# TM

# TESTING PSYCHOMETRIC PROPERTIES IN DYADIC DATA USING CONFIRMATORY FACTOR ANALYSIS: CURRENT PRACTICES AND RECOMMENDATIONS

SHANNON E. CLAXTON
HAYLEE K. DELUCA
MANFRED H. M. VAN DULMEN
KENT STATE UNIVERSITY

The current paper examines the challenges — and provides recommendations — related to testing the psychometric properties of measures using dyadic data (distinguishable cases). The paper provides a review of the procedures that researchers have been using when performing confirmatory factor analyses (CFAs) on data that is dyadic in nature. In addition to reviewing common methods currently used, we provide examples of how different methods of testing measurement invariance with dyadic data affect results. Specifically, we utilize three dyadic samples (containing 336 couples, $N = 672$) to examine and cross-validate techniques across two measures: the Adult Self Report (Achenbach & Rescorla, 2003) and the Relationship Assessment Scale (Hendrick, 1988). Finally, we present specific recommendations for examining measurement invariance using a CFA framework with dyadic data.

Substantial attention has been devoted in the literature to applying complex techniques that take advantage of the nature of dyadic data (see, e.g., Kenny, Kashy, & Cook, 2006) and to accounting for the nature of non-independent data (e.g., Acock, van Dulmen, Allen, & Piercy, 2004). This is important given that violations of the assumption of non-independence (i.e., when scores from two individuals are interrelated) can lead to biased parameter estimates and standard errors (e.g., Kenny et al., 2006). Common statistical frameworks such as the analysis of variance framework and the regression framework assume that the individual is the sampling unit. Many family and relationship scholars, however, would argue that best practices in relationships research (e.g., research on romantic couples, parent-child dyads) suggest collecting data from dyads — thus the dyad, and not the individual, is the sampling unit. Statistical advancements and reports of best practices with relationship data have largely focused on accounting for — or modeling of — dyadic data in multivariate frameworks that investigate associations among variables, for instance, using the Actor-Partner Interdependence Model to examine actor and partner effects (Kenny et al., 2006), or the Hierarchical Linear Modeling (HLM; Raudenbush & Bryk, 2002). The purpose of the current paper is to extend this body of literature by focusing on the development and validation of measures for use by dyads. In particular, we will evaluate the current practices employed in the literature, discuss various choices researchers must make when using confirmatory factor analysis

(CFA) to develop and/or validate measures for use with dyads, and make several recommendations that we consider to reflect best practices in developing — and enhancing — psychometric properties of measures particularly useful for empirical research with couple dyads.

## INITIAL WAYS TO EVALUATE/DEMONSTRATE NON-INDEPENDENCE

The recognition that data collected from dyads (e.g., marital couples) is inherently non-independent is not new. For several decades researchers have recognized and provided solutions for dealing with dyadic data in, for example, analysis of variance (e.g., Kenny & Judd, 1986) or correlational/regression frameworks (e.g., Gonzalez & Griffin, 1999; Griffin & Gonzalez, 1995). With the rise and increasing flexibility/user-friendliness of statistical software packages such as Mplus (Muthén & Muthén, 1988-2012) and HLM (Raudenbush, Bryk, & Congdon, 2004), the focus in multivariate analyses has shifted toward analytic techniques that can inherently control and model the dyadic nature of couple-level data. Thus, the fact that the individual is not the sampling unit is no longer a problem for statistical analysis.

## IMPORTANCE OF CFA IN THE DEVELOPMENT OF PSYCHOMETRIC PROPERTIES OF MEASURES

Before measures can be used in multivariate analyses, it is essential their psychometric properties are solid and meet scientific standards for reliability and validity (e.g., Carmines & Zeller, 1979). Unless measures — and key constructs derived from these measures — have solid reliability and validity properties, it is unclear what associations between constructs of interest may reflect. At a minimum, the associations would reflect noise in the associations (i.e., the strength of the association is not a good reflection of the actual size of the association between constructs) and noise in the constructs themselves (i.e., the constructs are not measuring what we say they are measuring). In the case of using measures with dyadic data, it is important that researchers establish the measure does not only work well — in terms of psychometric properties — for each individual in the dyad, but also that the psychometric properties are similar across dyad members, or if not account for those differences in statistical analyses. Thus, potential issues of problematic psychometric properties are substantially inflated with dyadic data as we not only work under the assumption that the measure works well across individuals within a particular sample (and thus are an accurate/close reflection of population properties), but also across members of each dyad for a given sample.

Factor analysis techniques are widely used to test psychometric properties when scholars develop new measures/constructs, or when scientists want to validate existing measures with a new sample. In contrast to exploratory factor analysis techniques, which focus on determining the number of latent factors or dimensions present in a measure, CFA allows one to test a proposed factor structure (and thus the construct validity of a measure; see Floyd & Widaman, 1995). Therefore, CFA techniques allow researchers to test theories about the arrangement and connections among variables in a model and specifically evaluate the internal structure of a scale. Further, CFA models can be used to test if a construct is measured the same way across groups (i.e., measurement invariance/equivalence) and to examine associations among measures (thereby evaluating convergent and discriminant validity). Overall, CFA techniques are paramount to es-

tablishing both validity and reliability of measures (for more detailed discussions see Floyd & Widaman, 1995; Furr, 2011).

CFA models should be particularly attractive to relationship scholars because CFA models provide researchers with the ability to (a) test whether measures work similarly across dyad members and (b) model the non-independence inherent to dyadic data. Furthermore, CFA techniques form the cornerstone to the measurement aspect of many other latent variable models, including structural equation models. Thus, carefully considering CFA techniques is an important first step in many studies using dyadic data.

### IMPORTANCE OF CORRECTLY HANDLING CFA WITH DYADIC DATA

It is important to model non-independence during the measurement development phase because (a) not accounting for non-independence can lead to inaccurate standard errors and (b) treating dyadic data as separate samples (e.g., in heterosexual couples separating the sample out by females vs. males) results in a loss of power as well as an inability to account for the non-independence within dyads. Furthermore, studies that only consider individual level data are not fully able to capture if the measure works the same way across different levels of the dyad (e.g., for heterosexual couples the measurement structure may differ for females and males).

Romantic relationships, for example, are inherently interdependent, and couple-level processes (e.g., social support, relationship satisfaction, and commitment) are important for fully understanding the way these relationships function. Consistent with this assumption, most theories of relationships, for instance, Adult Attachment Theory (Hazan & Shaver, 1987), Equity Theory (Messick & Cook, 1983), Interdependence Theory (Kelley & Thibaut, 1978) recognize the importance of mutual influence (i.e., interdependence). For these reasons, interdependence is a core concept in the majority of theories used to understand close relationships.

Overall, the dyadic nature of romantic relationships calls into question the appropriateness of relationship measures that do not utilize couple level data — or have not been validated with couple level data. Yet, many measures designed to evaluate aspects of romantic relationships have been developed based on responses from individuals, ignoring interdependence. It is important, therefore, to examine how we can better develop and validate relationship measures using couples, rather than individuals, as the unit of analysis. Furthermore, we argue that non-relationship measures (that is individual-level measures) that lend themselves well for research with dyads — for example behavioral measures that have a cross-informant component (e.g., Adult Self Report and Adult Behavior Checklist; Achenbach & Rescorla, 2003) — also would benefit from testing psychometric properties within a dyadic framework. Multivariate applications that consider the non-independent nature of the data are already represented in the literature (e.g., van Dulmen & Goncy, 2010). In the current manuscript we extend this work by focusing on testing the psychometric properties of such measures within a dyadic framework.

### INVARIANCE

In addition to accounting for non-independence, an important assumption of research using dyads is that a measure works the same way for both individuals in a dyad. Unfortunately, in-

TPM Vol. 22, No. 2, June 2015
181-198 – Special Issue
© 2015 Cises

Claxton, S. E., DeLuca, H. K.,
& van Dulmen, M. H. M.
Dyadic CFA

variance is often assumed without a formal test to determine equivalency. Measurement invariance, however, can be evaluated within the context of a CFA framework. In understanding invariance, it is important to note that there are two overarching naming schemes for invariance, one based on the degree of invariance (Meredith, 1993): *weak*, *strong*, *strict*, and *structural factorial invariance*; and a second that names separate hypotheses for testing invariance (see Vandenberg & Lance, 2000): *configural*, *metric*, *scalar*, *uniquenesses*, *factor variances*, *factor covariance*, and *factor means invariance* (for a comprehensive discussion of the terminology see Bontempo & Hofer, 2007). The two naming schemes are not mutually exclusive as metric and scalar invariance together comprise strong factorial invariance. In the current paper we will primarily use Vandenberg and Lance's recommended terminology, but note the corresponding factorial invariance met based on Meredith' comparability prerequisites.

### ACCOUNTING FOR NON-INDEPENDENCE WHEN EXAMINING INVARIANCE

Testing invariance in dyads involves additional considerations compared to examining invariance among independent groups. Specifically, one must account for the non-independence of the data while conducting invariance tests. This leaves researchers with several options. One way that many researchers attempt to account for non-independence is to run analyses separated by gender. As mentioned earlier, this technique has a number of flaws including a loss of important information regarding interrelationships between members of the dyad and thus is not preferable (see Kenny et al., 2006). Other, more accurate techniques, allow for non-independence to be accounted for at the factor or item level (or both). In general, researchers control for non-independence at the factor level. Ultimately, how to account for non-independence can be guided by both conceptual and analytic considerations.

At the conceptual level, we pose that relationship-level measures (i.e., measures designed to capture constructs that exist as part of the relationship such as relationship satisfaction, relationship cohesion, or dyadic adjustment — also known as dyadic constructs) should always be tested while accounting for non-independence at both the factor and item level. While such data is inherently non-independent due to the sampling nature of the data (i.e., dyads are the sampling unit) the measures themselves are also dyadic measures (i.e., they ask for information at the dyad level) and as such, conceptually, each of the indicators should also reflect the dyadic nature of the data.

On the other hand, individual-level measures (i.e., measures designed to capture individual differences such as depression or personality traits for partner and self) are conceptually different. Even though the sampling unit here is still the dyad (in couple studies), conceptually, the non-independence could be modeled at the factor level but not at the level of the factor indicators. Statistically, by modeling the non-independence at the factor level, the assumption that the data are inherently non-independent is being modeled/corrected. At the same time, the data may reveal that in order to have a model that reasonably fits the data it is essential to include non-independent correlations at both the factor and indicator level. In other words, methodologically, some may argue one should always test whether to also model non-independence at the indicator level. That is, regardless of the nature of the measure, if model fit improves by adding intercorrelations across indicators to account for non-independence, above and beyond intercorrelations across factors to account for such non-independence, then the former associations should be included in the model. While it is important to identify a model that is a "good" representation of the data, we also pose that these de-

cisions cannot simply be data-driven. As such we urge researchers to balance conceptual and methodological issues and we will also illustrate this with our empirical examples.


## CURRENT PRACTICES FOR CFAS WITH DYADIC DATA

While there is substantial information regarding the treatment of non-independence in a multivariate framework, few practical guides to the treatment of dyadic data when conducting confirmatory factor analysis exist (see Kenny et al., 2006, for suggestions regarding CFA with distinguishable cases; Olsen & Kenny, 2006, for examples with interchangeable dyads; Tagliabue & Lanz, in press). In particular Kenny and colleagues provide suggestions for examining CFA with distinguishable cases (e.g., setting factor loadings to be the same for each dyad member, constraining the variances of the latent variables and observed variable errors to be equal across dyad members, and/or setting the intercepts to be equal for dyad members), whereas Tagliabue and Lanz have more recently suggested using a modified correlated uniqueness model. The literature, however, is lacking (a) an in-depth discussion of the various choices researchers have to make when conducting CFA with dyadic data or (b) what the best practices are in conducting CFAs with dyadic data. The current paper aims to fill this gap.

However, before we discuss the choices researchers have to make and evaluate best practices, we also wanted to investigate the current practices in the field. In order to examine the procedures that researchers use when working with dyadic data we examined all quantitative studies published in the past 10 years (2004-2013) that used dyadic data with a distinguishable case in the Journal of Personal and Social Psychology, Journal of Social and Personal Relationships, Personal Relationships, Journal of Family Psychology, Journal of Marriage and Family, and Family Relations. Articles were specifically coded regarding the methods they used for accounting for non-independence in the data as well as if they examined measurement invariance. If measurement invariance was examined, we also coded the level of invariance reached. In total, 464 articles met our inclusion criteria. Of the articles coded, 335 modeled the non-independent nature of the data. The primary methods for handling non-independence were using Hierarchical Linear Modeling (Multi-Level Modeling) (42.99%) and APIM techniques (35.22%). Fortunately, about three-quarters of all publications with dyadic data accounted for non-independence in some capacity. Of the 129 articles that did not account for non-independence, a full 25.58% separated analyses by gender, whereas the remaining articles failed to properly account for non-independence.

For the current paper, however, we were largely interested to what degree authors tested and reported the psychometric properties of the core measures across dyad members. Only nine (2.7%) articles examined if measures worked the same way for both individuals of the dyad (i.e., tested for measurement invariance). Of these nine articles, all showed metric or weak invariance (factor loadings were constrained to be equal), zero showed or tested for scalar or uniqueness invariance (strong or strict invariance). One article also examined structural invariance (e.g., examined differences in factor means/variances across dyads).


## AIMS

The results of our literature review clearly demonstrate that while many researchers have adopted multivariate techniques that account for non-independence, a large number continue to

use less than optimal techniques for handling non-independence. Furthermore, the results suggest that very few consider the importance of demonstrating measurement invariance across dyads. Given the lack of attention paid in the literature to the non-independence of dyadic data when assessing psychometric properties, the current paper aims to examine various ways to account for non-independence (i.e., dyadic nature of data) and recommends best practices in validating dyadic measures using CFA. Specifically, our work will extend the suggestions provided by previous authors (e.g., Kenny et al., 2006; Tagliabue & Lanz, in press) by testing measurement invariance with two common measures: the Relationship Assessment Scale (RAS; Hendrick 1988) and the Adult Self Report (ASR; Achenbach & Rescorla, 2003).

## MEASURE SELECTION

We selected the ASR and the RAS for the purposes of demonstrating the process of testing measurement invariance in a CFA framework using dyadic data for two primary reasons. First, the ASR is an individual level measure (measuring individual psychological functioning) whereas the RAS measures a relationship level construct (relationship satisfaction). Furthermore, the RAS is a one dimensional scale whereas the ASR provides an example of a multidimensional scale and thus allows for a demonstration of the decisions necessary when working with multiple factors. As such, these measures will provide examples of the process of validating measures for use with couple data.

## METHOD

### Participants

Participants include 336 heterosexual couples who had been dating for at least one month across three dyadic samples collected over several years (Spring 2007-Spring 2008, Fall 2009-Spring 2010, and Spring 2012-Spring 2013). All samples consisted of individuals recruited through an undergraduate research subject pool at a large Midwestern university and their romantic partners. Primary participants received partial course credit and their romantic partners were financially compensated for their participation in the study ($20-25 depending on semester of participation). The sample was primarily Caucasian/White (86%) and had a mean age of 20.07 (1.76) years. Most of the couples (50.6%) had been in a relationship for more than 12 months and roughly 9% were cohabitating at the time of assessment.

### Procedure

As part of larger multi-method protocols both members of the couples completed several self-report measures including a demographic questionnaire, the RAS, and the ASR. All study protocols were approved by the university's Institutional Review Board.

Measures

*Adult Self Report* (ASR; Achenbach & Rescorla, 2003). The ASR for ages 18-59 is a self-report, paper-and-pencil survey used to elicit information regarding psychological functioning. Items are assessed on a 3-point Likert scale (0 = *Not true*, 1 = *Somewhat true or sometimes true*, and 2 = *Very often or very true*). For the current study three narrowband scales were utilized as indicators for the latent constructs of internalizing and externalizing behavior problems, respectively. Specifically, the indicators of externalizing behavior problems were: Aggressive Behavior (e.g., "I argue a lot" and "I am mean to others"), Rule-Breaking Behavior (e.g., "I destroy my own things" and "I act without stopping to think"), and Intrusive ("I brag" and "I try to get a lot of attention"). Similarly, indicators of internalizing behavior problems were: Anxious/Depressed (e.g., "I feel that no one loves me" and "I cry a lot"), Withdrawn (e.g., "I am not liked by other kids" and "I keep from getting involved with others"), and Somatic Complaints (e.g., "I feel dizzy or lightheaded" and "I feel overtired without good reason"). Research has demonstrated good reliability and validity for the scales of the ASR. Associations between cross-informant data have been examined for the ASR but, to our knowledge, dyadic properties have not been examined.

*Relationship Assessment Scale* (RAS; Hendrick, 1988)*.* The RAS is a self-report, paper-and-pencil survey used to assess relationship satisfaction. The survey was administered to each participant/partner and consists of seven items (e.g., "In general, how satisfied are you with your relationship?"). All items were rated on a 5-point Likert scale format, ranging from 1 = *never/ poorly/unsatisfied* to 5 = *extremely well/extremely satisfied/very often/completely/excellent*. Higher scores correspond with greater relationship satisfaction. Research has demonstrated good reliability and validity for the RAS (Hendrick, Dicke, & Hendrick, 1998). The RAS measures a relationship or dyad level construct, yet, to our knowledge, CFAs have not been conducted to validate invariance across partners.

ANALYSIS PLAN

Below we provide two examples to illustrate methods for performing measurement invariance tests in a CFA framework using dyadic data. Specifically, following steps for testing invariance outlined by previous researchers (e.g., Bontempo & Hofer, 2007; van de Schoot, Lugtig, & Hox, 2012), we tested a series of models from fewest constraints to most constraints in order to establish measurement invariance. Overall model fit was evaluated using several different model fit indices: comparative fit index (CFI), root mean square error of approximation (RMSEA), and standardized root mean square residual (SRMR). CFI values closer to 1 indicate better model fit: CFI values between .95 and 1.0 are considered excellent fit, and CFI values less than .90 indicate poor model fit (Hancock & Mueller, 2011; Hu & Bentler, 1999; Kenny et al., 2006). RMSEA values less than .08 suggest reasonable fit whereas values below .06 suggest good fit (Kenney et al., 2006), and SRMR values of .08 or less indicate good model fit (Hu & Bentler, 1999).

Comparisons among these nested models utilized the chi-square difference test (also known as the likelihood-ratio test), in which a statistically significant change in chi-square indicates worsening model fit (i.e., invariance is not met). Given that this test is sensitive to sample size we also examined relative goodness-of-fit indices (see Cheung & Rensvold, 2002). Specifically, we

examined the change in CFI between models, such that CFI changes greater than .01 indicated that invariance was not met. We used the following sequence for testing invariance.

1. Configural invariance: we first examined a model with no equivalency constraints in which the same indicators load on the common factor for both males and females. This model serves as a baseline model against which the other models were compared.

2. Metric invariance (i.e., weak factorial invariance) was examined by constraining factor loadings to be equal for men and women by placing equality constraints on the common factor loadings. This specifically tests the assumption that factor loadings are proportionally equivalent.

3. Scalar invariance (i.e., strong factorial invariance) was tested with the addition of constraints on the intercepts of the manifest variables (i.e., indicators) for men and women across common factors. This evaluates the assumption that the means of the different indicators are similar across men and women.

4. Invariant uniquenesses (i.e., strict factorial invariance) was tested by adding constraints to the specific-factor variances (residuals) in addition to the constraints for metric and scalar invariance. These constraints allow for an examination of the equivalency of the error variances for each manifest indicator (i.e., this examines if the explained variance is the same for each item for males and females and thus if the latent construct is measured equally across groups).

If invariance at one level was not supported, partial invariance was tested by sequentially releasing the constraints in order to determine which specific parameters lacked invariance (see Byrne, Shavelson, & Muthén, 1989; Steenkamp & Baumgartner, 1998; van de Schoot et al., 2012). If at least partial scalar invariance was met, we also evaluated if the mean and variance of the common factors were invariant (i.e., structural invariance) by constraining the common factor means and variances to be equal respectively. Additionally, we considered factor covariance invariance for measures with multiple factors (i.e., ASR).

Notably, for model scaling purposes, when examining invariance regarding the first four steps we constrained the latent variable variance to 1 and the latent variable means to 0, whereas when comparing latent means/variances across groups we constrained one factor loading to 1 and the corresponding intercept to 0 (see van de Schoot et al., 2012). All analyses were conducted in Mplus version 7.11 (Muthén & Muthén, 2012).

RESULTS

Descriptive Statistics and Evaluation of Non-Independence

One important step in working with dyadic data is to assess the degree of non-independence. The couples in this data set are conceptually distinguishable (e.g., they can be distinguished by a single variable, in this case sex). For distinguishable dyads, non-independence can be statistically measured using a Pearson product-moment correlation (Kenny et al., 2006). As evident in the bivariate correlations (Tables 1 and 2), in general the study variables were significantly related for males and females. While some of the correlations were small in size, the majority were medium for the RAS (roughly $r = .30$) and around $r = .20$ for the ASR. Additionally, a canonical correlation with female data predicting male data was statistically significant for both the RAS, Wilks's $\lambda = .52$, $F(49, 1623.93) = 4.82$, $p < .001$, and the ASR, Wilks's $\lambda = .64$, $F(6, 320) =$

TPM®

30.09, $p < .001$, suggesting that the scores for males and females are related at the multivariate level. Overall, these findings suggest that the data is non-independent, or that the scores for individuals within a dyad are more similar than scores for individuals who are not in the same dyad (Kenny et al., 2006). Because non-independence has been demonstrated both conceptually (i.e., individuals are in a romantic relationship and thus mutually influence each other) and statistically, the dyad, rather than the individual, should be the unit of analysis. Therefore, data for these analyses were structured at the dyad level.

TABLE 1
Bivariate correlations, RAS

|  | Item 1 | Item 2 | Item 3 | Item 4r | Item 5 | Item 6 | Item 7r |
|---|---|---|---|---|---|---|---|
| Item 1 | .26*** | .36*** | .34*** | .16** | .28*** | .15** | .30*** |
| Item 2 | .28*** | .44*** | .36*** | .12* | .35*** | .08 | .37*** |
| Item 3 | .27*** | .34*** | .42*** | .15** | .32*** | .13** | .28*** |
| Item 4r | .14** | .24*** | .17*** | .19*** | .27*** | .05 | .17*** |
| Item 5 | .22*** | .27*** | .25*** | .04 | .26*** | .03 | .23*** |
| Item 6 | .10* | .11* | .13** | .03 | .21*** | .24*** | .00 |
| Item 7r | .12* | .25*** | .25*** | .10* | .21*** | −.01 | .39*** |

*Note.* Female scores are on the vertical axis and male scores are on the horizontal axis. r = reverse scored.
* $p < .05$. ** $p < .01$. *** $p < .001$.

TABLE 2
Bivariate correlations, ASR

|  | Anx_Dep | Withdraw | Somatic | Aggression | Rule_Breaking | Intrusive |
|---|---|---|---|---|---|---|
| Anx_Dep | .15** | .23*** | .09* | .12* | .11* | .01 |
| Withdraw | .19*** | .20*** | .15** | .10* | .12* | .00 |
| Somatic | .07 | .03 | .10* | .05 | .07 | −.04 |
| Aggression | .14** | .17*** | .06 | .13** | .06 | .11* |
| Rule_Breaking | .15** | .15** | .11* | .08 | .25*** | .03 |
| Intrusive | −.03 | −.02 | −.02 | .05 | −.02 | .14** |

*Note.* Female scores are on the vertical axis and male scores are on the horizontal axis.
* $p < .05$. ** $p < .01$. *** $p < .001$.

## RAS

*Configural invariance.* We first examined a CFA model including no constraints and including a covariance between common factors for males and females (RAS Model 1a; see Figure 1 without C1-C7). In other words, we only accounted for non-independence at the factor level. This model provided a reasonable fit: $\chi^2(76) = 244.13$, $p < .001$. While the chi-square value was significant, the CFI for this baseline model was reasonable (.90), the RMSEA was adequate (.08), and the SRMR was .06 (with values under .08 representing good fit). Based on these fit indexes, the baseline model provides adequate fit (see Table 3 for all model results).
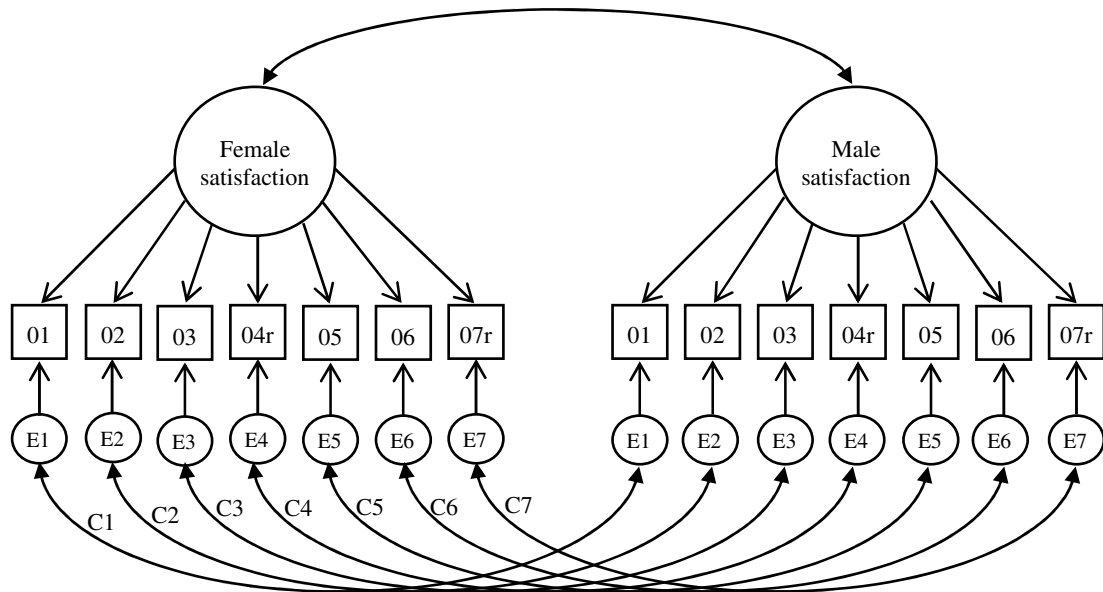
*Note.* r = reverse scored.

FIGURE 1
RAS model.

TABLE 3
RAS model fit

| Model | $\chi^2$ | df | $\Delta\chi^2$ | CFI | $\Delta$CFI | RMSEA | SRMR | AIC | BIC |
|---|---|---|---|---|---|---|---|---|---|
| 1a. Configural | 244.125 | 76 | – | .894 | – | .081 | .059 | 8993.966 | 9158.102 |
| 1b. Configural – including correlated error terms | 172.574 | 69 | 71.55(7)*** | .935 | .041 | .067 | .055 | 8936.415 | 9127.271 |
| 2a. Initial metric invariance | 190.139 | 76 | 17.57(7)* | .928 | .007 | .067 | .106 | 8939.980 | 9104.116 |
| 2b. Final partial metric invariance | 181.551 | 74 | 8.98(5) | .932 | .003 | .066 | .090 | 8935.392 | 9107.162 |
| 3. Partial scalar invariance | 192.224 | 79 | 10.67(5) | .929 | .003 | .065 | .094 | 8936.066 | 9088.750 |
| 4a. Initial partial uniqueness invariance | 220.913 | 84 | 28.69(5)*** | .914 | .015 | .070 | .135 | 8954.755 | 9088.354 |
| 4b. Final partial uniqueness invariance | 198.121 | 82 | 5.89(3) | .927 | .002 | .065 | .120 | 8935.963 | 9077.196 |

*Note*. CFI = comparative fit index; RMSEA = root mean square error of approximation; SRMR = standardized root mean square residual; AIC = Akaike information criterion; BIC = Bayesian information criterion.
* $p < .05$. *** $p < .01$.

We compared this model to a model including correlated error terms between common indicators for males and females (i.e., adding C1-C7 to Figure 1, Model 1b). The statistically sig-

TPM Vol. 22, No. 2, June 2015
181-198 – Special Issue
© 2015 Cises

Claxton, S. E., DeLuca, H. K.,
& van Dulmen, M. H. M.
Dyadic CFA

nificant difference between these two models, $\Delta\chi^2(7) = 71.55$, $p < .001$; $\Delta$CFI = .04, suggests that the model including correlated error terms provides significantly better model fit, supporting the need to control for non-independence at both the item and factor level for this measure. Furthermore, this finding is consistent with our earlier premise that it is important to model non-independence at both the factor and indicator level for relationship measures. This model also provided better model fit. While the $\chi^2$ was still significant, the CFI for this baseline model was acceptable (.94), the RMSEA was good (.07), and the SRMR was good (.06). As such, all subsequent analyses included correlated error terms for common indicators.

*Metric invariance.* Full metric invariance (Model 2a) was tested by adding equality constraints on common factor loadings (i.e., constraining the loading of Item 1 for females to be equal to the loading of Item 1 for males, etc.). While the chi-square difference tests suggested a statistically significant difference between the two models $\Delta\chi^2(7) = 17.57$, $p =.01$, the change in CFI ($\Delta$CFI = .007) indicated that the models were not meaningfully different. Partial invariance was tested (Model 2b) by sequentially relaxing constraints. Results suggest that the potential non-invariance was due to significant differences in the loading for the first and seventh indicator for males and females. Specifically, the factor loading for the item "How well does your partner meet your needs?" was significantly higher for females than males. This suggests that having a partner meet one's needs is more salient for relationship satisfaction for females than for males. On the other hand the factor loading for the item "How many problems are there in your relationship?" (reverse scored) was significantly higher for males than for females, suggesting that the number of problems in a relationship is more salient for males' relationship satisfaction than females'. When these constraints were relaxed, the fit of Model 2b was not significantly worse than Model 1b, $\Delta\chi^2(5) = 8.98$, $p =.11$; $\Delta$CFI = .003.

*Scalar invariance.* Given that only partial metric invariance was achieved, only the intercepts of invariant factor loadings were constrained to be equal in Model 3 when testing scalar invariance (e.g., constraints were added to hold the intercept of Item 2 to be equal for males and females, etc.). Scalar invariance was supported compared to Model 2b, $\Delta\chi^2(5) = 10.67$, $p =.06$; $\Delta$CFI = .003, and model fit remained similar. When compared to the baseline model (Model 1b) the chi-square difference was statistically significant, $\Delta\chi^2(10) = 19.65$, $p = .03$. However, the CFI did not change substantially ($\Delta$CFI = .006) whereas the RMSEA actually improved (see Table 3). As such, partial scalar invariance was supported.

*Uniqueness invariance.* We next examined the equality of residual variance across each item (uniqueness invariance, Model 4a), again only adding constraints on the residual variances for the items found to demonstrate metric invariance (for example constraining the uniqueness for Item 2 to be equal for men and women). Uniqueness invariance was not supported compared to Model 3, $\Delta\chi^2(5) = 28.69$, $p <.001$; $\Delta$CFI = .015. This is not surprising given that invariance at this level is difficult to achieve (see Schmitt & Kuljanin, 2008). Analyses revealed that residual variances differed between men and women for Items 2 and 4. Specifically, residual variance was larger for males than females on both Item 2 ("In general, how satisfied are you with your relationship?") and Item 4 ("How often do you wish you hadn't gotten into this relationship?" — reverse scored). This indicates that there may be more measurement error for these items for males than for females, but may also suggest that these items are not as strong of indicators of relationship satisfaction for males as they are for females. With these constraints released (Model 4b), partial invariance was supported compared to Model 3, $\Delta\chi^2(3) = 5.90$, $p = .12$; $\Delta$CFI = .002.

When compared to the baseline model (Model 1b) the chi-square difference was significant, $\Delta\chi^2(13) = 25.55$, $p = .02$. However, the CFI did not change substantially ($\Delta$CFI = .008). As such, partial uniqueness invariance was supported.

*Structural invariance.* Given that partial scalar invariance was met, we next compared the latent means for relationship satisfaction across males and females (Byrne et al., 1989; Steenkamp & Baumgartner, 1998). The mean for males was 4.56 whereas the mean for females was 4.54. To test factor mean invariance, we constrained the factor loading for Item 3 (which demonstrated invariance across all previous tests) to 1 and the corresponding intercept to 0 for males and females. The addition of constraints on the latent variable means revealed no significant differences compared to a model without these constraints, $\Delta\chi^2(1) = .49$, $p = .48$; $\Delta$CFI = 0. As an alternative test of these differences, the mean for females can also be set to zero, and the significance of the factor mean for males thus represents a test of the difference between males and females (i.e., whether the mean for males differs significantly from zero; Schmitt & Kuljanin, 2008; van de Schoot et al., 2012). No significant difference was found for the mean factor score for males and females ($p = .07$) when using this method.

Factor variance invariance was tested by constraining the latent variable variances to be equal. Factor variance invariance was not demonstrated compared to a model without these constraints, $\Delta\chi^2(1) = 5.83$, $p = .02$; $\Delta$CFI = .003. Specifically, the variance in the latent variable for females was higher than the variance in the latent variable for males, suggesting that the construct of relationship satisfaction has more unexplained variability for females than males.

*Conclusions from the RAS.* Only partial measurement invariance was demonstrated for the RAS, indicating that there were differences between males and females in terms of factor loadings as well as the residuals for the indicators. Furthermore, while the latent means demonstrated invariance, the factor variance was not invariant.

ASR

*Configural invariance.* As with the RAS, we first examined a model including no constraints and including a covariance between common factors for males and females (ASR Model 1a; see Figure 2 without C1-C6). This model provided adequate fit $\chi^2(48) = 178.45$, $p < .001$. While the chi-square value was statistically significant, the CFI (.91) and RMSEA (.09) values for this baseline model were adequate and the SRMR was good (.06) (see Table 4 for all model results).

We compared this model to a model including correlated error terms between common indicators for males and females (Model 1b, adding C1-C6 to Figure 2). The statistically significant difference between these two models, $\Delta\chi^2(6) = 42.02$, $p < .001$; $\Delta$CFI = .02, suggests that the model including correlated error terms provides significantly better model fit, again suggesting the need to control for non-independence at both the item and factor level (even though conceptually the ASR is an individual level measure). This model also provided better model fit. While the difference in $\chi^2$ was still significant, the CFI for this alternative baseline model was good (.94), the RMSEA was adequate (.08), and the SRMR was good (.05). As such, all subsequent analyses included correlated error terms for common indicators.
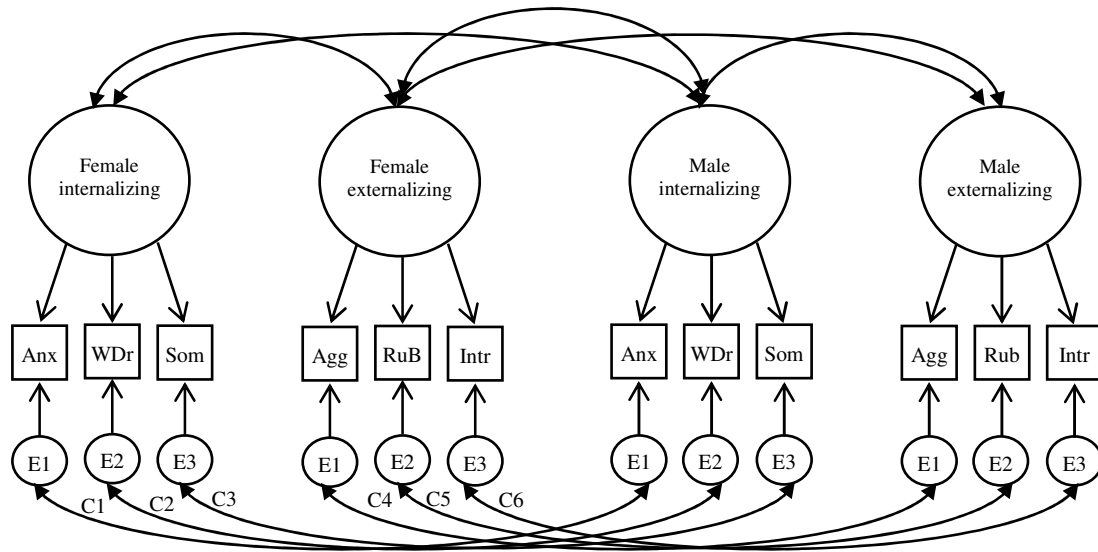
TPM Vol. 22, No. 2, June 2015
181-198 – Special Issue
© 2015 Cises

Claxton, S. E., DeLuca, H. K.,
& van Dulmen, M. H. M.
Dyadic CFA

FIGURE 2
ASR model.

TABLE 4
ASR model fit

| Model | $\chi^2$ | $df$ | $\Delta\chi^2$ | CFI | $\Delta$CFI | RMSEA | SRMR | AIC | BIC |
|---|---|---|---|---|---|---|---|---|---|
| 1a. Configural | 178.488 | 48 | – | .911 | – | .090 | .057 | 20213.66 | 20373.97 |
| 1b. Configural – Including correlated error terms | 136.470 | 42 | 42.02(6)*** | .935 | .0248 | .082 | .053 | 20183.63 | 20366.86 |
| 2a. Initial metric invariance | 167.640 | 48 | 31.17(6)*** | .918 | .017 | .086 | .070 | 20202.808 | 20363.126 |
| 2b. Final partial metric invariance | 142.605 | 46 | 6.14(4) | .934 | .001 | .079 | .058 | 20181.772 | 20349.725 |
| 3a. Initial partial scalar invariance | 200.381 | 50 | 57.78(4)*** | .897 | .037 | .095 | .062 | 20231.547 | 20384.232 |
| 3b. Final partial scalar invariance | 143.392 | 47 | .787(1) | .934 | .000 | .078 | .057 | 20180.559 | 20344.695 |
| 4. Initial partial uniqueness invariance | 145.600 | 48 | 2.21(1) | .933 | .001 | .078 | .056 | 20180.766 | 20341.085 |

*Note*. CFI = comparative fit index; RMSEA = root mean square error of approximation; SRMR = standardized root mean square residual; AIC = Akaike information criterion; BIC = Bayesian information criterion.
***$p < .01$.

*Metric invariance.* Full metric invariance (Model 2a) was tested by adding equality constraints on common factor loadings. Chi-square difference tests as well as change in CFI, $\Delta\chi^2(6)$ = 31.17, $p < .001$; $\Delta$CFI = .02, indicated that invariance was not met, suggesting that there was non-equivalence of factor loadings across males and females. Given these findings, partial invari-

TPM Vol. 22, No. 2, June 2015
181-198 – Special Issue
© 2015 Cises

Claxton, S. E., DeLuca, H. K.,
& van Dulmen, M. H. M.
Dyadic CFA

ance was tested (Model 2b) by sequentially relaxing the constraints. Results suggest that the absence of invariance was due to significant differences in the loading for somatic complaints (on internalizing behavior problems) and rule breaking behavior (on externalizing behavior problems) for males and females. Specifically, the factor loading for somatic complaints was significantly higher for females than males and the factor loading for rule-breaking behavior was significantly higher for males than females. This suggests that somatic complaints are more salient for internalizing behavior problems for females than males and that rule breaking behavior is more salient for externalizing behavior problems for males than females. When these constraints were relaxed, the fit of Model 2b was not significantly worse than Model 1b.

*Scalar and uniqueness invariance.* Given that only partial metric invariance was achieved, only the intercepts of invariant factor loadings were constrained to be equal in Model 3a when testing scalar invariance. Scalar invariance was not supported compared to Model 2b, $\Delta\chi^2(4) = 57.78$, $p < .001$; $\Delta$CFI = .037, and model fit deteriorated markedly. Subsequent analyses revealed that the intercepts of Anxious/Depressed and Withdrawn (for internalizing) and Intrusive (for externalizing) lacked invariance. Specifically, the mean for Anxious/Depressed was higher in females than males whereas the means for Withdrawn and Intrusive were higher for males than females. Thus, the only intercept that demonstrated invariance across males and females was Aggressive Behavior. When comparing this model (Model 3b) to the model for partial metric invariance the difference was non-significant, $\Delta\chi^2(1) = .78$, $p = .34$; $\Delta$CFI = .00. Similarly, a significant difference was not found when compared to the baseline model (Model 1b), $\Delta\chi^2(5) = 6.92$, $p = .23$; $\Delta$CFI = .001. Partial error variance invariance (uniqueness invariance) was evident only for the residual for Aggressive Behavior (Model 4) compared to Model 3b, $\Delta\chi^2(1) = 2.21$, $p = .14$; $\Delta$CFI = .001.

*Structural invariance.* Given that only one item demonstrated invariant factor loadings and invariant intercepts for externalizing behavior problems, and that no items demonstrated scalar invariance for internalizing behavior problems, comparisons of the factor mean, factor intercepts, and the factor covariances (which would be tested by placing constraints on the correlations between internalizing and externalizing behavior problems for males and females) are not meaningful (see Byrne et al., 1989) and thus were not examined.

*Conclusions from the ASR.* Only partial metric invariance was demonstrated for the ASR, indicating that there were a number of differences between males and females in terms of factor loadings, as well as the intercepts and residuals for the indicators.

## DISCUSSION

Despite a sizable literature on the topic of non-independence (e.g., Griffin & Gonzalez, 1995; Kenny et al., 2006) our literature review clearly demonstrates that properly accounting for non-independence is still an issue in relationship research (i.e., it is common practice but not consistent in quantitative research using multivariate techniques). Furthermore, our literature review revealed that most quantitative studies do not examine measurement invariance when using a dyadic sample, or in some way consider non-independence when evaluating psychometric properties. Given that dyadic samples are seen as a best practice in relationship research, this is an important limitation to the field because measurement invariance has implications for how constructs are conceptualized and understood. In order to further the understanding of relationships

through proper implementation of advanced statistical techniques, a consistent method and understanding of these techniques is required.

Therefore, this paper extends the literature by providing a primer on how to perform CFA analyses using dyadic data. We examined the challenges related to conducting factor analyses with non-independent data. We also illustrated the importance of proper handling of non-independence when establishing the psychometric properties of relationship measure and when evaluating measurement invariance. Given the negative consequences of incorrectly handled dyadic data (e.g., unstable estimates and violation of statistical assumptions), the findings of this paper enhance the best practices for dyadic data analysis using CFA.

One important issue when conducting a CFA using dyadic data is how to best account for the non-independence in the data. We have discussed how to model non-independence and made a distinction between individual and relationship measures. Conceptually, we argue, there exists a difference between measures designed to capture relationship constructs versus individual level constructs. However, decisions about whether one should account for non-independence at the factor level only or at both the factor and indicator level should be guided by theory as well as model fit statistics as illustrated in this paper.

We also discussed and illustrated the various levels of invariance that a measure can meet and the importance for examining these different levels of invariance (e.g., factor loadings, intercepts, and uniqueness). Importantly for both the ASR as well as the RAS we demonstrated that the measures did not demonstrate full invariance across different levels. This suggests that the measures do not work in the same way for different members of the dyad. Conceptually, for example, if different items load differently (in terms of strength of factor loading) on a construct then the construct itself may mean different things for males and females. For example, in the ASR we found that Somatic Complaints are more salient for internalizing behavior problems for females than males in our sample. In the RAS for our sample we found "how well a partner meets one's needs" is more salient for relationship satisfaction for females than for males. Differences such as these can advance theory on relationship processes, as well as proper measurement of a given construct, by allowing researchers to statistically test these relationships and how they differ across males and females. Testing these latent associations through CFAs can help us understand and test the generalizability of our theories and constructs.

The finding that neither measure was fully invariant supports the argument that measurement properties for measures to be used with dyadic samples cannot be decided based on individual level data, but rather need to be examined for both members of the dyad. This finding also highlights the importance of using latent variable modeling techniques when working with dyadic data. If analyses with these measures were conducted using a regression framework, it would be assumed that each item contributed equally to the underlying construct for males and females. For the RAS the item "How many problems are there in your relationship?" loaded higher on relationship satisfaction for men, suggesting that the number of problems in a relationship is more salient for males' relationship satisfaction than females'. A regression framework would not take into account this difference, whereas structural equation modeling (SEM) allows one to both detect as well as model differences in the underlying measurement structure for different members of a dyad.

## LIMITATIONS AND FUTURE DIRECTIONS

One important aspect of the current study to note is that for the RAS measurement invariance was tested at the item level whereas for the ASR measurement invariance was tested using subscales. While this choice was made to ease the presentation of the basic technique, using subscales or parcels as an indicator makes the assumption that the reliability of the various subscales is equal and can lead researchers to conclude that measurement invariance exists when it actually does not (see Meade & Kroustalis, 2006). Thus, when testing invariance it is important for researchers to be aware of the choices they make about the level of measurement utilized.

Additionally, an important next step for researchers is to examine the specific effects of using a measure when various levels of invariance are not met for the members of the dyad. Specifically, how are conclusions altered when invariance is not met or tested (and thus the measures are assumed to be invariant across dyad members) versus when full and partial measurement invariance is established? For example, researchers could examine the results of a study examining associations between constructs shown to meet only partial measurement invariance when (a) measures are assumed to be invariant and (b) when the partial measurement invariance is accounted for either through use of SEM or by removing non-invariant indicators. Additionally, simulation studies examining the effects of non-invariance could shed light on the specific negative effects of not meeting invariance.

## CONCLUSIONS

Overall, the current paper contributes to the literature by (a) exploring the current methods of addressing non-independence and prevalence of testing for measurement invariance with dyadic data and (b) demonstrating the steps of conducting CFAs with dyadic data while (c) discussing the relevant decisions researchers need to make when conducting the analyses. Specifically, we illustrate the importance of considering non-independence when evaluating the psychometric properties of measures for use with dyadic data to avoid unstable estimate and the violation of statistical assumptions. Furthermore, our findings confirm that it is important to test psychometric properties at a dyadic level, examining first if the measure works in a similar manner across dyad members and, if it does not, accounting for these differences in statistical analyses. Proper implementation of these techniques and careful consideration by researchers of the conceptual and methodological issues at hand will inform the development and validation of measures for use by dyads.

## REFERENCES

Achenbach, T. M., & Rescorla, L. A. (2003). *Manual for the ASEBA adult forms & profiles*. Burlington, VT: University of Vermont, Research Center for Children, Youth, & Families.

Acock, A., van Dulmen, M. H. M., Allen, K., & Piercy, F. (2004). Contemporary and emerging research methods in studying families. In V. Bengtson, A. Acock, K. Allen, P. Dilowrth-Anderson, & D. Klein (Eds.), *Sourcebook of family theory and research* (pp. 59-89). Thousand Oaks, CA: Sage Publications.

Bontempo, D. E., & Hofer, S. M. (2007). Assessing factorial invariance in cross sectional and longitudinal studies. In A. D. Ong & M. H. M. van Dulmen (Eds.), *Oxford handbook of methods in positive psychology* (pp. 307-359). New York, NY: Oxford University Press.

TPM Vol. 22, No. 2, June 2015
181-198 – Special Issue
© 2015 Cises

Claxton, S. E., DeLuca, H. K.,
& van Dulmen, M. H. M.
Dyadic CFA

# TPM®

Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin*, *105*, 456-466. doi:10.1037/0033-2909.105.3.456

Carmines, E. G., & Zeller, R. A. (Eds.). (1979). *Reliability and validity assessment* (Vol. 17). Thousand Oaks, CA: Sage.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, *9*, 233-255. doi:10.1207/S15328007SEM0902_5

Floyd, F. J., & Widaman, K. F. (1995). Factor analysis in the development and refinement of clinical assessment instruments. *Psychological Assessment*, *7*, 286-299. doi:10.1037/1040-3590.7.3.286

Furr, M. (2011). *Scale construction and psychometrics for social and personality psychology*. Thousand Oaks, CA: Sage.

Gonzalez, R., & Griffin, D. (1999). The correlational analysis of dyad-level data in the distinguishable case. *Personal Relationships*, *6*, 449-469. doi:10.1111/j.1475-6811.1999.tb00203.x

Griffin, D., & Gonzalez, R. (1995). Correlational analysis of dyad-level data in the exchangeable case. *Psychological Bulletin*, *118*, 430-439. doi:10.1037/0033-2909.118.3.430

Hancock, G. R., & Mueller, R. O. (2011). The reliability paradox in assessing structural relations within covariance structure models. *Educational and Psychological Measurement*, *71*, 306-324. doi:10.1177/0013164410384856

Hazan, C., & Shaver, P. R. (1987). Romantic love conceptualized as an attachment process. *Journal of Personality and Social Psychology*, *52*, 511-524. doi:10.1037/0022-3514.52.3.511

Hendrick, S. S. (1988). A generic measure of relationship satisfaction. *Journal of Marriage and the Family*, *50*, 93-98. doi:10.2307/352430

Hendrick, S. S., Dicke, A., & Hendrick, C. (1998). The Relationship Assessment Scale. *Journal of Social and Personal Relationships*, *15*, 137-142. doi:10.1177/0265407598151009

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*, 1-55. doi:10.1080/10705519909540118

Kelley, H. H., & Thibaut, J. W. (1978). *Interpersonal relations: A theory of interdependence*. New York, NY: Wiley.

Kenny, D. A., & Judd, C. M. (1986). Consequences of violating the independence assumption in analysis of variance. *Psychological Bulletin*, *99*, 422-431. doi:10.1037/0033-2909.99.3.422

Kenny, D. A., Kashy, D. A., & Cook, W. L. (2006). *Dyadic data analysis.* New York, NY: Guilford Press.

Meade, A. W., & Kroustalis, C. M. (2006). Problems with item parceling for confirmatory factor analytic tests of measurement invariance. *Organizational Research Methods*, *9*, 369-403. doi:10.1177/1094428105283384

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, *58*, 525-543.

Messick, D. M., & Cook, K. S. (1983). *Equity Theory: Psychological and sociological perspectives.* New York, NY: Praeger Publishers.

Muthén, L. K., & Muthén, B. O. (1998-2012). Mplus user's guide (7th ed.). Los Angeles, CA: Author.

Olsen, J. A., & Kenny, D. A. (2006). Structural equation modeling with interchangeable dyads. *Psychological Methods*, *11*, 127-141. doi:10.1037/1082-989X.11.2.127

Raudenbush, S. W., Bryk, A. S. (2002). *Hierarchical Linear Models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage Publications, Inc.

Raudenbush, S. W., Bryk, A. S, & Congdon, R. (2004). HLM 6 for Windows [Computer Software]. Skokie, IL: Scientific Software International, Inc.

Schmitt, N., & Kuljanin, G. (2008). Measurement invariance: Review of practice and implications. *Human Resource Management Review*, *18*, 210-222. doi:10.1016/j.hrmr.2008.03.003

Steenkamp, J. B. E., & Baumgartner, H. (1998). Assessing measurement invariance in cross national consumer research. *Journal of Consumer Research*, *25*, 78-107. doi:10.1086/209528

Tagliabue, S., & Lanz, M. (in press). Exploring social and personal relationships: The issue of measurement invariance of non-independent observations. *European Journal of Social Psychology*.

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, *3*, 4-70. doi:10.1177/109442810031002

van de Schoot, R., Lugtig, P., & Hox, J. (2012). A checklist for testing measurement invariance. *European Journal of Developmental Psychology*, *9*, 486-492. doi:10.1080/17405629.2012.686740

van Dulmen, M. H. M., & Goncy, E. A. (2010). Extending the Actor-Partner Interdependence Model to include cross-informant data. *Journal of Adolescence*, *33*, 869-877. doi:10.1016/j.adolescence.2010.07.002