Title

Multilevel structural equation modeling (MSEM) for social work researchers: An introduction and an application to healthy youth development

Authors

Lee, Juyeon
Shapiro, Valerie
Kim, BKE
et al.

Publication Date

2018-12-18

Peer reviewed

Multilevel Structural Equation Modeling for Social Work Researchers: An Introduction and Application to Healthy Youth Development

Juyeon Lee, University of California, Berkeley
Valerie B. Shapiro, University of California, Berkeley
B. K. Elizabeth Kim, University of Southern California
Joan P. Yoo, Seoul National University

CITATION: Lee, J., Shapiro, V.B., Kim, B.K.E., & Yoo, J.P. (2018). Multilevel structural equation modeling (MSEM) for social work researchers: An introduction and an application to healthy youth development. *Journal of the Society for Social Work and Research*, *9*(4), 689–719.

AVAILABLE ONLINE: https://www.journals.uchicago.edu/doi/pdfplus/10.1086/701526

Author Notes

Juyeon Lee, MSW, is a doctoral student at the University of California, Berkeley, School of Social Welfare.

Valerie B. Shapiro, PhD, is an associate professor at the University of California, Berkeley, School of Social Welfare.

B. K. Elizabeth Kim, PhD, is an assistant professor at the University of Southern California, USC Suzanne Dworak-Peck School of Social Work.

Joan P. Yoo, PhD, is a professor at Seoul National University, Department of Social Welfare.

Correspondence regarding this article should be directed to Juyeon Lee, 120 Haviland Hall, #7400, Berkeley, CA 94720 or via e-mail to juyeon.lee@berkeley.edu

---

Acknowledgments

This project, and the associated data collection and analyses, was made possible by the support of the Stuart Foundation. The authors wish to acknowledge the teachers, staff, and administrators who participated in this project. The authors also wish to thank colleagues at the Center for Prevention Research in Social Welfare who played essential roles in the completion of this work. The authors appreciate consultation from Kristopher J. Preacher, Sophia Rabe-Hesketh, and Anders Skrondal on statistical methods; Mark Collin and Chuck Fisher on the TOOLBOX program; and Paul LeBuffe, Jennifer Fleming Robitaille, Jessica Adamson, and Marc Kirsch on the DESSA-Mini. We also thank the three anonymous reviewers whose generous suggestions improved the final version of the manuscript. The opinions expressed in this article do not necessarily reflect the views of any funding agency, resource or method developer, or community partner. No author of this paper receives any financial remuneration from the sale of the assessments or resources mentioned within this manuscript.

Manuscript submitted: October 24, 2017
First revision submitted: February 24, 2018
Second revision submitted: April 11, 2018
Accepted: May 6, 2018
Electronically published: November 29, 2018

Abstract

Objective: To achieve the grand challenge goal of unleashing the power of prevention, we must determine how and under what conditions an intervention leads to desired outcomes. These questions remain largely unknown partly due to analytical challenges involving testing mediation and moderation hypotheses with multiple dependent variables in nested data. This paper introduces multilevel structural equation modeling (MSEM) and demonstrates multilevel mediation and moderation analysis to understand the mechanisms by and contexts in which preventive interventions work. ***Method:*** Using illustrative research questions, we review the conceptual backgrounds of multilevel modeling and structural equation modeling and explain how MSEM combines these methods. We then analyze longitudinal data from a quasi-experimental study of a social and emotional learning program to examine how classroom teachers' baseline social–emotional competence (SEC) relates to students' year-end SEC, focusing on the mediation of instruction and the moderation of implementation leadership. ***Results:*** Teachers' SEC was directly related to students' year-end SEC (95% CI [.04, .95]) and indirectly related through the number of lessons delivered (95% CI [.01, .35]), controlling for students' baseline SEC and grade level. When teachers reported more implementation leadership at baseline, however, teachers' own SEC contributed less to the number of lessons they delivered (95% CI for interaction effect [−2.50, −.27]). MSEM techniques enabled examination of how teachers' SEC relates to their implementation behaviors and, in turn, to students' social and emotional development, and how these relationships are modified by implementation contexts. ***Conclusions:*** Identifying mechanisms and contexts in which students benefit from classroom-level interventions can help refine interventions and/or target implementation supports for taking preventive interventions to scale.

Keywords: multilevel structural equation modeling (MSEM), mediation and moderation, prevention, implementation, social and emotional learning (SEL)

Social workers play a critical role in developing and delivering interventions to promote healthy youth development (Shapiro, Kim, Robitaille, LeBuffe, & Ziemer, 2018). A large body of scientific evidence has demonstrated that mental, emotional, and behavior problems in children and youth can be prevented through societal action (Hawkins et al., 2015). Many interventions intended to ensure the healthy development and well-being of young people have been tested and found to be effective (O'Connell, Boat, & Warner, 2009). Social and emotional learning (SEL) programs are one type of school-based preventive intervention. SEL programs aim to prevent behavioral problems by promoting students' social–emotional competence (SEC), or the child's ability to make decisions and connect with others in ways that reflect the successful recognition and management of emotions (e.g., social awareness, goal-directed behavior, relationship skills; Shapiro, Accomazzo, & Robitaille, 2017). Experimental and quasi-experimental studies have demonstrated the benefits of many SEL programs on students' SEC and other indicators of positive development (Durlak, Weissberg, Dymnicki, Taylor, & Schellinger, 2011; Taylor, Oberle, Durlak, & Weissberg, 2017). The focus of such studies has primarily been examination of mean differences in target outcomes between intervention and control/comparison conditions under ideal (efficacy) or relatively more routine (effectiveness) implementation conditions (Flay et al., 2005).

To achieve the grand challenge goal of unleashing the power of prevention (i.e., scaling clinical outcomes for population impact), however, we also need to determine how and under what conditions an intervention leads to desired outcomes *within* an intervention condition in order to disseminate it well (Gottfredson et al., 2015; Shapiro, et al., 2013; Spoth et al., 2013). A data analytic technique called multilevel structural equation modeling (MSEM) can be used to examine the mechanisms of implementation delivery (i.e., mediation hypotheses) and the contexts that might shape intervention effects (i.e., moderation hypotheses) in routine practice settings. This method combines two different modeling approaches—multilevel modeling (MLM) and structural equation modeling (SEM)—to overcome the limitations of each by facilitating the analysis of multiple dependent variables in nested data (e.g., observations that are nonindependent by nature of their shared group membership). Prevention researchers have suggested that data analysis techniques such as MSEM may be well-suited to answer practice-relevant questions important for refining and disseminating prevention practice at scale (Shapiro, Hawkins, & Oesterle, 2015).

Since MSEM was integrated into existing software (e.g., Mplus Version 5, Stata Version 8) as a generalized modeling framework (B. Muthén & Asparouhov, 2008; Rabe-Hesketh, Skrondal, & Pickles, 2004), it has been introduced to social science researchers as either an extension of MLM (e.g., Christ et al., 2017; Hox, 2010) or an extension of SEM (e.g., Kaplan, 2008; Kaplan, Kim, & Kim, 2009). Several papers have also illustrated how to apply MSEM for different purposes, including confirmatory factor analysis (Davidov, Dülmer, Schlüter, Schmidt, & Meuleman, 2012; Geldhof, Preacher, & Zyphur, 2014; Mehta & Neale, 2005), measurement and structural models (Cheung & Au, 2005; Rabe-Hesketh, Skrondal, & Zheng, 2007), mediation models (MacKinnon & Valente, 2014; Preacher, Zyphur, & Zhang, 2010), and moderation models (Preacher, Zhang, & Zyphur, 2016). However, these resources may not be readily accessible to social work researchers seeking to apply MSEM to the social problems and interventions they typically examine if they lack highly contemporary and sophisticated statistical training. This paper responds to calls for workforce development activities specifically intended for, but not limited to, social work researchers studying the implementation of preventive interventions that advance the well-being of all youth (Hawkins, Shapiro, & Fagan, 2010).

In this paper—which may be most helpful to social work researchers who already have some familiarity with MLM and SEM—we provide specific examples of MSEM used to examine research questions that simultaneously require MLM and SEM analysis techniques. To orient readers to the terms, notations, and diagram techniques we use, we first present the conceptual backgrounds of MLM and SEM, separately, using illustrative research questions. To facilitate an accessible, initial understanding of MSEM, we then explain how MSEM can combine MLM and SEM in a single modeling framework, focusing on diagram representations for model specification instead of multiple matrix equations. Finally, we use data from a quasi-experimental study of an SEL program to illustrate the analytic procedures and results generated through the use of MSEM. In doing so, we demonstrate the utility of MSEM for addressing the grand challenge to ensure healthy youth development (Uehara et al., 2013). Possible applications and limitations of MSEM are discussed.

Multilevel Modeling

Analysis of Nested Data

Let us consider data collected during a yearlong implementation of a school-based SEL program in which classroom teachers deliver the lessons. Even when students in an intervention condition, on average, had higher SEC at the end of the year than students in a control/comparison condition, it is likely that variations in student outcomes exist within the intervention condition such that some students show higher year-end SEC than others. The variations in students' year-end SEC scores can be partitioned into two levels: within- and between-classroom variations. A *within-classroom variation* would reflect a student's SEC score being higher or lower than his or her classroom average (i.e., cluster mean). A *between-classroom variation* might occur when the average SEC score of the student's classroom is higher or lower than the average SEC score of all classrooms (i.e., grand mean). A larger between-classroom variation indicates a stronger correlation in SEC scores between a student and his/her classmates in the same classroom. Simply put, students in the same classrooms are likely to be more similar to each other than to students in different classrooms based on shared norms or experiences. This violates the assumption of independent observations traditional in regression analysis using ordinary least squares (OLS) estimation and therefore requires more sophisticated methods, such as MLM.

Also known as hierarchical linear modeling or mixed effects modeling, MLM handles the dependence of clustered observations (e.g., students nested in classrooms) by partitioning residuals (i.e., unexplained sources of variance) into within-cluster and between-cluster residuals. OLS regression analysis assumes that all observations are independent of each other and therefore does not partition total variation of a variable into within-cluster and between-cluster variability. By estimating the between-cluster residual variance separately from the within-cluster residual variance, however, MLM can account for the nonindependence of observations and prevent biased estimation.

Random Intercept Model

In our example, we have a set of SEC scores for a student (*i*) in a classroom (*j*); the deviations of individual student scores from the grand mean (i.e., average score of the full sample) are partitioned into within- and between-classroom levels. This is done by setting the regression intercept (β_{0j}; see Equation 1) as a random intercept that can vary across clusters—in this case classrooms, notated as u_{0j} (see Equation 2). Note that variables, coefficients, and residuals with subscript *ij* vary across students and classrooms, whereas variables, coefficients, and residuals with subscript *j* only vary across classrooms and have no within-classroom variation. Equation 3 combines Equations 1 and 2, indicating that the SEC score of a student (*i*) in a classroom (*j*; Y_{ij}) can be decomposed into a grand mean (γ_{00}), between-classroom residual (u_{0j}), and within-classroom residual (e_{ij}). In other words, all students share the same intercept (γ_{00}), with students in the same classroom sharing the same deviation from the grand mean (u_{0j}) and each student also having an individual deviation from her or his classroom mean (e_{ij}). The MLM equations introduced in this paper follow the conventions of Raudenbush and Bryk (2002).

$$Y_{ij} = \beta_{0j} + e_{ij} \quad (1)$$

$$\beta_{0j} = \gamma_{00} + u_{0j} \quad (2)$$

$$Y_{ij} = \gamma_{00} + u_{0j} + e_{ij} \quad (3)$$

Equations 1, 2, and 3 specify a *random intercept model*, represented in Figure 1 as a full diagram on the left (Boxes A and B) and as a simplified diagram on the right, providing an illustrative example in which the variable is a student's SEC (Boxes C and D). The MLM diagrams in this paper follow the conventions of the *Mplus User's Guide* (L. K. Muthén & Muthén, 1998–2017), with minor adaptations.

The proportion of between-cluster residual variance in a variable's total variance is called the intraclass correlation (ICC; Raudenbush & Bryk, 2002). In our example, the ICC can be interpreted as the degree to which individual SEC scores within a classroom are correlated with one another. Often, the first step of MLM is to examine the ICC of an outcome variable of interest to assess the proportion of the variation in the variable that exists across clusters (Woltman, Feldstain, MacKay, & Rocchi, 2012). Although there are no widely accepted criteria for assessing the value of ICC, it has been suggested that ICCs equal to or greater than .05 render

MLM necessary (Geldhof et al., 2014; Huta, 2014). ICC values between .05 and .20 have commonly been reported in the MLM literature (Bliese, 2000; Preacher, Zhang, & Zyphur, 2011).

Random Intercept and Fixed (or Random) Slopes Model

Continuing with our example, let us say that the ICC for students' SEC score is .15, indicating that 15% of variance in SEC scores is due to between-classroom variations and the remaining 85% is due to within-classroom variations (see, e.g., Shapiro, Kim, Accomazzo, & Roscoe, 2016; Smith-Millman et al., 2017). The next step would be to find predictors that explain these different sources of variation in students' SEC scores. We hypothesize that within-classroom variations may be explained by individual student characteristics, such that students' preintervention SEC scores may predict their postintervention scores. Between-classroom variations may be explained by features that are shared by all students in the classroom but vary among different classrooms, such as teachers' SEC and the number of SEC lessons delivered within a given classroom. The preintervention SEC score (X_{ij}) is included in Equation 4 as a student-level predictor, and the teacher SEC score (W_{1j}) and the number of SEL lessons delivered (W_{2j}) are included in Equation 5 as classroom-level predictors. By adding these predictors, the within- and between-classroom residuals (e_{ij} and u_{0j}) now reflect residuals left unexplained after accounting for the effects of these predictors. Equation 6 indicates that the regression slope of the preintervention SEC score (β_{1j}) in Equation 4 is assumed not to vary across classrooms (i.e., fixed slope) given that there is no residual term (u_{1j}). In the combined Equation 7, the effect of each predictor on students' year-end SEC, controlling for other predictors, is noted as γ_{10}, γ_{01}, and γ_{02}. Depending on the research hypothesis, however, the slope of a student-level predictor can be specified to vary across classrooms (i.e., random slope). For example, one might hypothesize that the relationship between baseline SEC scores and year-end SEC scores would be different across classrooms and that these variations in the slope could be explained by other classroom-level predictors.

$$Y_{ij} = \beta_{0j} + \beta_{1j}X_{ij} + e_{ij} \quad (4)$$

$$\beta_{0j} = \gamma_{00} + \gamma_{01}W_{1j} + \gamma_{02}W_{2j} + u_{0j} \quad (5)$$

$$\beta_{1j} = \gamma_{10} \quad (6)$$

$$Y_{ij} = \gamma_{00} + \gamma_{10}X_{ij} + \gamma_{01}W_{1j} + \gamma_{02}W_{2j} + u_{0j} + e_{ij} \quad (7)$$

The model specified in Equations 4–7 is called a *random intercept and fixed slope model* (see Figure 2). In this model, teacher SEC scores as a classroom-level predictor cannot explain the within-classroom variations, since all students in the same classroom would share the same level of teacher SEC. A classroom-level predictor, therefore, can reduce only the between-classroom residual variance. In contrast, students' preintervention SEC scores—the student-level predictor—can explain both the within- and between-classroom variations. Similar to the outcome variable, students' preintervention SEC scores vary across students within the same classroom as well as across different classrooms. The effect of preintervention SEC scores on year-end SEC scores (γ_{10}), therefore, conflates the within-classroom effect (e.g., how much the baseline individual student SEC is related to the year-end individual student SEC) and the between-classroom effect (e.g., how much the baseline classroom-level SEC is related to the year-end classroom-level SEC).

Limitations of MLM

MLM has advantages over OLS regression when analyzing nested data by partitioning the residuals of an outcome variable into within- and between-cluster residuals. Conventional MLM techniques, however, have several limitations. First, MLM is not easily applied to more complex models that seek to examine several correlated dependent variables simultaneously (e.g., multiple related constructs, a single construct measured through several related indicators). For example, one might hypothesize that the number of SEL lessons delivered would mediate the relationship between teachers' SEC (independent variable) and students' year-end SEC (dependent variable). In this case, the teachers' SEC is predicting both the number of lessons delivered (mediator) and students' year-end SEC at the same time, which is also predicted by the mediator. In conventional MLM, these relationships would be examined in multiple steps that do not account for the relationships between variables simultaneously. In the next section, we will review structural equation modeling, one of the most widely used multivariate methods for modeling several equations simultaneously.

Furthermore, conventional MLM techniques that separately estimate the within-cluster and between-cluster effects can produce biased estimates. In an MLM approach, these two conceptually different effects are often separated by including both observed cluster means (e.g., aggregated individual characteristics) and cluster-mean-centered scores as predictors. The methodological literature, however, suggests that this approach may produce a biased estimation of between-cluster effects, and one way to overcome this problem is to use a multilevel latent covariate (MLC) approach (e.g., Christ et al., 2017; Lüdtke et al., 2008). We will discuss later how MSEM integrates the MLC approach rather than requiring an additional, often-overlooked step in the analytic process.

Structural Equation Modeling

Analysis of Multiple Equations Simultaneously

So far, our illustrative models only included observed variables. But we may be interested in measuring outcome variables that are latent (i.e., not directly observable, and thus, not directly measurable). For example, if we assumed that students' SEC is a latent construct measured by eight teacher-rated indicators intended to reflect SEC, we would want to first examine the extent to which all of these indicators appear to be actually measuring SEC. We would also want to understand how SEC, as a construct, relates to other variables after partialling out (i.e., segregating the effect of) any measurement errors (see the next section for explication of this term). When we use a mean or sum score of these eight indicators as a single outcome variable, as is often the case for OLS regression and MLM, there is no mechanism to parse out the measurement error from a total residual.

SEM, also known as covariance structure modeling (Kline, 2015), is designed to consider the unknown measurement errors embedded in each indicator and is used for, but not limited to, path analysis with latent variables. Neither the OLS regression nor conventional MLM approaches allow for multiple dependent variables (either observed indicators or latent constructs) in a single model. By estimating more than one equation simultaneously, however, SEM can partition the variance of each observed indicator into the portion explained by an underlying latent construct and the portion left unexplained (due to unknown measurement errors), which in turn permits more accurate estimates of relationships among constructs.

Measurement Model

We continue with our example to consider the exploration of a measurement model. If the eight indicators of student SEC are designed well, a student with higher SEC would consistently get higher scores on each indicator of the construct, just as another student with lower SEC would consistently get lower scores across indicators. In other words, the common variance underlying the observed variances in these indicators reflects students' underlying SEC. Measurement errors can also exist in each indicator, and such errors are theoretically unrelated to the underlying construct but are produced for a variety of unknown reasons related to the indicators themselves. For example, if a teacher is asked how frequently she or he has observed the performance of a specific behavior, a student rating may be influenced by constraints on the opportunity to observe the student's behavior rather than student's ability or propensity to skillfully execute the behavior. Additionally, some of the indicators might be more influenced by a related, yet distinct, construct, such as cognitive abilities or personality.

SEM separates the common variance in these eight indicators and measurement errors in each indicator by estimating factor loadings (i.e., the degree to which an indicator is explained by the construct it intends to measure) and the measurement error variances (i.e., the residual variances left unexplained in the indicator). Equation 8 shows how multiple observed variables (y_{1i}–y_{8i}), relate to a single latent construct (η_i). The degree to which each indicator is related to the latent construct is expressed by λ_1–λ_8, and the unknown measurement errors are expressed by ε_{1i}–ε_{8i}. The SEM equations introduced in this paper follow the conventions of Kaplan (2008) with minor adaptations.

$$\begin{pmatrix} y_{1i} \\ y_{2i} \\ y_{3i} \\ y_{4i} \\ y_{5i} \\ y_{6i} \\ y_{7i} \\ y_{8i} \end{pmatrix} = \begin{bmatrix} \lambda_1\eta_i + \varepsilon_{1i} \\ \lambda_2\eta_i + \varepsilon_{2i} \\ \lambda_3\eta_i + \varepsilon_{3i} \\ \lambda_4\eta_i + \varepsilon_{4i} \\ \lambda_5\eta_i + \varepsilon_{5i} \\ \lambda_6\eta_i + \varepsilon_{6i} \\ \lambda_7\eta_i + \varepsilon_{7i} \\ \lambda_8\eta_i + \varepsilon_{8i} \end{bmatrix} \tag{8}$$

The model specified in Equation 8 is called a *measurement model* or *confirmatory factor model*, represented as a diagram in Figure 3. The SEM diagrams in this paper follow the conventions of the *Mplus User's Guide* (L. K. Muthén & Muthén, 1998–2017).

If all eight indicators adequately measure what they intended to measure, the model would produce a good fit to the observed covariances among the indicators, with consistently high factor loadings and low measurement error variances. Often, the first step of SEM is to evaluate the goodness of fit using various model fit indices, such as chi-squared statistic (χ^2), standardized root mean square residual (SRMR), comparative fit index (CFI), Tucker-Lewis index (TLI), and root mean square error of approximation (RMSEA; Hu & Bentler, 1999; Kaplan, 2008; Kline, 2015). Bowen and Guo (2011) recommended fit criteria for social work research that include a nonsignificant chi-squared statistic, CFI ≥ .95, TLI ≥ .95, GFI ≥ .90, and RMSEA ≤ .08. Hair, Black, Babin, and Anderson (2010) provided fit assessment guidelines that account for sample size and model complexity.

Structural Model

When the measurement model shows an acceptable fit, one may want to examine the relationships among constructs while accounting for the measurement errors. Referring back to our example, Equations 9 and 10 illustrate one possible structural model that simultaneously examines the relationships between teacher's SEC (X_{1i}) and students' year-end SEC (η_i), mediated by the number of SEL lessons delivered (X_{2i}). In Equation 9, the outcome variable (η_i) is explained by two predictors, X_{1i} and X_{2i}, and has residuals ($\zeta_{\eta i}$). In Equation 10, the number of lessons (X_{2i})—one of the predictors in Equation 9—is set as an outcome variable predicted by X_{1i} with residuals (ζ_{Xi}).

$$\eta_i = \beta_{31}X_{1i} + \beta_{32}X_{2i} + \zeta_{\eta i} \quad (9)$$

$$X_{2i} = \beta_{21}X_{1i} + \zeta_{xi} \quad (10)$$

Figure 4 represents the model specified in Equations 9 and 10, in combination with Equation 8.

Limitations of SEM

SEM can be used to simultaneously examine a set of interrelations among multiple variables, accounting for measurement errors. Yet, with the exception of repeated observations nested within individuals, conventional SEM techniques are rarely applied to nested data. To analyze the example model (Figure 4) in a conventional SEM framework, one could disaggregate teacher data to the student level or aggregate student data to the classroom level. Both approaches, however, can lead to a loss of information and biased estimates of the associations among variables. In the following section, we will introduce an analytic technique that combines the features of MLM and SEM, allowing analysis of measurement and structural models with nested data.

Multilevel Structural Equation Modeling

Analysis of Multiple Equations Simultaneously in Nested Data

There have been considerable efforts to synthesize MLM and SEM into a single generalized modeling framework (B. Muthén & Asparouhov, 2008; Rabe-Hesketh et al., 2004). MSEM is a combination of MLM and SEM that can address the limitations of MLM for multivariate modeling (e.g., cannot have several dependent variables simultaneously nor account for measurement errors) and those of SEM for multilevel modeling (e.g., violates the assumption of independent observation). MSEM not only partitions the observed variations into within- and between-cluster residuals as in MLM, but it also separates measurement errors from a total residual as in SEM. Moreover, the procedure of partitioning variations from different levels (i.e., within- and between-cluster) and different sources (i.e., the latent construct and its measurement error) can be applied to any number of dependent variables and predictors, as needed.

Example Model of MSEM

To illustrate, let us examine a multilevel mediation model demonstrated in Figure 5. The dependent variable is students' SEC (measured by an 8-item indicator), the independent variable is classroom teachers' SEC (a single-item indicator), and the mediating variable is the number of SEL lessons delivered during the intervention year. Students' baseline SEC scores (a single-item indicator) and their grade level (K–2 or 3–5) are included as covariates, given program differences between these grade level groups. In this model, the 8-item indicator of students' SEC at year-end and the students' baseline SEC scores have variations at both within- and between-classroom levels. All other variables have no within-classroom variation. Assuming that no classroom is shared by more than one teacher or students from different grades, students in the same classroom would share the same value for their teacher's SEC, the number of SEL lessons delivered, and the grade level.

In Figure 5, the variances in the eight indicators of year-end SEC, as well as the single indicator of baseline SEC, are partitioned into two levels (*Between* and *Within*) to account for the multilevel nature of the data. The variations in the eight indicators of student SEC at each level are partitioned into two parts: variations explained by the underlying construct and variations due to measurement errors. In the Within section, the latent variable in an oval (labeled SEC_W) reflects the common variance underlying the within-classroom variations in the eight indicators. The short horizontal arrows pointing to each item indicate measurement errors. The same procedure is applied to the between-classroom variations in the eight indicators: the latent variable in an oval labeled SEC_B reflects the common variance, and the short horizontal arrows indicate measurement errors.

After accounting for the measurement errors at each level, the relationships among variables are specified as hypothesized and the residuals left unexplained are represented with short vertical arrows. In the Within section, students' year-end SEC at the student level (SEC_W) is predicted by their baseline SEC score ($PRESEC_W$) with some residuals accounting for measurement errors. In the Between section, students' year-end SEC at the classroom level (SEC_B) is predicted by the teachers' SEC (TSEC), the number of SEL lessons delivered (NLES), students' grade level, and their baseline SEC scores at the classroom level ($PRESEC_B$); NLES is also predicted by other variables at the classroom level. In this way, we can examine the mediational relationships among TSEC, NLES, and SEC_B, controlling for students' baseline SEC score and grade level, and accounting for the nonindependence of observations and measurement errors embedded in the dependent variable.

Advantages of MSEM

MSEM provides a great deal of flexibility in specifying and testing complex models with nested data. Note that the model represented in Figure 5 is only one specific example of MSEM. For example, one might further test if the mediational path specified in Figure 5 is moderated by an implementation context (moderated mediation) or if the mediation model is maintained across student subsamples (multigroup analysis). From an MLM perspective, MSEM can be used to examine multilevel models that involve two or more correlated dependent variables at any level, one or more mediators at any level, or one or more latent constructs measured by multiple indicators at any level. From an SEM perspective, MSEM can be used to test measurement or structural models with data structured across two or three nested levels. For a variety of possibilities and examples of MSEM, refer to Chapter 9 of the *Mplus User's Guide*, which is available for free online (L. K. Muthén & Muthén, 1998–2017).

One of the unique features of MSEM compared to conventional MLM is that the intercept of students' baseline SEC scores is also considered as a latent variable randomly varying across classrooms, and its effects on the outcome variables are separated into within- and between-classroom effects. To differentiate these two effects from a conventional MLM approach, one would include classroom means and mean-centered scores, but this may yield biased estimates when the observed classroom means are not a reliable measure of the unobserved classroom-level characteristic, especially with small classroom size and low ICC (Hox, 2010; Lüdtke et al., 2008). One might consider incorporating the MLC method in a conventional MLM approach to account for the unreliability of the classroom mean and obtain more accurate estimates of within- and between-classroom effects. But this requires a secondary procedure only available for models of limited complexity. In an MSEM framework, however, the MLC method is applied in a model-based way, meaning that the unobserved group characteristic for an independent variable is treated as a cluster-level latent variable by default, leading to a more accurate and efficient way of estimating the within- and between-cluster effects separately yet simultaneously (Asparouhov & Muthén, 2006).

The general principles of MSEM can be summarized as follows:

- All observed variations in lower-level variables are included in both the within- and between-cluster parts of a research model by allowing their intercepts to vary across clusters, whereas cluster-level variables are only included in the between-cluster part of the model.
- The within- and between-cluster parts of the model can be the same or different, but variables in one part cannot directly or indirectly predict variables in another part because the within- and between-cluster variations are assumed to be unrelated.

- Any regression slope of the within-cluster part of the model, however, can be related to between-cluster variables, by specifying the slope as a random coefficient explained by variables in the between-cluster part of the model. This is called a cross-level interaction.

Demonstration of MSEM

We will demonstrate the analytic procedures and results of MSEM in the following section using real-world data from the implementation of a preventive intervention. The example models are premised on the assertion that “students’ social, emotional, and academic competencies are enhanced through coordinated classroom, school, family, and community strategies” (Weissberg, Durlak, Domitrovich, & Gullotta, 2015, p. 8) and that “SEL is as much about adult change as it is about improvements in student performance” (Schonert-Reichl & O’Brien, 2012, p. 319). We selected the data used in the example for their multilevel structure and utility for depicting a measurement model, a mediation model, and a moderated mediation model. These data have the potential to provide information about the mechanisms and contexts of prevention program implementation that are associated with desired student outcomes.

To this end, we will first examine the multilevel mediation model in Figure 5, where teachers’ own SEC is hypothesized to relate to students’ year-end SEC mediated through their delivery of SEL lessons. Jennings and Greenberg (2009) as well as Jones and Bouffard (2012) suggested conceptual frameworks explaining that teachers’ own SEC influences effective SEL program implementation behaviors, which in turn influence students’ performance of social and emotional skills. We will then examine the moderated mediation model in Figure 6, where implementation leadership is hypothesized to moderate the relationship between teachers’ SEC and their SEL lesson delivery. Durlak and DuPre’s (2008) “Framework for Effective

Implementation" suggested that leadership is an important organizational capacity in a successful prevention delivery system. An empirical study underlying this claim found significant interaction effects between school leadership and teacher implementation of an SEL curriculum on student behavioral outcomes (Kam, Greenberg, & Walls, 2003). To our knowledge, however, no study has empirically examined the path model linking teachers' SEC to students' SEC through SEL implementation behaviors nor the effect of implementation contexts (e.g., implementation leadership) on these mediational relationships.

Methods

Example Data, Sample, and Measures

Example data come from the 2015–16 TOOLBOX Implementation Research Project (TIRP). TOOLBOX is a school-based SEL program that promotes children's social and emotional development through the instruction and reinforcement of 12 tools (e.g., Breathing Tool, Courage Tool; Collin, 2015). The TIRP aimed to understand naturally occurring variation in the routine implementation of TOOLBOX to explore the relationship between 1 year of TOOLBOX implementation and year-end student outcomes (Shapiro, Whitaker, Kim, & Roscoe, 2018). All data collection was approved by the University of California, Berkeley, Institutional Review Board.

TIRP data are derived from a quasi-experimental study in which four elementary schools adopted TOOLBOX and two comparison schools did not (all from within the same California school district). In this paper, we will focus on the variations in students' social and emotional outcomes within the intervention schools, as explained by classroom characteristics (e.g., teachers' SEC, the number of lessons delivered). The TIRP intervention sample consists of 1,897 K–5 students (48% in grades K–2; 46.2% female; 59% Hispanic/Latinx, 28% Asian/Filipinx, 13.2% Black/African American, and 8% White/European American) nested in 82 primary-assignment classrooms (teacher characteristics: 95.1% female; 61% White/European American, 9.8% Asian/Filipinx, and 8.5% Hispanic/Latinx; 44.4% had teaching credential for 11–20 years; average classroom size = 23) nested within four schools in one school district. The school district serves a community where 42% of students primarily speak a language other than English, 70% of students have household incomes of less than $44,123 annually for a family of four in a region with a high cost of living, and students meeting or exceeding the state educational standards in language arts/literacy is 27% and is 24% in mathematics), representing 84% of enrolled students. Cases were eliminated when students enrolled after baseline or when classroom teachers did not consent to inclusion in research. No difference was found in the assessment of student baseline SEC between consent and nonconsent classrooms ($t = .87$, $p = .39$). No classroom was shared by two or more surveyed teachers or contained students across different grade level groups.

In this sample data, students' SEC was measured using the Devereux Student Strengths Assessment-Mini (DESSA-Mini), a standardized teacher-completed behavior rating scale (Naglieri, LeBuffe, & Shapiro, 2011/2014). The DESSA-Mini is composed of eight items (e.g., show good judgment, focus on task despite problems or distraction) rated on a scale ranging from 0 (*never*) to 4 (*very frequently*; LeBuffe, Shapiro, & Robitaille, 2017). End-of-year scores were generated with the DESSA-Mini Form 3. The mean scores of each item at the end of the year ranged from 2.65–2.95 with standard deviations of .89–1.06 (some items missing on 236 cases). Students' baseline SEC was measured with DESSA-Mini Form 1 at the beginning of the year. For simplicity of illustrating the model, we used the DESSA-Mini Social Emotional Total (a composite T score, as would be generated in practice settings) as our control variable. The baseline student SEC in this sample ($M = 50.98$, $SD = 11.80$) was determined to be typical of students in the United States (*M* = 50, *SD* = 10). Alternate form reliability across Form 1 and Form 3 had previously been assessed as .92 (Naglieri et al., 2011/2014).

Classroom teachers' SEC was measured at the beginning of the year using a 1-item self-reported indicator from the Social and Emotional Learning Implementation Survey (SEL-IS; Shapiro et al., 2018). The SEL-IS defined SEC as "an awareness of and ability to manage emotions in a context-appropriate manner" followed by a single question asking respondents to rate their own SEC, as it shows up at work, from 0 (*very low*) to 4 (*very high*). Teachers, on average, rated their own SEC moderately (*M* = 2.76, *SD* = .71; 6% missing). The number of SEL lessons delivered throughout the year was measured by a summative score from teacher reports of endorsing particular lessons they had delivered to students (of 17 possible lessons for Grades K–2 and 18 possible lessons for Grades 3–5). Teachers, on average, reported teaching about two thirds of the lessons they were invited/asked to deliver (*M* = 12.65, *SD* = 3.87; 1% missing). Teacher-reported proactive implementation leadership was measured at the beginning of the year using the average score of the proactive subscale of the Implementation Leadership Scale (Aarons, Ehrhart, & Farahnak, 2014) which was adapted to the school implementation context (Accomazzo, Ziemer, Kim, & Shapiro, 2018). The proactive Implementation Leadership Scale consists of three items—"Has our school leadership removed obstacles to implementation of TOOLBOX?"; "To what extent were written plans/timelines for the 2015–2016 implementation of TOOLBOX at our school shared?"; and "Has our school leadership provided clear standards for the implementation of TOOLBOX?"—rated on a scale ranging from 0 (*not at all*) to 4 (*to a very great extent*). Teachers, on average, assessed their school implementation leadership as moderately proactive (*M* = 2.10, *SD* = .95; 16% missing). Grade level was coded as 1 (*Grades K–2*) and 0 (*Grades 3–5*; 48% in Grades K–2).

Analytic Procedures

Analyses were conducted in three steps. In the first step, we estimated the ICCs of the eight DESSA-Mini items measuring students' SEC for both classroom and school levels. Note that if the ICC for a cluster-level is zero, it suggests that the observations within that cluster may not be correlated with each other. In our second step, we fit the multilevel measurement model of the outcome variable (SEC) to examine whether all eight DESSA-Mini items were measuring one underlying construct at both the student and classroom levels. Again, if the model does not fit the data at one level, modifications to the measurement model may be needed to better represent the construct (ideally, followed by confirmations across diverse samples). When the measurement model adequately fits the data at both levels, one can estimate the relationships between the construct and other variables at each level more accurately by accounting for the measurement errors. In the third step, we fit the multilevel structural model to test hypothesized relationships among variables at each level. We first examined the direct relationship between teachers' SEC and students' year-end SEC, and then tested the mediation effect of the number of SEL lessons delivered on the relationship between teacher and student SEC. Finally, we tested the moderation effect of proactive implementation leadership on the relationship between teacher SEC and the delivery of SEL lessons.

We conducted all analyses demonstrated in this paper using Mplus (Version 8) with the combination add-on. We used the FIML method with a sandwich estimator (MLR) to address missing data. MLR also allows combinations of different types of variables (e.g., continuous, ordinal, and nominal) and computes standard errors that are robust to nonnormality and nonindependence of observations (L. K Muthén & Muthén, 2017). To test the indirect effect, we applied the Monte Carlo method for assessing mediation, which produces a sampling distribution of the indirect effect through a large number of repeated simulations (Selig & Preacher, 2008). Advantages of this method are similar to those of nonparametric bootstrapping, but it can also be used in cases where bootstrapping is not feasible, such as in multilevel modeling (Preacher & Selig, 2012). The Mplus codes used for each step can be found in the Appendix (online); these will work in Mplus (Versions 5 and later) with the multilevel or combination add-on.

Results

Step 1: Intraclass Correlations

Table 1 summarizes the ICC estimates of each DESSA-Mini item for classroom and school levels. Classroom-level ICCs ranged from .08 (“focus on a task despite a problem or distraction”) to .23 (“prepare for school, activities, or upcoming events”), indicating that 8%–23% of variance in students’ year-end SEC existed across classrooms. However, school-level ICCs were nearly zero after rounding to three decimal places, suggesting that students’ SEC did not vary at the school level. For parsimony, we proceeded with two-level modeling focused only on within- and between-classroom variations in students’ SEC.

Step 2: Multilevel Measurement Model

The eight indicators were hypothesized to measure a single latent construct: students’ year-end SEC. To differentiate the within- and between-classroom SEC, we named the construct SECW for the within-cluster model and SECB for the between-cluster model. The measurement model fit was acceptable—$\chi^2 = 267.94$ ($df = 40$, $p < .001$), CFI = .97, TLI = .95, RMSEA = .06, $SRMR_{within} = .02$, $SRMR_{between} = .05$—according to Hair et al.’s (2010) model-fit criteria for the corresponding sample size and number of variables. Table 2 presents the standardized factor loadings of each item at each level. All of the standardized factor loadings were higher than .80 and significant at the $p < .001$ level at both within- and between-classroom levels. These results show that the eight indicators construct a coherent outcome variable. Thus, all indicators were included in structural model analysis as observed variables measuring the SEC latent construct.

Step 3: Multilevel Structural Model

Before testing the mediation model, we first examined the direct relationship between teachers' SEC (TSEC) and students year-end SEC at classroom-level (SECB), controlling for students' baseline SEC (SECPRE) and grade level (GRAD). The structural model fit was acceptable: $\chi^2 = 373.05$ ($df = 68$, $p < .001$), CFI = .97, TLI = .96, RMSEA = .05, $SRMR_{within} = .01$, $SRMR_{between} = .05$. Table 3 presents the estimated regression slopes. At the classroom level, teachers' initial SEC was positively related to their students' year-end SEC ($b = .57$, $p < .05$). The relationship between students' baseline and postintervention SEC was significant at both the student ($b = .10$, $p < .001$) and classroom level ($b = .14$, $p < .001$). The difference in students' year-end SEC between lower and upper grades was marginally significant ($b = .55$, $p < .10$).

Next, we included the number of SEL lessons delivered (NLES) as a mediating variable linking TSEC and SECB. The structural model fit after including NLES was still acceptable: $\chi^2 = 386.65$ ($df = 75$, $p < .001$), CFI = .97, TLI = .96, RMSEA = .05, $SRMR_{within} = .01$, $SRMR_{between} = .07$. Table 4 presents the estimated regression slopes. Teachers' initial SEC was positively related to the number of SEL lessons delivered throughout the year ($b = 1.33$, $p < .05$), and the number of lessons delivered was also positively related to students' year-end SEC at the classroom level ($b = .11$, $p < .05$). The direct relationship between teachers' SEC and students' year-end SEC was still significant ($b = .50$, $p < .05$).

Using the regression slopes of TSEC on NLES and NLES on SECB, as well as the variances and covariance of these slopes, the 95% Monte Carlo confidence interval for the indirect effect was calculated as [.01, .35]. Given that this interval does not contain zero, we can conclude that the teachers' initial SEC is directly related to students' year-end SEC and indirectly related through the delivery of SEL lessons at the 95% confidence level. The direct path between teacher SEC and student SEC remained significant, indicating a partial mediation of SEL lesson delivery in the path between teacher and student SEC.

Lastly, we included proactive implementation leadership (PROSL) as a moderator of the path between TSEC and NLES. After including PROSL, the structural model fit was acceptable: $\chi^2 = 411.46$ ($df = 91$, $p < .001$), CFI = .97, TLI = .96, RMSEA = .05, $SRMR_{within} = .02$, $SRMR_{between} = .06$. Table 5 presents the estimated regression slopes. Proactive implementation leadership ($b = 4.78$, $p < .01$) as well as teachers' own SEC ($b = 4.78$, $p < .01$) were each positively related to the number of SEL lessons delivered. The interaction effect of PROSL and TSEC on NLES was also significant ($b = -1.38$, $p < .05$), indicating that proactive leadership moderated the relationship between teachers' SEC and the number of SEL lessons delivered. When proactive leadership was higher, the relationship between teachers' SEC and the number of SEL lessons delivered was weaker.

Discussion

Our illustrative analysis first revealed the extent to which student SEC was correlated (i.e., nonindependent) within classrooms during the implementation of a preventive intervention. Depending on the indicator, 8%–23% of variance in student year-end SEC could be explained at the classroom level. Some indicators varied more among students (e.g., "focus on a task despite a problem or distraction"), whereas other indicators varied more among classrooms (e.g., "prepare for school, activities, or upcoming events"). School-level ICCs, however, were nearly zero, indicating that students' year-end SEC did not vary by their school membership. The values of classroom-level and school-level ICCs found in our example are consistent with typical findings in other school-based preventive intervention studies (e.g., Murray et al., 2004; Scheier, Griffin, Doyle, & Botvin, 2002; Siddiqui, Hedeker, Flay, & Hu, 1996).

Based on an assessment of factor loadings and fit indices, we found the eight indicators measuring student SEC to be an acceptable representation of the construct. Furthermore, we found that the structural model postulating relations between teacher SEC, proactive implementation leadership, number of SEL lessons delivered, and student year-end SEC—controlling for student grade category and student baseline SEC—adequately represented the data. In MSEM analysis, most model-fit indices test a global goodness of fit, including both *within* and *between* parts of the model. The methodological literature, however, suggests that these global fit indices may fail to detect a lack of fit at one level, especially at the cluster level, and therefore recommends calculating level-specific fit indices (Ryu & West, 2009; Yuan & Bentler, 2007). In Mplus, only the SRMR reports a level-specific model fit, which has been found to be more sensitive to model misspecification at each level than other global fit indices (Hsu, Kwok, Lin, & Acosta, 2015). In our example, the $SRMR_{within}$ and $SRMR_{between}$ were both acceptable (i.e., ≤ .08) in addition to other conventional global fit indices. That said, one might desire further evaluations of the model fit (for one approach, see Ryu, 2014) beyond the scope of this analysis.

Substantively, we learned that student baseline SEC predicted student year-end SEC. In other words, where a student started was predictive of where a student ended. Yet, what happened during the year did matter. As several studies of SEL programs have found (Domitrovich, Gest, Jones, Gill, & DeRousie, 2010; Reyes, Brackett, Rivers, Elbertson, & Salovey, 2012)—although never before in the study of TOOLBOX—the number of lessons delivered by teachers predicted student year-end SEC, controlling for baseline SEC and grade category. As conceptualized by Jennings and Greenberg (2009) as well as Jones and Bouffard (2012), teachers' self-perceived SEC was also directly related to the number of lessons taught and directly related to year-end student SEC. We found evidence for partial mediation of the relationship between teacher SEC and student year-end SEC by the number of lessons taught. In addition, we found a moderating effect of proactive implementation leadership on this mediation, suggesting that when teacher reports of proactive leadership were higher at the beginning of the year, teachers' own SEC did not explain the number of SEL lessons that they ultimately delivered to the same extent as when the leadership was reported as less proactive.

To our knowledge, using an MSEM technique has enabled the first empirical examination of how leadership behaviors and teachers' own SEC influence students' social and emotional development in the context of SEL implementation. Prior research has shown that teacher characteristics that vary between classrooms influence SEL program implementation and outcomes (Wanless & Domitrovich, 2015). Yet, studies exploring how teacher-level characteristics relate to student outcomes are often constrained by the analytic tools available to the researcher. Some studies use MLM to account for nested data but then encounter analytic limitations that lead to running separate models for each outcome variable and not specifying structural relationships between predictors, such as how teacher training relates to implementation quality or dosage (e.g., Reyes et al., 2012). Studies evaluating interventions designed to improve student outcomes by promoting teacher SEC have examined the program effects on teachers using analysis of covariance; thus far, however, they have been unable to test how those teacher-level changes relate to student-level outcomes (e.g., Jennings et al., 2013). Also, studies exploring the relationships between implementation characteristics and student outcomes in school-based preventive interventions have used analysis of covariance without accounting for the nested structure of the data (e.g., Kam et al., 2003). By applying the MSEM method, however, we learned not only that teacher SEC relates to student outcomes, but we also identified a mechanism (number of lessons delivered) through which it operates and a contextual moderator (the lack of proactive leadership) in which the mechanism is strongest. Identifying mechanisms of intervention effectiveness and contexts in which they are salient raises the possibility of refining interventions and/or targeting implementation supports for taking preventive interventions to scale.

Despite the advantages and potentials of this method, the actual practice of MSEM techniques can be quite challenging. Success may be dependent on data quality (e.g., number of clusters, magnitude of ICCs, extent of missingness), and as models become increasingly complex, computational problems (e.g., nonconvergence, unstable or biased estimation) can arise (Rabe-Hesketh et al., 2007). The maximum likelihood approach using the expectation-maximization algorithm with a sandwich estimator (i.e., FIML, or the MLR estimator in Mplus) is often used with MSEM to produce unbiased estimates and robust standard errors (Kaplan, 2008; Schminkey, von Oertzen, & Bullock, 2016), but this approach requires large sample sizes to obtain estimates with asymptotically useful properties (e.g., bias goes to 0 as sample size goes to infinity; Holtmann, Koch, Lochner, & Eid, 2016). Low ICCs can also lead to model-fit and parameter-estimation problems, especially at the cluster level (Hsu, Lin, Kwok, Acosta, & Willson, 2017). Furthermore, the FIML approach to handling missing data assumes missing at random (i.e., missingness for an outcome variable does not depend on the value of outcome itself, after adjusting for a set of predictors), which, unfortunately, is not statistically testable (Allison, 2009). Requirements for the use of MSEM in regard to sample size, ICC, and missingness are largely unavailable, but these questions are being explored and recommendations are forthcoming (Preacher et al., 2010; Christ et al., 2017). In addition to these technical challenges, it should be noted that MSEM findings do not automatically indicate causal evidence unless all the model assumptions (e.g., about the relationships among variables, residuals) are correct and conditions for causal inference (e.g., time-order, no omitted confounders) are met.

The measurement and structural models we suggest fit the routine-practice data without any convergence problems or unstable results (e.g., negative variances), with a cluster-level sample size of 82, outcome variable ICC estimates ranging from .08–.23, and some missing data. Yet, we acknowledge that adding more complexity to these models could easily lead to computational problems. For example, we were unable to model baseline SEC as a latent construct (indicated by the eight items comprising the DESSA-Mini rather than the single DESSA-Mini total score used in clinical practice) or additional student-level covariates due to nonconvergence. Further, we acknowledge the imperfections of the data used in our illustration for making broad claims (e.g., a single self-report indicator of teacher SEC, data derived from a single school district). We strongly suggest testing assumptions we made (e.g., the program is similarly effective across student differences by race, gender, and socioeconomic status) in future studies of this and other school-based prevention programs.

To achieve the grand challenge goal of unleashing the power of prevention, or to take any effective social work intervention to scale, we need to look beyond mean differences in outcomes under optimal conditions. In order to determine how and under what conditions an intervention leads to desired outcomes, sophisticated analytic approaches such as MSEM should be used to enable the testing of mediation and moderation hypotheses with nested data. This introduction to MSEM may help researchers answer practice-relevant questions important for refining preventive interventions to ensure the healthy development of all youth and disseminating those interventions at scale.

References

Aarons, G. A., Ehrhart, M. G., & Farahnak, L. R. (2014). The implementation leadership scale (ILS): Development of a brief measure of unit level implementation leadership. *Implementation Science*, *9*(1), 45. https://doi.org/10.1186/1748-5908-9-45

Accomazzo, S., Ziemer, K. L., Shapiro, V. B., & Kim, B. K. E. (2018). Teacher's assessment of "implementation leadership" during a new social emotional learning initiative. Manuscript under review.

Allison, P. D. (2009). Missing data. In R. Millsap & A. Maydeu-Olivares (Eds.), *The SAGE handbook of quantitative methods in psychology* (pp. 73–90). London, England: SAGE Publications.

Asparouhov, T., & Muthén, B. (2006). Constructing covariates in multilevel regression. *Mplus Web Notes: No. 11*. Retrieved from http://statmodel2.com/download/webnotes/webnote11.pdf

Bliese, P. D. (2000). Within-group agreement, nonindependence, and reliability: Implications for data aggregation and analysis. In K. J. Klein & S. W. J. Kozlowski (Eds.), *Multi-level theory, research and methods in organizations: Foundations, extensions, and new directions* (pp. 349–381). San Francisco, CA: Jossey-Bass.

Bowen, N. K., & Guo, S. (2011). *Structural equation modeling*. New York, NY: Oxford University Press.

Bridgeland, J., Bruce, M., & Hariharan, A. (2013). *The missing piece: A national teacher survey on how social and emotional learning can empower children and transform schools. A report for CASEL.* Washington, DC: Civic Enterprises.

Cheung, M. W.-L., & Au, K. (2005). Applications of multilevel structural equation modeling to cross-cultural research. *Structural Equation Modeling*, *12*(4), 598–619. http://dx.doi.org/10.1207/s15328007sem1204_5

Christ, O., Hewstone, M., Schmid, K., Green, E. G., Sarrasin, O., Gollwitzer, M., & Wagner, U. (2017). Advanced multilevel modeling for a science of groups: A short primer on multilevel structural equation modeling. *Group Dynamics: Theory, Research, and Practice*, *21*(3), 121–134. http://dx.doi.org/10.1037/gdn0000065

Collin, M. A. (2015). *TOOLBOX™ primer*. Sebastopol, CA: Dovetail Learning.

Davidov, E., Dülmer, H., Schlüter, E., Schmidt, P., & Meuleman, B. (2012). Using a multilevel structural equation modeling approach to explain cross-cultural measurement noninvariance. *Journal of Cross-Cultural Psychology*, *43*(4), 558–575. http://dx.doi.org/10.1207/s15328007sem1204_5

Domitrovich, C. E., Gest, S. D., Jones, D., Gill, S., & DeRousie, R. M. S. (2010). Implementation quality: Lessons learned in the context of the Head Start REDI trial. *Early Childhood Research Quarterly*, *25*(3), 284–298. https://doi.org/https://doi.org/10.1016/j.ecresq.2010.04.001

Durlak, J. A., & DuPre, E. P. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology*, *41*(3–4), 327–350. https://doi.org/http://dx.doi.org/10.1007/s10464-008-9165-0

Durlak, J. A., Weissberg, R. P., Dymnicki, A. B., Taylor, R. D., & Schellinger, K. B. (2011). The impact of enhancing students' social and emotional learning: A meta-analysis of school-based universal interventions. *Child Development*, *82*(1), 405–432. doi:10.1111/j.1467-8624.2010.01564.x

Flay, B. R., Biglan, A., Boruch, R. F., Castro, F. G., Gottfredson, D., Kellam, S., ... Ji, P. (2005). Standards of evidence: Criteria for efficacy, effectiveness and dissemination. *Prevention Science*, *6*(3), 151–175. https://doi.org/10.1007/s11121-005-5553-y

Geldhof, G. J., Preacher, K. J., & Zyphur, M. J. (2014). Reliability estimation in a multilevel confirmatory factor analysis framework. *Psychological Methods*, *19*(1), 72. https://doi.org/http://dx.doi.org/10.1037/a0032138

Gottfredson, D. C., Cook, T. D., Gardner, F. E., Gorman-Smith, D., Howe, G. W., Sandler, I. N., & Zafft, K. M. (2015). Standards of evidence for efficacy, effectiveness, and scale-up research in prevention science: Next generation. *Prevention Science, 16*(7), 893–926. https://doi.org/10.1007/s11121-015-0555-x

Hair, J. K., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). Structural equations modeling overview. In *Multivariate data analysis* (7th ed., pp. 611–667). Upper Saddle River, NJ: Pearson.

Hawkins, J. D., Jenson, J. M., Catalano, R. F., Fraser, M. W., Botvin, G. J., Shapiro, V. B., ... Stone, S. (2015). Unleashing the power of prevention (Discussion Paper). Washington, DC: Institute of Medicine and National Research Council. Retrieved from http://nam.edu/wp-content/uploads/2015/06/DPPowerofPrevention.pdf

Hawkins, J. D., Shapiro, V. B., & Fagan, A. A. (2010). Disseminating effective community prevention practices: Opportunities for social work education. *Research on Social Work Practice*, *20*(5), 518–527. https://doi.org/10.1177/1049731509359919

Holtmann, J., Koch, T., Lochner, K., & Eid, M. (2016). A comparison of ML, WLSMV, and Bayesian methods for multilevel structural equation models in small samples: A simulation study. *Multivariate Behavioral Research*, *51*(5), 661–680. https://doi.org/https://doi.org/10.1080/00273171.2016.1208074

Hox, J. J. (2010). *Multilevel analysis: Techniques and applications* (2nd ed.). New York, NY: Routledge.

Hsu, H. Y., Kwok, O. M., Lin, J. H., & Acosta, S. (2015). Detecting misspecified multilevel structural equation models with common fit indices: A Monte Carlo study. *Multivariate Behavioral Research*, *50*(2), 197–215. https://doi.org/10.1080/00273171.2014.977429

Hsu, H. Y., Lin, J. H., Kwok, O. M., Acosta, S., & Willson, V. (2017). The impact of intraclass correlation on the effectiveness of level-specific fit indices in multilevel structural equation modeling: A Monte Carlo study. *Educational and Psychological Measurement*, *77*(1), 5–31. https://doi.org/10.1177/0013164416642823

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

Huta, V. (2014). When to use hierarchical linear modeling. *The Quantitative Methods for Psychology*, *10*(1), 13–28. https://doi.org/10.20982/tqmp.10.1.p013

Jennings, P. A., Frank, J. L., Snowberg, K. E., Coccia, M. A., & Greenberg, M. T. (2013). Improving classroom learning environments by Cultivating Awareness and Resilience in Education (CARE): Results of a randomized controlled trial. *School Psychology Quarterly*, *28*(4), 374–390. https://doi.org/http://dx.doi.org/10.1037/spq0000035

Jennings, P. A., & Greenberg, M. T. (2009). The prosocial classroom: Teacher social and emotional competence in relation to student and classroom outcomes. *Review of Educational Research*, *79*(1), 491–525. https://doi.org/https://doi.org/10.3102/0034654308325693

Jones, S. M., & Bouffard, S. M. (2012). Social and emotional learning in schools: From programs to strategies. *Social Policy Report*, *26*(4), 3–22Ann Arbor, MI: Society for Research in Child Development. Retrieved from https://files.eric.ed.gov/fulltext/ED540203.pdf

Kam, C.-M., Greenberg, M. T., & Walls, C. T. (2003). Examining the role of implementation quality in school-based prevention using the PATHS curriculum. *Prevention Science*, *4*(1), 55–63. https://doi.org/https://doi.org/10.1023/A:1021786811186

Kaplan, D. (2008). *Structural equation modeling: Foundations and extensions* (2nd ed.). Thousand Oaks, CA: SAGE Publications.

Kaplan, D., Kim, J.-S., & Kim, S.-Y. (2009). Multilevel latent variable modeling: Current research and recent developments. In R. E. Millsap & A. Maydeu-Olivares (Eds.), *The SAGE Handbook of Quantitative Methods in Psychology* (pp. 592–612). Thousand Oaks, CA: SAGE Publications.

Kline, R. B. (2015). *Principles and practice of structural equation modeling* (4th ed.). New York, NY: Guilford Publications.

LeBuffe, P. A., Shapiro, V. B., & Robitaille, J. L. (2017). The Devereux Student Strengths Assessment (DESSA) comprehensive system: Screening, assessing, planning, and monitoring. *Journal of Applied Developmental Psychology*, *55*, 62–70. https://doi.org/10.1016/j.appdev.2017.05.002

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*(3), 203–229. http://dx.doi.org/10.1037/a0012869

MacKinnon, D. P., & Valente, M. J. (2014). Mediation from multilevel to structural equation modeling. *Annals of Nutrition and Metabolism*, *65*(2–3), 198–204. https://doi.org/10.1159/000362505

Mehta, P. D., & Neale, M. C. (2005). People are variables too: Multilevel structural equations modeling. *Psychological Methods*, *10*(3), 259–284. http://dx.doi.org/10.1037/1082-989X.10.3.259

Murray, D. M., Catellier, D. J., Hannan, P. J., Treuth, M. S., Stevens, J., Schmitz, K. H., ... Conway, T. L. (2004). School-level intraclass correlation for physical activity in adolescent girls. *Medicine and Science in Sports and Exercise*, *36*(5), 876–882. https://doi.org/10.1249/01.mss.0000126806.72453.1c

Muthén, B., & Asparouhov, T. (2008). Growth mixture modeling: Analysis with non-Gaussian random effects. In G. Fitzmaurice, M. Davidian, & G. V. Geert (Eds.), *Longitudinal Data Analysis* (pp. 143–165). Boca Raton, FL: Chapman & Hall/CRC Taylor & Francis Group.

Muthén, L. K., & Muthén, B. O. (1998–2017). *Mplus user's guide.* (8th ed.). Los Angeles, CA: Authors.

Naglieri, J. A., LeBuffe, P. A., & Shapiro, V. (2011/2014). *Devereux student strengths assessment—Mini (DESSA-Mini): Assessment, technical manual, and user's guide*. Charlotte, NC: Apperson.

O'Connell, M. E., Boat, T., & Warner, K. E. (Eds.). (2009). *Preventing mental, emotional, and behavioral disorders among young people: Progress and possibilities*. Washington DC: National Academies Press.

Preacher, K. J., & Selig, J. P. (2012). Advantages of Monte Carlo confidence intervals for indirect effects. *Communication Methods and Measures, 6*, 77–98. https://doi.org/10.1080/19312458.2012.679848

Preacher, K. J., Zhang, Z., & Zyphur, M. J. (2011). Alternative methods for assessing mediation in multilevel data: The advantages of multilevel SEM. *Structural Equation Modeling*, *18*(2), 161–182. https://doi.org/https://doi.org/10.1080/10705511.2011.557329

Preacher, K. J., Zhang, Z., & Zyphur, M. J. (2016). Multilevel structural equation models for assessing moderation within and across levels of analysis. *Psychological Methods*, *21*(2), 189–205. http://dx.doi.org/10.1037/met0000052

Preacher, K. J., Zyphur, M. J., & Zhang, Z. (2010). A general multilevel SEM framework for assessing multilevel mediation. *Psychological Methods*, *15*(3), 209. http://dx.doi.org/10.1037/a0020141

Rabe-Hesketh, S., Skrondal, A., & Pickles, A. (2004). Generalized multilevel structural equation modeling. *Psychometrika, 69*(2), 167–190. https://doi.org/10.1007/BF02295939

Rabe-Hesketh, S., Skrondal, A., & Zheng, X. (2007). Multilevel structural equation modeling. In S. Lee (Ed.), *Handbook of latent variable and related models* (pp. 209–227). Amsterdam, The Netherlands: Elsevier.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: SAGE Publications.

Reyes, M. R., Brackett, M. A., Rivers, S. E., Elbertson, N. A., & Salovey, P. (2012). The interaction effects of program training, dosage, and implementation quality on targeted student outcomes for the RULER approach to social and emotional learning. *School Psychology Review*, *41*(1), 82–99. Retrieved from http://ei.yale.edu/wp-content/uploads/2013/06/pub318_Reyesetal2012_SPR.pdf

Ryu, E. (2014). Model fit evaluation in multilevel structural equation models. *Frontiers in Psychology*, *5*, 1–9. https://doi.org/https://doi.org/10.3389/fpsyg.2014.00081

Ryu, E., & West, S. G. (2009). Level-specific evaluation of model fit in multilevel structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, *16*(4), 583–601. https://doi.org/https://doi.org/10.1080/10705510903203466

Scheier, L. M., Griffin, K. W., Doyle, M. M., & Botvin, G. J. (2002). Estimates of intragroup dependence for drug use and skill measures in school-based drug abuse prevention trials: An empirical study of three independent samples. *Health Education & Behavior*, *29*(1), 85–103. https://doi.org/10.1177/109019810202900109

Schminkey, D. L., von Oertzen, T., & Bullock, L. (2016). Handling missing data with multilevel structural equation modeling and full information maximum likelihood techniques. *Research in Nursing & Health*, *39*(4), 286–297. https://doi.org/10.1002/nur.21724

Schonert-Reichl, K. A., & O’Brien, M. U. (2012). Social and emotional learning and prosocial education: Theory, research, and programs. In P. Brown, M. W. Corrigan, & A. Higgins-D’Alessandro (Eds.), *Handbook of Prosocial Education* (Vol. 1, pp. 311–345). Lanham, MD: Rowman & Littlefield.

Selig, J. P., & Preacher, K. J. (2008). Monte Carlo method for assessing mediation: An interactive tool for creating confidence intervals for indirect effects [Computer software]. Available from http://quantpsy.org/medmc/medmc.htm

Shapiro, V. B., Accomazzo, S., & Robitaille, J. L. (2017). In the same ballpark or a whole new ball game? Staff as raters of youth behavior. *Journal of Child and Family Studies*, *26*(4), 1051–1055. https://doi.org/10.1007/s10826-016-0632-1

Shapiro, V. B., Hawkins, J. D., & Oesterle, S. (2015). Building local infrastructure for community adoption of science-based prevention: The role of coalition functioning. *Prevention Science*, *16*(8), 1136–1146. https://doi.org/10.1007/s11121-015-0562-y

Shapiro, V. B., Hawkins, J. D., Oesterle, S., Monahan, K. C., Brown, E. C., & Arthur, M. W. (2013). Variation in the effect of Communities That Care on community adoption of a scientific approach to prevention. *Journal of the Society for Social Work and Research*, *4*(3), 154–164. https://doi.org/10.5243/jsswr.2013.10

Shapiro, V. B., Kim, B. K. E., Accomazzo, S., & Roscoe, J. N. (2016). Predictors of rater bias in the assessment of social-emotional competence. *International Journal of Emotional Education*, *8*(2), 25.

Shapiro, V. B., Kim, B. K. E., Robitaille, J. L., LeBuffe, P. A., & Ziemer, K. L. (2018). Efficient implementation monitoring in routine prevention practice: A grand challenge for schools. *Journal of the Society for Social Work and Research*, *9*(3), 377–394. https://doi.org/10.1086/699153

Shapiro, V. B., Whitaker, K., Kim, B. K. E., & Roscoe, J. N. (2018). *Introducing the Social Emotional Learning Implementation Survey (SEL-IS) for the routine implementation monitoring of universal school-based programs.* Manuscript under review.

Siddiqui, O., Hedeker, D., Flay, B. R., & Hu, F. B. (1996). Intraclass correlation estimates in a school-based smoking prevention study: Outcome and mediating variables, by sex and ethnicity. *American Journal of Epidemiology*, *144*(4), 425–433. https://doi.org/10.1093/oxfordjournals.aje.a008945

Smith-Millman, M. K., Flaspohler, P. D., Maras, M. A., Splett, J. W., Warmbold, K., Dinnen, H., & Luebbe, A. (2017). Differences between teacher reports on universal risk assessments. *Advances in School Mental Health Promotion*, *10*(4), 235–249. https://doi.org/https://doi.org/10.1080/1754730X.2017.1333914

Spoth, R., Rohrbach, L. A., Greenberg, M., Leaf, P., Brown, C. H., Fagan, A., ... Society for Prevention Research Type 2 Translational Task Force Members and Contributing Authors. (2013). Addressing core challenges for the next generation of type 2 translation research and systems: The translation science to population impact (TSci Impact) framework. *Prevention Science*, *14*(4), 319–351. https://doi.org/10.1007/s11121-012-0362-6

Taylor, R. D., Oberle, E., Durlak, J. A., & Weissberg, R. P. (2017). Promoting positive youth development through school-based social and emotional learning interventions: A meta-analysis of follow-up effects. *Child Development*, *88*(4), 1156–1171. http://doi.org/10.1111/cdev.12864

Uehara, E., Flynn, M., Fong, R., Brekke, J., Barth, R. P., Coulton, C., ... Walters, K. (2013). Grand challenges for social work. *Journal of the Society for Social Work and Research*, *4*(3), 165–170. https://doi.org/10.5243/jsswr.2013.11

Wanless, S. B., & Domitrovich, C. E. (2015). Readiness to implement school-based social-emotional learning interventions: Using research on factors related to implementation to maximize quality. *Prevention Science*, *16*(8), 1037–1043. https://doi.org/10.1007/s11121-015-0612-5

Weissberg, R. P., Durlak, J. A., Domitrovich, C. E., & Gullotta, T. P. (2015). Social and emotional learning: Past, present, and future. In J. A. Durlak, C. E. Domitrovich, R. P. Weissberg, & T. P. Gullotta (Eds.), *Handbook of social and emotional learning: Research and practice* (pp. 3–19). New York, NY: Guilford Publications.

Woltman, H., Feldstain, A., MacKay, J. C., & Rocchi, M. (2012). An introduction to hierarchical linear modeling. *Tutorials in Quantitative Methods for Psychology*, *8*(1), 52–69. https://doi.org/10.20982/tqmp.08.1.p052

Yuan, K.-H., & Bentler, P. M. (2007). Multilevel covariance structure analysis by fitting multiple single-level models. *Sociological Methodology*, *37*(1), 53–82. doi:10.1111/j.1467-9531.2007.00182.x

Table 1
ICCs of the Eight Items Measuring Students' SEC

	SEC1	SEC2	SEC3	SEC4	SEC5	SEC6	SEC7	SEC8
ICCs for classroom level	.11	.09	.11	.08	.23	.15	.15	.20
ICCs for school level	.00	.00	.00	.00	.01	.00	.00	.00

Note. ICC = intraclass correlation; SEC = social–emotional competence.

Table 2
Factor Loadings of the Eight Items Measuring Students' SEC

Variable	Standardized Estimate	Standard Error	*t* Statistic	*p* Value
Within level: SECW				
SEC1	.89	.01	84.76	< .001
SEC2	.92	.01	120.44	< .001
SEC3	.90	.01	79.13	< .001
SEC4	.89	.01	102.53	< .001
SEC5	.89	.01	73.97	< .001
SEC6	.89	.01	74.37	< .001
SEC7	.91	.01	90.24	< .001
SEC8	.86	.02	47.34	< .001
Between level: SECB				
SEC1	.93	.03	33.65	<.001
SEC2	.95	.02	46.72	< .001
SEC3	.90	.03	33.22	< .001
SEC4	.90	.05	19.98	< .001
SEC5	.88	.05	17.12	< .001
SEC6	.97	.02	52.70	< .001
SEC7	.83	.09	10.13	< .001
SEC8	.85	.06	15.04	< .001

Note. SEC = social–emotional competence; SECW = students' year-end SEC at individual level; SECB = students' year-end SEC at classroom level.

Table 3

Relationships Among Variables (Before Including NLES)

	Unstandardized Estimate	Standard Error	*t* Statistic	*Pp* Value
Within level: SECW				
SECPRE	.10	.00	22.04	< .001
Between level: SECB				
TSEC	.57	.23	2.48	.013
SECPRE	.14	.04	3.90	< .001
GRADK2	.55	.32	1.72	.086

Note. SECW = students' year-end social–emotional competence (SEC) at individual level; SECB = students' year-end SEC at classroom level; TSEC = teachers' baseline SEC; NLES = number of lessons delivered; SECPRE = students' baseline SEC; GRADK2 = students' grade level (1 = Grades K–2, 0 = Grades 3–5).

Table 4
Relationships Among Variables (After Including NLES)

	Unstandardized Estimate	Standard Error	*t* Statistic	P Value
Within level: SECW				
SECPRE	0.10	.00	22.04	< .001
Between level: SECB				
TSEC	0.50	.23	2.15	.032
NLES	0.11	.05	2.47	.014
SECPRE	0.11	.04	3.13	.002
GRADK2	0.76	.35	2.16	.031
Between level: NLES				
TSEC	1.33	.54	2.44	.015
SECPRE	0.33	.09	3.63	< .001
GRADK2	−1.44	.85	−1.70	.089

Note. SECW = students' year-end social–emotional competence (SEC) at individual level; SECB = students' year-end SEC at classroom level; TSEC = teachers' baseline SEC; NLES = number of lessons delivered; SECPRE = students' baseline SEC; GRADK2 = students' grade level (1 = Grades K–2, 0 = Grades 3–5).

Table 5

Relationships Among Variables (After Including PROSL)

	Unstandardized Estimate	**Standard Error**	***t* Statistic**	***p* Value**
Within level: SECW				
SECPRE	0.10	0.01	20.14	< .001
Between level: SECB				
TSEC	0.44	0.24	1.85	.065
NLES	0.12	0.05	2.44	.015
SECPRE	0.11	0.04	3.02	.003
GRADK2	0.73	0.38	1.95	.051
Between level: NLES				
TSEC	4.78	1.39	3.45	.001
PROSL	4.78	1.78	2.68	.007
INTERACT	−1.38	0.57	−2.43	.015
SECPRE	0.24	0.11	2.30	.021
GRADK2	−1.30	0.84	−1.55	.121

Note. SECW = students' year-end social–emotional competence (SEC) at individual level; SECB = students' year-end SEC at classroom level; TSEC = teachers' baseline SEC; NLES = number of lessons delivered; PROSL = teacher-reported proactive school leadership at baseline; INTERACT = interaction between TSEC and PROSL; SECPRE = students' baseline SEC; GRADK2 = students' grade level (1 = Grades K–2, 0 = Grades 3–5).

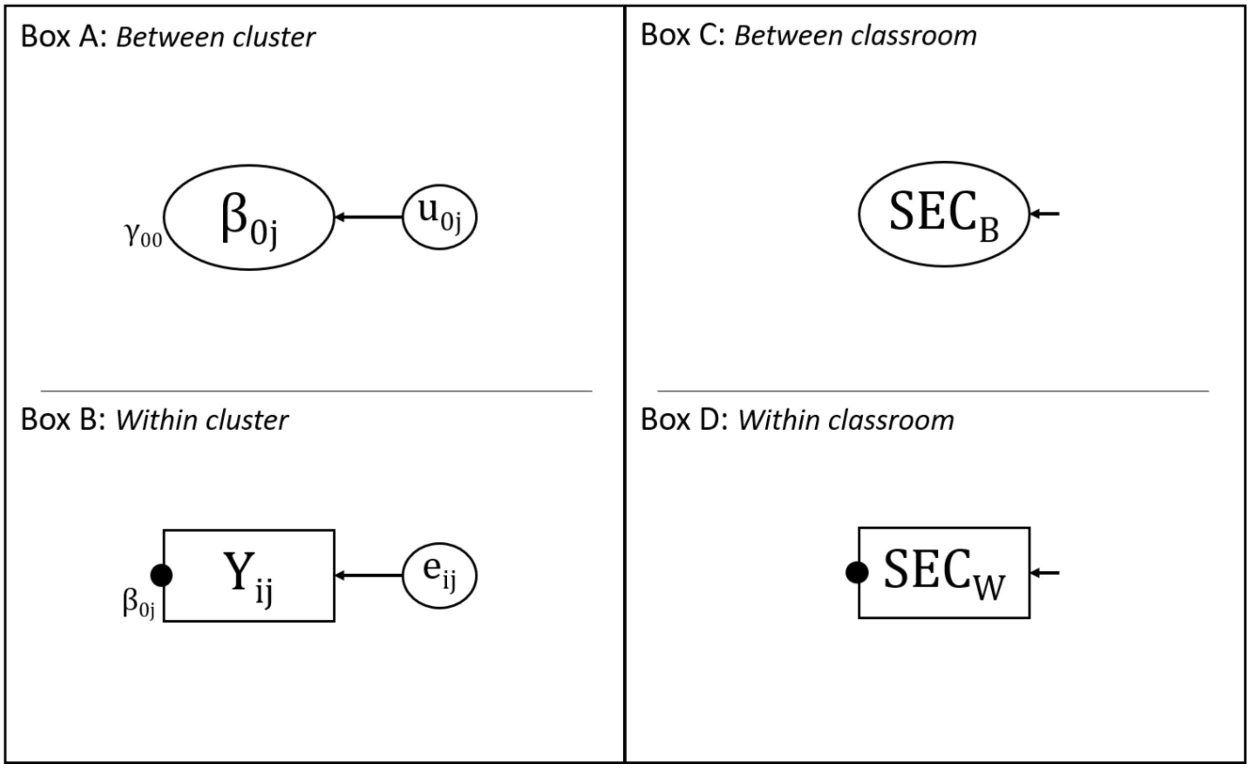

Figure 1. Random intercept model with no predictor. In Box B, Y_{ij}, an observed variable, is denoted as a rectangle with an intercept (β_{0j}) and within-cluster residuals (e_{ij}). Variation of the intercept across clusters is represented by the solid dot left of the rectangle in Box B and the oval with random intercept (β_{0j}) in Box A. The random intercept is not a set of observed cluster means but a random variable that has an unobservable variance across clusters. This random intercept consists of an intercept (γ_{00}) and between-cluster residuals (u_{0j}). The subscript *B* in Box C represents "between"; in Box D, the subscript *W* represents "within."

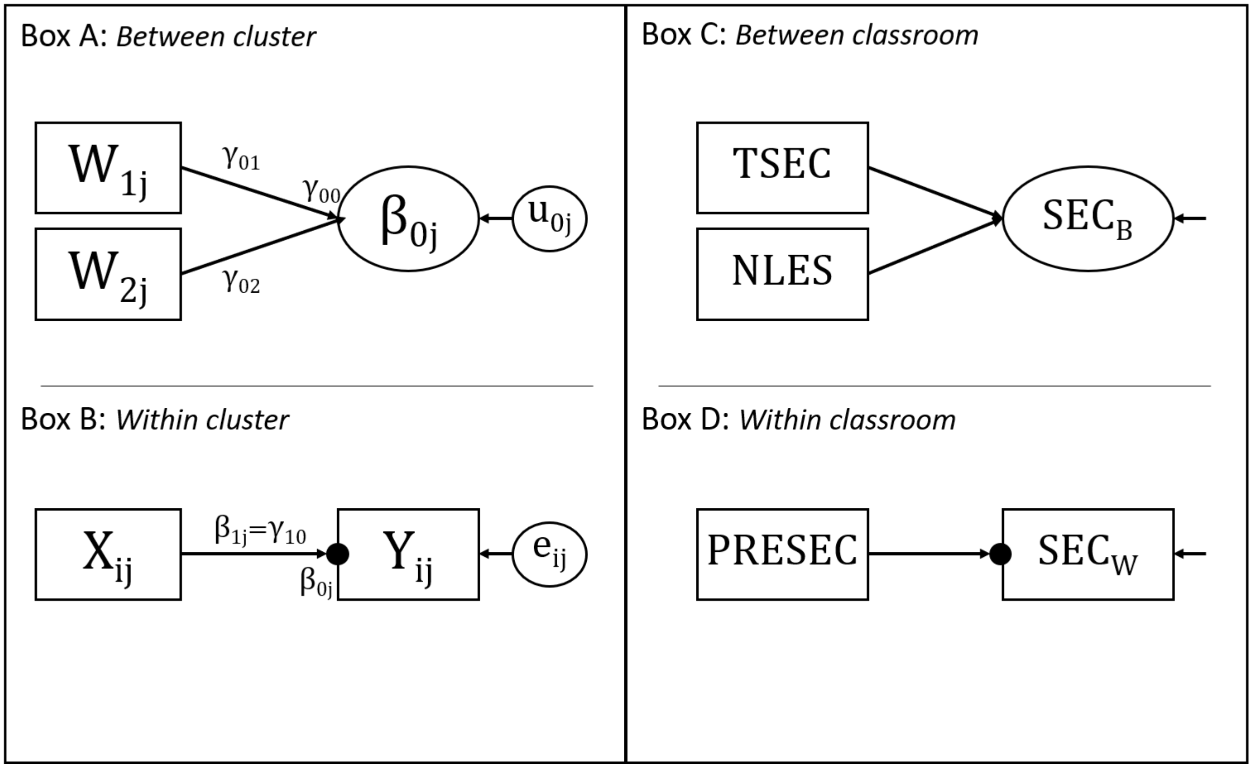

Figure 2. Random intercept model with predictors. In Box A, the classroom-level predictors (W_{1j} and W_{2j}) point to the random intercept (β_{0j}), which has its own intercept (γ_{00}) and between-cluster residuals (u_{0j}). In Box B, the student-level predictor (X_{ij}) in a rectangle points to the outcome variable (Y_{ij}), which has a random intercept (β_{0j} with a solid dot) and within-cluster residuals (e_{ij}). The arrows pointing to the outcome variable at each level represent the regression slopes of the predictors (γ_{10}, γ_{01}, and γ_{02}). Boxes C and D depict example predictors: students' year-end social–emotional competence (SEC) and baseline SEC (PRESEC) at the student level, and the teacher's SEC (TSEC) and the number of lessons delivered (NLES) at the classroom level.

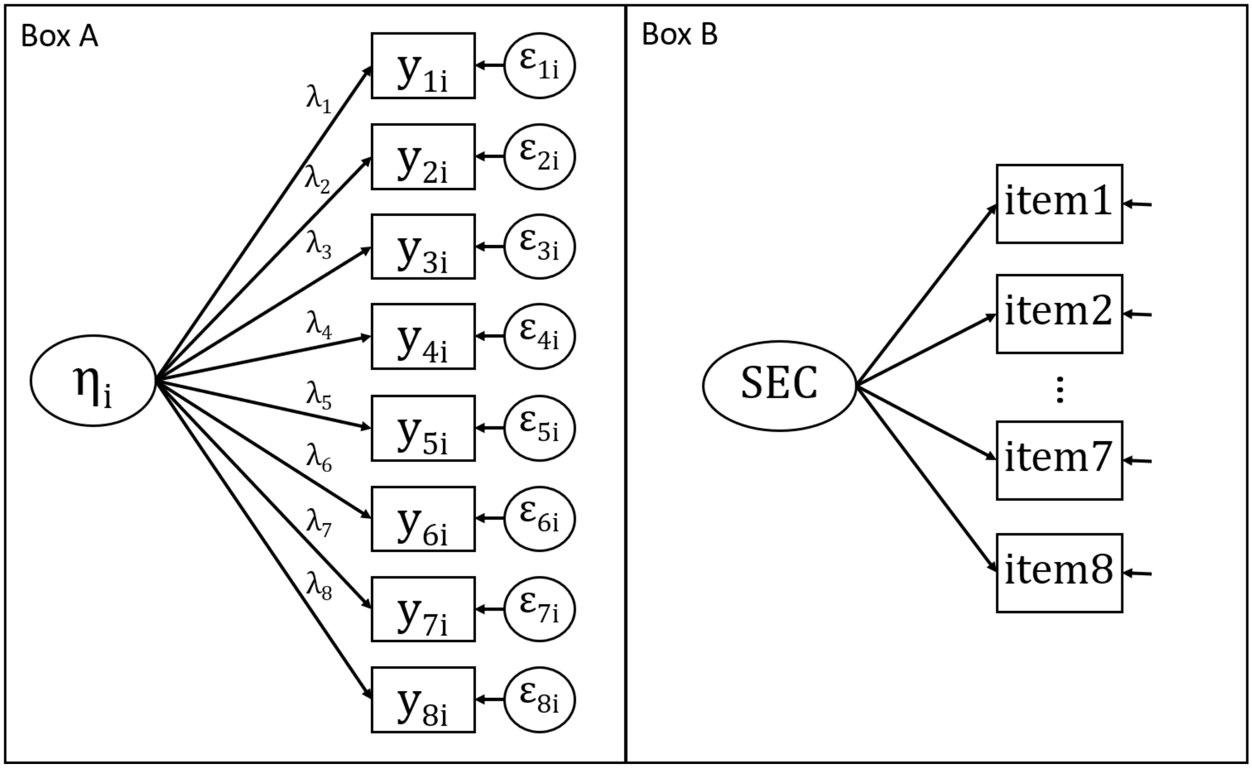

Figure 3. Measurement model. In Box A, each indicator is depicted as a rectangle, pointed to by two arrows, one from a single latent construct and the other from its own source of unknown measurement errors (both unmeasured and depicted as ovals). Box B shows a simplified diagram with illustrative example variables where each indicator is named as teacher-rated Items 1–8 with the points of ellipsis in the middle representing Items 3–6. SEC = social–emotional competence.

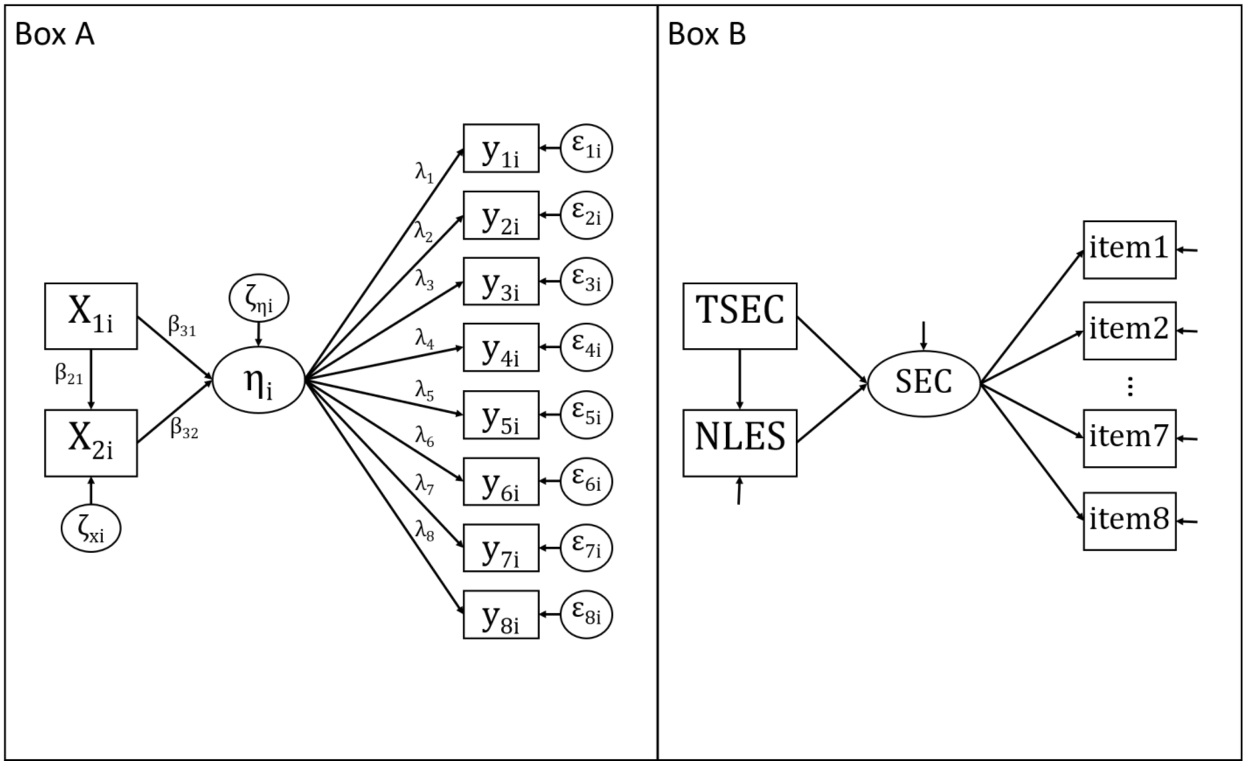

Figure 4. Structural model. The latent construct, measured by eight indicators, is predicted directly by X_{1i}, with a slope of β_{31}, and also is predicted indirectly through the mediating variable (X_{2i}). This indirect relationship is noted as X_{1i} predicting X_{2i} with a slope of β_{21} as well as X_{2i} predicting η_i with a slope of β_{32}. Box B represents a simplified diagram where the teacher's social–emotional competence (TSEC) is predicting students' social–emotional competence (SEC) directly and indirectly through the number of social and emotional learning lessons delivered during the intervention year (NLES).

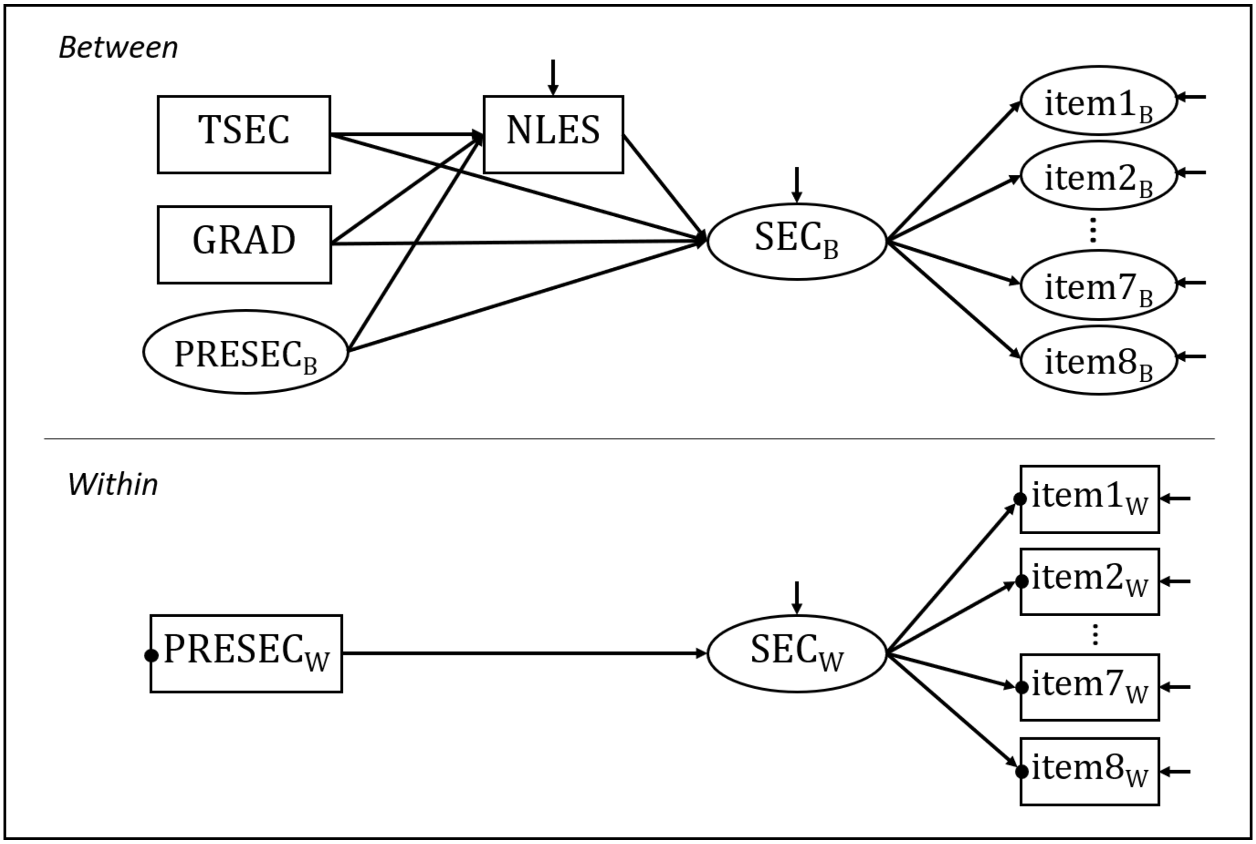

Figure 5. Multilevel mediation model. In the *Within* section, the observed indicators of dependent variable (item1_W–item8_W) and the independent variable ($PRESEC_W$ = students' baseline SEC) in rectangles represent the within-classroom variations of each observed variable. The solid dots to the left of these rectangles indicate that the intercepts of these variables randomly vary across classrooms; these random intercepts are represented in ovals in the *Between* section (item1_B–item8_B and $PRESEC_B$). The teacher's social–emotional competence (TSEC), the number of social and emotional learning lessons delivered (NLES), and the grade level (GRAD) are represented in rectangles only in the *Between* section because they have no variation within classrooms.

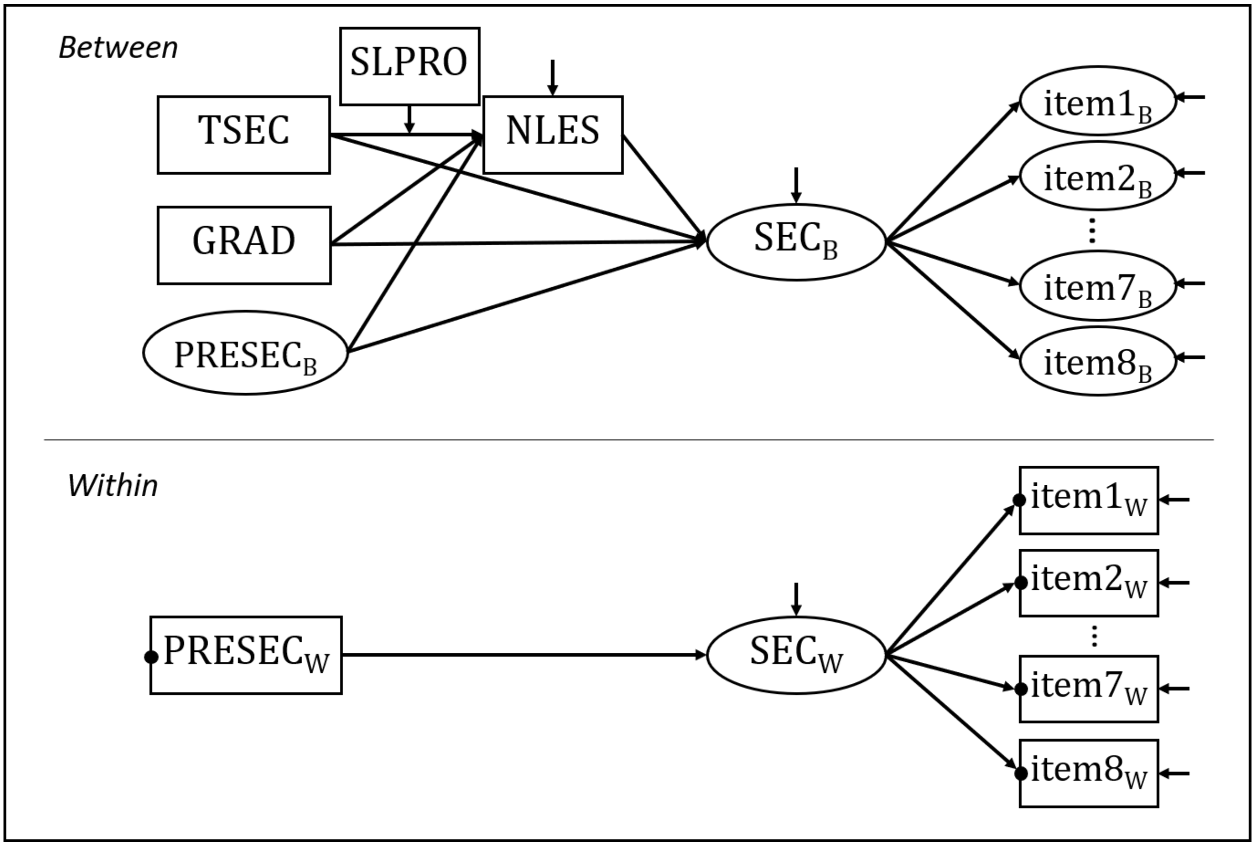

Figure 6. Multilevel moderated mediation model. In the *Within* section, the observed indicators of dependent variable ($item1_W$–$item8_W$) and the independent variable ($PRESEC_W$ = students' baseline SEC) in rectangles represent the within-classroom variations of each observed variable. The solid dots to the left of these rectangles indicate that the intercepts of these variables randomly vary across classrooms; these random intercepts are represented in ovals in the *Between* section ($item1_B$–$item8_B$ and $PRESEC_B$). The teacher's social–emotional competence (TSEC), the number of social and emotional learning lessons delivered (NLES), school proactive implementation leadership (SLPRO), and the grade level (GRAD) are represented in rectangles only in the *Between* section because they have no variation within classrooms.

Appendix: Mplus Codes for the Demonstrated Analyses

Step 1: Intraclass Correlations

```
TITLE: SEL working mechanism model - Intraclass correlations;

DATA:
FILE = toolbox_example.dat;
FORMAT = 15F5.0;

VARIABLE:
NAMES = stuid classid schoolid
sec1 sec2 sec3 sec4 sec5 sec6 sec7 sec8
secpre gradk2
tsec nles prosl;
MISSING = ALL(999);

USEVARIABLES = sec1;
CLUSTER = schoolid classid;

ANALYSIS:
TYPE = THREELEVEL;
ESTIMATOR= MLR;

MODEL:
%WITHIN%
sec1;

%BETWEEN classid%
sec1;

%BETWEEN schoolid%
sec1;
```

Note: First, repeat these codes eight times for each indicator by changing the variable name in the USEVARIABLES and MODEL commands. Repeat this procedure to get two-level intraclass correlations by changing the model type into TYPE = TWOLEVEL; and deleting the %BETWEEN schoolid% part of the MODEL.

Step 2: Multilevel Measurement Model

```
TITLE: SEL working mechanism model - measurement model;

DATA:
FILE = toolbox_example.dat;
FORMAT = 15F5.0;

VARIABLE:
NAMES = stuid classid schoolid
sec1 sec2 sec3 sec4 sec5 sec6 sec7 sec8
secpre gradk2
tsec nles prosl;
MISSING = ALL(999);

USEVARIABLES = sec1 sec2 sec3 sec4 sec5 sec6 sec7 sec8;
CLUSTER = classid;

ANALYSIS:
TYPE = TWOLEVEL;
ESTIMATOR= MLR;

MODEL:
%WITHIN%
secw BY sec1* sec2 sec3 sec4 sec5 sec6 sec7 sec8;
secw@1;

%BETWEEN%
secb BY sec1* sec2 sec3 sec4 sec5 sec6 sec7 sec8;
secb@1;

OUTPUT: STDYX;
```

Step 3: Multilevel Structural Model

Relationship between teachers' social–emotional competence (SEC) and students' year-end SEC.

```
TITLE: SEL working mechanism model - structural model 1;

DATA:
FILE = toolbox_example.dat;
```

```
FORMAT = 15F5.0;

VARIABLE:
NAMES = stuid classid schoolid
sec1 sec2 sec3 sec4 sec5 sec6 sec7 sec8
secpre gradk2
tsec nles prosl;
MISSING = ALL(999);

USEVARIABLES = sec1 sec2 sec3 sec4 sec5 sec6 sec7 sec8 secpre gradk2 tsec;
CLUSTER = classid;
BETWEEN = gradk2 tsec;

ANALYSIS:
TYPE = TWOLEVEL;
ESTIMATOR = MLR;

MODEL:
%WITHIN%
secw BY sec1* sec2 sec3 sec4 sec5 sec6 sec7 sec8;
secw@1;
secw ON secpre;

%BETWEEN%
secb BY sec1* sec2 sec3 sec4 sec5 sec6 sec7 sec8;
secb@1;
secb ON secpre gradk2 tsec;

OUTPUT: STDYX;
```

Mediation of the number of lessons delivered.

```
TITLE: SEL working mechanism model - structural model 2;

DATA:
FILE = toolbox_example.dat;
FORMAT = 15F5.0;

VARIABLE:
NAMES = stuid classid schoolid
```

```
sec1 sec2 sec3 sec4 sec5 sec6 sec7 sec8
secpre gradk2
tsec nles prosl;
MISSING = ALL(999);

USEVARIABLES = sec1 sec2 sec3 sec4 sec5 sec6 sec7 sec8 secpre gradk2 tsec nles;
CLUSTER = classid;
BETWEEN = gradk2 tsec nles;

ANALYSIS:
TYPE = TWOLEVEL;
ESTIMATOR = MLR;

MODEL:
%WITHIN%
secw BY sec1* sec2 sec3 sec4 sec5 sec6 sec7 sec8;
secw@1;
secw ON secpre;

%BETWEEN%
secb BY sec1* sec2 sec3 sec4 sec5 sec6 sec7 sec8;
secb@1;
secb ON secpre gradk2 tsec nles;
nles ON secpre gradk2 tsec;

OUTPUT: STDYX TECH1 TECH3;
```

Note: You can use Selig and Preacher’s (2008) Monte Carlo method for assessing mediation using the TECH1 and TECH3 outputs.

Moderation of proactive school leadership.

```
TITLE: SEL working mechanism model - structural model 3;

DATA:
FILE = toolbox_example.dat;
FORMAT = 15F5.0;

VARIABLE:
NAMES = stuid classid schoolid
sec1 sec2 sec3 sec4 sec5 sec6 sec7 sec8
```

```
secpre gradk2
tsec nles prosl;
MISSING = ALL(999);

USEVARIABLES = sec1 sec2 sec3 sec4 sec5 sec6 sec7 sec8 secpre gradk2 tsec nles
prosl interact;
CLUSTER = classid;
BETWEEN = gradk2 tsec nles prosl interact;

DEFINE:
interact = tsec * prosl;

ANALYSIS:
TYPE = TWOLEVEL;
ESTIMATOR = MLR;

MODEL:
%WITHIN%
secw BY sec1* sec2 sec3 sec4 sec5 sec6 sec7 sec8;
secw@1;
secw ON secpre;

%BETWEEN%
secb BY sec1* sec2 sec3 sec4 sec5 sec6 sec7 sec8;
secb@1;
secb ON secpre gradk2 tsec nles;
nles ON secpre gradk2 tsec prosl interact;

OUTPUT: STDYX TECH1 TECH3;
```