Journal of Consulting and Clinical Psychology
1981, Vol. 49, No. 2, 269-277

0022-006X/81/4902-0269$00.75

Spouses as Observers of the Events in Their Relationship

Neil S. Jacobson
University of Washington

Danny Moore
University of Florida

This study concerns itself with the reliability of spouses as observers of the behaviors that occur in their own marital relationships. Distressed and nondistressed couples were paid to collect data in the home on 21 consecutive evenings. Once per day, spouses completed a behavioral checklist in which they independently indicated which of 409 behaviors had occurred during the preceding 24 hours. Consensus was measured by calculating both percentage agreement and kappa statistics. The average agreement for all couples was 47.8%. Across the entire checklist, nondistressed couples exhibited significantly greater consensus than did distressed couples, based on both percentage agreement and kappa. However, when comparisons were made on selected individual categories of behavior, differences between distressed and nondistressed couples held up only for percentage agreement. Suggestive evidence was presented that inferential items were less reliably coded than noninferential items. Discussion of these results centers on the implications for viewing spouse-collected data as observational data, possible innovations in data collection procedure that could result in more reliable recording, and theoretical implications of the low rate of consensus between spouses.

The hallmark of behavioral assessment is the direct observation of relevant behavior in the environment of interest (Goldfried & Sprafkin, 1974). Unfortunately, there are often practical problems associated with the implementation of observational assessment strategies in the natural environment, in addition to knotty methodological issues such as observer bias and reactivity (cf. Johnson & Bolstad, 1973). These obstacles to direct observation have led many investigators to compromise this behavioral assessment ideal by adopting strategies such as behavioral interviews (Linehan, 1977), self-monitoring (Ciminero, Nelson, & Lipinski, 1977), and laboratory analogue observation procedures.

In the assessment of marital interaction, investigators are blessed with an advantage that has the potential to mitigate these problems. Each spouse in a marital relationship is equipped with a ready-made observer to all behavioral exchanges—the partner (Jacobson, Elwood, & Dallas, 1981; Margolin & Jacobson, in press; Weiss & Margolin, 1977). The utilization of spouses as observers allows for an ongoing record of all transactions that occur in a marital relationship. Already, in the short history of behavioral research on marital interaction, spouse observation has become a common assessment strategy. Spouse observations have been utilized as outcome measures in investigations of the efficacy of behavioral marital therapy (Jacobson, 1977, 1979; Margolin & Weiss, 1978). The discriminant validity of spouse observations has already been demonstrated: Distressed and nondistressed couples can be distinguished on the basis of the frequency of pleasing and displeasing spouse behaviors (Barnett & Nietzel, 1979; Birchler, Weiss, & Vincent, 1975; Margolin, in press). Finally, spouse

Portions of this paper were presented at the meeting of the Western Psychological Association, Honolulu, Hawaii, May 1980. The study was conducted while both authors were at the University of Iowa. Support was provided by a University of Iowa Seed Grant and by National Institute of Mental Health Grant 1 RC3 MH32085-01, both awarded to the first author. Preparation of the manuscript was supported by grants from the Graduate School Research Fund at the University of Washington and by NIMH Grant 6 RO1 MH33722-01, both awarded to the first author.

The authors wish to acknowledge invaluable assistance in the collection and reduction of data provided by Marcia Stamps, Holly Waldron, Judy Perlman, Mercedes Dallas, Erling Anderson, and Jane Deichler.

Requests for reprints should be sent to Neil Jacobson, Department of Psychology NI-25, University of Washington, Seattle, Washington 98195.

observations have been used to test hypotheses derived from behavioral models of marital satisfaction (Jacobson & Moore, in press; Jacobson, Waldron, & Moore, 1980; Vincent, Cook, & Messerly, 1980; Wills, Weiss, & Patterson, 1974).

Research utilizing spouse observations presupposes that we can trust spouses to report accurate, reliable data. However, there is no reason to presume that spouses reliably report on the events of their relationship. Robinson and Price (1980) provided suggestive evidence that spouses are not inclined to agree with observers regarding the frequency of positive interactional events. Moreover, Robinson and Price (1980) found that distressed couples were significantly less inclined to agree with observers than were their nondistressed counterparts. If distressed couples are significantly less reliable than nondistressed couples, all studies that compare distressed and nondistressed couples are suspect, since any behavioral differences are confounded with differential accuracy.

Unfortunately, it is difficult to draw conclusions from the Robinson and Price (1980) study due to the small sample size, the measurement of a limited range of behaviors, and the use of an observational coding system that requires extensive training for reliable coding. The Robinson and Price study may underestimate spouses' true potential for accurate data collection.

Christensen and Nies (1980) recently reported another assessment of spouse reliability, measuring a larger pool of behaviors and comparing the extent of agreement between spouses. The results were once again discouraging, with percentage agreement statistics well below minimally acceptable standards for observational research. However, this study may not be a fair test for assessing spouses' capacity for reliable data collection, since each couple reported on relationship events for 1 day only. It is not particularly surprising that agreement percentages were low, just as one would expect any pair of observers to be unreliable initially. Moreover, Christensen and Nies assessed spouse accuracy by bringing couples into the laboratory and asking them to recount the events of the previous 24 hours retrospectively, without advance warning. Thus, they were asked to report on a period of time for which there was no reason for them to have been carefully tracking relationship behaviors.

The present study provided a fair yet stringent test of interspouse reliability by having spouses observe and record behaviors over a number of days. It also compared the interspouse agreement of distressed and nondistressed couples.

In addition to the investigation of interspouse reliability, other issues were explored in a preliminary manner in the present study. One issue concerns the characteristics of spouses as observers along with their rate of agreement on the occurrence or nonoccurrence of relationship behaviors. Within the constraints of our experimental procedures, we attempted to determine when and why disagreements occur and whether these disagreements can be accounted for by the behaviors being observed.

Method

Subjects

Thirty six couples were selected as subjects for the study. They were selected by advertising in local newspapers and by articles in newspapers around the state of Iowa promoting the research. Couples were paid $20 each for their participation; in addition, during a debriefing following their participation, they were presented with a profile of their relationship based on the data that they provided for the study.

Of the 36 couples, 20 were designated as nondistressed and 16 as distressed based on their scores on the Dyadic Adjustment Scale (DAS; Spanier, 1976). A cutoff score of 100 was used to differentiate the two categories of couples. If at least one of the two spouses scored below 100, the couple was designated as distressed. Mean scores for distressed wives and husbands were 82.9 and 80.6, respectively, whereas the corresponding totals for nondistressed spouses were 118.3 and 119.2. Differences between distressed and nondistressed spouses were highly significant, $t(35) = 6.80$, $p < .001$.[1]

[1] It should be noted that although DAS scores indicated that the couples in our distressed sample were significantly distressed, they were not seeking clinical assistance, and as a result, they cannot be considered a clinical population. However, it is our belief that they resembled a clinical population on most important dimensions, and, in fact, one third of the sample volunteered because they were looking for help. They were either treated in the Psychology Department Clinic (four couples) or referred to therapists in the community (two couples).

There were no significant differences between distressed and nondistressed couples in age, years of education, duration of marriage, or number of children. Husbands averaged 34.6 years old; wives averaged 34.0. In about half of the cases, both spouses had completed at least a bachelor's level of education; in a large majority (84%) of the couples, at least one spouse was college educated. Couples had been married for an average of 10.1 years, and they had an average of 1.7 children.

Measures

Data collection involved use of the Spouse Observation Checklist (SOC; Patterson, 1976; Weiss & Perry, Note 1). The SOC, consisting of 409 items, is a comprehensive list of potential events that can occur in a marital relationship. Items on the SOC are divided into twelve content categories: companionship, affection, consideration, sex, communication, coupling activities, child care and parenting, household responsibilities, financial decision making, employment and education, personal habits and appearance, and self and spouse independence. The checklist is described in detail elsewhere (Jacobson, Elwood, & Dallas, 1981; Jacobson & Margolin, 1979; Margolin & Jacobson, in press; Weiss, Hops, & Patterson, 1973; Weiss & Margolin, 1977), including its psychometric properties. The SOC has been shown to be sensitive to marital distress both in its ability to discriminate between distressed and nondistressed couples (Barnett & Nietzel, 1979; Birchler et al., 1975) and its ability to predict subjective marital satisfaction within couples on a day-by-day basis (Jacobson, Waldron, & Moore, 1980; Margolin, in press; Wills et al., 1974). The measure has also been shown to be sensitive to improvements in marital satisfaction resulting from behavioral marital therapy (Jacobson, 1979; Margolin & Weiss, 1978).

The SOC consists primarily of items that require each spouse to track the behavior of the partner (e.g., "Spouse complimented me"). The exceptions to this general characteristic involve certain items comprising joint activities (e.g., "We watched television together"). For the purposes of this study, a companion checklist was designed called the Self-Monitoring Checklist (SMC), which consists primarily of items requiring each spouse to track his or her own behavior (e.g., "I complimented my partner"). In construction of the SMC, all joint activity items on the SOC were retained, and the wording of all other items was altered so that instead of reporting on the other's behavior, each spouse was reporting on his or her own.

Procedure

Participating couples were invited in groups of five to the lab, where they were instructed in the use of both the SOC and SMC. The basic task involved nightly data collection for 21 consecutive evenings. Using either the SOC or SMC, at a designated time each evening, each spouse was required to complete a checklist independently. Specifically, spouses were instructed to recall the previous 24-hour period and simply note each item on the checklist that had occurred during the past 24 hours. As a check to make sure that spouses were collecting the data each evening, experimenters telephoned subjects at night to collect summaries of the data.

The SOC and SMC were combined in the following way. For the first 6 nights of data collection, one spouse was randomly assigned to complete the SOC, and the other completed the SMC. Thus, on these first 6 evenings, both spouses were tracking the behavior of the same person. Then, for an intervening 9-day period, each spouse completed the SOC. Finally, for the final 6 days, tasks were the reverse of what they had been for the first 6 days; that is, the spouse who had been completing the SOC during the first 6 days switched to the SMC, and vice versa.

This study is concerned with the 12 days in which both spouses were tracking the behavior of the same person (Days 1–6 and 16–21). Results are reported only for those 12 days.

Results

Overall Consensus

The primary purpose of this study was to examine the degree of consensus between spouses regarding the relationship behaviors that had occurred during a 24-hour period. To determine degree of consensus for a particular couple, percentage agreement statistics were computed. These were based on the following formula: (number of items simultaneously checked by both spouses on the same day/number of items checked by both spouses + items checked by one spouse but not by the other) × 100, that is, agreements/agreements + disagreements. The resulting percentage agreement statistics were then averaged across 33 couples. (Three of the original sample had to be dropped because of incomplete data). The mean agreement was 47.8%; couples ranged from 31% agreement to 78.6% agreement. Thus, on the average, spouses agreed slightly less than half of the time regarding whether a particular event had occurred. These figures are well below the minimum acceptable reliability standards for observational research.

Overall mean percentage agreement statistics were separately calculated for distressed and nondistressed couples. Distressed couples agreed on an average of 42%, whereas nondistressed couples averaged 52% agreement. To test the significance of these differences, an analysis of variance was conducted, with the overall percentage agreement statistic for each couple as the unit of analysis. Level of distress served as

the between-subjects factor, with sex of spouse and days (Days 1–6 vs. Days 16–21) as within-subject factors. There were main effects for level of distress, $F(1, 29) = 5.25$, $p < .05$, and for days, $F(1, 29) = 4.56$, $p < .05$. None of the interactions even approached significance. In general, nondistressed couples achieved greater consensus than their distressed counterparts, and couples tended to reach greater consensus during the second data collection period than they did during the first (Days 1–6 $M =$ 46%; Days 16–21 $M = 50\%$). However, it should be noted that agreement rates were unacceptably low even for nondistressed couples and even during the second data collection period.

As low as these percentage agreement statistics are, they may actually be inflated, since there is a certain probability that spouses will agree on the occurrence of a particular behavior simply by chance. Although the probability of chance agreement (Pc) is generally low for a checklist with 409 items, Pc increases as the total number of endorsed items increases. Once it was discovered that nondistressed couples tended to endorse more items than did distressed couples,[2] another analysis was conducted comparing distressed and nondistressed couples after Pc had been partialed out. The kappa statistic was used for this purpose (Hartman, 1977; Hollenbeck, 1978). The mean kappas for distressed and nondistressed couples, respectively, were .38 and .47. This difference remained statistically significant, $F(1, 29) = 4.44$, $p < .05$, as did the main effect for days, $F(1, 29) = 6.33$, $p < .02$. Thus, even with a statistic that removed any spurious advantage attained by nondistressed couples due to their greater frequency of item endorsement, they still produced agreement rates that were significantly higher than those produced by distressed couples. Again, however, it should be noted that the magnitude of agreement was unimpressive even for nondistressed couples.

Analysis of Individual Behavioral Categories

To shed light on the variables influencing degree of consensus, analyses were conducted for selected individual behavioral categories from the SOC. Mean percentage agreement statistics and kappa values were calculated for distressed and nondistressed couples, and analyses of variance were used to determine the significance of differences between distressed and nondistressed couples. Only main effects for the distress factor are discussed here.

Table 1 lists the mean percentage agreement statistics and kappa values for distressed and nondistressed couples for five behavioral categories. These categories were chosen for further analysis, because they were the only ones for which events occurred with sufficient frequency to justify such an analysis.[3]

As Table 1 shows, for all five of the analyzed categories, consensus was higher for nondistressed than it was for distressed couples. This was true for both percentage agreement and kappa. The differences were statistically significant for three of the five comparisons based on percentage agreement. However, comparisons based on kappa approached statistical significance for only one of the five categories. Since kappa reflects consensus after Pc has been partialed out, it is likely that a substantial proportion of the differences in percentage agreement between distressed and nondistressed couples is artifactual, attributable to the tendency on the part of nondistressed couples to report higher frequencies of positive behaviors. Analyses of variance based on the frequency of reported pleasing behaviors for these five individual categories support this interpretation. These analyses are summarized in Table 2. For four of the five categories, nondistressed couples reported significantly higher frequencies of the behaviors

[2] Distressed couples endorsed an average of 44.2 items per day, compared with 45.2 for nondistressed couples. Although this difference was not significant, $F(1, 29) = 3.20$, $p < .10$, there was a statistically significant correlation between level of distress and number of items endorsed by at least one spouse ($r = .25$, $p < .05$).

[3] For all other categories, two or more couples failed to report even a single occurrence of items in that category. It should also be noted that all five of the categories included in Table 1 constitute groupings of pleasing behavior, that is, behaviors that the creators of the SOC assumed to be generally pleasing when they occur. Displeasing behaviors in each category occurred with insufficient frequency to justify category analysis.

Table 1
Percentage Agreement and Kappa Values for Selected Behavioral Categories

Category	% agreement			Kappa		
	Distressed	Nondistressed	$F(1, 29)$	Distressed	Nondistressed	$F(1, 29)$
Companionship	.59	.63	*ns*	.57	.61	*ns*
Affection	.56	.72	10.79****	.41	.50	*ns*
Sex	.60	.67	*ns*	.55	.59	*ns*
Consideration	.36	.46	5.26**	.30	.37	*ns*
Communication	.32	.46	7.74***	.21	.30	4.15*

* $p < .06$. ** $p < .05$. *** $p < .01$. **** $p < .005$.

in question than did distressed couples. Moreover, correlation coefficients examining the relationship between frequency of reported occurrence of behaviors and consensus, summarized in Table 3, reflect a significant relationship between percentage agreement and reported frequency for four of the five categories. On the other hand, none of the correlations attained significance when kappa was used as a measure of consensus.

To summarize, although nondistressed couples tended to reach greater consensus than their distressed counterparts on a number of behavioral categories, these differences may be largely artifactual. It is probable that despite the artifact, there is a slight tendency for nondistressed couples to agree on the occurrence of behaviors at a higher rate; this tentative conclusion is based on the significant difference in overall consensus based on kappa as well as the consistently higher kappa values for nondistressed couples even where significance was not attained. However, what is most striking about these findings is the generally low degree of consensus for all couples rather than the differences between distressed and nondistressed couples.

Table 2
Mean Daily Frequencies of Pleasing Behaviors for Distressed and Nondistressed Couples

Category	Distressed couples	Nondistressed couples	$F(1, 29)$
Companionship	2.71	2.83	*ns*
Affection	3.58	5.21	9.33**
Sex	3.13	4.86	4.72*
Consideration	6.85	9.24	5.13*
Communication	4.17	5.79	5.33*

* $p < .05$. ** $p < .005$.

The examination of consensus for individual categories sheds some light on another variable of potential significance. Prior to the analysis of data, categories were ranked a priori on the basis of the degree of inference required by the observer in order to determine whether an event had occurred. As might be expected, it was predicted that consensus would be higher for the categories requiring little inference. In fact, for those categories where item descriptions include inferences regarding the intent or feeling state of the perpetrator, one would not expect a high degree of consensus. Consensus for such categories does not, strictly speaking, reflect reliability as the term is usually used, but rather reflects the ability to decode and interpret the partner's behavior. This task requires more than accurate, objective observation.

Of the five categories that were individually analyzed, three (companionship, affection, sex) required minimal or no inferences on the part of the observing spouse. Examples of items in each of these categories are shown in Table 4. Most of the items in these categories consist either of discrete, joint activities or behaviors whose occurrence can be unambiguously determined. Items under the categories of consideration and communication are strikingly different; although some can be unambiguously determined, many require considerable inference. As expected, consensus was substantially greater

for categories comprising noninferential items, as an examination of Table 1 indicates. In fact, for nondistressed couples, percentage agreement statistics for these categories are fairly close to conventional criteria of acceptability. In contrast, consensus figures for categories composed of highly inferential items are much lower and well below conventional criteria of acceptability. Thus, it can be tentatively concluded that the level of inference required to determine the occurrence of relationship behaviors is an important contributor to the generally low rates of agreement between spouses.

Discussion

The major finding of this study was that married couples disagree as often as they agree regarding the occurrence or nonoccurrence of events in their relationship. Since 70% agreement is usually considered the absolute minimum level of interrater agreement required for observational research, it is clear that spouses cannot be construed as collecting reliable observational data. Since spouses have already been extensively utilized as observers in marital research, all studies which assume that spouses collect data as accurately as do uninvolved observers must be reinterpreted in light of this finding.

Distressed couples tended to disagree a greater percentage of the time than nondistressed couples. However, when kappa statistics were calculated to measure consensus, thus partialing out the probability of chance agreement, the differences between distressed and nondistressed couples for individual categories were no longer statistically significant. This suggests that a substantial component of the consensus differences between distressed and nondistressed couples was an artifact of the greater overall frequency of item endorsement on the part of nondistressed couples. The significant differences in overall consensus using kappa as the consensus statistic documents that real differences exist beyond the artifact, although the differences are relatively small. It should be pointed out that the differences may have been attenuated by the fact that distressed couples did not constitute a true clinical population; it is conceivable that if our sample of distressed couples had been composed of couples seeking therapy, the differences would have been more striking (see Footnote 1).

In attempting to account for the greater

Table 3
Correlations Between Spouse Consensus and Frequency of Occurrence

Category	% agreement	Kappa
Companionship	.02	−.03
Affection	.56**	.06
Sex	.38**	.15
Consideration	.27*	.01
Communication	.45***	.06

* $p < .05$. ** $p < .005$. *** $p < .0001$.

Table 4
Examples of Items From the Spouse Observation Checklist

Category	Item
Companionship	We watched TV.
	We took a walk.
	We attended a sporting event.
	We went to a bar.
	We laughed together.
Affection	We held each other.
	We took a shower or bath together.
	Spouse hugged or kissed me.
Sex	We engaged in sexual intercourse.
	Spouse petted me.
	Spouse caressed me with his/her mouth.
Consideration	Spouse said he was glad to see me.
	Spouse agreed strongly with something I said.
	Spouse was tolerant when I made a mistake.
	Spouse answered my questions with respect.
Communication	We talked about personal feelings.
	Spouse confided in me.
	Spouse expressed feelings and thoughts to me.
	We had a humorous conversation.
	Spouse asked me about my feelings.

degree of consensus found in happy couples, we investigated a hypothesis based on "selective tracking," the notion that distressed couples selectively attend to their partners' negative behavior and therefore underestimate the frequency of the partner's positive behavior. This tendency would be reflected in a particular type of disagreement, namely, a disagreement resulting from positive items endorsed by the self-monitoring spouse but not endorsed by the spouse observing the other. No support was found for the selective tracking hypothesis. Distressed couples were no more inclined to disagree in the manner predicted by this hypothesis than were their nondistressed counterparts.

Perhaps the most striking characteristic of the data is the general lack of consensus rather than the rather small differences in consensus between distressed and nondistressed couples. One factor that may improve consensus is the elimination or the reduction of the number of inferential items. The data on individual categories suggest that consensus is substantially reduced when items require an inference on the part of the spouse–observer. Elimination of such items may indeed improve the accuracy of recording, but perhaps at the cost of construct validity. More will be said of this below. Another potential strategy for improving consensus would be to reduce the length of the checklist so that spouses have fewer items to peruse and consider for endorsement. Christensen and Nies (1980) reported that a substantial proportion of SOC items are endorsed rarely or not at all. If the number of items were substantially reduced and frequency of occurrence in a normative sample used as a criterion for inclusion, one result should be the collection of more reliable data. Finally, more extensive training should facilitate accuracy and consensus in the same way that the training of behavioral observers usually renders them more reliable.

Moreover, we did find a significant effect for days, indicating that spouses reached greater consensus over time. This finding suggests that although practice may not make perfect, it may, along with more rigorous training, significantly increase consensus.

Unfortunately, most of the available methodological strategies for improving consensus are double-edged swords. In the quest for greater reliability, one must be careful that the utility of spouse-collected data is not sacrificed. For example, inferential items may be very important in a marital relationship, and eliminating them simply to improve reliability may be self-defeating. It was noted in the Results section that consensus was relatively low for items in the category of communication, yet communication items have been shown to be highly correlated with subjective marital satisfaction and to be among the leading discriminators between distressed and nondistressed couples (Jacobson et al., 1980; Margolin, in press).

Such conflicts between reliability and validity raise more general considerations regarding the purposes of assessment strategies. Although it has been generally assumed that reliability is a prerequisite to validity, whether this is, in fact, true depends on the uses to be derived from a particular assessment instrument. Certainly if data collected through spouse observations are to be treated as objective and accurate indices of the true state of affairs in the day-to-day life of a married couple, then spouses must be able to agree on whether the particular events have occurred as a necessary but not sufficient condition to regarding the data as trustworthy. On the other hand, if we stop demanding that such data accurately mirror overt behavior in an objective sense, it can readily be seen that such data are useful despite this lack of consensus. For example, the ability of spouse-collected data to discriminate between distressed and nondistressed couples can define the usefulness of such data independent of the accuracy or reliability. More generally, if spouse reports of the events in their relationship are considered important in their own right, independent of their accuracy, then validity remains an issue that is unencumbered by standard reliability considerations. It is our belief that the phenomenology of married couples, their perceptions, appraisals, and beliefs regarding the relationship, constitute important sources of assessment data and therefore should not be discarded simply

because such data are unreliable in the classical sense.

In other words, our methodological loss may turn out to be a source of theoretical enrichment. The fact that two spouses living in the same environment perceive such different worlds suggests that in functional terms, spouses are operating in vastly different environments. The limitations of spouses as reliable observers have vast theoretical significance. The characteristics of spouses as observers constitute a viable area of study within behavioral marital therapy. How spouses in a marital relationship perceive and process information relevant to the relationship, the kinds of attributions they make regarding their own as well their partner's behavior, and the factors that influence these cognitive events potentiate the development of more useful assessment instruments, a richer understanding of the characteristics of marital distress, and the creation of a more effective treatment technology. Clinicians need to know not just what is actually happening in a distressed marital relationship but also what each spouse perceives as happening. Standard reliability considerations are largely irrelevant to this latter point. Marital therapy may require an extensive focus on perceptual modification in addition to its emphasis on behavior change. Thus, there is good news as well as bad in the results of this study. Spouses are not passive and objective observers of their partners' behavior; rather, they are participants with biases, and these biases need to be understood.

Reference Note

1. Weiss, R. L., & Perry, B. A. *Assessment and treatment of marital dysfunction.* Eugene: Oregon Marital Studies Program, 1979. (Available from Robert L. Weiss, Department of Psychology, University of Oregon, Eugene, Oregon 97403).

References

Barnett, L. R., & Nietzel, M. T. Relationship of instrumental and affectional behaviors and self-esteem to marital satisfaction in distressed and nondistressed couples. *Journal of Consulting and Clinical Psychology*, 1979, *47*, 946–957.

Birchler, G. R., Weiss, R. L., & Vincent, J. P. A multimethod analysis of social reinforcement exchange between maritally distressed and nondistressed spouse and stranger dyads. *Journal of Personality and Social Psychology*, 1975, *31*, 349–360.

Christensen, A., & Nies, D. C. The Spouse Observation Checklist: Empirical analysis and critique. *American Journal of Family Therapy*, 1980, *8*, 69–79.

Ciminero, A. R., Nelson, R. O., & Lipinski, D. P. Self-monitoring procedures in behavioral assessment. In A. R. Ciminero, K. S. Calhoun, & H. E. Adams (Eds.), *Handbook of behavioral assessment.* New York: Wiley, 1977.

Goldfried, M. R., & Sprafkin, J. N. *Behavioral personality assessment.* Morristown, N.J.: General Learning Press, 1974.

Hartman, D. P. Considerations in the choice of interobserver reliability estimates. *Journal of Applied Behavior Analysis*, 1977, *10*, 103–116.

Hollenbeck, A. R. Problems of reliability in observational research. In G. P. Sackett (Ed.), *Observing behavior*, (Vol. 2). Baltimore, Md.: University Park Press, 1978.

Jacobson, N. S. Problem solving and contingency contracting in the treatment of marital discord. *Journal of Consulting and Clinical Psychology*, 1977, *45*, 92–100.

Jacobson, N. S. Increasing positive behavior in severely distressed adult relationships. *Behavior Therapy*, 1979, *10*, 311–326.

Jacobson, N. S., Elwood, R., & Dallas, M. The behavioral assessment of marital dysfunction. In D. H. Barlow (Ed.), *Behavioral assessment of adult disorders.* New York: Guilford Press, 1981.

Jacobson, N. S., & Margolin, G. *Marital therapy: Strategies based on social learning and behavior exchange principles.* New York: Brunner/Mazel, 1979.

Jacobson, N. S., & Moore, D. Behavior exchange theory of marriage: Reconnaisance and reconsideration. In J. P. Vincent (Ed.), *Advances in family intervention, assessment, and theory.* Greenwich, Conn.: JAI Press, in press.

Jacobson, N. S., Waldron, H., & Moore, D. Toward a behavioral profile of marital distress. *Journal of Consulting and Clinical Psychology*, 1980, *48*, 696–703.

Johnson, S. M., & Bolstad, O. D. Methodological issues in naturalistic observation: Some problems and solutions for field research. In L. A. Hamerlynck, L. C. Handy, & E. J. Mash (Eds.), *Behavior change: Methodology, concepts, and practice.* Champaign, Ill.: Research Press, 1973.

Linehan, M. M. Issues in behavioral interviewing. In J. D. Cone & R. P. Hawkins (Eds.), *Behavioral assessment: New directions in clinical psychology.* New York: Brunner/Mazel, 1977.

Margolin, G. Behavior exchange in distressed and nondistressed marriages: A family life cycles perspective. *Behavior Therapy*, in press.

Margolin, G., & Jacobson, N. S. Assessment of marital dysfunction. In A. S. Bellack & M. Hersen (Eds.), *Behavioral assessment: A practical handbook.* London: Pergamon Press, in press.

Margolin, G., & Weiss, R. L. A comparative evaluation of therapeutic components associated with behavioral marital treatment. *Journal of Consulting and Clinical Psychology*, 1978, *46*, 1476–1486.

Patterson, G. R. Some procedures for assessing changes in marital interaction patterns. *Oregon Research Institute Bulletin*, 1976, *16*(7).

Robinson, E. A., & Price, M. G. Pleasurable behavior in marital interaction: An observational study. *Journal of Consulting and Clinical Psychology*, 1980, *48*, 117–118.

Spanier, G. B. Measuring dyadic adjustment: New scales for assessing the quality of marriage and similar dyads. *Journal of Marriage and the Family*, 1976, 38, 15–28.

Vincent, J. P., Cook, N. I., & Messerly, L. A social learning analysis of couples during the second postnatal month. *American Journal of Family Therapy*, 1980, *8*, 49–68.

Weiss, R. L., Hops, H., & Patterson, G. R. A framework for conceptualizing marital conflict, a technology for altering it, some data for evaluating it. In L. A. Hamerlynck, L. C. Handy, & E. J. Mash (Eds.), *Behavior change: Methodology, concepts, and practice*. Champaign, Ill.: Research Press, 1973.

Weiss, R. L., & Margolin, G. Assessment of marital conflict and accord. In A. R. Ciminero, K. D. Calhoun, & H. E. Adams (Eds.), *Handbook of behavioral assessment*. New York: Wiley, 1977.

Wills, T. A., Weiss, R. L., & Patterson, G. R. A behavioral analysis of the determinants of marital satisfaction. *Journal of Consulting and Clinical Psychology*, 1974, *42*, 802–811.

Received September 25, 1980 ■