JAMES M. GRAHAM *Western Washington University*

YENLING J. LIU AND JENNIFER L. JEZIORSKI *Argosy University at Chicago**

The Dyadic Adjustment Scale: A Reliability Generalization Meta-Analysis

We conducted a reliability generalization meta-analysis to examine the internal consistency of Dyadic Adjustment Scale (DAS; Spanier, 1976) scores across 91 published studies with 128 samples and 25,035 participants. The DAS was found to produce total and Dyadic cohesion, Consensus, and Satisfaction scores of acceptable internal consistency, although lower than those originally reported by Spanier (1976). Reliability estimates of these scores did not differ by the sexual orientation, gender, marital status, or ethnicity of the sample. The Affective Expression subscale was found to produce scores with poor Cronbach's alpha across studies. Reliability estimates of Affective Expression scores were highly influenced by sample characteristics. The implications of these results are discussed as they relate to the use of the DAS in research.

The quality of romantic relationships can have a profound impact on the quality of life of the individuals comprising the romantic dyad. The presence of distress in a romantic relationship has been linked to increased risk for depression (Kurdek, 1998), anxiety (McLeod, 1994), and a wide variety of health problems (Prigerson, Maciejewski, & Rosenheck, 1999). Children of parents with high-quality relationships are more likely to have higher self-esteem (Amato, 1986) and to subsequently form high-quality romantic relationships (Amato & Sobolewski, 2001) than their peers with parents in distressed relationships. Given the importance of relationship quality across a wide variety of contexts, it has justly been the focus of a vast body of research. The quality and utility of this research depend partly on the reliable and accurate assessment of relationship quality. To this end, a wide variety of instruments have been developed to measure the quality of a romantic relationship, including the Locke-Wallace Marital Adjustment Test (Locke & Wallace, 1959), the Marital Satisfaction Inventory—Revised (Snyder, 1997), the Marital Satisfaction Scale (Roach, Frazier, & Bowden, 1981), the Kansas Marital Satisfaction Scale (Schumm, Scanlon, Crow, Green, & Buckler, 1983), and the Quality Marriage Index (Norton, 1983), to name but a few. This study focuses on one such measure, the Dyadic Adjustment Scale (DAS; Spanier, 1976), and examines the reliability of DAS scores across studies.

Department of Psychology, Western Washington University, 516 High Street, Bellingham, WA 98225-9089 (Jim.Graham@wwu.edu).

*Illinois School of Professional Psychology, Argosy University at Chicago, 350 N Orleans Street, Chicago, IL 60654.

Key Words: measurement, meta-analysis, quality, relationship adjustment, satisfaction.

THE DAS

The DAS is perhaps the most widely used measure of relationship quality in the social and behavioral sciences literature. Spanier (1985) noted that the DAS had been used in more than 1,000 studies,

within 10 years of its creation. Since that time, that number has continued to grow.

Scale Description

The DAS is a 32-item measure in a variety of response formats developed to measure dyadic adjustment. Spanier (1976) defined dyadic adjustment as "… a process, the outcome of which is determined by the degree of: (1) troublesome dyadic differences; (2) interpersonal tensions and personal anxiety; (3) dyadic satisfaction; (4) dyadic cohesion; and (5) consensus on matters of importance to dyadic functioning" (p. 17).

Spanier's (1976) development of the instrument was based on a review of over 15 preexisting measures of marital satisfaction and adjustment. An initial pool of approximately 300 items was subsequently paired down through expert ratings, the items' abilities to differentiate between married and divorced samples, and exploratory principal axis factoring. The 32 items that were retained to form the final DAS are summed to create a total score ranging from 0 to 151, with higher scores indicating more positive dyadic adjustment. Spanier (1976) also identified four subscales: Dyadic Consensus (13 items; the degree to which the couple agrees on matters of importance to the relationship), Dyadic Satisfaction (10 items; the degree to which the couple is satisfied with their relationship), Dyadic Cohesion (5 items; the degree of closeness and shared activities experienced by the couple), and Affective Expression (4 items; the degree of demonstrations of affection and sexual relationships). Spanier argued that the subscales could be used alone "without losing confidence in the reliability and validity of the measure" (1976, p. 22).

Since its inception, the DAS has sparked a vast amount of research examining the validity and factor structure of the instrument. Total DAS scores have been consistently shown to discriminate between distressed and nondistressed couples and have been shown to identify couples with a high likelihood of divorce (Crane, Busby, & Larson, 1991; Schumm, Paff-Bergen, Hatch, & Obiorah, 1986; Spanier, 1988; Spanier & Thompson, 1982). Typically, cutoff scores between 92 and 107 are used to differentiate between distressed and nondistressed couples (Sabourin, Valois, & Lussier, 2005). The DAS has been translated into multiple languages, including Chinese (Shek, 1994), French (Baillargeon, Dubois, & Marineau, 1986; Vandeleur, Fenton, Ferrero, & Preisig, 2003), Italian (Gentili, Contreras, Cassaniti, & D'Arista, 2002), Korean (Lee & Kim, 1996), and Turkish (Fisiloglu & Demir, 2000). Additionally, a number of "short forms" of the DAS have been developed, including versions with 4 (Sabourin et al., 2005), 6 (Hunsley, Pinsent, Lefebvre, James-Tanner, & Vito, 1995), 7 (Hunsley, Best, Lefebvre, & Vito, 2001; Hunsley et al., 1995; Sharpley & Rogers, 1984), 10 (Kurdek, 1992), and 14 items (Busby, Crane, Larson, & Christensen, 1995). Other researchers have examined the utility of using single items from the DAS as valid measures of relationship adjustment and satisfaction (Goodwin, 1992; Hunsley et al.).

Factor Structure

Ever since Spanier initially identified the DAS as being comprised four interrelated factors, the validity of the proposed factor structure of the DAS has been hotly contested. A number of confirmatory (Eddy, Heyman, & Weiss, 1991; Sabourin, Lussier, Laplante, & Wright, 1990; Spanier, 1989) and exploratory (Baillargeon et al., 1986; Spanier & Thompson, 1982) factor analyses have supported the factor structure of Spanier's (1976) original data. Although Crane et al.'s (1991) exploratory factor analysis provided support for the majority of Spanier's factors, it provided less support for the construct validity of the Dyadic Satisfaction subscale. Other exploratory factor analyses have supported the validity of the four DAS factors in men only (Antill & Cotton, 1982; Sabourin, Bouchard, Wright, Lussier, & Boucher, 1988).

The vast majority of other (exploratory) factor analyses point toward different three-factor models, typically consisting of two of the original subscales as unique factors and a third factor comprised items pulled from the remaining two original subscales (Antill & Cotton, 1982; Kazak, Jarmas, & Snitzer, 1988; Sabourin et al., 1988; Sharpley & Cross, 1982). In four of the five exploratory factor analyses conducted in the four studies supporting various three-factor models (the exception being the analysis of women's scores conducted by Antill & Cotton), the items comprising the Affective Expression subscale emerged as a unique factor. In those same four analyses, items from the Dyadic Satisfaction subscale were combined in a factor with items from another of Spanier's original subscales. In two analyses, items from the Dyadic Satisfaction subscale were combined with items from the Dyadic Consensus

subscale (women in Sabourin et al.; Sharpley & Cross), and in one analysis, the satisfaction items were combined with items from the Dyadic Cohesion subscale (women from Kazak et al., 1988). In the fourth analysis (men from Kazak et al.), the items from the Dyadic Satisfaction subscale were separated into two factors, one containing items from the Dyadic Consensus and the other containing items from the Dyadic Cohesion subscales.

The Dyadic Satisfaction subscale is of particular interest in an examination of the DAS, as it has been responsible for one of the chief critiques of the DAS: that the DAS confounds dyadic "satisfaction" (Bradbury, Fincham, & Beach, 2000; Eddy et al., 1991) or "happiness" (Norton, 1983) with the determinants of satisfaction/happiness. That is, oftentimes researchers are looking at how variables such as the level of conflict or the amount of time spent in conjoint activities influence satisfaction with one's relationships. If researchers chose the total DAS score as their dependent variable, they would likely find very strong relationships between these variables, not only because those variables actually are related to one another but also because the total DAS score measures not just satisfaction but also agreement (Dyadic Consensus) and shared activities (Dyadic Cohesion)! Critiques of the DAS using this rationale are actually critiques of common usage of the DAS, as Spanier did not originally intend to create a measure of dyadic satisfaction; rather, he attempted to create a measure of dyadic adjustment. Spanier (1976) even stated in his original article that researchers interested in a single component of dyadic adjustment may wish to use a single subscale from the DAS: "For example, researchers interested specifically in dyadic satisfaction may use the 10-item subscale for this purpose" (p. 22). In many cases, this is exactly what researchers have done. In other cases, they have erroneously used the total DAS score as a measure of dyadic satisfaction. In either case, the blame (or the credit) lies with the researcher, not with the instrument.

In a novel validity study, Kurdek (1992) examined the correlations between the DAS subscales and measures of love, liking, and marital satisfaction across heterosexual married and gay and lesbian cohabiting couples. Unequivocally, the Dyadic Satisfaction subscale accounted for a large portion of the variance of these other variables, and the hierarchical addition of the other three variables after the Dyadic Satisfaction subscale added very little to the prediction of the criterion variables. This provided support for the notion that the DAS is a multidimensional measure and that marital satisfaction is but one component of what is being measured by the DAS. This may also help explain the proliferation of three-factor models that tie items from the Dyadic Satisfaction subscale to items from other subscales. Satisfaction may be a separate component that has consistently strong relationships with the determinants of satisfaction: closeness and conflict.

It seems somewhat unfortunate that so much energy has been put forth in trying to replicate the results of Spanier's original exploratory factor analysis, rather than to directly compare competing models through confirmatory factor analysis. For example, Kurdek's (1992) results seem to suggest a factor structure in which the Dyadic Cohesion, Dyadic Consensus, and Affective Expression subscales are first-order factors measured directly by their own items, and the Dyadic Satisfaction subscale is a mixed first- and second-order factor that is measured directly by the items of the Dyadic Satisfaction subscale and indirectly by the other three latent factors. Such a model may account for the high degree of relatedness between the Dyadic Satisfaction and the other scores, as well as account for the critiques that satisfaction is separate from, but related to, the determinants of satisfaction.

Weighting Issues

The DAS has also been critiqued because, though it measures several different latent constructs, those latent constructs are given unequal weighting in the total DAS score (Norton, 1983). For example, the 4 items comprising the Affective Expression subscale have much less influence on the total DAS score than the items comprising the 13 items of the Dyadic Consensus subscale. Furthermore, because of the different response formats used for the items comprising the DAS, some items account for a proportionately larger amount of the variance in total DAS scores than other items. Items 29 and 30, which are dichotomously scored, will produce less variance than Item 31, which is measured on a 7-point Likert-like scale. The use of different item formats and unequal weighting of subscales has been pointed out as making it possible for two couples to have quantitatively identical scores on the DAS but to be qualitatively very different from one another (Norton). Some researchers have attempted to avoid this issue by converting items to *z*-scores before

summing them to create total DAS scores (Fitzpatrick, Salgado, Suvak, King, & King, 2004). The use of different response formats is also of concern in determining score reliability, as will be discussed in greater detail later.

Reliability

Within the tenets of classical test theory, reliability (ρ_{xx}) is characterized as the proportion of observed score (X) variance accounted for by the true score (T) variance, or $\rho_{xx} = Var_T/Var_X$. Although part of the variance in a given set of observed scores is the result of variance in the underlying trait being measured, other influences such as sampling error, measurement error, and model specification error may also account for part of the observed score variance. Reliability estimates, therefore, attempt to characterize the proportion of variance in observed scores that is the result of nonerror factors. Specifically, measures of internal consistency such as Cronbach's coefficient alpha, split-half reliabilities, and KR-20 attempt to characterize the amount of the variance of a test score that is not accounted for by measurement error. The present discussion is limited to internal consistency, one source of measurement error, and excludes consideration of other types of reliability, such as test-retest and interrater.

It should be noted that reliability is a property of test scores and not measurement instruments themselves; that is, to say that tests are not reliable and data are reliable (Gronlund & Linn, 1990; Thompson & Vacha-Haase, 2000). It is perfectly reasonable that a test could result in highly reliable scores when used with one sample and highly unreliable scores when used with another sample. As stated by Wilkinson and the APA Task Force on Statistical Inference (1999), "[i]t is important to remember that a test is not reliable or unreliable. Reliability is a property of test scores on a test for a particular population of examinees" (p. 596).

Because reliability is an artifact of test scores and not test instruments, it is important for researchers to report the reliability of the scores that are the focus of analysis. Reliability estimates are extremely important in any research, as they form the upper limit for any effect size using those scores. Specifically, the upper value of any r^2 statistic between two test scores is the product of the reliabilities of the two measures. Wilkinson and the APA Task Force on Statistical Inference (1999) comment on this issue by stating that "[i]nterpreting the size of observed effects requires an assessment of the reliability of the scores" (p. 596) and that "… authors should provide reliability coefficients of the scores for the data being analyzed even when the focus of the research is not psychometric" (p. 596). The reliability of test scores is especially important to take into account when one considers that inferences drawn from results may be influenced by the reliability of the scores. For example, a study might show that a specific predictor variable is related to the relationship quality of women but not to the relationship quality of men. This finding may be the result of actual differences between men and women. Conversely, it might be that the smaller correlation seen in men is a result of relationship quality being measured less reliably in men than in women, serving to cap the maximum possible effect size of men at a lower level than women.

Nonetheless, the vast majority of published literature fails to report reliability estimates for the data in hand (Vacha-Haase, Ness, Nilsson, & Reetz, 1999). Because of misunderstandings regarding the meaning of score reliability, many individuals erroneously assume that the reliability estimates reported for the scores used in the development of a measure apply equally to their own data, even when their sample differs widely from the original development sample. One must invoke inductive reasoning to apply information from a very specific situation (the reliability of test scores for a specific sample) to a more general conclusion (any scores produced by this test will be reliable); therefore, this phenomenon has been labeled reliability induction (Vacha-Haase, Kogan, & Thompson, 2000). Reliability induction occurs when researchers assume that their data are reliable because the measure that they used produced reliable scores in a previous sample.

Inductive reasoning relies in part on the number and accuracy of the previous observations. It seems logical that one would not want to base an inductive argument on a single observation (such as the reliability of test scores reported when a measure is developed). Furthermore, inductive reasoning requires that the previously and currently observed events be highly similar to one another. Although reliability induction might be a safe practice in some instances, it would certainly not seem to be so if the characteristics of the samples used in the previous and present observations are widely different from one another.

Additionally, it is often the case that the reliabilities reported when a measure is initially developed

are based on the scores that were used to determine the factors, or subscales, of the measure. Because most exploratory factor analytic techniques result in a factor structure that best fits the data, these techniques tend to capitalize sampling error (Gorsuch, 1983). The resultant factor structures are therefore never typically as strong in other samples as they are in the initial sample. This capitalization on sampling error also affects reliability estimates. Because the factor (or true score) variance is maximized in the development sample, internal consistency estimates of those factors will likely be higher than they will be with subsequent samples. Put another way, it is highly likely that the reliability coefficients reported from the sample used to create the factor structure represent the upper limit of score reliability. In these cases, reliability induction will tend to overestimate the actual score reliability of the data in hand.

Reliability Generalization

Because reliability is a characteristic of scores, because the reliability of scores serves as the upper limit to measures of size of effect for any statistical procedure using those scores, and because the role of relationship quality has been shown to be extremely important in a variety of salient contexts, it is important to be cognizant of the reliability of test scores used in relationship research. This study utilizes the reliability generalization framework described by Vacha-Haase (1998) to examine the reliability of DAS scores across published studies. Reliability generalization is a meta-analytic technique that can characterize the reliability of test scores for a given measure across a wide variety of applications. Additionally, reliability generalization procedures allow one to examine the variance in score reliability estimates across studies. If it were true that tests (not scores) were reliable, then it would be expected that the reliability estimates of scores for a given test would be the same across different samples and administrations. Any variance in reliability estimates across administrations and samples could not be the result of the test itself (as the test format and content is a constant); rather, the variance would be the result of characteristics unique to the sample and study conditions. As such, reliability generalization procedures allow one to determine what amount of the variance in reliability estimates can be accounted for by specific characteristics of the sample and data. Such information can be of great use in determining the situations under which a given test is likely to produce scores with high degrees of reliability and the situations under which they are not.

Given the impact of the quality of relationships on a wide range of aspects of individuals' lives, it is of the utmost importance that relationship quality be measured reliably. Though not without its critics, the DAS continues to be a widely used measure of relationship quality. This study seeks to use a reliability generalization meta-analytic framework to (a) characterize the reported internal consistency of DAS scores across studies and (b) determine the influence of sample characteristics on DAS score reliability.

METHOD

Sample

A sample of 554 studies was initially identified through *PsycINFO* using *Dyadic Adjustment Scale* as a keyword search term. Although a much larger number of studies in the existing literature have used the DAS as a variable, the chosen search strategy had the benefit of producing a large number of representative studies and minimizing the large number of false hits that would be anticipated using broader search terms such as *marital quality* or *marital satisfaction*. Of the 554 initial studies identified, 61 (11%) were discarded because they were published in a language other than English, and 8 (1.4%) were unable to be obtained through normal interlibrary loan channels. Each article was then reviewed for suitability. Thirty-nine (7%) studies were discarded as false hits because they did not involve original data collection of DAS scores, and 43 (7.8%) articles were discounted because they utilized a modified or incomplete version of the DAS.

A final sample of 403 articles was determined to involve original data collection in which either the DAS or one or more of its original four subscales were used in entirety. Of these articles, 34.0% ($n = 137$) made no mention of the reliability of the DAS or the reliability of DAS scores. Eighteen percent ($n = 72$) of these articles used reliability induction by stating that the DAS had been shown to be a reliable instrument but did not report any reliability estimates. Twenty-four percent ($n = 97$) of these articles also used reliability induction by stating that the DAS had been shown to be a reliable instrument and by reporting reliability estimates found in other studies.

The reported reliability values were most often cited from Spanier's (1976) original study. An additional 1.5% ($n = 6$) of articles reported reliability estimates for the data in hand in a format that made it unusable for this study (e.g., reported a range of reliabilities, stated all reliabilities were greater than .70). The remaining 22.6% ($n = 91$) of articles reported internal consistency reliability estimates of their data for either the total DAS scores or one or more DAS subscale scores. Because reliability estimates are often reported for multiple groups in a given study, a total of 128 samples with one or more coefficient alpha reliability estimates from 91 studies were retained for the final analyses. The final 128 samples represented 25,035 participants.

Coding of Study Characteristics

Each of the articles reporting original reliability coefficients was examined in detail, and the following study and sample characteristics were recorded: (a) sample size—the number of participants in the sample; (b) percent White—the percentage of the total sample that was identified as White, Caucasian, or nonethnic European; (c) percent heterosexual—the percentage of the total sample that was identified as being in a romantic relationship with a partner of the opposite gender; (d) percent men—the percentage of the total sample identified as men; (e) percent married—the percentage of the total sample identified as being currently married.

Initially, a number of other study characteristics, such as mean relationship length, mean age, mean income, and percentage of sample recruited from a clinic, were also coded; because of inconsistency in reporting of sample and study characteristics, however, the inclusion of these variables in the final sample of articles would have resulted in an insufficient sample size for the analyses chosen. As a result, these characteristics were excluded from the analysis. In all instances, the sample characteristics were coded at the same level as the reported reliability estimates. For example, if a study reported separate reliability coefficients for husbands and wives, separate demographic information was coded for husbands and wives. If a study reported a single reliability estimate, demographic information was coded for the entire sample. In the instances where reliability estimates were reported for the total sample and demographic characteristics were coded for separate groups, aggregate demographic characteristics for the total sample were computed. Some information, such as whether the couples were same- or opposite-gender pairings were infrequently reported explicitly but could be deduced from other study characteristics. For example, a sample of married couples in which both spouses complete measures and there is an equal number of men and women is likely a heterosexual sample.

Results

Reliability generalization procedures allow reliability estimates to be compared across studies. In addition to allowing one to characterize a measure's internal consistency across studies, it also allows potential sources of variability in score reliability to be examined across studies. Because coefficient alpha is a variance-accounted-for statistic, it is equivalent to a squared correlation (Thompson & Vacha-Haase, 2000). The square roots of the reliability coefficients were therefore treated as r-equivalent statistics or correlations. Because correlations are not normally distributed and have problematic standard errors, it is necessary to apply a transformation before they can be used in an analysis (Lipsey & Wilson, 2001). As is standard in meta-analytic procedure, Fisher's r to Z transformation, $z_r = (.5)\,[\log_e (1 + r) - \log_e (1 - r)]$, was applied to the square root of all reliability estimates. Unless otherwise noted, all analyses were conducted using these transformed reliability estimates and by weighting each study by a function of the standard error of the statistic ($n - 3$) to take into account the effects of different sample sizes across studies. When appropriate, the resultant statistics are transformed back into the metric of coefficient alpha for ease of interpretation. All reliability coefficients are reported in the metric of coefficient alpha.

Descriptive Statistics

The descriptive statistics for the DAS total and subscale score reliability estimates are shown in Table 1. As shown in Table 1, reliability of total DAS scores ranged from .58 to .96, with a mean score reliability of .915. A 95% confidence interval was constructed about the mean and found to range from .906 to .922. In the social and behavioral sciences, .70 is often considered the minimum value for a measure to be considered to have an acceptable degree of internal consistency (Nunnally, 1978). The average reliability estimate and the lower bound of the confidence interval

Table 1. *Reliability of Dyadic Adjustment Scores Across Studies*

Scale	*N*	Spanier α	Mean α	95% CI Lower	95% CI Upper	Minimum	Maximum	*Q*	*p*
Total	98	.96	.915	.906	.922	.58	.96	1077.77	<.001
Dyadic Consensus	39	.90	.872	.855	.886	.73	.93	288.47	<.001
Dyadic Satisfaction	38	.94	.848	.823	.869	.70	.96	547.25	<.001
Dyadic Cohesion	35	.86	.789	.755	.820	.58	.89	478.31	<.001
Affective Expression	34	.73	.714	.681	.746	.50	.80	251.34	<.001

Note: Spanier = α values reported by Spanier (1976); C I. = confidence interval. Estimates use a random-effects model.

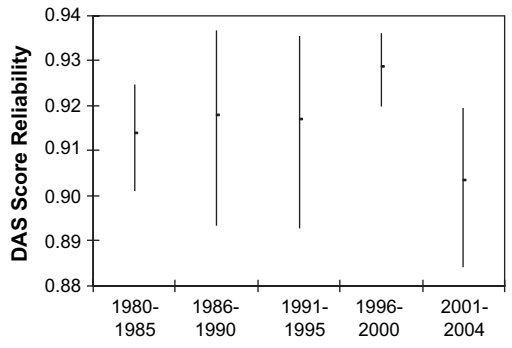
suggest that DAS scores have an acceptable level of reliability across studies, though the mean reliability was statistically significantly lower than the value of .96 reported by Spanier (1976). Similarly, the means and confidence intervals of the reliability estimates for the Dyadic Consensus, Dyadic Satisfaction, and Dyadic Cohesion subscales, all fell within an acceptable range, though each was statistically significantly lower than the reliability of subscale scores reported by Spanier (1976). Although the mean reliability of Affective Expression subscale scores was greater than .70, the 95% confidence interval constructed about that mean did extend to as low as .681. This suggests that the Affective Expression subscale scores do not possess adequate internal consistency as measured by Cronbach's alpha across studies. Spanier's (1976) estimate of Affective Expression subscale score reliability did fall within the 95% confidence interval.

The present meta-analysis does not include all published and unpublished studies that have used the DAS. As previously described, reliability coefficients were not reported in many studies, and a large number of published studies were likely not obtained because of the search strategy. Additionally, it is likely that there are a number of unpublished studies that used the DAS; it is possible that many of these were not published because of statistically insignificant findings, potentially in part as a result of low reliability coefficients. Furthermore, some published works may have initially examined the score reliabilities but failed to report them because they were spuriously low. As such, it is necessary to examine how including additional studies with low reliabilities would influence the present results. The robustness of each of the mean reliability estimates was tested using Orwin's (1983) fail-safe *n* for effect sizes. A reliability of .69 was chosen as an unacceptable level. The results of this analysis indicated that it would take 30 studies with an average total DAS score reliability of .50 to bring the mean reliability below the level of acceptability. It would similarly take 10, 8, 4, and 2 studies with subscale score reliabilities of .50 to bring the Dyadic Consensus, Dyadic Satisfaction, Dyadic Cohesion, and Affective Expression mean reliabilities below .70, respectively. Therefore, it seems likely that the mean total DAS reliability estimates reported in this study will be likely to remain robust, save for the Affective Expression subscale.

Change Over Time

To examine how reported reliabilities of the DAS have changed over time, 95% confidence intervals were constructed about the mean DAS total score reliabilities of studies published in 5-year intervals. These confidence intervals are shown in Figure 1. As shown in Figure 1, reliability rates are fairly consistent from time period to time period, with the exception of consistently high reliabilities reported in work published from 1996

FIGURE 1. DYADIC ADJUSTMENT SCALE SCORE RELIABILITIES AND 95% CONFIDENCE INTERVALS BY YEAR



to 2000. The overall rate of reporting reliability for data in hand of 22.6% across years is fairly typical of those found in reliability generalization studies. The number of relevant studies published in any given 5-year span and the percentage of studies reporting reliability estimates for data in hand are shown in Table 2. As shown in Table 2, reporting rates have been increasing slowly over time, likely in response to changes in journal policies and seminal writings (e.g., Wilkinson & the APA Task Force on Statistical Inference, 1999).

Reliability Heterogeneity

The degree of dispersion of reliability estimates about the mean reliability was tested for the total DAS and for each of the subscales using Q tests of homogeneity. The Q test uses the χ^2 distribution and tests whether it is reasonable to assume that the estimates from different studies are all estimating the same population reliability coefficient (Hedges & Olkin, 1985). As shown in Table 1, the results of each of these tests were statistically significant at the $p < .001$ level, suggesting heterogeneous variances. As such, it is assumed that the studies do not all provide reliability estimates from the same population; rather, it is likely that several different populations are being represented in the present data.

Meta-analytic procedures typically employ either fixed- or random-effects models, each of which makes different assumptions about the effect sizes (or in the present case, reliability coefficients) being examined. Fixed-effects models assume that the reliability coefficients of the populations used in different studies are homogenous, or the same. Conversely, random-effects models assume that population reliability coefficients vary randomly from study to study (Hedges & Vevea, 1998). Put another way, fixed-effects models assume that the entire universe of studies is included in the meta-analysis, and random-effects models assume that only a sample of the universe of studies is included in the meta-analysis (Hunter & Schmidt, 2000). Consequently, the results of random-effects models tend to be more easily generalized to studies not included in the meta-analysis than the results of fixed-effects models. Additionally, fixed-effects models tend to produce results with inflated Type I errors when the population effect size variances are heterogeneous (Field, 2003a). Therefore, because the sampling method used in this study did not include all published studies using the DAS, because the Q tests suggested heterogeneous variance, and because random-effects models have been noted as being a more realistic representation of the real world (Field, 2003b), random-effects models were used in all the following analyses.

Table 2. *Rates of Reporting Reliability Estimates for Data in Hand*

Time Period	Studies, *n*	Reporting, *n* (%)
1976 – 1980	7	1 (14.3)
1981 – 1985	72	12 (16.7)
1986 – 1990	77	12 (15.6)
1991 – 1995	80	17 (21.3)
1996 – 2000	68	15 (22.1)
2001 – 2004	99	34 (34.3)
Total	403	91 (22.6)

Determinants of Reliability

In an attempt to explain some of the variance in reliability estimates across studies, the transformed reliability estimates were regressed on the percentage of the study sample that included White, heterosexual, men, and married individuals. For the reasons described above, the regression analyses used the random-effects weighted maximum likelihood regression analysis described by Lipsey and Wilson (2001). Separate regression analyses were run for each of the subscale score reliabilities and the total score reliability. For each of the analyses, the use of a random-effects model resulted in statistically nonsignificant Q tests of the residual variance, indicating that the variability in the reliability coefficients between studies was statistically nonsignificant once the predictor variables were taken into account. Despite statistical nonsignificance, the residual Q statistics were often larger than those for the fixed components of the model; as such, it was determined that a random-effect model best represents the data. In addition to examining beta-weights in interpreting the regression results, structure coefficients were also calculated. Structure coefficients, or the correlations between the predictor variables and the predicted dependent variable, were used to assess for the presence of suppressor effects and high degrees of shared variance (Courville & Thompson, 2001).

The results of the regression analyses are shown in Table 3. As shown in Table 3, the regression

Table 3. *Random-Effects Weighted Maximum Likelihood Regression of Score Reliability on Sample Characteristics*

		Model			Residual	% White		% Heterosexual		% Men		% Married	
Subscale	*n*	R^2	*Q*	*p*	*Q*	β	r_s	β	r_s	β	r_s	β	r_s
Total	36	.060	2.25	.69	35.16	−.048	−.023	.111	.742	.076	.008	.174	.907
Dyadic Consensus	19	.148	3.13	.54	17.94	−.180	−.475	.229	.576	−.197	−.801	−.057	.219
Dyadic Satisfaction	20	.040	0.82	.94	19.67	.166	.723	.058	.061	−.077	−.285	−.154	−.455
Dyadic Cohesion	18	.255	5.97	.20	17.46	.268	.422	.239	.369	−.399	−.773	−.323	−.172
Affective Expression	17	.424	12.25	.02	16.61	−.490	−.747	.383	.191	−.242	−.537	−.334	−.227

Note: Dependent variable = score reliability; $r_s = r_{x\hat{y}}$ = structure coefficient.

model using total score reliability as the criterion failed to achieve statistical significance, with a low overall model effect size ($R^2 = .06$). This suggests that the variability in DAS score reliability across studies is not explained by marital status, sexual orientation, gender, or ethnicity. Total DAS scores are likely to be equally reliable across diverse samples.

The regression analysis using Dyadic Satisfaction subscale score reliabilities as the criterion was similarly statistically nonsignificant with a low effect size. Although the analyses predicting Dyadic Consensus and Dyadic Cohesion subscale reliabilities were not statistically significant, both had moderate overall effect sizes ($R^2 = .148$ and $R^2 = .255$, respectively). Despite the lack of statistically significant overall models, higher Dyadic Consensus and Cohesion subscale score reliabilities were most strongly associated with higher proportions of women in the study samples.

Finally, the total model using Affective Expression subscale score reliabilities as the criterion was statistically significant with a very large effect size ($R^2 = .424$). An examination of the beta-weights and structure coefficients suggests that lower Affective Expression subscale score reliabilities were most strongly associated with samples that included higher proportions of Whites. Lower Affective Expression subscale score reliabilities were less strongly associated with samples that included more men, fewer heterosexuals, and more married individuals. The reliability of Affective Expression subscale scores appears to be largely dependent on sample characteristics.

Discussion

This study examined the reliability of DAS total and subscale scores across a sample of published studies.

DAS Total

The reliability of DAS total scores was strong, with a mean score reliability of .915 (range .58 – .96) across studies. Although this number was lower than that reported by Spanier (1976) in the development of the measure, it is, nonetheless, acceptable in psychometric terms. In fact, Spanier's reported total score reliability of .96 represented the highest reliability estimate found across all the studies sampled. As previously discussed, this is likely an artifact of the fact that the reliability estimates presented by Spanier (1976) were based on the same scores that were used to determine the factor structure of the DAS. As such, Spanier's original reliability estimates were likely somewhat inflated because of a capitalization on sampling error.

The reliability of DAS scores did not give evidence of being substantially influenced by the marital status, ethnicity, sexual orientation, or gender of the sample. Spanier initially set out to design a global measure of relationship adjustment that could be used across a wide variety of cohabiting couples. In this study, the DAS gives evidence of producing internally consistent scores across a wide variety of romantic relationships. Although other study or sample characteristics may account for variance in total DAS score reliability, this study suggests that the score reliability is generalizable across diverse samples.

Dyadic Satisfaction

The Dyadic Satisfaction subscale was shown to produce scores with acceptable internal consistency, with a mean alpha of .848. The variance in score reliabilities across studies was not explained by sample characteristics, suggesting that the Dyadic Satisfaction subscale score reliabilities are stable across diverse samples. Although

acceptable, the mean reliability of the Dyadic Satisfaction scores was much lower across published studies than that reported by Spanier (1976).

Dyadic Consensus

The reliability of DAS Consensus scores was also acceptable across studies, with a mean score reliability of .872, the highest of all the subscale scores. The model used to predict score reliability with sample characteristics did not emerge as statistically significant, though it did produce a moderate effect size. This suggests that Dyadic Consensus subscale score reliabilities were relatively stable across diverse samples. The results also indicated that studies using reliability induction by referring to Spanier's reliability estimates are likely to overestimate their scores' own reliability.

Dyadic Cohesion

The Dyadic Cohesion subscale of the DAS was found to produce scores with acceptable, although somewhat low, internal consistency, with a mean alpha of .789 across studies. The variance in score reliabilities across studies was not explained by sample characteristics despite a moderate effect size, suggesting that the Dyadic Cohesion subscale score reliabilities are stable across diverse samples. The mean reliability estimate was lower than was found in Spanier's (1976) original data.

The regression for both the Dyadic Consensus and Dyadic Cohesion subscales yielded statistically nonsignificant results; both yielded moderate effect sizes, however. Given the relatively low number of studies reporting reliability coefficients and sample characteristics for these studies, the possibility of a Type II error resulting from low power cannot be ruled out. Further investigation into the equivalence of these scales across genders appears warranted. Based solely on current evidence, the reliabilities of Dyadic Consensus and Cohesion subscales appear stable across gender.

Affective Expression

Although the mean Affective Expression score reliability of .714 was above the generally accepted cutoff of .70, a 95% confidence interval constructed about the average score reliability did include .70. This suggests that the reliabilities of the Affective Expression subscale scores demonstrate borderline acceptable internal consistency that was comparable with the reliability estimate reported in Spanier's (1976) study. Furthermore, sample characteristics did predict score reliability, with studies including lower proportions of Whites, men, and married individuals and higher proportions of heterosexual relationships resulting in higher score reliability estimates. The overall model was statistically significant, with a large effect size. In general, the Affective Expression subscale produces scores with borderline internal consistency that is more likely to be unacceptable in studies including large numbers of White participants.

The lower reliability of the Affective Expression scores across studies can likely be attributed to a combination of three factors. First, the Affective Expression subscale consists of only four items. Because reliability is partly a function of test length, with greater test length being equated with greater internal consistency, it is not surprising that the shortest subscale of the DAS also produces scores with the least reliability.

Second, it is highly likely that the Affective Expression subscale violates the assumption of tau-equivalence required by Cronbach's alpha. Cronbach's alpha assumes that all items measure the same latent construct on the same scale, with the same degree of accuracy and with possibly different errors (Miller, 1995; Raykov, 1997). Because the Affective Expression subscale consisted of items that are measured on two different scales (two dichotomous items and two items scored on 6-point scales), it is likely that Cronbach's alpha is an underestimate of the scores' true reliability. An examination of congeneric measures of reliability used Affective Expression subscale scores from actual data to test the violation of the tau-equivalence assumption (Graham, in press). The Affective Expression scores used in this examination were determined to have a Cronbach's alpha of .72; the data gave evidence of violating the tau-equivalence assumption, however, and the subscale was determined to have a congeneric reliability of .83.

Finally, it is possible that the construct underlying what Spanier called Affective Expression takes on different meanings in samples with different characteristics. This notion is particularly interesting when one notes that two of the items comprising the Affective Expression subscale refer explicitly to sex and the other two could be interpreted as not referring to sex, rather to affection and love. It is possible that non-White, nonmarried, heterosexual couples, and women tend

to respond more consistently to the items of the Affective Expression subscale. Perhaps, the items referring to love and affection are interpreted as sexual (or vice versa). Alternatively, affection and sex may be viewed as highly similar constructs in these groups. These two constructs might be viewed as more distinct in men and married, homosexual, and White couples. The present data do not provide a direct answer regarding why the reliability of Affective Expression subscale scores differs greatly from group to group but does point out some interesting possibilities for future investigation.

Conclusion

An examination of a select sample of published studies suggests that the DAS consistently produces scores with acceptable internal consistency, although lower than what Spanier reported in the development of the measure. Furthermore, the variation in reliability estimates is not explained by sample characteristics, suggesting that the DAS meets its goal of producing a measure that can be used to assess a wide range of romantic relationships (Spanier, 1976).

The Dyadic Consensus, Dyadic Satisfaction, and Dyadic Cohesion subscales also show evidence of producing scores of acceptable reliability across studies. The Affective Expression subscale, however, gives evidence of producing borderline unacceptable internal consistency according to Cronbach's alpha. It is likely that Cronbach's alpha provides an underestimate of the Affective Expression subscale's actual reliability, as the subscale appears to give evidence of violating the assumption of tau-equivalence (Graham, in press). Although the true reliability of the Affective Expression subscale is likely higher than is suggested here, this study gives the evidence that the reliability is highly dependent on sample characteristics. Thus, although Spanier (1976) suggested that researchers interested in specific components of dyadic adjustment might choose any of the subscale scores as variables of interest, the evidence here suggests otherwise. Researchers are strongly cautioned that the Affective Expression subscale not be used in isolation, lest researchers subject their analyses to potentially possible bias and lower levels of reliability when used with samples consisting predominately of Whites, married couples, same-gender couples, and men. The evidence suggests that the other subscales may be used in isolation with less concern.

Although global relationship quality may often be a variable of interest, researchers are strongly cautioned to consider the aims of their study before selecting the DAS as a measure. Because the DAS measures relationship *quality* and not just relationship *satisfaction*, it is easy to find spuriously high relationships between variables that measure similar constructs. Researchers who are interested in examining relationship satisfaction unconfounded by some of its determinants are encouraged to use either the Dyadic Satisfaction subscale of the DAS, or one of the many, specific measures of relationship quality (Norton, 1983; Roach et al., 1981; Schumm et al., 1983). Although measures in standard use such as the DAS are often attractive to researchers given the relative wealth of information about the measure and the ease with which the measure is accepted and understood by others, careful consideration is warranted. It is the individual researcher's place to determine whether the problems with the DAS preclude its use and whether other, newer measures with more sound psychometric properties might be better suited to the task at hand.

Additionally, it is hoped that this study underscores the importance of always reporting reliability estimates for data in hand. Although individuals often mistake internal consistency as a property of a test, it is most certainly not; reliability is a function of scores that is partly determined by the test but also determined by characteristics of the sample and the way in which the data were collected. If the findings of this meta-analysis are applied to the studies that were examined and did not report reliability estimates for data in hand, then the over 75% of the studies examined that either explicitly or by omission induced the reliability of their data from that of Spanier's (1976) greatly overestimated the reliability of their data.

Finally, it should be noted here that this study concerns itself solely with measures of internal consistency; other types of reliability, such as test-retest reliability, are not addressed. Different sources of measurement error are quite distinct from one another and often behave differently in the same measure. Thus, one could not expect the present results to pertain to the consistency of DAS scores over time. Similarly, this study does not examine the validity of DAS scores across diverse studies or samples. Although one could expect validity coefficients to be lower in populations with low internal consistency, validity is not directly assessed here.

Nearly 30 years after its inception, the DAS continues to be a widely popular and oft-used measure in relationship research. The measurement of relationship quality is important because the correlates and consequences of relationship quality are so widespread and encompassing. Although it is hoped that the field will continue to produce and refine methods of measurement, the DAS remains a viable force.

NOTE

The authors would like to thank Fay Kallista for her invaluable assistance in obtaining many of the articles used in this study.

REFERENCES

References marked with an asterisk indicate studies included in the meta-analysis.

*Adam, D., & Gingras, M. (1982). Short- and long-term effects of a marital enrichment program upon couple functioning. *Journal of Sex and Marital Therapy*, *8*, 97 – 118.

Amato, P. R. (1986). Marital conflict, the parent-child relationship and child self-esteem. *Family Relations*, *35*, 403 – 410.

Amato, P. R., & Sobolewski, J. M. (2001). The effects of divorce and marital discord on adult children's psychological well-being. *American Sociological Review*, *66*, 900 – 921.

*Anderson, S. A. (1986). Cohesion, adaptability and communication: A test of an Olson Circumplex Model hypothesis. *Family Relations*, *35*, 289 – 293.

*Antill, J. K. (1983). Sex role complementarity versus similarity in married couples. *Journal of Personality and Social Psychology*, *45*, 145 – 155.

*Antill, J. K., & Cotton, S. (1982). Spanier's Dyadic Adjustment Scale: Some confirmatory analyses. *Australian Psychologist*, *17*, 181 – 189.

*Antill, J. K., Cotton, S., & Tindale, S. (1983). Egalitarian or traditional: Correlates of the perception of an ideal marriage. *Australian Journal of Psychology*, *35*, 245 – 257.

*Aron, A., & Henkemeyer, L. (1995). Marital satisfaction and passionate love. *Journal of Social and Personal Relationships*, *12*, 139 – 146.

*Askildsen, E. C., Watten, R. G., & Faleide, A. O. (1993). Are parents of asthmatic children different from other parents? Some follow-up results from the Norwegian PRAD project. *Psychotherapy and Psychosomatics*, *60*, 91 – 99.

*Babcock, J. C., Green, C. E., Webb, S. A., & Graham, K. H. (2004). A second failure to replicate the Gottman et al. (1995) typology of men who abuse intimate partners and possible reasons why. *Journal of Family Psychology*, *18*, 396 – 400.

*Badger, T. A. (1992). Coping, life-style changes, health perceptions, and marital adjustment in middle-aged women and men with cardiovascular disease and their spouses. *Health Care for Women International*, *13*, 43 – 55.

*Badr, H. (2004). Coping in marital dyads: A contextual perspective on the role of gender and health. *Personal Relationships*, *11*, 197 – 211.

Baillargeon, J., Dubois, G., & Marineau, R. (1986). Traduction française de l'Échelle d'ajustement dyadique. *Canadian Journal of Behavioural Science*, *18*, 25 – 34.

*Basolo-Kunzer, M., Diamond, S., Maliszewski, M., & Weyermann, L. (1991). Chronic headache patients' marital and family adjustment. *Issues in Mental Health Nursing*, *12*, 133 – 148.

*Benzies, K. M., Harrison, M. J., & Magill-Evans, J. (1998). Impact of marital quality and parent-infant interaction on preschool behavior problems. *Public Health Nursing*, *15*, 35 – 43.

*Besser, A., & Priel, B. (2003). A multisource approach to self-critical vulnerability to depression: The moderating role of attachment. *Journal of Personality*, *71*, 515 – 555.

*Bigras, M., LaFreniere, P. J., & Dumas, J. E. (1996). Discriminant validity of the parent and child scales of the parenting stress index. *Early Education and Development*, *7*, 167 – 178.

*Bogat, G. A., Levendosky, A. A., Theran, S., von Eye, A., & Davidson, W. S. (2003). Predicting the psychosocial effects of interpersonal partner violence (IPV): How much does a woman's history of IPV matter. *Journal of Interpersonal Violence*, *18*, 1271 – 1291.

*Bouchard, G., Lussier, Y., & Sabourin, S. (1999). Personality and marital adjustment: Utility of the five-factor model of personality. *Journal of Marriage and the Family*, *61*, 651 – 660.

*Bouchard, G., Sabourin, S., Lussier, Y., Wright, J., & Richer, C. (1998). Predictive validity of coping strategies on marital satisfaction: Cross-sectional and longitudinal evidence. *Journal of Family Psychology*, *12*, 112 – 131.

Bradbury, T. N., Fincham, F. D., & Beach, S. R. H. (2000). Research on the nature and determinants of marital satisfaction: A decade in review. *Journal of Marriage and the Family*, *62*, 964 – 980.

Busby, D. M., Crane, D. R., Larson, J. H., & Christensen, C. (1995). A revision of the Dyadic Adjustment Scale for use with distressed and nondistressed couples: Construct hierarchy and

multidimensional scales. *Journal of Marital and Family Therapy*, *21*, 289 – 308.

*Cano, A., O'Leary, K. D., & Heinz, W. (2004). Short-term consequences of severe marital stressors. *Journal of Social and Personal Relationships*, *21*, 419 – 430.

*Cano, A., & Vivian, D. (2003). Are life stressors associated with marital violence? *Journal of Family Psychology*, *17*, 302 – 314.

*Carey, M. P., Spector, I. P., Lantinga, L. J., & Krauss, D. J. (1993). Reliability of the Dyadic Adjustment Scale. *Psychological Assessment*, *5*, 238 – 240.

*Cook, J. M., Riggs, D. S., Thompson, R., Coyne, J. C., & Sheikh, J. I. (2004). Posttraumatic stress disorder and current relationship functioning among World War II ex-prisoners of war. *Journal of Family Psychology*, *18*, 36 – 45.

Courville, T., & Thompson, B. (2001). Use of structure coefficients in published multiple regression articles: Beta is not enough. *Educational and Psychological Measurement*, *61*, 229 – 248.

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.

*deTurck, M. A., & Miller, G. R. (1986). The effects of husbands' and wives' social cognition on their marital adjustment, conjugal power, and self-esteem. *Journal of Marriage and the Family*, *48*, 715 – 724.

*Deutsch, F. M., Lozy, J. L., & Saxon, S. (1993). Taking credit: Couples' reports of contributions to child care. *Journal of Family Issues*, *14*, 421 – 437.

Eddy, J. M., Heyman, R. E., & Weiss, R. L. (1991). An empirical evaluation of the Dyadic Adjustment Scale: Exploring the differences between marital "satisfaction" and "adjustment." *Behavioral Assessment*, *13*, 199 – 220.

*Elek, S. M., Hudson, D. B., & Bouffard, C. (2003). Marital and parenting satisfaction and infant care self-efficacy during the transition to parenthood: The effect of infant sex. *Issues in Comprehensive Pediatric Nursing*, *26*, 45 – 57.

*Espina, A., de Alda, I. O., & Ortego, A. (2003). Dyadic adjustment in parents of daughters with an eating disorder. *European Eating Disorders Review*, *11*, 349 – 362.

Field, A. P. (2003a). Can meta-analysis be trusted? *Psychologist*, *16*, 642 – 645.

Field, A. P. (2003b). The problems in using fixed-effects models of meta-analysis on real-world data. *Understanding Statistics*, *2*, 77 – 96.

*Fisiloglu, H., & Demir, A. (2000). Applicability of the Dyadic Adjustment Scale for measurement of marital quality with Turkish couples. *European Journal of Psychological Assessment*, *16*, 214 – 218.

*Fitzpatrick, M. K., Salgado, D. M., Suvak, M. K., King, L. A., & King, D. W. (2004). Associations of gender and gender-role ideology with behavioral and attitudinal features of intimate partner aggression. *Psychology of Men and Masculinity*, *5*, 91 – 102.

*Friedman, D., Holmbeck, G. N., Jandasek, B., Zukerman, J., & Abad, M. (2004). Parent functioning in families of preadolescents with Spina Bifida: Longitudinal implications for child adjustment. *Journal of Family Psychology*, *18*, 609 – 619.

*Gartstein, M. A., & Fagot, B. I. (2003). Parental depression, parenting and family adjustment, and child effortful control: Explaining externalizing behaviors for preschool children. *Journal of Applied Developmental Psychology*, *24*, 143 – 177.

*Gattis, K. S., Berns, S., Simpson, L. E., & Christensen, A. (2004). Birds of a feather or strange birds? Ties among personality dimensions, similarity, and marital quality. *Journal of Family Psychology*, *18*, 564 – 574.

Gentili, P., Contreras, L., Cassaniti, M., & D'Arista, F. (2002). La Dyadic Adjustment Scale: Una misura dell'adattamento di coppia. *Minerva Psichiatrica*, *43*, 107 – 116.

Goodwin, R. (1992). Overall, just how happy are you? The magical Question 31 of the Spanier Dyadic Adjustment Scale. *Family Therapy*, *19*, 273 – 275.

Gorsuch, R. L. (1983). *Factor analysis* (2nd ed.). Hillsdale, NJ: Erlbaum.

*Gossmann, I., Julien, D., Mathieu, M., & Chartrand, E. (2003). Determinants of sex initiation frequencies and sexual satisfaction in long-term couples' relationships. *Canadian Journal of Human Sexuality*, *12*, 169 – 181.

Graham, J. M. (in press). Congeneric and (essentially) tau-equivalent estimates of score reliability: What they are and how to use them. *Educational and Psychological Measurement*.

*Green, R. G., & Sporakowski, M. J. (1983). The dynamics of divorce: Marital quality, alternative attractions and external pressures. *Journal of Divorce*, *7*, 77 – 88.

*Grekin, E. R., Sher, K. J., & Larkins, J. M. (2004). The role of behavioral undercontrol in the relation between alcohol use and partner aggression. *Journal of Studies on Alcohol*, *65*, 658 – 662.

Gronlund, N. E., & Linn, R. L. (1990). *Measurement and evaluation in teaching* (6th ed.). New York: Macmillan.

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

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.

Hedges, L. V., & Vevea, J. L. (1998). Fixed- and random-effects models in meta-analysis. *Psychological Methods*, *3*, 486 – 504.

*Heyman, R. E., Sayers, S. L., & Bellack, A. S. (1994). Global marital satisfaction versus marital adjustment: An empirical comparison of three measures. *Journal of Family Psychology*, *8*, 432 – 446.

*Horesh, N., & Fennig, S. (2000). Perception of spouses and relationships: A matched control study of patients with severe affective disorder in remission and their spouses. *Journal of Nervous and Mental Disease*, *188*, 463 – 466.

*Houseknecht, S. K., & Macke, A. S. (1981). Combining marriage and career: The marital adjustment of professional women. *Journal of Marriage and the Family*, *43*, 651 – 661.

*Howe, G. W., Levy, M. L., & Caplan, R. D. (2004). Job loss and depressive symptoms in couples: Common stressors, stress transmission, or relationship disruption. *Journal of Family Psychology*, *18*, 639 – 650.

*Hughes, F. M., Gordon, K. C., & Gaertner, L. (2004). Predicting spouses' perceptions of their parenting alliance. *Journal of Marriage and Family*, *66*, 506 – 514.

Hunsley, J., Best, M., Lefebvre, M., & Vito, D. (2001). The seven-item short form of the Dyadic Adjustment Scale: Further evidence for construct validity. *American Journal of Family Therapy*, *29*, 325 – 335.

*Hunsley, J., Pinsent, C., Lefebvre, M., James-Tanner, S., & Vito, D. (1995). Construct validity of the short forms of the Dyadic Adjustment Scale. *Family Relations*, *44*, 231 – 237.

Hunter, J. E., & Schmidt, F. L. (2000). Fixed effects vs. random effects meta-analysis models: Implications for cumulative research knowledge. *International Journal of Selection and Assessment*, *8*, 275 – 292.

*James, S., Hunsley, J., Navara, G. S., & Alles, M. (2004). Marital, psychological, and sociocultural aspects of sojourner adjustment: Expanding the field of enquiry. *International Journal of Intercultural Relations*, *28*, 111 – 126.

*Johnson, S. M., & Greenberg, L. S. (1985). Differential effects of experiential and problem-solving interventions in resolving marital conflict. *Journal of Consulting and Clinical Psychology*, *53*, 175 – 184.

Kazak, A. E., Jarmas, A., & Snitzer, L. (1988). The assessment of marital satisfaction: An evaluation of the Dyadic Adjustment Scale. *Journal of Family Psychology*, *2*, 82 – 91.

*King, L. A. (1993). Emotional expression, ambivalence over expression, and marital satisfaction. *Journal of Social and Personal Relationships*, *10*, 601 – 607.

*Knee, C., Raymond, P., Heather, V., Nathaniel, A., & Neighbors, C. (2004). Implicit theories of relationships: Moderators of the link between conflict and commitment. *Personality and Social Psychology Bulletin*, *30*, 617 – 628.

*Kurdek, L. A. (1992). Dimensionality of the Dyadic Adjustment Scale: Evidence from heterosexual and homosexual couples. *Journal of Family Psychology*, *6*, 22 – 35.

Kurdek, L. A. (1998). The nature and predictors of the trajectory of change in marital quality over the first 4 years of marriage for first-married husbands and wives. *Journal of Family Psychology*, *12*, 494 – 510.

*Kurdek, L. A., & Schmitt, J. P. (1986). Early development of relationship quality in heterosexual married, heterosexual cohabiting, gay, and lesbian couples. *Developmental Psychology*, *22*, 305 – 309.

*Langis, J., Sabourin, S., Lussier, Y., & Mathieu, M. (1994). Masculinity, femininity, and marital satisfaction: An examination of theoretical models. *Journal of Personality*, *62*, 393 – 414.

*LaSala, M. C. (2004). Extradyadic sex and gay male couples: Comparing monogamous and nonmonogamous relationships. *Families in Society*, *85*, 405 – 412.

Lee, M., & Kim, Z. (1996). A preliminary study on the standardization of the Korean Dyadic Adjustment Scale. *Korean Journal of Clinical Psychology*, *15*, 129 – 140.

*Lenz, E. R., Soeken, K. L., Rankin, E. A., & Fischman, S. H. (1985). Sex-role attributes, gender, and postpartal perceptions of the marital relationship. *Advances in Nursing Science*, *7*, 49 – 62.

*Lim, B. K., & Ivey, D. (2000). The assessment of marital adjustment with Chinese populations: A study of the psychometric properties of the Dyadic Adjustment Scale. *Contemporary Family Therapy*, *22*, 453 – 465.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage.

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

*Lussier, Y., Sabourin, S., & Turgeon, C. (1997). Coping strategies as moderators of the relationship between attachment and marital adjustment. *Journal of Social and Personal Relationships*, *14*, 777 – 791.

*Lussier, Y., Sabourin, S., & Wright, J. (1993). On causality, responsibility, and blame in marriage: Validity of the entailment model. *Journal of Family Psychology*, *7*, 322 – 332.

*Manne, S., Ostroff, J., Winkel, G., Goldstein, L., Fox, K., & Grana, G. (2004). Posttraumatic growth after breast cancer: Patient, partner, and couple perspectives. *Psychosomatic Medicine*, *66*, 442 – 454.

*Manne, S., & Schnoll, R. (2001). Measuring cancer patients' psychological distress and well-being: A factor analytic assessment of the Mental Health Inventory. *Psychological Assessment*, *13*, 99 – 109.

*Manne, S., Sherman, M., Ross, S., Ostroff, J., Heyman, R. E., & Fox, K. (2004). Couples' support-related communication, psychological distress, and relationship satisfaction among women with early stage breast cancer. *Journal of Consulting and Clinical Psychology*, *72*, 660 – 670.

*Manusov, V. (1990). An application of attribution principles to nonverbal behavior in romantic dyads. *Communication Monographs*, *57*, 104 – 118.

*Matsumoto, D., LeRoux, J. A., Iwamoto, M., Choi, J., Wook, R., David, T., & Haruyo, U. H. (2003). The robustness of the Intercultural Adjustment Potential Scale (ICAPS): The search for a universal psychological engine of adjustment. *International Journal of Intercultural Relations*, *27*, 543 – 562.

*McLean, R., Marini, I., & Pope, M. (2003). Racial identity and relationship satisfaction in African American gay men. *Family Journal*, *11*, 13 – 22.

McLeod, J. D. (1994). Anxiety disorders and marital quality. *Journal of Abnormal Psychology*, *103*, 767 – 776.

*McMillan, L. H. W., O'Driscoll, M. P., & Brady, E. C. (2004). The impact of workaholism on personal relationships. *British Journal of Guidance and Counseling*, *32*, 171 – 186.

*Meredith, W. H., Abbott, D. A., & Adams, S. L. (1986). Family violence: Its relation to marital and parental satisfaction and family strengths. *Journal of Family Violence*, *1*, 299 – 305.

Miller, M. B. (1995). Coefficient alpha: A basic introduction from the perspectives of classical test theory and structural equation modeling. *Structural Equation Modeling*, *2*, 255 – 273.

*Najman, J. M., Behrens, B. C., Andersen, M., Bor, W., O'Callaghan, M., & Williams, G. M. (1997). Impact of family type and family quality on child behavior problems: A longitudinal study. *Journal of the American Academy of Child and Adolescent Psychiatry*, *36*, 1357 – 1365.

*Northhouse, L. L., Jeffs, M., Cracchiolo-Caraway, A., & Lampman, L. (1995). Emotional distress reported by women and husbands prior to a breast biopsy. *Nursing Research*, *44*, 196 – 201.

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

Nunnally, J. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.

*Onyskiw, J. E., Harrison, M. J., & Magill-Evans, J. E. (1997). Past childhood experiences and current parent-infant interactions. *Western Journal of Nursing Research*, *19*, 501 – 518.

Orwin, R. G. (1983). A fail safe N for effect size in meta-analysis. *Journal of Educational Statistics*, *8*, 157 – 159.

*Parish, W. E. (1992). A quasi-experimental evaluation of the Premarital Assessment Program for premarital counseling. *Australian and New Zealand Journal of Family Therapy*, *13*, 33 – 36.

*Pittman, J. F., Price-Bonham, S., & McKenry, P. C. (1983). Marital cohesion: A path model. *Journal of Marriage and the Family*, *45*, 521 – 531.

Prigerson, H. G., Maciejewski, P. K., & Rosenheck, R. A. (1999). The effects of marital dissolution and marital quality on health and health service use among women. *Medical Care*, *37*, 858 – 873.

*Primomo, J., Yates, B. C., & Woods, N. F. (1990). Social support for women during chronic illness: The relationship among sources and types to adjustment. *Research in Nursing and Health*, *13*, 153 – 161.

Raykov, T. (1997). Scale reliability, Cronbach's coefficient alpha, and violations of essential tau-equivalence with fixed congeneric components. *Multivariate Behavioral Research*, *32*, 329 – 353.

*Remien, R. H., Wagner, G., Dolezal, C., & Carballo-Diéguez, A. (2003). Levels and correlates of psychological distress in male couples of mixed HIV status. *AIDS Care*, *15*, 525 – 538.

Roach, A. J., Frazier, L. P., & Bowden, S. R. (1981). The Marital Satisfaction Scale: Development of a measure for intervention research. *Journal of Marriage and the Family*, *43*, 537 – 546.

*Rosenzweig, J. M., & Dailey, D. M. (1989). Dyadic adjustment/sexual satisfaction in women and men as a function of psychological sex role self-perception. *Journal of Sex and Marital Therapy*, *15*, 42 – 56.

*Rudy, E. B., & Estok, P. J. (1990). Running addiction and dyadic adjustment. *Research in Nursing and Health*, *13*, 219 – 225.

*Rusbult, C. E., Martz, J. M., & Agnew, C. R. (1998). The Investment Model Scale: Measuring commitment level, satisfaction level, quality of alternatives, and investment size. *Personal Relationships*, *5*, 357 – 391.

*Russell, G. (1989). Work/family patterns and couple relationships in shared caregiving families. *Social Behaviour*, *4*, 265 – 283.

Sabourin, S., Bouchard, G., Wright, J., Lussier, Y., & Boucher, C. (1988). L'influence du sexe sur l'invariance factorielle de l'echelle d'adjustement dyadique. *Science et Comportement*, *18*, 187 – 201.

*Sabourin, S., Lussier, Y., Laplante, B., & Wright, J. (1990). Unidimensional and multidimensional models of dyadic adjustment: A hierarchical reconciliation. *Psychological Assessment*, *2*, 333 – 337.

Sabourin, S., Valois, P., & Lussier, Y. (2005). Development and validation of a brief version of the Dyadic Adjustment Scale with a nonparametric item analysis model. *Psychological Assessment*, *17*, 15 – 27.

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

Schumm, W. R., Scanlon, E. D., Crow, C. L., Green, D. M., & Buckler, D. L. (1983). Characteristics of the Kansas Marital Satisfaction Scale in a sample of 79 married couples. *Psychological Reports*, *53*, 583 – 588.

*Sedikides, C., Rudich, E. A., Gregg, A. P., Kumashiro, M., & Rusbult, C. (2004). Are normal narcissists psychologically healthy?: Self-esteem matters. *Journal of Personality and Social Psychology*, *87*, 400 – 416.

*Sharpley, C. F., & Cross, D. G. (1982). A psychometric evaluation of the Spanier Dyadic Adjustment Scale. *Journal of Marriage and the Family*, *44*, 739 – 747.

Sharpley, C. F., & Rogers, H. J. (1984). Preliminary validation of the abbreviated Spanier Dyadic Adjustment Scale: Some psychometric data regarding a screening test of marital adjustment. *Educational and Psychological Measurement*, *44*, 1045 – 1049.

*Sheeran, T., Marvin, R. S., & Pianta, R. C. (1997). Mothers' resolution of their child's diagnosis and self-reported measures of parenting stress, marital relations, and social support. *Journal of Pediatric Psychology*, *22*, 197 – 212.

*Shek, D. T. L. (1994). Psychometric properties of the Chinese version of the Dyadic Adjustment Scale. *Psychologia: An International Journal of Psychology in the Orient*, *37*, 7 – 17.

*Shek, D. T. L. (1995). The Chinese version of the Dyadic Adjustment Scale: Does language make a difference. *Journal of Clinical Psychology*, *51*, 802 – 811.

*Shek, D. T. L. (1999). Individual and dyadic predictors of family functioning in a Chinese context. *American Journal of Family Therapy*, *27*, 49 – 61.

*Shek, D. T., Lam, M. C., Tsoi, K. W., & Lam, C. M. (1993). Psychometric properties of the Chinese version of the Kansas Marital Satisfaction Scale. *Social Behavior and Personality*, *21*, 241 – 249.

Snyder, D. K. (1997). *Marital satisfaction inventory—Revised*. Los Angeles, CA: Western Psychological Services.

*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. (1985). Improve, refine, recast, expand, clarify: Don't abandon. *Journal of Marriage and the Family*, *47*, 1073 – 1074.

Spanier, G. B. (1988). Assessing the strengths of the Dyadic Adjustment Scale. *Journal of Family Psychology*, *2*, 92 – 94.

Spanier, G. B. (1989). *Dyadic Adjustment Scale manual*. North Tonawanda, NY: Multi-Health Systems.

*Spanier, G. B., & Thompson, L. (1982). A confirmatory analysis of the Dyadic Adjustment Scale. *Journal of Marriage and the Family*, *44*, 731 – 738.

*Spotts, E. L., Neiderhiser, J. M., Towers, H., Hansson, K., Lichtenstein, P., Cederblad, M., Pederson, N. L., & Reiss, D. (2004). Genetic and environmental influences on marital relationships. *Journal of Family Psychology*, *18*, 107 – 119.

*Sturge-Apple, M. L., Gondoli, D. M., Bonds, D. D., & Salem, L. N. (2003). Mothers' responsive parenting practices and psychological experience of parenting as mediators of the relation between marital conflict and mother-preadolescent relational negativity. *Parenting: Science and Practice*, *3*, 327 – 355.

*Sussman, L. M., & Alexander, C. M. (1999). How religiosity and ethnicity affect marital satisfaction for Jewish-Christian couples. *Journal of Mental Health Counseling*, *21*, 173 – 185.

Thompson, B., & Vacha-Haase, T. (2000). Psychometrics is datametrics: The test is not reliable. *Educational and Psychological Measurement*, *60*, 174 – 195.

*Ulbrich, P. M., Coyle, A. T., & Llabre, M. M. (1990). Involuntary childlessness and marital

adjustment: His and hers. *Journal of Sex and Marital Therapy*, *16*, 147 – 158.

Vacha-Haase, T. (1998). Reliability generalization: Exploring variance in measurement error affecting score reliability across studies. *Educational and Psychological Measurement*, *58*, 6 – 20.

Vacha-Haase, T., Kogan, L. R., & Thompson, B. (2000). Sample compositions and variabilities in published studies versus those in test manuals: Validity of score reliability inductions. *Educational and Psychological Measurement*, *60*, 509 – 522.

Vacha-Haase, T., Ness, C., Nilsson, J., & Reetz, D. (1999). Practices regarding reporting of reliability coefficients: A review of three journals. *Journal of Experimental Education*, *67*, 335 – 341.

Vandeleur, C. L., Fenton, B. T., Ferrero, F., & Preisig, M. (2003). Construct validity of the French version of the Dyadic Adjustment Scale. *Swiss Journal of Psychology—Schweizerische Zeitschrift für Psychologie—Revue Suisse de Psychologie*, *62*, 161 – 175.

*Van Egeren, L. A. (2004). The development of the coparenting relationship over the transition to parenthood. *Infant Mental Health Journal*, *25*, 453 – 477.

*Van Egeren, L. A., & Hawkins, D. P. (2004). Coming to terms with coparenting: Implications of definition and measurement. *Journal of Adult Development*, *11*, 165 – 178.

*Vaughn, M. J., Matyastik Baier, M. E. (1999). Reliability and validity of the Relationship Assessment Scale. *American Journal of Family Therapy*, *27*, 137 – 147.

*Walker, K. L., Dickson, F. C., & Orbuch, T. (2004). An exploration of illness-related narratives in marriage: The identification of illness-identity scripts. *Journal of Social and Personal Relationships*, *21*, 527 – 544.

*Whisman, M. A., Uebelacker, L. A., & Weinstock, L. M. (2004). Psychopathology and marital satisfaction: The importance of evaluating both partners. *Journal of Consulting and Clinical Psychology*, *72*, 830 – 838.

Wilkinson, L., & the APA Task Force on Statistical Inference (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist*, *54*, 594 – 604.