

University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

Family Assessment

Buros-Nebraska Series on Measurement and
Testing

1995

6. Assessing Marital Quality In Longitudinal And Life Course Studies

David R. Johnson

University of Nebraska-Lincoln, drjls@yahoo.com

Follow this and additional works at: <https://digitalcommons.unl.edu/burosfamily>



Part of the Educational Assessment, Evaluation, and Research Commons, Family, Life Course, and Society Commons, and the Quantitative, Qualitative, Comparative, and Historical Methodologies Commons

Johnson, David R., "6. Assessing Marital Quality In Longitudinal And Life Course Studies" (1995). *Family Assessment*. 10.

<https://digitalcommons.unl.edu/burosfamily/10>

This Article is brought to you for free and open access by the Buros-Nebraska Series on Measurement and Testing at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Family Assessment by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

ASSESSING MARITAL QUALITY IN LONGITUDINAL AND LIFE COURSE STUDIES

David R. Johnson

University of Nebraska-Lincoln

INTRODUCTION

Family researchers have been developing measures to assess the quality of the marital relationship for over six decades (e.g., Hamilton, 1929). Indeed, the quality of the husband-wife relationship has been the focus of more research than any other single topic in the field of family study (Spanier & Lewis, 1980). Embedded in these studies are hundreds of varied scales and measures that were designed to assess some aspect of the quality of a marriage (Touliatos, Perlmutter, & Straus, 1990). Lack of consensus on what constitutes marital quality and the absence of any widely accepted and used instruments have contributed to this proliferation of measures. Even scales that enjoy wide use have come under persistent theoretical and methodological criticism (Huston & Robins, 1982; Norton, 1983; Sabatelli, 1988). This state of affairs reflects the different aims of the researchers developing the measures and the evolution over the last several decades of the theoretical and conceptual definitions of the quality of a marriage.

This study was partially supported by grant 5 R01AG04146 from the National Institute on Aging.

The term "marital quality" has only recently been used to refer to concepts and measures that in the past have been called marital adjustment, satisfaction, and happiness (Spanier & Lewis, 1980). Marital satisfaction and happiness both refer to subjective evaluations of positive affect in the marital relationship by one (or both) of the spouses. Marital adjustment signifies both behavioral and evaluative aspects of a marital relationship. These include dyadic cohesion, satisfaction, consensus, interpersonal tensions, and troublesome dyadic differences (Spanier, 1976). A well-adjusted marriage is often characterized by high interaction and cohesion, low levels of disagreement, high levels of commitment to the relationship (i.e., a low likelihood of leaving the relationship), and good communication and problem-solving abilities. Adjustment is clearly seen as multidimensional, composed of several distinct, but closely related concepts (Spanier & Lewis, 1980). The behavioral and evaluative factors that define marital quality are assumed, based on experience in marital counseling and therapy, to be necessary for a harmonious relationship.

Marital quality measures have been created with two quite different aims: the identification of troubled marriages—primarily a clinical aim, and the desire to test theories related to marital functioning and behavior—a basic research aim. There are no necessary theoretical reasons why measures that function well in one capacity cannot also be valid in the other. Practical and methodological matters, however, often play a more important role. For example, it is unlikely that a 250-item marital assessment scale would be used in a national interview survey of married persons in which the quality of the marital relationship is only one focus. This difference in objectives has been a key factor accounting for variation in concepts and methods used to develop the measures and in the criteria applied to evaluate them. This review focuses on issues of marital quality assessment in nonclinical research settings that use quantitative methods. However, the strong link between family therapy and marital quality research studies—many key researchers are also family therapists—makes it necessary to consider the influence of marital therapy.

Research studies exploring marital quality have, with some notable exceptions, made use of interview or questionnaire data of married respondents collected in one-time (cross-sectional) surveys. This has been the case despite an increasingly awareness that valid answers to some key unsolved issues in the study of marriage over the life course require longitudinal data (Mattessich & Hill, 1987).

It might be expected that reliable and valid measures of marital quality used in cross-sectional studies would be equally applicable to

longitudinal samples. Unfortunately, this is not always the case. Many measurement and analysis issues are introduced when inferences are attempted from multiyear samples (Johnson, 1988). Panel analysis raises concerns about the reliability and stability of measures and their ability to reflect changes (Huston & Robins, 1982). The analysis of the dynamics of family development and change requires that the concepts and measures be analytically distinct, particularly when one aspect of the marital relationship is seen as having a causal effect on another (Norton, 1983). A study examining the effect of wife's employment on marital quality could not examine the intervening mechanisms, such as degree of marital interaction or disagreement, which mediate the effect of work on marital happiness or thoughts of divorce (Booth, Johnson, White, & Edwards, 1984) if these are combined in the measure of marital quality.

There have been several reviews of measurement and conceptual issues in assessing the quality of the marital relationship (Sabatelli, 1988; Spanier, 1976; Glenn, 1990; Huston & Robins, 1982; Johnson, White, Edwards, & Booth, 1986; Hicks & Platt, 1970). A recent inventory of marital quality scales is also available (Touliatos, Perlmutter, & Straus, 1990). None have focused on the conceptual and measurement issues raised by the increasing amount of life course research that focuses on the dynamics of the marital relationship. The purpose of this chapter is to critically examine a selected set of conceptual and methodological issues that have relevance to the study of marital quality over the life course.

Life course theory is concerned with explanation of psychological and social changes in individuals as they progress from birth to death within the context of their society (Featherman & Lerner, 1985). Marital life course studies identify factors that account for changes in the husband-wife relationship that reflect the chronological aging of the individuals and the marriage and the changing roles and structures of the family as the individuals move through their marital life cycle (Mattessich & Hill, 1987). The effects on the marital relationship of the birth of children, changes in health and well-being caused by aging, children leaving home, retirement, and changes in economic status and assets, are examples of variables that can be examined in a life course perspective. This perspective also focuses on how patterns of behavior and evaluations early in a marriage carry over into later stages of the relationship.

Research on marital quality over the life course has made use of both cross-sectional and longitudinal samples. In cross-sectional studies change can only be inferred by comparing marriages at

different life course stages. These findings will be biased to the extent that there are period, cohort, and selection effects (Glenn, 1991). In longitudinal panel and trend studies, such patterns of change can be observed more directly, but additional problems arise while estimating the effects, such as autocorrelated errors and separating reliability from stability.

This chapter begins to approach the issues of marital quality measurement by reviewing issues related to the definition of marital quality that have influenced assessment strategies. This is followed by a selective review of several scales and measures used in studies of the marital life course that exemplify the different conceptual perspectives on marital quality found in the literature. The focus is then turned to a specific examination of conceptual and methodological issues that have emerged as problems in the assessment of marital quality in life course research. Findings from a four-wave panel of marriages studied over a period of 12 years will be used to illustrate and in some cases provide tentative answers to some important methodological and measurement questions. Finally, conclusions are made about the adequacy of current conceptualization and measurement of marital quality for longitudinal studies.

ISSUES IN DEFINING MARITAL QUALITY

Although many different measures have been called marital quality, there has been more convergence at the level of operationalization than at the level of conceptualization. Scale items that are very similar if not identical are often shared by instruments with widely differing labels and conceptual definitions. Most of the measures have employed a self-report questionnaire or survey format responded to by married persons answering as individuals and not as couples. Many have been validated by comparing scale scores of persons in marital therapy with those not in therapy.

Although the available instruments share much in common in terms of the items used, samples studied, and criteria used to assess scale validity, a basic conceptual and theoretical issue repeatedly surfaces in the published reviews of marital quality measures. This issue concerns whether the definition of marital quality and its measurement should include both behavioral and evaluative components and whether single or distinct measures are needed to assess these components (Norton, 1983; Johnson, White et al., 1986). There have been three perspectives on what constitutes marital quality: marital adjustment, global evaluation, and marital quality as a set of variables.

Marital Adjustment Perspective

The concept of marital adjustment has an extensive history in family research and predates the use of the term marital quality (Lively, 1969). Although the concept has received careful theoretical and conceptual clarification (Spanier, 1976; Spanier & Cole, 1976), the general perspective towards assessment has been a pragmatic one. How well does a potential measure differentiate between “well-adjusted” and “poorly-adjusted” marriages? The definition of adjustment, as discussed earlier, includes not only the married person’s subjective evaluation of the marriage but also behavioral characteristics that signify adjustment. Married individuals who are satisfied or happy with their marriage are not necessarily in a well-adjusted marriage. The behavior of the couple in terms of their interaction, communication, consensus, agreement, and commitment is all viewed as important for the placement of a married person on an adjustment continuum (Spanier, 1976). An adequate measure must tap domains of individual subjective evaluation as well as dyadic behavior.

Specification of the appropriate domain of content for the universe of items to be included in a marital adjustment scale often begins with qualitative experience gained from working with distressed couples in family and marital therapy. Because the definition of adjustment includes both evaluative and behavioral traits, the universe of items tapping marital adjustment includes both. The ultimate criterion of whether an item is appropriate for inclusion is its ability to distinguish between maladjusted and normal marriages (Locke & Wallace, 1959; Spanier, 1976). There is an explicit recognition that although the concept of adjustment is multidimensional, a single ordering of marriages from high to low adjustment is possible. The two scales most commonly used in family research (as opposed to those whose primary function is to assist in marital therapy with individual marriages) that share this perspective are the Locke-Wallace Marital Adjustment Test (LWMAT; Locke & Wallace, 1959) and Spanier’s (1976) Dyadic Adjustment Scale (DAS). Both have been widely used in marital research, including longitudinal and marital life course studies.

A major critique of the adjustment perspective is that by defining the concept to include several behavioral and evaluative properties, its research utility is limited (Norton, 1983; Fincham & Bradbury, 1987), particularly in studies of the interrelationships between characteristics of the marriage. Scales created for prediction purposes can impose less rigorous standards on the content of the domain of

items than do those designed to test empirically the interrelationship among a set of theoretically derived and relevant concepts (Nunnally, 1967).

An example of the wide universe of items often allowed for scales focusing primarily on their ability to predict a trait is the Marital Prediction Test developed by Locke and Wallace (1959). The Marital Prediction Test is designed for "forecasting the likelihood of marital adjustment at a future time" (Locke & Wallace, 1959, p. 251). Among the 20 items in the scale are the respondents' educational attainment, age at marriage, church attendance, size of community in which they grew up, parents' approval of their marriage, and general attitudes toward sex. The combination of demographic, background, and evaluative items makes the concept and the measure virtually worthless for research purposes.

Although marital adjustment measures tap a narrower domain of content, choice of items is often guided more by the ability to differentiate among adjusted and maladjusted marriages than by the need to measure a theoretically coherent trait. Because marital adjustment is defined as a multidimensional concept encompassing a wide range of behaviors and attitudes, this conceptualization has limited utility both in the theoretical models of the dynamics of marital relations and in their empirical testing.

Marital Evaluation Perspective

A perspective that restricts the concept of marital quality solely to subjective global evaluations of the satisfaction/happiness of the married partners has gained increasing support in the marital quality literature (Fincham & Bradbury, 1987; Norton, 1983; Sabatelli, 1988). Advocates of the marital evaluation perspective view the concept as a tool for research and theory and not marital therapy. Fincham and Bradbury (1987) argue that combining behavioral and evaluative components in the same concept and scale confounds the description of the marriage with its evaluation. Attempts to explain marital quality with characteristics of the marital relationship are artifacts of the common variance of shared items in the independent and dependent measures. A researcher interested in the extent to which dyadic communication affects marital quality would be making a serious methodological error to use a measure of marital quality based on the adjustment perspective because good communication is in the domain of content of the adjustment concept and is tapped by its measures.

Several recent measures build on this concept of marital quality. Both the Kansas Marital Satisfaction Scale (KMSS; Schumm, Paff-Bergen, Hatch, Obiorah, Copeland, Meens, & Bugaighis, 1986) and the Marital Quality Index (MQI; Norton, 1983) are unidimensional measures of global satisfaction. Single-item measures of marital happiness have been used in many studies and conform to this perspective (Glenn, 1990). Although psychometrically suspect, the single-item measures of marital quality possess the pragmatic advantage of having been included for decades as the only indicator of marital quality on many large national longitudinal surveys (Glenn, 1990; Orden & Bradburn, 1968).

The problem with this perspective is that it takes the term marital quality that has been widely used to refer to a range of both evaluative and behavioral characteristics of the marital relationship and narrows its application to a much smaller set of concepts and measures. Even if there are compelling theoretical and conceptual reasons for restricting the meaning of the term, the practical matter is that the broader meaning of the term marital quality has already been established, a condition that is difficult to reverse in practice. Perhaps another term other than happiness or satisfaction needs to be selected to refer to the global subjective evaluation of the marriage.

Marital Quality as a Set of Traits

Rather than referring to a specific quality of the relationship that can be assessed by a single instrument, this perspective treats marital quality as an umbrella concept encompassing a *set* of marital behaviors and evaluations, each assessed by a separate measure. This is the most widely accepted meaning of the term in the current literature. Recent reviews of research on marital happiness, satisfaction, and adjustment have also adopted this usage (Spanier & Lewis, 1980; Glenn, 1990). The value of defining marital quality in this way can be seen in the conceptual and definitional confusion found in the field before the term was introduced. Hicks and Platt (1970), in a decade review of research on the same set of concepts that are currently called marital quality, reluctantly used the term "marital happiness" to refer to the set of measures because no other term was available.

The use of separate scales to measure the components of marital quality (Johnson, White et al., 1986) and the practice of breaking composite measures such as the Dyadic Adjustment Scale into subscales (Spanier, 1976) both fit this perspective. This allows for a broader definition of marital quality, similar to that used by advocates of the marital adjustment perspective. It also insists on separate definitions

and measures of behavioral and evaluative elements of the marital relationship that are needed for research into the dynamics of the marital relationship. It is this use of marital quality that is adopted in this chapter.

SCALES ASSESSING MARITAL QUALITY IN LIFE COURSE STUDIES

Many measures of marital quality have been used to assess change and stability in marriages as they pass through the life course. Cross-sectional studies predominate, but there also have been a few trend studies (separate samples with the same measures surveyed in different years) and panel studies (same sample surveyed two or more times). The measures selected for review were primarily designed for research rather than therapy, represent the range of definitions of marital quality discussed above, and illustrate some major methodological and conceptual issues in the assessment marital quality over the life course.

Orden and Bradburn's Marital Adjustment Balance Scale (MABS)

The Orden and Bradburn Marital Adjustment Balance Scale (MABS) is based on the theoretical model of psychological well-being that assumes that individual subjective happiness is a function of two independent dimensions, one of positive, the other of negative affect (Bradburn, 1969, p. 9). This theoretical model was applied to account for both overall individual and marital well-being.

To develop the MABS, a nine-item scale of marital tensions and a nine-item scale of marital satisfactions were created based on intensive interviews with a small sample of respondents. Other items were included based on their relationship to the general positive and negative affect scales also developed by Bradburn (1969). Respondents were asked to give a yes or no response to a checklist of items. The marital satisfactions measure included items measuring companionship and sociability, which were also treated as separate subscales (Orden & Bradburn, 1968). These included items tapped affection, sharing a good laugh, spending an evening chatting with one another, doing things together with friends, eating out together, and going out together for entertainment. The marital tensions subscale included a set of situations that caused disagreements or problems in the marriage in the last few weeks, such as being tired, irritating personal habits, household expenses, being away from home, and not showing love.

The subscales, which were a simple summation of the number of yes responses, were found to be correlated with a single-item measure

of marital happiness (Taking all things together, how would you describe your marriage? Would you say your marriage was very happy, pretty happy or not too happy?) but were not significantly correlated with one another (Orden & Bradburn, 1968). The marital satisfactions and tensions subscales were combined into a single composite 11-point scale to create the Marital Adjustment Balance Scale.

Orden and Bradburn validated the scale primarily based on its strong relationship to the marital happiness item and the similar correlations of MABS and happiness with variables such as socioeconomic status and gender. Although the word adjustment is used in the title, it was not constructed or validated by score comparisons of well and poorly adjusted couples as determined by therapists. They do not report any indices of internal consistency of test-retest reliability for the scale.

The study in which the scale was first used involved two to four waves of panel data over a period of one year. Because all items were included only on a later wave, patterns of change for the entire scale were only available for samples interviewed in Waves II and III. Test-retest correlations (computed from cell frequencies presented in tables [Bradburn, 1969]) for marital tensions (collapsed into three ordinal categories) were .4 for both men and women. It was not possible with the data presented to compute the correlations for the satisfactions subscale or the total MABS.

This scale is important because it was used in one of the first attempts to evaluate quantitatively in a panel study the relationship between change in different components of the marital relationship. Based on an analysis of cross-classifications, Bradburn (1969) concluded that change in marital tensions was associated with change in marital happiness but change in marital satisfactions was not. For this analysis, the scale was disaggregated into its components and was not treated as the single balance measure. This practice was also noted in other studies making use of the scales (e.g., Burke, Weir, & DuWors, 1979).

The MABS was developed from a specific theoretical model of well-being applied to marriages in which the ultimate dependent variable is a subjective global evaluation of the marriage (happiness). Thus, it appears to fit the conceptualization of marital quality as a global subjective evaluation of the marriage. Orden and Bradburn (1968) even discuss whether the researcher should use their scale or the single-item measure of marital happiness. They conclude that the MABS would be preferred primarily because of its greater precision

(more categories). Yet the scale itself does not include evaluative measures and might be seen primarily as assessing marital behavior. The measure of tensions comes closer to a marital problems scale (Johnson, White et al., 1986) and the positive affect measure primarily taps spousal interaction and, to some extent, intimacy.

The relatively low test-retest correlations of the tensions subscale in the MABS and the acknowledgement by Bradburn (1969) that the positive affect items performed poorly in accounting for change in marital happiness suggest that this scale and its components may not be useful for longitudinal studies. The observed independence of the tensions from marital sociability and companionship subscales is not necessarily consistent with findings from other studies using similar measures. Johnson, White et al. (1986) found strong correlations between a marital problems index (similar to the MABS tension subscales) and marital interaction. It is possible that application of the psychometric scaling techniques available to researchers today to data collected using these scales would help clarify some of the issues related to their reliability and stability. The MABS itself taps several marital behaviors and is multidimensional. This scale has the conceptual advantages of not combining behavioral and evaluative components and being derived from a theoretical model of psychological well-being. However, its problems and uncertainties outweigh these advantages.

The Locke-Wallace Marital Adjustment Test

This widely used scale was created to provide a short 15-item test of marital adjustment at a time when most of the available scales averaged around 150 items (Locke & Wallace, 1959). It was created to provide a short, easily administered scale for use in research settings. Items were selected from previous scales that best discriminated high and low adjustment in the original studies and covered the important domains of content as evaluated by the authors. Reliability was judged as high (.90 using the Spearman-Brown formula) and the scale discriminated well between respondents in mal- and well-adjusted marriages judged by clinical criteria.

The Locke-Wallace Marital Adjustment Test (MAT) has an important place in family research because it represents the first short instrument geared to researchers from the marital adjustment and marital therapy perspective. As a measure of marital quality it clearly fits the conceptual definition marital adjustment, because the domain of item content includes both behavioral and evaluative components.

It has often been used as a criterion to evaluate the validity of other marital quality scales (Spanier, 1976).

Included in the scale are a marital happiness item, a set of items about disagreements, marital interaction, and questions about whether the respondents would have ever married or would marry the same person again. A scaling system is present for weighting items, although there is little explanation of how these were derived. For example, the single item of marital happiness has a weight three to seven times greater than other items.

Use of the scale in other samples has confirmed its reliability (Margolin, 1978) and several reported studies have examined the factorial structure of the items in the scale (Kimmel & Van Der Veen, 1974; Locke & Williamson, 1958) and its overall reliability and validity (Cross & Sharpley, 1981). Several factor analyses all support the scale as multidimensional, although there has been little agreement on the number of dimensions (from one to eight). Kimmel and Van Der Veen (1974) found only one factor when men and women were combined in the same analysis, but found two distinct factors for husbands and wives when analyzed separately. They also reported that these factors have high test-retest stability. In a small sample of 44 couples tested a little over 2 years apart, the test-retest correlations were between .69 and .78 for the separate factors for husbands and wives. They concluded that both factors appear to tap stable and enduring characteristics of the marital relationship.

Because Locke has published several versions of the marital adjustment scale with varying sets of items, few of these validity and factorial structure studies report on the same scale. This has made it difficult to judge the dimensionality of the scale or provide information on how best to form subscales to separate out substantively important behavioral and evaluative components. The small number of items in the scale contributed to the wide use of the measure but made its potential disaggregation into useful subscales more difficult. As a result, the scale would not be very useful for life course studies examining the dynamics of the components of the marital relationship.

Spanier's Dyadic Adjustment Scale

The Dyadic Adjustment Scale (DAS) is the most widely used indicator of marital quality in the literature, with over 1,000 studies making use of the scale (Touliatos, Perlmutter, & Straus, 1990). It was also conceived in the marital adjustment tradition where the primary criterion for the scale was its ability to distinguish between well-adjusted and failing marriages (Spanier, 1976). A unique feature of

the scale was that items were worded in a way that made the scale appropriate for nonmarital dyads (e.g., a cohabiting couple).

The pool of items considered for the scale was selected from among all previously published adjustment instruments. Additional items were added to fill gaps in domains the author believed were not well represented in the pool. The final composite scale consists of 32 items and taps both behavioral and evaluative components of the relationship. The DAS includes a global happiness item and 15 items tapping agreement in different areas of the relationship, thoughts of divorce, temporary separations, quarreling, marital interaction, and displays of affection. Twelve of the 15 items in the Locke-Wallace MAT are included in the DAS. This results in a close correspondence between these two scales; Spanier (1976) reported a correlation of .86 between the DAS and the MAT.

Selection of items from the pool for inclusion in the DAS involved several criteria. A critical factor was the ability of the item to discriminate between a sample of divorced persons who answered the scale based on recollection of the last months of their failed marriage and a sample of currently married persons. Highly skewed items were also excluded. A final step excluded items with low factor loadings. A coefficient alpha reliability of .96 was reported for the total scale.

Subscales of the DAS were created to reflect the multidimensional nature of marital adjustment. These were developed by factor analysis and consist of four subscales: Dyadic Consensus (13 items), Dyadic Satisfaction (10 items), Dyadic Cohesion (5 items), and Affectional Expression (4 items). All the subscales except Affectional Expression had coefficient alpha reliabilities exceeding .85. Confirmatory factor analysis in a second sample of divorced and separated persons generally confirmed the four factors (Spanier & Thompson, 1982), but another factor analysis of married respondents did not (Sharpley & Cross, 1982).

Because of its widespread use, the DAS has often been the focus in critical reviews of the measurement of marital quality. Several problems with the scale have received the most emphasis. Because the DAS includes behavioral and evaluative items, the concern has been raised that this confounds and limits analysis of marital processes (Norton, 1983; Fincham & Bradbury, 1987). The practice observed in many studies of using the subscales rather than the composite measure partially alleviates this concern (e.g., Belsky, Spanier, & Rovine, 1983). However, the subscales do not separate behavioral and evaluative dimensions that need to be measured separately in causal and

longitudinal models of marital processes (Johnson, White et al., 1986). For example, the Dyadic Satisfaction subscale included behavioral reports (e.g., frequency of quarrels, discussion of divorce or separation, frequency of marital interaction) as well as evaluative items (marital happiness, feelings about the future of the relationship)

Norton (1983) raises concerns about the arbitrary weighting of items in the DAS. Although most of the items can contribute up to 5 points each to the scale, two can only contribute 1 point, two 4 points, and one 6 points. Their relative contribution reflects only the number of response categories and not the discriminating power of the item. Coupled with the variable number of items in each subscale, these lead to disproportionate contribution of certain domains of content to the total scales score that are unrelated to their conceptual importance or discriminating power. This is not a serious problem for the researcher who is willing to discard the recommended scoring of the DAS in favor of weighting to equalize the contributions of items and subscales to the total scale score (Norton, 1983).

Several methodological concerns have been directed to the definition and structure of the subscales. Because they were defined by factor analysis they can be questioned when factor analyses in other samples do not reproduce the same structure. Although a close fit to the four-factor structure was confirmed by Spanier and Thompson (1982), Sharpley and Cross (1982) found a very different factor structure. Crane, Busby, and Larson (1991) also failed to reproduce the four factor structure among both distressed and nondistressed couples. Unfortunately, none of these studies used large or representative samples. Spanier and Thompson's (1982) sample was of divorced persons responding about their failed marriages, Sharpley and Cross did not say where they got their 95 married respondents, and Crane, Busby, and Larson used a sample of 253, containing both couples in therapy and nondistressed couples. Because the subscales were defined by factor analysis, Sabatelli (1988) raises the concern that they are not true scales because no attempt was made to define a universe of content for the dimensions from which to select the items.

Methodological and conceptual concerns about the DAS raise questions about its utility in studies examining the dynamics of the marital relationship over the course of the marriage. However, many studies examining family life cycle issues have used the DAS. The DAS and its subscales were used in a longitudinal panel study of the effects of the birth of a child on the marital relationship (Belsky, Spanier, & Rovine, 1983). Couples were interviewed before the birth of a child, and 3 and 9 months after the birth, for three waves of data.

The study examined the changes in mean scores for wives and husbands over the three waves of the subscales and composite DAS. Stability of the subscale and total scores was also reported. Additional measures of marital functioning and marital interaction were also included in the study. Significant mean declines were found over the three waves for the total scale, cohesion, and affectional expression. The total scale score was found to be highly stable, particularly for wives. Correlations between the first and third waves were .82 for wives and .69 for husbands. The subscales were less stable, with Satisfaction the most stable for both genders ($r = .81$ for wives and .60 for husbands) and Affectional Expression and Gender Cohesion the least stable (r s from .69 for wives to .43 for husbands). They conclude that the study observed real and reliable mean declines in components of marital adjustment over the period of the birth of a child, but that the relative rank order of the married persons on marital adjustment changed little over this marital life cycle transition.

Because of criticism directed toward the DAS and its subscales, Belsky, Lang, and Rovine (1985) replicated the study, substituting another set of scales that separately measured different aspects of the marital relationship. No attempt was made in either study to examine the causal process through which the addition of a child influenced the dynamics of the relationship between the spouses. The findings of the two studies were remarkably similar, suggesting that some conceptual and methodological criticism of the DAS in longitudinal studies may be unjustified.

The Kansas Marital Satisfaction Scale

The Kansas Marital Satisfaction Scale (KMSS) is the shortest scale in the marital quality literature, being composed of only three items. Yet its validity and reliability have been very carefully and completely evaluated in published studies (Schumm, et al., 1986) and it has been used in family life cycle studies (Anderson, Russell, & Schumm, 1983). It strictly fits the conceptualization of marital quality as global evaluation of the marriage. The three items measure satisfaction with spouse, the marriage, and the marital relationship. The scale has the advantage of being short, and concurrent validity studies found it to be correlated highly with the Dyadic Adjustment Scale. It has also been shown to be stable over a 10-week period ($r = .71$) (Mitchell, Newell, & Schumm, 1983).

Anderson, Russell, and Schumm (1983) used the KMSS in a cross-sectional study to test research questions about the relationship between marital quality and stage in the family life cycle. The KMSS was used

to measure the global assessment of marital satisfaction and several other scales were included to measure other aspects of the marital relationship (regard, empathy, discussion, and self-disclosure) that were viewed as causes of marital satisfaction. They found that marital quality (using the five scales, including KMSS, as a set in a MANOVA) showed a curvilinear relationship with family life cycle (lowest levels when the oldest children were from 5 to 12 years of age), which replicated findings from previous studies with other scales.

Although no panel studies have been reported that include this scale, the availability of three items gives it several advantages over the single-item measure of marital happiness to be reviewed below. How well the three items would serve as multiple indicators of a latent variable of marital satisfaction in structural equation path models is not known but deserves further exploration.

The Nebraska Marital Quality Scales

As part of a panel study of a national sample of married persons designed to assess factors predicting marital instability over the life course, Johnson, White et al. (1986) devised a set of five scales to measure five theoretically important dimensions of marital quality. Combinations of these scales have been used in many research studies based on a four-wave panel of married persons followed over 12 years. The marriage characteristics assessed were selected because of their theoretical importance in a model of the marital process and were proposed to account for the effects of wife's paid employment in the labor force on the likelihood that the marriage would end in divorce or permanent separation (Booth, Johnson, White, & Edwards, 1984). Scale items were selected based on a review of the literature, seeking items that fell within the theoretical domain of content for the concepts. Pretest of a national sample of 300 married persons was used to evaluate and modify some scales. The final versions of the scales were developed from the larger study of over 2,000 married respondents through item analysis. The five measures are Marital Happiness, Marital Interaction, Marital Disagreement, Marital Problems and Marital Instability (or Divorce Proneness).

Marital Happiness was defined as an individual level property reflecting positive and negative feelings about the marriage, and is conceptually equivalent to the Kansas Marital Satisfaction Scale and other global evaluative measures. Eleven items were included in the scale. Seven measured happiness with aspects of the relationship and four were global assessments of the relationship. The scale had a coefficient alpha reliability of .86.

Marital Interaction was defined as the amount of interaction of the couple in day-to-day activities. It consists of five items tapping eating main meals together, shopping, visiting friends, working on projects around the house together, and going out. The alpha reliability of the scale was .63 and yields one factor in factor analysis.

Marital Disagreements was designed to test for the presence and severity of disagreements between the spouses. Four items tap disagreements, frequencies of quarrels, and physical abuse. Because of different numbers of response categories in the items, each item was z-scored before the items were summed. The reliability was relatively low (alpha = .54).

Marital Problems assesses the extent to which personal traits and behaviors of the spouses contribute to problems in the marriage. It measures a collective property of the relationship reflecting a dyadic condition. Respondents were asked to indicate if 13 potential trouble spots in the marriage caused problems in their marriage due to either their behavior or the behavior of their spouse. Included were such things as getting angry easily, won't talk to each other, has irritating habits, drinks, or uses drugs. The alpha reliability was .76.

Marital Instability (also called Divorce Proneness) is defined as the propensity to divorce and included both cognitive and behavioral components. This includes thoughts about divorce and specific actions such as talking to a friend or spouse, seeing an attorney, separating, etc. There are 13 items in the scale and its alpha reliability is .91 (Booth, Johnson, & Edwards, 1983). This scale was validated primarily by its ability to predict divorce or permanent separation. Those who scored high on the scale were nine times more likely to divorce within 3 years than those exhibiting no instability on the scale. The five subscales are substantially intercorrelated. A confirmatory factor analysis of the scales found two correlated factors, one included marital happiness and marital interaction, the other marital instability, marital problems, and marital disagreements. Because of the need to retain separate scales for the conceptually distinct aspects of the marital relationship, no attempt was made to combine them into two composite measures.

Several research studies have made use of these scales included on a national longitudinal telephone survey of a sample of married persons. Many have examined one or more aspects of marital quality over the marital life course with either cross-sectional data from one of the earlier panels or two and three wave panel data (Booth & Edwards, 1989; Booth, Johnson, White, & Edwards, 1986; Johnson, Amoloza, & Booth, 1992; White, 1983; White & Booth, 1985; White &

Booth, 1991; White & Edwards, 1990; Zuo, 1992). Three of these studies exemplify how these measures can be used to improve our knowledge about the dynamics of marital processes.

White (1983) examined the reciprocal relationship between marital happiness and marital interaction with cross-sectional data from the 1980 wave of the national study. Making use of two-stage least squares, she tested the reciprocal relationship between marital happiness and marital interaction. Because she was using cross-sectional data, certain untestable assumptions were required to identify mathematically the set of equations needed to test the reciprocal relationship. Her results suggested that marital happiness was more likely to influence marital interaction than vice versa.

Zuo (1992) replicated White's findings using the same sample but included the information gained in three waves of interviews conducted in 1980, 1983, and 1988 with the same respondents. Panel data allowed a different set of assumptions to identify the equations. Zuo also treated happiness and interaction as latent variables and used the scale items as multiple indicators. This adjusted for the biasing effect of measurement errors in panel models. A reciprocal effect was found in the second wave that confirmed White's (1983) findings. In the third wave, however, the findings suggested approximately equal effects of happiness on interaction and interaction on happiness.

Johnson, Amoloza, and Booth (1992) examined the degree of stability and developmental change in the five measures over the first three waves (8 years) of the panel study for 1,043 respondents continuously married over the period. Developmental change was measured by the mean changes in marital quality scores that were due to the passage of time. Marital Happiness and Interaction were found to undergo significant declines over the 8-year period. No significant changes were found for the other measures. This pattern of change was the same for both married men and women and for respondents at different marital durations, although short-term marriages (under 5 years of marriage in the first wave) showed significantly greater declines than other groups in happiness and interaction.

Johnson, Amoloza, and Booth (1992) also examined the stability of the marriages. A procedure that separates reliability from stability in panel studies with three or more waves was used (Wiley & Wiley, 1970). This assured that the differences in the reliability of the scales did not bias a comparison of their stabilities. Structural equation models were used to estimate the relative stability of gender and marital duration groups. All the measures were found highly stable

over the 8 years of the panel study. Correlations between waves over approximately 3 years when adjusted for attenuation due to reliability were found to be in the .8 to .9 range. Overall, no component of marital quality was more stable than the other. The only difference found for any of the five subscales was significantly less stability in marital problems among respondents' marriages of less than 5 years.

It is possible that the highly stable nature of the marital quality items may reflect stable characteristics of the individuals or the tendency of persons to consistently report similar evaluations regardless of their actual relationship. Johnson, Amoloza, and Booth (1992) examined the 37 persons in the panel study who had divorced and remarried by the third wave. The correlations of marital quality they reported while still in their first marriage were compared to the reports they provided for their second marriage. These correlations were very low, mostly negative, and nonsignificant, suggesting that persons appear to take the conditions of the dyad into account in their ratings.

Research studies making use of the Nebraska Marital Quality Scales in multiwave panel models would not have been possible, or would have been more limited, if the marital quality measures had not been separated into separate scales. A problem with the measures is that some scales do not meet normal criteria for satisfactory reliability ($r_{xx} > .8$). Use of large samples and models that incorporate assumptions about measurement error can go a long way to eliminate this as a serious concern. The five scales do not encompass all the characteristics of the marital relationship that are normally viewed as important. Measures of intimacy, communication, and cohesion are omitted and would need to be added for some models of the marital process.

Single-Item Measurement of Marital Happiness

Generally, reviews of assessment instruments discount or ignore single-item measures (e.g., Norton, 1983; Sabatelli, 1988). The difficulties in estimating reliability and obtaining sufficiently high levels of it, the lack of precision afforded by restricted response categories, and the limited domain of content that can be covered by single-item measures often leave little to recommend. In the field of marital quality, measurement of marital happiness by a single item not only has a long history but also a large and significant body of current use, particularly in issues related to cohort, period, and selection effects in marital quality over the life course (Glenn, 1991). This reflects the availability of only one item indicating marital happiness/satisfaction on many large national surveys.

The General Social Survey, an annual interview survey of a national sample of respondents, has included a one-item measure of marital happiness since 1973. This represents a unique and valuable source of trend data that can help separate cohort and period effects from changes in individuals as they traverse the marital life course (Glenn, 1990). The recent and widely available National Survey of Families and Households (NSFH; Sweet, Bumpass, & Call, 1988), which contains over 7,000 items on a national probability sample of over 13,000 persons, includes only one item tapping marital happiness/satisfaction and one item tapping divorce proneness. These data will be available soon in panel form because the second wave of a 5-year panel has recently been completed. Several other large national surveys contain only single indicators of important marital quality concepts (Kolb & Straus, 1974).

The single-item measure of happiness takes several forms and varies primarily in the number of response categories. The most common form asks the respondents to evaluate how happy they are with their marriages. Three response categories are most common (very happy, pretty happy, not too happy), although the NSFH data make use of a 7-point scale (from very happy to very unhappy). Although there has been no explicit attempt to estimate the reliability and validity of this item, evidence from several sources can be used for this purpose. A similar happiness item is found in most marital quality measures, including the Locke-Wallace MAT, Spanier's DAS, the Nebraska Marital Happiness Scale, and several other scales not reviewed here. Orden and Bradburn (1968) used the single-item report of marital happiness to validate their balance scale.

Factor analyses of these scales often show that the single global item of marital happiness generally has the highest communality of any item (Sharpley & Cross, 1982), suggesting it is the best single item indicator of the scales. Responses to the item appear stable. Orden and Bradburn (1968) found high test-retest correlations for the item in a short-term panel study (correlations using gamma between .82 and .94).

A major criticism of the single-item measure of marital happiness is that it is highly skewed. In most samples 60% to 80% of the respondents select the most happy category; a very small proportion, normally less than 3%, select the not too happy response in the most common three-category version of the item. Studies using this item over the last five decades generally find similar patterns (Orden & Bradburn, 1968; Glenn, 1991). The item normally remains skewed even when more categories are available. The NSFH contains seven

response categories, but the modal response is still in the highest happiness category.

The marital happiness item has been most extensively analyzed by Glenn and his associates, primarily using data collected in large national data sets (Glenn & Weaver, 1978a; Glenn & Weaver, 1978b; Glenn & McLanahan, 1982; Glenn, 1989; Glenn, 1991). Glenn (1991) concludes that the item is unlikely to be biased in assessing change in trend studies, although he acknowledges that it may be biased by social desirability. He compared annual trends in the percent responding very happy from 1973 through 1988 and found a significant decline in this percent over the period. This was the first study to find a trend in the United States in the last two decades toward lower reported levels of marital happiness.

Reliance on single-item measures of marital quality is not recommended. When the use of such measures is the only way to make inferences about trends or to access large, nationally representative, longitudinal samples, then more effort needs to be devoted toward assessing the psychometric properties of these indicators so they can be used in the most valid manner.

METHODOLOGICAL ISSUES IN ASSESSING MARITAL QUALITY IN LIFE COURSE RESEARCH

In this section four methodological issues in the assessment of marital quality are examined. These issues have been selected because of their relevance to the study of the marital relationship over the course of the marriage. The first issue examined is social desirability response tendency in marital quality scales. Some scholars have discounted any attempt to measure married persons' evaluation of their marriage because of the strong tendency to want to report the marriage in a positive light. The evidence for this is examined and the consequences for life course studies is explored. The second issue is the influence of selection effects on inferences made from research findings on married persons. Selection of persons out of the pool of married persons through divorce is an increasing problem that affects both cross-sectional and longitudinal studies. A third issue returns to the problem of single-item indicators of marital quality. Focusing on the marital happiness item, estimates of reliability and stability in panel studies are developed. The final issue examines problems in estimating the reliability and stability of marital quality in panel studies. For several of these issues, data from the four waves of a national sample of married persons were used to illustrate problems and suggest solutions.

Marital Conventionalization and Marital Quality

A study by Edmonds (1967) introduced the concept and measurement of marital conventionalization to the marital quality literature. Marital conventionalization is the tendency for married persons to rate their marriage in more positive terms than is actually true of the relationship. The method of assessing conventionalization was modeled after the techniques used to measure social desirability. Both include several statements that are unlikely to actually occur to which the respondent is asked to give a true or false response. Methodological concerns have been raised by the high correlations ($r = .3$ to $.7$) that have been observed between marital conventionalization and a variety of evaluative measures of marital quality (Fowers & Pomerantz, 1992). This has led some family scholars to question the value of subjective assessment of marital quality (Hicks & Platt, 1970; Edmonds, 1967). For example, some researchers have found that when marital conventionalization is controlled, the effect of other variables on marital satisfaction/adjustment is substantially reduced (Edmonds, Withers, & Dibatista, 1972).

The important question for the assessment of marital quality is whether this tendency to give improbably high ratings to marriages is itself an indicator of marital quality or is a contaminant that biases most marital quality measures. Recent research provides a strong indication that conventionalization is more a measure of marital quality than it is a measure of a marital social desirability response tendency. In an extensive review of the research on marital conventionalization, Fowers and Pomerantz (1992) conclude that it behaves more as another indicator of marital satisfaction than a social desirability response set. This argument is supported by factor analyses that find that the items load on the same factor as marital satisfaction items (Hansen, 1981), and by the low relationship of conventionalization to other social desirability measures.

There is some empirical evidence of a substantial relationship between religiosity and marital conventionalization (Edmonds, Withers, & Dibatista, 1972). Some researchers in marital happiness have discounted the relatively strong effects of the importance of religion in the respondent's life (Glenn & Weaver, 1978a) as reflecting merely a response bias. Concern that marital quality measures are highly biased by marital social desirability led Glenn (1991) to examine whether the decline in marital happiness observed over the last two decades reflects only increases in openness about intimate relationships and the subsequent lesser need to exaggerate. Although there have

not been trend studies of changes in marital conventionalization over time, indirect evidence suggests that declines in marital happiness are not the products of response biases. Glenn (1991) argued that if marital happiness is being more accurately measured in recent years, this should also increase the relationship between marital happiness and other variables such as general happiness. The virtually unchanged relationship between these two types of happiness over a 14-year period makes it unlikely that more honest reporting accounts for the observed decline in marital happiness.

Similar logic can be applied to the strong relationship between the perceived importance of religion in life and marital happiness. If the higher marital quality of more religious persons is primarily a reporting bias, one would expect that the relationship between evaluative marital quality variables and behavioral ones would be weaker among highly religious than among less religious married persons. Booth and Johnson (1992) examined the relationship between marital happiness in 1980 and the occurrence of a divorce or separation within the next 8 years. Marital happiness and importance of religion were both found to be significant and strong predictors of subsequent divorce. Much, but not all, of the effect of religious importance on divorce was through marital happiness. The effect of marital happiness on divorce, however, did not vary by level of religious importance as would be the case if marital happiness had a different meaning for religious and non-religious people. Therefore, it is unlikely that the high marital happiness levels of more religious persons can be discounted as a response bias.

One additional piece of evidence from life course research casts doubt on the likelihood that measures of marital quality are heavily biased by personal response tendencies unrelated to the nature of the marital relationship. If variance in marital quality was primarily a trait unrelated to the marital relationship, it would be expected that as a person moves from one marital relationship to another, there should be a consistency in their tendency to evaluate any marriage. The low, mostly negative and nonsignificant correlations over time between marital quality scale scores in 1980 when respondents were married to one spouse and in 1988 when they were married to another spouse make it unlikely that factors not related to the marriage are responsible for the responses (Johnson, Amoloza, & Booth, 1992).

Although much more research needs to be done on the intriguing tendency for people to view their marriages in very positive and exaggerated ways, particularly in longitudinal and trend studies, the body of evidence points to marital conventionalization as another

measure of marital quality and not a source of potential bias in drawing conclusions about patterns of change in marriages over the life course.

Selection Effects

There has been much research on marital quality over the family life cycle that makes use of cross-sectional samples to reach conclusions about how marital quality varies as a marriage moves through the stages of the family life course (Burr, 1970; Rollins & Cannon, 1974; Anderson, Russell, & Schumm, 1983; Spanier, Sauer, & Larzelere, 1979). It is undoubtedly clear to these researchers that the results could be seriously biased by selection and cohort effects. For example, differences in marital happiness between marriages of 5 and 25 years duration may not reflect the effect of duration on happiness. Not only is the group of couples who have been married for 25 years likely to be in a select group of surviving marriages, they also are likely to have gotten married in a period with different cultural, social, and economic climates than those married 5 years ago. Differences in marital happiness may not reflect a change at all, but represent a difference in marriage cohorts and the different social and marital characteristics of marriages that survive 25 years in a society with high divorce rates.

Most studies of the effects of socioeconomic, background, and structural variables on marital quality only study currently married persons. It is possible, however, that variables strongly related to marital quality may show no effects in such an analysis (Glenn & Weaver, 1978a). For example, if the presence of a premarital birth is strongly related to low marital quality in a subgroup of respondents, it is likely that this group would be selected out due to divorce. A study of the relationship between marital happiness and the presence of premarital birth using a sample of current married respondents may find no effect because the group with the largest effects has been selected out.

Glenn and Weaver (1978a) used a similar argument to account for the small or nonexistent effects of several social, economic, and demographic variables on marital happiness in several national surveys. For example, early age at marriage, which has been found to be a strong predictor of divorce (Bumpass & Sweet, 1972), was not significantly related to marital happiness. They argue that the surveys they examined were conducted (1973-1975) when divorce rates were increasing rapidly and selection of unhappy marriages out of the pool of currently married persons was high. This would attenuate the effect of such variables in the cross-sectional analysis.

The selection argument has also been used to explain the highly skewed distribution of many marital happiness and satisfaction variables (Glenn & Weaver, 1978a; Orden & Bradburn, 1968). The very small proportion reporting their marriages as not too happy may reflect that such persons move quickly out of marriage.

Because marital happiness is a strong predictor of divorce, selection should affect both the distribution of marital happiness scores and the relationship of happiness to social variables. However, other evidence raises serious doubts that selection alone is the basis for negative findings and the small proportion reporting low happiness. Donohue and Ryder (1982) examine both issues. Studies since 1938, when divorce rates were much lower, find nearly identical distributions on the responses to a global marital happiness item to those found in recent decades where disruption due to divorce is more common. If the selection argument were valid, earlier studies should find a larger proportion of unhappy persons, which they do not. They also replicate Glenn and Weaver's (1978a) regression analysis that used data from the 1970s with similar national survey data from the 1960s. Because divorce rates were lower in the 1960s, they argued that the selection effect should be smaller. The effects of social and demographic variables on marital happiness were very similar in both decades. This finding makes it unlikely that the higher selection into divorce in the 1970s was attenuating the findings.

One solution to the selection problem in making inferences from cross-sectional data is to study a closed population, one in which few people enter or leave. This is difficult in marital quality research because it makes no sense to assess the marital quality of persons who have not yet married or are no longer married. One possible solution is the concept of marital success (Glenn, 1990; Glenn, 1991). Marital success distinguishes marriages that are still intact with both partners viewing it as satisfactory from failed marriages or marriages in which at least one partner views it as unsatisfactory. Glenn (1991) combines information on divorce and separation with marital happiness rating to classify marriages as successful or not. He then empirically examines trends in marital success by years since first marriage and by period. Although the measure is relatively crude, some of his findings present a sobering view of the chances for marital success in the 1980s. For persons in the 1980s who were first married 20 to 24 years ago, only 32.5% are classified as successful in their marriage. Even lower rates are found for selected demographic groups (Glenn, 1989).

One of the most widely accepted findings in change in marital quality over the life course is that the likelihood of divorce declines

with marital duration (Bumpass & Sweet, 1972). However, there is evidence that this effect, with the possible exception of declines in the first 2 or 3 years, is primarily the result of selection. High-risk marriages are selected out early, leaving only those with relatively low risks in the pool of married persons. Johnson, Amoloza, and Booth (1992) found in a sample of married persons followed over 8 years that mean scores on the Nebraska Divorce Proneness Scale did not decline and that the scale was very stable. Use of panel data allows some control for the problem of selection, but even here care must be taken. Panel studies are susceptible to high attrition rates, particularly panels followed over many years. The marriages that remain may be selective in many ways that can bias the findings. For example, although there was no evidence that persons leaving the three-wave panel of marriages had higher scores on Divorce Proneness or any of the other marital quality scales, it is possible that unspecified factors select out persons more subject to change. If so, this would reduce the external validity of the findings.

Single-item Measures of Marital Quality

A significant portion of the research on marital quality relies upon single-item measures of marital happiness. Almost all of the research making use of large, nationally representative samples relies on single-item measures (Glenn, 1990). These items are frequently highly skewed in the positive direction, have only three to seven response categories, and have unknown reliability and stability.

Low reliability and limited response categories are not serious problems when the item is used as a dependent variable in regression-based models in large samples (Johnson & Creech, 1984). Both tend to introduce random errors that attenuate statistical power rather than bias the estimates of effects. Concerns continue to persist, however, that the low reliability and precision of single-item measures may contribute to the inability of studies using the single-item indicator to replicate findings from smaller samples that make use of multiple-item scales (Donohue & Ryder, 1982). The more serious problems occur in panel studies. Difficulties in estimating the reliability affect the ability to accurately estimate stability. Estimation of change is hampered by ceiling and floor effects introduced by skewed distributions and few response categories.

Methods for estimating the reliability of single-item measures in panel models have been developed, but have not been applied to the basic marital happiness item. Panel models designed to explore the causal linkages between marital quality and other aspects of the

marital relationship (e.g., Zuo, 1992) need to include information on the measurement error in the indicators used to avoid biasing the results. This section makes a contribution to these unsolved problems by comparing regression results from the single-item happiness measure with those from a more psychometrically sound scale and by exploring the reliability and stability of the item in a four-wave panel model.

When Donohue and Ryder (1982) ruled out selection as an explanation for the small and generally nonsignificant effects of socioeconomic and demographic variables on marital happiness found by Glenn and Weaver (1978a), another explanation for the generally negative findings from large national surveys regarding these variables was needed. One suggestion was that perhaps the single-item happiness measure was so flawed that it produced meaningless results. If this were the case, then regression models making use of the single item should yield weaker results than models that measure marital happiness with a reliable multi-item scale. A test of this is found in Table 1. Two regression models are computed for the 1980 wave of the four-wave panel study discussed elsewhere in this chapter. Most of the demographic and social variables used as predictors by Glenn and Weaver (1978a) and Donohue and Ryder (1982) are included. One model uses the Nebraska Marital Happiness Scale (Johnson, White et al., 1986) as the dependent variable. The other uses the standard global happiness item with three response categories (very happy, coded 3; pretty happy, 2; and not too happy, 1). Standardized regression coefficients (betas) are reported so the relative effects can be compared.

The analyses provide some evidence that the poorer measurement properties of the single-item measure attenuate the effects, but other findings cannot be so clearly interpreted. More variance is explained in the scale than in the item (5.6% to 4.9% but the difference is not substantial. Five of the independent variables were statistically significant related to the scale whereas only four reached significance when the item was the dependent variable. However, only two variables were significant in both models: respondent is nonwhite and religion is important in life. Gender, age, and total family income were only significant in the scale regression, whereas education and husband's occupational status were only significant for the item regression. The direction of all the effects (whether significant or not) was the same in both models. Generally, the results conclude that the single-item measure of marital happiness is quite robust. Differences found may be more substantive than methodological. There is little

Table 1. Comparison of regression models in which marital Happiness is measured by the single item global Happiness and measured by the Marital Happiness Scale. ($N = 1,888$)

Independent Variables	Global Happiness	Marital Happiness
	Item	Scale
	Beta	Beta
Age of Respondent in Years	-.045	-.109*
Total Annual Family Income	.033	.075*
Years of Schooling Completed	-.087*	-.046
Husband's Occupational Status	.071**	.033
Wife Work Full Time	-.015	-.037
Wife Work Part Time	.007	.002
Respondent's Gender (M=1)(F=2)	-.050	-.128*
Respondent is Non-White	-.116*	-.085*
Religion importance in life	-.164*	-.165*
Children under age 5 in HHold	.004	.001
Children under age 12 in HHold	-.040	-.053
Number of Children under 18	-.053	-.065
R-Squared	.049	.056

* Statistically significant at the .01 level.

indication in this analysis that findings from studies making use of the single-item measure of marital happiness are suspect.

Application of single-item measures in panel data may be more problematic. To date, the only panel analyses employing the single item were by Bradburn (1969) and Orden and Bradburn (1968). They report results of the test-retest stability of the global happiness item with three-response categories and make some inferences about change in marital happiness relative to change in positive and negative affect (Bradburn, 1969). Making use of cross-classification techniques, Bradburn (1969) concluded that changes in marital happiness over the short period of their panel were more likely to reflect changes in negative than in positive affect in the marriage. A nonconventional analysis method used in the study limits further exploration of these results for biases and methodological problems.

As was clearly shown by Duncan (1969), causal panel analyses that do not consider measurement error can produce results that are seriously biased. Therefore, estimates of the reliability and stability of single-item measures are needed. Heise (1969) presented a method for estimating the reliability and stability of indicators in panel studies

with three or more waves. This technique has been applied to estimate the reliability of single items on public opinion surveys (Alwin & Krosnick, 1991; Jagodzinski & Kühnel, 1987). Data from the national four-wave panel are used to provide reliability and stability estimates for the single-item happiness indicator.

Test-retest correlations have generally been the only method available for estimating the reliability of single-item measures. Such correlations, however, are likely to be affected by both the reliability and true score change in the measure. The method, originally proposed by Heise (1969) and modified by Wiley and Wiley (1970), separately estimates reliability and stability if three or more waves of panel data are available and if certain assumptions are made. Figure 1 presents a basic model for such an analysis. The variables in circles are unmeasured variables representing the true score component of marital happiness (MH) in each of the panel years. The indicators in the rectangles are the measures of marital happiness, in this case the global happiness item (GH). The λ coefficients are estimates of the relationship between the true score and the measure, and the ϵ coefficients are the measurement errors in the indicators. The MH variables are assumed to be related to one another in a simplex or lag-1 manner. This means that MH at time t is only directly related to marital happiness at time $t+1$. Any relationship between MH_1 and MH_3 is through MH_2 .

Further restrictions are required to identify mathematically the equations. Two alternative sets have been proposed. The first assumes that the reliability is the same in each wave (Heise, 1969). In Figure 1, all the ϵ coefficients (in their standardized form) would be assumed to be equal. This reduces the number of unknowns sufficiently to just identify a three-wave and to overidentify a four-wave model by 2 degrees of freedom. The second choice is to assume that the measurement error variances of the indicators are the same for each wave (Wiley & Wiley, 1970). The degrees of freedom are the same in both models. Both can be estimated with three or more waves of data available with LISREL VII (Jöreskog & Sörbom, 1988).

The correlations among the four waves of data for the global happiness item are presented in Table 2. Given that each wave is 3 to 4 years apart, the test-retest correlations are quite high, averaging around .5 for adjacent waves. The mean score declines steadily over the 12 years, which is a pattern also found for the Marital Happiness Scale.

Estimates of the reliability and stability for the two alternative models are presented in Table 3. The equal reliability model yielded a reliability score (calculated as the square of the standardized lambda

Figure 1. Four-wave path model for Global Happiness as single indicator of Marital Happiness.

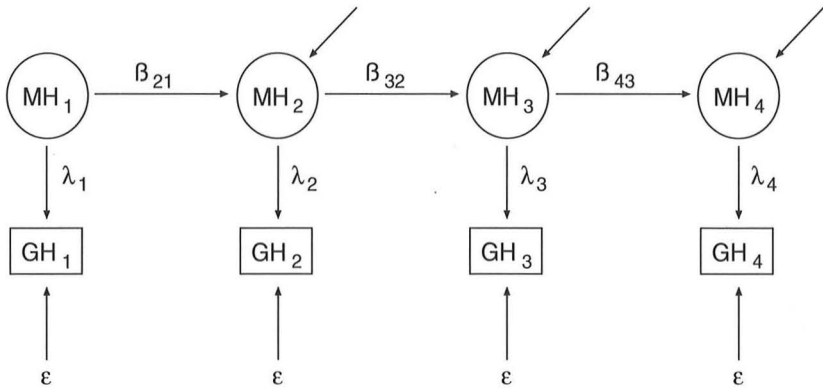


Table 2. Correlations, means and standard deviations among the for the global happiness item in the four waves. (N = 945)

	Wave 1	Wave 2	Wave 3	Wave 4
Wave 1	1.0000			
Wave 2	.5298	1.0000		
Wave 3	.4494	.4849	1.0000	
Wave 4	.4150	.4276	.5076	1.0000
Means	2.7027	2.6227	2.6081	2.5572
S. D.	.4817	.5141	.5332	.5692

coefficient in the model) of .563. The equal error variance model found the lowest reliability in Wave 1 and the highest in Wave 4. Stabilities were high in both models. The stability was highest from the first to the second wave. The standardized stability between these waves for the equal error variance model exceeded 1, an illogical value that suggests specification errors in the model. Other stability estimates were in the .85 to .95 range, which are still extraordinarily

Table 3. Reliability and stability coefficients for the single global happiness item using the equal reliability and the equal error variance models.

	Equal Reliability model	Equal Error variance model
<i>Reliability</i>		
Wave 1	.563	.484
Wave 2	.563	.549
Wave 3	.563	.578
Wave 4	.563	.630
<i>Stability</i>		
Waves 1 - 2	.942	1.019
Waves 2 - 3	.855	.854
Waves 3 - 4	.901	.848

high. Either MH is an extremely stable trait over a 12-year period, or the model is misspecified in some way and yields invalid results.

There is a good basis for questioning the equal error variance model. Because the marital happiness item is so highly skewed, the mean score is closely related to the standard deviation. The proportion of persons saying they are not too happy is so small that the item effectively behaves like a dichotomy. For dichotomous items the standard deviation is a perfect, but nonlinear, function of the mean score ($sd = \sqrt{pq}$). Because the standard deviations vary with the means scores it makes more sense to assume that reliabilities are equal and differences in variances are a function of error and not the latent trait. The more mathematically meaningful estimates from the equal reliability model support this view.

Because the global happiness item is part of the Nebraska Marital Happiness Scale (Johnson, White et al., 1986), an estimate of reliability can be computed from an item analysis of the scale. For the 1980 wave, the corrected item-total score correlation of the global item with the scale was .692. Correcting for the higher reliability of the multiple-item scale ($r_{xx} = .851$), this would yield an estimated reliability for the global item of .563. This is identical to that obtained in the four-wave model assuming equal reliabilities.

Although the convergence of estimates from internal consistency and test-retest methods should increase confidence in the accuracy of this estimate of the reliability, the very high stability over 3 or more years implied by such a reliability estimate questions this confidence. One resolution is to consider the latent variable implied by the marital

happiness item. The measurement model implies that MH is tapping global happiness, free of measurement error. However, the lag-1 model specifies that MH has a direct causal effect on MH in the next wave. It may be reasonable to assume that this is an incorrect model. Marital happiness, even free of measurement errors, may not be causally related to happiness in subsequent years. Instead, MH may be an outcome of other unmeasured characteristics of the marital environment that are quite stable and autocorrelated to a lag-1 process. If this were the case, the reliability of the global happiness indicator would be underestimated and the high stability coefficients would reflect that latent trait and not marital happiness net of measurement error.

Hargens, Reskin, and Allisson (1976) were confronted with a similar problem while trying to estimate measurement error in indicators of scientific productivity. They conclude that when only a single indicator is available, it is not possible to infer the nature of the unmeasured variable estimated by the model. Attempts by Jagodzinski and Kühnel (1987) to solve this problem making use of polychoric correlations suggests a possible solution. The solution proposed here builds on their work, but has not been presented in this form in literature.

A model assuming two latent traits is shown in Figure 2. Marital Happiness free of error (MH) is not assumed to directly affect itself in subsequent waves. Another latent variable, labelled Marital Environment (ME), has effects on MH and is causally related to itself in a lag-1 pattern. The problem with the model is developing a method of estimating both the measurement errors in the global happiness indicator and the effects of Marital Environment. The model is underidentified and no solution in the single variable case has been found in the literature. Combining a polychoric model with a four-wave path model is a key to estimating this model (Jöreskog & Sörbom, 1988; Jagodzinski & Kühnel, 1987).

A polychoric correlation provides an estimate of the relationship between two unmeasured continuous and normally distributed variables implied by crudely categorized indicators with a small number of ordered response categories (Jöreskog & Sörbom, 1988). The tetrachoric correlation is the version of this coefficient used for dichotomous variables. The cross-classification of the categories of the two indicators is fit to a model which assumes that this pattern was generated by two normally distributed, continuous variables. Polychoric correlations are correlations between the indicators after removing the effects of categorization errors.

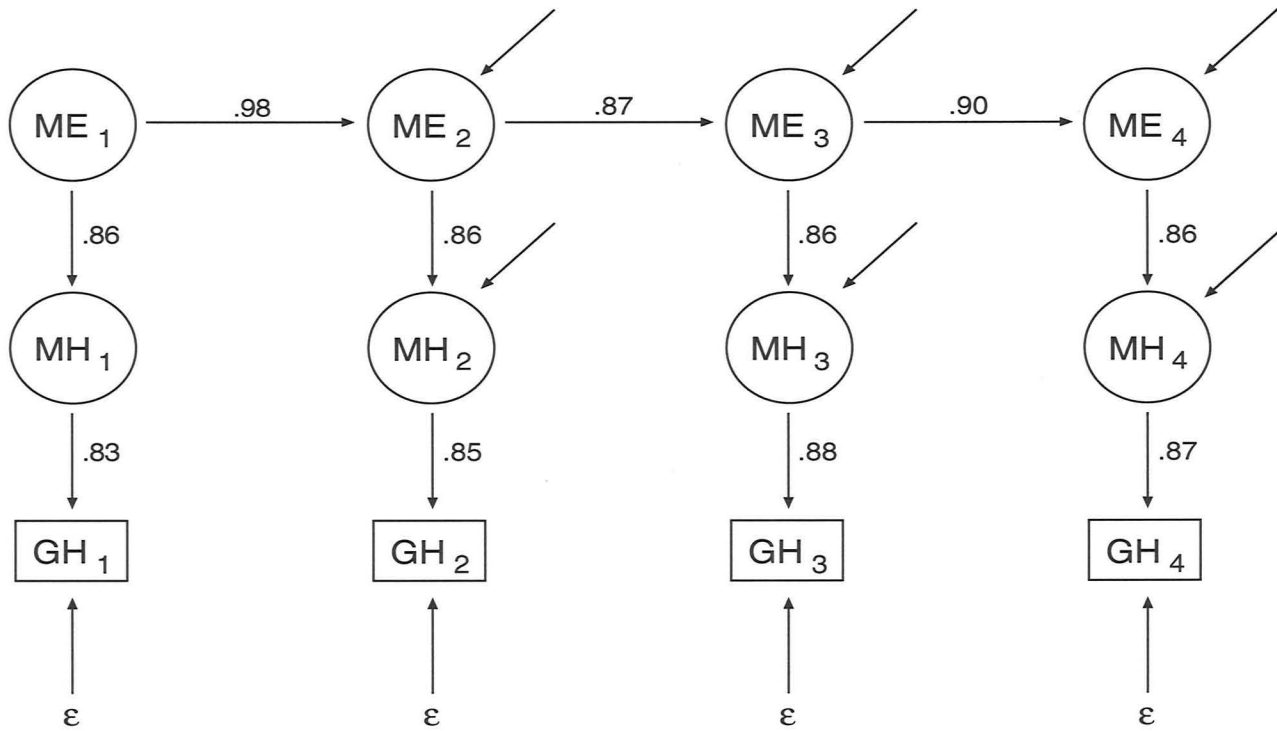


Figure 2. Four-wave path model for Global Happiness including latent Marital Environment.

In the four-wave panel model, polychoric correlations are used to estimate the relationship among the unmeasured MH variables. Fitting the Hiese model extended to four waves to these correlations yield estimates of the paths among the unmeasured variables. The parameters linking MH and the global happiness indicators were estimated in a second stage by a LISREL model for the correlations among the indicators. The estimates from the first stage were set as fixed values and the paths between MH and GH were estimated.

Standardized path coefficients estimated in this model are presented in Figure 2. Marital Environment was very stable between waves and strongly affected Marital Happiness (.86). The paths from MH to GH range from .825 to .875, which imply reliabilities from .68 to .76. These are higher than the previously estimated reliability of the marital happiness item. The model is also consistent with the theoretical expectation that happiness would be expected to show reasonable instability over time (Campbell, Converse, & Rodgers, 1976). Test-retest correlations for MH implied by the model are approximately .7 between adjacent waves. These are more reasonable than the correlations in the .8 to .9 range found in the first model.

It is clear from this analysis that establishing the reliability and stability of single-item marital quality indicators is not a simple task. Multiple indicators of marital happiness would have simplified the task and reduced the need for as many untestable assumptions. Other possible sources of error in these models, such as serially correlated measurement errors, which could not be addressed here, might also have been evaluated (Johnson & Amoloza, 1989). Although these qualifications suggest that single-item measures should be avoided whenever possible, the overall conclusion reached about this single-item measure of global marital happiness is that it is a reliable and robust indicator of happiness. Confidence can be placed in previous findings making use of the item and future use of this and similar items in life course studies appears warranted.

Marital Quality Scales Used in Panel Studies: Reliability and Stability Issues

Most marital quality measures are multiple-item scales that have been psychometrically evaluated to one degree or another (Sabatelli, 1988). Internal consistency reliability is normally (but not always) computed, the factor structure is examined, and occasionally test-retest reliability results are reported. None of these steps guarantee that the scale will behave acceptably when used in panel models

designed to assess the causal structure and process of marital change over the life course. In estimating such models it is normally necessary to incorporate estimates of measurement error to avoid biased estimates. Although internal consistency is generally adequate as an estimate of reliability in cross-sectional studies, reliability based on over-time correlations becomes very important in panel designs. Problems arise when internal consistency and test-retest reliability estimates do not coincide.

Johnson and Amoloza (1989) examined three marital quality scales (Marital Happiness, Marital Interaction, and Marital Disagreements) in a three-wave panel. The test-retest was higher than internal consistency reliability for one scale, the two estimates were approximately the same for another, and internal consistency was highest in a third. Serially correlated measurement errors were generally found responsible for the difference when test-retest exceeds internal consistency reliability. When internal consistency was highest, the misspecification of the causal lag among the unmeasured variables appears responsible. As shown in the analysis of the global happiness item, estimates of reliability can greatly affect those of stability.

Large differences between test-retest and internal consistency reliability estimates pose problems for the researcher. The normal solution to these problems is to have multiple indicators for all important variables. Although two or more scales may be used as indicators for each variable (Johnson, White et al., 1986), scales are often disaggregated into subscales or separate items (e.g., Zuo, 1992; Johnson & Amoloza, 1989). Disaggregation of scales that are multi-dimensional presents a problem because each dimension must be represented by a separate latent variable in the model analyzed. Because each latent variable should have at least two indicators, models can quickly become unmanageable. For example, assume a researcher is exploring the relationship between the quality of the marital relationship and psychological depression in a three-wave panel study. Measures of marital happiness, disagreements, and interaction would be needed to explore the reciprocal relationship to depression. Additionally, five or six control and background variables (marital duration, gender, socioeconomic factor, children, etc.) would be needed. If each of the marital quality and depression variables were measured by disaggregating them into the separate items, and the Nebraska Marital Quality Scales were used, then these three scales would require 18 indicators per wave, depression may take 7 more, and the controls at least 6 (assuming no multiple indicators for these). Over the three waves, this model would have 93 indicators.

The model could be simplified using only one indicator for each variable and correcting them for attenuation due to unreliability by estimating the error variances from internal consistency reliability estimates (a procedure often proposed—e.g., Hayduk, 1989). Alternatively, the error variances with single indicators could be estimated by imposing restrictions on the lagged process using models similar to those used to evaluate marital happiness indicators discussed above (Werts, Jöreskog, & Linn, 1971). Both approaches carry the risk of seriously misspecifying the model and producing biased results.

This problem is illustrated with an analysis of the reliability and stability of the Nebraska Marital Happiness Scale making use of the same four-wave dataset discussed above. Table 4 presents the correlations, means, and standard deviations for the scale in each of the four waves for all respondents with complete data and married to the same person over the 12 years of the study. Over this period, mean scale scores declined and the standard deviations increased.

Three separate estimates of reliability and stability were computed. The Heise model assumes equal reliabilities, the Wiley and Wiley model assumes equal error variances, and the third model corrects the covariance matrix for attenuation with coefficient alpha as the reliability estimate. A comparison of the alternative models is presented in Table 5. The most conspicuous difference is the large discrepancy in reliability, and subsequent stability, between the panel and the correction for attenuation models. The scale is much less reliable and far more stable when panel methods are used.

Table 4. Correlations, means, and standard deviations among the Marital Happiness Scale across the four waves. ($N = 900$)

	Wave 1	Wave 2	Wave 3	Wave 4
Wave 1	1.0000			
Wave 2	.6199**	1.0000		
Wave 3	.5546**	.6123**	1.0000	
Wave 4	.5183**	.5225**	.6334**	1.0000
Mean	29.17	28.50	28.22	27.97
S. D.	3.406	3.657	3.826	4.150

** Statistically significant at .01 level.

Table 5. Reliability and Stability of the Marital Happiness Scale across four waves for three alternative models.

	Model		
	Equal Reliability (Heise)	Equal Error Variance (Wiley & Wiley)	Correction For Attenuation (Alpha)
<i>Reliability</i>			
1980	.689	.621	.831
1983	.689	.671	.850
1988	.689	.706	.865
1992	.689	.745	.882
<i>Stability</i>			
Wave 1–2	.913	.964	.738
Wave 2–3	.878	.878	.714
Wave 3–4	.907	.870	.725

It is likely that the same problem noted for the analysis of the global happiness item may be occurring here (Figure 1). An unmeasured variable, labelled Marital Environment, may be driving the stability of Marital Happiness. With only one indicator of Marital Happiness, there are not enough degrees of freedom in the model to estimate the effects. The required multiple indicators can be obtained by disaggregating the scale into separate items. However, with 11 items in this scale, this would produce an unwieldy model. A compromise is to create subscales from among the items to yield at least two indicators.

The Marital Happiness Scale items were factor analyzed to aid in identifying two or three meaningful subscales. Although all items had their highest loadings on the first unrotated factor, a good indicator the scale is unidimensional, two- and three-factor rotated solutions were explored. A three-subscale solution was the most satisfactory. The items in each scale are shown in Table 6. Scales created were (A) intimacy, (B) companionship, and (C) relationship satisfaction. Correlations, means, and standard deviations among the subscales in all waves are given in Table 7. Figure 3 presents the path

model fit to the data. The model was fit to both the observed covariance and correlation matrices. Models based on covariance allow the retention of the metric of the indicators. This insures that the unmeasured variables of the same concept are equivalent across waves, and is generally preferred (Alwin & Jackson, 1980). In this case, however, the highly skewed subscales create a mathematical dependence between the group means and standard deviations. These artifactual fluctuations across the waves in the standard deviations affect the covariances, not the correlations. Here, the analysis of the

Table 6. Marital Happiness Scale items and subscales.

Subscale A

1. How happy are you with the amount of understanding you received from you (husband/wife)? Would you say you are very happy, pretty happy, or not too happy with this aspect of your marriage? (response categories the same for items 1 thru 7)
2. With the amount of love and affection you receive?
3. The extent to which you and your spouse agree about things?
4. With your sexual relationship?

Subscale B

5. With your spouse as someone to take care of things around the house?
6. With your spouse as someone to do things with?
7. With your spouse's faithfulness to you?

Subscale C

8. Taking all things together, how would you describe your marriage? Would you say that your marriage is very happy, pretty happy or not too happy?
9. Compared to other marriages you know about, do you think your marriage is better than most, about the same as most, or not as good as most?
10. Compared to your marriage three years ago, is your marriage getting better, staying the same, or getting worse?
11. Would you say the feeling of love you have for your (husband/wife) are extremely strong, very strong, pretty strong, not too strong, or not strong at all?

Table 7. Correlations, means and standard deviations among the three marital happiness subscales in the four panel waves. (N= 943)

	HA1	HB1	HC1	HA2	HB2	HC2	HA3	HB3	HC3	HA4	HB4	HC4
Correlations												
HA1	1.000											
HB1	.584	1.000										
HC1	.543	.432	1.000									
HA2	.537	.365	.421	1.000								
HB2	.421	.513	.396	.564	1.000							
HC2	.368	.297	.563	.597	.471	1.000						
HA3	.481	.347	.335	.537	.430	.411	1.000					
HB3	.378	.455	.296	.371	.517	.314	.619	1.000				
HC3	.384	.279	.437	.402	.348	.492	.638	.522	1.000			
HA4	.459	.324	.352	.481	.374	.358	.561	.413	.464	1.000		
HB4	.342	.387	.274	.302	.452	.273	.427	.507	.389	.635	1.000	
HC4	.365	.297	.417	.369	.301	.434	.460	.386	.552	.679	.566	1.000
Means	10.064	8.090	10.979	9.807	7.928	10.710	9.731	7.830	10.630	9.636	7.80	10.476
Standard												
Deviations	1.789	1.125	1.194	1.864	1.199	1.345	1.909	1.251	1.419	1.996	1.272	1.590

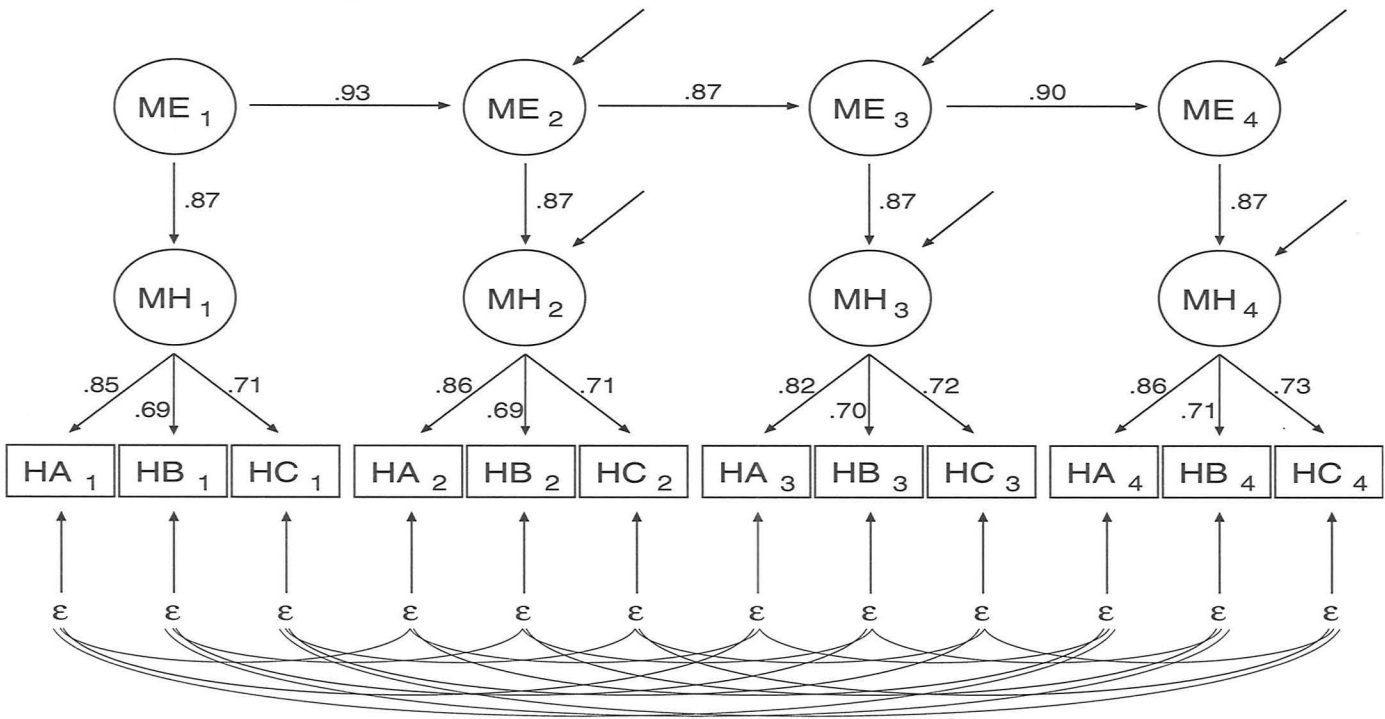


Figure 3. Four-wave path model for three-indicator Marital Happiness with latent Marital Environment variable.

correlation matrix is more reasonable. To ensure comparability across waves in the unmeasured variables, the paths from the MH to the indicators and their error terms are allowed to vary by indicator but are constrained to be the same for each wave.

The model in Figure 3 also assumes that measurement errors of the same scale are correlated across waves. This is the usual assumption in multiple indicator panel models (Jöreskog & Sörbom, 1988). It accounts for the part of the variance that indicators do not share in common that may be correlated across time. The estimates appearing on the model are from the analysis of the correlation matrix. Estimates for the error terms and their intercorrelations are omitted from the diagram to simplify the figure.

Comparisons of the estimates in Figure 3 with those from the model fit to the single happiness item in Figure 2 show remarkable similarities. The pattern of effects among Marital Environment (ME) and Marital Happiness (MH) are almost identical. Perhaps more surprising is the estimate of the relationship between the indicators and MH. The square of this estimate is the measure of reliability. The single-item global happiness scale is found to be about as reliable as the most reliable of the subscales (A: Intimacy). It is considerably more reliable than the four-item scale in which it is included.

This anomaly may reflect two things. Subscale C includes two items (9 and 10) with the lowest item-total score correlations in the Marital Happiness Scale that may be suppressing the subscale's reliability. The model in Figure 3 also includes autocorrelated measurement errors not found in the global happiness model. These errors can include part of the reliability variance in the scale that is not included in the effect from MH to the indicator (Alwin & Jackson, 1980).

An important outcome of this exercise is the stability estimates of marital happiness free of measurement error. The stability estimates are not present as parameters in the model, but can be calculated from the coefficients. The correlations of MH among adjacent waves are $r_{12} = .695$; $r_{23} = .670$; and $r_{34} = .693$. These moderately high estimates of stability raise doubts about the high levels of stability reported for the single indicator panel models in Table 5.

This analysis was designed to illustrate some issues and problems that arise in panel models that require attention in panel studies of marital quality over the life course. The influence of the latent Marital Environment variable is intriguing and is substantively as well as methodologically important. The findings suggest that there are very stable traits in marriages that are strongly linked to marital happiness.

Panel models that include other indicators of marital quality and other more direct measures of the marital environment are needed. Such models will need to demonstrate close attention to the reliability and measurement issues discussed throughout this chapter.

SUMMARY AND CONCLUSIONS

This chapter has examined in a selective way the assessment of marital quality. Although many ways of assessing the quality of marriage have been proposed, it is clear that the variety of measures reflects basic conceptual issues about the meaning of the term. Three perspectives were identified: the marital adjustment perspective with its roots in marital therapy and identification of troubled and well-adjusted marriages, the global satisfaction view that seeks to restrict the definition of marital quality to subjective evaluation of the whole marriage, and a more eclectic approach that groups a series of separate concepts under the umbrella term of marital quality but seeks separate measures for each. The third perspective has the widest acceptance and use in the literature and is consistent with researchers seeking to assess aspects of marital quality in causal life course models.

The next objective was to more closely examine specific measures of marital quality that might have utility in life course studies. The review was restricted to scales and measures with relatively small numbers of items that were designed primarily for research and not clinical purposes. Basic criticisms and concerns raised about the measures and examples of their use in life course research were discussed. More attention was given to an evaluation of the single-item measure of marital happiness than in previous reviews of marital quality measures because most studies making use of large and nationally representative samples employ crude, often highly skewed, single-item measures. Scales found most appropriate for life course work were those that measured only one trait of the marital relationship well.

Four methodological issues in the assessment of marital quality were reviewed. The conclusion that could be drawn from the discussion of methodological issues in the assessment of marital quality is that, with some care in selection of scales, analysis method, and the interpretation of the results, these problems do not seriously impair the ability of the researcher from making valid statements about the quality of marriages. Marital conventionalization, or the tendency of persons to report their marriage in a more positive light than it actually was, has cast doubt on the validity of evaluative measures of marital quality. More recent research suggests that this

tendency is not a typical survey response bias like social desirability, but may be a valid component of how people see these marriages. The fact that marital happiness scales, often viewed as highly contaminated by marital social desirability, are strong predictors of behavior such as divorce, even among groups susceptible to reporting high marital satisfaction, suggests that marital quality measures may not be biased enough to reduce their utility as research tools.

Selection effects from failed marriages are a special problem for researchers making use of cross-sectional data to make inferences about life course changes, but also present problems when longitudinal data are available. In some cases, especially with panel data, it is possible to estimate the effects of selection and take them into consideration in the analysis models.

The problems of single-item measures, quite prevalent in the analysis of marital quality, are examined in detail by concentrating on the global marital happiness item with three response categories. The review of previous work and an empirical analysis of the behavior of the item in four-wave panel suggests that the measure is quite robust and reliable. This reduces concerns that findings from studies using the single-item indicator should be discounted as not sufficiently valid.

The final issue examined concerned problems in estimating the reliability and stability of marital quality measures in panel studies. A panel model for multiple indicators of marital quality was proposed and partially applied to four-wave panel data. This analysis suggested that splitting scales into two or more indicators may be necessary to estimate stability in a valid manner.

Some conclusions can be drawn from this exploration of the measurement of marital quality. First, the debate over what should be called marital quality should be ended. Most researchers now recognize the need to assess the various components of marital quality in separate scales. Use of marital quality to refer only to global assessment of the marital relationship appears too limited and removes a term that has been useful in characterizing research on the marital relationship.

Second, because of resource constraints, work should concentrate on creating short, unidimensional scales for the components of marital quality. Family researchers should work to add these to some of the regular national surveys that are the only real source of large and nationally representative samples. Third, although scales should be short, the researcher should be able to subdivide the scales when two or more indicators of each construct are needed to estimate panel

models of marital quality. Fourth, there is evidence that a single-item scale of marital happiness may be more robust and less biased than expected and appears to be a valid replacement of complete scales of marital happiness. Research making use of single-item marital evaluation measures, particularly when large representative samples are available, has been found to be clearly worth pursuing when more complete scales are not available. Finally, multiwave, multivariate models of the causal relationships between marital quality and other marital variables require multiple indicators of constructs and careful specification of the models to avoid serious bias introduced by the reliability of the measures.

REFERENCES

- Alwin, D. F., & Jackson, D. J. (1980). Measurement models for response errors in surveys: Issues and applications. In K. F. Schuessler (Ed.), *Sociological methodology: 1980* (pp. 69-119). San Francisco: Jossey-Bass.
- Alwin, D. F., & Krosnick, J. A. (1991). Aging, cohorts and the stability of sociopolitical orientations over the life span. *American Journal of Sociology, 97*, 169-195.
- Anderson, S. A., Russell, C. S., & Schumm, W. R. (1983). Perceived marital quality and family life-cycle categories: A further analysis. *Journal of Marriage and the Family, 45*, 126-139.
- Belsky, J., Lang, M. E., & Rovine, M. (1985). Stability and change in marriage across the transition to parenthood: A second study. *Journal of Marriage and the Family, 47*, 855-865.
- Belsky, J., Spanier, G. B., & Rovine, M. (1983). Stability and change in marriage across the transition to parenthood. *Journal of Marriage and the Family, 45*, 567-577.
- Booth, A., & Edwards, J. N. (1989). Transmission of marital and family quality over the generations: The effect of parental divorce and unhappiness. *Journal of Divorce, 13*, 41-58.
- Booth, A., & Johnson, D. R. (1992). *Religious change and change in marital quality*. Unpublished manuscript.
- Booth, A., Johnson, D. R., & Edwards, J. N. (1983). Measuring marital instability. *Journal of Marriage and the Family, 45*, 387-394.
- Booth, A., Johnson, D. R., White, L. K., & Edwards, J. N. (1984). Women, outside employment and marital instability. *American Journal of Sociology, 90*, 567-583.
- Booth, A., Johnson, D. R., White, L. K., & Edwards, J. N. (1986). Divorce and marital instability over the life course. *Journal of Family Issues, 6*, 421-442.

Bradburn, N. M. (1969) *The structure of psychological well-being*. Chicago: Aldine.

Bumpass, L. L., & Sweet, J. L. (1972). Differentials in marital instability: 1970. *American Sociological Review*, 37, 754-766.

Burke, R. J., Weir, T., & DuWors, R. E. (1979). Type A behavior of administrators and wives' reports of marital satisfaction and well being. *Journal of Applied Psychology*, 64, 57-65.

Burr, W. R. (1970). Satisfaction with various aspects of marriage over the life cycle: A random middle class sample. *Journal of Marriage and the Family*, 32, 29-37.

Campbell, A., Converse, P. E., & Rodgers, W. L. (1976) *The quality of American life: Perceptions, evaluations, and satisfactions*. New York: Russell Sage Foundation.

Crane, D. R., Busby, D. M., & Larson, J. H. (1991). A factor analysis of the Dyadic Adjustment Scale with distressed and nondistressed couples. *American Journal of Family Therapy*, 19, 60-66.

Cross, D. G., & Sharpley, C. F. (1981). The Locke-Wallace marital adjustment test reconsidered: Some psychometric findings as regards its reliability and factorial validity. *Educational Psychology Measurement*, 41, 1303-1306.

Donohue, K. C., & Ryder, R. G. (1982). A methodological note on marital satisfaction and social variables. *Journal of Marriage and the Family*, 44, 743-747.

Duncan, O. D. (1969). Some linear models for two-variable panel analysis. *Psychological Bulletin*, 72, 177-182.

Edmonds, V. H. (1967). Marital conventionalization: Definition and measurement. *Journal of Marriage and the Family*, 29, 681-688.

Edmonds, V. H., Withers, G., & Dibatista, B. (1972). Adjustment, conservatism and marital conventionalization. *Journal of Marriage and the Family*, 34, 96-103.

Featherman, D. L., & Lerner, R. M. (1985). Ontogenesis and sociogenesis. *American Sociological Review*, 50, 659-690.

Fincham, F. D., & Bradbury, T. N. (1987). The assessment of marital quality: A reevaluation. *Journal of Marriage and the Family*, 49, 797-809.

Fowers, B. J., & Pomerantz, B. L. (1992, November). *Reconceptualizing marital conventionalization: Examining psychological, relational and cultural explanations*. Paper presented at the 22nd annual Theory Construction and Research Methodology Workshop, Orlando, FL.

Glenn, N. D. (1989). Duration of marriage, family composition, and marital happiness. *National Journal of Sociology*, 3, 3-24.

Glenn, N. D. (1990). Quantitative research on marital quality in the 1980s: A critical review. *Journal of Marriage and the Family*, 52, 818-831.

Glenn, N. D. (1991). The recent trend in marital success in the United States. *Journal of Marriage and the Family*, 53, 261-270.

Glenn, N. D., & McLanahan, S. (1982). Children and marital happiness: A further specification of the relationship. *Journal of Marriage and the Family*, 44, 63-72.

Glenn, N. D., & Weaver, C. N. (1978a). A multivariate, multisurvey study of marital happiness. *Journal of Marriage and the Family*, 39, 269-282.

Glenn, N. D., & Weaver, C. N. (1978b). The contribution of marital happiness to global happiness. *Journal of Marriage and the Family*, 43, 161-168.

Hamilton, G. V. (1929). *A research in marriage*. New York: Albert and Charles Boni.

Hansen, G. L. (1981). Marital adjustment and conventionalization: A reexamination. *Journal of Marriage and the Family*, 43, 855-863.

Hargens, L. L., Reskin, B. F., & Allisson, P. D. (1976). Problems in estimating measurement error from panel data: An example involving the measurement of scientific productivity. *Sociological Methods and Research*, 4, 439-458.

Hayduk, L. A. (1987). *Structural equation modelling with LISREL: Essentials and advances*. Baltimore: Johns Hopkins University Press.

Heise, D. R. (1969). Separating reliability and stability in test-retest correlations. *American Sociological Review*, 34, 93-101.

Hicks, M. W., & Platt, M. (1970). Marital happiness and stability: A review of the research in the sixties. *Journal of Marriage and the Family*, 32, 553-574.

Huston, T. L., & Robins, E. (1982). Conceptual and methodological issues in studying close relationships. *Journal of Marriage and the Family*, 44, 901-925.

Jagodzinski, W., & Kühnel, S. M. (1987). Estimation of reliability and stability in single-indicator multiple-wave models. *Sociological Methods and Research*, 15, 219-258.

Johnson, D. R. (1988). Panel analysis in family studies. *Journal of Marriage and the Family*, 50, 949-955.

Johnson, D. R., & Amoloza, T. O. (1989, April). *Assessing reliability using internal consistency and three-wave panel methods: A synthesis and empirical exploration*. Paper presented at the Midwest Sociological Society Annual Meeting, St. Louis, MO.

Johnson, D. R., Amoloza, T. O., & Booth, A. (1992). Stability and developmental change in marital quality: A three-wave panel analysis. *Journal of Marriage and the Family*, 54, 582-594.

Johnson, D. R., & Creech, J. C. (1984). Ordinal measures in multiple indicator models: A simulation study of categorization errors. *American Sociological Review*, 48, 398-407.

Johnson, D. R., White, L. K., Edwards, J. N., & Booth, A. (1986). Dimensions of marital quality: Towards methodological and conceptual refinement. *Journal of Family Issues*, 7, 31-49.

Jöreskog, K. I. G., & Sörbom, D. (1988). *LISREL 7: A guide to the program and applications*. Chicago: SPSS.

Kimmel, D., & Van Der Veen, F. (1974). Factors of marital adjustment in Locke's marital adjustment test. *Journal of Marriage and the Family*, 36, 57-63.

Kolb, T. M., & Straus, M. A. (1974). Marital power and marital happiness in relation to problem-solving ability. *Journal of Marriage and the Family*, 36, 756-766.

Lively, E. L. (1969). Towards a conceptual clarification: The case of marital interaction. *Journal of Marriage and the Family*, 31, 108-114.

Locke, H. J., & Wallace, K. M. (1959). Short marital adjustment and prediction tests: Their reliability and validity. *Marriage and Family Living*, 21, 251-255.

Locke, H. J., & Williamson, R. C. (1958). Marital adjustment: A factor analysis study. *American Sociological Review*, 23, 562-569.

Margolin, G. (1978). Relationships among marital assessment procedures: A correlational study. *Journal of Consulting and Clinical Psychology*, 46, 1556-1558.

Mattessich, P., & Hill, R. (1987). Life cycle and family development. In M. B. Sussman & S. K. Steinmetz (Eds.), *Handbook of marriage and the family* (pp. 437-469). New York: Plenum Press.

Mitchell, S. E., Newell, G. K., & Schumm, W. R. (1983). Test-retest reliability of the Kansas Marital Satisfaction Scale. *Psychological Reports*, 53, 545-546.

Norton, R. (1983). Measuring marital quality: A critical look at the dependent variable. *Journal of Marriage and the Family*, 45, 141-151.

Nunnally, J. C. (1967). *Psychometric theory*. New York: McGraw-Hill.

Orden, S. R., & Bradburn, N. M. (1968). Dimensions of marriage happiness. *American Journal of Sociology*, 73, 715-731.

Rollins, B. C., & Cannon, K. L. (1974). Marital satisfaction over the family life cycle: A reevaluation. *Journal of Marriage and the Family*, 36, 271-282.

Sabatelli, R. M. (1988). Measurement issues in marital research: A review and critique of contemporary survey instruments. *Journal of Marriage and the Family*, 50, 891-917.

Schumm, W. R., Paff-Bergen, L. A., Hatch, R. C., Obiorah, F. C., Copeland, J. M., Meens, L. D., & Bugaighis, M. A. (1986). Concurrent and discriminant validity of the Kansas Marital Satisfaction Scale. *Journal of Marriage and the Family*, *48*, 381-387.

Sharpley, C. F., & Cross, D. G. (1982). A psychometric evaluation of the Spanier dyadic adjustment scale. *Journal of Marriage and the Family*, *44*, 739-741.

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

Spanier, G. B., & Cole, C. L. (1976). Towards clarification and investigation of marital adjustment. *International Journal of Sociology of the Family*, *6*, 121-146.

Spanier, G. B., & Lewis, R. A. (1980). Marital quality: A review of the seventies. *Journal of Marriage and the Family*, *42*, 825-839.

Spanier, G. B., Sauer, W., & Larzelere, R. (1979). An empirical evaluation of the family life cycle. *Journal of Marriage and the Family*, *41*, 27-38.

Spanier, G. B., & Thompson, L. (1982). A confirmatory analysis of the dyadic adjustment scale. *Journal of Marriage and the Family*, *44*, 731-738.

Sweet, J. A., Bumpass, L. L., & Call, V. R. A. (1988). Design and content of the National Survey of Families and Households (Working Paper NSFH-1). Madison, WI: Center for Demography and Ecology, University of Wisconsin.

Touliatos, J., Perlmutter, B. F., & Straus, M. A. (1990). *Handbook of family measurement techniques*. Newbury Park, CA: Sage.

Werts, C. E., Jöreskog, K. G., & Linn, R. L. (1971). Comment on "Estimation of measurement error in panel data." *American Sociological Review*, *36*, 110-113.

White, L. K. (1983). Determinants of spacial interaction: Marital structure or marital happiness. *Journal of Marriage and the Family*, *45*, 511-519.

White, L. K., & Booth A. (1985). Transition to parenthood and marital quality. *Journal of Family Issues*, *6*, 435-449.

White, L. K., & Booth A. (1991). Divorce over the life course: The role of marital happiness. *Journal of Family Issues*, *12*, 5-21.

White, L. K., & Edwards, J. N. (1990). Emptying the nest and parental well-being: An analysis of national panel data. *American Sociological Review*, *55*, 235-242.

Wiley, D. E., & Wiley, J. A. (1970). The estimation of measurement error in panel data. *American Sociological Review*, *35*, 112-117.

Zuo, J. (1992). The reciprocal relationship between marital interaction and marital happiness: A three-wave study. *Journal of Marriage and the Family*, 54, 870-878.