

# Croon's Bias-Corrected Factor Score Path Analysis for Small- to Moderate-Sample Multilevel Structural Equation Models

Organizational Research Methods  
1-23

© The Author(s) 2019

Article reuse guidelines:

[sagepub.com/journals-permissions](https://sagepub.com/journals-permissions)

DOI: 10.1177/1094428119879758

[journals.sagepub.com/home/orm](https://journals.sagepub.com/home/orm)



Benjamin Kelcey<sup>1</sup> , Kyle Cox<sup>2</sup>, and Nianbo Dong<sup>3</sup>

## Abstract

Maximum likelihood estimation of multilevel structural equation model (MLSEM) parameters is a preferred approach to probe theories involving latent variables in multilevel settings. Although maximum likelihood has many desirable properties, a major limitation is that it often fails to converge and can incur significant bias when implemented in studies with a small to moderate multilevel sample (e.g., fewer than 100 organizations with 10 or less individuals/organization). To address similar limitations in single-level SEM, literature has developed Croon's bias-corrected factor score path analysis estimator that converges more regularly than maximum likelihood and delivers less biased parameter estimates with small to moderate sample sizes. We derive extensions to this framework for MLSEMs and probe the degree to which the estimator retains these advantages with small to moderate multilevel samples. The estimator emerges as a useful alternative or complement to maximum likelihood because it often outperforms maximum likelihood in small to moderate multilevel samples in terms of convergence, bias, error variance, and power. The proposed estimator is implemented as a function in *R* using *lavaan* and is illustrated using a multilevel mediation example.

## Keywords

Croon, limited information estimator, full information estimator, maximum likelihood, structural equation models, unreliability, measurement error, multilevel models, factor score regression

---

<sup>1</sup>University of Cincinnati, Cincinnati, OH, USA

<sup>2</sup>University of North Carolina at Charlotte, Charlotte, NC, USA

<sup>3</sup>University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

## Corresponding Author:

Benjamin Kelcey, University of Cincinnati, 2820 McMicken Circle, Cincinnati, OH 45221, USA.

Email: [ben.kelcey@gmail.com](mailto:ben.kelcey@gmail.com)

Organizational research often involves multilevel theories that probe complex systems of latent variables using observable but fallible indicators subject to measurement error. A common approach to operationalize such multilevel structures and address the measurement error associated with fallible indicators is to construct and connect multiple indicator factor models using multilevel structural equation models (MLSEMs; e.g., Li & Beretvas, 2013). Full information methods such as maximum likelihood (ML) estimation of all parameters has been shown to be an effective strategy to address the deleterious effects of measurement error (e.g., bias, inaccurate inferences, erroneous estimates of power) in large sample settings (e.g., Cheung & Lau, 2008, 2017; Cole & Preacher, 2014; Ledgerwood & Shrout, 2011; Li & Beretvas, 2013).

MLSEMs are however often heavily parameterized and, as a result, ML estimation typically requires large sample sizes at each level to dependably deliver stable and admissible solutions that provide unbiased estimates of the parameters and their standard errors (Li & Beretvas, 2013; McNeish, 2017). For instance, research on mediation with MLSEM has demonstrated that ML frequently encounters convergence or estimation errors and produces biased parameter estimates and standard errors even in simple models with samples of 40 organizations and 20 individuals per organization (e.g., Li & Beretvas, 2013). This research has established that a minimum of 80 to 100 organizations are required to ensure model convergence with simple multilevel models and suggested even larger samples are needed for ML to dependably deliver nearly unbiased parameter estimates and standard errors across common conditions (Gagne & Hancock, 2006; Hox & Maas, 2001; Li & Beretvas, 2013; McNeish, 2017).

Yet, substantive literature has emphasized that the benefits of studying theories, interventions, and intermediate processes are not limited to large-scale studies—small- to moderate-scale studies offer critical contributions to theory and social issues when they are well executed (e.g., Bodner & Bliese, 2018; Walton, 2014). In many areas of research, samples of fewer than 80 organizations are typical and samples greater than 80 may be considered prohibitively large (e.g., Schochet, 2011; Spybrook, Shi, & Kelcey, 2016). Perhaps because of the mismatch between the scale of many multilevel studies and the large-scale requirements of ML estimation in MLSEM, literature reviews have reported a widespread absence in the appropriate adjustment for measurement error (e.g., Aguinis, Edwards, & Bradley, 2017). This lack of treatment of measurement error may be in part attributable to the lack of suitable methods available for small- to moderate-scale studies rather than researcher oversight.

A historical alternative to ML estimation of MLSEM parameters is a type of limited information factor score path analysis. This approach reduces MLSEMs to the more parsimonious multilevel path models by using single equation ML to estimate each of the factor models and then substitutes the resulting factor scores in place of the measurement models to estimate the structural model parameters (Cheung & Lau, 2017; Devlieger, Mayer, & Rosseel, 2016; Devlieger & Rosseel, 2017; Hoshino & Bentler, 2013; Li & Beretvas, 2013; Lu, Kwan, Thomas, & Cedzynski, 2011; McNeish, 2017). This type of multilevel path analysis with factor scores has been shown to outperform ML in MLSEM in terms of power and in the stability and convergence of estimates but is well known to produce biased estimates of path coefficients and sampling variance when measurement error is present (Li & Beretvas, 2013).

Contemporary advances in single-level designs have, however, shown that the nature of this bias can be tracked and corrected by taking into account the results of the measurement model error associated with each latent construct (Croon, 2002). One recent development in this framework is Croon's bias-corrected version of factor score path analysis (Devlieger et al., 2016; Devlieger & Rosseel, 2017).

Croon's bias-corrected factor score path analysis can be implemented through four general steps: (a) separately estimate factor models and scores for each latent variable (e.g., for covariates, mediator, outcome) using ML; (b) estimate the variance-covariance matrix of the factor scores (e.g.,

covariates, mediator, outcome) and any observed variables (e.g., treatment status); (c) use the results of the measurement models in (a) to construct a method of moments type correction and use it to adjust the variance-covariance matrix; and (d) estimate the path model using the corrected variance-covariance matrix.

This approach parallels the conventional but suboptimal factor score path analysis method. However, it differs in one key way—the conventional path analysis approach uses the factor score variance-covariance matrix to estimate structural relationships whereas Croon’s approach introduces a method of moments adjustment based on the measurement model—in step (c) above—to correct the factor score variance-covariance matrix and obtain a consistent estimate of the variance-covariance matrix.

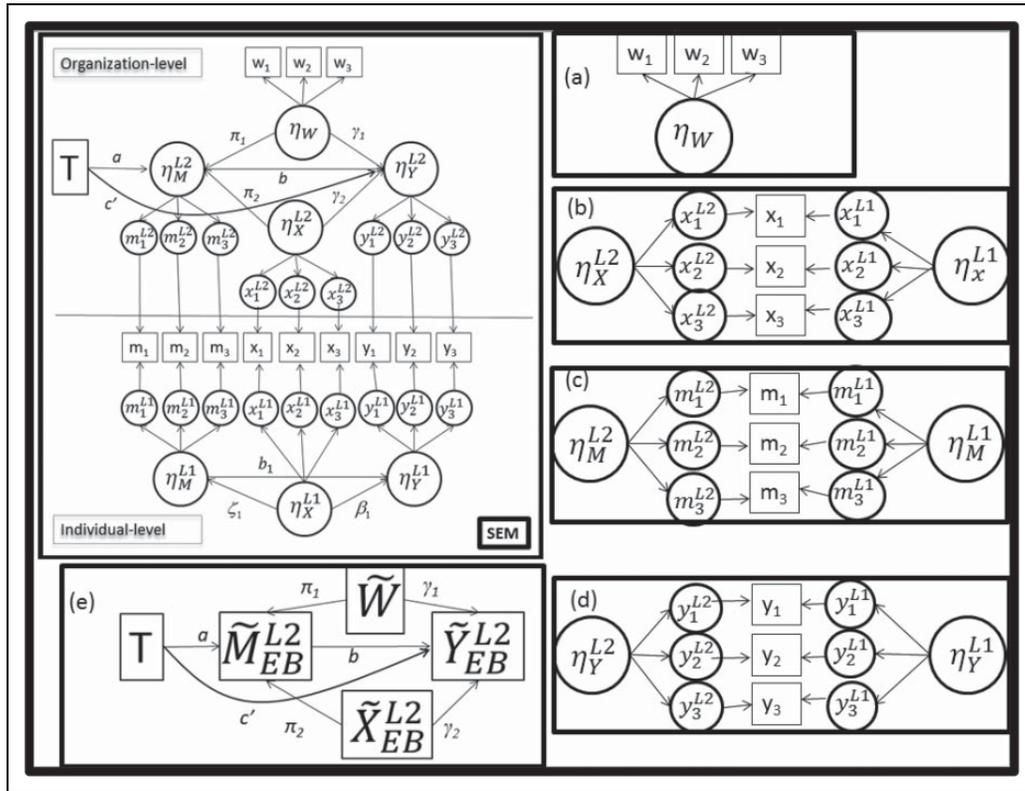
Recent developments have provided a theoretical and empirical basis for the application of Croon’s bias-corrected estimator to a broad range of settings and have identified the estimator as an important small to moderate sample estimation alternative for SEMs (Devlieger et al., 2016; Devlieger & Rosseel, 2017; Loncke et al., 2018). For instance, past literature with single-level structural equation models has shown that Croon’s estimator outperforms ML in a wide variety of settings including small to moderate samples, missing data, measurement misspecifications, structural misspecifications, and correlated error terms (Devlieger et al., 2016; Devlieger & Rosseel, 2017; Hayes & Usami, 2019; Kelcey, 2019; Loncke et al., 2018; Lu et al., 2011). Given the potential analytic advantages of Croon’s limited information estimator in buttressing evidence from small- to moderate-scale studies, an open set of questions is if and how these methods can be extended to multilevel settings where governing sample sizes (i.e., the number of organizations) tend to be small to moderate and model complexity tends to be high.

In this study, we extend the scope of Croon’s bias-corrected estimator by deriving its corrections for multilevel studies. We detail the estimation of relationships within a multilevel system. We then take up a case study of multilevel mediation to anchor and apply the developments and provide an *R* function to implement Croon’s method using the *lavaan* package. Finally we assess of the absolute and relative performance of the estimator with multilevel mediation models using a Monte Carlo simulation and conclude with a discussion of the results, implications and limitations.

## Working Example

We detail the development of the multilevel extension of Croon’s estimator through its implementation in a multilevel mediation application that is graphically depicted in the top-left frame of Figure 1 (notation and models subsequently detailed). We note that although we describe the estimator within the context of a common multilevel mediation model, the framework can accommodate a much wider range of mediation models and MLSEMs. In general, the estimator and method we develop can be applied to a broad range of standard MLSEMs, including, for example, sequential mediation models, multiple mediator models, models with multiple outcomes, models with multiple interventions, and models with combinations of these and other features. We outline some of the limitations and future work of the framework in the discussion section.

Our working example draws on multilevel mediation as it represents one prevalent type of MLSEM in organizational research and in research across almost every area of the social sciences (e.g., Aarons & Sawitzky, 2006; Allen, Herst, Bruck, & Sutton, 2000; Brough & O’Driscoll, 2005; Brough, O’Driscoll, Kalliath, Cooper, & Poelmans, 2009; Brown et al., 2014; Eby, Casper, Lockwood, Bordeaux, & Brinley, 2005; Kelcey, Hill, & Chin, 2019; Ketelaars, 2017; Sampson, Raudenbush, & Earls, 1997; Song, Tsui, & Law, 2009). There are of course many other types of MLSEMs that are common in the organizational sciences and for which the proposed estimator is suitable. However, we focus on multilevel mediation because it provides a clear introduction to the



**Figure 1.** Conceptual representation of a 2-1-1 MLSEM mediation model when the mediator, outcome, and covariate are subject to measurement error (top panel) and the steps to implementing this model using the bias-corrected estimator; panels (a), (b), (c), and (d) capture the respective measurement models for each latent trait in the system, while panel (e) captures the corrected path analysis.

corrections in multilevel settings and it allows us to provide a specific and targeted initial simulation assessment of the performance of the estimator.

Suppose we would like to assess the mediation effect of an organizational policy promoting a family-friendly workplace culture (organization-level intervention) on employee self-reported organizational commitment (individual-level outcome) as it operates through employee self-reported work-life balance (individual-level mediator; e.g., Brough et al., 2009; Eby et al., 2005). Let us consider an experiment that randomly assigns the family-friendly policy to organizations and is designed to assess the extent to which the policy improves organizational commitment by improving work-life balance—both of which are assessed with error using individual-level indicators.

Although random assignment of the policy directly addresses the potential for confounding variables regarding the main effect of the policy on employee organizational commitment, it does not address the potential for variables that confound the relationship between the mediator and the outcome (i.e., sequential ignorability; e.g., VanderWeele, 2010). Omission of variables that confound the mediator-outcome relationship can bias path coefficients and this bias can be further amplified when ignoring measurement error (Fritz, Kenny, & MacKinnon, 2016). In order to plausibly identify the mediation effect, we must adjust for measurement error and for variables that confound the relationship between the mediator and outcome. For this reason, our example additionally includes latent covariates at both levels—an organization-level covariate assessed with error

using organization-level indicators ( $W$ ) and an individual-level covariate measured with error using individual-level indicators ( $X$ ). Figure 1 provides an illustration that corresponds to this working example. Our subsequent analysis makes the assumption that conditioning on these covariates resolves confounding bias and further correctly identifies the mediation effect. However, we note that this assumption requires careful consideration in practice, and we probe the extent to which our results are sensitive to this assumption in the simulation.

## Model

We draw on a MLSEM formulation that uses multilevel and single-level common factor models (e.g., Preacher, Zyphur, & Zhang, 2010). We use the single-level common factor model (described below) for the organization-level latent covariate ( $\eta_W$ ) and the multilevel factor model (described below) for the individual-level outcome (decomposed into  $\eta_Y^{L2}$  and  $\eta_Y^{L1}$ ), mediator (decomposed into  $\eta_M^{L2}$  and  $\eta_M^{L1}$ ), and covariate (decomposed into  $\eta_X^{L2}$  and  $\eta_X^{L1}$ ). The structural model then organizes the variables such that the mediator path model is

$$\begin{aligned}\eta_{M_{ij}}^{L1} &= \zeta_0 + \zeta_1 \eta_{X_{ij}}^{L1} + \varepsilon_{ij}^M \quad \varepsilon_{ij}^M \sim N(0, \sigma_{M_1}^2) \\ \eta_{M_j}^{L2} &= \pi_0 + aT_j + \pi_1 \eta_{W_j} + \pi_2 \eta_{X_j}^{L2} + u_j^M \quad u_j^M \sim N(0, \tau_{M_1}^2)\end{aligned}\quad (1)$$

For variation among organizations ( $\eta_{M_j}^{L2}$ ),  $T_j$  is a treatment indicator for organization  $j$ ,  $a$  is the path coefficient capturing the impact of the policy on the mediator,  $\pi_1$  and  $\pi_2$  are the path coefficients for the latent organization-level covariate ( $\eta_{W_j}$ ) and the organization-level component of the individual-level covariate ( $\eta_{X_j}^{L2}$ ),  $\pi_0$  is the intercept, and  $u_j^M$  is the organization-specific random effect on the mediator. Similarly, for variation among individuals within organizations ( $\eta_{M_{ij}}^{L1}$ ),  $\zeta_1$  is the path coefficient for the individual-level component of the individual-level covariate ( $\eta_{X_{ij}}^{L1}$ ),  $\zeta_0$  is the intercept, and  $\varepsilon_{ij}^M$  is the individual-specific error for the mediator. Similarly, the outcome path model is

$$\begin{aligned}\eta_{Y_{ij}}^{L1} &= \beta_0 + b_1 \eta_{M_{ij}}^{L1} + \beta_1 \eta_{X_{ij}}^{L1} + \varepsilon_{ij}^Y \quad \varepsilon_{ij}^Y \sim N(0, \sigma_{Y_1}^2) \\ \eta_{Y_j}^{L2} &= \gamma_{00} + b \eta_{M_j}^{L2} + c' T_j + \gamma_1 \eta_{W_j} + \gamma_2 \eta_{X_j}^{L2} + u_j^Y \quad u_j^Y \sim N(0, \tau_{Y_1}^2)\end{aligned}\quad (2)$$

We use  $b$  as the path coefficient for the organization-level component of the latent mediator ( $\eta_{M_j}^{L2}$ ),  $c'$  as the path coefficient for the direct effect of the policy on the outcome,  $\gamma_1$  and  $\gamma_2$  as path coefficients for the organization-level latent covariate ( $\eta_{W_j}$ ) and the organization-level component of the individual-level covariate ( $\eta_{X_j}^{L2}$ ),  $\gamma_{00}$  is the intercept, and  $u_j^Y$  as the organization-specific random effect. For variation among individuals in the latent outcome ( $\eta_{Y_{ij}}^{L1}$ ),  $\beta_0$  is the intercept,  $b_1$  is the path coefficient for the individual-level component of the latent mediator ( $\eta_{M_{ij}}^{L1}$ ),  $\beta_1$  is the path coefficient for the individual-level component of the individual-level latent covariate ( $\eta_{X_{ij}}^{L1}$ ), and  $\varepsilon_{ij}^Y$  as the individual-level error.

Within this framework, we use the  $a$  path coefficient from the mediator model to quantify the changes in work-life balance brought about by a new family-friendly workplace policy and the  $b$  coefficient to quantify how changes in work-life balance and the new organizational culture are conditionally associated with improved organizational commitment. In our example, we assess the mediation effect ( $ME$ ) that quantifies the shared organization-level covariance among the organizational policy, work-life balance (mediator) and organizational commitment (outcome) at the organization-level. The effect is typically estimated using the product of the organization-level path coefficients for the intervention-mediator path ( $a$ ) and mediator-outcome ( $b$ ) paths:  $ME = ab$ .

## Croon's Bias-Corrected Estimator

We next provide a step by step outline of our developments along with an example analysis using the *lavaan* package (Rosseel, 2012). We first describe the measurement models and factor scores, and considerations for extending Croon's estimator to a multilevel framework—steps (a) and (b) above. We then detail the resulting multilevel corrections—step (c) above. Last, we discuss the path analysis with the bias-corrected estimator—step (d) above—and end with inferential considerations.

### Measurement Models

We begin by delineating step (a)—we detail how to fit separate factor models to each latent variable using ML and then follow with the prediction of the respective factor scores (or the covariance matrix of the factor scores; see a-d in Figure 1). Within the context of multilevel structures, we describe two primary types of measurement models that arise—single-level and multilevel factor models. In our two-level setting, single-level factor models are primarily used when we draw on organization-level indicators (i.e., constant within an organization) to measure a latent variable that varies only among organizations whereas multilevel factor models are employed when we draw on individual-level indicators that vary within and across organizations in order to track a latent variable that potentially varies both within and across organizations. We outline these models below and the considerations that help us to subsequently operationalize the bias-corrected estimator in multilevel settings.

**Organization-level variables.** We start by considering a single-level common factor model for organization-level latent variables that draws on observed indicators ( $w$ ). For instance, our measurement model for the latent covariate ( $\eta_w$ ) in the 2-1-1 example depicted in panel (a) of Figure 1 is

$$w_j = \mu_w + \Lambda_w \eta_{w_j} + \epsilon_j^w \quad (3)$$

We use  $j$  to index organizations,  $w_j$  as the observed indicators for the latent covariate  $\eta_w$ ,  $\Lambda_w$  as the factor loadings,  $\mu_w$  as the indicator intercepts,  $\epsilon_j^w$  as the error terms, and we set the scale by fixing the variance of  $\eta_{w_j}$  to one (or alternatively fixing the loading of the first indicator to one). Having fit a factor model for the organization-level latent variable using ML, we then follow with the prediction of the factor scores using, for example, the regression predictor method.

**Individual-level variables.** Now consider an individual-level latent variable such as the outcome ( $\eta_y$ ) depicted in panel (d) of Figure 1. Our multilevel measurement model is

$$y_{ij} = \mu_{y_j} + \Lambda_Y^{L2} \eta_{y_j}^{L2} + \Lambda_Y^{L1} \eta_{y_{ij}}^{L1} + \epsilon_j^{Y^{L2}} + \epsilon_{ij}^{Y^{L1}} \quad (4)$$

We use  $y_{ij}$  as the indicators of individual  $i$  in organization  $j$  for the latent variable,  $\eta_{y_j}^{L2}$  and  $\eta_{y_{ij}}^{L1}$  as the organization- and individual-level components of the latent variable,  $\Lambda_Y^{L2}$  and  $\Lambda_Y^{L1}$  as the organization- and individual-level factor loadings,  $\mu_{y_j}$  as intercepts that vary across organizations, and  $\epsilon_j^{Y^{L2}}$  and  $\epsilon_{ij}^{Y^{L1}}$  as the organization- and individual-level error terms. We set the scale by assigning unit variances to the organization- and individual-level factors (or by fixing the loading of the first indicator at each level to one). We can estimate similar measurement models for the mediator and the individual-level covariate ( $X$ ).

To operationalize the proposed bias-corrected estimator in multilevel settings, we need to estimate the covariances of the factor scores at the organization- and individual-level for pairs of constructs. For variables that capture variation solely among organizations (e.g., the organization-

level covariate  $W$ ), we can use single-level factor scoring methods. However, for variables that vary among and within organizations, we must take into account the organizational or clustered nature of its observed indicators. Conceptually, the organization-level factor scores for a latent variable (e.g., outcome) that varies both within an organization and among organizations can be expressed as

$$\tilde{Y}_j^{L2} = A_{Y_1}^{L2} y_{1j}^{L2} + A_{Y_2}^{L2} y_{2j}^{L2} + A_{Y_3}^{L2} y_{3j}^{L2} \quad (5)$$

Here we use  $\tilde{Y}_j^{L2}$  as the predicted factor score for organization  $j$ ,  $y_j^{L2}$  as the organization-level components or random intercepts of the indicators and  $A_{Y_i}^{L2}$  as the organization-level factor score matrix weights. When the variance of the factor is set to one, the  $A_{Y_i}^{L2}$  regression weights are

$$A_{Y_i}^{L2} = C_{y^{L2}}^{-1} \Lambda_{Y_i}^{L2} \quad (6)$$

with  $C_{y^{L2}}^{-1}$  as the inverse of the organization-level covariance matrix of the outcome indicators and  $\Lambda_{Y_i}^{L2}$  as the organization-level factor loadings from the multilevel measurement model in Equation 4.

When predicting organization-level factor scores for the organization-level component ( $\tilde{Y}_j^{L2}$ ) of a latent variable such as the outcome,  $C_{y^{L2}}^{-1}$  and  $\Lambda_{Y_i}^{L2}$  are directly available from the multilevel factor model. However, the organization-level components of the individual-level indicators ( $y^{L2}$ ) are also required but are themselves latent variables. As a result, the method requires suitable predictions of  $y^{L2}$  to operationalize the estimator. That is, in single-level studies the values of the indicators are observed and can be directly used to construct factor scores; with multilevel latent variables, however, the values of the indicators at the organization level (e.g.,  $y^{L2}$ ) are not directly observed and must also be predicted.

There are multiple univariate and multivariate approaches one could adopt to predict the organization-level values of indicators. Multivariate approaches that simultaneously decompose all of a factor's indicators potentially strengthen predictions because they leverage the interrelations of the indicators. However, such approaches are potentially susceptible (perhaps to a lesser degree) to the same types of convergence and estimation errors associated with ML because of their complexity relative to sample size. Below we outline univariate approaches that accept issues of factor score indeterminacy and uncertainty in exchange for less complexity and potentially fewer convergence issues. However, the function also allows for multivariate approaches.

We considered two univariate predictors of the organization-level components of indicators—organizational means and empirical Bayes means. Organizational means simply estimate the cluster-level components of the indicators with the observed organization-specific means (denoted  $\bar{y}_j$ ). This approach picks up both the organization-level variation ( $y^{L2}$ ) and a piece of the individual-level variation ( $y^{L1}$ ) that is proportional to the number of individuals sampled per organization (Muthén, 1991).

An alternative approach is to predict the organization components of the indicators using an empirical Bayes estimator. Empirical Bayes prediction forms precision weighted means on the basis of the reliability of organizational means (e.g., Raudenbush & Bryk, 2002). For instance, for the first indicator, an empirical Bayes prediction ( $y_{1j}^{L2EB}$ ) can be obtained as

$$y_{1j}^{L2EB} = \left( \frac{\tau_{y_1}^{L2}}{\tau_{y_1}^{L2} + \sigma_{y_1}^{L1}/n_{1j}} \right) \bar{y}_{1j} + \left( 1 - \frac{\tau_{y_1}^{L2}}{\tau_{y_1}^{L2} + \sigma_{y_1}^{L1}/n_{1j}} \right) \bar{y}_{1..} = R_{y_{1j}} \bar{y}_{1j} + (1 - R_{y_{1j}}) \bar{y}_{1..} \quad (7)$$

where  $\bar{y}_{1..}$  is the overall mean for the indicator,  $\tau_{y_1}^{L2}$  and  $\sigma_{y_1}^{L1}$  are its variances at the organization and individual levels, and  $R_{y_{1j}}$  is the reliability of the organization-level scores for the first indicator for organization  $j$  (note that  $R_{y_{1j}} = R_{y_1}$  when the number of individuals per organization are equal  $n_{1j} = n_1$ ). The empirical Bayes approach provides consistent estimates but still incurs finite sample

bias that is proportional to the intraclass correlation coefficient and the number of individual sampled per organization.

Multivariate versions of this approach can also be used. For example, for a given set of indicators (e.g.,  $y_1$ ,  $y_2$ , and  $y_3$ ), we can estimate the organization-level covariance matrix ( $\mathbf{T}$ ) and the individual-level covariance matrix ( $\mathbf{V}$ ). We then can then construct a multivariate reliability matrix ( $\mathbf{\Omega}_{y_j}$ ) as

$$\mathbf{\Omega}_{y_j} = \mathbf{T}(\mathbf{T} + \mathbf{V}/n_{1j})^{-1} \quad (8)$$

with empirical Bayes multivariate predictions ( $\mathbf{y}_j^{L2EB}$ ) as

$$\mathbf{y}_j^{L2EB} = \mathbf{\Omega}_{y_j} \bar{\mathbf{y}}_j + (\mathbf{I} - \mathbf{\Omega}_{y_j}) \bar{\mathbf{y}} \quad (9)$$

### Corrections

We next detail the core machinery of the proposed estimator—the covariance-based corrections. The factor score-based covariance matrices developed above will be biased because they neglect the uncertainty introduced by representing the latent constructs and indicators with factor scores (e.g., Bollen, 1989). We derive a Croon (2002) type method of moments correction that adjusts the covariances for the unaccounted measurement error to produce consistent estimates of the covariance matrices at each level.

The proposed estimator is similar in character to two-stage ML approaches (e.g., Savalei & Bentler, 2009), stepwise estimators developed for latent class models (e.g., Bakk & Vermunt, 2015), or the (improved) regression calibration approaches (e.g., Skrondal & Kuha, 2012) because it introduces a correction factor that accounts for the attenuating effects of the uncertainty in the factor scores stemming from both the measurement and the multilevel components. More conceptually, the corrections detailed below parallel classical test theory disattenuation formulas for the correlation between two unreliably measured constructs—principally the corrections we derive leverage the unreliabilities of the latent variables to adjust for the impact of measurement error on a covariance (Spearman, 1904). For example, the classical test theory adjustment disattenuates the correlation between two measurement error prone variables by dividing it by the square root of the product of the variable reliabilities. In MLSEM these reliabilities (and their uncertainties) are unknown and must be estimated through complex functions of the parameters in the measurement models. The corrections we propose in this study conceptually extend the classical test theory style adjustments because they first identify the expected relationship among the covariance of latent variables, the covariance of their factor scores, and the unreliabilities of the factors, and second develop empirical estimates of those unreliabilities using the estimated factor models. In turn, the proposed estimator blends the estimation stability of factor score path analysis with the measurement error-corrections typically associated with ML.

Below we outline the nature of these corrections (see Supplemental Material Part 2 for details) as classified by the levels of pairs of variables. We outline the corrections for (a) the covariance between two organization-level variables, (b) the covariance between an organization- and individual-level variable, (c) the organization-level covariance between two individual-level variables, and (d) the individual-level covariance between two individual-level variables.

**Organization-level variables.** The first case we consider is the correction associated with the covariance between two organization-level latent variables. Consider the covariance between two organization-level latent variables (e.g.,  $\eta_A$  and  $\eta_B$ ) as a function of the covariance between their organization-level covariate factor scores ( $\bar{A}$  and  $\bar{B}$ ) under the regression predictor method. Once we have the covariance

between the factor scores ( $\text{cov}(\tilde{A}, \tilde{B})$ ) based on the aforementioned single-level factor models, a corrected estimate of the covariance between the latent constructs ( $\text{cov}(\eta_A, \eta_B)$ ) can be obtained as

$$\text{cov}(\eta_A, \eta_B) = \frac{\text{cov}(\tilde{A}, \tilde{B})}{\mathbf{A}_A \mathbf{\Lambda}_A \mathbf{\Lambda}_B' \mathbf{A}_B'} \quad (10)$$

where  $\mathbf{A}$  is a factor score matrix and  $\mathbf{\Lambda}$  is a factor loading matrix. In essence and form, the correction replicates the adjustment necessary for single-level structural equation modeling (Devlieger et al., 2016). Equation 10 indicates that a corrected estimate of the covariance between two organization-level latent variables can be obtained by dividing the covariance of their factor scores by the matrix product of their factor score and loading matrices. More conceptually, the core intuition here is that the product of the factor score and loading matrices for each latent variable (e.g.,  $\mathbf{A}_A \mathbf{\Lambda}_A$ ) provides a type of model-based estimate of the latent variable's reliability. In turn, a disattenuated estimate of the covariance between two organization-level latent variables can be obtained by dividing the covariance of the factor scores by the product of the empirical reliabilities ( $\mathbf{A}_A \mathbf{\Lambda}_A \mathbf{\Lambda}_B' \mathbf{A}_B'$ ).

In our working 2-1-1 mediation example (Figure 1), consider the covariance between the organization-level latent covariate ( $\eta_W$ ) and the manifest treatment variable ( $T$ ). For manifest variables, adjustments to covariances are based solely on the factor model of the covarying latent variable. In our example, the correction for the covariance is reduced to adjusting only for the factor score and loading matrices of the latent covariate such that

$$\text{cov}(\eta_W^{L2}, T) = \frac{\text{cov}(\tilde{W}, T)}{\mathbf{A}_W \mathbf{\Lambda}_W \mathbf{\Lambda}_T' \mathbf{A}_T} = \frac{\text{cov}(\tilde{W}, T)}{\mathbf{A}_W \mathbf{\Lambda}_W} \quad (11)$$

*Organization- with individual-level variables.* Next, we detail the corrections for cross-level covariances—that is, the covariance between an organization- and individual-level latent variable. In our working 2-1-1 mediation example (Figure 1), consider the covariance between the organization component of outcome latent variable ( $\eta_Y^{L2}$ ) and the organization-level covariate ( $\eta_W$ ) as a function of the covariance their factor scores ( $\text{cov}(\tilde{Y}_{EB}^{L2}, \tilde{W})$ ) under the aforementioned univariate empirical Bayes approach for the indicators. For such cross-level relationships, the correction is (see Supplemental Material Part 2.1).

$$\text{cov}(\eta_Y^{L2}, \eta_W) = \frac{\text{cov}(\tilde{Y}_{EB}^{L2}, \tilde{W})}{\mathbf{A}_Y^{L2} \mathbf{R}_y \mathbf{\Lambda}_Y^{L2} \mathbf{\Lambda}_W' \mathbf{A}_W'} \quad (12)$$

Here  $\text{cov}(\eta_Y^{L2}, \eta_W)$  is the covariance between the organization-level component of an individual-level latent variable and an organization-level latent variable,  $\text{cov}(\tilde{Y}_{EB}^{L2}, \tilde{W})$  is the covariance of the estimated factor scores when using empirical Bayes estimates of the organization-level components of the indicators,  $\mathbf{A}$  is a factor score matrix, and  $\mathbf{\Lambda}$  is a factor loading matrix. Further,  $\mathbf{R}_y$  (or  $\mathbf{\Omega}_{y_j}$  if multivariate) is a vector (or matrix) of the average outcome indicator reliabilities. For example, under the univariate approach the average outcome indicator reliability for the first indicator of the outcome is

$$R_{y_1} = \frac{1}{n_2} \sum_{j=1}^{n_2} \left( \frac{\tau_{y_1^{L2}}^2}{\tau_{y_1^{L2}}^2 + \sigma_{y_1^{L1}}^2 / n_{1j}} \right) = \frac{1}{n_2} \sum_{j=1}^{n_2} R_{y_{1j}} \quad (13)$$

Conceptually, the results simply indicate that in addition to adjusting for the unreliability of the factors (as in the case for single-level or with two organization-level variables), we must further

correct for the unreliability of the indicators (e.g., using  $\mathbf{R}_y$  or  $\mathbf{\Omega}_y$ ). Similar derivations regarding the covariance between an organization- and an individual-level latent variable under the organizational means approach shows that its correction parallels that of the Empirical Bayes approach but does not require additional adjustment for the indicator unreliability—only the corrections in Equation 10 are necessary (see Supplemental Material Part 2.1).

*Organization-level component of individual-level variables.* We next consider the organization-level covariance between two latent variables measured at the individual level. When considering the covariance of the organization-level components of an outcome and mediator assessed at the individual level (denoted as  $\text{cov}(\eta_Y^{L2}, \eta_M^{L2})$ ) using the empirical Bayes estimates of the indicators, the resulting correction is (see Supplemental Material Part 2.2)

$$\text{cov}(\eta_Y^{L2}, \eta_M^{L2}) = \frac{\text{cov}(\tilde{Y}_{EB}^{L2}, \tilde{M}_{EB}^{L2})}{\mathbf{A}_Y^{L2} \mathbf{R}_y \mathbf{\Lambda}_Y^{L2} \mathbf{\Lambda}_M^{L2} \mathbf{R}_m \mathbf{A}_M^{L2}} \quad (14)$$

Here  $\text{cov}(\eta_Y^{L2}, \eta_M^{L2})$  is the covariance between the organization-level components of an outcome and mediator assessed with individual-level indicators,  $\text{cov}(\tilde{Y}_{EB}^{L2}, \tilde{M}_{EB}^{L2})$  is the covariance of the estimated factor scores when using a univariate empirical Bayes estimates,  $\mathbf{A}^{L2}$  and  $\mathbf{\Lambda}^{L2}$  are the factor score matrix and loading matrix at the organization level, and  $\mathbf{R}$  are the respective vectors of indicator reliabilities (replaced by  $\mathbf{\Omega}_y$  in the multivariate approach). Once again, the results suggest that Croon's bias-corrected estimator can be adapted by adjusting for the unreliability of the indicators in both the outcome ( $\mathbf{R}_y$ ) and mediator ( $\mathbf{R}_m$ ).

*Individual-level component of individual-level variables.* In complement to previous type of organization-level covariance, we can consider the individual-level covariance between two latent variables measured at the individual level. This covariance arises when, for example, we are interested in assessing the within organization relationship between the latent mediator and outcome (labeled  $b_1$ ). The correction for the covariance of the individual-level components of an outcome and mediator assessed at the individual level ( $\text{cov}(\eta_Y^{L1}, \eta_M^{L1})$ ) is the same as that in the single level case (i.e., Equation (10)).

Having corrected the covariances, we can then adjust the variance of each latent variable. When setting the scale of a latent variable by fixing its variance to unity, we can replace the variance of the factor scores for a given latent variable with this scaling (i.e., set factor variances to one). When fixing the loading of the first indicator to unity, we replace the variance of the factor scores with the latent variance estimate obtained in the factor model (e.g., for the outcome, we use the estimated level two variance from model (4)).

### ***Bias-Corrected Path Analysis***

Having formed corrected covariance matrices, we can then separately estimate the organization- and individual-level structural models described in Equation 1 using separate conventional single-level path analyses or a combined multilevel path analysis on the corrected covariance matrices.

### ***Inference***

In order to draw statistical inferences regarding parameter values, we need to track their error variance. Standard model-based approaches that use the information matrix to estimate the error variance fail in the current context because they do not account for the additional uncertainty that

arises from measurement error and its corrections. To address this limitation, we can use bootstrap-based methods (e.g., Efron & Tibshirani, 1993). There are several potentially applicable implementations of the bootstrap in the current context, but one simple approach is a parametric bootstrap based on the estimates obtained in the original analysis.

The flow of this resampling approach is as follows: (a) sample new values of the latent variables based on multivariate normality with covariances taken from the corrected covariance matrix, (b) sample new values of the latent variable indicators using the estimated measurement model parameters, (c) reestimate the MLSEM using Croon's method, and (d) replicate (a) to (c) a sufficient number of times to establish stable parameter distributions. We then draw on percentile-based confidence intervals to test hypotheses or form confidence intervals.

## Illustration

Continuing with our working example, suppose we would like to assess the indirect effect of an organizational policy promoting a family-friendly workplace culture (organization-level intervention) on employee self-reported organizational commitment (individual-level outcome) as it operates through employee self-reported work-life balance (individual-level mediator; e.g., Brough et al., 2009; Eby et al., 2005). Let us consider data that draws on a sample of 30 organizations with 5 employees per organization (150 total individuals).

To begin our example analysis in *R*, we download call the `bcfspace` function and download the data (see Supplemental Material) and read it into *R* using

```
d <- read.table(file="211 example.dat", h=TRUE)
```

We then specify the model depicted in Figure 1 using *lavaan* syntax. We identify the scales of the latent variables by fixing the loading of the first indicator to one yielding a specification of

```
sem1model <- '
  level: 1
  fyL1 =~ y1+y2+y3
  fmL1 =~ m1+m2+m3
  fxL1 =~ x1+x2+x3
  fyL1 ~ fmL1+fxL1

  level: 2
  fxL2 =~ x1+x2+x3
  fyL2 =~ y1+y2+y3
  fmL2 =~ m1+m2+m3
  fw =~ w1+w2+w3
  fyL2 ~ fmL2+T+fxL2+fw
  fmL2 ~ T+fxL2+fw'
```

Alternatively, we could free the loading of the first indicator for each factor at both levels and constrain the variance for each factor at both levels to be one. Having specified the model, we can then estimate the parameters using ML with

```
semML <- sem(sem1model, data = d, cluster="id")
```

To implement the multilevel version of Croon's estimator detailed in this study, we can call the `bcfspa()` function. The model and argument specification of the `bcfspa()` parallels that of the `sem()` function in the *lavaan* package

```
croon1 <- bcfspa(sem1model, data = d, cluster = "id")
```

A summary of the organization-level coefficients can be obtained using, for example, `croon1$`Level 2 Empirical Bayes corrected`[[1]]$`Coefficients Table`[[1]]`

The first argument in the `bcfspa` function (`sem1model`) specifies the model, the `data` argument indicates the data frame, `cluster` defines the level two membership or grouping variable.

In addition, a `method` argument selects either the default empirical Bayes ("eb") or the organizational means ("means") approach for tracking the decompositions of the indicators and a `univariate` argument indicates whether these indicator decompositions should be done in a univariate manner (default) or a multivariate approach, a `levels` argument indicates which level (i.e., "2" or "1") should be output (defaults to both levels), and the `uncorrected.cov` and `corrected.cov` arguments are logical indications of whether the (un)corrected covariance matrix and path coefficients should be returned. For instance, we can also obtain the uncorrected factor score path analysis estimates using

```
croon2 <- bcfspa(sem1model, data = d, cluster = "id", uncorrected.cov = T)
```

```
croon2$`Level 2 Empirical Bayes Uncorrected`[[1]]$`Coefficients Table`[[1]]
```

A different specification draws on metric invariance across levels by constraining indicator loadings to be equal across levels. In this context, indicators are equally good at discriminating among organizations and individuals on the latent variables. We can extend the syntax using the standard *lavaan* code that introducing labels such that

```
sem2model <- '
  level: 1
  fyL1 =~ Ly1*y1+Ly2*y2+Ly3*y3
  fmL1 =~ Lm1*m1+Lm2*m2+Lm3*m3
  fxL1 =~ Lx1*x1+Lx2*x2+Lx3*x3
  fyL1 ~ fmL1+fxL1

  level: 2
  fxL2 =~ Lx1*x1+Lx2*x2+Lx3*x3
  fyL2 =~ Ly1*y1+Ly2*y2+Ly3*y3
  fmL2 =~ Lm1*m1+Lm2*m2+Lm3*m3
  fw =~ w1+w2+w3
  fyL2 ~ fmL2+T+fxL2+fw
  fmL2 ~ T+fxL2+fw'
```

Here, labels such as `Ly1` are used to premultiply the indicators and by using the same label on both levels for an indicator the loadings are constrained to be equal. The equal loading model can be estimated using the `bcfspa()` function without any additional changes

```
croon3 <- bcfspa(sem2model, data = d, cluster = "id")
```

**Table 1.** Organization-Level Measurement and Structural Model Results for 2-1-1 Illustrative Example.

	Coefficient	True	Croon-EB	FS-EB	ML	
Measurement Model	$\lambda_{W_1}^{L2}$	1	1	1	1	
	$\lambda_{W_2}^{L2}$	1	0.99	0.99	1.00	
	$\lambda_{W_3}^{L2}$	1	0.88	0.88	0.89	
	$\lambda_{X_1}^{L2}$	1	1	1	1	
	$\lambda_{X_2}^{L2}$	1	1.28	1.28	1.10	
	$\lambda_{X_3}^{L2}$	1	0.83	0.83	0.76	
	$\lambda_{Y_1}^{L2}$	1	1	1	1	
	$\lambda_{Y_2}^{L2}$	1	0.95	0.95	0.95	
	$\lambda_{Y_3}^{L2}$	1	0.91	0.91	0.92	
	$\lambda_{M_1}^{L2}$	1	1	1	1	
	$\lambda_{M_2}^{L2}$	1	0.99	0.99	0.98	
	$\lambda_{M_3}^{L2}$	1	1.05	1.05	1.09	
		Ave abs bias	—	0.07	0.07	0.06
	Structural Model	a	0.93	0.81	0.68	0.76
		b	0.50	0.47	0.38	-0.13
c'		0.18	0.56	0.72	0.96	
$\gamma_1$		0.11	0.20	0.37	0.50	
$\gamma_2$		0.68	0.66	0.55	0.95	
$\pi_1$		0.68	0.69	0.58	0.54	
$\pi_2$		0.40	0.32	0.26	0.49	
		Ave abs bias	—	0.10	0.22	0.35

Note: ML is concurrent estimation of all parameters using maximum likelihood. FS-EB is factor score path analysis using empirical Bayes. Croon-EB is Croon's bias-corrected method using empirical Bayes.

## Results

The resulting organization-level estimates under the first specification (`croon1`) based on ML, Croon's, and uncorrected factor score path analysis as well as the true coefficient values are displayed in Table 1. In this analysis, we notice few differences among the loading coefficients produced by the different estimators and each is quite close to its true value. In contrast, the path coefficient estimates are highly variable across estimators. Each approach takes on bias, however, Croon's approach returns estimates with the smallest amount of bias across all coefficients. In terms of the mediation effect, for instance, while the true mediation effect was  $ab = 0.93(0.50) = 0.47$ , Croon's returned an estimate of  $0.81(0.47) = 0.38$  whereas ML returned  $0.76(-0.13) = -0.10$ , and the uncorrected method returned  $0.67(0.38) = 0.25$ .

Although the methods differentially incurred bias across structural coefficients, Croon's method had the lowest average absolute bias at 0.10 (relative bias of 21%). In contrast, ML took on nearly four times that with an average absolute bias of 0.35 (relative bias of 71%) while the uncorrected approach took on an average absolute bias of 0.22 (relative bias of 44%). Although this example was purposefully chosen to illustrate the differences among the approaches, these observed discrepancies align well with the subsequent simulation results. More generally, the theoretical and subsequent simulation results of our study and other studies suggest that such discrepancies are common in small to moderate samples.

## Simulation

To provide an initial assessment of the value and precision of the methods developed, we conducted Monte Carlo simulations that outline performance along several criteria. We generated data under the 2-1-1 mediation model described in Equations 1 and 2 and their respective multilevel and single-level measurement models. We first considered an organization-level sample size of 30, an individual-level sample size of 5 per organization, unit error variances for indicators with an intraclass correlation coefficients of 0.20 for multilevel factors, factor loadings of 0.50, 3 indicators per latent variable, and the following standardized path coefficients:  $a = 0.20$ ,  $b = 0.28$ ,  $c' = 0.14$ ,  $\gamma_1 = 0.27$ ,  $\gamma_2 = 0.27$ ,  $\pi_1 = 0.40$ , and  $\pi_2 = 0.40$ , and  $\zeta_1 = 0.50$ ,  $\beta_1 = 0.50$ , and  $b_1 = 0.50$ . We then held constant these parameter values and examined the influence of variation in the organization-level sample size (ranging from 20 to 100), individual-level sample size (ranging from 5 to 100), intraclass correlation coefficients for the indicator error terms (ranging from 0.10 to 0.70), factor loadings (ranging from 0.3 to 0.7), and the number of indicators per latent variable (3 and 5). We assessed the accuracy of the estimators (Croon's, ML, and uncorrected factor score path analysis) under a standardized approach across 1,000 samples in terms of bias, error variance, power, and sensitivity to misspecifications.

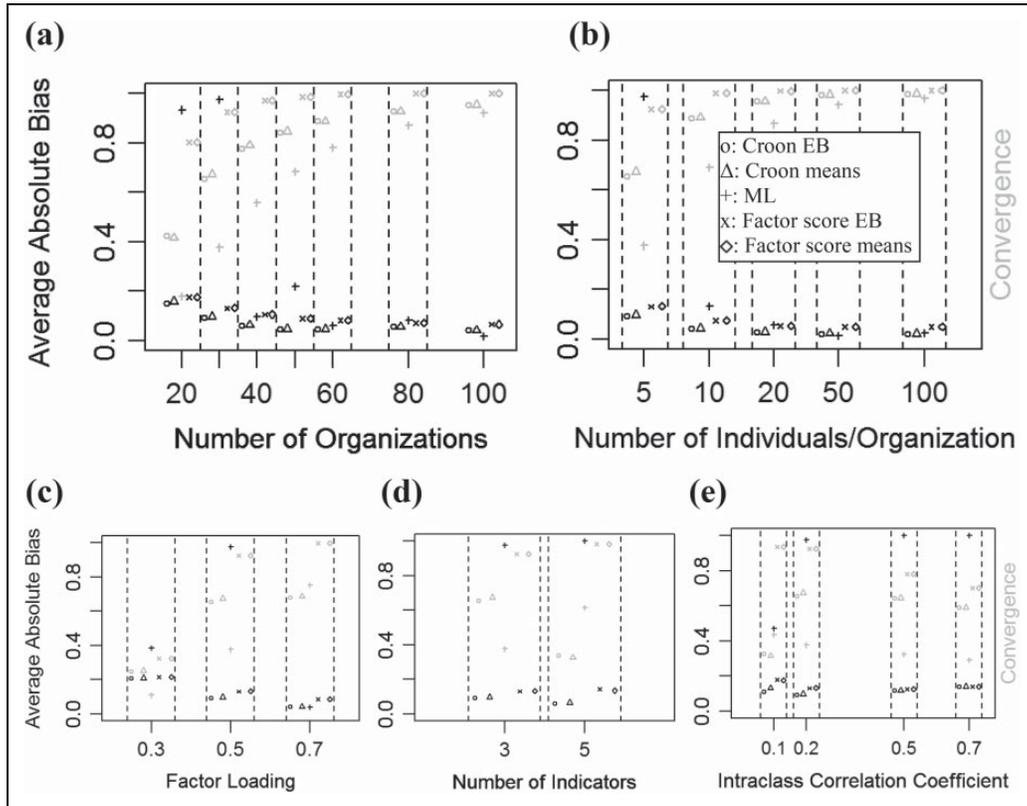
The results of our analyses are outlined in Figures 2 to 5 and in the Supplemental Material Part 1 Tables S1 and S2. Our analyses suggested that under many common conditions the properties of the estimators were similar with larger samples. Convergence and bias were similar for ML and Croon's once the sample sizes reach about 50 organizations or the number of individuals per organization exceeded 20 with at least 30 organizations; however, Croon's approach had lower root mean square error and higher power until samples exceeded 100 organizations with at least 20 individuals per organization. Similarly, differences among the estimators in terms of individual-level coefficients were typically quite small across samples. For these reasons, we outline the results with larger samples in the Supplemental Material (Part 1) and probe the results for organization-level coefficients with small to moderate samples (i.e., 30 to 100 organizations with 5 to 100 individuals per organization) below.

## Convergence

Figure 2 outlines the proportion of samples that converged (gray) for each condition by estimator. For instance, the gray markers in Figure 2 panel (a) identify the convergence rates for each estimator for samples of 20, 30, 40, 50, 60, 80, and 100 organizations (that are separated by vertical dashed lines) holding constant the other parameter values. Likewise, the gray markers in Figure 2 panel (b) identify the convergence rates for each estimator for individual-level sample sizes of 5, 10, 20, 50, and 100 organizations.

The results replicate prior literature and underscore a key advantage of Croon's method—it routinely converges in small to moderate samples when ML fails to converge. For example, ML converged in just 36% of the drawn samples when using 30 organizations and 5 individuals whereas Croon's method converged in 70% of these samples (Figure 2). The differences in convergence rates receded as a function of sample size; however, even when we double the number of organizations to 60, Croon's method (90% convergence) still notably outperformed ML (77% convergence).

A closer look at the results further reveals that Croon's method continued to yield the same relatively low bias parameter estimates when ML sustained estimation or convergence issues. With 30 organizations, for instance, of the 636 (of 1,000) samples that ML failed to converge in, Croon's method converged in 406 of them (64%). That is, in a large majority of the samples where ML encountered convergence problems, Croon's estimator was able to avoid such problems. Increased sample size depreciates this advantage but even with 60 organizations, of the 233 (of 1,000) samples



**Figure 2.** Convergence rate (gray) and average absolute bias (black) by estimator as a function of (a) organizations, (b) individual per organization, (c) factor loadings, (d) number of indicators, and (e) intraclass correlation coefficients of the indicator error terms.

ML failed to converge in, Croon’s method converged in 182 of them (78%). The cross-tabulations for the number of samples the estimators converged in were

$n_2 = 30$	ML		$n_2 = 60$	ML			
		0		1		0	1
	Croon	0		230	92	0	51
		1	406	272	1	182	718

Similar advantages in convergence rates were also observed for uncorrected factor score path analysis. With 30 organizations it converged 90% of the time and with 60 organizations it converged in 100% of the samples. Moreover, the uncorrected approach converged in 90% of the samples ML failed in with 30 organizations and in 99% with 60 organizations.

Discrepancies in convergence among the estimators were heavily influenced by sample size but mostly disappeared once the organization-level sample size reached about 50 or the number of individuals per organization exceeded 20 with at least 30 organizations. However, the intraclass correlation coefficients of the indicator error terms, factor loadings and the number of indicators also influenced convergence rates. As the intraclass correlation coefficients of the indicator error terms approached more extreme values of zero and one, convergence rates dipped. As factor loadings increased, the convergence rate for each estimator improved but did not impact the relative rank

order of the estimators (Figure 2). In contrast, increasing the number of indicators reduced the convergence rate for Croon's but improved the convergence rate for ML (Figure 2).

### **Bias**

Figure 2 further outlines the average absolute bias (black) of the path coefficients by estimator for small to moderate samples (see Supplemental Material Part 1 Tables S1-S2 for additional results). For example, the black markers in Figure 2 panel (a) identify the average absolute bias for each estimator for samples of 20, 30, 40, 50, 60, 80, and 100 organizations (that are separated by vertical dashed lines) holding constant the other parameter values.

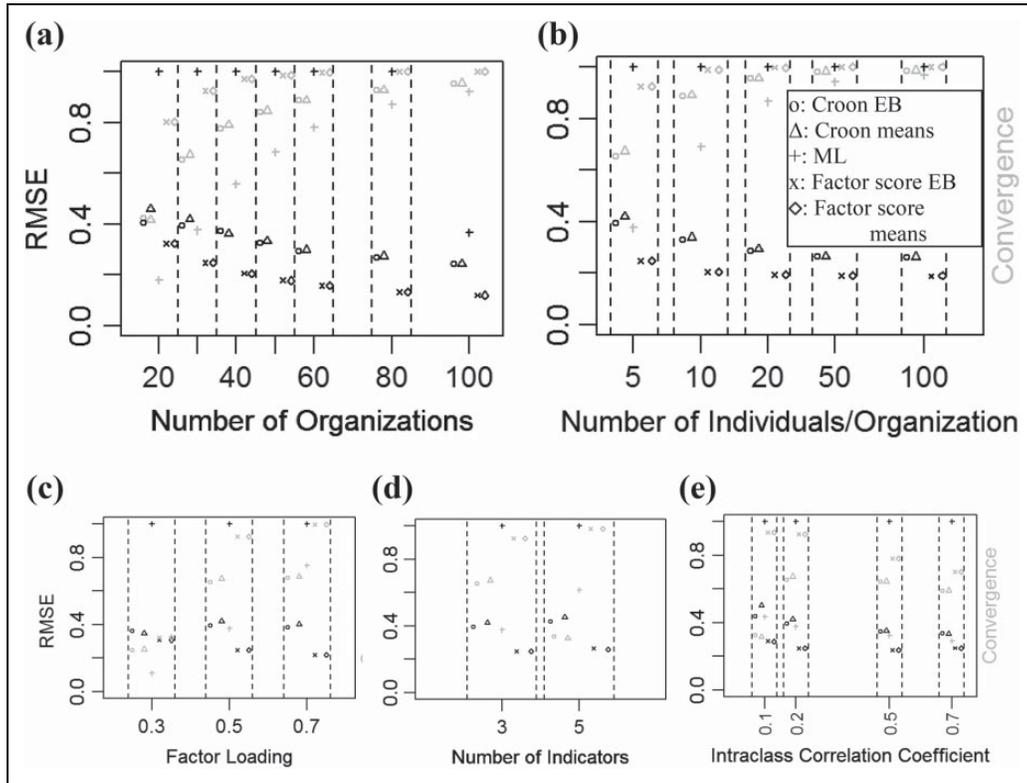
The results extended prior work with Croon's method in single-level settings in that Croon's method often outperformed ML and the uncorrected approach in terms of bias in small to moderate samples. Croon's method incurred some bias but regularly maintained the lowest level of bias across small to moderate samples. The bias of the ML estimator was plagued by convergence issues and intermittent extreme parameter estimates in small to moderate samples that faded in larger samples. Uncorrected factor score path analysis often returned estimates that were in a similar neighborhood as ML and Croon's but was typically more biased because of its neglect of measurement error.

An important practical consideration is the extent to which Croon's method can produce nearly unbiased estimates in samples where ML failed to converge. More specifically, the aforementioned convergence results suggested that a key utility of Croon's method is that it frequently circumvents convergence issues in samples where ML fails to converge. Further probing the intersection of the convergence and bias results indicated that Croon's method retained the same relatively low level of bias in samples where it converged but ML did not.

Returning to our earlier example with 30 organizations—of the 406 samples where ML did not converge but Croon did, the average absolute bias of Croon's method was identical to the overall absolute average bias for all samples (0.09). Similarly, with 60 organizations, of the 182 samples where ML failed to converge but Croon did, Croon's average absolute bias was still only 0.04—again, identical to the overall average absolute bias for all samples. Uncorrected factor score path analysis also maintained its bias level (e.g., 0.13 for 30 organizations) for the subset of samples in which ML (and/or Croon's) failed to converge but these higher convergence rates typically came at the price of increased bias.

Bias across all estimators was influenced by sample size at both levels but this relationship was much more pronounced for ML than for the uncorrected and Croon methods. For instance, the average absolute bias of the path coefficients in samples of 30 organizations and 5 individuals was 0.97 for ML and 0.13 for the uncorrected approach whereas for Croon's method it was only 0.09 (Figure 2). Croon's bias advantage declined as a function of sample size—for instance, with 60 organizations the bias associated with Croon's method was reduced to 0.04 whereas it was reduced to 0.06 with ML and 0.08 with uncorrected factor score path analysis. Paralleling the convergence results, once the organization-level sample size reached about 50 or the number of individuals per organization exceeded 20 with at least 30 organizations, Croon's and ML approached near zero bias while the uncorrected method retained bias proportional to the unreliability of the latent variables.

To a lesser extent, differences in bias among the estimators were also influenced by the other parameters considered but the degree and direction of the influence was not necessarily consistent across estimators (Figure 2). Regarding the different multilevel factor scoring approaches developed, the results suggested that the bias of the univariate empirical Bayes and the organization means approaches were largely indistinguishable. For factor loadings, increases tended to yield lower bias on average. Changes in the intraclass correlation coefficients of the indicator error terms and the number of indicators produced less notable and less consistent changes in bias across the estimators (Figure 2).

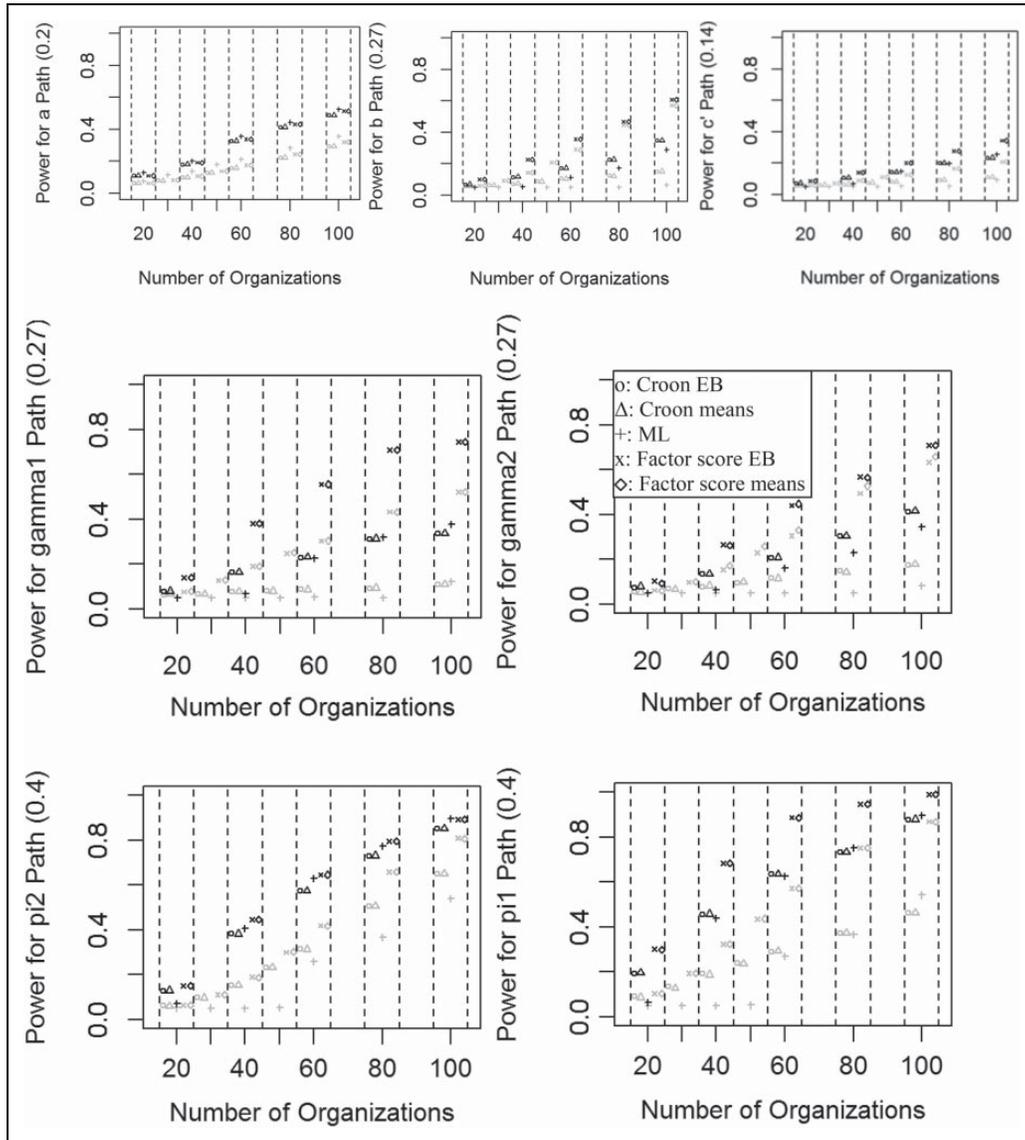


**Figure 3.** Convergence rate (gray) and root mean square error (black) by estimator as a function of (a) organizations, (b) individual per organization, (c) factor loadings, (d) number of indicators, and (e) intraclass correlation coefficients of the indicator error terms.

*Mean square error*

The results also suggested that Croon’s method was more efficient than the ML estimator in samples up to 100 organizations but that Croon’s was less efficient when compared to the uncorrected approach. The reduction in bias associated with Croon’s method was bought with an increase in uncertainty relative to the uncorrected factor score path analysis approach. Because Croon’s method pairs bias reduction with additional error variance, we describe the bias-variance trade-off using the root mean square error of the estimators in Figure 3. Like previous plots, the black markers in Figure 3 panel (a) identify the root mean square error for each estimator for samples of 20, 30, 40, 50, 60, 80, and 100 organizations (that are separated by vertical dashed lines).

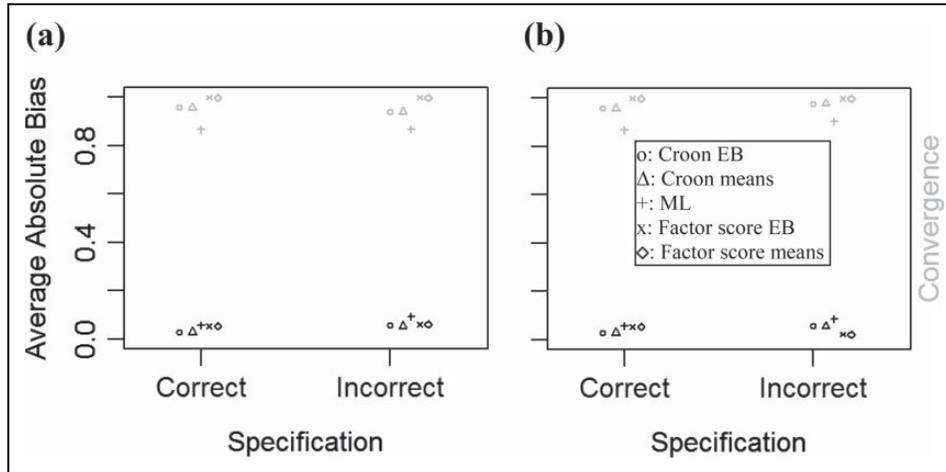
Two results emerged. First, Croon’s method outperformed ML in terms of root mean square error. The distribution of path coefficient estimates under Croon’s method was consistently more concentrated around the true values compared to those of ML and with the exception of sample size, the values of other parameters (e.g., factor loadings) appeared to have little influence. Second, the uncorrected approach consistently outperformed Croon’s method in terms of root mean square error. That is, although the uncorrected approach returned biased estimates, its dispersion is significantly smaller than that of ML or Croon’s. In contrast to the bias results, the root mean square error of the estimators was not comparable until the number of organizations exceeded 100 and 20 individuals per organization.



**Figure 4.** Power as a function of organization sample size (x-axis) by path (indicated on y-axis along with path coefficient). Black symbols indicate an individual-level sample size of 20, while gray symbols indicate an individual-level sample size of 5.

### Power

We surveyed the power with which each method could detect the presence of a nonzero relationship for each path (e.g., Kelcey, Dong, Spybrook, & Cox, 2017; Preacher & Selig, 2012). Figure 4 displays power as a function of organization-level sample size for an individual-level sample size of 5 (gray) and 20 (black). The analyses suggested two findings that parallel the previously outlined results. First, Croon's method was typically more powerful than ML in small to moderate samples but this depended on the path and the sample sizes. However, the uncorrected method was more powerful than Croon's and ML. Second, from an absolute perspective, designing well-powered



**Figure 5.** Convergence rate (gray) and average absolute bias (black) by estimator and (a) measurement misspecification and (b) omitted covariate.

studies will minimally require moderate sample sizes at both levels (e.g., 60 organizations and 20 individuals per organization) and medium to large effects (e.g., 0.3 to 0.5 on a standardized scale)

### Misspecifications

Prior work in single-level SEMs has suggested that Croon's method can be often less sensitive than ML to certain types of model misspecifications because of its stepwise nature (e.g., Devlieger & Rosseel, 2017). We probed two such misspecifications. The first was a misspecification in the measurement model for covariate  $X$  such that the first indicator (e.g.,  $x_1$ ) for the covariate  $\eta_x$  was driven by  $\eta_x$  and  $\eta_w$  (i.e., a type of cross-loading). The results suggested that Croon's method was no more or less sensitive to this type of misspecification than ML. An example summarizing the impact of the misspecification on convergence and bias is presented in Figure 5 for 30 organizations and 20 individuals/organization.

The second scenario we considered was missing variable bias. A key concern in mediation analyses involving latent variables is the potential for bias due to uncontrolled confounding (Fritz et al., 2016). We examined the sensitivity of the estimators to violations of the sequential ignorability assumption in the presence of measurement error by excluding the confounding organization-level latent covariate ( $\lambda_{M_3}^{L2}$ ) and its indicators from our estimation. None of the methods were expected to be robust to missing variable bias, but rather our purpose was to probe the extent to which the intersection of an omitted variable and measurement error may disproportionality accelerate or decelerate the bias of the estimator (Fritz et al., 2016). The results are displayed in Figure 5 for the samples of 30 organizations and 20 individuals/organization. Overall, there was no impact on the convergence rates, but both Croon's and ML experienced a similar level of increased bias while the uncorrected approach did not. The results provide no evidence that Croon's method is more susceptible to omitted variable bias in the presence of measurement error than is ML.

### Discussion

Recent reviews of the literature have suggested that measurement error rarely receives appropriate treatment, and we posited that this oversight may often owe to the mismatch between sample

requirements of extant analytic methods and the complexity and scale of many studies. Yet, literature has clearly documented the deleterious and widespread effects of measurement error. For this reason, we developed multilevel extensions to Croon's bias-corrected factor score-based path analysis estimator that is suitable for MLSEM.

The results of our study and the results of other studies suggest that Croon's estimator is a viable alternative or complementary estimator for small to moderate sample MLSEMs for three primary reasons. First, initial evidence suggests that Croon's estimator is likely to provide less biased estimates of path coefficients relative to ML and the uncorrected method in many common small-to moderate-scale studies.

Second, Croon's estimator more frequently converged in small to moderate samples than ML while also providing the least biased estimates in samples where ML failed to converge. In this way, Croon's estimator emerges as a key alternative or complementary estimator for settings in which ML estimation fails. The results also suggested that the uncorrected approach can play a similar complementary or backstop role—the uncorrected approach had the most reliable convergence rate but did so at a cost of more bias.

Third, Croon's estimator had less sampling variability, lower root mean square error, and tended to be more powerful when compared with ML. This result was somewhat unanticipated because past research has shown that in many cases ML is still quite efficient in finite sample sizes (e.g., Lüdtke et al., 2008; Wooldridge, 2002). However, the results of this study and another recent study have suggested that in finite sample sizes Croon's estimator can be more efficient (Kelcey, 2019). Comparisons between Croon's and the uncorrected approach were more equivocal on this front because while the uncorrected approach returned biased estimates, it had several desirable properties including lower root mean square error and higher power.

Despite the promising initial performance of the proposed methods, this work thus far has several limitations. From an absolute perspective, Croon's method may often retain slightly more power than concurrent ML but may still be underpowered with small to moderate sample sizes depending on the type and size of coefficient targeted and variable reliabilities. Moreover, both of these methods were clearly less powerful than the uncorrected approach. Future research should probe sample size limits and guidelines and the conditions that moderate the absolute and relative levels of power for the estimators.

A second limitation is that the proposed approach needs additional augmentations to generalize to settings with more complicated measurement models. The form of Croon's estimator we detailed is built on a factor-wise implementation—the current approach leverages the ML estimates of the individual measurement models. However, MLSEM can accommodate much more complicated measurement models such as those with within indicator multidimensionality or correlated residual errors (e.g., Hayes & Usami, 2019). Future research is needed to further develop and probe Croon-like corrections for a broad range of measurement structures to detail the extent to which Croon's method represents a versatile strategy.

Similarly, the current implementation of the proposed estimator is limited in the types of structural models it can serve. For instance, the proposed approach does not immediately accommodate random slopes (e.g., 1-1 -1 mediation with random slopes). Estimation of several or more random slopes within the context of MLSEM can be data intensive and will typically require more than the number of organizations we have considered. Extensions to the proposed bias-corrected estimator that can relax the large sample requirements of random slope models may be particularly valuable in many settings.

A third limitation is that, like ML, Croon's estimator is still subject to bias introduced through the omission of confounding variables. From a relative sense, our simulations suggested that in the presence of measurement error and an omitted confounding variable, Croon's incurs the same level of additional bias as the ML estimator. However, from an absolute sense, Croon's and ML still cannot fully correct for confounding variables that have been omitted from the model.

A final limitation of the proposed approach is that it currently lacks indices to assess the global fit of the model. More specifically, a key utility in MLSEM is the ability to test the extent to which the proposed model is plausibly consistent with the data and the degree to which it accounts for the observed covariances among variables. Recent work has begun to address this gap for single-level settings but more work is needed to develop a broad range of fit statistics that is comparable to those of conventional methods (Devlieger, Talloen, & Rosseel, 2019).

Even with these limitations, the results of our study provide important initial support for the utility of Croon's estimator in multilevel settings. The results suggest that the estimator potentially serves as an important alternative or complement to ML in small- to moderate-scale multilevel models, and more generally they encourage the continued development of this approach.

### 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 authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This study was supported by the National Science Foundation (grant numbers 1552535, 1760884). The National Science Foundation had no role or involvement in the conduct of the research or the preparation of the manuscript.

### ORCID iD

Benjamin Kelcey  <https://orcid.org/0000-0002-5083-1406>

### Supplemental Material

Supplemental material for this article is available online.

### References

- Aarons, G., & Sawitzky, A. (2006). Organizational climate partially mediates the effect of culture on work attitudes and staff turnover in mental health services. *Administration and Policy Mental Health, 33*, 289-301.
- Aguinis, H., Edwards, J. R., & Bradley, K. J. (2017). Improving our understanding of moderation and mediation in strategic management research. *Organizational Research Methods, 20*, 665-685.
- Allen, T. D., Herst, D. E. L., Bruck, C. S., & Sutton, M. (2000). Consequences associated with work-to-family conflict: A review and agenda for future research. *Journal of Occupational Health Psychology, 5*, 278-308.
- Bakk, Z., & Vermunt, J. (2015). Robustness of stepwise latent class modeling with continuous distal outcomes. *Structural Equation Modeling: A Multidisciplinary Journal, 23*, 20-31.
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using *lme4*. *Journal of Statistical Software, 67*, 1-48.
- Bodner, T. E., & Bliese, P. D. (2018). Detecting and differentiating the direction of change and intervention effects in randomized Trials. *Journal of Applied Psychology, 103*, 37-53.
- Bollen, K. A. (1989). *Structural equations with latent variables*. New York, NY: John Wiley.
- Brough, P., & O'Driscoll, M. (2005). Work-family conflict and stress. In A. Antoniou & C. Cooper (Eds.), *Research companion to organizational health psychology* (pp. 346-365). Cheltenham, UK: Edward Elgar.
- Brough, P., O'Driscoll, M., Kalliath, T., Cooper, C. L., & Poelmans, S. (2009). *Workplace psychological health: Current research and practice*. Cheltenham, UK: Edward Elgar.
- Brown, E. C., Hawkins, J. D., Rhew, I. C., Shapiro, V. B., Abbott, R. D., Oesterle, S., & Catalano, R. F. (2014). Prevention system mediation of communities that care effects on youth outcomes. *Prevention Science, 15*, 623-632.

- Cheung, G., & Lau, R. (2008). Testing mediation and suppression effects of latent variables: Bootstrapping with structural equation models. *Organizational Research Methods, 11*, 296-325.
- Cheung, G., & Lau, R. (2017). Accuracy of parameter estimates and confidence intervals in moderated mediation models: A comparison of regression and latent moderated structural equations. *Organizational Research Methods, 20*, 746-769.
- Cole, D. A., & Preacher, K. J. (2014). Manifest variable path analysis: Potentially serious and misleading consequences due to uncorrected measurement error. *Psychological Methods, 19*, 300-315.
- Croon, M. (2002). Using predicted latent scores in general latent structure models. In G. Marcoulides & I. Moustaki (Eds.), *Latent variable and latent structure modeling* (pp. 195-223). Mahwah, NJ: Lawrence Erlbaum.
- Devlieger, I., Mayer, A., & Rosseel, Y. (2016). Hypothesis testing using factor score regression: A comparison of four methods. *Educational and Psychological Measurement, 76*, 741-770.
- Devlieger, I., & Rosseel, Y. (2017). Factor score path analysis: An alternative for SEM? *Methodology, 13*, 31-38.
- Devlieger, I., Talloen, W., & Rosseel, Y. (2019). New developments in factor score regression: Fit indices and a model comparison test. *Educational and Psychological Measurement, 79*, 1017-1037.
- Eby, L. T., Casper, W. J., Lockwood, A., Bordeaux, C., & Brinley, A. (2005). Work and family research in IO/OB: Content analysis and review of the literature (1980-2002). *Journal of Vocational Behavior, 66*, 124-197.
- Efron, B., & Tibshirani, T. J. (1993). *An introduction to the bootstrap*. London, UK: Chapman and Hall.
- Fritz, M., Kenny, D., & MacKinnon, D. (2016). The combined effects of measurement error and omitting confounders in the single-mediator model. *Multivariate Behavioral Research, 51*, 681-697.
- Gagne, P., & Hancock, G. R. (2006). Measurement model quality, sample size, and solution propriety in confirmatory factor models. *Multivariate Behavioral Research, 41*, 65-83.
- Hayes, T., & Usami, S. (2019). Factor score regression in the presence of correlated unique factors. *Educational and Psychological Measurement*. Advance online publication. doi:10.1177/0013164419854492
- Hoshino, T., & Bentler, P. (2013). Bias in factor score regression and a simple solution. In A. R. de Leon & K. C. Chough (Eds.), *Analysis of mixed data: Methods & applications* (pp. 43-62). New York, NY: Chapman and Hall/CRC
- Hox, J. J., & Maas, C. (2001). The accuracy of multilevel structure equation modeling with pseudobalanced groups and small samples. *Structural Equation Modeling: A Multidisciplinary Journal, 8*, 157-174.
- Kelcey, B. (2019). A robust alternative estimator for small to moderate sample SEM: Bias-corrected factor score path analysis. *Addictive Behaviors, 94*, 83-98.
- Kelcey, B., Dong, N., Spybrook, J., & Cox, K. (2017). Statistical power for causally-defined indirect effects in group-randomized trials with individual-level mediators. *Journal of Educational and Behavioral Statistics, 42*, 499-530.
- Kelcey, B., Hill, H., & Chin, M. (2019). Teacher mathematical knowledge, instructional quality, and student outcomes: A multilevel mediation analysis. *School Effectiveness & School Improvement*. Advance online publication.
- Ketelaars, P. (2017). Tracing protest motives: The link between newspaper coverage, movement messages, and demonstrators' reasons to protest. *Sociological Forum, 32*, 480-500.
- Ledgerwood, A., & Shrout, P. (2011). The trade-off between accuracy and precision in latent variable models of mediation processes. *Journal of Personality and Social Psychology, 101*, 1174-1188.
- Li, X., & Beretvas, S. (2013). Sample size limits for estimating upper level mediation models using multilevel SEM. *Structural Equation Modeling: A Multidisciplinary Journal, 20*, 241-264.
- Loncke, J., Eichelsheim, V., Branje, S., Buysse, A., Meeus, W., & Loeys, T. (2018). Factor score regression with social relations model components: A case study exploring antecedents and consequences of perceived support in families. *Frontiers in Psychology, 9*, 1-19.
- Lu, I. R. R., Kwan, E., Thomas, D. R., & Cedzynski, M. (2011). Two new methods for estimating structural equation models: An illustration and a comparison with two established methods. *International Journal of Research in Marketing, 28*, 258-268.

- Lüdtke, O., Marsh, H. W., Robitzsch, A., Trautwein, U., Asparouhov, T., & Muthén, B. (2008). The multilevel latent covariate model: A new, more reliable approach to group-level effects in contextual studies. *Psychological Methods, 13*, 203-229.
- McNeish, D. (2017). Multilevel mediation with small samples: A cautionary note on the multilevel structural equation modeling framework. *Structural Equation Modeling: A Multidisciplinary Journal, 24*, 609-625.
- Muthén, B. (1991). Multilevel factor analysis of class and student achievement components. *Journal of Educational Measurement, 28*, 338-354.
- Preacher, K. J., & Selig, J. P. (2012). Advantages of Monte Carlo confidence intervals for indirect effects. *Communication Methods and Measures, 6*, 77-98.
- Preacher, K. J., Zyphur, M. J., & Zhang, Z. (2010). A general multilevel SEM framework for assessing multilevel mediation. *Psychological Methods, 15*, 209-233.
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*. Thousand Oaks, CA: Sage.
- Rosseeel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software, 48*, 1-36.
- Sampson, R. J., Raudenbush, S. W., & Earls, F. (1997). Neighborhoods and violent crime: A multilevel study of collective efficacy. *Science, 277*, 918-924.
- Savalei, V., & Bentler, P. M. (2009). A two-stage approach to missing data: Theory and application to auxiliary variables. *Structural Equation Modeling, 16*, 477-497.
- Schochet, P. Z. (2011). Do typical RCTs of education interventions have sufficient statistical power for linking impacts on teacher practice and student achievement outcomes? *Journal of Educational and Behavioral Statistics, 36*, 441-471.
- Skrondal, A., & Kuha, J. (2012). Improved regression calibration. *Psychometrika, 77*, 649-669.
- Song, L., Tsui, A., & Law, K. (2009). Unpacking employee responses to organizational exchange mechanisms: The role of social and economic exchange perceptions. *Journal of Management, 35*, 56-93.
- Spearman, C. E. (1904). "General intelligence" objectively determined and measured. *American Journal of Psychology, 15*, 201-292.
- Spybrook, J., Shi, R., & Kelcey, B. (2016). Progress in the past decade: An examination of the precision of cluster randomized trials funded by the U.S. Institute of Education Sciences. *International Journal of Research & Method in Education, 39*, 255-267.
- Takane, Y., & Hwang, H. (2018). Comparisons among several consistent estimators of structural equation models. *Behaviormetrika, 45*, 157-188.
- VanderWeele, T. J. (2010). Direct and indirect effects for neighborhood-based clustered and longitudinal data. *Sociological Methods & Research, 38*, 515-544.
- Walton, G. (2014). The new science of wise psychological interventions. *Current Directions in Psychological Science, 23*, 73-82.
- Wooldridge, J. M. (2002). *Econometric analysis of cross section and panel data*. Cambridge, MA: MIT Press.

### Author Biographies

**Benjamin Kelcey** is an associate professor of quantitative research methodologies at the University of Cincinnati. His research interests focus on causal inference and measurement methods within the context of multilevel and multidimensional organizational settings.

**Kyle Cox** is an assistant professor of research, measurement, and evaluation at the University of North Carolina at Charlotte. His research interests include experimental and quasi-experimental design, mediation, and measurement of group-level organizational constructs.

**Nianbo Dong** is an associate professor of quantitative methods at the University of North Carolina at Chapel Hill. His research interests include design and analysis of multilevel experiments and causal inference.