

Bias, Type I Error Rates, and Statistical Power of a Latent Mediation Model in the Presence of Violations of Invariance

Educational and Psychological
Measurement

2018, Vol. 78(3) 460–481

© The Author(s) 2017

Reprints and permissions:

sagepub.com/journalsPermissions.nav

DOI: 10.1177/0013164416684169

journals.sagepub.com/home/epm



**Margarita Olivera-Aguilar¹, Samuel H. Rikoon²,
Oscar Gonzalez³, Yasemin Kisbu-Sakarya⁴, and
David P. MacKinnon³**

Abstract

When testing a statistical mediation model, it is assumed that factorial measurement invariance holds for the mediating construct across levels of the independent variable X . The consequences of failing to address the violations of measurement invariance in mediation models are largely unknown. The purpose of the present study was to systematically examine the impact of mediator noninvariance on the Type I error rates, statistical power, and relative bias in parameter estimates of the mediated effect in the single mediator model. The results of a large simulation study indicated that, in general, the mediated effect was robust to violations of invariance in loadings. In contrast, most conditions with violations of intercept invariance exhibited severely positively biased mediated effects, Type I error rates above acceptable levels, and statistical power larger than in the invariant conditions. The implications of these results are discussed and recommendations are offered.

Keywords

measurement invariance, statistical mediation model, latent variables

¹The National Center of Evaluation of Higher Education, Mexico City, Mexico

²Educational Testing Service, Princeton, NJ, USA

³Arizona State University, Tempe, AZ, USA

⁴Koc University, Istanbul, Turkey

Corresponding Author:

Margarita Olivera-Aguilar, The National Center of Evaluation of Higher Education, Mexico City, Mexico.

Email: margarita.olag@gmail.com

Statistical mediation analysis studies the hypothesized effect of an independent variable (X) on a mediator variable (M), which in turn causes the observed values in a dependent variable (Y ; Baron & Kenny, 1986; Judd & Kenny, 1981; MacKinnon, 2008). Statistical mediation is particularly relevant in treatment and prevention science where it is used to evaluate if interventions changed a hypothesized mediator that is thought to be related to an outcome of interest (MacKinnon, 2008; MacKinnon & Dwyer, 1993). In randomized control trials, the independent variable X in the mediation model represents the groups being compared (i.e., treatment vs. control). For example, Gregg, Callaghan, Hayes, and Glenn-Lawson (2007) compared a group of Type 2 diabetes patients randomly assigned to receive a one-day education workshop with a group of patients receiving a combination of the education workshop and acceptance and commitment therapy. The researchers found that the changes in diabetes-related acceptance and self-management behavior mediated the impact of the treatment on changes in blood glucose. In a different randomized control trial comparing the effectiveness of an acceptance and commitment therapy intervention and a control group for increasing clinicians' willingness to use pharmacotherapy, Varra, Hayes, Roget, and Fisher (2008) found that the changes in psychological flexibility and the believability of barriers to using empirically supported treatments mediated the effect of the treatment on the targeted outcome. In another example, Stice, Presnell, Gau, and Shaw (2007) investigated two eating disorder prevention programs and found that a dissonance-based intervention partially affected several outcome measures such as body dissatisfaction, dieting, negative affect, and bulimic symptoms through reducing levels of a mediator (*thin-ideal internalization*).

Among the many assumptions needed to accurately test for statistical mediation (MacKinnon, 2008), the single mediator model assumes that the variables in the model are reliable and valid representations of the measured constructs. Therefore, as in the above examples, the statistical mediation model assumes that the same mediating construct is being measured across the treatment groups. That is, it is assumed that measurement invariance in the mediator holds across the treatment and control groups. Even though the importance of assessing measurement invariance has long been recognized in psychological testing (Chen, 2008; Cheung & Rensvold, 1999; Millsap, 2011; Vandenberg & Lance, 2000), few studies have examined the consequences of violations of measurement invariance in the estimation of the mediated effect. Therefore, the overarching goal of the current study is to provide information to researchers of the extent to which a lack of measurement invariance influences the detection and accurate estimation of the mediated effect. To achieve our goal, we first define statistical mediation and measurement invariance. Next, we describe previous research investigating the influence of measurement invariance on the mediated effect. Finally, we present results from a Monte Carlo simulation study in which the magnitude and the number of items exhibiting violations of measurement invariance are manipulated to examine their influence on the relative bias, statistical power, and Type I error rates of the mediated effect.

Statistical Mediation Analysis

Statistical mediation analysis investigates *how* two variables are related by introducing intermediate variables (known as mediators) to explain the relationship between an independent variable and an outcome (MacKinnon, 2008). The single mediator model can be described by the following three regression equations:

$$Y = i_1 + \nu X + e_1 \quad (1)$$

$$M = i_2 + \alpha X + e_2 \quad (2)$$

$$Y = i_3 + \nu' X + \beta M + e_3 \quad (3)$$

In the mediation model, the ν path represents the relationship between the independent variable X and the dependent variable Y . The α path represents the relationship from X to the mediator M ; the β path represents the relationship from M to Y , controlling for X ; and the ν' path represents the relationship from X to Y after controlling for M . The i_1 , i_2 , i_3 , and the e_1 , e_2 , e_3 are the intercepts and residual variances of the regression equations, respectively. Typically, the mediated effect is then computed as the product of the α and β paths (i.e., $\alpha\beta$), although there are alternative approaches to quantify the mediated effect (i.e., $\nu - \nu'$; MacKinnon, Lockwood, Hoffman, West, & Sheets, 2002). Furthermore, the significance of the mediated effect is computed by dividing the point estimate of the mediated effect, $\alpha\beta$, by its standard error calculated with the multivariate delta method and then comparing this ratio to the normal distribution (Sobel, 1982; 1986). Most structural equation modeling packages estimate the standard error based on the multivariate delta method by default. However, several other methods such as the distribution of the product (MacKinnon, Fritz, Williams, & Lockwood, 2007; Tofighi & MacKinnon, 2011; Valente, Gonzalez, Miočević, & MacKinnon, 2016), or resampling techniques (MacKinnon, Lockwood, & Williams, 2004) have been recommended for conducting significance testing of the mediated effect since $\alpha\beta$ may not be normally distributed (Kisbu-Sakarya, MacKinnon, & Miočević, 2014; MacKinnon et al., 2002).

As previously mentioned, the statistical mediation model assumes that there is a reliable and valid representation of the X , M , and Y variables in the model. However, the psychometric properties of the variables in the mediation model are often ignored (Gonzalez & MacKinnon, 2016; MacKinnon, 2008), and those can be studied with latent variable models. Our model of interest is presented in Figure 1, where the X variable represents a binary independent variable in which two groups are studied (e.g., treatment and control groups), M is a continuous, latent mediator variable measured with six indicators, and Y is a continuous, latent outcome variable measured with six indicators.

Measurement Invariance

An important psychometric prerequisite for using instruments to compare groups on a latent variable of interest is that the measurement properties of observed variables

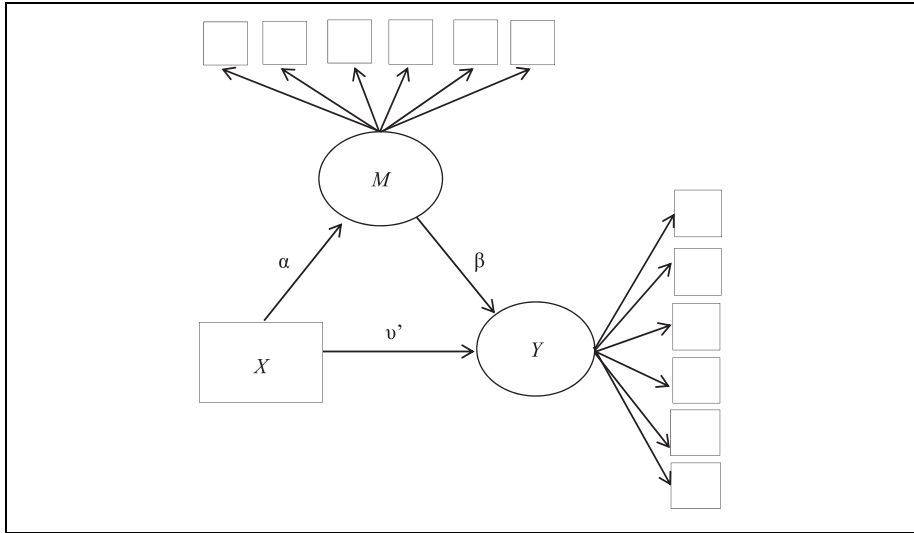


Figure 1. A single mediation model with independent variable (X) corresponding to a categorical observed variable measuring two or more groups, and two latent factors corresponding to the dependent (Y) and mediator (M) variables.

in relation to the latent variable must be the same across groups. That is, such comparisons assume the same construct M is being measured across the groups represented in X . When the measurement properties of the observed variables in relation to the target latent variable are the same across populations, measurement invariance is said to hold. In other words, knowledge about examinee group membership should not alter the relationships between observed and latent variables (Millsap, 2011). Measurement invariance is formally defined as (Mellenbergh, 1989),

$$P(\mathbf{V}|\xi, \mathbf{k}) = P(\mathbf{V}|\xi) \tag{4}$$

where \mathbf{V} is a $p \times 1$ vector of the observed variables, ξ is a $r \times 1$ vector of latent variables that we intend to measure, and \mathbf{k} corresponds to the groups assessed. Equation 4 indicates that the probability of observing scores on variables \mathbf{V} , conditional on the latent variables ξ , does not depend on the examinee belonging to group \mathbf{k} . If measurement invariance holds, group membership should not affect the probability of obtaining a score in the observed variables once the latent variables are taken into account. If Equation 4 does not hold, measurement bias is said to exist. Under measurement bias, the scores in the observed variables \mathbf{V} of two people with the same values in ξ will depend on their group membership (\mathbf{k}).

Factorial invariance denotes measurement invariance when relationships between the observed measures and latent variables are modeled through the common factor

model. The common factor model is extended to the multiple group case, with $k = 1, 2, \dots, K$ groups (Millsap, 2011),

$$\mathbf{V}_k = \boldsymbol{\tau}_k + \mathbf{\Lambda}_k \boldsymbol{\xi}_k + \boldsymbol{\delta}_k \quad (5)$$

where \mathbf{V} and $\boldsymbol{\xi}$ are defined as in Equation 4, $\boldsymbol{\tau}$ is a $p \times 1$ vector of latent factor intercepts, $\mathbf{\Lambda}$ is a $p \times r$ matrix of factor loadings, and $\boldsymbol{\delta}$ is a $p \times 1$ vector of unique factor scores. Factorial invariance is examined by comparing a series of nested models in which constraints on the item parameters within the common factor model are added sequentially (Jöreskog, 1971; Sörbom, 1974). Models are compared using a chi-square difference test in which a nonsignificant result indicates that the more constrained model fits the data as well as the less constrained model. The most basic form of factorial invariance is *configural* invariance, requiring that the same number of factors and same patterns of zero and nonzero loadings hold across the measured groups (Horn & McArdle, 1992). If the configural invariance model fits the data adequately well, the next more constrained model reflects *metric* invariance (Horn & McArdle, 1992), in which the values of the loadings are required to be the same across groups such that $\mathbf{\Lambda}_k = \mathbf{\Lambda}$. If the metric invariance model fits the data, it can be concluded that observed group differences in indicator covariances are due to the common factors and not differential functioning of the assessment across groups. If the hypothesis of metric invariance is not rejected, invariance constraints in the latent intercepts are tested $\boldsymbol{\tau}_k = \boldsymbol{\tau}$. This level of invariance is called *strong* measurement invariance (Meredith, 1993) or *scalar* invariance (Steenkamp & Baumgartner, 1998). If the hypothesis of strong measurement invariance is not rejected, observed differences in group means are fully explained by differences in the common factor means. Finally, equality across groups in the unique factor variances is tested in the *strict* factorial invariance model (Meredith, 1993). Invariance in the unique factor variances ensures that differences in the means and covariance structure of the observed variables are fully explained by differences in the common factors distributions. If invariance cannot be established in the evaluation of metric, strong and strict factorial invariance, an alternative is to test a model in which some of the item parameters are constrained to invariance while the others are allowed to vary between groups. Partial invariance is the term used to denote invariance in only a subset of parameters (Byrne, Shavelson, & Muthén, 1989).

Violations of Factorial Invariance in Mediation Models

The single mediation model described above assumes that measurement invariance holds, that is, it is assumed that the relationships between the observed items and the latent variable assessed in M are invariant across the groups represented in X . There have been only a few studies examining the impact of violations of invariance in mediation analysis. Two notable examples are the studies by Williams et al. (2010) and by Guenole and Brown (2014).

Williams et al. (2010) studied the relationship between anxiety (X) and heavy drinking as an outcome (Y) in a sample of Air Force and Navy trainees. The authors found that drinking motives was a significant mediator (M) in the relationship between anxiety and heavy drinking. The authors examined the drinking motives instrument for invariance across the Anxiety and Nonanxiety groups and found violations to both metric and strong measurement invariance. For example, the Anxiety group endorsed the item “drinking increases my self-confidence” at a higher level than the Nonanxiety group, after taking into account group differences in the underlying factor. To control for the lack of invariance, a Multiple Indicators Multiple Causes (MIMIC) model with direct paths from the X variable (Anxiety or Nonanxiety group) to the indicators measuring M were included. Once the significant paths relating X to each individual item were found, Williams et al. (2010) tested for statistical mediation. The α path coefficient, which describes the relationship between the anxiety groups and drinking motives, had a value of .11 ($SE = 0.05$). The authors also fit a mediation model with a latent mediator in which measurement invariance was assumed across groups. In this model, the α path coefficient was .20 ($SE = 0.04$). Williams et al. (2010) concluded that when violations of measurement invariance were not explicitly modeled in mediation analysis, estimates of the mediated effect might be inflated. The findings of Williams et al. (2010) indicated that measurement noninvariance may bias the conclusions of statistical mediation; however, this study was limited to investigating one substantive application.

Guenole and Brown (2014) conducted a simulation study to investigate the influence of measurement invariance in mediation models. For their study, Guenole and Brown (2014) simulated the independent (X), mediator (M), and dependent (Y) variables as continuous latent variables defined by multiple categorical indicators. Data were generated such that the loadings and/or thresholds were different across two groups in X , Y , or M . In the analysis phase, violations of invariance were ignored such that constraints in the loadings and/or thresholds were imposed and group differences in the estimated α and β path coefficients were examined. The results indicated that violating threshold invariance had a minimal effect on the confidence interval coverage and relative bias of path coefficients. In contrast, conditions with noninvariant loadings showed path coefficients that were under- or overestimated depending on the variable that showed violations of invariance. For example, noninvariance in the independent variable (X) led to the overestimation of path coefficients for the focal group; noninvariance in the dependent variable (Y) led to the underestimation of path coefficients for the focal group; and noninvariance in the mediator (M) led to underestimation of the α path coefficient and overestimation of the β path coefficient.

Although Guenole and Brown (2014) provided insights into the consequences of violations of invariance in α and β path coefficients, their study did not target our model of interest. The authors simulated the X , M , and Y variable as latent continuous variables; that is, the variable representing the groups being compared was not an explicit part of their model. We were interested in the case where the X variable represents two or more observed groups and where the M variable shows different

levels of violations of invariance across the groups measured in X . This model specification is typical in randomized controlled trials or observational studies where X represents observed groups (e.g., Williams et al., 2010). Furthermore, the magnitude of noninvariance in the loadings and thresholds was not systematically manipulated in Guenole and Brown (2014). That is, it was unclear in their findings what differences in impact should be expected under conditions of relatively trivial versus substantial noninvariance. Moreover, only one sample size of $N = 1,000$ per group was examined, which is much larger than typical controlled experiments in applied psychological research. Finally, Guenole and Brown examined the impact of noninvariance on the coverage and relative bias of the α and β path coefficients, but the overall impact of noninvariance on the mediated effect was not examined.

The studies by Williams et al. (2010) and Guenole and Brown (2014) indicate that inaccurate conclusions about mediated effects can be reached in the presence of noninvariance. However, given the limitations in Guenole and Brown's (2014) simulation and the difficulty of generalizing beyond the results of Williams et al. (2010), several important issues remain undetermined regarding the relationship between factorial invariance and mediation analysis. The purpose of the present study was to investigate the extent to which violations of invariance can be ignored in mediation analysis, and under what circumstances do violations of measurement invariance represent a threat to making accurate inferences when using the statistical mediation model. We examined the effects of violations of measurement invariance in mediation analysis by systematically manipulating the number of noninvariant items, the magnitude of the violations of invariance, and the sample size. We expected that as the number of items with violations of invariance and the magnitude of such violations increased, so would Type I error rates and relative bias of parameter estimates. Statistical power was also hypothesized to be affected by violations of invariance.

Method

A simulation study was conducted to examine the impact of violations of measurement invariance in the mediating construct on the estimation of the mediated effect in the single mediator model. The data-generating model for this Monte Carlo simulation is presented in Figure 1, where the X variable was a binary-indicator variable (e.g., 1 = Treatment, 0 = Control), and the M and Y were simulated as latent variables (Figure 1) indicated by six items. The indicators of the latent mediator were generated to violate measurement invariance, where the relationship between the items and the latent construct assessed in the mediator (M) differed between the two groups represented in X . To reflect the common situation wherein researchers fail to test for mediation invariance before conducting their intended analysis, a mediation model presuming measurement invariance was fit to the generated data. The impact of violations of invariance on estimates of the mediated effect was evaluated with the Monte Carlo outcomes of parameter bias (relative and standardized bias), Type I error rate, and statistical power.

Data Generation

The Monte Carlo procedure in Mplus 7.11 (Muthén & Muthén, 1998-2012) was used to generate data under multivariate normality. In each condition, 1,000 random samples were simulated. Data were generated using a multiple group approach in which M and Y were latent variables measured by six items each and mean group differences in M were used to simulate different effect sizes for the α path coefficient. The item parameters of the Y variable were simulated as invariant across the groups in X over all conditions. The item parameters for the mediator (M) were simulated under measurement invariance or with small, medium, or large violations (defined below) across groups in X in the intercepts or loadings. Because we were interested in examining the isolated effect of noninvariant intercepts or loadings, no condition was simulated with violations of invariance to both intercepts and loadings.

Effect Size for Measurement Noninvariance. In the current measurement invariance literature, there has not been a consensus on how to quantify *small*, *medium*, and *large* violations of measurement invariance. We took the following approaches. To define the magnitude of violations of invariance in the loadings, the approach suggested by Yoon and Millsap (2007) was followed. Since the same loading difference between groups of 0.1 could have a different practical implication (and represents a different amount of proportional change) if the loading shifted from 0.9 to 0.8 versus a shift from 0.3 to 0.2, it would not be adequate to define magnitudes of violations of invariance as simple fixed quantities. Instead, Yoon and Millsap (2007) followed a two-step procedure to calculate loading differences between groups. In the first step, effect sizes for group differences in the loadings were defined with respect to a specific item. In the present study, noninvariance was defined with respect to an item that had a loading of 0.6 in group one, and loadings of 0.7, 0.8, or 0.9 to exhibit *small*, *medium*, and *large* noninvariance, respectively.

The second step consisted of defining loadings for the remaining noninvariant items as a change proportional to the group differences defined in the first step. That is, in step one a small violation of invariance was defined as a change in an item loading from 0.6 in group one to a loading of 0.7 in group two. This represented a change ratio of 0.7/0.6 or 1.17. To determine the loadings of the remaining noninvariant items in group two, the item loadings in group one were multiplied by 1.17. In the same way, to create medium and large violations of invariance the loadings in group two were multiplied by 1.33 (0.8/0.6) and 1.50 (0.9/0.6), respectively.

The magnitude of noninvariance in the intercepts was defined as proposed by Millsap and Olivera-Aguilar (2012). As shown in Equation 6, the magnitude of noninvariance in the intercepts was defined as the ratio of the difference in intercepts to the difference in means.

$$d = \frac{\tau_{j1} - \tau_{j2}}{\mu_{j1} - \mu_{j2}} \quad (6)$$

Table 1. Generating Parameter Values.

	Proportion noninvariant items	Magnitude of noninvariance	Parameter values
Invariance	—	—	$\Lambda_{1=2} = [1, .65, .55, .60, .50, .80]$ $\tau_{1=2} = [.90, .40, .50, .50, .70, .40]$ $\theta_{1=2} = [1.3, 1.3, .80, .70, .90, .80]$
	1/3	Small	$\Lambda_2 = [1, .76, .64, .60, .50, .80]$
		Medium	$\Lambda_2 = [1, .87, .73, .60, .50, .80]$
Large		$\Lambda_2 = [1, .98, .83, .60, .50, .80]$	
Noninvariant loadings	2/3	Small	$\Lambda_2 = [1, .76, .64, .70, .58, .80]$
		Medium	$\Lambda_2 = [1, .87, .73, .80, .67, .80]$
		Large	$\Lambda_2 = [1, .98, .83, .90, .75, .80]$
	1/3	Small	$\tau_2 = [.90, .47, .56, .50, .70, .40]$
		Medium	$\tau_2 = [.90, .66, .72, .50, .70, .40]$
		Large	$\tau_2 = [.90, 1.44, 1.38, .50, .70, .40]$
Noninvariant intercepts	2/3	Small	$\tau_2 = [.90, .47, .56, .56, .75, .40]$
		Medium	$\tau_2 = [.90, .66, .72, .74, .90, .40]$
		Large	$\tau_2 = [.90, 1.44, 1.38, 1.46, 1.5, .40]$

where $\tau_{j1} - \tau_{j2}$ corresponds to the difference in intercepts of item j between two groups, and $\mu_{j1} - \mu_{j2}$ corresponds to the group difference in means for item j . Following Millsap and Olivera-Aguilar (2012), ratios of 0.2, 0.5, and 0.8 were defined as *small*, *medium*, and *large* effect sizes, respectively. In the present study, item means were determined using Equation 7 which shows the expected value for group k ,

$$E(\mathbf{V}_k) = \boldsymbol{\mu}_k = \boldsymbol{\tau}_k + \boldsymbol{\Lambda}_k \boldsymbol{\xi}_k \tag{7}$$

Substituting Equation 7 into Equation 6 and rearranging terms, intercepts for non-invariant items in group two were determined using Equation 8,

$$\tau_{j2} = \frac{\mathbf{d}\tau_{j1} + \mathbf{d}\boldsymbol{\Lambda}_1 \boldsymbol{\xi}_1 - \mathbf{d}\boldsymbol{\Lambda}_2 \boldsymbol{\xi}_2 - \boldsymbol{\tau}_{j2}}{\mathbf{d} - 1} \tag{8}$$

where the factor mean for the first group ξ_1 was set to zero, and the factor mean of the second group ξ_2 was set to 0.4. Note that since the factor mean for the second group was larger than the factor mean of the first group, the intercepts of noninvariant items increased in the second group.

The number of items demonstrating violations of invariance was manipulated to be zero, two, or four. These numbers were selected to represent conditions in which all items were invariant, or where 1/3 or 2/3 of the total number of items exhibited noninvariance (Yoon & Millsap, 2007). Final item parameter values for the latent mediator variable are shown in Table 1. Item parameter values were selected so that communalities were between 0.2 and 0.6.

Other manipulated variables were the sample size per group and values of the path coefficients. Group sample sizes were set to $n = 100, 250, \text{ or } 500$. These values were selected to represent total N values of 200 and 500 commonly found in social science research and to explore the case of a larger total sample size of $N = 1,000$. Population values of the regression parameters ($\alpha, \beta, \text{ and } \nu'$) were manipulated to be 0.00, 0.14, 0.39, and 0.59 corresponding approximately to values of Cohen's criteria for zero, small (2% of variance accounted for), medium (13% of the variance), and large (26% of the variance) effect sizes, respectively (Cohen, 1988; MacKinnon et al., 2002). In summary, we simulated 192 invariant conditions (3 sample sizes and 4 path values for $\alpha, \beta, \text{ and } \nu$), 1,152 conditions with noninvariant loadings (3 sample sizes, 4 path values for $\alpha, \beta, \text{ and } \nu, 3$ effect sizes for violations of invariance, and 2 proportions of items violating invariance), and 1,152 conditions with noninvariant intercepts (3 sample sizes, 4 path values for $\alpha, \beta, \text{ and } \nu, 3$ effect sizes for violations of invariance, and 2 proportions of items violating invariance).

Data Analysis

A single mediator model with latent variables for the mediator and the dependent variable was fit to the generated data using Mplus 7.11 (Muthén & Muthén, 1998-2012). Measurement invariance (i.e., loadings and intercepts constrained to equality) was assumed in the fitted model for both the mediator and dependent variable. The impact of ignoring the violations of invariance in the loadings and intercepts was evaluated looking at several criteria including Type I error rates, statistical power, and relative or standardized bias of the mediated effect coefficients.

To test statistical significance and obtain asymmetric confidence intervals for the mediated effect, we used the analytical solution for the distribution of the product of two random variables computed in the R-statistical platform using the package RMediation (Tofighi & MacKinnon, 2011). The distribution of the product method has comparable power to the computationally intensive resampling methods to test for significance of the mediated effect, and accounts for a possible correlation between α and β often present in mediation models with latent variables (MacKinnon, 2008; Valente et al., 2016). Type I error rates were defined as the proportion of replications in which the estimate of the mediated effect was statistically significant but the population value of the mediated effect was zero (i.e., when the confidence interval did not contain zero in conditions in which either or both α and β were simulated to equal zero). Statistical power was defined as the proportion of replications in which a significant mediated effect was correctly detected (i.e., when the confidence interval did not contain zero in conditions where both α and β were simulated to be nonzero).

Relative bias was computed for the estimated α, β, ν' and $\alpha\beta$ coefficients in the conditions in which population values for α and β were nonzero. Relative bias for each estimated coefficient ($\hat{\theta}_c$) was defined as the ratio of the difference between the true and estimated parameter value to the true value:

$$\text{RBias}(\hat{\theta}_c) = R^{-1} \sum_{r=1}^R \frac{(\hat{\theta}_{rc} - \theta_c)}{\theta_c} \quad (9)$$

where R refers to the total number of replications that converged to a solution, θ_c refers to the true value of the α , β , ν' or $\alpha\beta$ coefficients, and $\hat{\theta}_{rc}$ refers to the parameter estimate for replication r in condition c . Relative bias was judged to be acceptable when its absolute value was less than 0.05 (Hoogland & Boomsma, 1998).

A problem with relative bias is that it is not defined for conditions where population values for the parameter estimates are zero. Therefore, standardized bias was computed for the estimated α , β , ν' , and $\alpha\beta$ coefficients in the conditions in which α or β were zero. Standardized bias was defined as the ratio of the difference between the true and estimated parameter value to the standard deviation of the parameter estimate ($SD_{\hat{\theta}_c}$):

$$\text{SBias}(\hat{\theta}_c) = R^{-1} \sum_{r=1}^R \frac{(\hat{\theta}_{rc} - \theta_c)}{SD_{\hat{\theta}_c}} \quad (10)$$

An analysis of variance (ANOVA) was conducted for each of the estimated path coefficients to determine the effect of the independent variables (sample size, path values for α , β , and ν , number of items, proportion of noninvariant items, and magnitude of the violations of invariance) on relative and standardized bias. The effect size of ANOVA results was determined by η^2 , with Cohen's (1988) suggested values used to judge associations between the variables as small ($\eta^2 = 0.01$), medium ($\eta^2 = 0.06$), or large ($\eta^2 = 0.14$).

Results

All replications in every condition converged, with no error messages or negative variance estimates obtained. For ease of presentation, we present results averaged over the values of the direct effect, ν' , because varying the direct effect did not lead to substantial differences on simulation outcome measures.

Type I Error Rates

Type I error rates for the cases in which $\beta = 0$ were below or close to .05 in all conditions and hence not presented. The Type I error rates for the cases in which $\alpha = 0$ are shown in Table 2. It is clear that when measurement invariance holds, the proportion of replications in which a significant mediation effect was incorrectly detected was below or close to 5%. The same pattern was observed in conditions with noninvariant loadings where Type I error rates remained below 5.7% regardless of the proportion of noninvariant items or the magnitude of loading noninvariance.

In contrast, in the presence of noninvariant intercepts, Type I error rates were above 5% in most conditions. Interestingly, as the proportion of noninvariant items,

Table 2. Type I Error Rates for the Mediated Effect by Magnitude of Noninvariance, Proportion of Noninvariant Items, Sample Size, and Values of α and β .

n	α and β	Invariance	Noninvariant loadings						Noninvariant intercepts						
			Small		Medium		Large		Small		Medium		Large		
			$p = 1/2$	$p = 2/3$	$p = 1/2$	$p = 2/3$	$p = 1/2$	$p = 2/3$	$p = 1/2$	$p = 2/3$	$p = 1/2$	$p = 2/3$	$p = 1/2$	$p = 2/3$	
100	Zero, zero	0.000	0.002	0.002	0.000	0.000	0.001	0.001	0.000	0.000	0.001	0.002	0.006	0.023	0.034
	Zero, small	0.008	0.012	0.012	0.012	0.012	0.012	0.012	0.009	0.014	0.015	0.015	0.055	0.141	0.257
	Zero, medium	0.044	0.051	0.053	0.050	0.048	0.047	0.048	0.048	0.062	0.083	0.086	0.224	0.580	0.970
250	Zero, large	0.045	0.054	0.055	0.054	0.052	0.055	0.057	0.054	0.067	0.086	0.086	0.233	0.609	1.000
	Zero, zero	0.000	0.002	0.002	0.001	0.001	0.001	0.001	0.004	0.002	0.002	0.002	0.019	0.037	0.046
	Zero, small	0.026	0.025	0.024	0.028	0.025	0.022	0.020	0.025	0.044	0.077	0.077	0.273	0.577	0.634
500	Zero, medium	0.049	0.045	0.045	0.053	0.052	0.049	0.051	0.051	0.076	0.154	0.154	0.488	0.915	1.000
	Zero, large	0.046	0.046	0.046	0.053	0.052	0.046	0.048	0.049	0.075	0.150	0.150	0.483	0.920	1.000
	Zero, zero	0.000	0.000	0.001	0.000	0.000	0.003	0.003	0.002	0.001	0.006	0.006	0.032	0.035	0.048
	Zero, small	0.042	0.031	0.031	0.044	0.047	0.038	0.037	0.057	0.093	0.204	0.204	0.701	0.907	0.911
	Zero, medium	0.052	0.045	0.044	0.053	0.052	0.049	0.048	0.067	0.107	0.242	0.242	0.784	0.995	1.000
	Zero, large	0.052	0.044	0.044	0.052	0.052	0.048	0.048	0.067	0.110	0.248	0.248	0.786	0.995	1.000

Note. p = proportion of noninvariant items. For α and β , small = 0.14, medium = 0.39, and large = 0.59.

Table 3. Standardized Bias of the Mediated Effect in the Conditions With Noninvariant Intercepts by Magnitude of Noninvariance, Proportion of Noninvariant Items, Sample Size, and Values of β .

n	α and β	Violations of invariance					
		Small		Medium		Large	
		$p = 1/2$	$p = 2/3$	$p = 1/2$	$p = 2/3$	$p = 1/2$	$p = 2/3$
100	Zero, zero	0.002	0.001	-0.033	0.003	-0.038	-0.055
	Zero, small	0.078	0.171	0.314	0.660	0.952	1.360
	Zero, medium	0.112	0.264	0.499	1.056	1.775	3.165
	Zero, large	0.124	0.291	0.527	1.129	1.995	4.000
250	Zero, zero	0.015	-0.015	0.007	-0.021	-0.049	-0.026
	Zero, small	0.217	0.384	0.674	1.275	1.790	2.222
	Zero, medium	0.228	0.472	0.862	1.791	3.016	5.131
	Zero, large	0.236	0.483	0.886	1.855	3.296	6.442
500	Zero, zero	-0.043	0.015	-0.002	-0.011	-0.061	0.009
	Zero, small	0.309	0.612	1.042	1.955	2.655	3.202
	Zero, medium	0.342	0.674	1.219	2.579	4.322	7.308
	Zero, large	0.343	0.682	1.242	2.666	4.670	9.154

magnitude of noninvariance, and sample size increased, Type I error rates increased as well. In some cases, the Type I error rates reached 100%. For example, where the value of α was zero, the value of β was 0.39 ("medium"), 2/3 of the items had large violations of invariance and a sample size per group of 500, 100% of replications detected a significant mediated effect where none existed.

Standardized Bias

To explore the Type I error rate results in the conditions with noninvariant intercepts in more detail, standardized bias was computed for the path parameter estimates. On average, the α path showed a standardized bias of 2.61, the β path of 0.24, the ν' path of -0.77, and the $\alpha\beta$ path of 1.31. These results indicate that the bias in the α path inflated the estimated value of the $\alpha\beta$ path.

For ease of presentation, Table 3 shows standardized bias for the $\alpha\beta$ parameter estimate only. Similar to the Type I error rate results, as the proportion of noninvariant items, magnitude of noninvariance, and sample size increased, the standardized bias of the $\alpha\beta$ path increased. The most affected conditions were for those with a sample size of 500 per group, large violations of invariance and β values larger than zero; in the case with 2/3 of noninvariant items and a large value of β , standardized bias showed a value of 9.15. ANOVA results indicated large effect sizes ($\eta^2 > 0.14$) for the magnitude of violations of intercept invariance, for the β path effect size, and for the interaction between the β path effect size and the magnitude of violations of

invariance. The interaction indicates that the effect of the magnitude of violations on invariance increased as the β path coefficient increased.

Statistical Power

Results for statistical power are shown in Table 4. As expected, in the invariant conditions as the sample size and effect size of α and β increased, power to detect a mediated effect increased. All conditions with noninvariant loadings followed the same pattern of results as conditions generated under measurement invariance. That is, statistical power increased as the sample size and effect size of the α and β parameters increased.

In conditions with noninvariant intercepts, statistical power was higher than in the invariant conditions, in particular in the conditions with medium and large violations of invariance, and small and medium values for α . For example, in conditions with a sample size of 100, a small effect size for α , and a large effect size for β , statistical power was .133 for the invariant conditions. However, statistical power increased to 1 in the conditions with 2/3 of the items showing large violations of intercept invariance. Except for the conditions in which $\beta = 0$, most conditions with large violations of invariance had statistical power above .8 regardless of the value of α and sample size.

Relative Bias

Table 5 shows the relative bias for the mediated effect $\alpha\beta$, and for the α and β paths. Results are averaged over sample size conditions since all ANOVAs showed η^2 values below 0.01 for all sample sizes. For ease of presentation, results are displayed for conditions in which both coefficients had the same effect size (i.e., both α and β were simulated to exhibit small, medium, or large effects). As expected, in conditions in which measurement invariance held the relative bias of parameters was always below the suggested 0.05 cutoff.

In the presence of noninvariant loadings, while the β estimate did not show relative bias values above the suggested value of 0.05, the α , ν' , and the $\alpha\beta$ path coefficients were affected by violations of invariance in most conditions. As the magnitude of violations of loading invariance and the proportion of noninvariant items increased, the relative bias of the α , ν' , and the $\alpha\beta$ path coefficients tended to increase as well. ANOVA results showed η^2 values below 0.01 for all independent variables across all conditions with noninvariant loadings.

In conditions with noninvariant intercepts, path coefficient estimates were severely affected as demonstrated by relative bias values above 0.05 across most conditions. Even conditions with small violations of invariance and only 1/2 of items exhibiting noninvariant intercepts showed relative bias values above 0.05 for most path coefficients. The most affected conditions were for small (.14) values of α and β , wherein the $\alpha\beta$ estimate showed relative bias values up to 8.06 in the case of large violations

Table 4. Power for the Mediated Effect by Magnitude of Noninvariance, Proportion of Noninvariant Items, Sample Size, and Values of α and β .

n	α and β	Invariance	Noninvariant loadings						Noninvariant intercepts					
			Small		Medium		Large		Small		Medium		Large	
			$p = 1/2$	$p = 2/3$	$p = 1/2$	$p = 2/3$	$p = 1/2$	$p = 2/3$	$p = 1/2$	$p = 2/3$	$p = 1/2$	$p = 2/3$	$p = 1/2$	$p = 2/3$
100	Small, small	0.024	0.031	0.033	0.028	0.031	0.033	0.036	0.037	0.047	0.063	0.128	0.259	0.285
	Small, medium	0.123	0.140	0.144	0.142	0.149	0.141	0.154	0.164	0.203	0.296	0.528	0.844	0.966
	Small, large	0.133	0.147	0.151	0.154	0.160	0.147	0.160	0.174	0.215	0.312	0.548	0.866	1.000
	Medium, small	0.179	0.166	0.173	0.185	0.204	0.204	0.223	0.181	0.195	0.241	0.281	0.297	0.293
	Medium, medium	0.636	0.650	0.669	0.680	0.727	0.713	0.762	0.681	0.727	0.827	0.930	0.969	0.972
	Medium, large	0.666	0.680	0.702	0.691	0.737	0.734	0.782	0.709	0.757	0.843	0.948	0.993	1.000
	Large, small	0.280	0.274	0.284	0.289	0.305	0.317	0.337	0.281	0.278	0.294	0.298	0.313	0.283
	Large, medium	0.923	0.936	0.947	0.944	0.960	0.953	0.972	0.934	0.951	0.965	0.979	0.976	0.973
	Large, large	0.946	0.958	0.969	0.963	0.977	0.968	0.984	0.956	0.970	0.988	0.999	0.999	1.000
	Small, small	0.144	0.155	0.164	0.174	0.187	0.176	0.199	0.199	0.268	0.380	0.576	0.646	0.623
	Small, medium	0.262	0.282	0.293	0.309	0.329	0.319	0.355	0.359	0.457	0.616	0.911	0.995	1.000
	Small, large	0.254	0.283	0.297	0.306	0.324	0.320	0.355	0.359	0.462	0.610	0.908	0.997	1.000
Medium, small	0.616	0.626	0.634	0.659	0.673	0.664	0.679	0.632	0.644	0.665	0.679	0.658	0.629	
Medium, medium	0.965	0.973	0.978	0.978	0.987	0.982	0.989	0.981	0.991	0.997	1.000	1.000	1.000	
Medium, large	0.965	0.976	0.980	0.975	0.983	0.983	0.989	0.982	0.991	0.997	1.000	1.000	1.000	
Large, small	0.673	0.681	0.682	0.690	0.697	0.672	0.680	0.668	0.680	0.653	0.668	0.657	0.639	
Large, medium	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	
Large, large	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	
500	Small, small	0.433	0.455	0.476	0.471	0.511	0.483	0.542	0.558	0.678	0.814	0.909	0.927	0.912
	Small, medium	0.480	0.506	0.531	0.525	0.567	0.541	0.596	0.618	0.742	0.899	0.998	1.000	1.000
	Small, large	0.486	0.511	0.533	0.517	0.558	0.534	0.588	0.617	0.746	0.902	0.997	1.000	1.000
	Medium, small	0.926	0.926	0.928	0.931	0.935	0.928	0.929	0.917	0.921	0.930	0.926	0.920	0.910
	Medium, medium	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	Medium, large	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	Large, small	0.915	0.920	0.921	0.919	0.920	0.923	0.924	0.919	0.923	0.932	0.925	0.915	0.913
	Large, medium	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	Large, large	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Note. p = proportion of noninvariant items. For α and β , small = 0.14, medium = 0.39, and large = 0.59.

Table 5. Relative Bias of the Path Coefficients and Mediated Effect by Magnitude of Noninvariance, Proportion of Noninvariant Items, Sample Size, and Values of α and β .

	Proportion noninvariant items	Magnitude of noninvariance	$\alpha = \beta = \text{Small (0.14)}$						$\alpha = \beta = \text{Medium (0.39)}$						$\alpha = \beta = \text{Large (0.59)}$					
			α		β		$\alpha\beta$		α		β		$\alpha\beta$		α		β		$\alpha\beta$	
			α	β	α	β	$\alpha\beta$	ν'	α	β	$\alpha\beta$	ν'	α	β	$\alpha\beta$	ν'	α	β	$\alpha\beta$	ν'
Invariance	—	—	-0.015	.004	.004	-0.020	-.003	.005	-0.004	-0.001	-0.002	0.009	-0.002	0.009	-0.002	0.009	-0.002	0.003	-0.002	0.003
Non invariant loadings	1/2	Small	0.027	0.001	-0.004	0.024	0.019	0.005	-0.010	0.020	0.014	0.008	-0.026	0.019	0.035	0.012	-0.048	0.044	0.035	0.012
		Medium	0.058	0.006	-0.004	0.052	0.032	0.009	-0.023	0.037	0.035	0.012	-0.086	0.065	0.053	0.014	-0.086	0.065	0.053	0.014
	2/3	Large	0.057	0.004	-0.001	0.050	0.058	0.009	-0.036	0.064	0.053	0.014	-0.086	0.065	0.053	0.014	-0.086	0.065	0.053	0.014
		Small	0.045	0.002	-0.005	0.043	0.037	0.007	-0.021	0.040	0.031	0.011	-0.052	0.039	0.031	0.011	-0.052	0.039	0.031	0.011
		Medium	0.090	0.007	-0.007	0.086	0.064	0.012	-0.043	0.073	0.065	0.017	-0.094	0.079	0.065	0.017	-0.094	0.079	0.065	0.017
		Large	0.098	0.007	-0.004	0.095	0.098	0.014	-0.062	0.110	0.091	0.021	-0.147	0.110	0.091	0.021	-0.147	0.110	0.091	0.021
Non invariant intercepts	1/2	Small	0.169	0.016	-0.013	0.186	0.056	0.008	-0.044	0.060	0.036	0.011	-0.065	0.043	0.036	0.011	-0.065	0.043	0.036	0.011
		Medium	0.662	0.012	-0.050	0.669	0.227	0.029	-0.144	0.259	0.142	0.029	-0.219	0.171	0.142	0.029	-0.219	0.171	0.142	0.029
	2/3	Large	2.820	0.137	-0.242	3.368	0.982	0.195	-0.792	1.369	0.599	0.210	-1.241	0.932	0.599	0.210	-1.241	0.932	0.599	0.210
		Small	0.355	0.020	-0.026	0.384	0.121	0.015	-0.066	0.134	0.075	0.015	-0.119	0.088	0.075	0.015	-0.119	0.088	0.075	0.015
		Medium	1.371	0.033	-0.109	1.446	0.463	0.057	-0.311	0.542	0.283	0.065	-0.481	0.362	0.283	0.065	-0.481	0.362	0.283	0.065
		Large	4.779	0.579	-0.607	8.057	1.487	0.579	-1.680	2.899	0.875	0.553	-2.510	1.892	0.875	0.553	-2.510	1.892	0.875	0.553

of invariance and 2/3 of items demonstrating noninvariant intercepts. For the relative bias of all estimated path coefficients, ANOVA results indicated large effect sizes ($\eta^2 > 0.14$) for the magnitude of violations of intercept invariance and for the effect size of the α path coefficient. A large ANOVA effect size was also found for the interaction term between the α path coefficient and magnitude of violations of invariance when studying the relative bias of α and $\alpha\beta$ estimates, such that as the effect size of the α path coefficient increased, the magnitude of violations of invariance had a smaller effect on the relative bias of the α and $\alpha\beta$ estimates.

Discussion

Measurement invariance is a crucial assumption when using an instrument to examine group differences. However, violations of invariance are frequently found in practice (Schmitt & Kuljanin, 2008), and it is important to understand its implications for the conclusions derived from statistical analyses. The purpose of this study was to systematically explore the impact of violations of measurement invariance in a mediation model in which the independent variable (X) denotes two or more groups (e.g., control vs. treatment), and both the mediator (M) and dependent variable (Y) are latent variables measured via several observed indicators. We conducted a simulation study in which we manipulated the proportion of noninvariant items in the mediator as well as the magnitude of invariance violations in both its item loading and intercept values, and examined its impact on the mediated effect. Significance testing for the mediated effect was conducted using the distribution of the product to compute asymmetrical confidence intervals. The simulations indicated a different pattern of results when violations of invariance were simulated in item loadings versus when violations of invariance occurred in item intercepts. Conditions exhibiting violations of invariance in the loadings showed results similar to our control conditions (i.e., noninvariance) regarding Type I error rates, statistical power, and relative bias of the path coefficients. Under both scenarios, Type I error rates were below or close to 5% in most conditions. In line with MacKinnon et al. (2002), when α and β were generated to show large effect sizes, statistical power was close to 1.00 even with sample sizes of only 100 per group. Sample sizes of 250 per group were sufficient to detect a mediated effect when α and β were generated with medium effect sizes. Although some conditions with medium and large violations of invariance in item loadings produced relative bias estimates for the mediated effect ($\alpha\beta$) above the suggested 0.05 cutoff (Hoogland & Boomsma, 1998), this bias did not affect the Type I error rates or statistical power.

Contrary to the results found for measurement noninvariance in item loadings, in the presence of noninvariant intercepts Type I error rates were inflated in most conditions, even reaching values of 1.0 in conditions with large violations of invariance. Statistical power was larger in these conditions than in the invariant conditions, with $\alpha\beta$ estimates severely biased.

To explain these results, it is important to note that in conditions with violations of intercept invariance, the group with the larger intercept values also had larger factor scores. As shown in Equation 4, as a consequence of having larger intercepts, the expected value of the mediator in the group with violations of intercept invariance was overestimated. This led to the overestimation of the α path coefficients and, as a consequence, of the mediated effect $\alpha\beta$ as well.

In general, these results suggest that the mediated effect in a single mediator model is robust to violations of metric invariance, but not to violations of scalar invariance (i.e., noninvariance in their indicator intercepts). As noted by Vandenberg and Lance (2000), differences in item intercepts between groups could be indicative of systematic response bias (e.g., leniency). That is, differences in intercepts related to measurement invariance could reflect a bias in how participants in one group respond to items, rather than a treatment induced change in thresholds. However, differences in item intercepts could also be a consequence of known group differences. In situations in which group mean differences at the latent level are expected, such differences would likely be reflected at the item level. In the specific case examined in our study, in which the independent variable corresponds to two groups (e.g., treatment vs. control), differences in mediator item intercepts may be a reflection of the efficacy of the treatment. To distinguish between measurement bias and true population differences between groups, a pretest measurement would be required. That is, measurement invariance would be analyzed *before* the treatment was implemented to assess whether there were salient pretreatment differences. Measurement invariance between the two groups would then be tested again after the treatment was implemented. Posttreatment group differences in the intercepts would not necessarily indicate measurement bias if only the treatment group showed significant differences in its item intercepts over time. That is, in addition to measurement invariance *across* groups, longitudinal invariance *within* each group would also be compared (see Millsap & Cham, 2013, for a review on longitudinal invariance). In an experiment in which longitudinal measurement invariance was found in the control group but not in the treatment group, posttreatment violations of invariance between groups could be interpreted as a reflection of true differences caused by the treatment. Mediation analysis could then be conducted assuming measurement invariance in the mediator, with noninvariance due to the treatment effect absorbed by the mediation effect (as shown in the present study). However, if there were violations of invariance across groups prior to the treatment or the control group fails to demonstrate longitudinal invariance, then group differences in the intercepts posttreatment may be due to bias in one of the groups. In such cases, researchers run the risk of obtaining biased estimates of mediated effects when methods of handling noninvariance are not implemented (i.e., noninvariance is ignored in mediation analyses). Two methods available to mitigate effects of noninvariance are the estimation of MIMIC models to control for between-group differences and partial invariance models allowing some noninvariant item parameters to be freely estimated.

Limitations and Future Studies

Overall, our results indicate that Type I error rates, statistical power, and relative bias were not as severely affected in the presence of noninvariant loadings as in data exhibiting noninvariant intercepts. However, it is possible that this finding was a result of the particular way in how the magnitude of noninvariance was conceptualized. Although values of the intercepts and loadings in the noninvariant conditions were chosen following examples in the literature (Millsap & Olivera-Aguilar, 2012; Yoon & Millsap, 2007), the literature also lacks clarity on the issue of what effect size values should be considered *small*, *medium*, or *large* violations of invariance. It is possible that a different approach to conceptualize different magnitudes of measurement invariance would have resulted in larger relative bias values or the inflation of Type I error rates or statistical power.

It should be noted that in the study of measurement invariance, it is often the case that if an item has noninvariant loadings it also exhibits a noninvariant intercept. A simulation condition in which both the intercepts and loadings are generated with violations of invariance would be helpful. In this case, we would expect that the results for relative bias, Type I error rates, and statistical power would be larger than those observed in the current study.

In our study, violations of invariance were only simulated for indicators of the latent mediating variable. In most mediation studies, the dependent variable (e.g., depression, intelligence, drug use) is frequently measured with a series of indicators, and as such, it becomes relevant to test for measurement invariance. Simulation studies that examine conditions in which violations of invariance are also considered in the *Y* variable are important. Future studies should also explore the performance of different procedures for controlling or correcting for noninvariance, such as latent variable models allowing for partial invariance (Byrne et al., 1989) or MIMIC models.

Conclusions

To the best of our knowledge, this is the first study to systematically examine the impact of differing magnitudes and types of measurement noninvariance in a model in which a latent variable mediates the relationship between group membership and a dependent variable. Similar to Williams et al. (2010), we found that under some conditions the mediated effect was overestimated when noninvariance was ignored. Researchers run the risk of making inaccurate conclusions about mediated effects if noninvariance is ignored. This highlights the importance of integrating psychometric analyses within mediation studies, particularly the use of latent variable models to improve measurement quality (MacKinnon, 2008) and the incorporation of tests for different types of measurement invariance (Millsap, 2011).

Authors' Note

We dedicate this article to the memory of Roger E. Millsap.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was supported by the National Institute on Drug Abuse (Grant No. DA009757) and the National Science Foundation Graduate Research Fellowship (Grant No. DGE-1311230).

References

- Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51*, 1173-1182.
- Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin, 105*, 456-466.
- Chen, F. F. (2008). What happens if we compare chopsticks with forks? The impact of making inappropriate comparisons in cross-cultural research. *Journal of Personality and Social Psychology, 95*, 1005-1018.
- Cheung, G. W., & Rensvold, R. B. (1999). Testing factorial invariance across groups: A reconceptualization and proposed new method. *Journal of Management, 25*, 1-27.
- Cohen, J. (1988). *Statistical power for the behavioral sciences*. Hillsdale, NJ: Erlbaum.
- Gonzalez, O., & MacKinnon, D. P. (2016, June). *Measurement and psychometric issues in statistical mediation analysis*. E-poster presented at the 24th annual meeting of the Society for Prevention Research, San Francisco, CA.
- Gregg, J. A., Callaghan, G. M., Hayes, S. C., & Glenn-Lawson, J. L. (2007). Improving diabetes self-management through acceptance, mindfulness, and values: A randomized controlled trial. *Journal of Consulting and Clinical Psychology, 75*, 336-343.
- Guenole, N., & Brown, A. (2014). The consequences of ignoring measurement invariance for path coefficients in structural equation models. *Frontiers in Psychology, 5*, 1-16.
- Hoogland, J. J., & Boomsma, A. (1998). Robustness studies in covariance structure modeling: An overview and a meta-analysis. *Sociological Methods & Research, 26*, 329-367.
- Horn, J. L., & McArdle, J. J. (1992). A practical guide to measurement invariance in research on aging. *Experimental Aging Research, 18*, 117-144.
- Jöreskog, K. G. (1971). Simultaneous factor analysis in several populations. *Psychometrika, 36*, 409-426.
- Judd, C. M., & Kenny, D. A. (1981). Process analysis: Estimating mediation in treatment evaluations. *Evaluation Review, 5*, 602-619.
- Kisbu-Sakarya, Y., MacKinnon, D. P., & Miočević, M. (2014). The distribution of the product explains normal theory mediation confidence interval estimation. *Multivariate Behavioral Research, 49*, 261-268.
- MacKinnon, D. P. (2008). *Introduction to statistical mediation analysis*. New York, NY: Taylor & Francis/Erlbaum.

- MacKinnon, D. P., & Dwyer, J. H. (1993). Estimating mediated effects in prevention studies. *Evaluation Review, 17*, 144-158.
- MacKinnon, D. P., Fritz, M. S., Williams, J., & Lockwood, C. M. (2007). Distribution of the product confidence limits for the indirect effect: Program PRODCLIN. *Behavior Research Methods, 39*, 384-389.
- MacKinnon, D. P., Lockwood C. M., Hoffman, J. M., West, S. G., & Sheets, V. (2002). A comparison of methods to test mediation and other intervening variable effects. *Psychological Methods, 7*, 83-104.
- MacKinnon, D. P., Lockwood C. M., & Williams, J. (2004). Confidence limits for the indirect effect: Distribution of the product and resampling methods. *Multivariate Behavioral Research, 39*, 99-128.
- Mellenbergh, G. J. (1989). Item bias and item response theory. *International Journal of Educational Research, 13*, 127-143.
- Meredith, W. (1993). Measurement invariance, factor analysis, and factorial invariance. *Psychometrika, 30*, 419-440.
- Millsap, R. E. (2011). *Statistical approaches to measurement invariance*. New York, NY: Taylor & Francis.
- Millsap, R. E., & Cham, H. (2013). Investigating factorial invariance in longitudinal data. In B. Laursen, T. D., Little, & N. A. Card (Ed.), *Handbook of developmental research methods* (pp. 109-126). New York, NY: Guilford Press.
- Millsap, R. E., & Olivera-Aguilar, M. (2012). Investigating measurement invariance using confirmatory factor analysis. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 380-392). New York, NY: Guilford Press.
- Muthén, L. K., & Muthén, B. O. (1998-2012). *Mplus user's guide* (7th ed.). Los Angeles, CA: Muthén & Muthén.
- Schmitt, N., & Kuljanin, G. (2008). Measurement invariance: Review of practice and implications. *Human Resource Management Review, 18*, 210-222.
- Sobel, M. E. (1982). Asymptotic confidence intervals for indirect effects in structural equation models. *Sociological Methodology, 13*, 290-312.
- Sobel, M. E. (1986). Some new results on indirect effects and their standard errors in covariance structure models. *Sociological Methodology, 16*, 159-186.
- Sörbom, D. (1974). A general method for studying differences in factor means and factor structure between groups. *British Journal of Mathematical and Statistical Psychology, 27*, 229-239.
- Steenkamp, J. E. M., & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research, 25*, 78-90.
- Stice, E., Presnell, K., Gau, J., & Shaw, H. (2007). Testing mediators of intervention effects in randomized controlled trials: An evaluation of two eating disorder prevention programs. *Journal of Consulting and Clinical Psychology, 75*, 20-32.
- Tofighi, D., & MacKinnon, D. P. (2011). RMediation: An R package for mediation analysis confidence intervals. *Behavior Research Methods, 43*, 692-700.
- Valente, M. J., Gonzalez, O., Miočević, M., & MacKinnon, D. P. (2016). A note on testing mediated effects in structural equation models: Reconciling past and current research on the performance of the test of joint significance. *Educational and Psychological Measurement, 76*, 889-911.

- 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.
- Varra, A. A., Hayes, S. C., Roget, N., & Fisher, G. (2008). A randomized control trial examining the effect of acceptance and commitment training on clinician willingness to use evidence-based pharmacotherapy. *Journal of Consulting and Clinical Psychology*, 76, 449-458.
- Williams, J., Jones, S. B., Pemberton, M. R., Bray, R. M., Brown, J. M., & Vandermaas-Peeler, R. (2010). Measurement invariance of alcohol use motivations in junior military personnel at risk of depression or anxiety. *Addictive Behaviors*, 35, 444-451.
- Yoon, M., & Millsap, R. E. (2007). Detecting violations of factorial invariance using data-based specification searches: A Monte Carlo study. *Structural Equation Modeling*, 14, 435-463.