

Change the Referent? A Meta-Analytic Investigation of Direct and Referent-Shift Consensus Models for Organizational Climate

J. Craig Wallace

Oklahoma State University

Bryan D. Edwards

Oklahoma State University

Jeff Paul

Illinois State University

Michael Burke

Tulane University

Michael Christian

University of North Carolina–Chapel Hill

Gabi Eissa

Oklahoma State University

Based on earlier taxonomies of group composition models, aggregating data from individual-level responses to operationalize group-level constructs is a common aspect of management research. The present study contributes to the literature on group composition models by quantitatively integrating the climate literature via meta-analysis to determine which of the two most common methods of aggregation, direct consensus and referent-shift consensus, is the stronger predictor of group-level outcomes. We found that referent-shift consensus was a stronger predictor of job performance and customer service performance than direct consensus. However, we found that direct consensus was a stronger predictor of job attitudes than referent-shift consensus. We also found that climate-performance relationships were moderated by aggregation method of the performance criterion. The implications of these findings for advancing multi-level theory and research are discussed.

Acknowledgment: This article was accepted under the editorship of Deborah E. Rupp.

Corresponding author: J. Craig Wallace, Watson Graduate School of Management, Oklahoma State University, 700 N Greenwood, 306 North Hall, Tulsa, OK 74106, USA.

E-mail: craig.wallace@okstate.edu or craig@theAOEgroup.com

Keywords: *culture and climate; hierarchical linear modeling (HLM)/multi-level; groups/group processes/dynamics; teams*

Multi-level modeling techniques have refined the precision with which researchers might predict behaviors in organizations, yet these techniques have also opened the door to a myriad of questions concerning conceptualization, operationalization, and measurement of constructs across organizational levels. The operationalization of group-level constructs typically derives from a specified composition model (e.g., Chan, 1998; Kozlowski & Klein, 2000). Composition models describe how constructs operationalized at one level of analysis are related to other forms of the constructs at different levels of analysis. The specific composition model (or functional relations among constructs at different levels) is typically derived from theory and operationalized as some combination (e.g., aggregation) of the lower-level units. In addition to providing a theoretical rationale for the chosen composition model, researchers also provide empirical evidence to support the aggregation of lower-level units to produce a meaningful indicator of the higher-level construct. The specification of the composition model and method of composition are crucial for operationalizing higher-order constructs. Thus, it is standard for researchers to provide these two pieces of information to justify how a given construct is conceptualized and operationalized, when its measurement is based on data aggregated from lower-level units (e.g., individual perceptions) to higher levels (e.g., team/group).

Chan's (1998) typology of composition models was the first typology to guide researchers on methods of aggregation, and Chan used climate as his exemplar. This typology served to coordinate the efforts of multi-level researchers to use comparable operationalizations of group-level constructs and aggregation statistics to justify composition models. As a result, a sizeable amount of research can now be synthesized via meta-analysis to make meaningful comparisons across composition models used to operationalize work climates.

The most commonly used composition models in multi-level research, particularly in climate research, have been direct consensus and referent-shift consensus (van Mierlo, Vermunt, & Rutte, 2009). Direct consensus typically aggregates survey items that begin with an individual perception (e.g., "I believe ...") whereas referent-shift consensus typically aggregates survey items that reflect an individual's perception of some higher-level structure (e.g., "My team believes ..."). Although both have been used to operationalize work climates, it is not clear which composition method is most appropriate in this domain. Researchers have argued that using the appropriate referent at a given level of analysis would predict same-level outcomes better (Kozlowski & Klein, 2000; Mathieu & Chen, 2011; Rousseau, 1985). However, in climate research some authors leave the referent at the individual level whereas others changed the referent to reflect the higher-level structure. In some instances, the specific aggregation was driven by the most convenient data collection procedure, with little regard to bandwidth or theoretical relevance. However, if the level of measurement differs from the level of analysis, one needs to justify the aggregation strategy (Mathieu & Chen, 2011). In their review of the work climate literature, Kuenzi and Schminke (2009) highlighted that the work climate literature has not been consistent in how climate is measured such that no standard has emerged. We intend to help resolve this issue by using meta-analysis to compare climate studies that have changed the referent to a higher

organizational level and those that left the referent at the individual level prior to aggregation. Specifically, we conducted a comparative study of referent-shift consensus and direct consensus composition models for organizational climate.¹

Theoretical and Empirical Background

Composition Models

Kozlowski and Klein (2000) note, “[t]he first and foremost task in crafting a multi-level theory or study is to define, justify, and explain the level of each focal construct that constitutes the theoretical system” (p. 27). If there is a mismatch between the levels of theory, measurement, and/or statistical analysis, empirical results may not be interpretable with regard to theoretical or methodological linkages among the constructs (Klein, Conn, Smith, & Sorra, 2001; Klein, Dansereau, & Hall, 1994; Mathieu & Chen, 2011; Rousseau, 1985). In addition, many researchers select one of the composition models but fail to conceptually and/or methodologically justify their choice of composition model (van Mierlo et al., 2009). Lack of consideration of composition issues results in ambiguity that obscures interpretation of results. As such, our primary objective is to examine the referent in aggregation of individual perceptions used to operationalize the higher-level construct—using climate as an exemplar.

Direct Consensus Model

As presented by Chan (1998), the direct consensus composition model “uses within-group consensus of the lower level unit as the functional relationship to specify how the construct conceptualized and operationalized at the lower level is functionally isomorphic to another form of the construct at the higher level” (p. 237). The operationalization of constructs using the direct consensus composition model involves two defining characteristics. First, the constructs are conceptualized and defined at each level of analysis. Within work climate research, psychological climates are typically operationalized as the individuals’ responses to survey items (and generating a composite score at the individual level), and organizational (or group, team, work unit) climates are operationalized as the mean of individual responses within each group (James et al., 2008). Second, the researcher should explain a priori conditions for aggregation. The second component is the appropriate justification needed to aggregate the individual responses to the higher-level construct. Within-group agreement assessments examine the interchangeability of individual composite scores using statistical metrics (e.g., r_{wg} , ICC, AD) following recommendations provided by James, Demaree, and Wolf (1984) and Burke and Dunlap (2002). Sufficient agreement should exist in individual-level responses to demonstrate that the chosen aggregation method provides a reliable and valid group-level measure (van Mierlo et al., 2009).

Referent-Shift Consensus Model

Referent-shift consensus shares the same two defining characteristics as direct consensus (Chan, 1998). The primary difference between referent-shift consensus and direct

consensus arises from a shift in the referent from the individual to the collective. The shift in referent is accomplished by simply changing the referent in each survey item from “I” (in direct consensus) to “we” or “the group” (in referent-shift consensus—or some other indicator that refers to the higher level). Referent-shift consensus and direct consensus create indicators of two conceptually distinct constructs at the individual level of analysis. As such, a question arises as to whether aggregate data derived from one or both composition models can serve as meaningful indicators of a particular higher-level construct. Although there is some agreement in the research literature that referent-shift consensus is the most appropriate model for measuring some group-level constructs such as collective efficacy (Arthur, Bell, & Edwards, 2007; Whiteoak, Chalip, & Hort, 2004), there seems to be no consensus regarding its appropriateness for operationalizing other group constructs such as climates. There are examples of climates operationalized using referent-shift consensus (e.g., Glisson & James, 2002; Mason & Griffin, 2003) and examples using direct consensus (e.g., Wallace, Popp, & Mondore, 2006)

Comparative Primary Studies

In the years since Chan’s typology, there have been dozens of studies using either the referent-shift consensus or direct consensus models of aggregation. Because they have both been used to operationalize climates it would be helpful for researchers to know which of these is most appropriate because the specific composition model is not always evident. There seems to be broad agreement that the group consensus method of composition (i.e., all group members collectively respond to items related to group-level constructs) is likely the composition model that results in the highest validities with performance as the criterion (i.e., Gibson, Randel, & Earley, 2000; Kirkman, Tesluk, & Rosen, 2001; Quigley, Tekleab, & Tesluk, 2007). However, these and other authors (e.g., Arthur et al., 2007; Whiteoak et al., 2004) note that the group consensus method is a very laborious method for participants and researchers by requiring a great deal of time and resources. This is one reason why Arthur et al. (2007) focused exclusively on a comparative examination of referent-shift consensus and direct consensus. Arthur et al. (2007) found that the referent-shift consensus method of defining collective efficacy is a stronger predictor of performance over time than direct consensus.

Although a few primary studies directly compared composition models in the same study, there remains a great deal of discrepancy concerning the utility of changing the referent for aggregation (van Mierlo et al., 2009). As such, we examined the relevant literature and meta-analytically compared the referent-shift consensus model of aggregation to the direct consensus method of aggregation by focusing on climate. We chose climate as our exemplar for several reasons. First, as discussed in more detail below, there are theoretical reasons for expecting employees’ climate assessments made via direct consensus and referent-shift consensus methods to qualitatively differ, with referent-shift consensus assessments producing less affectively based appraisals of work environment characteristics (Burke, Borucki, & Hurley, 1992). Second, climate is a well-researched multi-level construct with researchers

using both the direct consensus approach (e.g., Hofmann & Stetzer, 1998; Neal & Griffin, 2006; Schulte, Ostroff, Shmulyian, & Kinicki, 2009; Wallace et al., 2006) and the referent-shift consensus approach (e.g., Arthur et al., 2007; Whiteoak et al., 2004; Zohar, 2000). Third, climate captures the ambient stimuli (Hackman, 1986) that pervade the majority of higher-order constructs that are indicated by shared perceptions. Ambient stimuli are the background factors that influence team or group functioning and cue group members of the appropriate and inappropriate behaviors within the group. Over time, ambient stimuli help to establish group norms that partially direct group members' focus. By comparing the inferences made from different compositional models, future researchers can make more informed decisions about which model is appropriate to use in a given context.

Climate as a Composition Exemplar

Organizational climate refers to the shared perceptions among members of an organization with regard to policies, procedures, and practices. In other words, climate is an "experientially based description of what people see and report happening to them in an organizational situation" (Ostroff, Kinicki, & Tamkins, 2003: 566). Climate can be conceptualized both at the individual level (psychological climate; e.g., Barling, Loughlin, & Kelloway, 2002) and at the group or unit level (e.g., Hofmann & Stetzer, 1996). At the group or unit level, climate is the sharedness of member perceptions commonly operationalized under a specific leader, supervisor, group, or other organizational unit (Zohar, 2002). The focus of the present study is on group- or unit-level climates, but we used the term "organizational climate" to describe our group-level foci of climates (cf. James et al., 2008).

Burke et al. (1992) explained that climate can be further conceptualized as higher-order and first-order climates (see Table 1). This hierarchical perspective is similar to the partitioning of climates into foundation and specific climates later proposed by Schneider and colleagues (Schneider & Bowen, 1993; Schneider, Bowen, Ehrhart, & Holcombe, 2000). The distinction between climates rests in the bandwidth of the climate (Cronbach & Gleser, 1965). That is, higher-order climates² refer to those shared perceptions for larger, more encompassing environments and entities (i.e., relative to one's overall, personal well-being or the well-being of another stakeholder group). Within a service context, Burke et al. (1992) defined two higher-order climates: concern for employees and concern for customers. First-order climate constructs refer to those shared perceptions that are more specific to functional areas or strategic foci (e.g., safety, justice) and are described in Table 1.

To address our research question of whether the referent in aggregation of climates makes a difference in terms of validity, we decided to use the higher-order climates. We made this decision for several reasons. First, science is driven by parsimony. Rather than looking at a multitude of smaller, specific climates that can vary across organizations and industries, we examined the two higher-order climates that are theoretically posited to be present in all organizations (Burke, Borucki, & Kaufman, 2002; Burke et al., 1992; James et al., 2008). Second, the affective component of an employee's climate assessments is only expected to meaningfully vary in regard to the stakeholder or stakeholder group that is the focus of the climate assessment (e.g., oneself, customers, suppliers, and so on). As discussed in more detail below, the affective component of

Table 1
Burke, Borucki, and Hurley's (1992) First- and Higher-Order Psychological Climate Factors

Hypothesized First-Order Factor	Description
<i>Concern for employees—higher-order factor</i>	
Goal emphasis	The extent to which employees perceive their immediate manager as setting clear-cut performance standards.
Management support	The extent to which employees perceive their immediate manager as assisting in performing their jobs and showing concern and respect for employees.
Nonmonetary reward orientation	The extent to which nonmonetary rewards (e.g., praise) are perceived to be linked to outstanding performance in serving customers.
Work group cooperation	The extent to which employees perceive other employees as cooperative and friendly.
Means emphasis	The extent to which employees perceive that their store provides the appropriate training and information necessary for sales personnel to effectively perform their work.
<i>Concern for customers—higher-order factor</i>	
Organizational service orientation	The extent to which employees perceive the store and its employees as assisting customers in finding, buying, and returning merchandise.
Human resource-related obstacles	The extent to which employees perceive the problems or barriers to providing outstanding customer service occurring in their store to be due to human resource-related practices.
Merchandise-related obstacles	The extent to which employees perceive the problems or barriers to providing outstanding customer service occurring in their store to be due to the supply, arrangement, or ticketing of merchandise.

employees' climate assessments is expected to be uniformly strong when assessing all aspects of the work environment that impact them but uniformly lower in affect intensity when evaluating work attributes in relation to another stakeholder group such as customers. This expectation suggests the need to focus on the examination of aggregated higher-order climate scores that relate to these stakeholders when testing predictive relationships. Third, a growing body of empirical evidence supports the tenability of these higher-order climates as predictors of organizationally relevant outcomes (e.g., Alexandrov, Babakus, & Yavas, 2007; Borucki & Burke, 1999; Burke et al., 1992; Burke, Rupinski, Dunlap, & Davison, 1996; Chuang & Liao, 2010; Takeuchi, Chen, & Lepak, 2009; Towler, Lezotte, & Burke, 2011; Vaslow, 1999; Yoon, Beatty, & Suh, 2001). Hence, for theoretical, empirical, and practical reasons, we used the two higher-order climates (concern for employees and customers) to compare direct consensus and referent-shift consensus.

Hypotheses and Expectations

As a test of our hypotheses, we selected the more prominent outcomes of interests to climate scholars (Bowen & Ostroff, 2004; Burke et al., 1992; Ostroff & Bowen, 2000; Ostroff et al., 2003; Schneider, Ehrhart, Mayer, Saltz, & Niles-Jolly, 2005; Schneider, White, & Paul, 1998); these include job performance, customer service (i.e., external customer service ratings), and work attitudes. We selected these outcomes for a variety of reasons.

First, one of the more important and common criterion variables in the organizational sciences is job performance (Austin & Crespín, 2006; Austin & Villanova, 1992). We operationalized performance using performance ratings and similar indices, which corresponded to internal performance of the units. A general argument within the work climate literature is that, within business units, general and human resource management practices (such as training and reward practices) are implemented to produce a consistent level of employee performance (see Kopelman, Brief, & Guzzo, 1990; Schneider et al., 1998).

Second, external customer service ratings and similar indices that correspond to external performance of the units were used. A number of authors have argued that work climate variables at the boundary of the organization underlie customer perceptions of services rendered, with emphasis placed on aspects of climates that pertain to organizational service orientation and barriers to the provision of high levels of service (e.g., Johnson, 1996; Schneider & Bowen, 1993; Towler et al., 2011). Notably, both customer perceptions of service and employee performance have been shown, at the business unit level, to relate to business unit financial performance (e.g., Allen & Wilburn, 2002; Borucki & Burke, 1999; Schneider, Macey, Lee, & Young, 2009), which is indicative of overall effectiveness. Furthermore, the relationships between work climate and the two performance criteria correspond, conceptually, to a climate for employees–job performance (internal ~ job performance) and a climate for customers–customer perceptions of service (external ~ customer service performance) relationship.

Third, climate researchers (Carr, Schmidt, Ford, & DeShon, 2003; Ostroff & Bowen, 2000; Ostroff et al., 2003) have argued for and supported the notion that climates help shape individual (e.g., Seibert, Silver, & Randolph, 2004) and shared employee attitudes (e.g., Richardson & Vandenberg, 2005). The general and long-standing conceptual argument is that work environments that foster higher levels of worker autonomy, influence, and control will engender higher levels of employee attitudes including but not limited to job satisfaction, involvement, and identification with the business unit or organization (see James, Demaree, Mulaik, & Ladd, 1992; Takeuchi et al., 2009). Therefore, our third outcome is employee attitudes, aggregated to the unit level.

The extant literature provides many excellent examples of research demonstrating significant relationships between various first-order climates and performance (e.g., Carr et al., 2003; Christian, Bradley, Wallace, & Burke, 2009; Patterson, Warr, & West, 2004; Zohar, 2000, 2002) and shared attitudes (e.g., Richardson & Vandenberg, 2005). We expect our findings to be similar, such that both a climate for employees and a climate for customers are positively related to job performance, customer service performance, and attitudes. Thus, positive relationships with our three outcomes are expected regardless of the manner in which the climate data were aggregated (direct consensus or referent-shift consensus).

Beyond the expected replications discussed in the preceding paragraph, we expect a differential pattern to emerge with regard to the relationships between methods of aggregation (referent-shift consensus versus direct consensus). Chan (1998) argued that the choice of aggregation is dependent on both the *conceptual* and the *empirical* justification for the specific measurement. Thus, we provide the conceptual justification and meta-analytic evidence for empirical justification for referent-shift consensus and direct consensus. There is a long history of distinction between aggregated psychological climate (consistent with the direct consensus approach) and aggregated measures of work-unit climate (consistent with

referent-shift consensus) (Glick, 1985; James, 1982). From a conceptual standpoint, the distinction between these two levels of measurement would lead to differential prediction of outcomes based on the alignment by level of the construct.

Burke et al. (1992) defined climates as cognitive appraisals of the work environment, which may differ in the extent to which they are driven by affect. The influence of affect is dependent upon whether or not a worker is appraising the work environment relative to oneself or with respect to others. Affective intensity will be greater for an individual's valuations of his or her own work attributes as opposed to other organizational constituents (e.g., groups/teams). This position is consistent with other conceptualizations of cognitive appraisals of event-based emotional experience such as work on the cognitive structure of emotional experience (Ortony, Clore, & Collins, 1988). Event-based emotional experiences are categorized as fortunes-of-self and fortunes-of-others emotions. The important implication is that assessments of concern for employees or concern for customers (as higher-order climate constructs as well as with respect to first-order factors or climates) are expected to qualitatively differ when the climate items are posed in terms of self-evaluations (direct consensus) versus employees as a collective (referent-shift consensus to the relevant group level). That is, an employee's assessment of the work environment relative to oneself is driven more by affective evaluation than is the employee's assessment of how the work environment impacts the goals and interests of others inside and outside of one's workgroup. As such, an employee's assessment of how the work environment impacts the shared goals and interests of others inside and outside of one's workgroup is less affectively based in comparison to his or her assessments of how the work environment impacts him or her personally (Burke et al., 1992; James et al., 2008; James & James, 1989; Mathieu & Chen, 2011; Ortony et al., 1988). As a result, we posit that aggregated climate scores determined via a referent-shift consensus would be expected to have a stronger cognitive base whereas aggregated climate scores determined via direct consensus would have a stronger affective base. Job performance and customer service performance are more cognitively laden and are driven more strongly by shared cognitive processes (DeChurch & Mesmer-Magnus, 2010). Therefore, aggregated climate scores determined via a referent-shift consensus method should be more strongly related to indices of effectiveness (e.g., job performance and customer service performance). In contrast, while we fully acknowledge that attitudes are cognitively derived, they are also affectively laden (Thoresen, Kaplan, Barsky, Warren, & de Chermont, 2003). Because aggregated climate scores determined via direct consensus methods are theoretically more affective based than performance, they should be more strongly related to attitudes. In sum, cognitively laden constructs (referent-shift consensus climate, performance, and customer service) should be matched conceptually as should affectively laden constructs (direct consensus climate, attitudes). We expected that climate-outcome relationships would be stronger when the respective predictor and criterion variables are more closely conceptually matched in terms of their cognitive and affective bases (e.g., referent-shift consensus and performance) than when they are less closely matched (e.g., direct consensus and performance). Thus, our hypotheses are as follows:

Climates and Performance

Hypothesis 1: Concern for employee climate via referent-shift consensus aggregation will be a stronger positive predictor of performance than concern for employee climate via direct consensus.

Hypothesis 2: Concern for customer climate via referent-shift consensus aggregation will be a stronger positive predictor of performance than concern for customer climate via direct consensus.

Climates and Customer Service Performance

Hypothesis 3: Concern for employee climate via referent-shift consensus aggregation will be a stronger positive predictor of customer service performance than concern for employee climate via direct consensus.

Hypothesis 4: Concern for customer climate via referent-shift consensus aggregation will be a stronger positive predictor of customer service performance than concern for customer climate via direct consensus.

Climates and Attitudes

Hypothesis 5: Concern for employee climate via direct consensus aggregation will be a stronger positive predictor of attitudes than concern for employee climate via referent-shift consensus.

Hypothesis 6: Concern for customer climate via direct consensus aggregation will be a stronger positive predictor of attitudes than concern for customer climate via referent-shift consensus.

The justification for Hypotheses 1–6 was that climate-outcome relationships would be stronger when they are theoretically and conceptually matched than when they are mismatched. A potentially important moderator of our hypothesized relationships is the climate-outcome match on levels of measurement. That is, stronger relationships will be observed when direct consensus climate is aligned with direct consensus unit-level outcomes and referent-shift climate is aligned with referent-shift unit-level outcomes. For example, Hypothesis 1 predicts a stronger correlation between performance and concern for employee climate when climate is aggregated via referent-shift than via direct consensus. However, we would also argue that when concern for employee climate is aggregated via referent-shift consensus, relationships will be stronger when performance is also aggregated via referent-shift consensus (compared to direct consensus). Likewise, although we expect overall lower validities when concern for employee climate is aggregated via direct consensus (compared to referent-shift), the relationship with performance will be stronger when performance is also aggregated via direct consensus than referent-shift consensus. A crossover interaction is unlikely though because we would posit that the theoretical match has a stronger influence on validity than the measurement match, much akin to the long-held observation that construct variance is more influential than method variance on observed validities. That is, the relationship between concern for employees via direct consensus matched with performance via direct consensus is not likely to exceed the relationship between concern for employees via referent-shift matched with performance via direct consensus. Thus, a referent-shift climate will still be a stronger predictor of performance (regardless of matched aggregation) than direct consensus climate.

The same rationale would hold for both customer service performance and attitudes. Specifically, customer service performance and attitudes aggregation (referent-shift versus

direct consensus) will moderate the climate-outcome relationships such that stronger relationships are observed for climate-outcome relationships matched by level of measurement. An examination of aggregation indices of widely used criteria such as performance and attitudes represents a significant contribution to the literature. Although most prior multi-level research has focused attention on justifying aggregation of the predictor (e.g., climate), authors have made it clear that alignment of constructs according to measurement and analysis would include justification for aggregating unit-level criterion constructs using individual-level scores (Kozlowski & Klein, 2000; Mathieu & Chen, 2011; Rousseau, 1985). The operationalization of multi-level criterion constructs, and performance in particular, has received very little research attention because performance data are often collected for non-research purposes (e.g., archived performance data provided by the organization). Thus, to test substantive multi-level theories, researchers are often forced to use the most convenient measure of unit-level performance, regardless of measurement aggregation. Nevertheless, conceptual and empirical justification for aggregated performance data is as important as for unit-level climate data.

Method

Literature Search

A search was conducted to identify all unpublished and published manuscripts about climates aggregated to the group, unit, or team level. Included in our definition of work climates were variations on the following generally accepted climate definition: shared employee perceptions of the policies, practices, procedures, and/or behaviors that are rewarded, supported, and/or expected with regard to some focal aspect of the workplace (e.g., safety, service, leadership, diversity, ethics). Keywords for the literature searches included combinations of the following keywords: *climate*, *composition model*, *direct consensus*, and *referent-shift consensus*. Electronic literature searches were conducted through August 2011 of databases including ABI-Inform, Google Scholar, PsycINFO, and Social Sciences Citation Index. In addition, we conducted manual searches of major journals relevant to organizational sciences (e.g., *Academy of Management Journal*, *Journal of Applied Psychology*, *Personnel Psychology*, *Journal of Management*, *Journal of Organizational Behavior*) to locate articles that did not surface in the database searches as well as conference programs for relevant presentations in recent years. We also consulted reference sections of recent climate review articles to identify additional studies (e.g., Kuenzi & Schminke, 2009). Finally, we reviewed the reference sections of our identified studies and conducted cited reference searches to identify any other relevant studies. The initial searches yielded 136 potential articles.

Criteria for Inclusion

Two of the study authors independently evaluated each study using the following inclusion criteria. To be included in the meta-analysis, we retained studies that (a) reported an effect size between one or more climates and one or more aggregated attitudes and/or one or

more performance indicators; (b) presented relationships for climate and criterion variables at the group, team, or unit levels; and (c) provided appropriate justification for aggregation of variables (or enough data to ascertain aggregation suitability). Initial agreement regarding study inclusion was 99%. In the few cases of disagreement, three of the study authors went back to the primary studies to reach consensus. After implementing our inclusion criteria 58 independent samples from 50 studies and 171 effect sizes were used in our analyses.

Coding of Studies

The coding process was conducted by three of the present study's authors. Initially, two individuals coded all studies, and the results were compared. Initial agreement was 92%. In instances of disagreement, a third researcher also coded the study, and all three researchers met to arrive at a consensus through discussion.

Categorization of Criterion Variables

To categorize criterion variables, two of the present study's authors independently sorted them into the pre-determined categories of group/team/unit level (a) attitudes, (b) job performance, and (c) customer service performance.

Group/Team/Unit Attitudes

Attitudes were defined as beliefs and evaluations that focus on specific aspects of work. It was relatively simple to identify and sort attitudes because most studies used consistent construct labels (e.g., job satisfaction was consistently used across studies to describe that variable). In studies where it was less obvious, we sorted them by examining the survey items. The variables sorted as attitudes were job satisfaction, organizational commitment, support, sense of empowerment, and cynicism. We recorded the specific attitude measured (e.g., satisfaction, commitment). Then, we generated a composite measure of attitudes (see Harrison, Newman, & Roth, 2006), consistent with the objectives of the present study. Relevant constructs were reverse-coded (e.g., cynicism) before creating the broader attitude composite. In addition, for primary studies with multiple attitudes, we created within-study composites before computing the attitude composite.

Job Performance

Based on construct labels and items used to measure performance in the primary studies, we sorted performance dimensions into the following categories: overall performance, task performance, contextual performance, safety performance, and sales performance. Consistent with the attitude variable, we combined all job performance dimensions into a single, omnibus composite. Relevant constructs were reverse-coded (e.g., counterproductive behaviors) before creating the broader performance composite. In addition, for primary studies with

multiple performance dimensions, we created within-study composites before computing the overall job performance composite.

Customer Service Performance

Measures of customer service were relatively easy to code because the same construct label was used consistently across studies. The primary difference between customer service performance and job performance (in the previous category) is that customer service was measured by sources external to the organization. Constructs were sorted into the dimensions of customer loyalty, customer service, and perceptions of quality (e.g., Gelade & Ivery, 2003; Liao & Chuang, 2004). Again all customer service constructs/dimensions were combined into a composite score for use in our analyses.

Categorization of Climate Variables

We sorted all climate categories into the framework proposed by Kuenzi and Schminke (2009). The 11 climate categories were based on the focal aspect of the climate (e.g., service, ethics, creativity, etc.). However, a few climates did not easily fit into Kuenzi and Schminke's framework, so we added categories to accommodate additional climates. Next, we sorted the specific climates into the higher-order factors of (a) concern for customers and (b) concern for employees.

Meta-Analytic Calculations

We performed a "bare-bones" meta-analysis using the method of meta-analysis proposed by Raju and colleagues (RBNL procedure; Raju, Burke, Normand, & Langlois, 1991). We chose not to correct for measurement error for several reasons. First, information on the reliability of group-level measures (derived from individual-level scores) is typically not available in the literature. Second, although internal consistency reliability estimates for measures (based on individuals' scores) are available in group-level studies, these reliabilities are inappropriate for correcting relationships based on group-level scores. Third, although ICC2 values (Bartko, 1976; James, 1982; Shrout & Fleiss, 1979) could have been used to correct correlations between group-level measures, we chose not to do so. ICC2 values were not uniformly reported in group-level studies, and arguably their use in the RBNL meta-analytic procedures would call for treating them as assumed fixed reliability estimates given how reliability was defined by Raju and colleagues (see Raju & Brand, 2003; Raju et al., 1991: Appendix A).

The RBNL procedure corrects for artifactual error (i.e., sampling error) and estimates appropriately defined standard errors for corrected correlations. Importantly, the RBNL procedures provide a random effects estimate of the standard error of mean rho when only simple sampling error (i.e., error due to N, the number of groups in this case) is taken into account. The reader is referred to Burke and Landis (2003) for the equation used to estimate

the standard error of the mean corrected correlation (assuming a random effects model) used in this meta-analysis. As noted by several authors, random effects models result in more accurate Type I error rates and confidence intervals than does a fixed effect model (e.g., Erez, Bloom, & Wells, 1996; Overton, 1998). We report 90% confidence intervals for each mean ρ . We also computed credibility intervals, which show the extent that correlations varied across studies for a particular analysis distribution (Hunter & Schmidt, 2004).

Finally, in cases where studies provided multiple correlations from the same sample and the same constructs, we created a single effect to represent the range of non-independent effects using sample size-weighted average correlations (Lipsey & Wilson, 2001). All computations were weighted by the number of groups and not the number of individuals.

Results

Table 2 presents the relationships between our two climate variables (concern for employees and customers—see “overall” category) and our three outcomes (attitudes, performance, customer service). Although we collapsed across climate operationalizations and present the results for an overall rating in Table 2, the tests of our hypotheses are a comparison of climates aggregated via referent-shift consensus and direct consensus. As expected, both climates were positively related to the outcomes, regardless of aggregation method.

Hypotheses 1 and 2 predicted that the relationships between concern for employee (Hypothesis 1) and concern for customers (Hypothesis 2) and job performance would be stronger when the climates were aggregated via referent-shift consensus compared to direct consensus. The relationship for referent-shift consensus ($M_\rho = .39$) was stronger than the relationship for direct consensus ($M_\rho = .30$). We should note that the confidence intervals did not overlap (the upper-bound confidence interval was .336 for direct consensus models, and the lower-bound confidence interval for the referent-shift models was .344), providing support for Hypothesis 1. For the test of Hypothesis 2, we only found one study of job performance that used direct consensus for operationalizing concern for customers. Thus, Hypothesis 2 could not be tested using available data. Nevertheless, the effect size for concern for customers via referent-shift consensus ($M_\rho = .41$) was relatively strong and consistent with our theoretical arguments.

Hypotheses 3 and 4 predicted that the relationships between concern for employees (Hypothesis 3) and concern for customers (Hypothesis 4) and customer service performance would be stronger when the climates were aggregated via referent-shift consensus compared to direct consensus. Concern for employees operationalized via referent-shift consensus ($M_\rho = .30$) was stronger than for direct consensus ($M_\rho = .24$); however, the confidence intervals overlapped, and thus Hypothesis 3 was not supported. Hypothesis 4 was not supported in that the two relationships were nearly identical ($M_\rho = .22$ for direct consensus and $M_\rho = .23$ for referent-shift consensus).

Finally, we hypothesized that concern for employees (Hypothesis 5) and concern for customers (Hypothesis 6) via direct consensus would be a stronger predictor of attitudes than referent-shift consensus. As shown in Table 3, we found support for Hypothesis 5 involving concern for employees, as direct consensus ($M_\rho = .64$) was stronger than referent-shift

Table 2
Meta-Analyses of Climate with Criteria with Climate Composition Model Moderator

Climate							90% Conf. Int.		80% Cred. Int.		% Due to Artifacts	
	Criterion	<i>k</i>	<i>N_i</i>	<i>N_g</i>	<i>SD_r</i>	<i>M_p</i>	<i>SE_{M_p}</i>	L	U	<i>SD_p</i>		L
<i>Concern for employees</i>												
Job attitudes												
Overall	16	78,674	3,628	.16	.61	.04	.54	.68	.16	.41	.81	9.56
Direct consensus	10	74,713	3,175	.13	.64	.04	.57	.70	.13	.47	.80	10.08
Referent-shift	8	5,499	541	.22	.43	.08	.31	.56	.20	.18	.69	19.75
Job performance												
Overall	41	204,073 ^a	5,621	.14	.34	.02	.30	.37	.12	.19	.49	28.38
Direct consensus	19	183,730 ^a	4,018	.10	.30	.02	.26	.34	.08	.19	.40	35.65
Referent-shift	25	110,904	2,268	.15	.39	.03	.34	.44	.15	.20	.57	33.70
Service performance												
Overall	11	99,300	3,370	.06	.24	.02	.21	.28	.04	.20	.29	67.92
Direct consensus	8	90,580	3,030	.06	.24	.02	.21	.27	.03	.20	.28	66.67
Referent-shift	5	9,847	421	.05	.30	.03	.26	.34	.00			100.00
<i>Concern for customers</i>												
Job performance												
Overall	9	49,207	1,193	.22	.50	.07	.38	.62	.21	.23	.77	9.96
Direct consensus	1	34,866	463		.64	.03 ^b	.59	.69				
Referent-shift	8	14,341	730	.24	.41	.09	.27	.55	.23	.12	.71	12.79
Service performance												
Overall	8	49,591	1,267	.07	.23	.03	.19	.27	.00			100.00
Direct consensus	3	38,238	752	.08	.22	.04	.15	.29	.04	.17	.28	65.66
Referent-shift	6	11,353	515	.06	.23	.03	.19	.28	.00			100.00

Notes: *k* = the number of independent effect sizes included in each analysis; *N_i* = number of individuals for studies providing this information; *N_g* = number of groups; *SD_r* = standard deviation of uncorrected correlations; *M_p* = mean correlation corrected for sampling error; *SE_{M_p}* = standard error of *M_p*; 90% Conf. Int. = 90% confidence interval for *M_p*; *SD_p* = standard deviation of estimated *ρ*'s; 80% Cred. Int. = 80% credibility interval. Also, due to low sample size for concern for customers and job attitudes, we do not report effects for these relationships. *k* between direct consensus and referent-shift consensus do not always sum to the total overall *k* because some studies provided estimates for both aggregation models, and in these cases, we averaged the effects to create a single overall effect.

^aIndividual N data are unavailable for one sample of 300 groups.

^bStandard errors and confidence intervals for single correlations are based on the fixed effects model.

consensus (*M_p* = .43). Hypothesis 6 could not be tested as there was only one relationship involving job attitudes for concern for customers.

We also predicted that aggregation of our outcome variables would moderate our hypothesized relationships. It is interesting to note that of our primary studies, only studies of job performance varied the criterion composition model. Thus, of 171 independent effects, only 42 (25%) differentiated among aggregation methods for our outcomes (i.e., had data for referent-shift and direct consensus), and all of them were for job performance. Furthermore, the vast majority originated from studies that measured concern for employee climates. Therefore, our moderator analyses (results presented in Table 3) are limited to concern for

Table 3
Meta-Analyses of Concern for Employees Climate and Job Performance
with Climate Composition Model Moderator and Job Performance
Composition Model Moderator

<i>Climate</i>												
Predictor												
Composition Model												
Criterion												
Composition Model	<i>k</i>	<i>N_i</i>	<i>N_g</i>	<i>SD_r</i>	<i>M_ρ</i>	<i>SE_{M_ρ}</i>	90% Conf. Int.		80% Cred. Int.		% Due to artifacts	
							L	U	<i>SD_ρ</i>	L	U	
<i>Concern for employees</i>												
Direct consensus												
Direct consensus	16	145,921 ^a	3,308	.06	.27	.01	.25	.29	.00			100.00
Referent shift	2	2,943	247	.03	.18	.02	.16	.22	.00			100.00
Referent shift												
Direct consensus	12	63,400	1,054	.13	.30	.04	.23	.36	.10	.17	.42	48.43
Referent shift	12	45,777	1,021	.12	.43	.04	.37	.49	.08	.32	.54	53.70

Notes: *k* = the number of independent effect sizes included in each analysis; *N_i* = number of individuals for studies providing this information; *N_g* = number of groups; *SD_r* = standard deviation of uncorrected correlations; *M_ρ* = mean correlation corrected for sampling error; *SE_{M_ρ}* = standard error of *M_ρ*; 90% Conf. Int. = 90% confidence interval for *M_ρ*; *SD_ρ* = standard deviation of estimated *ρ*'s; 80% Cred. Int. = 80% credibility interval.

^aIndividual N data are unavailable for one sample of 300 groups.

employees (direct consensus and referent-shift consensus) and job performance (direct consensus and referent-shift consensus). The results presented in Table 3 supported our expectations in that the climate-performance relationship was stronger when concern for employees and performance were matched by aggregation indices. That is, the relationship was stronger for concern for employee climate via direct consensus when performance was aggregated via direct consensus ($M_{\rho} = .27$) than via referent-shift ($M_{\rho} = .18$). Likewise, when concern for employee climate was aggregated via referent-shift, the relationship with performance was stronger when performance was also aggregated via referent shift ($M_{\rho} = .43$) than via direct consensus ($M_{\rho} = .30$).

Discussion

At the onset of this paper, we set out with a rather simple but important question: Does the referent matter when composing work climate? As with most research, the answer appears to be: It depends. First, our research supports the overall notion that work climates predict important outcomes of interest for organizational scholars (i.e., job performance, external service performance, and attitudes). These findings hold true regardless of whether climate was aggregated via direct consensus or via referent-shift consensus and thus support a long stream of primary research (e.g., Richardson & Vandenberg, 2005; Schneider et al., 2005; Wallace & Chen, 2006; Zohar, 2000) and meta-analyses (e.g., Christian et al., 2009). Our overall results suggest that climates matter in organizations due to their moderate

relationships with important indicators of internal stakeholder success (i.e., unit-level employee attitudes and performance) and external stakeholders (i.e., unit-level external customer service performance).

The interesting findings were our theoretically expected differences with regard to form of climate composition. Multi-level theory suggests that referent-shift consensus should be a stronger positive predictor of job performance (Hypothesis 1) and external customer service performance (Hypothesis 3) than direct consensus. For example, Klein and colleagues (Klein et al., 1994; Kozłowski & Klein, 2000) suggested that relationships would be stronger if the climate and outcome variables are composed using the same referent at the same level of analysis. Although confidence intervals overlapped in one case, in terms of the direction and magnitude of effects, our results for Hypothesis 1 and Hypothesis 3 were consistent with this theoretical perspective. Our moderator analysis that involved concern for employees' climate and job performance also supported Klein et al.'s (1994) theory. However, Hypotheses 5 and 6 proposed a qualitatively different pattern. We would argue that in addition to the levels of analyses, we need to consider the conceptualization of what the composition model captures. Due to the strong affective underpinnings of employee attitudes (Thoresen et al., 2003) and the more affectively laden aspect of aggregated data based on a direct consensus model (Burke, Chan-Serafin, Salvador, Smith, & Sarpy, 2008; Burke et al., 1992) we expected that direct consensus, not referent-shift consensus, would yield stronger relationships with attitudes. Our findings supported Hypothesis 5—direct consensus for climate for employees yielded stronger positive relationships with attitudes than referent-shift consensus methods. Due to a lack of primary studies we were unable to meta-analytically examine Hypotheses 2 and 6 (Hypothesis 2: Climate for Customers and Job Performance; Hypothesis 6: Climate for Customers and Attitudes).

Theoretical and Practical Implications

There are several theoretical implications of the present study. Primary research has often demonstrated that the group consensus method best captures group phenomena and outperforms other compositional methods (Gibson et al., 2000; Kirkman et al., 2001; Quigley et al., 2007). However, the consensus or group discussion model is not without criticisms. Namely, dominance by one individual who might intimidate other group members could result in a skewed view of the group phenomenon of interest (Anderson & Martin, 1999). There is also likely to be increased pressure on disagreeing members to quickly conform to group norms (Isenberg, 1986; Janis, 1971; Kiesler & Kiesler, 1969). Additionally, this method is much more labor intensive than alternatives due to the time necessary to reach consensus (Bettenhausen, 1991).

As a result, most researchers do not have the luxury of conducting multi-level studies following group consensus methods and are therefore often left deciding between direct consensus and referent-shift consensus methods. Indeed, we found only a few studies that used the group consensus composition model and were thus unable to include them in our meta-analysis. Although we found some evidence that the referent-shift consensus composition yielded stronger effect sizes than direct consensus, this was only true for

performance-related outcomes, but not for attitudes. Specifically, we obtained stronger effects when the outcome was conceptually matched with the appropriate conceptualization of climate (e.g., affective-affective; cognitive-cognitive).

Theoretically, our findings provide an interesting twist to assumed choices for composition models. Results supported our theory that referent-shift consensus may be used for more cognitively laden climate unit-level constructs whereas direct consensus appears to be the more appropriate composition model for more affectively laden unit-level constructs. Researchers should take note and begin to examine the cognitive and affective underpinnings of outcomes they are trying to predict and understand. The current research supports theory (e.g., Burke et al., 1992; DeChurch & Mesmer-Magnus, 2010; Ortony et al., 1988; Thoresen et al., 2003) that an employee's evaluation of his/her immediate work environment relative to oneself is driven more by affective relative to cognitive mechanisms. In contrast, an employee's evaluation of his/her immediate work environment relative to the shared goals and interests of others inside and outside of one's workgroup is driven more by cognitive mechanisms rather than affective mechanisms.

Also, our study highlights a methodological issue concerning how researchers might go about addressing reliability when measures of central tendency are considered as indicators of group-level constructs, and corrections for measurement unreliability in primary and meta-analytic studies. First, researchers could compute disattenuated group-level correlations using ICC2 values as predictor and criterion reliability estimates as noted above. If using the RBNL procedures (Raju & Brand, 2003; Raju et al., 1991) for such corrections within primary and meta-analytic studies, we suggest that researchers treat ICC2 values as assumed fixed reliability estimates given how Raju and colleagues defined reliability (see their Appendix A concerning reliability defined relative to true and observed scores). In addition, within primary studies, for example, researchers could estimate predictor and/or criterion reliability by administering two forms of a climate questionnaire, based on the same content, on one (or more) occasion(s) to the same employees. The correlation between the paired item or scale means on the two forms of the questionnaire would be the reliability estimate. This procedure would produce a coefficient of equivalence or a coefficient of equivalence and stability (if evaluated over multiple occasions), which arguably could be used for making reliability corrections within the RBNL meta-analytic framework.

In terms of practical implications, many practitioners develop and administer employee surveys to measure attitudes or morale. Our results suggest that stronger effect sizes would be obtained using direct consensus to measure climate. The added advantage of using direct consensus is that one could examine results at both the individual and the group level of analyses because the referent would be "I." This would also allow practitioners greater flexibility in rolling up data to other organizational levels of interest. For example, many work units might occupy space in a certain functional division, and retaining the self as the referent would likely more easily allow aggregation up to the functional divisional level. In contrast, when examining performance-related outcomes for groups, teams, or other units of interest in relation to policies, procedures, and practices (i.e., climate variables) practitioners would likely benefit from using referent-shift consensus items for climate. Just as we encouraged researchers to better understand their unit-level criterion to better select the method of climate composition, the same encouragement is shared with practitioners.

Limitations and Directions for Future Work

Unfortunately, we were unable to test Hypothesis 2 because we only found one study that used direct consensus for operationalizing concern for customers. We were also not able to test Hypothesis 6 because there were only two relationships involving job attitudes for concern for customers. Indeed, drawing strong inferences from our analyses for concern for customers climate would be premature, given the low k and the resulting effect size instability. However, we note that the standard deviations of the effects are not large for Customer Service (range: .06–.08), which suggests that the effects in the observed studies are consistent and less prone to error. Also, although the estimates are based on a small number of studies, we note a large number of groups (463 to 1267) and individuals (11,353 to 49,591) went into each effect in the Job Performance and Customer Service categories. Therefore, although our results are interesting and somewhat informative, they are not conclusive. It is clear from our review of the literature that although great progress has been made with regard to investigating composition models in climate research, more primary research is needed. One suggestion for future research is to investigate the extent to which our theoretical contribution extends to other constructs. For example, how does the choice of referent affect relationships involving team conflict or team process variables? DeChurch and Mesmer-Magnus (2010) demonstrated that team process variables have cognitive underpinnings and through the lens of our results would suggest that team process be measured using referent-shift consensus. Future research may also examine the extent to which climate may be defined as having both an affective and cognitive component and if both were measured if this would be a better measure of the intended climate construct. A strong test of our hypotheses would be a primary study comparing group consensus, direct consensus, referent-shift consensus, and dispersion models of climate and multiple outcomes in one study.

Perhaps not surprisingly, very few studies report aggregation indices for criterion constructs. Indeed, none of the studies measuring customer service or attitudes provided aggregation data, and only 41% (42 of 102) of our job performance effect sizes provided aggregation data. Although results involving job performance were consistent with our expectation that climate-performance relationships matched by aggregation method will be stronger than mismatched relationships, we have no empirical data to assess whether the same holds true for customer service and attitudes. Thus, we would echo previous calls for future multi-level research studies to provide aggregation data for relevant criteria.

Conclusion

Aggregating data from the individual level to the group level is now a common and growing aspect of applied psychological research and practice. The present investigation contributes to this area of inquiry by explicating why and how different composition models for data aggregation may lead to different predictor-criterion relationships at the group level of analysis. In doing so, the present study underscores the need for researchers and practitioners alike to attend to the meaning and utility of aggregated data. The validity of aggregated data is no less important than the validity of individual scores; both are a function of the items and the persons responding.

Notes

1. We follow prior research and use the term “organizational climate” to describe group-level climates (James et al., 2008). In the current study we use this term to describe our focal level of climate: group/team/unit and only use this level in our analyses.

2. Some researchers refer to foundation climates as “molar climate” or “general climate” (e.g., Carr, Schmidt, Ford, & Deshon, 2003).

References

References marked with an asterisk indicate studies included in meta-analytic calculations.

- Alexandrov, A., Babakus, E., & Yavas, U. 2007. The effects of perceived management concern for frontline employees and customers on turnover intentions. *Journal of Service Research*, 9: 356-371.
- Allen, D. R., & Wilburn, M. 2002. *Linking customer and employee satisfaction to the bottom line*. Milwaukee, WI: American Society for Quality.
- Anderson, C. M., & Martin, M. M. 1999. The relationship of argumentativeness and verbal aggressiveness to cohesion, consensus, and satisfaction in small groups. *Communication Reports*, 12: 21-31.
- Arthur, W., Jr., Bell, S. T., & Edwards, B. D. 2007. A longitudinal examination of the comparative criterion-related validity of additive and referent-shift consensus operationalizations of team efficacy. *Organizational Research Methods*, 10: 35-58.
- Austin, J. T., & Crespino, T. R. 2006. Problems of criteria in industrial and organizational psychology: Progress problems and prospects. In W. Bennett, C. E. Lance, & D. J. Woehr (Eds.), *Performance measurement: Current perspectives and future challenges*: 9-48. Mahwah, NJ: Lawrence Erlbaum.
- Austin, J. T., & Villanova, P. 1992. The criterion problem: 1917–1992. *Journal of Applied Psychology*, 77: 836-874.
- *Baer, M., & Frese, M. 2003. Innovation is not enough: Climates for initiative and psychological safety, process innovations, and firm performance. *Journal of Organizational Behavior*, 24: 45-68.
- Barling, J., Loughlin, C., & Kelloway, E. K. 2002. Development and test of a model linking safety-specific transformational leadership and occupational safety. *Journal of Applied Psychology*, 87: 488-496.
- Bartko, J. J. 1976. On various intraclass correlation reliability coefficients. *Psychological Bulletin*, 83: 762-765.
- Bettenhausen, K. L. 1991. Five years of group research: What have we learned and what needs to be addressed. *Journal of Management*, 17: 345-381.
- *Borucki, C. C., & Burke, M. J. 1999. An examination of service-related antecedents to retail store performance. *Journal of Organizational Behavior*, 20: 943-962.
- Bowen, D. E., & Ostroff, C. 2004. Understanding HRM-firm performance linkages: The role of the “strength” of the system. *Academy of Management Review*, 29: 203-221.
- Burke, M. J., Borucki, C. C., & Hurley, A. 1992. Reconceptualizing psychological climate in a retail service environment: A multiple stakeholder perspective. *Journal of Applied Psychology*, 77: 717-729.
- Burke, M. J., Borucki, C. C., & Kaufman, J. 2002. Contemporary perspectives on the study of psychological and organizational climate: A commentary. *European Journal of Work and Organizational Psychology*, 11: 325-340.
- Burke, M. J., Chan-Serafin, S., Salvador, R., Smith, A., & Sarpy, S. 2008. The role of national culture and organizational climate in safety training effectiveness. *European Journal of Work and Organizational Psychology*, 17: 133-154.
- Burke, M. J., & Dunlap, W. P. 2002. Estimating interrater agreement with the Average Deviation (AD) index: A user's guide. *Organizational Research Methods*, 5: 159-172.
- Burke, M. J., & Landis, R. 2003. Methodological and conceptual issues in applications of meta-analysis. In K. Murphy (Ed.), *Validity generalization: A critical review*: 287-310. Hillsdale, NJ: Lawrence Erlbaum.
- Burke, M. J., Rupinski, M. T., Dunlap, W. P., & Davison, H. K. 1996. Do situational variables act as substantive causes of relationships between individual difference variables? Two large-scale tests of “Common Cause” models. *Personnel Psychology*, 49: 573-598.
- Carr, J. Z., Schmidt, A. M., Ford, J. K., & DeShon, R. P. 2003. Climate perceptions matter: A meta-analytic path analysis relating molar climate, cognitive and affective states, and individual level work outcomes. *Journal of Applied Psychology*, 88: 605-619.

- 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., Kirkman, B., Kanfer, R., Allen, D., & Rosen, B. 2007. A multilevel study of leadership, empowerment, and performance in teams. *Journal of Applied Psychology*, 92: 331-346.
- *Chen, Z., Lam, W., & Zhong, J. A. 2007. Leader-member exchange and member performance: A new look at individual-level negative feedback-seeking behavior and team-level empowerment climate. *Journal of Applied Psychology*, 92: 202-212.
- Christian, M. S., Bradley, J. C., Wallace, J. C., & Burke, M. J. 2009. Workplace safety: A meta-analysis of the roles of person and situation factors. *Journal of Applied Psychology*, 94: 1103-1127.
- Chuang, C., & Liao, H. 2010. Strategic human resource management in service context: Taking care of business by taking care of employees and customers. *Personnel Psychology*, 63: 153-196.
- *Collins, C. J., & Smith, K. G. 2006. Knowledge exchange and combination: The role of human resource practices in the performance of high-technology firms. *Academy of Management Journal*, 49: 544-560.
- *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.
- Cronbach, L. J., & Gleser, G. C. 1965. *Psychological tests and personnel decisions*. Urbana, IL: University of Illinois Press.
- *Cullen, J., Parboteeah, K. P., & Victor, B. 2003. The effects of ethical climates on organizational commitment: A two-study analysis. *Journal of Business Ethics*, 46: 127-141.
- *Dawson, J. F., Gonzalez-Roma, V., Davis, A., & West, M. A. 2008. Organisational climate and climate strength in UK hospitals. *European Journal of Work and Organizational Psychology*, 17: 89-111.
- *de Jong, A., de Ruyter, K., & Lemmink, J. 2005. Service climate in self-managing teams: Mapping the linkage of team member perceptions and service performance outcomes in a business-to-business setting. *Journal of Management Studies*, 42: 1593-1620.
- DeChurch, L. A., & Mesmer-Magnus, J. R. 2010. The cognitive underpinnings of effective teamwork: A meta-analysis. *Journal of Applied Psychology*, 95: 32-53.
- *Dietz, J., Pugh, S. D., & Wiley, J. 2004. Service climate effects on customer attitudes: An examination of boundary conditions. *Academy of Management Journal*, 47: 81-92.
- *Ehrhart, M. G. 2004. Leadership and procedural justice climate as antecedents of unit-level organizational citizenship behavior. *Personnel Psychology*, 57: 61-94.
- Erez, A., Bloom, M. C., & Wells, M. T. 1996. Using random rather than fixed effects models in meta-analysis: Implications for situational specificity and validity generalization. *Personnel Psychology*, 49: 275-306.
- *Fay, D., Luhrmann, H., & Kohl, C. 2004. Proactive climate in a post-reorganization setting: When staff compensate managers' weakness. *European Journal of Work and Organizational Psychology*, 13: 241-267.
- *Gelade, G. A., & Ivery, M. 2003. The impact of human resource management and work climate on organizational performance. *Personnel Psychology*, 56: 383-404.
- *Gelade, G. A., & Young, S. 2005. Test of a service profit chain model in the retail banking sector. *Journal of Occupational & Organizational Psychology*, 78: 1-22.
- Gibson, C. B., Randel, A. E., & Earley, P. C. 2000. Understanding group efficacy: An empirical test of multiple assessment methods. *Group and Organization Management*, 25: 67-97.
- *Gil, F., Rico, R., Alcover, C. M., & Barrasa, A. 2005. Change-oriented leadership, satisfaction and performance in work groups. *Journal of Managerial Psychology*, 20: 312-328.
- Glick, W. H. 1985. Conceptualization and measuring organizational and psychological climate: Pitfalls of multi-level research. *Academy of Management Review*, 10: 601-610.
- Glisson, C., & James, L. R. 2002. The cross-level effects of culture and climate in human service teams. *Journal of Organizational Behavior*, 23: 767-794.
- *Gonzalez-Roma, V., Fortes-Ferreira, L., & Peiro, J. 2009. Team climate, climate strength and team performance. A longitudinal study. *Journal of Occupational and Organizational Psychology*, 82: 511-536.
- Hackman, J. R. 1986. The psychology of self-management in organizations. In M. S. Pallack & R. O. Perloff (Eds.), *Psychology and work: Productivity, change, and employment*: 89-136. Washington, DC: American Psychological Association.
- Harrison, D. A., Newman, D. A., & Roth, P. L. 2006. How important are job attitudes? Meta-analytic comparisons of integrative behavioral outcomes and time sequences. *Academy of Management Journal*, 49: 305-325.

- *Hofmann, D. A., & Mark, B. 2006. An investigation of the relationship between safety climate and medication errors as well as other nurse and patient outcomes. *Personnel Psychology*, 59: 847-869.
- *Hofmann, D. A., & Stetzer, A. 1996. A cross-level investigation of factors influencing unsafe behaviors and accidents. *Personnel Psychology*, 49: 307-339.
- *Hofmann, D. A., & Stetzer, A. 1998. The role of safety climate and communication in accident interpretation: Implications for learning from negative events. *Academy of Management Journal*, 41: 644-657.
- *Hui, C. H., Chiu, W. C. K., Yu, P. L. H., Cheng, K., & Tse, H. 2007. The effects of service climate and the effective leadership behaviour of supervisors on frontline employee service quality: A multi-level analysis. *Journal of Occupational & Organizational Psychology*, 80: 151-172.
- Hunter, J. E., & Schmidt, F. L. 2004. *Methods of meta-analysis: Correcting error and bias in research findings* (2nd ed.). Thousand Oaks, CA: Sage.
- Isenberg, D. J. 1986. Group polarization: A critical review and meta-analysis. *Journal of Personality and Social Psychology*, 50: 1141-1151.
- James, L. R. 1982. Aggregation bias in estimates of perceptual agreement. *Journal of Applied Psychology*, 67: 219-229.
- James, L. R., Choi, C. C., Ko, C. E., McNeil, P. K., Minton, M. K., Wright, M. A., & Kim, K. 2008. Organizational and psychological climate: A review of theory and research. *European Journal of Work and Organizational Psychology*, 17: 5-32.
- James, L. R., Demaree, R. G., Mulaik, S. A., & Ladd, R. T. 1992. Validity generalization in the context of situational models. *Journal of Applied Psychology*, 77: 3-14.
- 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. A., & James, L. R. 1989. Integrating work perceptions: Explorations into the measurement of meaning. *Journal of Applied Psychology*, 74: 739-751.
- Janis, I. L. 1971. Groupthink. *Psychology Today*, 5: 43-46.
- Johnson, J. W. 1996. Linking employee perceptions of service climate to customer satisfaction. *Personnel Psychology*, 49: 831-851.
- *Katz-Navon, T., Naveh, E., & Stern, Z. 2005. Safety climate in healthcare organizations: A multidimensional approach. *Academy of Management Journal*, 48: 1075-1089.
- Kiesler, C. A., & Kiesler, S. B. 1969. *Conformity*. Reading, MA: Addison-Wesley.
- *King, E. B., de Chermont, K., West, M., Dawson, J. F., & Hebl, M. 2007. How innovation can alleviate negative consequences of demanding work contexts: The influence of climate for innovation on organizational outcomes. *Journal of Occupational and Organizational Psychology*, 80: 631-645.
- Kirkman, B. L., Tesluk, P. E., & Rosen, B. 2001. Assessing the incremental validity of consensus ratings over aggregation of individual-level data in predicting team effectiveness. *Personnel Psychology*, 54: 645-667.
- *Klein, K. J., Conn, A. B., Smith, 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.
- Klein, K. J., Dansereau, F., & Hall, R. J. 1994. Levels issues in theory development, data collection, and analysis. *Academy of Management Review*, 19: 105-229.
- Kopelman, R. E., Brief, A. P., & Guzzo, R. A. 1990. The role of climate and culture in productivity. In B. Schneider (Ed.), *Organizational climate and culture*: 282-318. San Francisco: Jossey-Bass.
- Kozlowski, S. W. J., & Klein, K. J. 2000. A multilevel approach to theory and research in organization: Contextual, temporal, and emergent properties. In K. J. Klein & S. W. J. Kozlowski (Eds.), *Multilevel theory, research, and methods in organizations*: 3-90. San Francisco: Jossey-Bass.
- Kuenzi, M., & Schminke, M. 2009. Assembling fragments into a lens: A review, critique, and proposed research agenda for the organizational work climate literature. *Journal of Management*, 35: 634-717.
- *Liao, H., & Chuang, A. 2004. A multilevel investigation of factors influencing employee service performance and customer outcomes. *Academy of Management Journal*, 47: 41-58.
- *Lindell, M., & Brandt, C. 2000. Climate quality and climate consensus as mediators of the relationship between organizational antecedents and outcomes. *Journal of Applied Psychology*, 85: 331-348.
- Lipsey, M. W., & Wilson, D. B. 2001. *Practical meta-analysis*. Thousand Oaks, CA: Sage.
- *Luria, G. 2008. Controlling for quality: Climate, leadership, and behavior. *Quality Management Journal*, 15: 27-42.

- *Luria, G., & Yagil, D. 2008. Procedural justice, ethical climate and service outcomes in restaurants. *International Journal of Hospitality Management*, 27: 276-283.
- *MacCormick, J. S., & Parker, S. K. 2010. A multiple climates approach to understanding business unit effectiveness. *Human Relations*, 63: 1171-1806.
- *Martin, A. 2008. Service climate and employee well-being in higher education. *Journal of Management & Organization*, 14: 155-167.
- Mason, C. M., & Griffin, M. A. 2003. Identifying group task satisfaction at work. *Small Group Research*, 34: 413-442.
- Mathieu, J. E., & Chen, G. 2011. The etiology of the multilevel paradigm in management research. *Journal of Management*, 37: 610-641.
- *Mathisen, G. E., Torsheim, T., & Einarsen, S. 2006. The team-level model of climate for innovation: A two-level confirmatory factor analysis. *Journal of Occupational and Organizational Psychology*, 79: 23-35.
- *Mayer, D. M., Kuenzi, M., & Greenbaum, R. L. 2010. Examining the link between ethical leadership and employee misconduct: The mediating role of ethical climate. *Journal of Business Ethics*, 95: 7-16.
- *Maynard, M. T., Mathieu, J. E., Marsh, W. M., & Ruddy, T. M. 2007. A multilevel investigation of the influences of employees' resistance to empowerment. *Human Performance*, 20: 147-171.
- *McKay, P. F., Avery, D. R., & Morris, M. A. 2008. Mean racial-ethnic differences in employee sales performance: The moderating role of diversity climate. *Personnel Psychology*, 61: 349-374.
- *McKay, P. F., Avery, D. R., & Morris, M. A. 2009. A tale of two climates: Diversity climate from subordinates' and managers' perspectives and their role in store unit sales performance. *Personnel Psychology*, 62: 767-791.
- *Naumann, S., & Bennett, N. 2002. The effects of procedural justice climate on work group performance. *Small Group Research*, 33: 361-377.
- *Naveh, E., Katz-Navon, T., & Stern, Z. 2005. Treatment errors in healthcare: A safety climate approach. *Management Science*, 51: 948-960.
- *Neal, A., & Griffin, M. A. 2006. A longitudinal study of the relationships among safety climate, safety behavior, and accidents at the individual and group levels. *Journal of Applied Psychology*, 91: 946-953.
- Ortony, A., Clore, G. L., & Collins, A. 1988. *The cognitive structure of emotions*. New York: Cambridge University Press.
- Ostroff, C., & Bowen, D. E. 2000. Moving HR to a higher level: HR practices and organizational effectiveness. In K. J. Klein & S. W. J. Kozlowski (Eds.), *Multilevel theory, research, and methods in organizations: Foundations, extensions, and new directions*: 211-266. San Francisco: Jossey-Bass.
- *Ostroff, C., Kinicki, A., & Clark, M. 2002. Substantive and operational issues of response bias across levels of analysis: An example of climate-satisfaction relationships. *Journal of Applied Psychology*, 87: 355-368.
- Ostroff, C., Kinicki, A. J., & Tamkins, M. M. 2003. Organizational culture and climate. In W. C. Borman, D. R. Ilgen & R. J. Klimoski (Eds.), *Comprehensive handbook of psychology: Vol. 12. Industrial and organizational psychology*: 565-594. New York: Wiley.
- Overton, R. C. 1998. A comparison of fixed-effects and mixed (random-effects) models for meta-analysis tests of moderator variable effects. *Psychological Methods*, 3: 354-379.
- Patterson, M., Warr, P., & West, M. 2004. Organizational climate and company productivity: The role of employee affect and employee level. *Journal of Occupational and Organizational Psychology*, 77: 193-216.
- *Pirola-Merlo, A., Härtel, C., Mann, L., & Hirst, G. 2002. How leaders influence the impact of affective events on team climate and performance in R&D teams. *Leadership Quarterly*, 13: 561-581.
- *Probst, T. 2004. Safety and insecurity: Exploring the moderating effect of organizational safety climate. *Journal of Occupational Health Psychology*, 9: 3-40.
- Quigley, N. R., Tekleab, A. G., & Tesluk, P. E. 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.
- Raju, N. S., & Brand, P. A. 2003. Determining the significance of correlations corrected for unreliability and range restriction. *Applied Psychological Measurement*, 27: 52-71.
- Raju, N. S., Burke, M. J., Normand, J., & Langlois, G. 1991. A new meta-analytic approach. *Journal of Applied Psychology*, 76: 432-446.
- *Richardson, H. A., & Vandenberg, R. J. 2005. Integrating managerial perceptions and transformational leadership into a work-unit level model of employee involvement. *Journal of Organizational Behavior*, 26: 561-589.

- Rousseau, D. M. 1985. Issues of level in organizational research: Multi-level and cross-level perspectives. *Research in Organizational Behavior*, 7: 1-37.
- *Salanova, M., Agut, S., & Peiro, J. M. 2005. Linking organizational resources and work engagement to employee performance and customer loyalty: The mediation of service climate. *Journal of Applied Psychology*, 90: 1217-1227.
- *Schmit, M. J., & Allscheid, S. P. 1995. Employee attitudes and customer satisfaction: Making theoretical and empirical connections. *Personnel Psychology*, 48: 521-536.
- Schneider, B., & Bowen, D. 1993. The service organization: Human resources management is crucial. *Organizational Dynamics*, 21: 39-52.
- Schneider, B., Bowen, D., Ehrhart, M. E., & Holcombe, K. M. 2000. The climate for service: Evolution of a construct. In N. M. Ashkanasy, C. Wilderom, & M. F. Peterson (Eds.), *Handbook of organizational culture and climate*: 21-36. Thousand Oaks, CA: Sage.
- *Schneider, B., Ehrhart, M. G., Mayer, D. M., Saltz, J. L., & Niles-Jolly, K. 2005. Understanding organization-customer links in service settings. *Academy of Management Journal*, 48: 1017-1032.
- Schneider, B., Macey, W. H., Lee, W., & Young, S. A. 2009. Organizational service climate drivers of the American Customer Satisfaction Index (ACSI) and financial and market value performance. *Journal of Service Research*, 32: 3-14.
- *Schneider, B., White, S. S., & Paul, M. C. 1998. Linking service climate and customer perceptions of service quality: Test of a causal model. *Journal of Applied Psychology*, 83: 150-163.
- *Schulte, M., Ostroff, C., Shmulyian, S., & Kinicki, A. 2009. Organizational climate configurations: Relationships to collective attitudes, customer satisfaction, and financial performance. *Journal of Applied Psychology*, 3: 618-634.
- *Seibert, S. E., Silver, S. R., & Randolph, W. A. 2004. Taking empowerment to the next level: A multiple-level model of empowerment, performance, and satisfaction. *Academy of Management Journal*, 47: 332-349.
- Shrout, P. E., & Fleiss, J. L. 1979. Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, 86: 420-428.
- *Simons, T., & Roberson, Q. 2005. Why managers should care about fairness: The effects of aggregate justice perceptions on organizational outcomes. *Journal of Applied Psychology*, 88: 432-443.
- *Smith, K. G., Collins, C. J., & Clark, K. D. 2005. Existing knowledge, knowledge creation capability, and the rate of new product introduction in high-technology firms. *Academy of Management Journal*, 48: 346-357.
- *Sowinski, D. R., Fortmann, K. A., & Lezotte, D. V. 2008. Climate for service and the moderating effects of climate strength on customer satisfaction, voluntary turnover, and profitability. *European Journal of Work and Organizational Psychology*, 17: 73-88.
- *Stetzer, A., Morgeson, F., & Anderson, E. 1997. Organizational climate and ineffectiveness: Evidence from 25 outdoor workcrew divisions. *Journal of Quality Management*, 2: 251.
- Takeuchi, R., Chen, G., & Lepak, D. P. 2009. Through the looking glass of a social system: Cross-level effects of high-performance work systems on employees' attitudes. *Personnel Psychology*, 62: 1-30.
- *Tesluk, P., Farr, J., Mathieu, J., & Vance, R. 1995. Generalization of employee involvement training to the job setting: Individual and situational effects. *Personnel Psychology*, 48: 607-632.
- Thoresen, C. J., Kaplan, S. A., Barsky, A. P., Warren, C. R., & de Chermont, K. 2003. The affective underpinnings of job perceptions and attitudes: A meta-analytic review and integration. *Psychological Bulletin*, 129: 914-945.
- *Towler, A., Lezotte, D. V., & Burke, M. J. 2011. An examination of the service climate-firm performance chain: The role of customer retention. *Human Resource Management*, 50: 391-406.
- 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.
- Vaslow, J. 1999. *The multiple stakeholder model of psychological climate: Beyond employees and customers*. Unpublished doctoral dissertation, Tulane University, New Orleans, LA.
- *Veld, M., Paauwe, J., & Boselie, P. 2011. HRM and strategic climates in hospitals: Does the message come across at the ward level. *Human Resource Management*, 20: 339-356.
- *Wagner, S. H., Parker, C. P., & Christiansen, N. D. 2003. Employees that think and act like owners: Effects of ownership beliefs and behaviors on organizational effectiveness. *Personnel Psychology*, 56: 847-871.
- *Wallace, C., & Chen, G. 2006. A multilevel integration of personality, climate, self-regulation, and performance. *Personnel Psychology*, 59: 529-557.

- *Wallace, J. C., Popp, E., & Mondore, S. 2006. Safety climate as a mediator between foundation climates and occupational accidents. A group level study. *Journal of Applied Psychology*, 91: 681-688.
- *Walsh, B., Matthews, R., Tuller, M., Parks, K., & McDonald, D. 2010. A multilevel model of the effects of equal opportunity climate on job satisfaction in the military. *Journal of Occupational Health Psychology*, 15: 191-207.
- *Walumbwa, F. O., Wu, C., & Orwa, B. 2008. Contingent reward transactional leadership, work attitudes, and organizational citizenship behavior: The role of procedural justice climate perceptions and strength. *Leadership Quarterly*, 19: 251-265.
- Whiteoak, J. W., Chalip, L., & Hort, L. K. 2004. Assessing group efficacy: Comparing three measurement models. *Small Group Research*, 35: 158-173.
- *Yagil, D., & Luria, G. 2010. Friends in need: The protective effect of social relationships under low-safety climate. *Group & Organization Management*, 35: 727-745.
- Yoon, M. H., Beatty, S. E., & Suh, J. 2001. The effect of work climate on critical employee and customer outcomes. *International Journal of Service Industry Management*, 12: 500-521.
- *Zohar, D. 1980. Safety climate in industrial organizations: Theoretical and applied implications. *Journal of Applied Psychology*, 65: 96-102.
- Zohar, D. 2000. A group-level model of safety climate: Testing the effect of group climate on micro-accidents in manufacturing jobs. *Journal of Applied Psychology*, 85: 587-596.
- Zohar, D. 2002. Modifying supervisory practices to improve subunit safety: A leadership-based intervention model. *Journal of Applied Psychology*, 87: 156-163.
- *Zohar, D., & Luria, G. 2005. A multilevel model of safety climate: Cross level relationships between organization and group-level climates. *Journal of Applied Psychology*, 90: 616-628.