16

Multilevel Structural Equation Modeling

JOHN P. BARILE

Multilevel structural equation modeling (MSEM), an analytical technique that combines traditional multilevel regression and structural equation modeling (SEM), offers many advantages compared to traditional approaches in understanding community-based data. MSEM enables researchers to assess individual-level and higher level data simultaneously, while minimalizing ecological, atomistic, psychologistic, and sociologistic fallacies commonly present in evaluation and intervention research. Utilization of MSEM is often necessary to understand the diverse web of ecological determinants of individual and community well-being. This chapter will present the basic tenets of MSEM and identify circumstances in which this approach is most appropriate. It will then present an example of the use of MSEM in an evaluation of community coalitions, in which data from multiple sources at both the individual and collaborative levels were utilized to better understand the processes and outcomes associated with successful collaboration.

Although researchers have stressed the need to consider context for many years (Lewin, 1935), the development and accessibility of analytic tools that can simultaneously assess individuals and their environment have only recently emerged. Shinn and Rapkin (2000) outline the need for cross-level modeling (i.e., multilevel modeling) when researchers are interested in (a) the direct effects of a higher level variable on a lower level variable, (b) the level of deviation an individual has from a group standard, (c) the study of variables at multiple levels simultaneously, with one controlling for the other, (d) the moderating effect of a variable at one level on a relationship at another level, and (e) the effects of person-environment fit.

Shinn and Rapkin (2000) made strong cases for the need to measure both individual and contextual constructs when doing community-based research. These ideas were further trumpeted by Luke (2005), who urged community scientists to "get the big picture" by utilizing analytical tools that capture context by using a variety of methods, including multilevel modeling. These calls have been answered by a growing number of publications that have utilized multilevel methods. This chapter builds upon community-based methodologies presented by Jason and Glenwick (2012) by introducing MSEM as a means to study context. The following sections review traditional multilevel regression and SEM and then show how MSEM addresses limitations of the former methods. The chapter concludes with an application of MSEM to data on community collaboratives.

INTRODUCTION TO TRADITIONAL MULTILEVEL REGRESSION MODELING

MSEM is a pairing of traditional multilevel regression modeling and structural equation modeling. Traditional multilevel regression modeling incorporates data that exist on two or more levels. (Multilevel modeling can also include longitudinal analyses of multiple scores nested within an individual but for purposes of this chapter, we will focus only on contextual modeling of multilevel data.) For something to be considered multilevel, individuals (or the lowest level of measurement) must be nested within (i.e., be part of) a larger construct or group. Most often, the lowest level occurs at the individual level. Individual-level data can include a person's background, such as his or her age, education, and/or ethnicity, as well as his or her responses to survey questions (e.g., attitudes, perceptions) or an inventory of behaviors. Contextual variables include measures that describe a specific unit in which individuals are nested. Examples of individuals within nesting units include residents within neighborhoods, students within classrooms, or members within spiritual groups. Contextual variables of neighborhood nesting units can include, for example, estimates of social capital, the nature of local policies, or the number of parks; contextual variables of classroom nesting units can include measures of teacher quality, lesson plans, or absence policies. Measures at the nesting unit level can include variables that occur only at this higher level, such as the nature of local polices, levels of school funding, and diversity indexes. They can also be represented by aggregated individual responses, such as average teacher ratings or average fear of crime. For the purposes of this chapter, the highest level of a multilevel model will be referred to as the *group level*, but other sources also refer to it as the *organization level, cluster level, between level*, or *level two*.

Historically, the majority of research in social sciences has been measured, analyzed, and reported on an individual level. Unfortunately, ignoring contextual influences associated with individuals' perceptions, actions, and outcomes can lead to biased results and inappropriate interpretations. Subsequently, a growing number of researchers have utilized multilevel modeling techniques to better address the limitations of single-level analyses. For example, Russo, Roccato, and Vieno (2011) conducted a study predicting perceived risk of crime. They found that individual factors, such as age, gender, and perceptions of disorder, were related to risk of crime, but county-level factors, such as collective perceptions of disorder, unemployment rates, and actual crime rates, also predicted individuals' perceived risk of crime.

Because individual attitudes and behaviors are shaped by both personal attributes and shared environment, one of the chief concerns of ignoring contextual influences is the potential to commit one or more inferential fallacies (Diez-Roux, 1998). Atomistic fallacies occur when researchers utilize individual-level data to make inferences at the group level. For example, a researcher could determine that IQ is the strongest predictor of academic achievement (both measured at the individual level) and consequently conclude that improving educational environments was unnecessary (group level). However, this analysis does not consider students circumstances; it could be that individuals found to have lower IQ were funneled into remedial classrooms with substandard teachers and resources, while those with high IQ were funneled into enriching classes with the best teachers (Gibbons, 2008). Similarly, ecological fallacies occur when researchers collect data at the group level but interpret them at the individual level. For example, a researcher may determine that there is no association between average household income and mortality at the county level (i.e., high-income counties and low-income counties, as aggregates, have comparable mortality rates), and consequently conclude that low-income individuals do not face any additional health challenges, compared to high-income individuals. Unfortunately, it is also plausible that within each county, individuals with the lowest incomes are at the higher risk of mortality when they live in higher income counties due to increased discrimination, but, on the aggregate, no associations are found. In the case of psychologistic and sociologistic fallacies, the researcher measures and analyzes data on the appropriate level but fails to take into account the impact that other levels of information have on the associations of interest. Here, if a researcher were interested in the association between student study habits and academic success (individual level), but the researcher did not take into account the school and home environment of the students (group level), the researcher would be at risk of committing a psychologistic fallacy (i.e., not taking into account contextual variables). Finally, if a researcher were interested in evaluating school policies that mandated the use of a curriculum (group level) but did not account for differences in the implementation of the curriculum by the teachers (individual level), he or she would be committing a sociologistic fallacy (i.e., not taking into account individual-level variables). Committing these fallacies can result in misidentifying the source of an influence, including whether it is at the individual or group level, which can lead to misidentification of problems and/or solutions.

Multilevel modeling helps limit the chance of committing each of these fallacies. Specifically, multilevel modeling can limit the chance of committing a fallacy by (a) estimating standard errors that account for the clustering of individuals within a higher order grouping, (b) evaluating the influence of contextual variables impacting individual-level variables, (c) evaluating the influence of individual-level variables influencing contextual variables, and (d) assessing cross-level interactions between associations at each level. Multilevel modeling is often necessary when data are nested within a higher grouping, regardless of whether or not the grouping variable is of any interest. For example, a researcher might be interested in evaluating individual risk factors for dangerous drinking within a sample of college students that belong to fraternities. In this case, the researcher will need to account for the clustering of students within fraternities even if he or she is not interested in any fraternity-level variables because the members of the same fraternity are not independent of one another. Not accounting for the clustering through the use of multilevel models (e.g., a random intercept models) often leads to biased estimates due to unaccounted dependency among individuals in the same cluster. The dependency occurs whenever individuals within the same cluster present more similarly than individuals in a different cluster, a violation of independence of observation (Kenny & Judd, 1986). The best and easiest way to determine whether multilevel modeling is necessary is to consider the intraclass correlation (ICC).

Determining the ICC of variables in one's analysis is a critical step in determining whether multilevel analyses are appropriate. The ICC is computed by examining the amount of variance that exists at the individual and at the group level. The ICC is computed by dividing the group-level variance (τ^2) by the total of the individual-level variance and group-level variance ($\tau^2 + \sigma^2$). This will result in a number ranging from 0 to 1 (ICC = ρ). This represents the percent of variance that occurs at the group level. For example, if a researcher determines that the ICC = .20 on a measure of student exam grades by classroom, we can assume that 20% of the variability in student grades was associated with the classroom a student was in and 80% of their grades was due to individual-level factors.

Assuming that researchers have accurately conceptualized the level at which their research questions lie and how they want to test their model, we can consider the other three purposes of multilevel modeling: (a) determining the influence of contextual variables impacting individual-level variables, (b) determining the influence of individual-level variables impacting contextual variables and (c) assessing cross-level interactions between associations at each level. Researchers are often interested in determining the amount of variance that occurs at each level and how each level uniquely predicts a dependent variable. For example, in a study by Vieno, Perkins, Smith, and Santinello (2005), school sense of community was regressed on students' perception of their school's democratic climate at the individual, classroom, and school level. They found that individual perceptions of the school climate was the strongest predictor of school sense of community, but that aggregates (averages) of student perceptions at the classroom and school levels were also positively related to sense of community. Also of note, they found that a school-level aggregate of socioeconomic status (SES) was a strong predictor of sense of community, but individual-level SES was not. This suggests that disadvantaged adolescents, concentrated in the same schools, likely experience climate-level factors that inhibit the development of a strong sense of community, but, within the same classroom, low-SES students do not perceive the sense of community any differently than high-SES students.

As in the earlier example, multilevel methods enable the researcher to model the same variable at more than one level. A common example of this is income. This is done by measuring the variable at the individual level (e.g., personal income) and aggregating these individual scores within a grouping variable, such as neighborhood, to create a group-level variable (e.g., average neighborhood income). Here, if we include both the individual-level variable (personal income) and the group-level variable (average neighborhood income) in the analysis, we can determine the level of association between income at both levels and an outcome (e.g., health). This type of analysis is sometimes referred to as compositional (individual-level factor) and contextual (group-level factor) effects analysis (Macintyre, Ellaway, & Cummins, 2002).

The final major purpose of conducting multilevel analyses is to determine whether associations between individual-level variables depend upon group-level predictors. This is a form of moderation across multiple levels of analysis commonly referred to as a *cross-level interaction*. For example, we may find that, although both individual income and neighborhood income predict health, low-income individuals have better health outcomes when they live in predominantly low-income neighborhoods and worse health outcomes when they live in high-income neighborhoods. In fact, associations such as these have previously been identified when strong cultural ties are more readily available in the lower income neighborhoods (Roosa et al., 2009). However, it is also possible that lower income individuals do worse in lower income neighborhoods due to a form of "double jeopardy," especially when there are not distinct cultural advantages to living in a lower income neighborhood compared to a higher income neighborhood (Barile, 2010; Williams, 1999). In each of these cases, the relationship between an individual-level predictor (individual income) and outcome (health) is moderated by a group-level variable (neighborhood income). Cross-level interactions are particularly useful when researchers are interested in determining whether an association is context dependent or examining issues associated with person-environment fit.

STRUCTURAL EQUATION MODELING

As stated previously, MSEM is a combination of traditional multilevel regression modeling and structural equation modeling (SEM). SEM is an analytic technique that enables researchers to estimate and model the relationships between latent variables. Latent variables (also described as latent factors, derived from confirmatory factor analysis) represent a construct of interest that cannot be directly observed. Instead, they are estimated by a set of manifest variables. Manifest variables, sometimes referred to as observed variables, are variables that are directly measured by the researcher. For example, we cannot directly measure depression, but we can measure symptoms of depression. With SEM, we would model depression (latent variable) as estimated by items from an inventory of symptoms (observed variables).

Structural equation modeling has a number of advantages over traditional ordinal least squares (OLS) regression. Unlike simple sum scores or the averaging of items from a scale, estimation of latent variables using SEM techniques takes into account measurement error associated with each item; only the common variance found between the indicators is used to define the construct (Anderson & Gerbing, 1988). Additionally, unlike the simple independent/dependent variable dichotomy found in OLS regression, SEM permits the researcher to test complicated models that include multiple mediators within a single model. Also, SEM provides greater flexibility in the types of indicators that can be used in the model (e.g., dichotomous, ordinal, categorical, count) and offers more advanced means for addressing missing data (e.g., full-information maximum likelihood, multiple imputation). (For additional information of missing data, see Graham, 2009.)

Within the SEM framework, researchers also have access to exact and approximate fit indices. Fit indices allow the researcher to compare statistically how well his or her model fares compared to an unconstrained model and/or alternative theoretical models. For most SEM models, exact model fit is assessed by a chi-square statistic (χ^2). The chi-square is a goodness-of-fit statistic that tests the magnitude of discrepancy between the sample covariance matrix and the estimated covariance matrix (Hu & Bentler, 1999). Approximate fit indexes (e.g., Comparative Fit index [CF]), Root Mean Squared Error of Approximation [RMSEA]) provide the researcher with additional indications of whether his or her model fits the data along a continuum (Hu & Bentler, 1999). For example, Hu and Bentler recommend that CFI values above .95 and an RMSEA below .06 correspond to a well-fitted model, although it should be noted that both the chi-square and approximate fit indices are sensitive to sample size (Browne, MacCallum, Kim, Andersen, & Glaser, 2002; Hu & Bentler, 1999). Further discussion of fit statistics are outside the scope of this chapter, but the simple availability of statistics to compare competing models directly is a distinct advantage of SEM over traditional OLS regression (see Hu & Bentler, 1999, and Vernon & Eysenck, 2007, for an in-depth examination of fit indices).

Finally, SEM also allows the researcher to detect measurement invariance, which OLS regression techniques do not. Measurement invariance addresses the extent to which individuals from different backgrounds interpret and report on survey questions in similar manners (Gregorich, 2006). Measurement invariance testing is particularly important when utilizing community-based data where individuals within a sample may come from a range of cultures and/or backgrounds. Measurement invariance is established by examining whether the strength of the association between an indicator (the observed variable) and its latent variable are similar across different populations, genders, cultures, etc. (See Gregorich, 2006, for an in-depth discussion of measurement invariance.)

MULTILEVEL STRUCTURAL EQUATION MODELING

Recent developments in statistical methodology have led to the growing use of MSEM, which combines the advantages of traditional multilevel regression modeling and SEM. MSEM has been found to be a particularly useful technique in studying such topics as social climates in schools (Barile, Donohue, et al., 2012; Marsh et al., 2012), community collaboratives (Barile, Darnell, Erickson, & Weaver, 2012; Brown, Hawkins, Arthur, Abbott, & Van Horn, 2008), and even factories (Brondino, Pasini, & da Silva, 2013). MSEM provides the measurement advantages associated with SEM and the research design advantages associated with traditional multilevel modeling. With this combination, MSEM aids researchers by (a) limiting the susceptibility for measurement bias common in multilevel regression models, (b) limiting the potential for fallacies often common in single-level SEM, and (c) accounting for the unreliability of individual-level reports of group-level constructs (Marsh et al., 2009; Mehta & Neale, 2005).

Multilevel data can include *compilation, composition*, or *fuzzy composition* variables (Dyer, Hanges, & Hall, 2005). Compilation variables, similar to individual-level formative variables (e.g., gender, income, age), are constructs that occur only at the group level and/or do not have any corresponding individual variables. These include such variables as policies, diversity indexes, and crime rates. Composition variables represent similar constructs at both the individual and group levels and are often measured by surveying individuals within groups. The group-level score is computed as an aggregate of individual scores. For example, if one were to ask students in a class how much they liked their teacher, the aggregate of all the student responses might be used to estimate an average score for the classroom. Unfortunately, it is difficult to know whether factor structures at the individual level correspond to the same factor structures at the group level. Fuzzy composition variables are variables that mean different things at different levels. For example, if individuals are asked a set of questions regarding dangers in their neighborhood, a researcher may get very different responses depending on whether they ask residents if they fear crime versus if residents in their neighborhood fear crime. Although it may be appropriate for researchers to ask about individuals' fear of crime, if modeled at the neighborhood level the resulting factors may mean something quite different than they do at the individual level. Consequently, a researcher may also find that constructs that are relatively independent at the individual level (e.g., fear of assault, fear of vandalism), fall under a single factor at the neighborhood level (e.g., climate of fear). One advantage of MSEM over traditional multilevel modeling is the ability for researchers to identify differences in factor structure across levels.

MSEM is a particularly useful analytic approach when survey data are collected from multiple respondents nested within multiple settings, a common scenario in the evaluation of such groups as students in classrooms or members of community collaboratives. In these cases, it would be unwise to simply aggregate these responses. Nowell (2009) noted that, although obtaining multiple reports from individuals within groups can be desirable, if these reports are simply aggregated, the researcher often loses critical individual perspectives. For example, if women are found to report a greater fear of assault (compared to men) but a similar fear of vandalism, important differences in perspectives would be unaccounted for. MSEM provides the unique opportunity to identify differences in perspectives at the individual level while also capitalizing on the shared knowledge of the group at higher levels. Marsh et al. (2009) dubbed models in which individual-level items (such as items on survey) are used to create latent variables at the individual (traditional SEM) and group levels (MSEM) *doubly latent models*.

Doubly latent models enable the researcher to address both sampling bias and measurement error (Marsh et al., 2009, 2012). Sampling bias can occur when the number or likelihood that certain individuals were sampled within one group differs from those sampled in another group. Sampling bias is a common occurrence in nonrandom samples. For example, if a researcher is interested in surveying members regarding their organization's leadership, it is likely that not all of the sampled members will have attended the same number of meetings or had similar roles within the organization, both of which could impact how they rate the leadership. Furthermore, it is unlikely that the researcher will be able to sample the same number of members within each organization. Using traditional regression techniques, the researcher would simply aggregate the responses of members within each group, and there would be no way to adequately address differences in who responded to the survey or how many people responded to the survey within each organization. Creating latent variables at the individual and group levels based on individual responses allows the research to correct for sampling bias due to nonrandom sampling and measurement bias associated with imperfect measurement of latent constructs.

Along with addressing sampling bias and measurement error, MSEM enables the researcher to test for measurement invariance of items at the individual and group levels. For example, if individuals within a group have different backgrounds or roles, they may interpret questions differently. An organization board member may interpret a question such as "Does the organization have a clear mission?" in a different way than a new volunteer does. This difference in interpretation can result in a latent variable (e.g., organizational vision) that has slightly different meaning for each individual and potentially across organizations. Like SEM, MSEM allows the researcher to test for these potential differences through a comparison of the strength of factor loadings for each construct and across organizations (see Jak, Oort, & Dolan, 2014, for more information on testing for measurement invariance in MSEM).

Practical Issues

Sample Size

In order to conduct a test of a multilevel model, the researcher must have an adequate sample size at both the individual and the group levels. The number of individuals and groups needed to obtain unbiased estimates and to ensure sufficient power for the analysis depends on the primary level of inquiry (individual, group, or cross-level interaction), the size of the ICCs, and expected effect sizes. Maas and Hox (2005) reported that using data with fewer than 50 group-level units can lead to biased estimates of the standard error, but this largely depends on the average number of participants within each organization. The bulk of the research on power and multilevel modeling (e.g., Heck & Thomas, 2009) suggests that maximizing the number of nesting units or groups, even at the expense of fewer individuals per group, may led to a greater chance of obtaining adequate power (see Maas & Hox, 2005, for more information on power and multilevel modeling).

Centering

Another important consideration when conducting multilevel modeling is how variables are centered within the model. In multilevel modeling, the researcher must decide whether to center predictor variables around their group mean, the mean of the sample as a whole (i.e., grand mean), or leave them in their raw metric. These decisions critically impact how a researcher should interpret his or her findings (see Enders & Tofighi, 2007, for an overview of this issue).

Model Building

Lastly, although MSEM is a powerful tool for community-based research, multilevel statistical models can quickly become large and computationally demanding. This is particularly true when researchers incorporate multiple factors, noncontinuous indicators and outcomes, or any other advanced analytical functions. It is wise to start with very simple models and slowly build up to more complete, ecological models. If not, researchers may have difficulties getting their models to converge and/or find their modeling programs running for hours (or days) on end (see Preacher, Zyphur, & Zhang, 2010, for more information on model-building techniques using MSEM).

CASE STUDY

The goals of the community-based study presented here were to (a) provide insight into the development of latent factors associated with collaborative vitality at the individual and group levels and (b) report on associations between manifest predictors at the individual and group levels. This case study is based on data obtained by the Georgia Family Connection Partnership (GaFCP). GaFCP is a public/private nonprofit that supports the Family Connection network of collaboratives that are focused on improving child and family well-being across the state of Georgia. Since 2002, these collaboratives have been operating in each of the state's counties. (One collaborative serves three counties.) Additional background on GaFCP can be found on their Web site (http://www.

gafcp.org/) and in previously published papers (Barile, Darnell et al., 2012; Darnell et al., 2013; Emshoff et al., 2007; Harper, Kuperminc, Weaver, Emshoff, & Erickson, 2014).

Members of 152 collaboratives completed a collaborative vitality survey. The survey included questions that assessed five a priori subscales: community (five items), communication (four items), participation (three items), productivity (four items), accountability (three items), and synergy (five items). Survey items queried whether, on a scale from 1 to 7, the member *strongly disagreed* (1) to *strongly agreed* (7) with 24 statements about his or her collaborative. Survey items included statements such as "There is a lack of communication among collaborative members" [communication], and "Collaborative members have a sense of pride in our collaborative's accomplishments" [community]. The survey also inventoried what survey respondents' position in the collaborative was (general member, board member, or staff), how many years they had been involved with the collaborative, and how many meetings (general, committee, and board) they had attended in the past year.

In order to determine whether multilevel modeling methods were appropriate, the ICC for each of the survey items was calculated. Based on 2,521 surveys (average of 16.59 surveys per collaborative), the ICCs for the items ranged from .08 (Item 4) to .24 (Item 6). This indicates that Item 4 (Conflict is freely expressed when it is felt in our collaborative) had the least amount of variability at the collaborative level and Item 6 (Family members are involved in our collaborative) had the most variability at the collaborative level.

To determine the best-fitting factor structure at both the individual and the collaborative level, four different multilevel confirmatory factor structures were estimated and evaluated with respect to their fit to the data. These models were used to determine how many factors the 24 items represented and whether the factor structure of the latent variables was similar at the individual and collaborative levels. Manifest individual- and collaborative-level predictors were included to determine whether there were any associations between collaborative vitality and (a) members' position in the collaborative (general member, board member, or staff); (b) how long members had been involved in the collaborative; and (c) how many general meetings they had attended in the last year. A graphic depiction of Model 2 appears in Figure 16.1.

Model fit indices and overall parsimony were considered to determine whether the six dimensions of collaborative vitality under study were differentially associated with one another, and whether they could be organized under a higher order factor of collaborative vitality. Model 1 specified a single global collaborative vitality factor underlying all 24 survey indicators at both the individual and collaborative levels. Model 2 specified the six a priori factors at both levels of analysis. Model 3 was constructed with restrictions on the latent factor covariances of Model 2 at both levels to test whether the six first-order factors were specified as indicators of a second-order global vitality factor. Model 4 included the higher order factor at the collaborative level only. Survey respondents' position in the collaborative, how many years they had been involved with the collaborative, and how many meetings they had attended in the past year were included as covariates at the individual level, and an aggregate of the number of meetings attended by respondents and the proportion of respondents that identified as a board member and staff were included as covariates at the collaborative level for all models.

Findings from the four models found that all fit the data well based on Hu and Bentler's (1999) criteria. Model 2, which included the six subscales at each level, fit the data best, χ^2 (690) = 1,546.47, $p < .001$, CFI = .98, RMSEA = .03 and Model 4, which included a second-order vitality factor at the collaborative level but not at the individual level, was the next best-fitting model and had the advantage of being more parsimonious, χ^2 (699) = 1,558.97, $p < .001$, CFI = .98, RMSEA = .03. Taken together, the four models suggest that there are likely multiple, semi-independent subscales that may also serve as indicators of a second-order collaborative vitality factor. In particular, these findings suggest that the six subscales form a cohesive second-order vitality factor at the collaborative level but not at the individual level.

Table 16.1 presents the associations between the predictors and the latent factors at both levels. At the individual level, the results do not indicate that staff or board members differ from general members in their responses on any of the six subscales. However, higher scores on a number of subscales were reported by members who were more

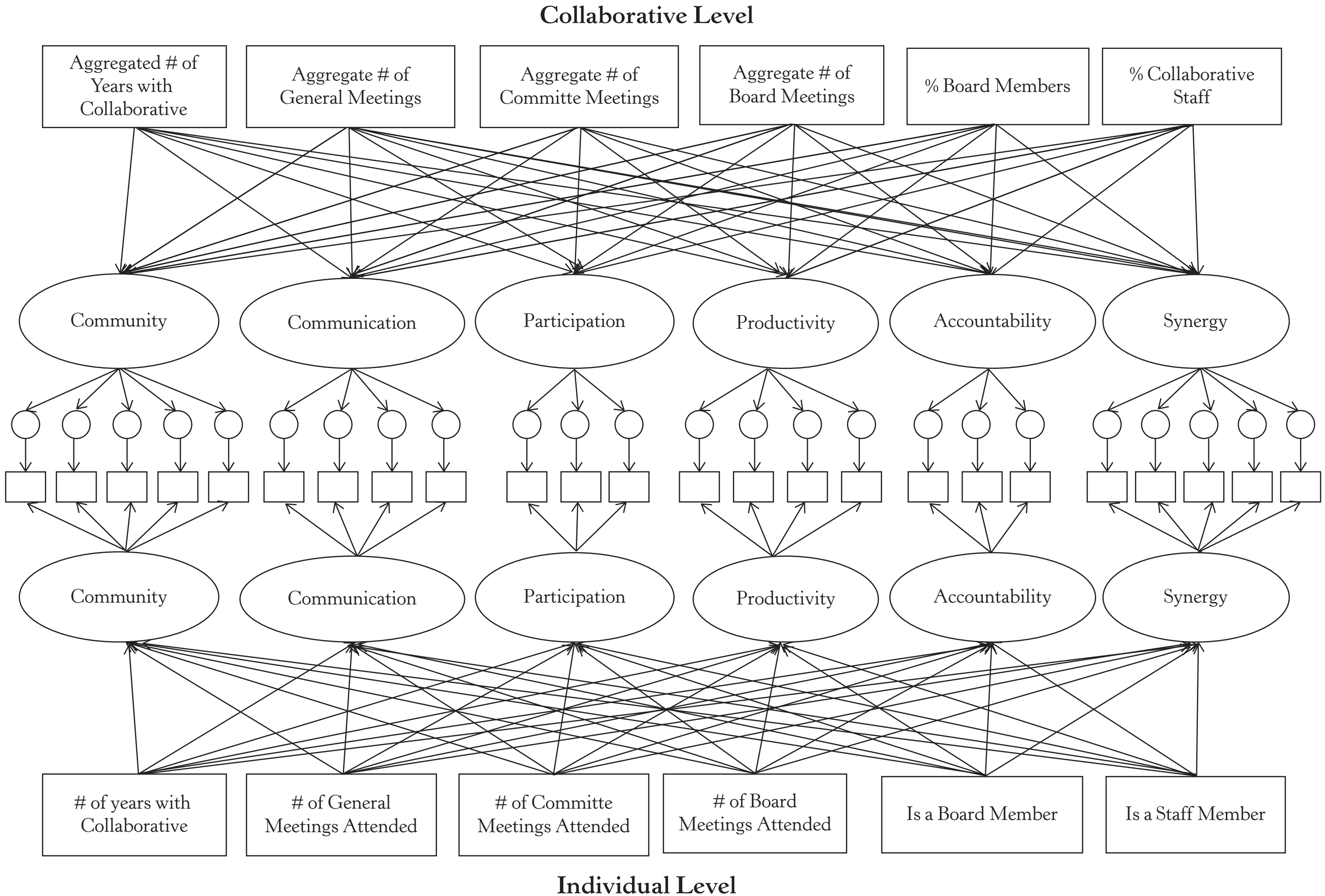

FIGURE 16.1: This represents the six latent factors at the individual and collaborative levels with six covariates at each level (Model 2). Boxes represent items from the collaborative vitality survey. Ovals represent latent variables that are estimated by the observed items.

TABLE 16.1: ASSOCIATIONS BETWEEN INDIVIDUAL AND COLLABORATIVE LEVEL PREDICTORS AND MULTILEVEL LATENT FACTORS

Covariates	Community		Communication		Participation		Productivity		Accountability		Synergy	
	b	*SE*	*b*	*SE*	*b*	*SE*	*b*	*SE*	*b*	*SE*	*b*	*SE*
Individual Level												
No. of years with collaborative	0.01	0.01	0.00	0.01	0.00	0.01	0.01	0.01	0.01	0.01	0.01	0.01
No. of general meetings	0.03**	0.01	0.01	0.01	−0.01	0.01	0.02	0.02	−0.01	0.01	0.00	0.01
No. of community meetings	0.03**	0.01	0.02	0.01	0.01	0.01	0.04*	0.02	0.01	0.01	0.04*	0.01
No. of board meetings	0.00	0.01	0.01	0.01	0.00	0.01	0.02	0.02	0.04*	0.02	0.01	0.01
Staff member 0/1	−0.13	0.12	−0.03	0.13	0.07	0.09	−0.12	0.19	−0.10	0.17	−0.02	0.14
Board member 0/1	0.12	0.08	0.07	0.09	0.07	0.07	0.05	0.15	0.18	0.13	0.08	0.10
Collaborative Level												
Aggregate years with collaborative	0.06*	0.02	0.13**	0.04	0.09**	0.03	0.12*	0.05	0.09*	0.04	0.04	0.03
Aggregate no. of general meetings	0.02	0.02	0.06	0.04	0.03	0.03	0.04	0.05	0.02	0.03	0.03	0.03
Aggregate no. of community meetings	−0.01	0.02	−0.01	0.04	0.02	0.04	0.01	0.05	0.05	0.04	0.03	0.04
Aggregate no. of board meetings	−0.01	0.03	0.03	0.06	0.02	0.05	0.04	0.07	−0.02	0.05	0.06	0.05
% Staff member	0.72*	0.37	1.08	0.62	0.97*	0.57	1.70*	0.85	1.04	0.56	1.47*	0.58
% Board member	−0.37	0.27	−0.89*	0.43	−0.74*	0.40	−1.21*	0.53	−0.52	0.40	−0.62	0.33

engaged in the collaborative through attending general meetings (community), committee meetings (community, productivity, synergy), and board meetings (accountability). At the collaborative level, the higher the average number of years members had been with the collaborative, the greater their scores on the community, communication, participation, productivity, and accountability subscales. This suggests that if collaboratives are able to keep the same members engaged over time, they will likely have greater collaborative vitality.

Additionally, the results suggest that the percentage of staff members responding (compared to general members and board members) was positively associated with higher scores on community, participation, productivity, and synergy. This is particularly interesting because at the individual level no differences between staff and general members were observed. It is possible that having a higher proportion of staff responding to the survey is indicative of collaboratives with greater funding (and subsequently more staff). This finding would support previous research on GaFCP that found positive associations between improved collaborative functioning (through systems change) and leveraged dollars. Finally, and interestingly, collaboratives with a higher percentage of board members reported lower scores on communication, participation, and productivity. This may suggest that collaboratives that are represented by fewer members (resulting in a higher percentage of board members reporting) are less able to develop strong indicators of collaborative vitality.

This example provides an illustration of how MSEM can be used with a simple community-based survey. In this example, multiple-factor structures were tested and compared. This process provides some indication of whether the factor structure at the collaborative level is similar to that on the individual level, providing some evidence of whether the survey items represent composition or fuzzy-composition factors. This example also tested associations with formative individual-level factors (number of meetings attended, years with the collaborative, role in the collaborative) and collaborative compilation variables (e.g., percent of staff members responding). This model could be expanded to include other county-level predictors of vitality (e.g., census data, measures of social capital), as well as other desired outcomes (e.g., child and family well-being). Additional invariance testing could also be undertaken to determine whether the factors identified on each level are consistent across all respondents and collaboratives.

CONCLUSION

MSEM techniques enable community-based researchers to disentangle individual and context-dependent variables using a robust methodology resistant to measurement and sampling error in ways that single-level regression, multilevel regression, and SEM do not. Furthermore, MSEM helps researchers avoid committing fallacies that can lead to inappropriate interventions and policies. MSEM techniques allow the researcher to incorporate individual- and group-level survey data, as well as archival data (e.g., census data), within the same model without having to choose whether to aggregate or not to aggregate. As such, MSEM is a flexible technique that allows the researcher to incorporate multiple predictors, mediators, moderators, and outcomes within a single model and can work in combination with other advanced techniques (e.g., latent class growth modeling). Community-based researchers interested in understanding the diverse web of ecological determinants of individual and community well-being should seriously consider this powerful tool when faced with nested data.

REFERENCES

Anderson, J. C., & Gerbing, D. W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. *Psychological Bulletin, 103*, 411–423.

Barile, J. P. (2010). *Health disparities in a diverse county: Investigating interactions between residents and neighborhoods*. Unpublished doctoral dissertation, Georgia State University, Atlanta.

Barile, J. P., Darnell, A. J., Erickson, S. W., & Weaver, S. R. (2012). Multilevel measurement of dimensions of collaborative functioning in a network of collaboratives that promote child and family well-being. *American Journal of Community Psychology, 49*, 270–282.

Barile, J. P., Donohue, D. K., Anthony, E. R., Baker, A. M., Weaver, S. R., & Henrich, C. C. (2012). Teacher-student relationship climate and school outcomes: Implications for educational policy initiatives. *Journal of Youth and Adolescence, 41*, 256–267.

Brondino, M., Pasini, M., & da Silva, S. C. A. (2013). Development and validation of an Integrated Organizational Safety Climate Questionnaire with multilevel confirmatory factor analysis. *Quality and Quantity, 47*, 2191–2223.

Brown, E. C., Hawkins, J. D., Arthur, M. W., Abbott, R. D., & Van Horn, M. L. (2008). Multilevel analysis of a measure of community prevention collaboration. *American Journal of Community Psychology*, *41*, 115–126.

Browne, M. W., MacCallum, R. C., Kim, C. T., Andersen, B. L., & Glaser, R. (2002). When fit indices and residuals are incompatible. *Psychological Methods*, *7*, 403–421.

Darnell, A. J., Barile, J. P., Weaver, S. R., Harper, C. R., Kuperminc, G. P., & Emshoff, J. G. (2013). Testing effects of community collaboration on rates of low infant birthweight at the county level. *American Journal of Community Psychology*, *51*, 398–406.

Diez-Roux, A. V. (1998). Bringing context back into epidemiology: Variables and fallacies in multilevel analysis. *American Journal of Public Health*, *88*, 216–222.

Dyer, N., Hanges, P., & Hall, R. (2005). Applying multilevel confirmatory factor analysis techniques to the study of leadership. *Leadership Quarterly*, *16*, 149–167.

Emshoff, J. G., Darnell, A. J., Darnell, D. A., Erickson, S. W., Schneider, S., & Hudgins, R. (2007). Systems change as an outcome and a process in the work of community collaboratives for health. *American Journal of Community Psychology*, *39*, 255–267.

Enders, C. K., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychological Methods*, *12*, 121–138.

Gibbons, L. M. (2008). Nature+ nurture> 100%: Genetic and environmental influences on child obesity. *American Journal of Clinical Nutrition*, *87*, 1968–1968.

Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology*, *60*, 549–576.

Gregorich, S. E. (2006). Do self-report instruments allow meaningful comparisons across diverse population groups? Testing measurement invariance using the confirmatory factor analysis framework. *Medical Care*, *44*, S78–94.

Harper, C. R., Kuperminc, G. P., Weaver, S. R., Emshoff, J., & Erickson, S. (2014). Leveraged resources and systems changes in community collaboration. *American Journal of Community Psychology*, *54*, 348–357.

Heck, R. H., & Thomas, S. L. (2009). *An introduction to multilevel modeling techniques* (2nd ed.). New York, NY: Routledge.

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

Jak, S., Oort, F. J., & Dolan, C. V. (2014). Measurement bias in multilevel data. *Structural Equation Modeling*, *21*, 31–39.

Jason, L. A., & Glenwick, D. S. (Eds.). (2012). *Methodological approaches to community-based research*. Washington, DC: American Psychological Association.

Kenny, D. A., & Judd, C. M. (1986). Consequences of violating the independence assumption in analysis of variance. *Psychological Bulletin*, *99*, 422–431.

Lewin, K. (1935). *A dynamic theory of personality*. New York, NY: McGraw-Hill.

Luke, D. A. (2005). Getting the big picture in community science: Methods that capture context. *American Journal of Community Psychology*, *35*, 185–200.

Maas, C. J., & Hox, J. J. (2005). Sufficient sample sizes for multilevel modeling. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, *1*, 86–92.

Macintyre, S., Ellaway, A., & Cummins, S. (2002). Place effects on health: How can we conceptualise, operationalise and measure them? *Social Science Medicine*, *55*, 125–139.

Marsh, H. W., Lüdtke, O., Nagengast, B., Trautwein, U., Morin, A. J. S., Abduljabbar, A. S., & Köller, O. (2012). Classroom climate and contextual effects: Conceptual and methodological issues in the evaluation of group-level effects. *Educational Psychologist*, *47*, 106–124.

Marsh, H. W., Lüdtke, O., Robitzsch, A., Trautwein, U., Asparouhov, T., Muthén, B., & Nagengast, B. (2009). Doubly-latent models of school contextual effects: Integrating multilevel and structural equation approaches to control measurement and aampling error. *Multivariate Behavioral Research*, *44*, 764–802.

Mehta, P., & Neale, M. (2005). People are variables too: Multilevel structural equations modeling. *Psychological Methods*, *10*, 259–284.

Nowell, B. (2009). Profiling capacity for coordination and systems change: The relative contribution of stakeholder relationships in interorganizational collaboratives. *American Journal of Community Psychology*, *44*, 196–212.

Preacher, K. J., Zyphur, M. J., & Zhang, Z. (2010). A general multilevel SEM framework for assessing multilevel mediation. *Psychological Methods*, *15*, 209–233.

Roosa, M. W., Weaver, S. R., White, R. M., Tein, J. Y., Knight, G. P., Gonzales, N., & Saenz, D. (2009). Family and neighborhood fit or misfit and the adaptation of Mexican Americans. *American Journal of Community Psychology*, *44*, 15–27.

Russo, S., Roccato, M., & Vieno, A. (2011). Predicting perceived risk of crime: A multilevel study. *American Journal of Community Psychology*, *48*, 384–394.

Shinn, M., & Rapkin, B. D. (2000). Cross-level research without cross-ups in community psychology. In J. Rappaport & E. Seidman (Eds.), *Handbook of* *community psychology* (pp. 669–693). New York, NY: Kluwer Academic.

Vernon, T., & Eysenck, S. (2007). Structural equation modeling [Special issue]. *Personality and Individual Differences*, *42*, 811–898.

Vieno, A., Perkins, D. D., Smith, T. M., & Santinello, M. (2005). Democratic school climate and sense of community in school: a multilevel analysis. *American Journal of Community Psychology*, *36*, 327–341.

Williams, D. (1999). Race, socioeconomic status, and health the added effects of racism and discrimination. *Annals of the New York Academy of Sciences*, *896*, 173–188.