TPM®

# THE RECIPROCAL RELATIONSHIP BETWEEN HUSBANDS AND WIVES' MARITAL FORGIVINGNESS: A TWO-WAVE CROSS-LAGGED LATENT DIFFERENCE SCORE ANALYSIS OF TEN-YEAR DATA

F. GIORGIA PALEARI
UNIVERSITY OF BERGAMO

FRANK D. FINCHAM
FLORIDA STATE UNIVERSITY

Cross-lagged latent difference score (LDS) models complement cross-lagged regression models, but are better suited to detecting differences in intrapersonal change and examining the relation between changes in different variables across time. In this article we present cross-lagged LDS models as a method for conceptualizing and measuring change in two-wave dyadic data. The statistical analysis of these models is illustrated using data on marital forgivingness collected from 61 couples at two time points separated by a 10-year interval. The results support the view that cross-lagged LDS models can be an appropriate means to analyze within-person change over two occasions in the context of nonindependent couple data. This is true even when sample size prohibits the estimation of cross-lagged LDS models through common factors using multiple indicators. Conditions that increase model reliability in the absence of multiple indicators are described.

Key words: Latent change; Cross-lagged design; Dyadic data; Forgiveness; Marriage.

*Correspondence concerning this article should be addressed to F. Giorgia Paleari, Department of Human and Social Sciences, University of Bergamo, Piazza S. Agostino 2, 24129 Bergamo (BG), Italy. Email: francesca-giorgia.paleari @unibg.it*

Many questions of interest to marital researchers involve longitudinal processes (Karney & Bradbury, 1995). For example, descriptive questions ask whether spouses/marriages change over time on one or more focal variables (e.g., "Does marital satisfaction decline over time?"), whereas explanatory questions seek to understand these changes by considering possible predictors (e.g., "Do conflict management strategies anticipate declines in marital satisfaction?") and/or mechanisms of change (e.g., "Does forgiveness mediate the relationship between conflict management strategies and declines in marital satisfaction?") (Chan, 1998). In relationships such as marriage, it is important to identify predictors of change not only within, but also between partners; one partner's emotions, cognitions, or behaviors are likely to affect the other partner, both concurrently and over time. Because partners are interdependent (Kelley & Thibaut, 1978), each spouse's standing on one or more focal variables can predict subsequent partner change on one or more related variables (e.g., "Husbands' dissatisfaction or conflict management strategies are likely to anticipate wives' declines in satisfaction"). Also, changes in one partner can be assumed to be accompanied by changes in the other (e.g., "Husbands' declines in satisfaction are likely to be associated with wives' declines").

In order to examine issues such as those outlined above, data from both partners (dyadic data) are needed. The analysis of longitudinal data from married dyads is challenging because it requires us to take into account at least two types of nonindependence simultaneously: autocorrelation and dyadic nonindependence (Kenny, Kashy, & Cook, 2006). Autocorrelation, sometimes called lagged correlation, refers to the association between measurements of the same variable taken at two different points in time; it reflects the fact that the best predictor of present behavior is past behavior. Autocorrelation is usually positive, suggesting a form of "stability," a tendency for a behavior to remain the same from one observation to the next. Dyadic nonindependence indicates the association between two measurements of the same variable taken from the two members of the dyad; it also denotes the association between any parameter (e.g., a slope or an intercept) estimated across dyad members. It reflects the fact that the two members of a dyad, are more similar (or different) to one another than two people who are not members of the same dyad. Dyadic nonindependence is usually positive, reflecting similarity, but it can also be negative, indicating discrepancy.

## OVERVIEW OF CROSS-LAGGED LATENT CHANGE MODELS

The goal of the present article is to present cross-lagged latent difference score (LDS) models. The analysis of these models is complementary to that used for cross-lagged regression models, and we will show that it can be used to examine two-wave dyadic data even when samples are quite small. It is not unusual to have small samples in two-wave designs involving couples, especially when they cover a long period of time and there is missing data. The problem of missing data is magnified when working with dyads as loss of data relating to one dyad member in one wave may result in all data for the dyad being unusable.

Cross-lagged latent difference score models are an extension of cross-lagged regression models that incorporate elements of growth curve analysis. We therefore present these models and their features in relation to better-known cross-lagged regression models. Thus, we begin by providing a brief introduction to the analysis of two-wave dyadic data through cross-lagged regression models estimated via structural equation modelling (SEM). We then consider cross-lagged latent difference score models before showing how to use these models to investigate reciprocal relationships between the tendency to forgive in husbands and wives.

### Cross-Lagged Regression Models

Cross-lagged regression models are widely used in the analysis of two-wave data to investigate the direction and the strength of prospective effects between two variables, while controlling for their stability. In these models, each variable is regressed on both its own lagged score and the lagged score of the other variable at the first measurement occasion ($t_1$); the two variables and their error terms are also allowed to covary at $t_1$ and $t_2$, respectively. The parameters of greatest interest in cross-lagged models are the autoregressive coefficients, that is autocorrelation of each variable with its lagged measurement, and cross-lagged regression coefficients or lagged effects of each variable on the other, net of autocorrelations and covariations specified in the model. Conceptually, autoregressive effects denote stability of the same variable over time, whereas cross-lagged effects reflect the degree to which a variable at $t_1$ predicts change in the other variable from $t_1$ to $t_2$, after controlling for the stability of variables and their association at $t_1$. Thus, if significant cross-lagged coefficients are found for both variables, this means that each variable has an effect on the other over time, supporting a bidirectional or reciprocal effects model. If only one cross-lagged effect is statistically significant, this provides support consistent with a unidirectional effect model. If neither of the cross-lagged effects are significant, the two variables can be inferred to be unrelated over time (Berrington, Smith, & Sturgis, 2006).

When analyzing prospective data from dyad members, the basic structure of two-wave cross-lagged models is shown in Figure 1.

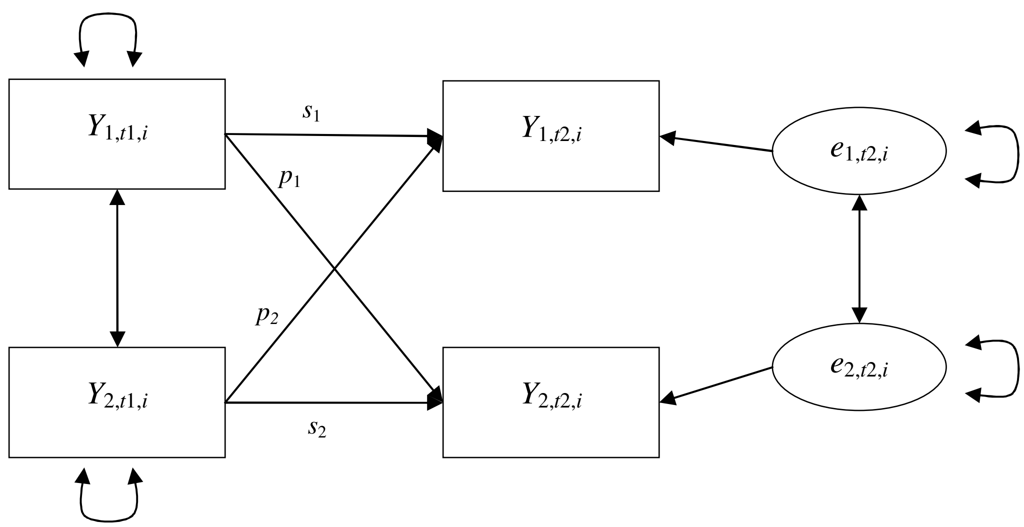



FIGURE 1
Cross-lagged regression model for two-wave dyadic data.

In this model, we have one variable of interest, $Y$, assessed at two times, $t_1$ and $t_2$, in two persons, 1 and 2, belonging to the same couple $i$. For example, $Y$ might be the level of happiness reported by a husband and his wife, and the two times at which their happiness is evaluated might be two points a year apart. Thus, for each couple $i$ we have two equations:

$$Y_{1,t2,i} = c_{1i} + s_{1i}Y_{1,t1,i} + p_{2i}Y_{2,t1,i} + e_{1,t2,i} \quad (1)$$

$$Y_{2,t2,i} = c_{2i} + s_{2i}Y_{2,t1,i} + p_{1i}Y_{1,t1,i} + e_{2,t2,i} \quad (2)$$

and six parameters:

$s_{1i}$, which is interpreted as the stability effect of $Y$ for person 1 (e.g., the stability of husband's happiness over one year);

$s_{2i}$, which is interpreted as the stability effect of $Y$ for person 2 (e.g., the stability of wife's happiness over one year);

$p_{1i}$, which represents partner effect from person 1 to person 2 (e.g., the effect of husband's happiness at baseline on wife's change in happiness over one year);

$p_{2i}$, which represents partner effect from person 2 to person 1 (e.g., the effect of wife's happiness at baseline on husband's change in happiness over one year). Overall, $p_{1i}$ and $p_{2i}$ indicate cross-partner or reciprocity effects;

$c_{1i}$, which is the intercept for person 1 (e.g., the predicted value of husband's happiness at $t_2$ when both husband's and wife's happiness at $t_1$ equal zero);

$c_{2i}$, which is the intercept for person 2 (e.g., the predicted value of wife's happiness at $t_2$ when both husband's and wife's happiness at $t_1$ equal zero).

As Kenny et al. (2006) note, it is very important that predictor variables have a meaningful zero and if not they should be centered. In a temporal model involving dyadic data the variables must be centered by subtracting a common value from person 1's and person 2's data, namely the average of both person 1's and person 2's scores at $t_1$. This makes the path coefficients comparable across dyad members and allows the variance of the six parameters to be meaningfully interpreted. In fact, each of the above six parameters may vary across dyads. For example, if $s_{1i}$ varies it means that $Y$ is more stable for some persons 1 than for others (e.g., happiness is more stable for some husbands than for others); if $p_1$ varies, it indicates that some persons 1 have greater effects on persons 2 than other persons 1 (e.g., the effect of husbands' happiness at baseline on wives' changes in happiness over one year is stronger for some husbands than for others).

As in standard two-wave cross-lagged models, each equation has a residual or error term ($e_{1,t2,i}$ and $e_{2,t2,i}$) representing the effect of all other predictor variables that have not been included in the equation (i.e., the extent to which the $Y_{t2}$ variables are not explained by either of the $Y_{t1}$ variables), plus measurement errors. Error terms are allowed to correlate to control for other sources of nonindependence such as family effects. The correlation between error terms tests the extent to which the dyad members are similar to one another at $t_2$ (e.g., to the extent that husband's happiness at $t_2$ is higher than would be expected on the basis of happiness experienced by husband and wife at $t_1$, then his wife's happiness would also be higher at $t_2$ than one would expect on the same basis). The predictor variables (e.g., husband's and wife's happiness at $t_1$) are also allowed to correlate so that partner effects are estimated while controlling for stability effects and vice versa.

Two-wave cross-lagged dyadic models are usually analyzed via SEM, which allows one to test differences between parameters by imposing constraints on them. This is important because it allows us to determine whether stability effects, partner effects, variances, intercepts or means differ for the members of the dyad. For example, one can test whether stability or actor effects vary across partners (e.g., "Is husbands' happiness more stable than wives' happiness?") and whether partner effects vary (e.g., "Does husbands' happiness predict their wives' later happiness more than wives' happiness predicts husbands' later happiness?"). If members of the dyads show no difference in relation to any parameters estimated, it means that partners are not empirically distinguishable with respect to the variable investigated (Kenny et al., 2006).

The structure presented in Figure 1 can be extended to include variables other than $Y$ (e.g., by assuming that $Y_{1,t2,i}$ and $Y_{2,t2,i}$ are predicted not only by $Y_{1,t1,i}$ and $Y_{2,t1,i}$ but also by another variable $X$ assessed at $t_1$ in both members of the dyad, i.e., $X_{1,t1,i}$ and $X_{2,t1,i}$). It is also possible to model the same structure with latent constructs using multiple indicators instead of observed variables.

*Limitations.* Cross-lagged regression models have important limitations. For example, they are not suitable for detecting differences in patterns of intrapersonal change or for examining the relation between changes in different variables across time. Do persons differ in their change in one domain? Is change in one domain accompanied by change in another domain?

Simple difference scores have been used to address the above questions. A simple difference score, sometimes called a gain score or change score, can be designated as

$$D_{1,i} = Y_{1,t2,i} - Y_{1,t1,i}$$

where $D_{1,i}$ stands for a difference score for person 1 of couple $i$, $Y_{1,t1,i}$ is the person's observed score at the initial assessment, and $Y_{1,t2,i}$ is the same person's observed score at the subsequent assessment. A number of influential scholars, however, have argued that difference scores are less reliable than the two scores from which they are derived, especially if the two scores are positively correlated (Williams & Zimmerman, 1977), as is often the case in longitudinal research because of autoregressive or stability effects. The basis for this assertion may be found in the formula

$$r_{DD} = \frac{r_{11} + r_{22} - 2r_{12}}{2(1 - r_{12})},$$

as well as in the more recent formula

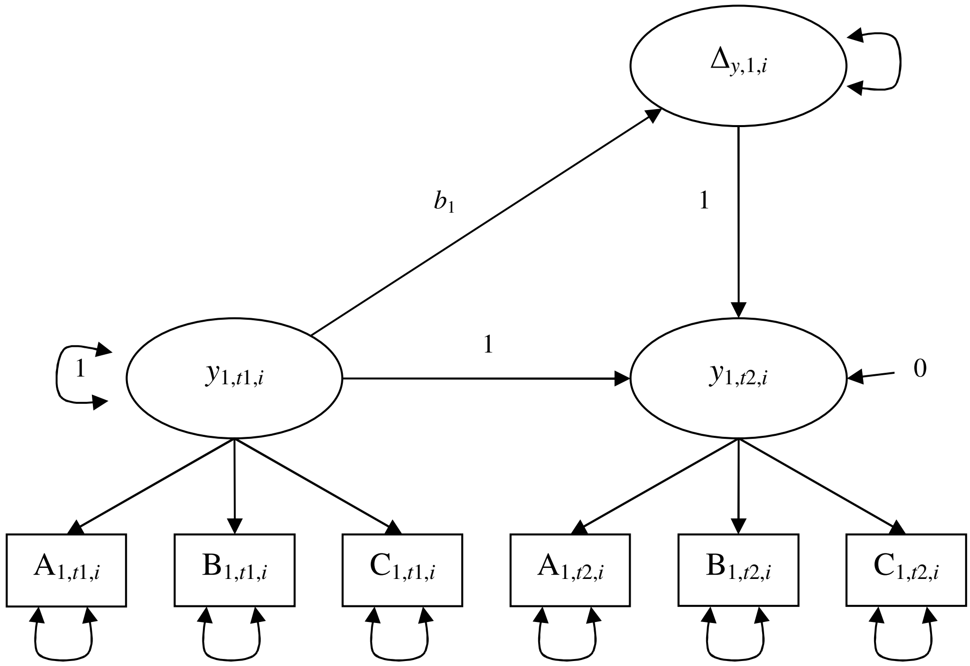
$$r_{DD} = \frac{\frac{SD_1}{SD_2} r_{11} + \frac{SD_2}{SD_1} r_{22} - 2r_{12}}{\frac{SD_2}{SD_1} + \frac{SD_1}{SD_2} - 2r_{12}}$$

which takes into account the dispersion of the scores at $t_1$ and $t_2$ (Williams & Zimmerman, 1996). In these formulas $r_{DD}$ is the reliability of the difference score, $r_{11}$ and $r_{22}$ are the reliability of the measure at $t_1$ and $t_2$ respectively, $r_{12}$ is the correlation between $t_1$ and $t_2$ scores, and $SD_1$ and $SD_2$ are the standard deviations of scores at $t_1$ and $t_2$ (King et al., 2006). According to both formulas, the reliability of the difference score decreases as the correlation between its two component scores increases, assuming other elements are held constant.

On the other hand, when longitudinal data are analyzed using classical approaches other than difference scores, for example using residual-change scores or repeated-measures ANOVA, restrictive assumptions must be met (e.g., sphericity of variances) and the negative effects of unreliability are significantly amplified (Rogosa & Willett, 1983).

## Cross-Lagged Latent Difference Score Models

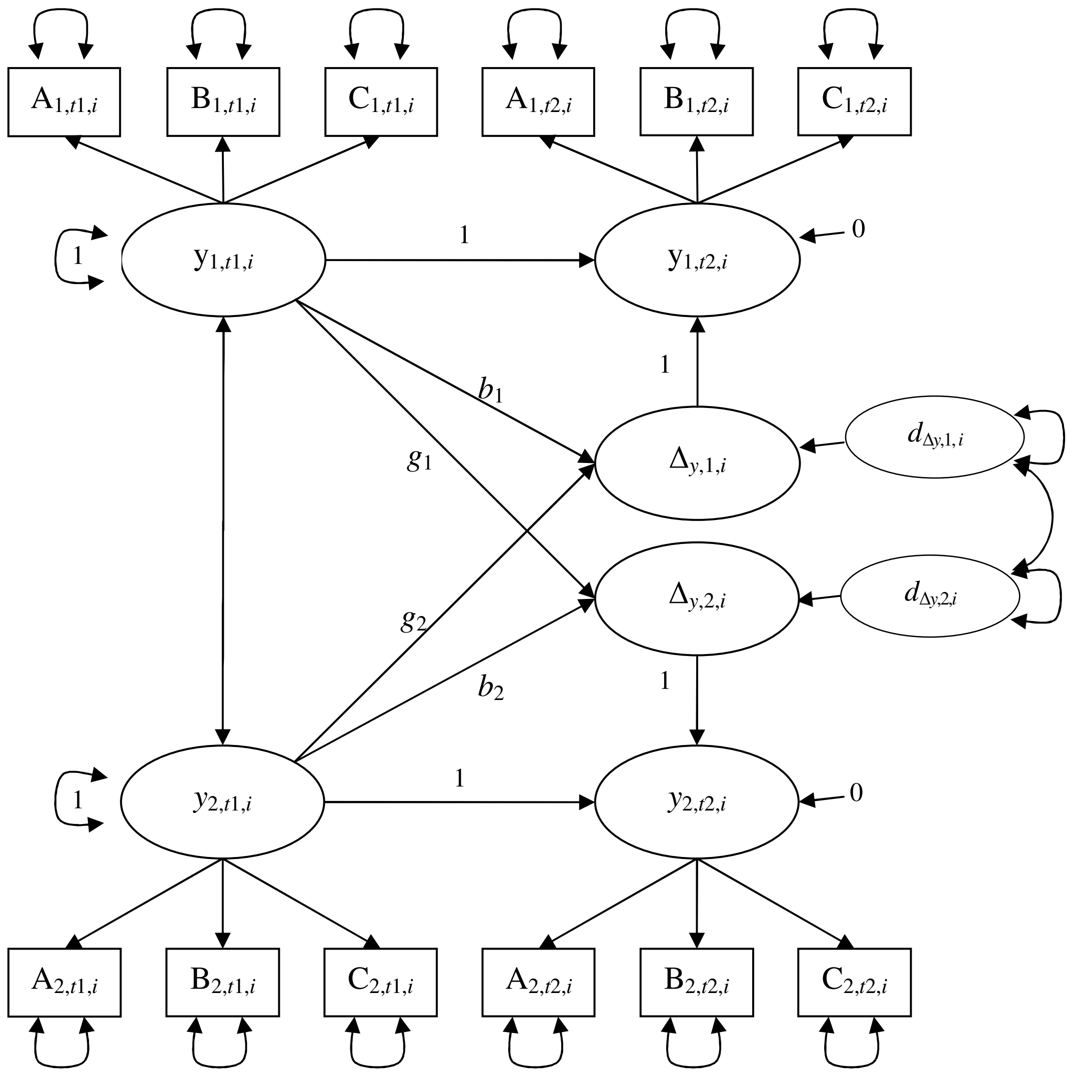
To overcome the problems outlined above, McArdle (2001, 2009) proposed the use of latent difference score (LDS) models that have less restrictive assumptions and are less sensitive to issues associated with measurement error. Inspection of the above $r_{DD}$ formulas reveals that improving the reliabilities of components of the difference score ($r_{11}$ and $r_{22}$), with other elements held constant, will improve the reliability of the difference score ($r_{DD}$). Thus, LDS models first partition the observed scores $Y_{1,t1,i}$ and $Y_{1,t2,i}$ into true scores $y_{1,t1,i}$ and $y_{1,t2,i}$, which are perfectly reliable, and measurement errors $e_{1,t1,i}$ and $e_{1,t2,i}$, so that $Y_{1,t1,i} = y_{1,t1,i} + e_{1,t1,i}$ and $Y_{1,t2,i} = y_{1,t2,i} + e_{1,t2,i}$. LDS models then define the difference between these two latent variables for person 1 of couple $i$ as $\Delta_{y,1,i} = y_{1,t2,i} - y_{1,t1,i}$. Stated differently, $y_{1,t2,i}$ is equal to $y_{1,t1,i}$ plus some change in $y_{1,t1,i}$ or $\Delta_{y,1,i}$, which is not directly measured, therefore it can be considered a latent difference score. To estimate independently the true score and error variance when there are only two measurement occasions, constructs must be assessed by multiple indicators at each time point; in this case variance in each latent factor is error free and, consequently, the change in variance is also measured without error (Little, Bovairds, & Slegers, 2006; McArdle, 2009). Figure 2 shows a LDS model for individual data which estimates change score across two measurement occasions by splitting the variance of $y_{1,t2,i}$ into two components: the $y_{1,t1,i}$ variance and the $\Delta_{y1i}$ variance, or variance associated with the difference from $t_1$.




*Note*. $y_{1,t2,i}$ disturbance is fixed at zero to identify model parameters as well as force the two-wave decomposition. $y_{1,t1,i}$ and $y_{1,t2,i}$ are measured with multiple indicators that are correlated and invariant over time (correlations are not reported for the sake of clarity).

FIGURE 2
Basic latent difference score model for two-wave individual data.

The model includes fixed-unit-value coefficients (= 1), so that the second latent factor $y_{1,t2,i}$ is defined as the sum of $y_{1,t1,i}$ and $\Delta_{y,1,i}$. Given that the model assumes a perfect linear relationship between $y_{1,t1,i}$ and $y_{1,t2,i}$, $y_{1,t1,i}$ will be related to $\Delta_{y,1,i}$ to the extent that this relationship differs from 1. Also, the model estimates a base-free measure of change (i.e., change due to initial level is removed). Because the latent difference score is part of the model, the model parameters include the variation and mean in latent changes across individuals, that is, $\Delta_{y,1,i}$ variance and mean, as well as the relation of change with the initial latent factor $y_{1,t1,i}$. The $\Delta_{y,1,i}$ mean captures the mean-level change and is similar to what would be estimated by repeated measures ANOVA with the exception that the mean-difference is corrected for measurement error and does not assume homogeneity of variances over time. The $\Delta_{y,1,i}$ variance captures individual differences around the mean change which are nearly impossible to describe with cross-lagged regression models. A requirement for LDS models is that the latent or common factors $y_{1,t1,i}$ and $y_{1,t2,i}$ have the same metric at each time of measurement; therefore factor loadings and intercepts of their indicators must be forced to be invariant across the two waves (Meredith, 1993). This simple LDS model provides the basis for more complex models like the cross-lagged LDS model for dyadic data portrayed in Figure 3.




*Note*. $y_{1,t2,i}$ and $y_{2,t2,i}$ disturbances are fixed at zero to identify model parameters as well as force the two-wave decomposition. $y_1$ and $y_2$ are measured with multiple indicators that are correlated and invariant both over time and across partner (correlations are not reported for clarity sake).

FIGURE 3
Cross-lagged latent difference score model with multiple indicators for two-wave dyadic data.

The model can be conceptualized as consisting of two parts: a) a longitudinal factor model that defines the latent variables for the two partners at two occasions and b) a structural model that specifies latent level and change factors for each partner and how changes in latent variables are related within and across partners. The longitudinal factor model assumes that the same configuration of relationships between multiple indicators (or observed variables) and the latent or common factor exists both across waves and across partners. Given a satisfactory longitudinal factor model, the structural model decomposes each partner common factor at $t_2$ into its corresponding common factor and latent difference factor, as required by the basic LDS model, by fixing its residual variance to zero. The structural model assumes that each partner's initial common factor predicts its corresponding LDS ($b_1$ and $b_2$) as well as the other partner's LDS ($g_1$ and $g_2$). By modelling a covariance between the disturbances, $d_{\Delta y,1,i}$ and $d_{\Delta y,2,i}$, which represent the variation in $\Delta_{y,1,i}$ and $\Delta_{y,2,i}$ unexplained by the set of the model paths, the model also assumes that one partner's LDS ($\Delta_{y,1,i}$) covaries with the other partner's LDS ($\Delta_{y,2,i}$), indicating that changes are correlated across partners.

The cross-lagged LDS model for dyadic data presented here emphasizes the need for (a) explicit hypotheses regarding measurement invariance of common factors, (b) explicit structural hypotheses about means and covariances, and (c) inclusion of latent difference scores to express specific temporal hypotheses about within and across partner changes. SEM offers a powerful technique to simultaneously test these specific hypotheses as well as to compare them with alternative ones (McArdle, 2009). For example, SEM allows us to determine whether parameters like the means of LDS, the residual variance of LDS or cross-lagged coefficients $g_1$ and $g_2$ are significantly different from zero or are equivalent across partners. At the same time cross-lagged LDS models require large samples in order to be estimated via SEM, a condition that is seldom met in two-wave, dyadic designs that span long time periods. To overcome this difficulty one can estimate the cross-lagged LDS models by including only the two latent variables that assess change. This strategy, already adopted by a few scholars (e.g., Gerstorf, Röcke, & Lachman, 2010; McArdle & Prindle, 2008), assumes that simple difference scores are not inherently unreliable. As Williams and Zimmerman (1996) showed, the reliability of difference scores is also a function of the standard deviations of their component scores: reliability increases as long as standard deviations are not the same. Actual research data suggest that discrepancies in the values of standard deviations (e.g., due to differential developmental change) are quite common (e.g., Collins, 1996; Nesselroade & Baltes, 1979; Nesselroade & Cable, 1974; Williams & Zimmerman, 1996). Moreover, when we study change we are primarily interested in intraindividual variability, or individual differences in change scores, not on interindividual variability, or individual differences per se. Clearly classical definitions of reliability do not take into account intraindividual variability. This means that even though an instrument perfectly measures change over time, it is possible for it to show poor reliability, and vice versa. For this reason, one can question, as Collins (1996) did, whether it is advisable to have reliable difference scores, "if this says nothing about whether they are precise measures of change" (p. 290).

In light of the above considerations, we therefore respecify the model represented in Figure 3 as one that assumes latent factors for difference scores only (see Figure 4). This LDS model starts with the same observed data $Y_{1,t1,i}$, $Y_{1,t2,\mathrm{i}}$, $Y_{2,t1,i}$, and $Y_{2,t2.,i}$ considered in simple differences scores used in cross-lagged regression models (i.e., it does not imply multiple indicators and therefore posits $Y_{1,t1,i}$ and $Y_{2,t1,i}$ are measured without error). However, this LDS model assumes that the differences between $Y_{1,t2,i}$ and $Y_{1,t1,i}$ and between $Y_{2,t2,i}$ and $Y_{2,t1,i}$ are unobserved or latent variables such that

$$Y_{1,t2,i} = Y_{1,t1,i} + \Delta_{Y,1,i} \text{ and } \Delta_{Y,1,i} = c_{1i} + d_1 Y_{1,t1,i} + g_2 Y_{2,t1,i} + d_{\Delta y,1,i} \quad (3\text{-}4)$$

$$Y_{2,t2,i} = Y_{2,t1,i} + \Delta_{Y,2,i} \text{ and } \Delta_{Y,2,i} = c_{2i} + d_2 Y_{2,t1,\mathrm{i}} + g_1 Y_{1,t1,\mathrm{i}} + d_{\Delta y,2,\mathrm{i}} \quad (5\text{-}6)$$

$\Delta_{Y,1,i}$ and $\Delta_{Y,2,i}$ are unobserved variables. In other words, difference scores are not directly measured but are inferred from other model relations. The model is simply a reinterpretation of the cross-lagged regression model depicted in Figure 1 and formalized through Equations (1) and (2). In fact $b_1 = s_1 - 1$ and $b_2 = s_2 - 1$ (see Figures 1 and 4) and intercepts $c_{1i}$ and $c_{2i}$ are the same for the two models (as previously outlined for cross-lagged regression models, also in this case it is important to center the observed variable means around both partners' grand mean in order to obtain coefficients comparable across dyad members). Thus, having the same number of parameters and achieving the same fit, the cross-lagged regression model of Figure 1 and the cross-lagged latent score model of Figure 4 are not alternatives that can be tested against each other. Nonetheless, they differ in that within-person changes are not described in cross-lagged regression models, whereas they are parameters of cross-lagged difference score models (McArdle, 2009).

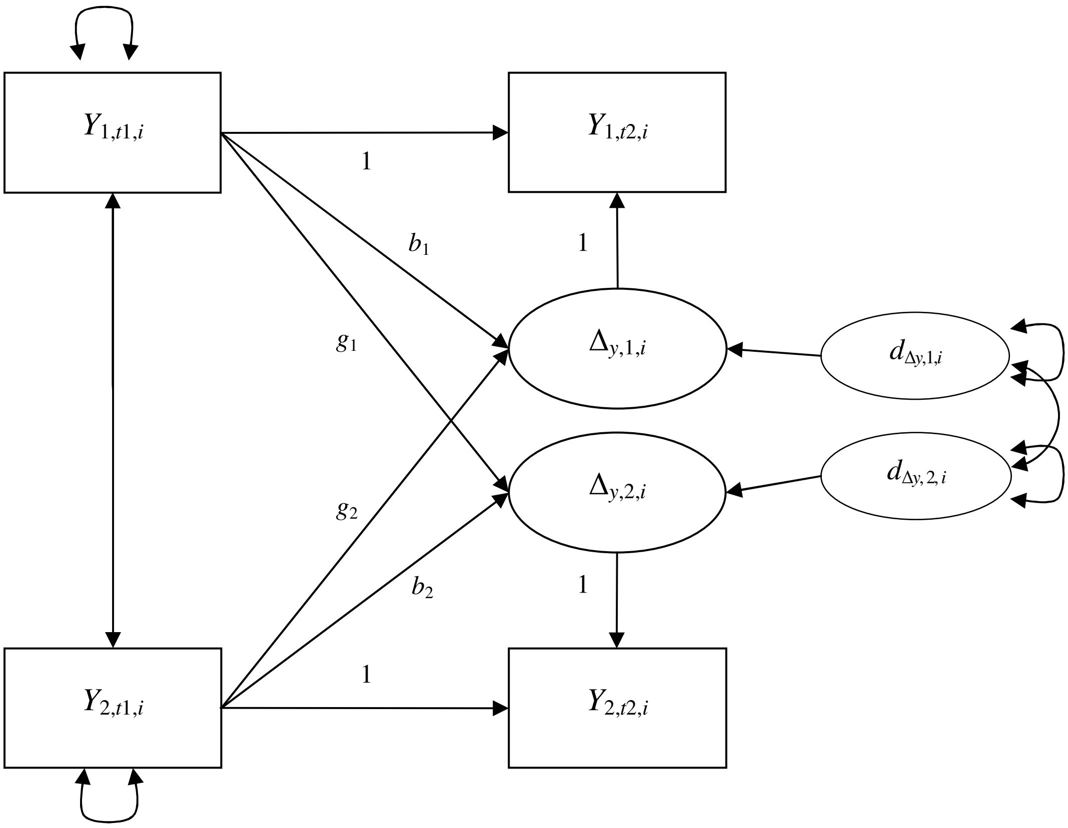



FIGURE 4
Cross-lagged latent difference score model without multiple indicators for two-wave dyadic data.

## ILLUSTRATION

We now illustrate how to fit a series of cross lagged-models to investigate the longitudinal relationships between husbands and wives' marital forgivingness, which is defined as the tendency to forgive the spouse across different offences and occasions. Despite the recent, dramatic growth in research on forgiveness, few studies have investigated reciprocity of forgiveness in close relationships, where partners are likely to repeatedly offend and forgive (or not forgive) each other (for exceptions see Hoyt, Fincham, McCullough, Maio, & Davila, 2005; Paleari, Regalia, & Fincham, 2011). Equity theory suggests that spouses tend to reciprocate partner forgiveness; nonetheless, existing evidence indicates that imbalance between the granting and receiving of marital forgiveness is a more common experience at least over short time periods (Paleari et al., 2011). The present data were therefore collected to examine whether and to what extent partners' forgivingness was related over a long time frame.

Data come from a 10-year follow-up study involving a small sample of married couples. Given the limited number of cases available, we first estimated a series of nested cross-lagged regression models and then we compared them with corresponding cross-lagged LDS models positing within-partner difference scores as the only latent factors. The two types of cross-lagged models allow one to evaluate the degree to which each spouse's marital forgivingness predicts their own as well as their partner's forgivingness over a 10-year period. They also permit one to determine whether these actor and partner effects are significantly different across husbands and wives.

Nonetheless, in addressing the above questions, the two types of cross-lagged models emphasize complementary aspects of within-partner longitudinal relations: cross-lagged regression models evaluate within-person stability in marital forgivingness (stability effects), whereas cross-lagged LDS models evaluate within-person change in marital forgivingness (change or proportional coefficients). Moreover, because within-person changes are explicit parameters of cross-lagged LDS models, these models evaluate the proportion of variance in marital forgivingness change accounted for by the set of paths in the model, whereas cross-lagged regression models evaluate the proportion of variance accounted for in marital forgivingness at $t_2$.

Finally, we estimated a series of nested cross-lagged LDS models with multiple indicators, or cross-lagged common factor LDS models, with the sole intent of providing the reader additional information about the modelling method, as the limited sample size does not allow estimation of reliable parameters in this last case. Cross-lagged common factor LDS models allow one to deal with the same research questions addressed by cross-lagged LDS models positing within-partner difference scores as the only latent factors, but they provide a more accurate evaluation of change over time by partialling out measurement error from the computation of change scores.

## Participants and Procedure

Married couples ($n$ = 122) in Northern Italy were identified through their children's high school to participate in a study on marriage. Sixty one couples agreed to a 10-year follow-up survey (conducted in 2010) in return for a petrol coupon. At $t_1$, spouses had been married an average of 19.42 years ($SD$ = 6.97) and had 2.11 children ($SD$ = 0.73); both husbands and wives were in their mid-forties ($M$ = 46.33 and 44.08, respectively; $SD$ = 6.95 and 6.17, respectively).

Spouses who provided data at $t_2$ represent 50% of the original sample (27% could not be contacted, 15% of couples refused to take part in the follow-up, and 8% were no longer eligible to participate because of divorce or death). Participants who provided data at both $t_1$ and $t_2$ did not differ from those who provided only $t_1$ data in terms of demographics or any of the variables investigated. Nonetheless, the findings reported must be interpreted with caution because of the selective survival typically found in longitudinal studies that span long periods of time.

## Measure

Within the context of a larger longitudinal study on marital forgiveness and well-being, each spouse completed the same marital forgivingness measure at $t_1$ and $t_2$. The measure assessed the tendency to forgive the partner when hurt or wronged by him/her using nine items from the Marital Offence-Specific Forgiveness Scale (MOFS; Paleari, Regalia, & Fincham, 2009). Items were modified so that they referred to marital transgressions in general rather than to a single transgression (e.g., "Although she/he hurt me, I definitely put what happened aside so that we could resume our relationship"; see Paleari et al., 2011) and were rated on a 6-point Likert scale (1 = *very strong disagreement*, 6 = *very strong agreement*).

## Preliminary Analyses: Establishing Factorial Invariance

Before estimating any model, analyses were performed to test the invariance of the forgivingness measure across time and gender. Exploratory factor analyses revealed that all forgivingness items loaded on the same factor except for two items, which loaded on a different factor for husbands at $t_2$ and were dropped. Confirmatory factor analyses (CFAs) were then performed on the remaining seven items to confirm the unidimensional solution for husbands and wives as well as to test for factorial invariance across measurement waves. Establishing factorial invariance is important to assure that the same construct is assessed and that scale scores fall on the same metric across time, so that change can be estimated unambiguously (Widaman, Ferrer, & Conger, 2010). For this purpose, a hierarchy of increasingly stringent tests of factorial invariance was used (Meredith, 1993). Specifically, we first tested configural invariance for husbands and wives by estimating baseline models which only require the number and pattern of factors to be equal across waves. These baseline models were then compared to progressively more constrained models, namely weak factorial invariance models which impose equality constraints for factor loadings, strong factorial invariance models which add equality of manifest variable intercepts, and finally strict factorial invariance models which add equality of manifest variable error term variance. In all these models manifest variable error terms were allowed to correlate across waves (Bijleveld et al., 1998). When there was evidence of one noninvariant measurement parameter in any model, but the remaining parameters were invariant, analyses proceeded in the context of partial measurement invariance. Models representing strong or strict factorial invariance, even though partial, must fit the data better than the less constrained ones in order to identify the same latent construct over time.

As the previous sequential models are all nested, we compared their fit through the likelihood chi-square difference test. Given that our data were nonnormal, incomplete in a few cases, and came from a small sample, we relied on the Bartlett correction, which performs well under these three conditions, in rescaling the Yuan-Bentler chi-square statistics (generally used with nonormal, incomplete data), into the Bartlett corrected Yuan-Bentler chi-square (Y-B$\chi^2_b$) (Savalei, 2010).[1] Bartlett corrected Yuan-Bentler chi-square difference test statistics were then computed following the steps suggested by Satorra (Bryant & Satorra, 2011; Satorra & Bentler, 2011).[2] The comparative fit index (CFI; Bentler, 1990) and the root-mean-square error of approximation (RMSEA; Bentler, 2006) were also adjusted for nonnormality by incorporating the Y-B$\chi^2_b$ into their calculations.

Table 1 presents fit indexes and comparisons of models testing measurement invariance over time. A partial strong invariance model for husbands and the strict factorial invariance model for wives provided better fits for the data than the less constrained models, Y-B$\chi^2_b$ (80) = 101.40, $p$ = .053; CFI$_b$ = .992; RMSEA$_b$ = .066, and Y-B$\chi^2_b$(88) = 84.05, $p$ = .599; CFI$_b$ = 1.000;

RMSEA$_b$ = .000, for husbands and wives respectively. These findings show that a minority of parameters are noninvariant, thereby supporting the assumption that the forgivingness scale captures the same latent construct over time.

TABLE 1
Model fit of various across waves invariance models for husbands and wives

| Model | Y-B$\chi^2_b$ | *df* | $\Delta$Y-B$\chi^2_b$ | CFI$_b$ | RMSEA$_b$ |
|---|---|---|---|---|---|
| | | Husbands | | | |
| Configural invariance | 115.63*** | 69 | – | .997 | .106 |
| Weak factorial invariance | 100.61* | 75 | 3.176 | .999 | .075 |
| Strong factorial invariance | 123.37** | 81 | 13.665* | .983 | .093 |
| Partial strong factorial invariance[a] | 101.40 | 80 | .235 | .992 | .066 |
| | | Wives | | | |
| Configural invariance | 85.22 | 69 | – | .991 | .063 |
| Weak factorial invariance | 86.85 | 75 | 3.376 | .993 | .051 |
| Strong factorial invariance | 93.73 | 81 | 6.833 | .989 | .051 |
| Strict factorial invariance | 84.05 | 88 | 1.540 | 1.000 | .000 |

[a]Intercepts were equal for six items out of seven.
*$p < .05$. **$p < .01$. ***$p < .001$.

Following Kenny et al.'s (2006) recommendations, we also tested the factorial invariance of the forgivingness scale across gender by adopting the same strategy used to establish measurement invariance over time. Fit indexes and comparisons of model testing measurement invariance across gender are reported in Table 2.

TABLE 2
Model fit of various across gender invariance models at $t_1$ and $t_2$

| Model | Y-B$\chi^2_b$ | *df* | $\Delta$Y-B$\chi^2_b$ | CFI$_b$ | RMSEA$_b$ |
|---|---|---|---|---|---|
| | | $t_1$ | | | |
| Configural invariance | 82.80 | 69 | – | .936 | .058 |
| Weak factorial invariance | 78.87 | 75 | 2.086 | .982 | .029 |
| Strong factorial invariance | 87.30 | 81 | 10.033 | .973 | .036 |
| Strict factorial invariance | 86.70 | 88 | 3.800 | 1.000 | .000 |
| | | $t_2$ | | | |
| Configural invariance | 110.60** | 69 | – | .913 | .100 |
| Weak factorial invariance | 104.95* | 75 | 4.565 | .939 | .080 |
| Strong factorial invariance | 114.94** | 81 | 17.005** | .925 | .084 |
| Partial strong factorial invariance[a] | 101.81 | 80 | .847 | .955 | .067 |

[a]Intercepts were equal for six items out of seven.
*$p < .05$. **$p < .01$.

The strict factorial invariance model and a partial strong invariance model were supported for $t_1$ data and $t_2$ data, respectively, Y-B$\chi^2_b(88) = 86.70$, $p = .519$; $CFI_b = 1.000$; $RMSEA_b = .000$, and Y-B$\chi^2_b(80) = 101.81$, $p = .051$; $CFI_b = .955$; $RMSEA_b = .067$, showing that the latent construct captured by the forgivingness scale did not substantially differ across gender. The internal consistency reliability was good for both husbands ($\alpha$ =.93 at $t_1$ and .84 at $t_2$) and wives ($\alpha$ =.92 at $t_1$ and .89 at $t_2$). Consequently, responses to items were averaged to form indexes of marital forgivingness. Descriptive statistics for these indexes are reported in Table 3.

TABLE 3
Means, standard deviations, and correlations

| | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| 1. Husbands forgivingness at $t_1$ | – | | | |
| 2. Wives forgivingness at $t_1$ | .22 | – | | |
| 3. Husbands forgivingness at $t_2$ | .67*** | .39** | – | |
| 4. Wives forgivingness at $t_2$ | .14 | .68*** | .39*** | – |
| Mean (*SD*) | 4.47 (1.10) | 4.19 (1.06) | 3.94 (1.11) | 3.83 (0.95) |

***p* < .01. ****p* < .001.

## Cross-Lagged Regression Analyses

Both cross-lagged regression and latent difference score analyses were conducted using a SEM approach because it offers the simplest and most direct way to analyze data for distinguishable dyads like married couples. The EQS program was used for all SEM analyses (Bentler, 2006).

When using SEM to estimate cross-lagged relations in dyadic data, one should center each variable by using a common meaningful value computed across partners to have parameters comparable across dyad members and waves. Accordingly, we first centred forgivingness scores by using the $t_1$ grand mean computed across the entire sample ($M = 4.33$).

We then estimated a series of nested, cross-lagged models in order to explore the relationships between husbands and wives' forgivingness over time. Specifically, we first estimated a highly constrained regression model ($RM_{HC}$) in which (a) means for all the variables of interest were set to be equal between husbands and wives, (b) variances for all the variables were set to be equal between husbands and wives, (c) intrapersonal covariances or stability effects ($s_1$, $s_2$) for husbands and wives were constrained to be equal, and (d) interpersonal covariances or partner effects ($p_1$, $p_2$) for husbands and wives were constrained to be equal. This comprises the Omnibus Test of Distinguishability (Kenny et al., 2006), because it tests whether dyad members are empirically distinguishable. Conceptually, partners in heterosexual relationships are distinguishable on the basis of sex but conceptually distinguishable partners may not be empirically distinguishable. This is the case if the highly constrained model just described fits the data better than less constrained models; if so then one could, in principle, treat the individual rather than the couple as the unit of analysis. Next, we estimated a less constrained model ($RM_{LC}$) in which we released those constraints that, on the basis of the Lagrange Multiplier test (Bentler, 2006), were shown to have been improperly imposed. This model tests the hypothesis that the two partners differ on some parameter estimates. Because our main purpose was to evaluate differences in the longitudinal paths within and across partners, releasing constraints on stability and partner effects could be particularly informative. Finally, the previous less constrained model was compared with alternative ones in which partner effects were alternatively or simultaneously set equal to zero ($RM_{LC\text{-NO HUSBAND PARTNER EFFECT}}$; $RM_{LC\text{-NO WIFE PARTNER EFFECT}}$; $RM_{LC\text{-NO HUSBAND AND WIFE PARTNER EFFECT}}$). These last models test the hypotheses of unidirectional or no longitudinal paths across partners, respectively. Because all the models were nested they were compared using the likelihood chi-square difference test. The sample size requirements are no different than for ordinary regression analysis (Kenny & Cook, 1999), and were fully satisfied in the present case. The highly constrained model (RMHC) displayed a good fit (see Table 4), showing that spouses were empirically indistinguishable with respect to forgivingness when assessed over time.

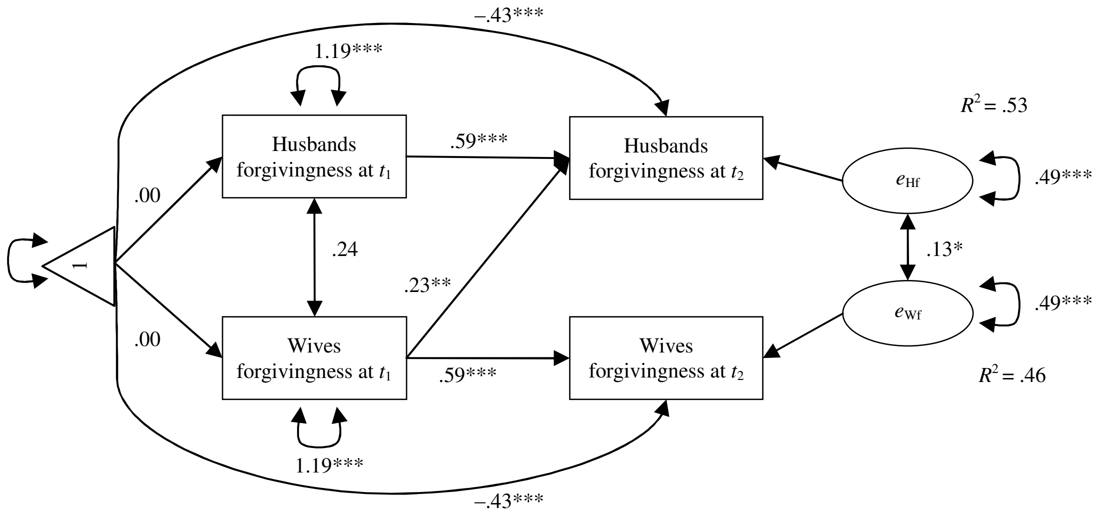
TABLE 4
Model fit of nested cross-lagged regression models or cross-lagged LDS models without multiple indicators

| Comparison model | $\chi^2(df)$ | $p$ | Baseline model | $\Delta\chi^2$ | $p$ |
|---|---|---|---|---|---|
| $RM_{HC}$ or $LDSM_{HC}$ | $\chi^2(6) = 7.326$ | .292 | – | – | – |
| $RM_{LC}$ or $LDSM_{LC}$ | $\chi^2(5) = 2.913$ | .713 | $RM_{HC}$ or $LDSM_{HC}$ | $\chi^2(1) = 4.413$ | .036 |
| $RM_{LC\text{-NO HUSBAND PARTNER EFFECT}}$ or $LDSM_{LC\text{-NO HUSBAND PARTNER EFFECT}}$ | $\chi^2(6) = 2.915$ | .814 | $RM_{LC}$ or $LDSM_{LC}$ | $\chi^2(1) = -0.002$ | *ns* |
| $RM_{LC\text{-NO WIFE PARTNER EFFECT}}$ or $LDSM_{LC\text{-NO WIFE PARTNER EFFECT}}$ | $\chi^2(6) = 10.721$ | .097 | $RM_{LC}$ or $LDSM_{LC}$ | $\chi^2(1) = -7.804$ | .005 |
| $RM_{LC\text{-NO HUSBAND AND WIFE PARTNER EFFECT}}$ or $LDSM_{LC\text{-NO HUSBAND AND WIFE PARTNER EFFECT}}$ | $\chi^2(7) = 10.751$ | .150 | $RM_{LC}$ or $LDSM_{LC}$ | $\chi^2(2) = -7.834$ | .020 |

*Note*. Fit indices for cross-lagged LDS models without multiple indicators are equal to those of cross-lagged regression models. When the *df* are small, as in present cases, both the CFI and the RMSEA can be misleading (the CFI can be large and the RMSEA very small despite a good fit). For this reason fit measures were not reported (see Kenny et al., 2006).

Inspection of the Lagrange Multipliers test however revealed that the partner effect constraint was improperly imposed and should be released. As expected, the model with partner effects freely estimated ($RM_{LC}$) fitted the data significantly better than the previous one. It also had a better fit than models estimating no wife to husband partner effect ($RM_{LC\text{-NO WIFE PARTNER EFFECT}}$) or no partner effects at all ($RM_{LC\text{-NO HUSBAND AND WIFE PARTNER EFFECT}}$). Conversely, the model with partner effects freely estimated ($RM_{LC}$) fitted the data equally well compared to the one positing no husband partner effect ($RM_{LC\text{-NO HUSBAND PARTNER EFFECT}}$). As it is more parsimonious, parameters for the last model that assumed no husband to wife partner effect were computed (see Figure 5; unstandardized parameters are reported, because SEM standardizes parameters separately for each member type and therefore yields coefficients that are not comparable across husbands and wives).

Overall the model explained 53% and 46% of variance in husbands and wives' forgivingness at $t_2$, respectively. For husbands and wives *stability* effects were equally large, positive, and statistically significant (.59), indicating that marital forgivingness was quite stable over time.




*$p$ < .05. **$p$ < .01. ***$p$ < .001.

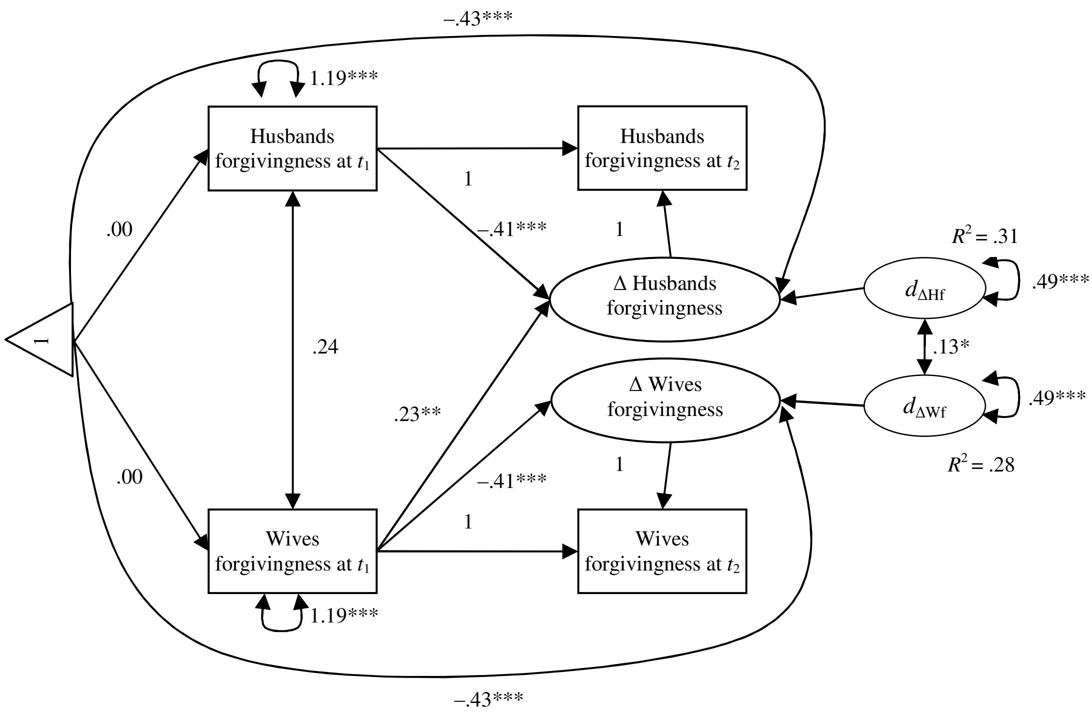
FIGURE 5
Unstandardized parameter estimates for the cross-lagged regression model positing no husband to wife partner effect ($RM_{\text{LC-NO HUSBAND PARTNER EFFECT}}$).

The wife to husband partner effect was positive and statistically significant (.23), whereas there was no significant husband to wife partner effect. Thus, support for a unidirectional influence model was found: wives' forgivingness at baseline predicted an increase in husbands' forgivingness over a 10-year period, whereas husbands' forgivingness at baseline was unrelated to wives' changes in forgivingness over the same period. Specifically, for each unit increase in the wife's forgivingness, the husband's forgivingness increased by .23 units 10 years later, whereas a unit increase in the husband's forgivingness was unrelated to the wife's change in forgivingness 10 years later. The model also indicates that husbands' forgivingness did not significantly covary with wives' forgivingness at $t_1$ (.24), but husbands' and wives' forgivingness were significantly related at $t_2$ (.13), after controlling for stability and partner effects. Finally, the model reveals that spouses' average forgivingness was significantly lower at $t_2$ (–.43) than at $t_1$ (.00) and it did not vary by sex at either time point.

## Cross-Lagged Latent Difference Score Analyses without Multiple Indicators

Cross-lagged latent difference scores were modelled using the same strategy outlined above. That is, we started by fitting a highly constrained model in which means, variances, and both intrapersonal and interpersonal covariances were forced to be equal across partners ($LDSM_{HC}$) and then compared this model to a less constrained one ($LDSM_{LC}$) in which we released those constraints that were shown to have been improperly imposed. Finally, the previous, less constrained model was compared with alternative ones in which partner effects were alternatively or simultaneously set equal to zero ($LDSM_{\text{LC-NO HUSBAND PARTNER EFFECT}}$; $LDSM_{\text{LC-NO WIFE PARTNER EFFECT}}$; $LDSM_{\text{LC-NO HUSBAND AND WIFE PARTNER EFFECT}}$). The sample size requirements of at least five cases for every estimated parameter (Bentler & Chou, 1987) are met for these models; in fact the ratio of sample size to estimated parameters varied between 7.63 : 1 and 8.71 : 1 for the LDSM$_{LC}$ and the LDSM$_{\text{LC-NO HUSBAND AND WIFE PARTNER EFFECT}}$, respectively.

As they are reinterpretations of cross-lagged regression models, all these LDS models had the same fit as the cross-lagged regression models presented earlier. Thus, the LDSM$_{\text{LC-NO HUSBAND PARTNER EFFECT}}$ was again the best fitting model (see Table 4). Its parameters are the same as the corresponding cross-lagged regression model with the exception of regression parameters of *change* in one partner's forgivingness on the same partner's baseline forgivingness (see Figure 6).




*$p$ < .05. **$p$ < .01. ***$p$ < .001.

FIGURE 6
Unstandardized parameter estimates for the cross-lagged latent difference score model positing no husband to wife partner effect (LDSM$_{\text{LC-NO HUSBAND PARTNER EFFECT}}$) and having no multiple indicators.

These parameters, sometimes called proportional coefficients, indicate the linear and proportional within-person change in forgivingness among husbands and wives: for a unit increase in spouses' forgivingness, their within-person change in forgivingness decreased by –.41 units over a 10-year period, net of any controls specified in the model. Thus, marital forgivingness decreased in both husbands and wives over time, even though this decline is less pronounced among men married to wives who were more prone to forgive at baseline (.23). Also, changes in forgivingness were on average negative and significantly different from zero (–.43), were significantly associated (.13) and did not differ across husbands and wives, as indicated by the previous cross-lagged regression model. Overall, the model explained 31% and 28% of variance in husbands and wives' forgivingness change, respectively.

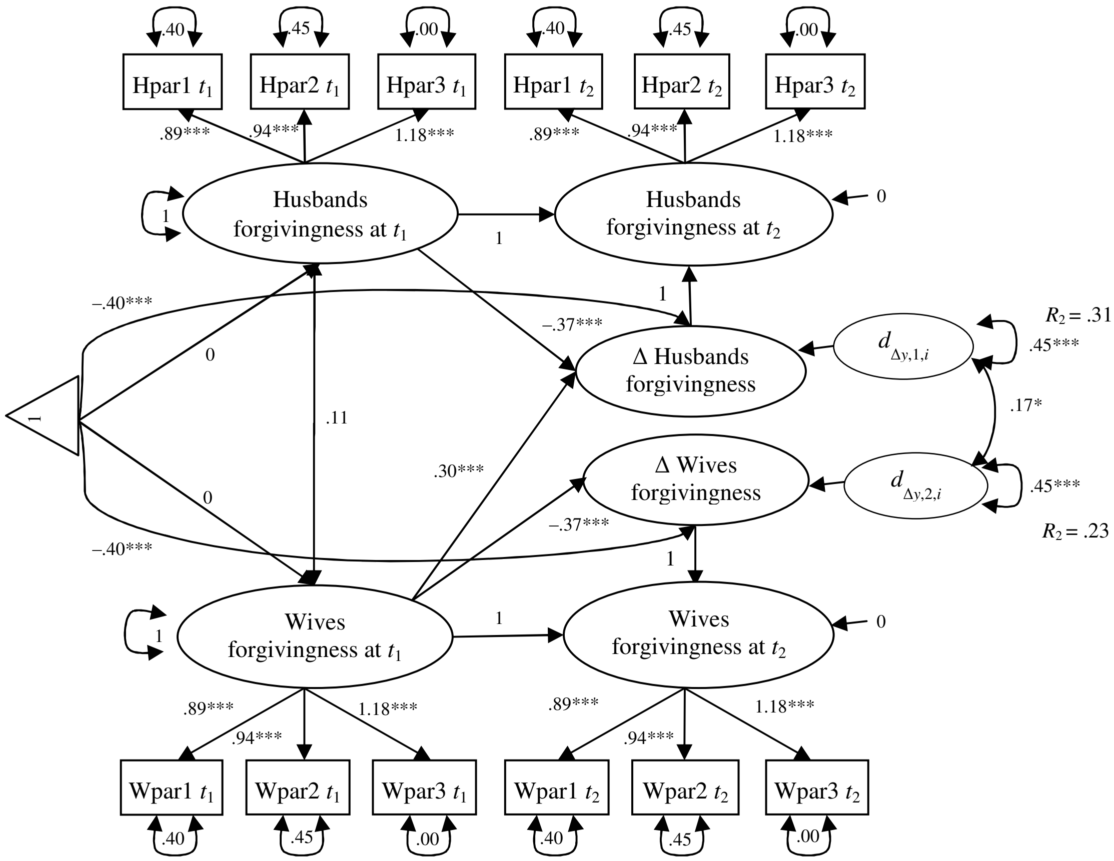
Cross-Lagged Latent Difference Score Analyses with Multiple Indicators

Finally we re-estimated the same cross-lagged latent difference score model assuming no husband partner effect ($LDSM_{\text{LC-NO HUSBAND PARTNER EFFECT}}$) with multiple indicators, thereby modelling a cross-lagged common factor latent difference score model, so as to take into account possible measurement errors in difference scores. In this model, common factors for forgivingness were estimated by use of parcel indicators as measured variables to ensure a parsimonious model (for a discussion of the pros and cons of the practice of parceling see Little, Cunningham, & Shahar, 2002); the dataset was therefore different from the one used in previous models. Parcels can be defined as aggregate-level indicators comprised of the average (or sum) of two or more items, responses or behaviors. In the present case, three parcel indicators were generated for each factor by averaging the two items having the highest and the lowest item-to-construct loadings and then assigning the resulting six items randomly to parcels, each computed as a two-item mean score. Parcels were kept consistent across the two partners and the two occasions. The model was computed under constraints of strict measurement invariance across waves and partners, which caused only slight and insignificant reductions of overall model fit. This assured that factor scores had the same unit of measurement across waves and partner. Also, means and variances were fixed to 0 and 1, respectively, in the level factors (factors which do not represent change) in order to identify the model. Hence, means and variances in the change factors can be directly interpreted as differences from the level estimates. The model showed an acceptable fit, $\chi^2(63) = 81.975$, $p = .054$; CFI = .971; RMSEA = .077. However, it must be noted that sample size requirements are far from satisfactory; the ratio of sample size to estimated parameters was only 3.39 : 1 in the present case. For this reason, findings were presented for heuristic purposes and must be considered with caution.

When compared to the corresponding cross-lagged LDS model without multiple indicators, this model explained a somewhat lower proportion of change variance for wives (23%), but it identified parameters for structural relations quite similar to those of the previous model with no common factors. This result is consistent with the hypothesis that cross-lagged LDS models can be an appropriate strategy to detect change over time even for sample sizes where it is not possible to use multiple indicators. Specifically, controlling for measurement errors leads to only a marginal decrease in model parameters, with the exception of the covariance between changes and the wife to husband partner effect which became slightly stronger (see Figure 7). As systematic measurement error that is stable over time manifests in autoregressive models in the form of upwardly biased stability estimates for the concept of interest, the decrease in stability parameters was somewhat expected.

Overall the model confirms that, even when latent change is estimated through common factors partialling out measurement errors, marital forgivingness significantly and equally decreased within husbands and wives over a 10-year period; this decline was reduced when husbands (but not wives) were married to a more forgiving partner at baseline. When compared to alternative models in which the wife or both spouses partner effects were set equal to zero ($LDSM_{\text{LC-NO WIFE PARTNER EFFECT}}$; $LDSM_{\text{LC-NO HUSBAND AND WIFE PARTNER EFFECT}}$), the $LDSM_{\text{LC-NO HUSBAND PARTNER EFFECT}}$ yielded the best fitting model also in the present case (see Table 5).

The presence of only one partner effect from wives to husbands suggests that husbands were dependent on their wife with respect to forgivingness: husbands became less unforgiving as long as their partner was more disposed to forgive them 10 years earlier. Conversely, even though wives' decline in forgivingness was similar to that of husbands, it was not significantly affected by their partners' previous tendency to forgive.




*Note*. Means and variances of forgivingness at $t_1$ were fixed to 0 and 1, respectively, and disturbances of forgivingness at $t_2$ were fixed at zero to identify model parameters as well as force the two-wave decomposition. Forgivingness was measured with multiple indicators that were correlated and invariant both over time and across partner (correlations between indicators and indicators intercepts are not reported for clarity sake).

FIGURE 7
Unstandardized parameter estimates for the cross-lagged latent difference score model positing no husband to wife partner effect (LDSM$_{\text{LC-NO HUSBAND PARTNER EFFECT}}$) and having multiple indicators.

## CONCLUSION

This article has presented the use of cross-lagged LDS models to conceptualize and measure change in two-wave dyadic data. In contrast to cross-lagged regression models, cross-lagged LDS models permit us to estimate and describe within-person changes as well as their covariation across partners. Specifically, they assume that each partner score at follow-up is composed of the same partner score on the same variable at baseline plus an unobserved change score; they also

TABLE 5
Model fit of nested cross-lagged LDS models with multiple indicators

| Comparison model | $\chi^2(df)$ | $p$ | CFI | RMSEA | Baseline model | $\Delta\chi^2$ | $p$ |
|---|---|---|---|---|---|---|---|
| $RM_{HC}$ or $LDSM_{HC}$ | $\chi^2(63) = 88.122$ | .020 | .961 | .089 | – | – | – |
| $RM_{LC}$ or $LDSM_{LC}$ | $\chi^2(62) = 81.973$ | .046 | .969 | .080 | $RM_{HC}$ or $LDSM_{HC}$ | $\chi^2(1) = 6.149$ | .013 |
| $RM_{\text{LC-NO HUSBAND PARTNER EFFECT}}$ or $LDSM_{\text{LC-NO HUSBAND PARTNER EFFECT}}$ | $\chi^2(63) = 81.975$ | .054 | .971 | .077 | $RM_{LC}$ or $LDSM_{LC}$ | $\chi^2(1) = –0.002$ | *ns* |
| $RM_{\text{LC-NO WIFE PARTNER EFFECT}}$ or $LDSM_{\text{LC-NO WIFE PARTNER EFFECT}}$ | $\chi^2(63) = 93.058$ | .008 | .953 | .097 | $RM_{LC}$ or $LDSM_{LC}$ | $\chi^2(1) = –11.085$ | .000 |
| $RM_{\text{LC-NO HUSBAND AND WIFE PARTNER EFFECT}}$ or LDS $M_{\text{LC-NO HUSBAND AND WIFE PARTNER EFFECT}}$ | $\chi^2(64) = 93.077$ | .010 | .955 | .095 | $RM_{LC}$ or $LDSM_{LC}$ | $\chi^2(2) = –11.104$ | .004 |

posit that each partner change score is a dependent variable in a simultaneous equation with intercept, lagged autoregression from the same partner baseline score and cross-lagged regression from the other partner baseline score. Thus, these models suggest that some part of the change in each partner score may be due to both partners' prior starting point. Because of the time-order relationships among the variables, cross-lagged effects can be properly interpreted as predictive paths, but cannot be considered causal paths. In fact some other unmeasured variable may be the causal mechanism driving the observed pattern of influences as well as $t_1$ constructs, which are assumed to be exogenous, but may not represent the true beginning of the time ordered sequence of relationships among the constructs (Little et al., 2006).

Our illustration shows that cross-lagged LDS models can be a suitable means to analyze within-partner change over time even when it is not feasible to estimate common factors owing to sample size. In this case, however, it is advisable to adopt the following two procedures in order to have parameters that are comparable across both waves and dyad members. First, the reliability and invariance of measures across time and gender must be established, and, second, variables must be centered by using a common meaningful value computed across dyad members.

Although cross-lagged LDS models with multiple indicators have sometime been used to analyze change in two-wave dyadic data (e.g., Cong & Silverstein, 2011; Lindwall, Larsman, & Hagger, 2011; Magee, Miller, & Heaven, 2013; Schilling, Wahl, & Wiegering, 2013), we hope to have shown that they are equally applicable to situations in which small sample sizes require them to be estimated without multiple indicators. Of course, cross-lagged common factor LDS models with multiple indicators (sample size permitting) are preferable because they partial out measurement error from the computation of change scores. We must however note that, whenever data from more than two time points are available, it is possible to estimate independently true change score and its error variance using models that do not imply common factors. These models are known as dual change models because, beside autoregressive effects, they take into consideration another internal source of change, not considered by cross-lagged LDS model, that is nonstationarity or natural change (for more details, see King et al., 2006).

Notwithstanding these alternatives, cross-lagged LDS models are appropriate methods of estimating change in the context of two-wave nonindependent data like that provided by married couples. Finally, it is worth noting that cross-lagged LDS models complement rather than compete with other better known methods of analysis, including cross-lagged regression and growth curve models.

## NOTES

1. The Bartlett corrected Yuan-Bentler chi-square was computed as follows:
$$\mathrm{Y-B}\chi_b^2 = \left\{1 - \frac{2p+4k+5}{6(n-1)}\right\}\mathrm{Y-B}\chi^2,$$
where $k$ is the number of latent factors in the model. The adjustment depends on sample size, and disappears asymptotically.
2. The difference in Bartlett corrected Yuan-Bentler chi-square for nested models does not correspond to a chi-square distribution. For this reason, it is not possible to directly compare Y-B$\chi^2_b$ of nested models by subtracting the Y-B$\chi^2_b$ value for the less restrictive baseline model ($M_1$) from the Y-B$\chi^2_b$ value for more restrictive comparison model ($M_0$), as researchers do for traditional chi-square difference tests. To overcome a similar difficulty related to the Satorra-Benteler scaled difference Bryant and Satorra (2011) suggest going through five steps which can also applied to the Bartlett corrected Yuan-Bentler chi-

square difference test. First, because Y-B$\chi^2_b$ is a standard goodness-of-fit chi-square value divided by a scaling correction factor, to recover the scaling correction factor ($c$) for a model, for use in scaled difference test, the standard goodness-of-fit chi-square value must be divided by the corrected chi-square value for that model (Y-B$\chi^2_b$ in the present case). Thus, scaling correction factors for models $M_0$ and $M_1$ can be obtained by dividing $M_0$ and $M_1$ standard chi-square by the Y-B$\chi^2_b$ value for each model. Second, the scaling correcting factor for each model must be multiplied by the model's *df*. Third, this product for $M_1$ model must be subtracted from the same product for $M_0$ model. Fourth, the result must be divided by $m$ (i.e., the difference in *df* between $M_0$ and $M_1$) to obtain the scaling factor for the scaled difference test ($c_d$). Finally, the difference in standard chi-square values of models $M_0$ and $M_1$ must be divided by the scaling factor ($c_d$), with *df* for the scaled difference test $m = df$ for model $M_0$-*df* for model $M_1$. In the event that the Bartlett corrected Yuan-Bentler chi-square difference test is negative, Bryant and Satorra (2011) suggest replacing $M_1$ with $M_{10}$ in the previous computations, where $M_{10}$ is the baseline model with number of interactions fixed at zero and the final parameter estimates for $M_0$ as starting values. Thus, $c$ for $M_{10}$ is then computed by dividing the standard chi-square value for $M_{10}$ by Y-B$\chi^2_b$ for $M_{10}$ and $c$ for $M_{10}$ is used in place of $c$ for $M_1$ to compute the correction factor for $c_d$ (for more details see Satorra and Bentler, 2011).

## REFERENCES


Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, *107*, 238-246. doi:10.1037/0033-2909.107.2.238

Bentler, P. M. (2006). *EQS6 structural equations program manual*. Encino, CA: Multivariate Software.

Bentler, P. M., & Chou, C. P. (1987). Practical issues in structural modeling. *Sociological Methods & Research*, *16*, 78-117. doi:10.1177/0049124187016001004

Berrington, A., Smith, P. W. F., & Sturgis, P. (2006). *An overview of methods for the analysis of panel data*. ESRC National Centre for Research Methods. Retrieved from http://eprints.ncrm.ac.uk/415/1/MethodsReviewPaperNCRM-007.pdf

Bijleveld, C. C. J. H., van der Kamp, L. J. T., Mooijaart, A., Van Der Kloot, W., van der Leeden, R., & dan der Burg, E. (1998). *Longitudinal data analysis. Designs, models, and methods*. Thousand Oaks, CA: Sage Publications Ltd.

Bryant, F. B., & Satorra, A. (2011). Principles and practice of scaled difference chi-square testing. *Structural Equation Modeling*, *19*, 2012. doi:0.1080/10705511.2012.687671

Chan, D. (1998). The conceptualization and analysis of change over time: An integrative approach incorporating longitudinal means and covariance structures analysis (LMACS) and multiple indicator latent growth modeling (MLGM). *Organizational Research Methods*, *1*, 421-483. doi: 10.1177/1059601102250814

Collins, L. M. (1996). Is reliability obsolete? A commentary on "Are simple gain scores obsolete?" *Applied Psychological Measurement*, *20*, 289-292. doi:10.1177/014662169602000308

Cong, Z., & Silverstein, M. (2011). Parents' depressive symptoms and support from sons and daughters in rural China. *International Journal of Social Welfare*, *20*, S4-S17. doi: 10.1111/j.1468-2397.2011.00823.x

Gerstorf, D., Röcke, C., & Lachman, M. E. (2010). Antecedent-consequent relations of perceived control to health and social support: Longitudinal evidence for between-domain associations across adulthood. *Journal of Gerontology: Psychological Sciences*, *66B*, 61-71, doi:10.1093/geronb/gbq077.

Hoyt, W. T., Fincham, F., McCullough, M. E., Maio, G., & Davila, J. (2005). Responses to interpersonal transgressions in families: Forgivingness, forgivability, and relationship-specific effects. *Journal of Personality and Social Psychology*, *89*, 375-394. doi:10.1037/0022-3514.89.3.375

Karney, B. R., & Bradbury, T. N. (1995). Assessing longitudinal change in marriage: An introduction to analysis of growth curves. *Journal of Marriage and the Family*, *57*, 1091-1108. doi:10.2307/353425

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

Kenny, D. A., & Cook, W. L. (1999). Partner effects in relationship research: Conceptual issues, analytic difficulties, and illustrations. *Personal Relationships*, *6*, 433-448. doi:10.1111/j.1475-6811.1999.tb00202.x

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

King, L. A., King, D. W., McArdle, J. J., Saxe, G. N., Doron-LaMarca, S., & Orazem, R. J. (2006). Latent difference score approach to longitudinal trauma research. *Journal of Traumatic Stress*, *19*, 771-785. doi:10.1080/10705511.2012.713275

Lindwall, M., Larsman, P., & Hagger, M. S. (2011). The reciprocal relationship between physical activity and depression in older European adults: A prospective cross-lagged panel design using SHARE data. *Health Psychology*, *30*, 453-462. doi:10.1037/a0023268

Little, T. D., Bovairds, J. A., & Slegers, D. W. (2006). Methods for the analysis of change. In D. Mroczek & T. D. Little (Eds.), *Handbook of personality development* (pp. 181-212). Mahwah, NJ: Erlbaum.

Little, T. D., Cunningham, W. A., & Shahar, G. (2002). To parcel or not to parcel: Exploring the question, weighing the merits. *Structural Equation Modeling*, *9*, 151-173. doi:10.1207/S15328007SEM0902_1

Magee, C. A., Miller, L. M., & Heaven, P. C. L. (2013). Personality trait change and life satisfaction in adults: The roles of age and hedonic balance. *Personality and Individual Differences*, *55*, 694-698. doi:10.1016/j.paid.2013.05.022

McArdle, J. J. (2001). A latent difference score approach to longitudinal dynamic structural analyses. In R. Cudeck, S. DuToit, & D. Sorbom (Eds.), *Structural equation modeling: Present and future* (pp. 342-380). Lincolnwood, IL: Scientific Software International.

McArdle, J. J. (2009). Latent variable modeling of differences and changes with longitudinal data. *Annual Review of Psychology*, *60*, 577-605. doi: 10.1146/annurev.psych.60.110707.163612.

McArdle, J. J., & Prindle, J. J. (2008). A latent change score analysis of a randomized clinical trial in reasoning training. *Psychology and Aging*, *23*, 702-719. doi:10.1037/a0014349

Meredith, W. M. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, 58, 525-543. doi:10.1007/BF02294825

Nesselroade, J. R., & Baltes, P. B. (Eds.). (1979). *Longitudinal research in the study of behavior and development*. New York, NY: Academic Press.

Nesselroade, J. R., & Cable, D. G. (1974). "Sometimes it's okay to factor difference scores" — The separation of state and trait anxiety. *Multivariate Behavioral Research*, *9*, 273-284. doi:10.1207/s15327906mbr0903_3

Paleari, F. G., Regalia, C., & Fincham, F. D. (2009). Measuring offence-specific forgiveness in marriage: The marital offence-specific forgiveness scale (MOFS). *Psychological Assessment*, *21*, 194-209. doi: 10.1037/a0016068

Paleari, F. G., Regalia, C., & Fincham, F. D. (2011). Inequity in forgiveness: Implications for personal and relational well-being. *Journal of Social and Clinical Psychology*, *30*, 297-324. doi:10.1521/jscp.2011.30.3.297

Rogosa, D. R., & Willett, J. B. (1983). Demonstrating the reliability of the difference score in the measurement of change. *Journal of Educational Measurement*, *20*, 335-343. doi:10.1111/j.1745-3984.1983.tb00211.x

Satorra, A., & Bentler, P. M. (2011). *Ensuring positiveness of the scaled difference chi-square test statistic*. Retrieved from UCLA Los Angeles, Department of Statistics website: https://escholarship.org/uc/item/76f0d1qz

Savalei, V. (2010). Small sample statistics for incomplete nonnormal data: Extensions of complete data formulae and a Monte Carlo comparison. *Structural Equation Modeling*, *17*, 241-264. doi:10.1080/10705511003659375

Schilling, O. K., Wahl, H.-W., & Wiegering, S. (2013). Affective development in advanced old age: Analyses of terminal change in positive and negative affect. *Developmental Psychology*, *49*, 1011-1020. doi:10.1037/a0035500

Widaman, K. F., Ferrer, E., & Conger, R. D. (2010). Factorial invariance within longitudinal structural equation models: Measuring the same construct across time. *Child Development Perspectives*, *4*, 10-18. doi:10.1111/j.1750-8606.2009.00110.x

Williams, R. H., & Zimmerman, D. W. (1996). Are simple gain scores obsolete? *Applied Psychological Measurement*, *20*, 59-69. doi:10.1177/014662169602000106

Williams, R. H., & Zimmerman, D. W. (1977). The reliability of difference scores when errors are correlated. *Educational and Psychological Measurement*, *37*, 679-689. doi:10.1177/001316447703700310