Available online at www.sciencedirect.com

SCIENCE DIRECT®

The Leadership Quarterly 16 (2005) 149–167

The Leadership Quarterly

# Applying multilevel confirmatory factor analysis techniques to the study of leadership

Naomi G. Dyer[a,*], Paul J. Hanges[a], Rosalie J. Hall[b]

[a]Department of Psychology, University of Maryland, College Park, MD 20742, United States
[b]Department of Psychology, University of Akron, United States

**Abstract**

Statistical issues associated with multilevel data are becoming increasingly important to organizational researchers. This paper concentrates on the issue of assessing the factor structure of a construct at aggregate levels of analysis. Specifically, we describe a recently developed procedure for performing multilevel confirmatory factor analysis (MCFA) [Muthen, B.O. (1990). Mean and covariance structure analysis of hierarchical data. Paper presented at the Psychometric Society, Princeton, NJ; Muthen, B.O. (1994). Multilevel covariance structure analysis. *Sociological Methods and Research, 22*, 376–398], and provide an illustrative example of its application to leadership data reflecting both the organizational and societal level of analysis. Overall, the results of our illustrative analysis support the existence of a valid societal-level leadership construct, and show the potential of this multilevel confirmatory factor analysis procedure for leadership research and the field of I/O psychology in general.


*Keywords:* Multilevel confirmatory factor analysis; Leadership; GLOBE project

The inherent hierarchical structure of most organizations has motivated researchers to increasingly pay attention to statistical issues associated with the analysis of multilevel data (cf. Klein, Dansereau, & Hall, 1994; Klein, Tosi, & Cannella, 1999). Research on leadership is no exception. For example, in a typical leadership study, surveys are administered to multiple subordinates belonging to multiple

* Corresponding author.
*E-mail address:* ndyer@psyc.umd.edu (N.G. Dyer).

1048-9843/$ - see front matter 
doi:10.1016/j.leaqua.2004.09.009

workgroups with each subordinate completing the survey to describe their workgroup leader. The data from this typical research design are said to be hierarchically structured—the responses are collected from persons who share membership in a common group (i.e., the workgroup). Thus, leadership research is usually collected from persons nested within a variety of levels, such as dyads, workgroups, departments, organizations, or even different social cultures.

Given the hierarchical structure or multilevel nature associated with typical leadership data, it is not surprising that concerns regarding multilevel issues have a long tradition in this area (e.g., Dansereau & Yammarino, 1998a, 1998b; Dansereau, Yammarino, & Markham, 1995; Den Hartog, House, Hanges, Ruiz-Quintanilla, & Dorfman, 1999; Hall & Lord, 1995, 1998; Hanges & Dickson, 2004; Hanges, Dickson, & Sipe, 2004; House, Hanges, Javidan, Dorfman, & Gupta, 2004; House et al., 1999; Yammarino & Bass, 1991). Leadership researchers have questioned the extent to which relationships among constructs would vary at different levels of analysis as well as the extent to which constructs would have different meanings or factor structures at different levels of analysis. It is this latter concern that is the focus of this paper. More specifically, the purpose of this paper is to discuss the impact of hierarchically structured data on the factor structure of a set of items/variables designed to measure a unidimensional construct at some aggregate (i.e., workgroup leader, department leader, organizational leader, etc.) level of analysis. Specifically, we discuss and illustrate Muthen's multilevel confirmatory factor analysis (MCFA) procedure using data from the Global Leadership and Organizational Behavior Effectiveness (GLOBE) Project (House et al., 2004).

## 1. Factor analysis and problems caused by hierarchically structured data

The term factor analysis refers to a set of statistical techniques that are used to either explore or confirm the underlying structure among a set of items/variables to determine those items/variables that tap a factor or latent construct. With exploratory techniques, the objective is to identify the underlying structure, while confirmatory factor analysis seeks to validate some a priori hypothesized structure among the items/variables (Nunnally & Bernstein, 1994).

Given the utility of factor analysis in the identification and validation of unidimensional scales, it is not surprising that leadership researchers rely on this statistical methodology for developing their scales.

Unfortunately, the hierarchical structure often inherent in the data of leadership researchers (e.g., multiple respondents describing the same leader) can result in incorrect conclusion about the factor structure of these scales and hence the constructs and the interrelationships among them. There are at least three reasons for this statement.

First, when observations are correlated rather than independent—as is likely when either survey responses for each leader come from multiple subordinates or when leaders are repeatedly surveyed over time (e.g., Ployhart, Holtz, & Bliese, 2002)—the fundamental independence assumption underlying many commonly used statistical techniques, including factor analysis, is violated. When the degree of non-independence is substantial, violation of this assumption biases parameter estimates and standard errors, as well as affects the power of statistical significance tests (e.g., Bliese & Hanges, 2004; Kenny & Judd, 1986). In other words, even when the items comprising a scale are truly unidimensional, the items will still be correlated after removing the influence of the single latent variable because the hierarchical structure of the data violated the independence assumption. This can lead the researcher to incorrectly conclude that his/her leadership scale is not unidimensional, when it truly is. Alternatively, it may lead the researcher to conclude that his/her items are not good at measuring the latent variable because the hierarchical structure can bias parameter estimates such as factor loadings.

For example, assume that a researcher asks soldiers to rate the leadership behavior of their officers. Since the scale was designed to assess differences in leadership style, individual soldier differences in how they describe the same leader is error or noise. The only meaningful source of variation for this construct is attributable to differences between actual leaders. If the researcher ran the factor analysis on the full data matrix, s/he would have a very large sample size but also very misleading results because the factor structure would appear very weak due to the substantial amount of noise (i.e., individual level variation) in the data. Even the typical approach of averaging the soldiers' responses to the leader level of analysis and then running the factor analysis on this aggregated data, does not solve this problem since Muthen (1990, 1994) has demonstrated that the aggregated correlation matrix is actually a combination of the individual and group influences. It has been known in the statistical community since the mid-1970s that a factor analysis of means can produce misleading results (Cronbach, 1976; Harnqvist, 1978).

Second, in organizational research, there is an increasing interest in building aggregate-level and multilevel theories. These theories usually include constructs that operate uniquely at aggregate levels of analysis (Bliese, 2000; Chan, 1998; Klein et al., 1994; Klein & Kowzlowski, 2000; Klein et al., 1999; Morgeson & Hofmann, 1999). More specifically, multilevel researchers have differentiated between compilation, composition, and fuzzy composition group level constructs (Bliese, 2000; Kozlowski & Klein, 2000). Compilation variables are constructs that only occur at the group level. They have no conceptual meaning at lower levels of analysis. For example, "gender diversity" is a construct that is meaningful only at the group level of analysis. It has no meaning at an individual level.

Composition variables, on the other hand, are constructs that emerge from responses of individuals within groups. Even though these constructs are a function of the cognition, affect, and personality of individuals, the psychometric properties emerge only at the group level of analysis. For example, the "culturally endorsed leadership theory" (CLT) constructs measured in the GLOBE Project are examples of composition variables (Hanges & Dickson, 2004). While individuals have their own beliefs about the attributes of leaders, the beliefs that are shared only occur or are identifiable at the group level of analysis (Dorfman, Hanges, & Brodbeck, 2004). Composition variables, however, are believed to operate almost identically or isomorphically at the group level and lower levels of analysis (Bliese, 2000).

Finally and in contrast, fuzzy composition variables are partially isomorphic constructs in that while they operate and are meaningful at multiple levels of analysis, their factor structure can differ across the levels. For instance, Avolio and Bass (1998) propose varying definitions for individualized consideration at the individual, team, and organizational level of analysis. Similarly, Hall and Lord (1998) suggest that both affective and cognitive reactions to a leader may operate differentially at the individual, dyad, and group level of analysis.

Since the nature of the construct can differ across levels of analysis, it is critical to ensure that a scale measuring a group level construct exhibits the desired dimensionality properties as well as convergent/discriminant validity at the aggregate level of analysis.

Finally, the third reason for conducting analyses that reflect the hierarchical structure of the data focuses around construct validity issues. The literature discussing how to address the *construct validity* of aggregate measures is very sparse. Most typically, either theoretical arguments are referenced to verify the construct, or an argument is made that the construct is valid because an adequate level of agreement in responses exists within the group. Few studies have used empirical techniques such as factor analysis to explore the validity of aggregate constructs in a manner that explicitly acknowledges the aggregate nature of the measure, while allowing for a simultaneous assessment of measurement qualities (e.g., factor loadings, factor intercorrelations) at both the aggregate and disaggregate levels of analysis.

Concern regarding this gap was expressed by Chan (1998), who worried that "...despite the existence of broad theoretical frameworks and methodological advances, the fundamental substantive issue of construct validation in multilevel research has not been addressed adequately" (p. 234). Importantly, the lack of empirically based studies examining the construct validity of aggregate measures means that we often do not know whether a given construct is structurally isomorphic (i.e., having an identical structure) across different levels of analysis, or whether its structure is fuzzy or varies across levels. This is a critical shortcoming because some statistical analyses assume isomorphism. For example, Mumford (1998) argues that WABA requires that the aggregate and disaggregate measures "...have similar reliability across levels and the same pattern of item loadings on the constructs being measured. When these conditions are not met, any effects observed in a WABA analysis may reflect nothing more than methodological artifacts" (p. 425). Unfortunately, good options for assessing the factor loadings of a scale at an aggregate level, much less options for assessing the similarity of the factor loading patterns across levels of analysis, have not been available until recently.

### 1.1. Focus of current paper

Recent works, especially those by Muthen (1990, 1994) and by Peter Bentler and colleagues (Chou, Bentler, & Pentz, 2000; Liang & Bentler, in press) provide promising approaches to performing multilevel confirmatory factor analyses. Unfortunately, this work is not widely known in organizational research (for exceptions, see Hall, Hanges, & Dyer, in press; Hall, Makiney, Marchioro, Tan, & Klein, 1999; Hanges & Dickson, 2004), and is complex enough that good illustrative examples are likely to be of help to many researchers considering its application. Thus, an important purpose of this paper is to describe steps that one might go through to assess the factor structure of constructs intended to reflect group-level phenomena even though the data was collected from lower level units. For simplicity, the remainder of this paper focuses on verifying constructs at the aggregate level of analysis. However, this technique may also be used to account for group-level influences when verifying individual-level constructs collected hierarchically.

In this paper, we first review a protocol for performing a MCFA suggested by Muthen (1994). Second, we present an empirical demonstration of MCFA with leadership data intended to describe a unidimensional construct at both the individual and societal level of analysis. Finally, the potential of this technique for enriching leadership research is discussed.

## 2. Description of multilevel confirmatory factor analysis procedures

The typical procedure that researchers use when assessing the factor structure of a group level measure is to either (a) conduct the factor analysis on the total covariance matrix derived from the entire data set (i.e., ignoring the hierarchical structure of the data) or to (b) average the item responses to the group level and then perform a factor analysis on the sample between-group covariance matrix, $\mathbf{S}_B$. Regardless of which of these two procedures is conducted, Muthen (1994) has shown that these approaches produce problematic results. In particular, when the total covariance matrix is factor analyzed, the fit of the group level factor structure as well as any factor loading estimates will be biased since it is a mixture of the factor structure operating at the between-group and within-group levels. Typically, this total factor structure will primarily be a function of the within-group factor structure.

The second procedure is also problematic since, while not widely known, the sample $\mathbf{S}_B$ matrix is a biased estimate of the population between-group covariance matrix, $\mathbf{\Sigma}_B$. Muthen (1994) showed that the sample between-group matrix is a function of not only the population between-group covariance matrix ($\mathbf{\Sigma}_B$) but also a function of the group-size-weighted, *within-group* covariance matrix ($\mathbf{\Sigma}_W$) (Muthen, 1994). While the degree of bias in this second analysis is not as dramatic as in the total covariance analysis, the analysis can still be substantially misleading. When the within-group covariance matrix ($\mathbf{\Sigma}_W$) is a function of error or noise variance, as is assumed in aggregate level composition constructs (Kozlowski & Klein, 2000), the subsequent factor analysis will underestimate the fit of the group level factor structure and produce conservatively biased factor loadings.

To counter this problem, Muthen (1994) developed the MCFA procedure. Fig. 1 illustrates a generic single-factor MCFA model with four observed indicators ($y_{w1}$–$y_{w4}$) depicted by squares. These indicators are the observed respondent ratings for the four items in a scale. The lower half of

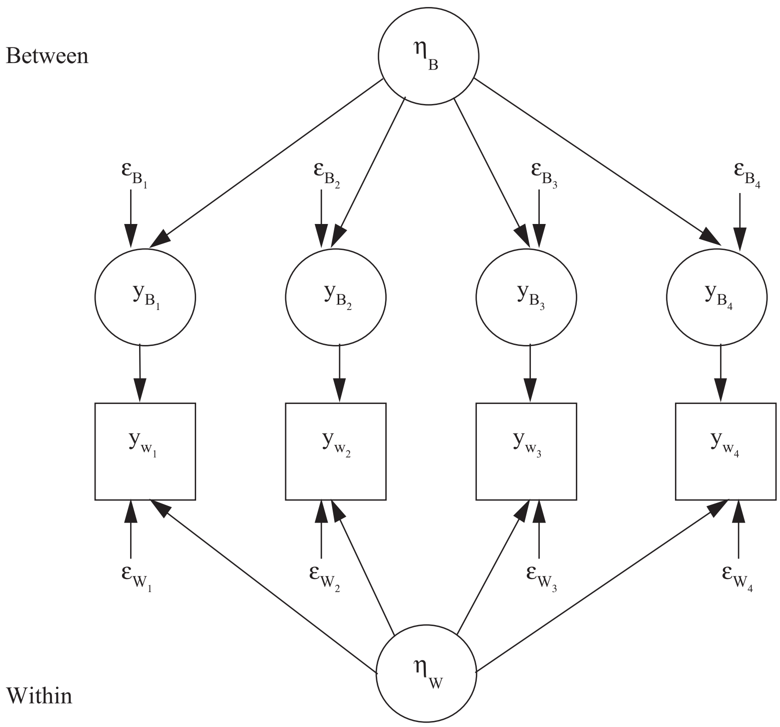


Fig. 1. Path diagram of a sample one-factor multilevel model.

Fig. 1, labeled "within", is consistent with a traditional confirmatory factor analysis on disaggregate data. As shown in this figure, the four observed variables load onto a single latent factor ($\eta_W$) at the "within" level. There are also four random errors ($\varepsilon_{W1}-\varepsilon_{W4}$) associated with each item at this level.

The upper half of Fig. 1, labeled "between", shows four indicators represented by the circled $y_{B1}-y_{B4}$. These are not observed/raw data, but rather represent the group means for each observed indicator ($y_{w1}-y_{w4}$). These group means load onto the aggregate latent variable ($\eta_B$) and are associated with their respective random error terms ($\varepsilon_{B1}-\varepsilon_{B4}$). The full model connects the disaggregate and corresponding aggregate indicators. Thus, the observed values of the original indicators ($y_{W1}-y_{W2}$) are considered to be a function of both the within- and between-level latent constructs ($\eta_W$ and $\eta_B$, respectively). (See Hall et al., in press, for a more detailed discussion.) The MCFA consists of a simultaneous analysis of both of the within- and between-group covariance matrices.

In Fig. 1, the between and within components are explained by a single latent factor, however, this need not be the case. For example, one could test a model that proposes a single factor at the aggregate level and two factors at the disaggregate level, or many other similar non-isomorphic structures. If the hypothesized factor structure proposes more than one factor at a given level, the model may also include covariances among those same-level factors (by definition in this type of model, no covariances are allowed among factors at different levels). Similarly, the model may suggest that some indicators are valid at one level only, indicating a fuzzy composition model. Furthermore, the model may show some important covariates (e.g., age, type of organization) that might be included in the model, relate to the focal latent construct at only one level. Estimation of these models yields both indicators of model fit, and parameter estimates of the factor loadings, factor variances, and uniquenesses (residuals). Thus, although our illustration presents only a very simple case, the MCFA technique in general promises some flexibility in the type of model that can be specified and tested.

Muthen (1994) noted that the MCFA sometimes does not converge and thus, he recommends a five-step procedure for conducting this analysis. The first four steps provide initial information about the factor structure of the scale at different levels of analysis as well as pertinent information used to justify multilevel analyses. The former information is useful for setting starting values for the estimation of the MCFA if it fails to converge. The fifth and final step consists of the actual MCFA, and thus most directly provides evidence for or against the hypothesized multilevel factor model. These steps may be performed with virtually any structural equation modeling software (e.g., LISREL, EQS), though the final step is best performed using either Mplus (Muthen & Muthen, 1998, 2002) or version 6.0 of EQS which have features that specifically address multilevel analysis. In the next section, we briefly review the recommended analytic steps, culminating in the test of the full MCFA model.

## 3. Step 1: perform a conventional confirmatory factor analysis on the sample total covariance matrix ($S_T$)

This step is identical to the first aforementioned procedure that researchers use to test the factor structure of group-level constructs. As indicated above, it is important to remember that the parameter estimates and fit statistics resulting from the Step 1 model may be biased when there is appreciable non-independence, especially when group sizes are large (Muthen, 1994) or when the factor structure at the within-group level of analysis is different from the factor structure at the between-group level. Thus, if in the next step, substantial between-group variation is found in the data, estimating a multilevel model using separate within- and between-group covariance matrices will produce more accurate results than those obtained in using the total covariance matrix as done in this step.

## 4. Step 2: estimate between-group level variation

This step addresses the question: "Is multilevel analysis appropriate for your data?" by estimating the proportion of systematic between-group variation for each observed variable in the model. A variety of analytic approaches have been devised to address such issues when working with multilevel data. For example, James, Demaree, and Wolf's (1984, 1993) $r_{wg}$, the intraclass correlation coefficient (ICC) and, WABA I (Dansereau, Alutto, & Yammarino, 1984) are all techniques designed to deal with detecting the multilevel nature of the data (see Klein & Kowzlowski, 2000 for a thorough discussion of aggregation techniques). Muthen (1994) suggests estimating a unique type of ICC to determine potential group influences. Muthen's ICC assumes different things about the nature of the group level effect than the other procedures (e.g., $r_{wg}$, WABA I, $\eta^2$). That is, WABA and $\eta^2$ assume fixed level effects, whereas Muthen's ICC assumes random level effects. Muthen's MCFA also assumes random level effects; hence, the use of this ICC. This ICC, however, is conceptually similar to $ICC_{(1)}$, just the way of estimating the two are slightly different. That is, Muthen's ICC is calculated from a ratio of the maximum likelihood estimates of the latent within and between variance components, assuming random level effects. In comparison, when estimating $ICC_{(1)}$ with ANOVA in SPSS or a similar package, fixed level effects are assumed (see Hall et al., in press, for more information).

The ICC ranges in value from 0 to 1, with higher values of the ICC indicating greater proportions of between-level variance and thus likely greater bias if the multilevel nature of the data is not taken into account. In practice, multilevel modeling may provide few benefits when ICCs are less than 0.05 (and, in fact, when ICCs are this small, multilevel models may be difficult or impossible to estimate).

## 5. Step 3: perform a factor analysis on the sample pooled-within covariance matrix ($\mathbf{S}_{PW}$)

If aggregation and multilevel analysis appear justified, it is beneficial to separately analyze the within- and between-level submodels of Fig. 1 before trying to simultaneously estimate the full two-level model. Thus, this step and the next one are focused on performing factor analyses of the submodels. The data used for the analysis of the disaggregate factor structure are in the form of the sample within-group covariance matrix, $\mathbf{S}_{PW}$, which is an unbiased and consistent estimator of the *population* within-group covariance matrix. Unlike the $\mathbf{S}_T$ matrix, the values in the $\mathbf{S}_{PW}$ matrix are adjusted to remove between-group differences by subtracting relevant group means from individual scores. The resulting values in the matrix reflect the factor structure at the within-group level only. If the predominance of the construct-relevant variance is at the between-group level, then the model estimated using $\mathbf{S}_{PW}$ may show a worse fit than that using $\mathbf{S}_T$. On the other hand, if the construct-relevant variance is primarily at the within-group level, or the factor structures differ substantially at the two levels, then the model estimated using $\mathbf{S}_{PW}$ may show an improved fit compared to that estimated using $\mathbf{S}_T$.

## 6. Step 4: perform a factor analysis on the sample between-group covariance matrix ($\mathbf{S}_B$)

The factor structure obtained from the within-group level analysis cannot be assumed to also hold at the between-group level of analysis (Cronbach, 1976; Harnqvist, 1978; Muthen, 1994). At this step, the appropriateness of the between-group factor structure is examined. Either an estimate of the population between-group covariance matrix is used, as done in Mplus, or if there are practical problems in the analysis when using this matrix (e.g., lack of convergence), the sample between-group covariance matrix, $\mathbf{S}_B$, may be used. ($\mathbf{S}_B$ is the covariance matrix of observed group means, corrected for the grand mean. The values of $\mathbf{S}_B$ are not only a function of the between-group population covariance matrix, but also reflect a group-size-weighted, *within-group* effect (Muthen, 1994).) In the case that the proposed factor structure at the between level is unsupported, the researcher may want to perform an exploratory factor analysis using either the estimated population between-group correlation matrix or the sample between-group correlation matrix. Results from step 4 provide preliminary information about the appropriate group-level factor structure.

## 7. Step 5: perform multilevel confirmatory factor analysis

Based on the information gathered in the previous steps, the researcher is now prepared to perform the MCFA. As with previous steps, the multilevel confirmatory factor analysis can be conducted using most structural equation modeling (SEM) software such as LISREL (Joreskog & Sorbom, 1989), EQS (Bentler, 1995), or Mplus (Muthen & Muthen, 1998, 2002). However, unless the particular SEM software package is specifically designed for multilevel analysis, the data analyst must specify the MCFA model as a multigroup model and include some non-conventional model specifications in order to correctly perform this analysis (see Muthen, 1994 for additional details). We provide sample code in Mplus for performing all the steps of a MCFA (see Appendices A–D). Due to space constraints, the complexity of performing these analyses with software without MCFA capabilities built in, as well as such capabilities becoming more readily available in software packages (e.g., EQS v.6), we provide only sample Mplus code. In the next section, we demonstrate the application of Muthen's procedure to a leadership measure.

## 8. Illustrative example of multilevel modeling procedure

### *8.1. Description of sample and measure*

For this demonstration, a subset of the data from the GLOBE Research Project was used (House et al., 2004, 1999). The GLOBE Project is a long-term project designed to assess the influence of societal and organizational culture on the content of leadership mental models shared by individuals within an organization or society (House et al., 2004). It was originally designed by Robert House and it currently has over 170 social scientists and management scholars from around the world working as co-investigators.

The GLOBE leadership scales were developed to assess societal level differences in the content of shared leadership mental models. Hanges and Dickson (2004) conducted MCFA of the GLOBE scales to verify that the items of the leadership scales exhibited their intended factor structure at the appropriate level of analysis. The MCFA results presented in this paper consists of survey responses of middle managers ($N$=13,412) from 61 different societies ($G$=61). We report the MCFA results for the 'Procedural' leadership scale. Specifically, respondents were asked to indicate the extent to which being 'formal', 'habitual', 'cautious', 'procedural', and 'ritualistic' are traits that would relate to effective leadership (see House et al., 1999). A more thorough description of the scale development process can be found in Hanges and Dickson.

### 8.2. Theory-based expectations

Our research hypothesis was that these data should fit a multilevel factor structure, with a unidimensional factor at both levels. Information processing theories of leadership perceptions (e.g., Lord, Foti, & DeVader, 1984; Lord & Maher, 1991) would suggest that responses to the five trait items would reflect individual differences in mangers' implicit leadership schemas, and thus would vary meaningfully from person to person regardless of the individual's nationality. However, several researchers have argued that people from the same organizational or societal culture would hold common beliefs about the attributes of a typical leader (e.g., Bass, 1990; Dorfman et al., 2004; Gerstner & Day, 1994; Hanges, Lord, & Dickson, 2000; House, Wright, & Aditya, 1997). The commonality in these mental models is believed to result from repeated exposure to organizational as well as societal policies, practices, and procedures. Thus, the responses to these items not only reflect individual variability but also systematic society-wide differences in beliefs about effective leadership traits and behaviors. Such society-wide differences should be captured in the patterns of mean responses across societies. A multilevel confirmatory factor analysis can provide evidence about which traits are particularly reflective of the latent construct at each level of analysis (i.e., the individual and the society levels), by an inspection of the relevant factor loadings—with higher factor loadings indicating those traits that are particularly reflective of the latent construct at the given level of analysis.

### 8.3. Model estimation and assessment of fit

Currently, Mplus (all versions) and version 6.0 of EQS are designed to expedite the analysis of multilevel data. MLwiN (Rasbash, Browne, Goldstein, & Yang, 2000) may also be used to estimate some varieties of multilevel CFA models. In our example, all analyses were performed using Mplus version 2.12 (Muthen & Muthen, 2002). The default estimator for the multilevel analysis in Mplus is the MUML estimator, which for balanced data (i.e., where group sizes are equal) yields maximum likelihood parameter estimates, robust standard errors, and a mean-adjusted chi-square goodness-of-fit statistic. For unbalanced data, this is a robust limited-information estimator (Muthen & Muthen, 1998).

Multiple indices were referenced to determine model fit (Raykov & Marcoulides, 2000). First, model fit was assessed using the chi-square goodness-of-fit statistic. A well-fitting model would be expected to have a small (relative to its degrees of freedom), non-significant value of chi-square. However, when sample sizes are large, the chi-square statistic may be statistically significant even though the model is substantially correct. Thus, we also used the comparative fit index (CFI), the root mean square error of approximation (RMSEA), and the standardized root mean square residual (SRMR) as guides in assessing fit. For lack of any standard protocol with MCFA, we followed Hu and Bentler's (1999) recommendations in regards to fit indices. We looked for values of CFI in the mid 0.90's or higher, and RMSEA and SRMR of 0.05 or less. However, strict application of these standards should be avoided until simulations using multilevel models are conducted to confirm Hu and Bentler's recommendations.

### *8.4. Preliminary analyses*

Prior to the structural equation modeling analyses, the individual responses to the items were screened to determine if there was substantial skewness or kurtosis, as well as inspected for outliers. The values of skewness and kurtosis for all five items were acceptable. Two of the items showed a heavy concentration of outlier responses at the high extreme. In both cases, the outliers appeared in part to be systematic by society. More specifically, an unusually large proportion of Bolivian respondents (21%) rated in the highest category for 'habitual'. An unusually large proportion of Egyptian (15%) and black South African (13%) respondents rated in the highest category for 'ritualistic'. These outliers were retained in the dataset, as they appeared to reflect legitimate societal differences.

## 9. Step 1: Conventional confirmatory factor analysis of the sample total covariance matrix

An a priori one-factor model with paths from the latent construct to all five leadership items was tested, using the total sample matrix (see Appendix A for sample Mplus code to perform this analysis). This model appeared generally plausible but was not extremely well-fitting. Specifically, the chi-square goodness-of-fit statistic was large compared to the degrees of freedom, $\chi^2_{(5)}=181.63$, $p<0.001$, although the values of the CFI, the RMSEA, and the SRMR were in a range suggesting adequate fit (see Table 1).

Table 1
Model fit for a priori single- and multilevel models

| | $\chi^2$ | *df* | CFI | RMSEA | SRMR |
|---|---|---|---|---|---|
| Models | | | | | |
| Step 1: Total | 181.63 | 5 | 0.98 | 0.05 | 0.02 |
| Step 3: Within | 221.15 | 5 | 0.96 | 0.06 | 0.03 |
| Step 4: Between | 23.67 | 5 | 0.89 | 0.25 | 0.08 |
| Step 5: Multilevel | 394.66 | 10 | 0.96 | 0.05 | $W$=0.03, $B$=0.07 |

All chi-square values are statistically significant at $p<0.05$. *df*=degrees of freedom, CFI=comparative fit index, RMSEA=root mean square error of approximation, SRMR=standardized root mean square residual. $W$=within-group portion of the model. $B$=between-group portion of the model.

Table 2
Standardized factor loadings from step 1 and step 5 models and intraclass correlations from step 2 by scale item

| Item | Standardized loadings | | | |
|---|---|---|---|---|
| | Step 1: Total | Step 5: Within | Step 5: Between | ICC |
| Formal | 0.58 | 0.50 | 0.87 | 0.19 |
| Procedural | 0.71 | 0.61 | 0.92 | 0.26 |
| Habitual | 0.45 | 0.39 | 0.69 | 0.20 |
| Ritualistic | 0.46 | 0.48 | 0.56 | 0.10 |
| Cautious | 0.48 | 0.39 | 0.68 | 0.24 |

Table 2 shows the standardized factor loadings for this model. All loadings were statistically significant and suggest that all five items adequately reflect the latent construct.

As discussed earlier, this analysis fails to incorporate the hierarchical nature of the data and therefore the analysis is potentially misleading. Because the total covariance matrix contains both between- and within-group level information, the results from step 1 will be influenced by the group-level factor structure when there is substantial systematic between-group variance, even though disaggregate observations are being analyzed (Kreft & De Leeuw, 1998). Thus, the next step is to assess the amount of between-group level information contained in the total matrix.

## 10. Step 2: Estimation of between-group variance

For this step, the intraclass correlation coefficients were calculated for each indicator in order to determine the extent of systematic group-level variance (see Appendix B for sample Mplus code to obtain ICC values). Estimates of the ICC are automatically generated when a two-level analysis is requested in Mplus. As shown in the right-hand column of Table 2, the ICC values for these five items ranged from 0.10 to 0.26, with an average ICC of 0.20. As there is no set standard for the Muthen ICC, we compared it to the standard ICC(1) values reported by James (1982). In his review of climate studies, he found that ICC(1)'s tend to range between 0.00 and 0.50 with a median value of 0.12. Given our relatively high ICC values, we concluded that there was sufficient between-group variation to statistically warrant the use of multilevel analyses.

## 11. Step 3: Within-group factor structure

In the third step, a single-level CFA model is tested, this time using the covariance matrix (SPW) based on individual-level scores, adjusted for their respective group means. The first task is to create this matrix. If the analysis is performed in Mplus or EQS v.6 (or alternatively, a preprocessor such as Gustafsson & Stahl's, 2000 STREAMS), the relevant matrix can be easily output to an external file (see Appendix B for sample Mplus code used to create this matrix). However, the SPW matrix can also be created with conventional software by first subtracting group means from individuals' item responses to create the appropriate deviation score. A variance–covariance matrix is then created and its values corrected to reflect division by the appropriate denominator. (The typical devisor for the covariance matrix is $N-1$, but for the current purposes it should be $N-G$, where $G$=the number of groups. Thus, each element of the matrix needs to be transformed by multiplying by $N-1$ and then dividing by $N-G$.) It is this matrix that may then be used to assess the within-group factor structure (see Appendix C for sample Mplus code to perform this step).

The fit indices results from step 3 analyses are displayed in Table 1. As was the case in step 1, in this step, the a-priori model showed room for improvement in fit, $\chi^2_{(5)}$=221.13, $p$<0.001. Interestingly, the model chi-square values of step 1 are smaller (i.e., better fitting) for the same number of degrees of freedom than are the values from step 3. Because step 1 results are based on the total covariance matrix, it may be that the systematic between-group relationships that have been eliminated from the Within covariance matrix used for step 3 analyses contributed to the better fit of the step 1 model. In other words, this may show support for the constructs operating at the between-group level of analysis. The parameter estimates (factor loadings) for this model ranged from 0.39 to 0.61. These factor loadings for steps 3 and 4 are not reported in Table 2 because they were (and will typically be) very similar in value to the estimates from the final multilevel model.

## 12. Step 4: Between-group factor structure

In this step, the fit of a society-level CFA model to the estimated between-group population matrix is investigated. One must also create this matrix. Again, if the analysis is performed in Mplus, EQS v. 6, or STREAMS, the relevant matrix can be easily output to an external file (see Appendix B for sample Mplus code used to create this matrix). However, this matrix can also be created with conventional software by first obtaining the variance–covariance matrix of the group means. This matrix must also be corrected to reflect the appropriate denominator or divisor. To do this, one should multiple the elements of the matrix by the default divisor ($N-1$) and then divide the appropriate divisor, in this case, the between-group level, $G-1$ (where $G$=the number of groups). This corrected matrix is then used to assess the between-group factor structure (see Appendix C for sample Mplus code to perform this step).

The fit indices reported in Table 1 show that the step 4 a-priori model has substantially poorer fit than the model from the previous step. In particular, the RMSEA of 0.25 and CFI of 0.89 indicate room for improvement. The chi-square value, $\chi^2_{(5)}$=23.67, $p$<0.001, is much smaller than those seen for steps 1 and 3, but this is in part because of the substantially smaller sample size used in step 4 (i.e., the number of societies versus the number of individuals). Given that this is one of the typical ways that researchers have analyzed group-level constructs, the overall conclusion is that the proposed factor structure does not fit the data very well.

## 13. Step 5: Multilevel confirmatory factor analysis

Specifying the multilevel confirmatory factor analysis in Mplus or EQS is relatively straightforward (see Appendix D for an example of the Mplus code used in our MCFA). Occasionally, however, estimation problems arise, most frequently a failure to converge to a solution. When this happens, the unstandardized parameter estimates for the factor loading and item variance estimates of steps 3 and 4 can be provided as start values, or there may be a need to fix error variances to zero in order to avoid a negative variance estimate.

As shown in Table 1, the results from a test of the a priori model of step 5 suggest adequate fit at the within level, and adequate, but slightly worse fit at the between level as indicated by the SRMR of 0.07. Parameter estimates from this model included factor loadings at both the within and between level, as can be seen in Table 2. The items load strongly onto the single factor at the between level, ranging from 0.56 ('ritualistic') to 0.92 ('procedural'). Consistent with our initial belief that this scale was primarily operating at the societal level of analysis, the factor loadings of the items at the within level, ranging from 0.39 ('cautious' and 'habitual') to 0.61 (procedural), are not as strong as the between factor loadings (see Fig. 2). As the between-group level is the theoretical level of interest in our data, we found strong support for our one factor model. Notice the much higher loadings at the between-group level than both the results of the step 1 analyses, as well as from the within-group MCFA level loadings. Had we been interested in the within-group level, our factor loadings might have led us to form a three-item

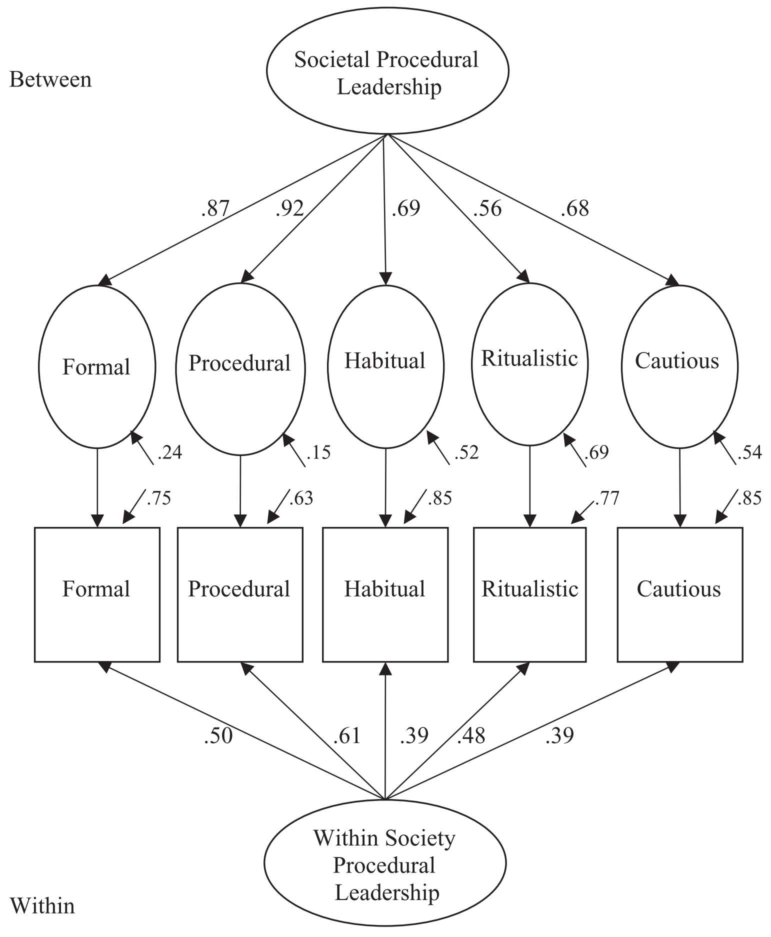


Fig. 2. Path diagram of final GLOBE one-factor multilevel model.

factor (deleting the two items with low factor loadings) rather than the five-item factor predicted and found at the between-group level.

While the factor loadings obtained from steps 3 and 4 were not reported in Table 2, it is important to reiterate that they will typically be very similar in value to the estimates obtained from this step. However, what does differ is the standard errors of these estimates. Namely, the standard errors of the estimates obtained taking into consideration the multilevel nature of the data are and typically will be substantially lower than those for the factor loadings in steps 3 and 4.

## 14. Discussion

Leadership researchers often collect data from multiple individuals who belong to the same workgroup. These researchers deliberately gather their data in this manner to provide reliable assessment of the characteristics of the leaders of these workgroups. Unfortunately, while much has been written on how to test multilevel relationships between constructs, there has been no discussion in the leadership literature about the way to verify the factor structure of aggregate-level scales.

In the present paper, we described Muthen's (1994) MCFA procedure and attempted to illustrate its usefulness for leadership research. This procedure progressively allows researchers to assess the factor structure of a scale at multiple levels of analysis. In our applied example, we used MCFA to investigate the societal level factor structure of a five-item 'procedural' leadership scale. This scale was designed to assess societal level differences in the shared leadership mental models. Our results illustrated that the societal level factor structure of this scale conformed to expectations. Also, we found that the factor loadings of the items were stronger at the between (i.e., societal level) than the within (i.e., within-society) levels of analysis. Indeed, our illustrative example typifies a fuzzy composition model. These findings were consistent with the intended nature of the scale.

We strongly believe that this technique has many different applications in the leadership literature. In particular, this technique was used extensively in the GLOBE project (Hanges & Dickson, 2004; House et al., 2004). Hanges and Dickson (2004) confirmed the factor structure of the GLOBE societal and organizational culture scales as well as the factor structure of the scales believed to measure organizational or societal level conceptualizations of leadership. MCFA provided initial construct validity evidence for these group level scales in that the factor structure of these scales were shown to be operating as expected at the organizational or societal level of analysis.

MCFA can also be used to test whether the structure of any construct differs across levels of analysis. Indeed, the work by Schwartz (1992, 1994) on values has documented that the same items can cluster differently at the individual versus the societal levels of analysis. This differentially clustering of items can happen with leadership constructs as well. The MCFA provides a methodology for systematic exploration of these issues. Further, Mayer's (2004) study of group level effects of LMX on justice and customer satisfaction could have applied MCFA to test the robustness of the LMX scale at the group level of analysis. Indeed, MCFA comes into play whenever multiple respondents rate a common target (e.g., team leadership, team climate, organizational climate), or, in other words, whenever the data is hierarchically structured.

MCFA is just now becoming more commonly accessible with software packages. That is, in addition to Mplus, EQS v.6 just recently incorporated Muthen's MUML as well as the Bentler, Liang, and Chou, Bentler, and Pentz approaches to multilevel estimation. Other conventional SEM software packages will no doubt be soon to follow, making MCFA easier to perform.

Finally, while MCFA has substantial potential for helping leadership researchers, it is important to recognize that this research is still relatively new. With the emergence of software packages that can now handle a multitude of multilevel analyses, we need to expand our knowledge and understanding of MCFAs. More specifically, we need to more fully understand acceptable limits for the parameters obtained from multilevel factor analyses. That is, what types of factors influence the values we get? For instance, what role do sample size, group size, and ICC magnitude have on the analysis? With the statistical capability to now account for hierarchical influences, we hope that future research, via simulations and studies, will begin to examine the issue of what characteristics might impact the analytic results. One initial study, discussed in Mok (1995), indicated that MCFA works reasonably well when the multilevel dataset consists of a total of 800 or more observations. Interestingly, it appears not to matter whether this 800 observation threshold comes from a large number of groups (and fewer individuals within each group) or vice versa (Mok, 1995). It is our hope that this paper will lead to a more widespread use and understanding of MCFAs.

## 15. Conclusions

Researchers have demonstrated that leadership constructs do indeed operate at the person, dyad, group, and collective levels of analysis (Dansereau & Yammarino, 1998a, 1998b; Yammarino, Dansereau, & Kennedy, 2001). Yet, "non-independence among individuals is a central fact of real groups,. . . [that is] norms, climate and social interactions cannot be always eliminated when groups are studied. . .The statistical requirement of independence should not block progress to a field in which interaction among persons necessarily results in dependence" (Kenny & LaVoie, 1985, p. 339). The MCFA has potential to allow leadership research to progress. That is, we may be better able to show that some indicators are clearly operating at the group level, or better viewed at the individual level. Such results will help in the development of composition models of the constructs often used in leadership research. It is necessary to reiterate, however, the importance of beginning any analyses with a solid definition of the construct at the theoretically appropriate level of analysis. That is, while we now have the statistical means to examine multiple levels of analysis, it is imperative that the level chosen is driven by theory, and not by the data. In other words, with statistical means such as the MCFA, we can now allow our theory to dictate the analyses we opt to perform, and not our analyses determining our theory (Guzzo, 1998). As such, multilevel research methods arguably have great potential and applicability for leadership research.

## Acknowledgements

We would like to thank Dr. Robert House and the GLOBE project for giving us permission to use a portion of their data to illustrate the multilevel confirmatory factor analysis technique. This work was partially funded by the National Science Foundation under Grant No. 9711143.

## Appendix A

Sample Mplus code to perform a conventional confirmatory factor analysis (CFA).

| | |
|---|---|
| TITLE | Confirmatory factor analysis sample |
| DATA | File is procedural.dat; (provide the file name of interest, for the total matrix it will be the raw data); format is f2.0, 5f1.0; (example of fixed formatting, other formatting options exist) |
| VARIABLE | Names are country formal cautious habitual procedrl ritlistc; (provide the names of all variables in the data set); usev formal cautious habitual procedrl ritlistc; (provide the names of those variables you would like to use in the CFA); missing is blank; (specify values for missing data) |
| MODEL | F1T by formal cautious habitual procedrl ritlistc; (specify the which variables will load onto the latent factor, in this case, formal through ritlistc load are indicators for F1T) |
| OUTPUT | Mod Stand Sampstat Res; (provides modification indices, standardized coefficients, residuals) |

## Appendix B

Sample Mplus program to create within and between level covariance matrices and obtain ICC values

INPUT INSTRUCTIONS

| | |
|---|---|
| TITLE | Create within and between matrices and obtain ICC values |
| DATA | File is procedural.dat; format is f2.0, 5f1.0; |
| VARIABLE | Names are country formal cautious habitual procedrl ritlistc; missing is blank; cluster is country; (*specify grouping variable*) |
| ANALYSIS | TYPE=MEANSTRUCTURE TWOLEVEL BASIC; |
| OUTPUT | sampstat; |
| SAVEDATA | FILE (SIGB) IS BetCov.dat; (*saves the corrected between covariance matrix*) FILE (SAMPLE) IS WinCov.dat; (*saves the corrected within covariance matrix*) |

## Appendix C

Sample Mplus program to perform a confirmatory factor analysis on the within/between matrix

| | |
|---|---|
| TITLE | Procedural factor estimated with within or between matrix |
| DATA | File is WinCov.dat; (*specifies the matrix of interest, either the corrected within or between*) TYPE IS COVA; (*specifies the matrix as variance/covariance matrix*)FORMAT IS FREE; NOBSERVATIONS IS 13351; |
| VARIABLE | Names are formal cautious habitual procedrl ritlistc; |
| MODEL | F1W by formal cautious habitual procedrl ritlistc; |
| OUTPUT | Stand Mod Res; |

## Appendix D

Sample Mplus code to perform the MCFA

| | |
|---|---|
| TITLE | Multilevel confirmatory factor analysis for GLOBE procedural factor |
| DATA | File is procedural.dat; (*note—may use the raw, total data set to perform this step*) format is f2.0, 5f1.0; |
| VARIABLE | Names are country formal cautious habitual procedrl ritlistc; missing is blank; cluster is country; |
| ANALYSIS | Type is twolevel; iterations is 5000; |
| MODEL | %between% f1Bet by formal cautious habitual procedrl ritlistc; %within% f1With by formal cautious habitual procedrl ritlistc; |
| OUTPUT | Stand Res; |

## References


Avolio, B. J., & Bass, B. M. (1998). Individual consideration viewed at multiple levels of analysis: A multi-level framework for examining the diffusion of transformational leadership. In F. Dansereau, & F. J. Yammarino (Eds.), *Leadership: The multiple-level approaches, contemporary and alternative* (pp. 53–74). Stamford, CT: JAI Press.

Bass, B. M. (1990). *Bass and Stogdill's handbook of leadership: Theory, research, and managerial applications* (3rd ed.). New York: Free Press.

Bentler, P. M. (1995). *EQS: Structural equations program manual*. Encino, CA: Multivariate Software.

Bliese, P. D. (1995). Within-group agreement, non-independence, and reliability: Implications for data aggregation and analysis. In K. J. Klein, & S. W. J. Kozlowski (Eds.), *Multilevel theory, research, and methods in organizations: Foundations, extensions, and new directions* (pp. 349–381). San Francisco, CA: Jossey-Bass.

Bliese, P. D., & Hanges, P. J. (2004). Being both too liberal and too conservative: The perils of treating grouped data as though it is independent. *Organizational Research Methods, 7*, 400–417.

Chan, D. (1998). Functional relations among constructs in the same content domain in different levels of analysis: A typology of composition models. *Journal of Applied Psychology*, *83*, 234–246.

Chou, C. P., Bentler, P. M., & Pentz, M. A. (2000). A two-stage approach to multilevel structural equation models: Application to longitudinal data. In T. D. Little, K. U. Schnabel, & J. Baumert (Eds.), *Modeling longitudinal and multilevel data: Practical issues, applied approaches, and specific examples* (pp. 33–50). Mahwah, NJ: Lawrence Erlbaum Associates.

Cronbach, L. J. (1976). *Research on classrooms and schools: Formulation of questions, design, and analysis*. California: Stanford University. Unpublished manuscript.

Dansereau, F., Alutto, J. A., & Yammarino, F. J. (1984). *Theory testing in organizational behavior: The varient approach*. Englewood Cliffs, NJ: Prentice-Hall.

Dansereau, F., & Yammarino, F. J. (1998a). *Leadership: The multiple-level approaches, classical and new wave*. Samford, CT: JAI Press.

Dansereau, F., & Yammarino, F. J. (1998b). *Leadership: The multiple-level approaches, contemporary and alternative*. Stamford, CT: JAI Press.

Dansereau, F., Yammarino, F. J., & Markham, S. E. (1995). Leadership: The multiple-level approaches. *The Leadership Quarterly*, *6*, 251–263.

Den Hartog, D. N., House, R. J., Hanges, P. J., Ruiz-Quintanilla, S. A., & Dorfman, P. W. (1999). Culture specific and cross-culturally generalizable implicit leadership theories: Are attributes of charismatic/transformational leadership universally endorsed? *The Leadership Quarterly*, *10*(2), 219–256.

Dorfman, P. W., Hanges, P. J., & Brodbeck, F. (2004). Cultural endorsed leadership. In R. J. House, P. J. Hanges, M. Javidan, P. W. Dorfman, & V. Gupta (Eds.), *Cultures, leadership, and organizations: The GLOBE study of 62 cultures* (pp. 669–720). Thousand Oaks, CA: SAGE Publications.

Gerstner, C. R., & Day, D. V. (1994). Cross-cultural comparison of leadership prototypes. *The Leadership Quarterly*, *5*, 121–134.

Gustafsson, J. E., & Stahl, P. A. (2000). *STREAMS user's guide, version 2.5 for Windows*. Mölndal Sweden: MultivariateWare.

Guzzo, R. A. (1998). Leadership, self-management, and levels of analysis. In F. Dansereau, & F. J. Yammarino (Eds.), *Leadership: The multiple-level approaches, classical and new wave* (pp. 213–219). Stamford, CT: JAI Press.

Hall, R. J., Hanges, P. J., & Dyer, N. G. (forthcoming). Two-level structural equation modeling with latent variables: An illustration of confirmatory factor analysis with applications to organizational research. *Organizational Research Methods*.

Hall, R. J., & Lord, R. G. (1995). Multi-level information-processing explanations of followers' leadership perceptions. *The Leadership Quarterly*, *6*, 265–287.

Hall, R. J., & Lord, R. G. (1998). Multi-level information-processing explanations of followers' leadership perception. In F. Dansereau, & F. J. Yammarino (Eds.), *Leadership: The multiple-level approaches, contemporary and alternative* (pp. 159–184). Stamford, CT: JAI Press.

Hall, R. J., Makiney, J. D., Marchioro, C. A., Tan, J. A., & Klein, K. J. (1999). Applying multilevel structural equation modeling techniques to the study of organizational behavior. *Paper presented at the 14th Annual Conference of the Society for Industrial and Organizational Psychology, Atlanta, GA*.

Hanges, P. J., & Dickson, M. W. (2004). The development and validation of the GLOBE culture and leadership scales. In R. J. House, P. J. Hanges, M. Javidan, P. W. Dorfman, & V. Gupta (Eds.), *Cultures, leadership, and organizations: The GLOBE study of 62 cultures* (pp. 122–151). Thousand Oaks, CA: SAGE Publications.

Hanges, P. J., Dickson, M. W., & Sipe, M. (2004). Rationale for GLOBE statistical analyses: Societal rankings and tests of hypotheses. In R. J. House, P. J. Hanges, M. Javidan, P. W. Dorfman, & V. Gupta (Eds.), *Culture, leadership, and organizations: The GLOBE study of 62 cultures* (pp. 219–234). Thousand Oaks, CA: SAGE Publications.

Hanges, P. J., Lord, R. G., & Dickson, M. W. (2000). An information processing perspective on leadership and culture: A case for connectionist architecture. *Applied Psychology: An International Review*, *49*, 133–161.

Harnqvist, K. (1978). Primary mental abilities of collective and individual levels. *Journal of Educational Psychology*, *70*, 706–716.

House, R. J., Hanges, P. J., Javidan, M., Dorfman, P. W., & Gupta, V. (Eds.). (2004). *Culture, leadership, and organizations: The GLOBE study of 62 cultures*. Thousand Oaks, CA: SAGE Publications.

House, R. J., Hanges, P. J., Ruiz-Quintanilla, S. A., Dorfman, P. W., Javidan, M., Dickson, M., et al. (1999). Cultural influences on leadership and organizations: Project GLOBE. *Advances in Global Leadership*, *1*, 171–233.

House, R. J., Wright, N. S., & Aditya, R. N. (1997). Cross-cultural research on organizational leadership: A critical analysis and a proposed theory. In P. C. Earley, & M. Erez (Eds.), *New perspectives in international industrial organizational psychology* (pp. 535–625). San Francisco: New Lexington.

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

James, L. R. (1982). Aggregation bias in estimates of perceptual agreement. *Journal of Applied Psychology*, *67*, 219–229.

James, L. R., Demaree, R. G., & Wolf, G. (1984). Estimating within-group interrater reliability with and without response bias. *Journal of Applied Psychology*, *69*, 85–98.

James, L. R., Demaree, R. G., & Wolf, G. (1993). Rwg: An assessment of within-group interrater agreement. *Journal of Applied Psychology*, *78*, 306–309.

Joreskog, K. G., & Sorbom, D. (1989). *LISREL 7: A guide to the program and applications* (2nd ed.). Chicago: SPSS.

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

Kenny, D. A., & LaVoie, L. (1985). Separating individual and group effects. *Journal of Personality and Social Psychology*, *48*, 339–348.

Klein, K. J., Dansereau, F., & Hall, R. J. (1994). Levels issues in theory development, data collection, and analysis. *Academy of Management Review*, *19*, 195–229.

Klein, K. J., & Kowzlowski, S. W. J. (2000). From micro to meso: Critical steps in conceptualizing and conducting multilevel research. *Organizational Research Methods*, *3*, 211–236.

Klein, K. J., Tosi, H., & Cannella, A. A., Jr. (1999). Multilevel theory building: Benefits, barriers, and new developments. *Academy of Management Review*, *24*, 243–248.

Kozlowski, S. W. J., & Klein, K. J. (2000). A multilevel approach to theory and research in organizations: Contextual, temporal, and emergent processes. In K. J. Klein, & S. W. J. Kozlowski (Eds.), *Multilevel theory, research, and methods in organizations: Foundations, extensions, and new directions* (pp. 3–90). San Francisco, CA: Jossey-Bass.

Kreft, I., & De Leeuw, J. (1998). *Introducing multilevel modeling*. Thousand Oaks, CA: SAGE Publications.

Liang, J., & Bentler, P. M. (in press). An EM algorithm for fitting two-level structural equation models. *Psychometrika*.

Lord, R. G., Foti, R., & De Vader, C. (1984). A test of leadership categorization theory: Internal structure, information processing, and leadership perceptions. *Organizational Behavior and Human Performance*, *34*, 343–378.

Lord, R. G., & Maher, K. J. (1991). *Leadership and Information Processing: Linking Perceptions and Performance*, *vol. 1*. Cambridge: Unwin Hyman.

Mayer, D. M. (2004). Are you in or out? A group level examination of the effects of LMX on justice and customer satisfaction. Unpublished doctoral dissertation. University of Maryland, College Park, MD.

Mok, M. (1995). *Sample size requirements for 2-level designs in educational research*. Sydney, Australia: Macquarie University.

Morgeson, F. P., & Hofmann, D. A. (1999). The structure and function of collective constructs: Implications for multilevel research and theory development. *Academy of Management Review*, *24*, 249–265.

Mumford, M. D. (1998). Situations, interactions, and followers: The case of individualized leadership. *Leadership: The multiple-level approaches, classical and new wave* (pp. 421–428). Stamford, CT: JAI Press.

Muthen, B. O. (1990). *Mean and covariance structure analysis of hierarchical data*. Paper presented at the Psychometric Society, Princeton, New Jersey.

Muthen, B. O. (1994). Multilevel covariance structure analysis. *Sociological Methods and Research*, *22*, 376–398.

Muthen, B. O., & Muthen, L. K. (1998). *Mplus: User's guide*. Los Angeles: Muthen & Muthen.

Muthen, L. K., & Muthen, B. O. (2002). *Mplus VERSION 2.12: Addendum to The Mplus Users's Guide*. Los Angeles: Muthen & Muthen.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric Theory*. New York: McGraw Hill.

Ployhart, R. E., Holtz, B. C., & Bliese, P. D. (2002). Longitudinal data analysis: Applications of random coefficient modeling to leadership research. *The Leadership Quarterly*, *13*, 455–486.

Rasbash, J., Browne, W., Goldstein, H., & Yang, M. (2000). *A user's guide to MLwiN* (2nd ed.). London: Institute of Education.

Raykov, T., & Marcoulides, G. A. (2000). *A first course in structural equation modeling*. Mahwah, NJ: Lawrence Erlbaum Associates.

Schwartz, S. (1992). Universals in the content and structure of values: Theoretical advances and empirical tests in 20 countries. In M. Zanna (Ed.), *Advances in Experimental Social Psychology*, *vol. 25*. (pp. 1–65). New York: Academic Press.

Schwartz, S. (1994). Beyond individualism-collectivism: New cultural dimensions of values. In U. Kim, H. C. Triandis, C. Kagitcibasi, S. C. Choi, & G. Yoon (Eds.), *Individualism and collectivism: Theory, method, and applications* (pp. 85–119). London: SAGE.

Yammarino, F. J., & Bass, B. M. (1991). Person and situation views of leadership: A multiple levels of analysis approach. *The Leadership Quarterly*, *2*, 121–139.

Yammarino, F. J., Dansereau, F., & Kennedy, C. J. (2001). A multiple-level multidimensional approach to leadership: Viewing leadership through an elephant's eye. *Organizational Dynamics*, *29*, 149–163.