsamples is not an objective of the research. Perceived responsibility judgments are moderately correlated across telephone and interview formats in terms of rank order and absolute values, especially for wives. This result suggests that the collection of perceived responsibility data via telephone interviews is a reasonable alternative to questionnaire data.

In generalizing to a larger methodological question regarding the consistency of data collected via telephone interviews and written questionnaires, this study suggests that telephone interviews may be an appropriate alternative to

written questionnaires. However, the viability of this alternative may depend on the nature of the measure as well as the manner in which the data are used. The present study suggests that the administrative format is most likely to influence measures containing global questions that lack clear referents and data analyses that involve examination of absolute scores. Until the factors that influence responses across telephone interview and questionnaire response formats are investigated further, it seems most prudent to demonstrate consistency of responses across data collection formats before accepting their equivalence.

## NOTE

This study was conducted as part of a master's thesis completed by the first author at the University of Illinois. Preparation of this paper was supported by a Faculty Scholar Award from the W. T. Grant Foundation, a grant from the Harry Frank Guggenheim Foundation, and grant R01 MH44078-01 from the National Institute of Mental Health awarded to the second author. The authors wish to thank the two anonymous reviewers for their helpful comments in revising this manuscript. Correspondence should be addressed to Susan Gano-Phillips.

## REFERENCES

- Baucum, D. H., Notarius, C. I., Burnett, C. K., & Haefner, P. (1990). Gender differences and sex-role identity in marriage. In F. D. Fincham & T. N. Bradbury (Eds.), *The psychology of marriage: Basic issues and applications* (pp. 150-171). New York: Guilford Press.
- Booth, A., & Johnson, D. R. (1985). Tracking respondents in a telephone interview panel selected by random digit dialing. *Sociological Methods and Research, 14*, 53-64.
- Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Boston: Houghton Mifflin.
- Crowther, J. H. (1985). The relationship between depression and marital adjustment: A descriptive study. *The Journal of Nervous and Mental Disease, 173*, 227-231.
- Edwards, A. L. (1957). *The social desirability variable in personality assessment and research*. New York: Dryden.
- Fincham, F. D., & Bradbury, T. N. (1989). Perceived responsibility for marital events: Egocentric or partner-centric bias? *Journal of Marriage and the Family, 51*, 27-35.
- Groves, R. M., & Kahn, R. L. (1979). *Surveys by telephone: A national comparison with personal interviews*. New York: Academic Press.
- Krokoff, L. J. (1987). A method for recruiting representative marital samples. *Journal of Social and Personal Relationship, 4*, 317-328.
- Krokoff, L. J. (1989). Predictive validation of a telephone version of the Locke-Wallace Marital Adjustment Test. *Journal of Marriage and the Family, 51*, 767-775.
- Locke, H. J., & Wallace, K. M. (1959). Short marital adjustment prediction tests: Their reliability and validity. *Marriage and Family Living, 21*, 251-255.
- 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.
- Rosenthal, R. (1966). *Experimenter effects in behavioral research*. New York: Appleton-Century-Crofts.
- Ross, M., & Sicoly, F. (1979). Egocentric biases in availability and attribution. *Journal of Personality and Social Psychology, 37*, 322-336.
- Thompson, S. C., & Kelley, H. H. (1981). Judgments of responsibility for activities in close relationships. *Journal of Personality and Social Psychology, 41*, 469-477.