Published in final edited form as:

Psychol Assess. 2012 September ; 24(3): 676–684. doi:10.1037/a0026637.

Incremental Validity of Spouse Ratings versus Self-Reports of Personality as Predictors of Marital Quality and Behavior during Marital Conflict

Jenny M. Cundiff, Timothy W. Smith, and Clay A. Frandsen

Department of Psychology, University of Utah.

Abstract

The personality traits of neuroticism and agreeableness are consistently related to marital quality, influencing the individual's own (i.e., actor effect) and the spouse's marital quality (i.e., partner effect). However, this research has almost exclusively relied on self-reports of personality, despite the fact that spouse ratings have been found to have incremental validity over self-reports for a variety of other important outcomes. In a study of 300 middle-aged and older married couples, we examined the incremental validity of spouse ratings of neuroticism and agreeableness in predicting concurrent levels of self-reported marital quality, observations of behavior during a marital disagreement task, and depressive symptoms. Neuroticism and agreeableness had expected actor and partner effects on each of these outcomes. Spouse ratings of these traits demonstrated incremental validity in estimates of actor and partner effects on marital quality, marital behavior, and depressive symptoms. Results suggest that spouse ratings of personality may be important additions to the typical reliance on self-reports for research and clinical assessment in marriage.

Keywords

neuroticism; agreeableness; spouse ratings; marriage; marital quality

Marriage and similar intimate relationships are central features in the lives of most adults, and the quality of such relationships influences emotional adjustment, physical health, and well-being (Fincham & Beach, 2010; Kiecolt-Glaser & Newton, 2001; Proulx, Helms, & Buehler, 2007). Personality traits are an important influence on marital quality. Of the Five-Factor Model (FFM; Digman, 1990) traits, neuroticism (vs. emotional stability) and agreeableness (vs. antagonism) are most closely associated with marital quality (Malouff, Thorsteinsson, Schutte, Bhullar & Rooke, 2010). Higher neuroticism (i.e., proneness to

Correspondence regarding this article should be addressed to Timothy W. Smith, Department of Psychology, University of Utah, 390 South 1530 East (room 502), Salt Lake City, UT, 84112. tim.smith@psych.utah.edu.

Publisher's Disclaimer: The following manuscript is the final accepted manuscript. It has not been subjected to the final copyediting, fact-checking, and proofreading required for formal publication. It is not the definitive, publisher-authenticated version. The American Psychological Association and its Council of Editors disclaim any responsibility or liabilities for errors or omissions of this manuscript version, any version derived from this manuscript by NIH, or other third parties. The published version is available at www.apa.org/pubs/journals/pas

¹Models controlling for income and number of years of marriage produced the exact same pattern of results in terms of incremental validity across all four outcome variables, and the R^2 did not increase by more than .01 for any outcome for either husbands or wives.

negative affect, worry, emotional vulnerability) predicts lower levels of the individual's own marital quality (i.e., actor effects) and the spouse's marital quality (i.e., partner effects). Agreeableness has similar actor and partner effects (Malouff et al., 2010), such that higher agreeableness (i.e., trust, tender-mindedness, straightforwardness, cooperativeness, modesty) predicts greater marital quality.

Studies of these associations have typically used self-reports to assess personality traits. However, ratings of a target's personality provided by spouses or other informed persons are often found to have superior predictive validity (Vazire & Carlson, 2011), accounting for as much as twice the variance in a variety of outcomes, including academic or job performance (Connelly & Ones, 2010, p. 1114 - 1115) and physical health (Smith et al., 2008). In perhaps the most relevant example, South, Turkheimer, and Oltmanns (2008) found that when self-reports of personality disorder symptoms were used to estimate actor and partner effects, personality disorder symptoms accounted for 10% of the variance in marital adjustment. When parallel spouse ratings of the target's symptoms were added to the model to estimate actor and partner effects, spouse ratings accounted for an additional 13% of variance in marital adjustment. Further, when self-report and spouse ratings were considered simultaneously, significant actor and partner effects were obtained only for spouse ratings. The incremental validity (Hunsley & Meyer, 2003) and greater overall predictive utility of spouse reports could reflect their greater accuracy or the reduced influence of self-presentational concerns (Vazire & Carlson, 2011). Hence, prior studies using only self-reports may have underestimated associations of neuroticism and agreeableness with marital quality (Malouff et al., 2010).

Given that high neuroticism and low agreeableness are the strongest FFM trait-level correlates of personality disorder symptoms (Samuel & Widiger, 2008), the South et al. (2008) results suggest that spouse ratings of these FFM traits might similarly be significant incremental and stronger predictors of marital functioning relative to self-reports. However, to our knowledge, researchers have not examined the incremental validity of spouse ratings of FFM traits in predicting marital quality.

Further, researchers typically use self-reports of marital quality when examining the association of personality with marital quality. Although self-reports of marital quality have well-established validity (Snyder, Heyman & Haynes, 2005), when participants' provide self-reports of both personality and marital quality, estimates of actor effects are likely to be inflated by common method variance. Similarly, when spouses rate a target's personality and their own marital quality, estimates of partner effects can be inflated. Ratings of marital quality by independent observers provide a valuable method of assessment in this regard, as they do not share common method variance with either self-reports or spouse ratings of personality. Further, behavioral observations during marital disagreements are related to concurrent and future marital quality, and to relationship outcomes such as separation and divorce (Heyman, 2001; Snyder et al., 2005; Snyder, Heyman, & Haynes 2008; Gottman, 1979).

To address these issues, we tested the incremental validity (Hunsley & Meyer, 2003) of spouse ratings of neuroticism and agreeableness over self-reports of these same traits in

estimating actor and partner effects of personality on: a) self-reports of marital quality, and b) independent behavioral ratings of affiliation (i.e., warmth vs hostility) and control (i.e., dominance) during marital disagreements. Self-reports and spouse ratings for a given personality trait are closely related, and these expected associations are seen as evidence of convergent validity (Costa & McCrae, 1992). Hence, the independent associations of self-reports and spouse ratings of a given personality trait with marital quality likely underestimate the association of that trait with marital quality, as their overlapping association with marital quality is not considered. Therefore, in an initial effort to replicate previous research on actor and partner effects of neuroticism and agreeableness on marital quality (Malouf et al., 2010), we combined self-reports and spouse ratings to form composite measures of neuroticism and agreeableness, and examined the association of these composite scales with marital functioning (i.e., marital quality and marital behavior). After this replication, we then examined the incremental validity of spouse ratings of personality, through analyses considering separate self-reports and spouse ratings of neuroticism and agreeableness.

Evidence of the incremental validity of spouse ratings of FFM traits over self-reports of personality could reflect not only their greater accuracy or a reduced influence of self-presentation concerns, but also the fact that spouse ratings are primarily based on the target's behavior when the spouse is present (i.e. marital interactions), whereas self-reports of personality likely reflect a broader range of situations and experiences. If spouse ratings of personality narrowly reflect marital behavior and not broader aspects of personality as typically defined, they may show incremental validity only when predicting marital outcomes. To address this issue, we also examined the incremental validity of spouse ratings of personality in predicting depressive symptoms.

Method

Participants

The Utah Health and Aging Study enrolled 146 middle-aged (wives, $M = 43.9$ years old, range = 32-54 years; husbands, $M = 45.8$, range = 37-59) and 154 older (wives, $M = 62.2$ years old, range = 50-71; husbands, $M = 64.7$, range = 52-76) couples during 2001-2005 (Smith et. al, 2009). The protocol was approved by the University of Utah IRB. Participants were recruited through a polling firm, newspaper ads, and community programs. Screening criteria included: 1) at least one member between 40 and 50 years old (middle-aged group) or between 60 and 70 years old (older group), 2) no more than 10 year age difference between spouses, and 3) no history of cardiovascular disease. Middle-aged couples had been married for an average of 18.4 years ($SD = 6.2$, range = 5 - 31), older couples for an average of 36.4 years ($SD = 10.2$, range = 5 - 53); and 21% of participants had been married previously. Consistent with local demographics, 93% of participants were Caucasian. Median household income was \$50,000 – 75,000 per year.

Measures

Personality—Participants completed neuroticism and agreeableness items from the self-report (Form S) and observer-rating (Form R) versions of the NEO-PI-R (Costa & McCrae,

1992a). These scales have been found previously to have adequate internal consistency (Cronbach's $\alpha > .85$) and convergent and discriminant validity in middle-aged and older adults, for both self-reports and spouse ratings, (Costa et al., 1992a). In the present sample, Cronbach's α exceeded .80 for all scales. Expected patterns of convergent and discriminant validity across husbands' and wives' self-reports and spouse ratings of neuroticism and agreeableness are evident in Table 1.

Self-reported marital quality—Marital quality comprises positivity (i.e., support, warmth) and negativity (i.e., conflict, hostility) dimensions (e.g., Herrington et al., 2008), as well as the traditionally examined single quality dimension. Hence, prior to the first session, participants independently completed the Locke-Wallace (1959) Marital Adjustment Test (MAT), and the support from spouse and conflict subscales of the Quality of Relationship Inventory (QRI) (Pierce, Sarason, & Sarason, 1991). The MAT has high internal consistency ($\alpha = .90$) and substantial evidence of validity (Snyder et al., 2005), although is not without limitations and may not fully capture marital quality when used in isolation (Snyder et al., 2005; 2008). The QRI support and conflict subscales also have acceptable internal consistency ($\alpha = .80$ and $.89$, respectively) and evidence of construct validity in the present sample (Smith et al., 2010). Participants independently completed the Areas of Disagreement Questionnaire (Fincham, 1985), where the degree of disagreement was rated for common problem topics (e.g., household duties, finances). Topics with the greatest spouses' combined score were used for the disagreement task.

Depressive symptoms—Participants also completed the Center for Epidemiological Studies Depression Scale (CES-D) (Radloff, 1977), which has well-documented reliability and construct validity (Joiner et al., 2005).

Behavioral assessments—Both positive and negative behaviors also should be measured for comprehensive behavioral assessment of marital quality, and negative behaviors should include both hostile and controlling behaviors (Heyman, 2001; Ehrensaft et al., 1999). Toward this end, videotaped couple interactions were coded using the Structural Analysis of Social Behavior (SASB) (Benjamin, Rothwiler, & Critchfield, 2006), specifically the SASB-Composite Observational Coding Scheme (SASB-COMP; Florsheim & Benjamin, 2001). SASB is a refinement of the interpersonal circumplex (for a review, see Fournier, Moskowitz, & Zuroff, 2011), the latter describing interpersonal behavior as varying in affiliation (i.e., warm and friendly vs. cold and quarrelsome) and control (i.e., dominant and directive vs. submissive and yielding). Although SASB and the IPC differ in the conceptualization of the control dimension, dominant or controlling behavior is specifically assessed in SASB (see Smith et al., 2009, for a more detailed discussion of marital behavior assessment in this sample). The SASB and IPC affiliation dimensions are similar.

Frequencies for each SASB code for husbands and wives were recorded separately for each minute of the initial 6-minute disagreement discussion. Videotapes were rated by two teams of coders who received a minimum of 75 hours of training in SASB coding and 20 hours of training with SASB-COMP. All coders attained a criterion level of reliability (SASB: Cohen's weighted kappa > 0.70 ; SASB-COMP: intraclass correlation $> .60$). Twenty percent

of tapes were randomly selected for reliability coding; coders were blind to which tapes were selected for reliability coding. The threshold for acceptable inter-coder agreement was $\alpha = .60$. Segments for which coders achieved less than $\alpha = .60$ were consensus coded; consensus codes were used for analyses. Average inter-rater reliability (Shrout & Fleiss, 1979) was .88 for wives, and .89 for husbands. Task totals for each code were calculated. To control differences in overall frequency of behavior, scores for the disagreement task were converted to proportions. Proportions were standardized to produce similar ranges and variances. As described elsewhere (Smith et al., 2011), individual SASB codes were combined - separately for husbands and wives - to form measures of warm, hostile, and controlling (i.e., dominant) behavior. Specifically, SASB codes for active and reactive warmth (i.e., affirm, active love, protect, disclose, reactive love, trust), and active and reactive hostility (i.e., blame, attack, ignore, sulk, recoil, wall-off) were used to create the affiliation dimension through weighted combination of the subscales, and active control (protect, control, blame) codes were combined to form the control composite. Evidence for the construct validity of these composites is also presented elsewhere (Smith et al., 2011).

Procedure

Prior to attending a laboratory session, couples completed marital questionnaires independently. During that session participants engaged in the disagreement task (see Smith et al., 2009). Participants were asked to confirm that the topic selected as the highest rated conflict was, “currently an issue rather than something that has been resolved,” and that it was an issue that, “you can discuss for the full time period with both of you contributing to the discussion, and both of you need to feel comfortable discussing it here.” Couples were instructed that, “we are not expecting you to solve the particular issue right now; you can think of this as an opportunity to work toward making progress on the issue.” Similar procedures have been widely used to study marital conflict and adjustment (Snyder, et al., 2005, 2008). A 6-minute unstructured discussion was used for behavioral coding. Following the laboratory sessions, participants independently completed the NEO PI-R (Forms S and R).

Overview of Analyses

Given the dependency in spouses' levels of marital quality, behavior during disagreement, and depressive symptoms, we performed dyadic analyses (Kenny, Kashy, & Cook, 2006). Specifically, we performed path analyses in Amos 7.0. Initially, we combined standardized self-reports and spouse ratings to form composite measures of both neuroticism and agreeableness, and tested actor and partner effects of these traits on self-report and behavioral measures of marital quality, and depressive symptoms.

We used the model comparison approach (Bollen, 1989) to assess the incremental validity of spouse ratings over self-reports. There were four steps in these models (see Figure 1). The first examined associations of self-reported neuroticism and agreeableness with the individual's own outcomes (i.e., actor effects of self-reported traits). The second examined the incremental utility of spouse ratings of personality in predicting those same outcomes (i.e., actor effects of spouse-rated traits). In the third, we allowed self-reports of personality to also predict the spouse's outcome (i.e., partner effects of self-reported traits). In the fourth,

we allowed spouse reports of personality to predict those same partner outcomes (i.e., partner effects of spouse-rated traits).

Model fit was evaluated with Chi-square values, Root Mean Square Error of Approximation (RMSEA) and the Comparative Fit Index (CFI) (Browne & Cudeck, 1993). Smaller χ^2 values indicate better model fit and p -values *greater than* .05 are considered a good indication that the proposed model does not significantly differ from the data (i.e. good model fit). For RMSEA a value .05 is a close model fit, with .08 indicating adequate fit (Browne & Cudeck, 1993). For CFI a value of .95 or greater is considered adequate model fit. We report two indications of incremental validity: whether added parameters produced a significant improvement in model fit (χ^2 difference p -value < .05), and change in outcome variance (R^2) accounted for by the model.

Results

Preliminary Analyses

Correlations among the study variables are presented in Table 1. Measures of self-reported marital quality were closely inter-correlated (absolute values from $r = .59$ to $r = .70$ for women, and $r = .51$ to $r = .67$ for men). Hence, prior to the main analyses, we reduced self-reported marital quality measures through principle components analysis for men and women separately. Using the QRI support and conflict scales and the MAT, we obtained a one-factor solution (i.e., one eigenvalue greater than 1.0) in both cases. Loadings for all 3 variables on the factor we will call self-reported *Marital Quality* had an absolute value of .80 or greater for both men and women. We created Marital Quality scores for each participant through unit weighting; higher scores reflect greater spouse support and satisfaction, and less conflict.

We also transformed the control variable for both men and women using log10 transformation, as this variable showed significant positive skew (2.2 and 1.5, respectively) as well as kurtosis (6.0 and 2.0, respectively). After transformation skew (−.08 and .18, for men and women respectively) and kurtosis (.06 and −.37, respectively) for the behavioral control variable were within acceptable limits for both men and women.

Actor and Partner Effects of Composite Neuroticism and Agreeableness

Self-reported Marital Quality—The full model with actor and partner effects for composite measures of neuroticism and agreeableness was fully saturated, hence model fit was perfect (RMSEA = .00; CFI = 1.0); R^2 for husbands = .36; R^2 for wives = .34. Further, all paths were significant. Husbands' and wives' neuroticism was negatively associated with both their own ($B = -.24$ and $-.19$ respectively, both $p < .001$) and their partners' Marital Quality ($B = -.18$, $p < .001$ and $B = -.15$, $p < .01$, respectively). Husbands' and wives' agreeableness was positively associated with both their own ($B = .16$, $p < .01$ and $B = .23$, $p < .001$ respectively) and their partners' Marital Quality ($B = .25$ and $.27$ respectively, both $p < .001$). These results replicate prior meta-analytic findings (Malouf et al., 2010).

Observer-rated marital behavior—The full model was fully saturated for both affiliation and control, hence model fit was perfect for both (both RMSEA = .00; both CFI =

1.0); R^2 for husbands' affiliation = .04; R^2 for wives' affiliation = .07; R^2 for husbands' control = .07; R^2 for wives' control = .05. There were two significant paths for the affiliation model: the actor effect of husbands' composite agreeableness scores on their own affiliative behavior during disagreement ($B = .18, p < .01$) and the partner effect of husbands' agreeableness on wives' affiliative behavior during disagreement ($B = .19, p < .01$). There were also two significant paths for the control model: the actor effects of both husbands' and wives' composite agreeableness on their own controlling behavior during disagreement ($B = -.29, p < .001$ and $B = -.19, p < .01$, respectively).

Depressive symptoms—The full model was fully saturated, hence model fit was perfect (RMSEA = .00; CFI = 1.0); R^2 for husbands = .34; R^2 for wives = .37. Significant paths included actor effects of husbands' and wives' composite neuroticism scores on their own depressive symptoms ($B = .59$ and $.58$, both $p < .001$), and a partner effect of husbands' neuroticism on wives' depressive symptoms ($B = .12, p < .05$).

Do spouse ratings of personality have incremental utility above self-reports?

Self-reported Marital Quality—Results are presented in Table 2. Compared to the model estimating actor effects using only self-reported personality, the model adding the estimates of actor effects when personality was also measured via spouse ratings resulted in improved model fit (χ^2 difference (4) = 16.72, $p < .01$), with increases in R^2 of .06 for husbands' and .07 for wives' Marital Quality, respectively.

Compared to this model using both self-reports and spouse ratings to estimate actor effects, the model adding self-reports of neuroticism and agreeableness to estimate partner effects on marital quality resulted in an improved model fit (see Table 2; χ^2 difference (4) = 38.58, $p < .001$), with increases in R^2 of .09 for husbands' and .08 for wives' Marital Quality, respectively. Finally, the model adding partner effects of spouse ratings of personality resulted in an improved fit (χ^2 difference (4) = 98.38, $p < .001$), with increases in R^2 of .12 for husbands' and .17 for wives' Marital Quality, respectively. Hence, spouse ratings of neuroticism and agreeableness had significant incremental validity over self-reports of these traits in estimates of both actor and partner effects on self-reported Marital Quality.

Observer rated marital behavior—Results of incremental validity analyses for observer-rated marital behavior are also presented in Table 2. Compared to the model estimating actor effects using only self-reported personality, the model adding spouse ratings to estimate actor effects resulted in improved model fit for both affiliation and control (χ^2 difference (4) = 14.2 and χ^2 difference (4) = 13.6 respectively, both $p < .01$), with increases in R^2 of .03 and .02 respectively for husbands and .02 and .03 respectively for wives.

Compared to this model using both self-reports and spouse ratings of personality to estimate actor effects, the model adding self-reports to estimate partner effects did not result in improved model fit for either affiliation or control (χ^2 difference (4) = 4.8 and χ^2 difference (4) = 4.2 respectively, both $p > .05$). Finally, the model adding spouse ratings of personality to estimate partner effects resulted in improvement in model fit for affiliation (χ^2 difference (4) = 14.7, $p < .01$) but only approached significance for control (χ^2 difference (4) = 8.8, p

= .07). The non-significance of this increment for control may be due to the excellent fit of the previous model, making a significant difference (i.e., χ^2 difference (4) > 9.5) impossible (see Table 2). The addition of spouse ratings resulted in the doubling of R^2 in the case of affiliation and an increase of .02 in R^2 in the case of control for both husbands and wives. Hence, spouse ratings of neuroticism and agreeableness again had significant incremental validity over self-reports in estimates of actor effects. Spouse ratings also provided improved prediction of partner effects for affiliation, though only approached statistical significance for control.

Depressive symptoms—Results of the incremental validity analyses for depressive symptoms are also presented in Table 2. Compared to the model estimating actor effects using only self-reports of personality, the model adding spouse reports to estimate actor effects resulted in improved model fit (χ^2 difference (4) = 23.30, $p < .001$), with increases in R^2 of .04 for both husbands' and wives' depressive symptoms.

Compared to this model using self-reports and spouse ratings of personality to estimate actor effects, the model adding self-reports to estimate partner effects did not result in improved model fit (χ^2 difference (4) = 1.40, $p > .05$). Finally, the model adding spouse reports of personality to estimate partner effects resulted in a significant improvement in model fit (χ^2 difference (4) = 13.15, $p < .05$). However, the additional variance explained was small, with increases in R^2 of .01 for husbands and .02 for wives.

Hence, despite the strong associations of self-reports of personality with the individual's own depressive symptoms, spouse ratings of these same traits had significant incremental validity for estimates of actor effects. Spouse ratings had incremental validity for partner effects, as well, but the increments in explained variance were small.

Do self-reports have incremental validity over partner ratings?

To this point, we have demonstrated that spouse ratings have incremental validity beyond self-reports of personality in predicting self-reported marital quality, behavior during marital disagreements, and depressive symptoms. To determine if self-reports of personality have incremental validity beyond spouse ratings, we altered the model comparisons. In step one, we estimated actor effects using partner ratings and in step two examined the incremental validity of self-reports for these actor effects. In step three, partner effects were estimated using partner ratings, and step 4 examined the incremental validity of self-reports to estimate partner effects. Results are presented in Table 3.

Self-reported Marital Quality—Compared to the model estimating actor effects using only spouse-rated personality, the model adding self-reports to estimate actor effects resulted in improved model fit (χ^2 difference (4) = 56.30, $p < .001$), with increments in R^2 of .11 and .07 for husbands' and wives' Marital Quality, respectively.

Compared to this model using both self-reports and spouse ratings of personality to estimate actor effects, the model adding spouse ratings to estimate partner effects resulted in an improved model fit (χ^2 difference (4) = 132.10, $p < .001$). The increment in R^2 was .21 and .25 for husbands' and wives' Marital Quality, respectively. Finally, the full model, adding

self-reports to estimate partner effects did not show significantly better fit (χ^2 difference (4) = 4.90, $p > .05$). Change in R^2 was .00 for both husbands and wives. Hence, self-reports of personality demonstrated incremental validity over spouse ratings for actor effects on Marital Quality, but not partner effects.

Observer rated marital behavior—Compared to the model estimating actor effects using only spouse-rated personality, the model adding self-reports to estimate actor effects did not result in improved model fit for affiliation (χ^2 difference (4) = 4.7, $p > .05$); changes in R^2 of .01 for both husbands and wives. However, adding self-reports to estimate actor effects *did* result in improved model fit for control (χ^2 difference (4) = 12.4, $p < .05$); changes in R^2 of .03 and .02 for husbands and wives, respectively.

Compared to this model using both self-reports and spouse ratings of personality to estimate actor effects, the model adding spouse ratings to estimate partner effects resulted in improved model fit for affiliation (χ^2 difference (4) = 14.6, $p < .01$), but not for control (χ^2 difference (4) = 5.2, $p > .05$). The increment in R^2 was .04 and .05 for husbands' and wives' affiliation and .01 for both husbands' and wives' control. Finally, the full model, adding self-reports to estimate partner effects, did not show significantly better fit for either affiliation or control (χ^2 difference (4) = 4.9 and χ^2 difference (4) = 7.2, both $p > .05$). Hence, self-reports of neuroticism and agreeableness did not provide significant incremental validity over spouse ratings of these traits for either actor or partner effects on affiliative behavior during marital disagreement. However, self-reports did provide incremental validity when estimating actor effects, but not partner effects, on controlling behavior during marital disagreement.

Self-reported depressive symptoms—Compared to the model estimating actor effects using only spouse-reported personality, the model adding self-reports to estimate actor effects resulted in improved model fit (χ^2 difference (4) = 165.70, $p < .001$). The increments in R^2 were .20 for both husbands' and wives' depressive symptoms. Compared to this model using both self-reports and spouse ratings of personality to estimate actor effects, the model adding spouse ratings to estimate partner effects resulted in an improved model fit (χ^2 difference (4) = 12.60, $p < .05$). However, the increments in R^2 were small: .01 and .02 for husbands and wives respectively. Finally, the addition of self-reports of personality to estimate partner effects did not result in an improved fit (χ^2 difference (4) = 1.9, $p > .05$). Hence, self-reports demonstrated incremental validity for actor effects of personality on depressive symptoms, but not partner effects.

Model Fit and Effects Size for Self-Reports Only and Spouse Ratings Only

As a final indication of the relative predictive validity of self-reports versus spouse ratings of personality, we examined full models estimating both actor and partner effects on each of the four outcomes, using only self-reports of neuroticism and agreeableness or only spouse ratings of these traits. The fit of these models cannot be compared directly as they have different predictors and are both saturated models (thus have zero degrees of freedom and a χ^2 of zero), but a comparison of the variance accounted for across the two models is informative. Spouse ratings of neuroticism and agreeableness predicted more variance in

Marital Quality ($R^2 = .34$ and $.41$ for husbands and wives, respectively) than did self-reports of these same personality traits ($R^2 = .25$ and $.23$ for husbands and wives, respectively). Also, spouse ratings of personality predicted more variance in affiliation during marital disagreement ($R^2 = .07$ and $.08$ for husbands and wives, respectively) than did self-reports of personality ($R^2 = .02$ and $.04$ for husbands and wives, respectively). Spouse ratings of personality also predicted more variance in control during marital disagreement but only for wives ($R^2 = .06$ for both husbands and wives) than did self-reports ($R^2 = .06$ and $.04$ for husbands and wives, respectively). However, self-reports of personality were more closely related to depressive symptoms ($R^2 = .37$ and $.36$ for husbands and wives, respectively) than were spouse ratings ($R^2 = .24$ and $.25$ for husbands and wives, respectively). Hence, spouse ratings of personality are more closely related to Marital Quality and behavior during conflict than are self-reports, with the exception of behavioral ratings of husbands' control. However, although spouse reports show incremental validity in predicting depressive symptoms (Table 2), self-reports of neuroticism and agreeableness are more closely related to this outcome than are spouse ratings of these traits.

Discussion

Consistent with past research (Malouf et al., 2010), each of the actor and partner effects of the composite neuroticism and agreeableness measures on husbands' and wives' self-reports of marital quality were significant. For observer-rated behavior during marital disagreement, husbands' and wives' composite agreeableness scores were associated with their own control during disagreement (actor effects). Husbands' composite agreeableness was also significantly associated with both their own and their wives' affiliation during disagreement (actor and partner effect). We found no effects of composite neuroticism on marital behavior. Using the composite measures, there were also significant actor effects of husbands' and wives' neuroticism on their own depressive symptoms, and a significant partner effect of husbands' neuroticism on their wives' depressive symptoms.

As in past research, self-reports of neuroticism and agreeableness were significant in many of these actor and partner effects. Importantly, spouse ratings of these traits demonstrated incremental validity in seven out of eight possible instances – both actor and partner effects on self-reported marital quality, actor and partner effects on affiliation, actor effect on control, and both actor and partner effects on depressive symptoms. Further, although the improvement in model fit only approached statistical significance in the case of partner effects on control, inclusion of spouse ratings did result in notable increases in effect size ($R^2 = .02$) for both husbands' and wives' controlling behavior during disagreement. Hence, there was clear evidence across outcomes for the incremental validity of spouse ratings when examining the actor and partner effects of neuroticism and agreeableness on marital functioning and emotional adjustment. Thus, previous studies that relied solely on self-reports of personality may have underestimated these associations. The fact that the incremental validity of spouse ratings was evident not only for marital outcomes but also for depressive symptoms suggests that their incremental validity is not simply due to the fact that they are narrowly based on the target's behavior during marital interactions.

By comparison, self-reports of neuroticism and agreeableness demonstrated incremental validity over spouse ratings of these traits in only three out of the eight possible instances – actor effects on self-reports of marital quality, actor effects on observer rated control during marital disagreement, and actor effects on depressive symptoms. Importantly, there was no evidence of incremental validity of self-reports of personality in predicting partner's reports of marital quality, and only limited evidence of incremental validity of self-reports of personality in predicting observer ratings of marital behavior. Thus, the incremental validity of self-reports of personality may be most evident in the prediction of subjective outcomes assessed via a single source. In contrast, the incremental validity of spouse ratings was evident for both subjective and objective outcomes, and for both associations involving a single source (e.g., spouse ratings of personality in estimating partner effects on reported marital quality) and those associations involving two sources (e.g., spouse ratings of personality in estimating actor effects).

There are some notable limitations of this study. First, the majority of the sample was White and middle or upper SES; findings may not generalize to other populations. Second, we examined middle-aged and older couples, the vast majority of whom were in long-term marital relationships. Most of these couples had decades of knowledge about one another and our findings may not generalize to younger couples or couples who are recently married. Further, our findings may not extend to close relationships beyond heterosexual marriage. Also, we did not examine other traits (e.g., conscientiousness) that have important associations with marital functioning (Roberts et al, 2007; Kosek, 1996), as we did not include measures of other FFM traits. It is possible that results for spouse ratings of other traits might differ. Finally, all of the associations tested here were concurrent, and a stronger test of incremental validity of spouse reports of personality would involve changes in marital functioning and emotional adjustment (Watson & Humrichhouse, 2006). It also should be noted that participants' fluctuating judgments of marital quality and/or transient depressive symptoms (i.e., states) may have biased ratings of their partners' trait neuroticism and agreeableness (e.g., a depressed husband might rate his wife as less agreeable).

These limitations notwithstanding, the present results add to a growing body of research suggesting that spouse ratings of personality can be a valuable addition to the more commonly used strategy of gathering only self-reports, in both the specific context of marital functioning and in several other important contexts (Connelly & Ones, 2010; Vazire & Carlson, 2011; Smith et al., 2008; South et al., 2008). Further these results suggest that if only one rating can be completed, more information is gained from spouse ratings than self-reports, and this is true not only for self-reported marital quality but independently rated marital behavior – and important predictor of future marital quality and outcome (Gottman, 1979; Gottman, Swanson, & Swanson, 2002). Future research should examine whether this statistically significant increment in predictive validity translates into incrementally useful clinical information. For example, spouse reports might provide a more accurate picture of clients' personality as contributions to maladaptive marital processes, and thus potentially suggest additional or refined avenues for intervention.

Processes leading to differences between self-reports and spouse ratings of personality are an important topic for future research (McCrae, Stone, Fagan, & Costa, 1998; Vazire &

Carlson, 2011), especially since such efforts could help account for the apparent incremental validity of spouse ratings. In the meantime, for researchers examining effects of personality on marital processes and other outcomes, spouse ratings of personality traits clearly warrant serious consideration. Reliance on self-reports could produce underestimates of important associations. Further, when designing marital assessment protocols (Snyder et al., 2005), clinicians should at least consider spouse ratings of personality as a potentially useful addition to traditional approaches.

Acknowledgments

This research was supported by NIH grant # R01 AG018903.

References

- Benjamin L, Rothweiler J, Critchfield K. The use of structural analysis of social behavior (SASB) as an assessment tool. *Annual Review of Clinical Psychology*. 2006; 2:83–109.
- Bollen, KA. *Structural equations with latent variables*. Wiley; New York: 1989.
- Browne, MW.; Cudeck, R. Alternative ways of assessing model fit.. In: Bollen, K.; Long, JS., editors. *Testing structural equation models*. Sage; Newbury Park: 1993. p. 136-162.
- Connelly BS, Ones DS. An other perspective on personality: meta-analytic integration of observers' accuracy and predictive validity. *Psychological Bulletin*. 2010; 136:1092–1122. [PubMed: 21038940]
- Costa, PT., Jr.; McCrae, RR. *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI): Professional manual*. Psychological Assessment Resources; Odessa, FL: 1992a.
- Digman J. Personality structure: Emergence of the five-factor model. *Annual Review of Psychology*. 1990; 41:417–440. Retrieved from PsycINFO database.
- Ehrensaft M, Langhinrichsen-Rohling J, Heyman R, O'Leary K, Lawrence E. Feeling controlled in marriage: A phenomenon specific to physically aggressive couples? *Journal of Family Psychology*. 1999; 13(1):20–32.
- Fincham FD. Attribution processes in distressed and nondistressed couples: Responsibility for marital problems. *Journal of Abnormal Psychology*. 1985; 94:183–190. [PubMed: 3998286]
- Fincham FD, Beach SRH. Marriage in the new millennium: A decade in review. *Journal of Marriage and Family*. 2010; 72:630–649.
- Florsheim, P.; Benjamin, LA. The Structural Analysis of Social Behavior observational coding scheme.. In: Kerig, P.; Lindahl, LM., editors. *Family observational coding systems: resources for systematic research*. Erlbaum; Hillsdale, NJ: 2001. p. 101-113.
- Fournier, MA.; Moskowitz, DS.; Zuroff, DC. Origins and applications of the interpersonal circumplex.. In: Horowitz, LM.; Strack, S., editors. *Handbook of interpersonal psychology: Theory, research, assessment, and therapeutic interventions*. John Wiley & Sons; Hoboken, NJ: 2011. p. 57-73.
- Gottman JM. Detecting cyclicity in social interaction. *Psychological Bulletin*. 1979; 86:338–348.
- Gottman JM, Swanson C, Swanson K. A general systems theory of marriage: Nonlinear difference equation modeling of marital interaction. *Personality and Social Psychology Review*. 2002; 6:326–340.
- Herrington R, Mitchell A, Castelli A, Joseph J, Snyder D, Gleaves D. Assessing disharmony and disaffection in intimate relationships: Revision of the Marital Satisfaction Inventory scales. *Psychological Assessment*. 2008; 20:341–350. [PubMed: 19086757]
- Heyman R. Observation of couple conflicts: Clinical assessment applications, stubborn truths, and shaky foundations. *Psychological Assessment*. 2001; 13:5–35. [PubMed: 11281039]

- Hunsley J, Meyer GJ. The incremental validity of psychological testing and assessment: Conceptual, methodological, and statistical issues. *Psychological Assessment*. 2003; 15:446–455. [PubMed: 14692841]
- Joiner TE, Walker RL, petit JW, Perez M, Cukrowicz KC. Evidence-based assessment of depression. *Psychological Assessment*. 2005; 17:267–277. [PubMed: 16262453]
- Kenny, DA.; Kashy, DA.; Cook, WL. *Dyadic data analysis*. Guilford Press; New York, NY: 2006.
- Kiecolt-Glaser J, Newton T. Marriage and health: His and hers. *Psychological Bulletin*. 2001; 127:472–503. [PubMed: 11439708]
- Kosek RB. The quest for a perfect spouse: Spousal ratings and marital satisfaction. *Psychological Reports*. 1996; 79:731–735.
- Locke JJ, Wallace KM. Short marital adjustment and prediction tests: Their reliability and validity. *Marriage and Family Living*. 1959; 21:251–255.
- Malouff JM, Thorsteinsson E, Schutte N, Navjot B, Rooke S. The Five-Factor Model of personality and relationship satisfaction of intimate partners: A meta-analysis. *Journal of Research in Personality*. 2010; 44:124–127.
- McCrae R, Stone S, Fagan P, Costa P. Identifying causes of disagreement between self-reports and spouse ratings of personality. *Journal of Personality*. 1998; 66:285–313. [PubMed: 9615420]
- Pierce G, Sarason I, Sarason B. General and relationship-based perceptions of social support: Are two constructs better than one? *Journal of Personality and Social Psychology*. 1991; 61:1028–1039. [PubMed: 1774625]
- Proulx CM, Helms HM, Buehler C. Marital quality and personal well-being: A meta-analysis. *Journal of Marriage and Family*. 2007; 69:576–593.
- Radloff L. The CES-D Scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement*. 1977; 1(3):385–401.
- Roberts BW, Kuncel NR, Shiner RL, Caspi A, Goldberg LR. The power of personality: The comparative validity of personality traits, socio-economic status, and cognitive ability for predicting important life outcomes. *Perspectives on Psychological Science*. 2007; 2:313–345. [PubMed: 26151971]
- Samuel DB, Widiger TA. A meta-analytic review of the relationships between the five-factor model and DSM-IV-TR personality disorders: A facet level analysis. *Clinical Psychology Review*. 2008; 28:1326–1342. [PubMed: 18708274]
- Shrout PE, Fleiss JL. Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*. 1979; 86:420–428. [PubMed: 18839484]
- Smith TW, Berg CA, Florsheim P, Uchino BN, Pearce G, Hawkins M, et al. Conflict and collaboration in middle-aged and older couples: I. Age differences in agency and communion during marital interaction. *Psychology and Aging*. 2009; 24:259–273. [PubMed: 19485646]
- Smith TW, Traupman E, Uchino BN, Berg CA. Interpersonal circumplex descriptions of psychosocial risk factors for physical illness: Application to hostility, neuroticism, and marital adjustment. *Journal of Personality*. 2010; 78:1011–1036. [PubMed: 20573134]
- Smith TW, Uchino BN, Berg CA, Florsheim P, Pearce G, Hawkins M, et al. Associations of self-reports versus spouse ratings of negative affectivity, dominance, and affiliation with coronary artery disease: Where should we look and who should we ask when studying personality and health?. *Health Psychology*. 2008; 27:676–684. [PubMed: 19025262]
- Smith TW, Uchino BN, Florsheim P, Berg CA, Butner J, Hawkins M, et al. Affiliation and control during marital disagreement, history of divorce, and asymptomatic coronary artery calcification in older couples. *Psychosomatic Medicine*. 2011; 73:350–357. [PubMed: 21364198]
- Snyder DK, Heyman RE, Haynes SN. Evidence-based approaches to assessing couple distress. *Psychological Assessment*. 2005; 17:288–307. [PubMed: 16262455]
- Snyder, DK.; Heyman, RE.; Haynes, SN. Couple distress.. In: Hunsley, J.; Mash, E., editors. *A Guide to Assessments That Work*. Oxford University Press; New York: 2008. p. 439-463.
- South S, Turkheimer E, Oltmanns T. Personality disorder symptoms and marital functioning. *Journal of Consulting and Clinical Psychology*. 2008; 76:769–780. [PubMed: 18837594]
- Vazire S, Carlson EN. Others sometimes know us better than we know ourselves. *Current Directions in Psychological Science*. 2011; 20:104–108.

Watson D, Humrichouse J. Personality development in emerging adulthood: Integrating evidence from self-and spouse-ratings. *Journal of Personality and Social Psychology*. 2006; 91:959–974. [PubMed: 17059313]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 1.

Path Analysis Model. Dashed lines signify actor effects and solid lines signify partner effects. "e" denotes measurement error. Single-headed arrows signify a unidirectional relationship (regression) and double-headed arrows signify a correlation between two variables (e.g. husbands' and wives' outcomes). Integers denote which additional paths were allowed to be estimated in the associated step. 1's signify paths in which actor's self-reports of personality are used to estimate actors' own outcomes. 2's signify paths in which spouse ratings of actor's personality are used to estimate actors' own outcomes. 3's signify paths in which self-reports are used to estimate partners' outcomes. 4's signify paths in which spouse ratings are used to estimate partners' outcomes. Though not shown pictorially, all predictor variables were allowed to correlate with one another.

Table 1

Correlation matrix of study variables.

	WSRN	WPRN	HSRN	HPRN	WSRA	WPRA	HSRA	HPRA	WMQ	HMQ	Waff	Haff	Wcon	Hcon	WCESD	HCESD
WSRN																
WPRN	.57***															
HSRN	.21***	.41***														
HPRN	.33***	.23***	.51***													
WSRA	-.29***	-.14*	-.08	-.26***												
WPRA	-.15***	-.37***	-.23***	-.25***	.43***											
HSRA	-.15***	-.31***	-.29***	-.15*	.08	.34***										
HPRA	-.29***	-.18***	-.12*	-.46***	.33***	.22***	.46***									
WMQ	-.35***	-.37***	-.24***	-.46***	.33***	.39***	.26***	.49***								
HMQ	-.24***	-.42***	-.38***	-.36***	.30***	.44***	.30***	.35***	.59***							
Waff	-.13*	-.09	-.01	-.19***	.11	.18**	.15**	.24***	.34***	.20***						
Haff	-.08	-.03	.02	-.18**	.05	.12*	.10	.23***	.23***	.15**	.49***					
Wcon	.00	-.04	-.07	.11	-.16**	-.21***	-.06	-.17**	-.28***	-.14**	-.35***	-.23***				
Hcon	.06	.06	-.03	.14*	-.04	-.09	-.18**	-.25***	-.15**	-.23***	-.25***	-.24***	.27***			
WCESD	.58***	.43***	.22***	.34***	-.07	-.18***	-.13*	-.21***	-.40***	-.18**	-.11	-.07	-.01	.07		
HCESD	.17**	.32***	.60***	.45***	-.09	-.20***	-.20***	-.20***	-.26***	-.34***	.00	-.02	.02	.07	.26***	

WSRN = wives' self-reported neuroticism, WPRN = wives' partner-rated neuroticism, HSRN = husbands' self-reported neuroticism, HPRN = husbands' partner-rated neuroticism, WSRA = wives' self-reported agreeableness, WPRA = wives' partner-rated agreeableness, HSRA = husbands' self-reported agreeableness, HPRA = husbands' partner-rated agreeableness, WMQ = wives' Marital Quality, HMQ = husbands' Marital Quality, Waff = wives' affiliation, Haff = husbands' affiliation, Wcon = wives' control, Hcon = husbands' control, WCESD = wives' depressive symptoms, HCESD = husbands' depressive symptoms.

* p<.05

** p<.01

*** p<.001

Table 2

Results of model comparisons testing incremental validity of spouse ratings of personality above self-reports of the same traits.

	χ^2	<i>p</i>	RMSEA	Adequate Fit	Superior to Previous Step	R ² Husband	R ² Wife
<u>Marital Quality</u>							
Paths 1	153.7	<.001	.19	no	--	.10	.09
Paths 1 + 2	137.0	<.001	.22	no	yes	.16	.16
Paths 1, 2 + 3	98.4	<.001	.25	no	yes	.25	.24
Paths 1, 2, 3 + 4	0	--	--	yes	yes	.37	.41
<u>Marital Behavior</u>							
<u>Affiliation</u>							
Paths 1	33.7	.001	.08	no	--	.00	.01
Paths 1 + 2	19.5	.01	.07	no	yes	.03	.03
Paths 1, 2 + 3	14.7	<.01	.09	no	no	.04	.05
Paths 1, 2, 3 + 4	0	--	--	yes	yes	.08	.10
<u>Control</u>							
Paths 1	26.6	<.01	.06	no	--	.05	.03
Paths 1 + 2	13.0	>.10	.05	yes	yes	.07	.06
Paths 1, 2 + 3	8.8	>.05	.06	yes	no	.08	.06
Paths 1, 2, 3 + 4	0	--	--	yes	no	.10	.08
<u>Depressive Symptoms</u>							
Paths 1	37.9	<.001	.08	no	--	.35	.34
Paths 1 + 2	14.6	.10	.05	yes	yes	.39	.38
Paths 1, 2 + 3	13.2	<.05	.07	no	no	.39	.38
Paths 1, 2, 3 + 4	0	--	--	yes	yes	.40	.40

Path labels refer to corresponding paths in Figure 1. Paths 1 = Self-reports estimating actor effects; Paths 2 = Spouse ratings estimating actor effects; Paths 3 = Self-reports estimating partner effects; Paths 4 = Spouse ratings estimating partner effects. RMSEA = Root mean squared error of approximation. *p*-values greater than .05 signify adequate model fit based on the Chi-squared value. When all paths (1-4) are entered the model is just-identified, thus χ^2 is always 0 for these models and no *p*-values or RMSEA are reportable.

Table 3

Results of model comparisons testing incremental validity of self-reports of personality above spouse ratings of the same traits.

	χ^2	<i>p</i>	RMSEA	Adequate Fit	Superior to Previous Step	R ² Husband	R ² Wife
<u>Marital Quality</u>							
Paths 2	193.3	<.05	.22	no	--	.05	.09
Paths 2 + 1	137.0	<.05	.22	no	yes	.16	.16
Paths 1, 2 + 4	4.9	.42	0.0	yes	yes	.37	.41
Paths 1, 2, 4 + 3	0	--	--	yes	no	.37	.41
<u>Marital Behavior</u>							
<u>Affiliation</u>							
Paths 1	24.2	<.05	.06	no	--	.02	.02
Paths 1 + 2	19.5	<.05	.07	no	no	.03	.03
Paths 1, 2 + 3	4.9	>.10	.03	yes	yes	.07	.08
Paths 1, 2, 3 + 4	0	--	--	yes	no	.08	.10
<u>Control</u>							
Paths 1	25.4	.01	.06	no	--	.04	.04
Paths 1 + 2	13.0	>.10	.05	yes	yes	.07	.06
Paths 1, 2 + 3	7.2	>.10	.05	yes	no	.08	.07
Paths 1, 2, 3 + 4	0	--	--	yes	no	.10	.08
<u>Depressive Symptoms</u>							
Paths 2	180.2	<.05	.21	no	--	.19	.18
Paths 2 + 1	14.5	.10	.05	yes	yes	.39	.38
Paths 1, 2 + 4	1.9	.87	0.0	yes	yes	.40	.40
Paths 1, 2, 4 + 3	0	--	--	yes	no	.40	.40

Path labels refer to corresponding paths in Figure 1. Paths 1 = Self-reports estimating actor effects; Paths 2 = Spouse ratings estimating actor effects; Paths 3 = Self-reports estimating partner effects; Paths 4 = Spouse ratings estimating partner effects. RMSEA = Root mean squared error of approximation. *p*-values greater than .05 signify adequate model fit based on the Chi-squared value. When all paths (1-4) are entered the model is just-identified, thus χ^2 is always 0 for these models and no *p*-values or RMSEA are reportable.