Organizational Research
Methods

Dispersion-Composition Models in Multilevel Research : A Data-Analytic Framework
Michael S. Cole, Arthur G. Bedeian, Robert R. Hirschfeld and Bernd Vogel

The online version of this article can be found at:

Published by:

$SAGE

On behalf of:

RMD
Academy of Management

The Research Methods Division of The Academy of Management

Additional services and information for *Organizational Research Methods* can be found at:

Email Alerts: http://orm.sagepub.com/cgi/alerts

Subscriptions: http://orm.sagepub.com/subscriptions

Reprints: http://www.sagepub.com/journalsReprints.nav

Permissions: http://www.sagepub.com/journalsPermissions.nav

Citations: http://orm.sagepub.com/content/14/4/718.refs.html

Dispersion-Composition Models in Multilevel Research: A Data-Analytic Framework

Michael S. Cole[1], Arthur G. Bedeian[2], Robert R. Hirschfeld[3], and Bernd Vogel[4]

Abstract

Multilevel researchers have predominantly applied either *direct consensus* or *referent-shift consensus* composition models when aggregating individual-level data to a higher level of analysis. This prevailing focus neglects both theory and empirical evidence, suggesting that the variance of group members' responses may complement the absolute mean level of group members' judgments. The goals of this article are to demonstrate the application of dispersion-composition models for capturing variability among group members' collective judgments and highlight the statistical challenges (and inherent constraints) of using group means and variances as predictors of study criteria. To this end, the authors present and illustrate a six-step sequential framework for applying dispersion-composition models using data from two independent field samples. The authors contend that the application of dispersion-composition models not only will strengthen a study's conclusions by eliminating potential rival data interpretations but may also shed new light on past findings, potentially opening new doors to a more complete understanding of multilevel phenomena.

Keywords

multilevel research, composition models, dispersion theory, group-level constructs

The proper measurement of collective phenomena that emerge from bottom-up processes is a perennial concern among multilevel researchers (see, e.g., Gooty & Yammarino, 2011; van Mierlo, Vermunt, & Rutte, 2009). This concern is highlighted in multilevel research when data from a lower level are used to compose higher level constructs. In such situations, researchers have commonly

[1] M. J. Neeley School of Business, Department of Management, Texas Christian University, Fort Worth, TX, USA
[2] College of Business Administration, Department of Management, Louisiana State University, Baton Rouge, LA, USA
[3] College of Business Administration, Department of Management and Quantitative Methods, University of Colorado at Colorado Springs, Colorado Springs, CO, USA
[4] Henley Business School, School of Management, University of Reading, Oxfordshire, UK

Corresponding Author:
Michael S. Cole, M. J. Neeley School of Business, Department of Management, Texas Christian University, Fort Worth, TX 76129, USA
Email: m.s.cole@tcu.edu

used Chan's (1998) typology of composition models to specify the functional relationship between phenomena at different levels of analysis. Chan's typology describes five basic or ideal forms of composition and, in essence, provides a framework for mapping the transformation of constructs across levels. In three of these forms (*direct consensus, referent-shift consensus,* and *dispersion*), level of agreement or homogeneity across individual group members' judgments is a central consideration. Conversely, *additive composition models* operationalize higher level constructs as the simple sum or average of scores on lower level variables, regardless of within-unit variance. As typically applied, the validity of an additive index (either a simple sum or average) provides empirical support for the transformation of a functional relationship across levels. The remaining form (*process composition*) focuses on the mechanisms by which a construct associated with episodes of or changes in behaviors occurring at a lower level of conceptualization emerge at a higher level. Given that our objective is to underscore the notion that variability among group members' perceptions is both theoretically and practically relevant, in the present instance we focus on the three composition models for which agreement is a consideration: direct consensus, referent-shift consensus, and dispersion-composition.

Chan's Composition Models and Within-Group Agreement

In aggregating lower level scores to index a target construct, a majority of multilevel researchers interested in group or team processes have adopted either a direct consensus or a referent-shift consensus model (Klein, Conn, Smith, & Sorra, 2001; van Mierlo et al., 2009).[1] Direct consensus models typically use averaged individual members' responses to operationalize group-level scores (Chan, 1998). For example, team members may be asked whether their work is psychologically empowering (e.g., "I make a difference in this organization"). Assuming some minimal level of within-team interrater agreement (IRA) and interrater reliability (IRR) consensus (Bliese, 2000; LeBreton & Senter, 2008), individual team member responses have generally then been aggregated (typically using the simple mean) to represent a group-level empowerment construct.

Referent-shift consensus models mirror their direct consensus counterparts in that they also require computing IRA and IRR indices across team or group members. Given sufficient IRA and IRR consistency, referent-consensus models typically use an index of central tendency (i.e., the mean) of lower level (individual) scores to approximate a group's standing on a higher level construct. There is an essential difference, however, between these models. Rather than simply averaging individual team members' scores, as is done in direct consensus models, referent-shift consensus models require individual team members to respond to survey items in reference to a higher level unit (Chan, 1998). Researchers interested in a team's empowerment beliefs might ask individual team members whether their team's work matters to their company (e.g., "My team makes a difference in this organization"). Thus, rather than asking team members about their individual perceptions, referent-consensus models incorporate a different referent (i.e., a team as a whole). In doing so, such models derive a theoretically distinct higher level form (e.g., team empowerment) of a lower level construct (e.g., psychological empowerment) by shifting content across levels.

Although direct consensus and referent-shift consensus models are widely used in multilevel research, an important limitation of both forms of composition is their reliance on mean scores. Lindell and Brandt (2000), for example, have shown that the exclusive use of the simple average of lower level (individual) scores to approximate group-level phenomena obscures the true distribution of the underlying constituent responses. In other words, the use of aggregate individual-level responses fails to consider that variation among team members' collective judgments may provide meaningful information. As a result, the use of aggregated responses as surrogates for meso-oriented constructs provides an insufficient basis for summarizing phenomena of or within groups. Indeed, prior results suggest that using only mean-based variables derived from direct consensus and

referent-shift models may oversimplify group-level phenomena and result in biased (i.e., understated) estimates and equivocal findings (e.g., Colquitt, Noe, & Jackson, 2002; Dineen, Noe, Shaw, Duffy, & Wiethoff, 2007; Naumann & Bennett, 2002).

In a further criticism of studies applying mean-based approaches, it has been noted that only groups with high within-group agreement (or alternatively, minimal dispersion) are judged to be appropriate when analyzing multilevel data. Consequently, published research on group-level phenomena largely generalizes to high-agreement groups only; that is, groups in which, based on scores among individual members, there is a sufficient level of agreement to justify aggregation (Quigley, Tekleab, & Tesluk, 2007; van Mierlo et al., 2009). Whereas most multilevel researchers seemingly view high agreement on group-level constructs as a favorable finding, Meade and Eby (2007) and DeRue, Hollenbeck, Ilgen, and Feltz (2010) have described situations in which too much within-team agreement may be detrimental to a team's effectiveness.

In a third criticism, Mathieu, Maynard, Rapp, and Gilson (2008) have challenged the tacit assumption that group members necessarily perceive constructs in a relatively uniform manner. In doing so, they have speculated that unpacking variance residing across group members' perceptions may offer valuable insights. Although researchers (e.g., Mossholder & Bedeian, 1983) have long acknowledged that the extent of dispersion among group members' perceptions is "an integral element in the definition of [a] group-level construct" and, thus, "more than a statistical hurdle" (Klein et al., 2001, p. 4), dispersion-composition models have been rarely applied in multilevel research. Further, in those instances when dispersion-based constructs are considered, the accompanying analyses are often conducted improperly, or relevant information for judging their tenability is not reported. This suggests that despite calls for the increased application of dispersion-composition models in validating and understanding multilevel constructs (e.g., Meade & Eby, 2007), how to best incorporate dispersion constructs into multilevel analyses is not widely understood.

Given these criticisms and concerns, the goal of the present article is to demonstrate the application of dispersion-composition models for capturing variability among team members' collective judgments and, thus, highlight the statistical challenges (and inherent constraints) of using group means and variances as predictors of study criteria. To this end, we present a six-step sequential framework (Table 1) for applying dispersion-composition models. We integrate relevant measurement, design, and analytical considerations to provide a nontechnical tutorial and methodological resource for designing or evaluating dispersion-composition models. Of particular significance is that an integrative framework for promoting consistency in the way in which this research is conducted and reported is not readily available. This has implications when studies employing different data analytic procedures yield contradictory and noncomparable findings. We illustrate the proposed framework with two examples incorporating field data. These examples demonstrate that, if the steps we outline are improperly applied, a study's findings may be at best ambiguous and at worst erroneous. Although both examples focus on individuals nested in teams, our underlying logic also applies to other collective entities such as departments in organizations, geographically dispersed facilities within multiunit organizations, and organizations within industries or networks.

The notion that variability among group members' judgments may provide meaningful information is consistent with the basic logic underlying within- and between-entities analysis (WABA; Dansereau & Yammarino, 2000). The first step of the WABA procedure involves assessing whether a construct of interest varies (a) primarily between groups (suggesting group members are homogeneous within groups), (b) within groups (suggesting groups members are heterogeneous within groups), or (c) independently (suggesting groups members are independent of groups). As such, a beneficial aspect of WABA is that rather than taking a dichotomous perspective (i.e., a higher order construct cannot be said to exist without high within-unit agreement; Klein et al., 2001), different types of unit-level effects may be inferred. On the other hand, it has been shown that WABA conclusions may be erroneous when unit scores suffer from range restriction (George & James, 1993)—

Table 1. Suggestions for Measurement, Research Design, and Analysis of Dispersion Constructs

	Proposed Analysis Sequence	Reference
Theory	Step 1: Develop appropriate theory that describes how, when, and why a dispersion construct might influence study criteria.	Harrison & Klein (2007)
Measurement and design	Step 2a: Scrutinize the pros and cons of alternative dispersion indices to determine the appropriate statistical index of dispersion and explicitly state how dispersion will be assessed; recognize that subtle differences in alternative dispersion indices can influence results.	Roberson, Sturman, & Simons (2007)
	Step 2b: Explicitly state how dispersion will be calculated; determine the extent to which varying sample sizes result in biased dispersion scores; in the presence of a biased measure of dispersion, employ adjusted formulas when calculating a dispersion index.	Sin & Newman (2005); Biemann & Kearney (2010)
	Step 3: Use all available data. Within-group agreement will differ across groups, even when there is evidence of construct validity at the group level; low r_{wg} should not be interpreted to imply that a group-level construct does not exist. Excluding groups with low r_{wg} values has two practical disadvantages: (a) reduced statistical power because fewer groups are available for analysis and (b) greater restriction of range in dispersion index values for the remaining groups.	Lindell & Brandt (2000)
Data analysis	Step 4: Examine the degree of statistical interdependence between the level and dispersion components of an isomorphic group-level construct by computing the correlation between the level (aggregate mean) and dispersion (variance) of the phenomena in question. Interpret the magnitude and direction of the correlation coefficient for systematic range restriction.	Lindell & Brandt (2000)
	Step 5: Because the level and dispersion of a group-level construct are not statistically independent, control for the absolute mean level in testing dispersion hypotheses.	Bliese & Halverson (1998)
	Step 6: Test for curvilinear relationships between the level and dispersion component of a proposed interaction by using analytical models that consider higher order (i.e., squared) terms.	Dineen, Noe, Shaw, Duffy, & Wiethoff (2007)

a common occurrence in research using dispersion-composition models (Lindell & Brandt, 2000). Readers interested in WABA's applications may wish to consult Dansereau and Yammarino (2000), Dansereau, Cho, and Yammarino (2006), and Gooty and Yammarino (2011).

Exploring Dispersion Constructs: A Framework and Methodological Considerations

Multilevel researchers have typically applied dispersion constructs in two ways. First, they have used dispersion constructs to discern whether variability (or, alternatively, within-group agreement or consensus) in group member perceptions has a main or direct effect on study criteria. Alternatively, the variability associated with dispersion constructs has been used to identify interactive or moderator effects. Table 1 presents a six-step sequential framework that is appropriate for applying dispersion-based constructs for either purpose. The table highlights three basic phases of a research project: (a) theory development, (b) measurement and design, and (c) data analysis. Further,

within each phase, the table offers a proposed sequence for analyzing or evaluating dispersion-composition models. The six steps in Table 1 are presented in a particular order; thus, each step should be conducted (and reported) in the arrangement shown. The reader is cautioned against avoiding or missing any of the proposed steps because doing so will likely yield incorrect multilevel inferences (as our empirical examples illustrate). The column labeled *reference* directs interested readers to further reading related to the six sequential steps.

Dispersion as a Main (Direct) Effect

In a small number of studies, multilevel researchers have applied dispersion-based constructs to establish a direct or main effect (e.g., Bliese & Halverson, 1998). Although these researchers have recognized the importance of variability in group member perceptions, considerable inconsistency exists regarding the procedures used for this purpose. In an effort to shape future applications, we apply the framework outlined in Table 1 to illustrate how to determine whether the extent of dispersion among group members' perceptions has a significant direct effect on study criteria.

Step 1: theoretical issues. As Chan (1998) notes, in far too many instances the logic underlying the mapping of functional relationships among phenomena at different levels has been inadequate. This failure has generated confusion and controversy, as the levels to which results should be appropriately generalized have been open to question. The resulting theoretical ambiguity has, in turn, obstructed the efficient accumulation of research findings. Thus, the first step identified in Table 1 is to develop appropriate theory that describes how, when, and why a dispersion construct might influence study criteria. In the present context, it would be incumbent on a researcher to articulate the theoretical basis for conceptualizing a focal construct in terms of its variability (see Harrison & Klein, 2007, for a detailed review). A research study's purpose and ever-present practical considerations would also require consideration.

Steps 2 and 3: measurement and study design. In Steps 2a and 2b, a researcher should carefully scrutinize the pros and cons of alternative dispersion indices (e.g., coefficient of variation, r_{wg}-based indexes, and standard deviation) and explicitly state how dispersion will be assessed and calculated. This is important because subtle differences in dispersion indices (Roberson, Sturman, & Simons, 2007) and variations in group sizes (Biemann & Kearney, 2010) can influence research results. Further, Sin and Newman (2005) have illustrated that because the common practice of averaging over item-level variances introduces a measurement artifact, it is more appropriate to operationalize a dispersion construct using the variance associated with individual group members' overall construct scores and not the mean of the individual item variances.

Concerning Step 3 in Table 1, we agree with others who advocate that all available data (e.g., Carron et al., 2003), rather than only data from groups with high within-group agreement, should be used for multilevel analyses. Although the measurement of group-level constructs is a two-step process—that is, demonstrate IRA and IRR consistency and then aggregate individual responses by group (LeBreton & Senter, 2008)—this process (for reasons noted above) should not be used to exclude groups that do not exceed arbitrary ''hurdle rates'' (e.g., r_{wg} $\geq$ .70; Lance, Butts, & Michels, 2006). Further, given the low statistical power of dispersion variables (Roberson et al., 2007), excluding groups with low r_{wg} values has two practical disadvantages: (a) a further reduction of statistical power because fewer groups are available for analysis and (b) greater restriction of range in dispersion index values for the remaining groups (Bliese & Halverson, 1998).

Steps 4 and 5: data analysis considerations. Steps 4 and 5 involve addressing the inherent nonindependence of observations within groups, which lead to biases in standard errors of estimated mean scores (Bliese & Hanges, 2004). It has been shown that, under conditions routinely encountered by multilevel researchers, mean-based (level) and variance-based (dispersion) components of the same group-level construct are not statistically independent; the systematic relationship between level and dispersion is symmetrically curvilinear (i.e., ∩-shaped; Dineen et al., 2007; Lindell & Brandt, 2000). This statistical interdependence is particularly salient in situations where all of a group's members respond with an extreme rating on the survey items representing a construct. For example, given a 1-to-5 Likert-type response ramp, the only way for a group's mean score on a construct to equal 1.0 is when all group members' responses are 1.0, so that the standard deviation (i.e., dispersion) is equal to 0. Similarly, there is only one way for a group's mean construct score to equal 5.0; that is, all group members' responses must equal this maximum value, so that the standard deviation is again equal to 0. Thus, because the level and dispersion of ratings on discrete response scales (assuming bottom-up aggregation) are functionally dependent, floor and ceiling constraints exist for group-level scores as dispersion decreases.

The effect of such range restriction carries several methodological implications. First, a restriction in the potential range of values may cause the relationship between a dispersion variable and a criterion to be systematically underestimated. Further, pronounced range restriction may produce such strong intercorrelations—between mean-based (level) and variance-based (dispersion) components—that unique variance attributable to a dispersion variable may be statistically inaccessible (Lindell & Brandt, 2000). Hence, Bliese and Halverson (1998) and Lindell and Brandt (2000) have cautioned that when multilevel researchers are interested in direct (linear) relationships, the degree of interdependence between the level and dispersion components of the same group-level variable must be examined.

The degree of interdependence may be examined by computing the correlation between the level (aggregate mean) and dispersion (variance) of the phenomenon or construct in question. A significant and strong correlation would indicate systematic range restriction. The direction of the correlation would reveal whether the range restriction is for either high- or low-scoring groups. If there is considerable range restriction among teams with, for example, the lowest empowerment beliefs, then the left-hand portion of the symmetrically curvilinear (i.e., ∩-shaped) distribution between level and dispersion will disappear so that only the right-hand portion remains. As a result, an increasingly negative correlation between empowerment level and empowerment dispersion will be observed as the minimum score on empowerment level approaches the midpoint of an underlying response scale (see Lindell & Brandt, 2000, pp. 334-335). Step 4, in the suggested framework, includes computing this correlation coefficient and interpreting its magnitude and direction.

Finally, because of statistical dependence, absolute-level effects should be treated as a covariate when exploring the relationship between a dispersion variable and a criterion (Step 5). It is impossible to determine the extent to which a dispersion variable is actually related to the study criteria being examined without establishing *and* controlling for the degree of interdependence between the level and dispersion components of a group-level predictor. Steps 4 and 5 are crucial because failing to consider and control for absolute level effects leaves ''open the possibility that observed variance effects are a spurious by-product of absolute level effects'' (Bliese & Britt, 2001, p. 433) and, by extension, doubt as to whether a dispersion effect actually exists.

Level × Dispersion Interactions

As noted, multilevel researchers have also combined direct consensus (or referent-shift) composition models and dispersion-composition models to test for new multilevel effects (e.g., Dineen et al., 2007). Such applications involve exploring possible interactions. Indeed, Harrison and Klein

(2007) have argued that a mean level-by-variability interaction is implicit when specifying a dispersion variable. This suggests possible curvilinear considerations and, thus, adds a sixth step to the Table 1 framework.

Step 6: curvilinear considerations. Of particular concern in investigating mean level-by-dispersion interactions is the interdependence between the level and dispersion components of aggregate group-level variables. Specifically, in testing for a level-by-dispersion interaction, the degree of interdependence between the level and degree components of a proposed interaction may mask a higher order exponential term that stems from one or both predictor variables having a curvilinear relationship with study criteria (e.g., Cortina, 1993; Lubinski & Humphreys, 1990). Because of this possibility, Step 6 involves accounting for potential nonlinear relationships by using regression techniques to determine whether higher order effects (i.e., squared terms) are present. In instances where there is insufficient theory to predict the specific form of a potential curvilinear relationship, multilevel researchers may have no choice but to note that they are proceeding in a post hoc or exploratory fashion, with the need to cross-validate results in a follow-up study.

Two Illustrative Examples

The following examples incorporate real-world data on established constructs that exist at the group (viz., group cohesion) and manager (viz., active management-by-exception, or MBEA) levels of analysis. Various theory, measurement, and design issues related to Steps 1 to 3 are generally well known, and their implications have been discussed here and in prior works. We thus refer readers who desire more detail to the articles cited in Table 1. Given our interest in highlighting the statistical challenges (and inherent constraints) of using group means and variances as predictors of study criteria, we focus on less appreciated data-analytic issues (i.e., Steps 4-6) related to establishing the statistical dependence or deterministic relatedness of the level and dispersion components of an aggregate group-level variable. Whereas both examples incorporate a single-unit multilevel model that we test using hierarchical (ordinary least squares) regression, the same data-analytic framework would likewise apply to other multilevel applications (e.g., cross-level models).

Example 1: Group Cohesion and Negative Group-Affective Tone

As the association between group cohesion and group-affective outcomes is widely acknowledged (George, 2002; Spoor & Kelly, 2004; Terry et al., 2000), it was selected as the basis for an initial illustration of the proposed framework (Table 1). *Group cohesion* is the tendency for group members to "stick together and remain united in the pursuit of its instrumental objectives and/or for the satisfaction of members' affective needs" (Carron & Brawley, 2000, p. 94). It is among the most thoroughly documented group-level phenomena (Mathieu et al., 2008) and is typically operationalized by multilevel researchers using direct consensus or referent-shift composition models (Hirschfeld & Bernerth, 2008; Quigley et al., 2007). Following George (1990), *negative group-affective tone* may be defined as "consistent or homogenous [negative] affective reactions within a group" (p. 108). In the current context, *cohesion level* may be taken to refer to group members' aggregated responses and *cohesion dispersion* as variation or lack of within-group agreement among individuals' cohesion judgments (Chan, 1998). Thus, whereas cohesion level and negative group-affective tone are operationalized using a referent-shift consensus model, cohesion dispersion is instantiated by applying a dispersion-composition model.

Study design. Data for this example come from 1,000 members of 79 functional work groups of an industrial manufacturing company operating within the global marketplace. We administered

Web-based surveys to group members at two points in time. At Time 1, we collected individual group members' responses to a cohesion measure. At Time 2 ($t_{1 + 7 \text{ months}}$), we collected individual group members' responses to a measure assessing negative affective tone. Group sizes were similar across both survey administrations, Time 1 ($M = 16.5$, $SD = 13.3$) and Time 2 ($M = 16.5$, $SD = 12.7$).

Measures. We assessed cohesion level ($\alpha_{\text{group}} = .93$) with four items from Riordan and Weatherly's (1999) work group cohesiveness measure (1 = *strongly disagree;* 5 = *strongly agree*). We assessed negative group-affective tone using 10 negatively valenced emotions (1 = *never;* 5 = *extremely often or always*; Van Katwyk, Fox, Spector, & Kelloway, 2000). An analysis of variance (ANOVA) test using "group" as the independent factor demonstrated that member ratings on cohesion and negative group-affective tone differed significantly ($p < .05$) across groups, indicating non-independence associated with group membership (Bliese, 2000, pp. 357-358). Following Bliese (2000) and Chen, Mathieu, and Bliese (2004), both of whom suggest that in a multilevel setting such a statistically significant ANOVA F statistic implies that group membership affects individual members' observations, we thus deemed it appropriate to aggregate individual ratings to the group level. Intraclass correlation coefficient, ICC(1), values indicated that there was a small to medium group-level effect on cohesion (.02) and negative group-affective tone (.09; LeBreton & Senter, 2008). As we discuss in a later section, ICC(1) values of this modest, though sufficient, magnitude were not entirely unexpected given that dispersion-composition assumes there will be individual differences within groups (Chan, 1998).

Selection of a dispersion index. A number of indices have been used to measure dispersion constructs (e.g., r_{wg}, r^*_{wg}, a_{wg}, and standard deviation). Roberson et al. (2007) have reported that despite being very similar, the aforementioned indices performed differently when used as predictor variables. They found that the standard deviation is one of the most effective indices for assessing within-group dispersion. Accordingly, cohesion dispersion scores were based on the square root of the variance for each set of team scores. Finally, variations in group sizes did not affect the inferences drawn with respect to the ensuing analyses (cf. Biemann & Kearney, 2010), and all available data were used.

Dispersion as a direct predictor of negative group-affective tone. To reiterate, we emphasize Steps 4 through 6 because these steps deal with key data-analytic issues rarely reported in the relevant literature. This it is not meant, however, to minimize theoretical issues associated with establishing the underlying nature of a construct when indexed in an alternative form (i.e., mean or variance). As we have stressed, the development of appropriate theory that describes how, when, and why a dispersion construct might influence study criteria is an essential first step when aggregating data from a lower level to a higher level of analysis.

To begin (Step 4, Table 1), we explored the statistical dependence between the absolute mean level and dispersion of cohesion by inspecting the association between cohesion level and dispersion scores (see Table 2). The correlation $r = -.70$ between these variables suggested a degree of range restriction among work groups with the lowest possible mean values on cohesion. This result is not atypical; Dineen et al. (2007) likewise reported that the absolute mean level and dispersion of team satisfaction were related in an ∩-shaped fashion.

Moving to Step 5 in our data-analytic framework, because the components of level and dispersion are statistically related, we regressed negative group-affective tone on cohesion dispersion, controlling for cohesion level (Lindell & Brandt, 2000). Inspecting the Model 1 main effects reported in Table 3, cohesion dispersion was not associated with negative group-affective tone ($B = 0.01$, $p > .05$, $R^2 = .00$) when controlling for cohesion level. As noted, however, some degree of range restriction was observed (as indicated by the correlation $r = -.70$ between cohesion level and

Table 2. Descriptive Statistics and Intercorrelations Among Study Variables

Variable	M	SD	1	2	3
				r	
1. Cohesion level	3.86	0.32	–		
2. Cohesion dispersion	1.00	0.59	–.70**	–	
3. Negative group-affective tone	2.36	0.28	–.05	.04	–

Note: $n = 79$ work groups. As cohesion dispersion is measured by its standard deviation, smaller scores imply less dispersion.
** $p < .01$.

Table 3. Hierarchical Multiple Regression for Group Cohesion

Variable entry order	Negative Group-Affective Tone			
	Model 1	Model 2	Model 3	Model 4
Direct effects				
Cohesion level	–0.03	–0.03	–0.00	0.03
Cohesion dispersion	0.01	0.14	0.15	0.16
R^2	.00			
Curvilinear terms				
Cohesion level squared			–0.65**	–1.16*
Cohesion dispersion squared			–0.05	–0.15
ΔR^2			.15	
Two-way interaction				
Cohesion Level × Cohesion Dispersion		0.38**		–0.48
ΔR^2		.10		.02

Note: $n = 79$ work groups. As cohesion dispersion is measured by its standard deviation, smaller scores imply less dispersion. Unstandardized regression coefficients (B) are shown. All tests are two-tailed. For Model 2, ΔR^2 indicates the change in variance explained by Model 2 relative to that by Model 1. For Model 4, ΔR^2 indicates the change in variance explained by Model 4 relative to that by Model 3.
* $p < .05$.
** $p < .01$.

dispersion scores, as reported in Table 2). Given this range restriction, unique variance in negative group-affective tone attributable to cohesion dispersion may be statistically inaccessible (Lindell & Brandt, 2000).

Interaction of Level × Dispersion on negative group-affective tone. Continuing with Step 6, we explored the Cohesion Level × Cohesion Dispersion interaction. As indicated by Model 2 in Table 3, the interaction between cohesion level and cohesion dispersion ($B = 0.38$, $p < .01$, $\Delta R^2 = .10$) was significant in predicting negative group-affective tone. These findings suggest a cohesion level–by–cohesion dispersion interaction. Nevertheless, Step 6 of our framework recommends an examination of higher order (i.e., squared) terms corresponding to each component of a two-way interaction. This is an important consideration because cohesion level and cohesion dispersion are statistically interdependent and, thus, "their joint effects on [study criteria] do not necessarily take the form of simple linear relations" (Lindell & Brandt, 2000, p. 336). In other words, a linear-by-linear interaction may actually result from one or both predictor variables having a nonlinear relationship with group-affective tone. Step 6, thus, examines this possibility when a significant interaction is present.

As indicated by Model 3 in Table 3, when both squared terms are included in the regression model, the cohesion level squared term was associated ($B = –0.65$, $p < .01$, $\Delta R^2 = .15$) with negative

group-affective tone. Upon entering the Cohesion Level $\times$ Cohesion Dispersion cross-product term in a final hierarchical step (Model 4), the cohesion dispersion squared term continued to be associated ($B = -1.16$, $p < .05$) with negative group-affective tone, but the two-way interaction was not significantly related ($B = 0.48$, $p > .05$, $\Delta R^2 = .02$) to the study criterion when the squared terms were considered. We were, therefore, able to conclude that the previously observed two-way interaction was, in fact, a spurious by-product of a curvilinear effect associated with a cross-product component term. Had we not accounted for the possibility of statistical dependence (through the curvilinear terms), we would have misinterpreted the two-way interaction between cohesion level and cohesion dispersion on negative group-affective tone. Thus, when an interaction between the mean level and dispersion of a group-level construct is of interest, the possibility of nonlinear relationships must be considered by examining higher order terms.

Example 2: Leadership Behavior and Team Performance

The second example we selected to illustrate Steps 4 through 6 of our data-analytic framework (Table 1) considers the relationship between managers' leadership behavior (viz., MBEA) and their teams' performance. In general, management-by-exception is the degree to which a leader "takes corrective action on the basis of results of leader–follower interactions" (Judge & Piccolo, 2004, p. 755). In terms of MBEA, a manager (as a leader) aggressively monitors subordinates' behavior, anticipates problems, and takes corrective actions before the behavior creates serious difficulties. Whereas *MBEA level* refers to individual members' shared responses about their managers' MBEA behavior (i.e., a direct consensus model), *MBEA dispersion* references the differences among subordinates' impressions (lateral position on a response scale) of their managers' MBEA behavior.

Study design. Data for this example were collected from the managers ($n = 75$) and subordinates ($n = 317$) of 75 work teams employed by a manufacturing company headquartered in Germany. Supervisors rated their teams' overall performance; no supervisor provided more than one team performance rating. Subordinates (i.e., team members), on the other hand, were asked to judge their respective managers' leadership behavior (i.e., a direct-consensus model). On average, five team members ($SD = 2.2$) completed the subordinate survey. By administering different survey items to managers and their subordinates, we reduced concerns associated with common method variance.

Measures. Team members were asked to indicate, using a 5-point frequency scale ranging from 1 (*not at all*) to 5 (*frequently, if not always*), how often their respective managers exhibited four behaviors that tap MBEA ($\alpha_{manager} = .84$). This measure was taken directly from the Multifactor Leadership Questionnaire Form 5X (Bass & Avolio, 2000). An ANOVA using "manager" as the independent factor demonstrated that member ratings on MBEA differed significantly ($p < .01$) across managers, suggesting substantive clustering in the data (Bliese, 2000, pp. 357-358). Accordingly, we aggregated member ratings to the manager level of analysis. The ICC(1) value of .20 for MBEA indicated an appreciable group-level effect (LeBreton & Senter, 2008). Finally, managers provided global *team performance* ($\alpha = .85$) ratings using a 5-item measure (1 = *strongly disagree* to 5 = *strongly agree;* Conger, Kanungo, & Menon, 2000).

Selection of a dispersion index. As in our initial example, we computed *leadership dispersion* based on the average squared distance (i.e., standard deviation) of a set of team ratings from its mean (Roberson et al., 2007). Variations in group sizes did not affect the inferences drawn, and all available data were used.

Dispersion as a direct predictor of team performance. Table 4 presents descriptive statistics and intercorrelations for all study variables. An inspection of the coefficients suggests a positive correlation

Table 4. Descriptive Statistics and Intercorrelations Among Study Variables

				r	
Variable	M	SD	1	2	3
1. MBEA level	2.92	0.64	–		
2. MBEA dispersion	0.74	0.34	.07	–	
3. Team performance	3.74	0.59	–.14	.25*	–

Note: $n = 75$ work teams. MBEA = active management-by-exception. As dispersion is measured by its standard deviation, smaller scores imply less dispersion.
* $p < .05$.

between MBEA dispersion and team performance ($r = .25$, $p < .05$); however, this implied association should be interpreted cautiously given that the absolute mean associated with MBEA level has yet to be taken into account.

We proceeded (with Step 4, Table 1) to investigate the possibility of statistical dependence between the absolute level and dispersion of MBEA. The near-zero correlation ($r = .07$; Table 4) suggests minimal restriction of range and that the functional relationship between level and dispersion is not ∩-shaped (i.e., teams' responses are equally distributed; Lindell & Brandt, 2000). We then moved on to Step 5 by regressing team performance on both MBEA level and MBEA dispersion. As indicated in Model 1 of Table 5, MBEA dispersion was associated with team performance ($B = 0.46$, $p < .05$, $R^2 = .09$) when MBEA level was included as a predictor.

Interaction of Level × Dispersion on team performance. Continuing with Step 6 in our data-analytic framework, we tested for possible interaction and curvilinear effects (Step 6, Table 1). We first examined the level-by-dispersion interaction, while taking into account the direct effects. As indicated by Model 2 in Table 5, the interaction between MBEA level and MBEA dispersion was significant ($B = 0.87$, $p = .01$, $\Delta R^2 = .08$) in predicting team performance. We then explored whether curvilinear relationships were present: Model 3 in Table 5 indicates that neither of the corresponding curvilinear terms was significantly related to team performance ($B = 0.09$ and $B = 0.22$; both $p > .05$). Finally, as also shown in Table 5 (Model 4), the MBEA Level × MBEA Dispersion cross-product was significantly related to team performance ($B = 0.89$, $p < .05$, $\Delta R^2 = .08$) when the curvilinear terms were included in the regression model. Thus, because we have ruled out the presence of a curvilinear effect, we are more confident in the veracity of the two-way interaction between MBEA level and MBEA dispersion in predicting team performance.

Discussion

A familiar issue tackled by multilevel researchers is the measurement of collective phenomena that emerge from bottom-up processes. Researchers have typically used a mean-based approach to aggregating individual members' responses to a higher level of analysis, according to either a direct consensus or a referent-shift consensus composition model (Chan, 1998). In doing so, tests of IRA and between-group differences have been customarily computed to determine whether estimating aggregate scores from individual-level data is empirically justifiable; unexplained variance has been regarded as noise or measurement error (Kozlowski & Klein, 2000). In contrast to this predominant view, it has been increasingly argued that unexplained variance among group members' assessments may represent a meaningful higher level construct and not just reflect error variance (Harrison & Klein, 2007; Mathieu et al., 2008). Such dispersion-composition models, however, have been seldom used in multilevel research.

Table 5. Hierarchical Multiple Regression for Active Management-by-Exception (MBEA)

	Team Performance			
Variable entry order	Model 1	Model 2	Model 3	Model 4
Direct effects				
MBEA level	−0.15	−0.08	−0.13	−0.08
MBEA dispersion	0.46*	0.55**	0.45*	0.52**
R^2	.09			
Curvilinear terms				
MBEA level squared			0.09	−0.03
MBEA dispersion squared			0.22	0.19
ΔR^2			.01	
Two-way interaction				
MBEA Level × MBEA Dispersion		0.87**		0.89*
ΔR^2		.08		.08

Note: $n = 75$ work teams. As MBEA dispersion is measured by its standard deviation, smaller scores imply less dispersion. Unstandardized regression coefficients (*B*) are shown. All tests are two-tailed. For Model 2, ΔR^2 indicates the change in variance explained by Model 2 relative to that by Model 1. For Model 4, ΔR^2 indicates the change in variance explained by Model 4 relative to that by Model 3.
* $p < .05$.
** $p < .01$.

Meade and Eby (2007) argue that dispersion models have a meaningful role to play in specifying functional relationships among phenomena, and the authors encourage multilevel researchers to incorporate dispersion models in their analyses. As an increasing number of multilevel researchers take up this charge, we believe that an easy-to-follow framework and modeling sequence will be helpful, if not essential, for comparing and building upon published results. Toward this end, we have highlighted various issues and judgment calls that may confront multilevel researchers.

Further Methodological Issues

Whereas our purpose was to highlight statistical challenges (and inherent constraints) of using group means and variances as predictors of study criteria, we suspect that multilevel researchers will confront additional methodological issues surrounding dispersion-composition measurement. We identify what we believe are a few of the more salient issues or challenges, acknowledging that in doing so we have not exhausted the full range of possibilities.

Appropriate grouping variable. When a research question involves multilevel phenomena, one must designate a focal unit of analysis. Thus, along with Mathieu and Chen (2011), we believe the process by which researchers identify an appropriate unit of analysis is oftentimes underdiscussed in published reports. As Mathieu and Chen further note, researchers should ''devote far more attention to the identification of focal units of analysis'' (p. 15) because unit designation ''becomes the hinge variable for the estimation of within-group agreement or variance, as well as the focal point for estimating interrater reliability and between-group variance'' (p. 13). Given the centrality of within-group agreement when evoking consensus and dispersion-composition models, we maintain that specifying an appropriate ''grouping variable'' is an important challenge that multilevel researchers must explicitly tackle if they are to advance an understanding of multilevel phenomena.

Subgroups may exist. Beyond identifying an appropriate grouping variable, there is a relative dearth of conceptual and methodological treatments regarding the existence of possible subgroups

within a unit of analysis. Organizational units are seldom as neatly nested as is implicitly assumed (Mathieu & Chen, 2011). For example, Gooty and Yammarino (2011) have shown how dyadic theorizing and testing can be used to complement and, thus, expand the ever-so-popular teams-based perspective. They present a thoughtful discussion of the multilevel implications associated with neglecting the fact that lower level units (i.e., individuals) are nested not only within groups or teams but also within dyadic configurations (i.e., subgroup nestings within teams). Further, Gooty and Yammarino illustrate the application of random coefficient modeling and WABA techniques for testing various multilevel questions involving individuals nested in dyads that in turn are nested in teams. Moreover, the challenge of subcultures or coalitions within groups is not limited to formal organizational units. Newman, Hanges, Duan, and Ramesh (2008) have explored how social processes may lead to the emergence of distinct subgroup climates within a firm's organization-level climate.

Missing data implications. Multilevel researchers are only beginning to understand how missing data may influence the precision with which the relations among *level* and *dispersion* constructs and study criteria may be estimated. Nonetheless, what we do know (Newman & Sin, 2009) is that the estimation of group-level properties is complicated when within-group response rates fall below 100% (a very common occurrence in field research). Across two studies (Allen, Stanley, Williams, & Ross, 2007; Timmerman, 2005), it has been shown that random and not-at-random missing data created within-group range restriction in both level (i.e., absolute means based on aggregated data) and dispersion scores (based on standard deviation), and thus, significant "true" relationships with criteria were reduced in magnitude (i.e., attenuated). Newman and Sin (2009) extended these findings, showing that (a) missing data bias estimates based on within-group dispersion (SD_{wg}) do not depend on team size and (b) direction of the effect sizes representing the strength of systematic missingness is irrelevant. This latter finding suggests to the extent that dissatisfied team members are less likely to respond to a survey, the SD_{wg} estimate for the team will be underestimated to the same degree as if they were nonresponding members of a team who are satisfied. Taken together, Timmerman (2005), Allen et al. (2007), and Newman and Sin (2009) suggest that the magnitudes of observed relationships between study criteria and dispersion variables (and Level × Dispersion interactions) are generally *underestimated*. Thus, future research that relies on surveys to collect data on level and dispersion constructs should be cognizant of the potential impact that missing data (random and systematic) can have on parameter estimates.

Statistical justification and the aggregation process. As has been discussed here and elsewhere (e.g., Harrison & Klein, 2007; Lindell & Brandt, 2000), dispersion and consensus composition models are oftentimes intimately intertwined. This connection introduces a unique challenge when testing both forms of composition (i.e., a level-by-dispersion interaction). In applying consensus composition models, multilevel researchers have relied on minimum IRA (e.g., r_{wg}) and IRR—ICC(1) and ICC(2)—estimates (i.e., "hurdle rates") to evidence construct validity at the group level (Chen et al., 2004; LeBreton & Senter, 2008). In doing so, researchers are implicitly considering within-group variability to be error variance. In contrast, dispersion-composition considers within-group variance as a focal construct in its own right. At present, the appropriate treatment of within-group variance both as error variance and as a theoretically significant phenomenon—within a single study—remains an open question, as there exists little guidance on computing IRA and IRR estimates for a group-level construct in a model containing its dispersion-variable counterpart.

Building on Bliese (2000) and Chen et al. (2004), one might suggest that a nonzero ICC(1) provides sufficient (albeit a low hurdle rate) evidence of group effects. Consistent with this suggestion, and as described in both of the foregoing empirical illustrations, we argued that a statistically significant ANOVA *F* statistic indicates that group membership is affecting members'

ratings (i.e., there is a nesting structure). We thus concluded it was "justified" to aggregate individuals' ratings to create our group-level, mean-based constructs. In addition, we recommend that multilevel researchers continue to report ICC(1) values associated with a consensus-based construct (or constructs), as ICC(1) may be interpreted as an effect size estimate of the extent to which individuals' ratings are attributable to group membership (Bliese, 2000). Accordingly, LeBreton and Senter (2008) have recommended that "traditional conventions" be adopted when interpreting ICC(1) as an effect size estimate: "small effect" = .01, "medium effect" = .10, and "large effect" = .25 (p. 838). In this regard, Bliese (1998) has shown that even where only 1% of the variance in a focal construct is attributable to group membership, ICC(1) = .01, strong group-level relationships may still be present.

We likewise suggest caution in following established conventional hurdle rates in interpreting ICC(2) estimates (which assess the reliability of unit-level means; Bliese, 2000) because a dispersion-composition model implicitly assumes that a collective phenomenon has yet to solidify in some (or all) of a study's focal units. Finally, we add that commonly applied thresholds for interpreting within-group agreement (e.g., r_{wg}-based) estimates are less important for empirical studies incorporating dispersion-composition models. Lindell and Brandt (2000) have shown that r_{wg} and standard deviation as a measure of spread (i.e., used to index our dispersion variables) are perfectly nonlinearly related (see also Roberson et al., 2007) in research applications and, thus, provide redundant information.

If researchers follow convention and, thus, choose to establish aggregation hurdles, they may wish to set a low threshold because an underlying assumption in testing level-by-dispersion interactions is that there will be groups with low, moderate, and high levels of relative consistency and absolute consensus vis-à-vis study variables (see LeBreton, Burgess, Kaiser, Atchley, & James, 2003; LeBreton & Senter, 2008, for detailed reviews). Further complicating matters, Beal and Dawson (2007) have shown that ICC estimates are underestimated when using Likert-type response formats. Given these concerns, either reporting confidence intervals (automatically reported in statistical programs such as Stata) in conjunction with ICC values or using resampling procedures to bootstrap ICC estimates may be appropriate. As LeBreton and Senter (2008) have noted, these are "tough decisions" that, for the present, must be guided by theory and informed judgment (p. 838).

Conclusion

The failure of researchers to articulate adequate composition models that fully specify the functional relationships among focal constructs has hampered multilevel research. As shown in our two examples, consideration of the level and dispersion of phenomena may offer insights that are of theoretical importance and relevant to practice. In considering whether dispersion variables are viable as an alternative application, researchers will strengthen a study's conclusions by eliminating potential rival data interpretations. Indeed, we believe that the application of dispersion-composition models will both shed new light on past findings and open new doors to a more complete understanding of all multilevel phenomena. We, thus, encourage researchers who conduct multilevel studies to explore alternative composition models by carefully examining the components underlying aggregate-level constructs.

Declaration of Conflicting Interests

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

Funding

Notes

1. Although there are potential differences between work groups and teams (Chan, 1998, p. 235), for simplicity, we view them similarly, as a clustering of individuals who are interdependent, based on a set of common expectations or hierarchical structuring, and who interact with one another as if they are a group. As such, we use the terms *group, unit,* and *team* interchangeably.

References

Allen, N. J., Stanley, D. J., Williams, H. M., & Ross, S. J. (2007). Assessing the impact of nonresponse on work group diversity effects. *Organizational Research Methods*, *10*, 262-286.

Bass, B. M., & Avolio, B. J. (2000). *Manual for the Multifactor Leadership Questionnaire (Form 5X)*. Redwood City, CA: Mindgarden.

Beal, D. J., & Dawson, J. F. (2007). On the use of Likert-type scales in multilevel data: Influence on aggregate variables. *Organizational Research Methods*, *10*, 657-672.

Biemann, T., & Kearney, E. (2010). Size does matter: How varying group sizes in a sample affect the most common measures of group diversity. *Organizational Research Methods*, *13*, 582-599.

Bliese, P. D. (1998). Group size, ICC values, and group-level correlations: A simulation. *Organizational Research Methods*, *1*, 355-373.

Bliese, P. D. (2000). Within group agreement, non-independence and reliability: Implications for data 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., & Britt, T. W. (2001). Social support, group consensus and stressor–strain relationships: Social context matters. *Journal of Organizational Behavior*, *22*, 425-436.

Bliese, P. D., & Halverson, R. R. (1998). Group consensus and psychological well-being: A large field study. *Journal of Applied Social Psychology*, *28*, 563-580.

Bliese, P. D., & Hanges, P. J. (2004). Being both to liberal and too conservative: The perils of treating grouped data as though they were independent. *Organizational Research Methods*, *5*, 362-387.

Carron, A. V., & Brawley, L. R. (2000). Cohesion: Conceptual and measurement issues. *Small Group Research*, *31*, 89-106.

Carron, A. V., Brawley, L. R., Eys, M. A., Bray, S. R., Dorsch, K. D., Estabrooks, P. A., et al. (2003). Do individual perceptions of group cohesion reflect shared beliefs? An empirical analysis. *Small Group Research*, *34*, 468-496.

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

Chen, G., Mathieu, J. E., & Bliese, P. D. (2004). A framework for conducting multi-level construct validation. *Research in Multi-Level Issues*, *3*, 273-303.

Colquitt, J. A., Noe, R. A., & Jackson, C. L. (2002). Justice in teams: Antecedents and consequences of procedural justice climate. *Personnel Psychology*, *55*, 83-109.

Conger, J. A., Kanungo, R. N., & Menon, S. T. (2000). Charismatic leadership and follower effects. *Journal of Organizational Behavior*, *21*, 747-767.

Cortina, J. M. (1993). Interaction, nonlinearity, and multicollinearity: Implications for multiple regression. *Journal of Management*, *19*, 915-922.

Dansereau, F., Cho, J., & Yammarino, F. J. (2006). Avoiding the "fallacy of the wrong level": A within and between analysis (WABA) approach. *Group and Organization Management*, *31*, 536-577.

Dansereau, F., & Yammarino, F. J. (2000). Within and between analysis: The varient paradigm as an underlying approach to theory building and testing. In K. J. Klein & S. W. J. Kozlowski (Eds.), *Multilevel theory,*

research, and methods in organizations: Foundations, extensions, and new directions (pp. 425-466). San Francisco, CA: Jossey-Bass.

DeRue, D. S., Hollenbeck, J., Ilgen, D., & Feltz, D. (2010). Efficacy dispersion in teams: Moving beyond agreement and aggregation. *Personnel Psychology*, *63*, 1-40.

Dineen, B. R., Noe, R. A., Shaw, J. D., Duffy, M. K., & Wiethoff, C. (2007). Level and dispersion of satisfaction in teams: Using foci and social context to explain the satisfaction and absenteeism relationship. *Academy of Management Journal*, *50*, 623-643.

George, J. M. (1990). Personality, affect, and behavior in groups. *Journal of Applied Psychology*, *75*, 107-116.

George, J. M. (2002). Affect regulation in groups and teams. In R. G. Lord, R. J. Klimoski, & R. Kanfer (Eds.), *Emotion in the workplace: Understanding the structure and role of emotions in organizational behavior* (pp. 183-217). San Francisco, CA: Jossey-Bass.

George, J. M., & James, L. R. (1993). Personality, affect, and behavior in groups revisited: Comment on aggregation, levels of analysis, and a recent application of within and between analysis. *Journal of Applied Psychology*, *78*, 798-804.

Gooty, J., & Yammarino, F. J. (2011). Dyads in organizational research: Conceptual issues and multilevel analyses. *Organizational Research Methods, 14*, 456-483.

Harrison, D. A., & Klein, K. J. (2007). What's the difference? Diversity constructs as separation, variety, or disparity in organizations. *Academy of Management Review*, *32*, 1199-1228.

Hirschfeld, R. R., & Bernerth, J. B. (2008). Mental efficacy and physical efficacy at the team level: Inputs and outcomes among newly formed action teams. *Journal of Applied Psychology*, *93*, 1429-1437.

Judge, T. A., & Piccolo, R. F. (2004). Transformational and transactional leadership: A meta-analytic test of their relative validity. *Journal of Applied Psychology*, *89*, 755-768.

Klein, K. J., Conn, A. B., Smith, D. B., & Sorra, J. S. (2001). Is everyone in agreement? An exploration of within-group agreement in employee perceptions of the work environment. *Journal of Applied Psychology*, *86*, 3-16.

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.

Lance, C. E., Butts, M. M., & Michels, L. C. (2006). The sources of four commonly reported cutoff criteria. What did they really say? *Organizational Research Methods*, *9*, 202-220.

LeBreton, J. M., Burgess, J. R. D., Kaiser, R. B., Atchley, E. K., & James, L. R. (2003). The restriction of variance hypothesis and interrater reliability and agreement: Are ratings from multiple sources really dissimilar? *Organizational Research Methods*, *6*, 80-128.

LeBreton, J. M., & Senter, J. L. (2008). Answers to 20 questions about interrater reliability and interrater agreement. *Organizational Research Methods*, *11*, 815-852.

Lindell, M. K., & Brandt, C. J. (2000). Climate quality and climate consensus as mediators of the relationship between organizational antecedents and outcomes. *Journal of Applied Psychology*, *85*, 331-348.

Lubinski, D., & Humphreys, L. G. (1990). Assessing spurious "moderator effects": Illustrated substantively with the hypothesized ("synergistic") relation between special and mathematical ability. *Psychological Bulletin*, *107*, 385-393.

Mathieu, J. E., & Chen, G. (in 2011). The etiology of the multilevel paradigm in management research. *Journal of Management*, *37*, 610-641.

Mathieu, J. E., Maynard, M. T., Rapp, T., & Gilson, L. (2008). Team effectiveness 1997-2007: A review of recent advancements and a glimpse into the future. *Journal of Management*, *34*, 410-476.

Meade, A. W., & Eby, L. T. (2007). Using indices of group agreement in multilevel construct validation. *Organizational Research Methods*, *10*, 75-96.

Mossholder, K. W., & Bedeian, A. G. (1983). Cross-level inference and organizational research: Perspectives on interpretation and application. *Academy of Management Review*, *8*, 547-558.

Naumann, S. E., & Bennett, N. (2002). The effects of procedural justice climate on work group performance. *Small Group Research*, *33*, 361-377.

Newman, D. A., Hanges, P. J., Duan, L., & Ramesh, A. (2008). A network model of organizational climate: Friendship, clusters, subgroup agreement, and climate schemas. In D. B. Smith (Ed.), *The people make the place: Dynamic linkages between individuals and organizations* (pp. 101-126). Mahwah, NJ: Lawrence Erlbaum.

Newman, D. A., & Sin, H. P. (2009). How do missing data bias estimates of within-group agreement? Sensitivity to SD_{wg}, CV_{wg}, $r_{wg(j)}$, $r_{wg(j)}*$, and ICC to systematic nonresponse. *Organizational Research Methods, 12,* 113-147.

Quigley, N., Tekleab, A. G., & Tesluk, P. (2007). Comparing consensus- and aggregation-based methods of measuring team-level variables: The role of relationship conflict and conflict management processes. *Organizational Research Methods*, *10*, 589-608.

Riordan, C. E., & Weatherly, E. W. (1999). Defining and measuring employees' identification with their work groups. *Educational and Psychological Measurement*, *59*, 310-324.

Roberson, Q. M., Sturman, M. C., & Simons, T. L. (2007). Does the measure of dispersion matter in multilevel research? A comparison of the relative performance of dispersion indexes. *Organizational Research Methods*, *9*, 564-588.

Sin, H.-P., & Newman, D. A. (2005). Variance of means versus mean of variances: A contrarian view on operationalizing group dispersion. In K. M. Weaver (Ed.), *Proceedings of the Sixty-Fourth Annual Meeting of the Academy of Management. Academy of Management Best Paper Proceedings.* Briarcliff Manor, NY: Academy of Management.

Spoor, J. R., & Kelly, J. R. (2004). The evolutionary significance of affect in groups: Communication and group bonding. *Group Processes and Intergroup Relations*, *7*, 398-412.

Terry, P. C., Carron, A. V., Pink, M. J., Lane, A. M., Jones, G., & Hall, M. (2000). Perceptions of group cohesion and mood in sport teams. *Group Dynamics*, *4*, 234-243.

Timmerman, T. A. (2005). Missing persons in the study of groups. *Journal of Organizational Behavior*, *26*, 21-36.

Van Katwyk, P. T., Fox, S., Spector, P. E., & Kelloway, E. K. (2000). Using the Job-Related Affective Well-Being Scale (JAWS) to investigate affective responses to work stressors. *Journal of Occupational Health Psychology*, *5*, 219-230.

van Mierlo, H., Vermunt, J. K., & Rutte, C. G. (2009). Composing group-level constructs from individual-level survey data. *Organizational Research Methods*, *12*, 368-392.

Bios

Michael S. Cole is an assistant professor in the Department of Management at the M. J. Neeley School of Business at Texas Christian University. His professional interests focus on multilevel theories, research, and methods as they relate to behavior in organizations.

Arthur G. Bedeian is a Boyd Professor at Louisiana State University and A&M College. A past-president of the Academy of Management, a former dean of the Academy's Fellows Group, and a recipient of the Academy's Distinguished Service Award, he is a fellow of the American Psychological Association and the International Academy of Management.

Robert R. Hirschfeld is a graduate of Tulane University with a PhD in management from Auburn University. He is currently an assistant professor in the Department of Management and Quantitative Methods at the University of Colorado at Colorado Springs. His research involves issues of motivation, development, and performance within and among individuals and teams.

Bernd Vogel is an assistant professor at the Henley Business School at the University of Reading, United Kingdom. His research involves issues of leadership, followership, organizational energy, and emotions.