Is Everyone in Agreement? An Exploration of Within-Group Agreement in Employee Perceptions of the Work Environment

Katherine J. Klein, Amy Buhl Conn, D. Brent Smith, and Joann Speer Sorra
University of Maryland at College Park

Multilevel researchers often gather individual-level data to measure group-level constructs. Within-group agreement is a key consideration in the measurement of such constructs, yet antecedents of within-group agreement have been little studied. The authors found that group member social interaction and work interdependence were significantly positively related to within-group agreement regarding perceptions of the work environment. Demographic heterogeneity was not significantly related to within-group agreement. Survey wording showed a complex relationship to agreement. Both evaluative items and socially undesirable items generated high within-group agreement. The use of a group rather than individual referent increased within-group agreement in response to descriptive items but decreased within-group agreement in response to evaluative items. Items with a group referent showed greater between-group variability than items with an individual referent.

As multilevel research grows increasingly common within the organizational sciences, researchers more and more frequently gather and summarize individual-level data to operationalize not only individual-level constructs but also group- and organization-level constructs. Thus, for example, researchers assess and aggregate individual organizational members’ perceptions of psychological climate to operationalize organizational climate (e.g., Hofmann & Stetzer, 1996). Similarly, researchers assess and aggregate individual subordinates’ assessments of their leader to operationalize leader charisma (e.g., Howell & Avolio, 1993). And researchers assess and summarize measures of group members’ individual demographic characteristics to operationalize group demographic heterogeneity (e.g., Harrison, Price, & Bell, 1998).

In gathering and summarizing individual-level data to operationalize unit-level (group or organization) constructs, researchers implicitly or explicitly propose composition models for their constructs (Kozlowski & Klein, 2000; Rousseau, 1985). Composition models “specify the functional relationships among phenomena or constructs at different levels of analysis . . . that reference essentially the same content but that are qualitatively different at different levels” (Chan, 1998, p. 234). Composition models, in other words, specify how individual psychological climate is related to organizational climate, how individual perceptions of a leader are related to the charisma of the leader, and how individual demographics are related to organizational demographic heterogeneity.

A central consideration in many composition models is agreement or homogeneity among unit members (Kozlowski & Klein, 2000; Klein, Dansereau, & Hall, 1994). To what extent, for example, do unit members agree about the climate of the unit? In this study, we examined three possible antecedents of agreement among group members’ perceptions of the work environment: (a) demographic homogeneity within the group, (b) social interaction and work interdependence among group members, and (c) the wording of survey items used to measure perceptions of the work environment. Before describing our specific hypotheses, methodology, and results, we summarize key elements of Chan’s (1998) recent typology of composition models, highlighting the significance of within-group agreement in each of the models.

Composition Models and Within-Unit Agreement

Building on prior discussions of composition (e.g., Brown, Kozlowski, & Hattrup, 1996; House, Rousseau, & Thomas-Hunt, 1995; Klein et al., 1994; Rousseau, 1985), Chan (1998) proposed a typology of five composition models. Within-group agreement emerges as a particularly important consideration in the three models outlined below.

Direct Consensus Models

The direct consensus model of composition is “probably the most familiar and popular form of composition among multilevel researchers” (Chan, 1998, p. 237). This model “uses within-group consensus of the lower level units 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” (Chan, 1998, p. 237). Constructs...
describing isomorphic or shared properties of a unit (Kozlowski & Klein, 2000)—such as organizational climate, organizational culture, group norms, and leadership style—typically rest on direct consensus models of composition. The very definition of such constructs asserts that unit members agree in their perceptions of the relevant characteristics of the unit. In the absence of substantial within-unit agreement, the unit-level construct is untenable, moot. Thus, for example, in the absence of substantial agreement among the members of a unit about the unit’s norms, the unit simply has no shared norms.

**Referent-Shift Consensus Models**

The referent-shift consensus model is very similar to the group consensus composition model, except that “in referent-shift consensus composition, the lower level attributes being assessed for consensus are conceptually distinct though derived from the original individual-level construct” (Chan, 1998, p. 238). Consider a leadership example. To operationalize a direct consensus model of leader charisma, a researcher might ask the subordinates of a leader to rate the extent to which they agree with the statement “I am inspired by my leader’s vision for our work group.” In contrast, to operationalize a referent-shift consensus model of leader charisma, the researcher might ask the subordinates of a leader to rate the extent to which they agree with the statement, “The members of my group are inspired by our leader’s vision for our work group.” In the latter case, the individual-level referent for the group-level construct has shifted from an individual’s report of the extent to which he or she is inspired by the leader to an individual’s perception of the extent to which group members, in general, are inspired by the leader. In both the direct consensus and referent-shift consensus models, however, within-group agreement is key. Unit members must show substantial within-unit agreement in their perceptual ratings or the unit-level measure—the aggregation of individual responses to the unit level—has no construct validity.

**Dispersion Models**

The dispersion model of composition is quite different. Here, “the degree of within-group agreement of scores from the lower-level units or attributes . . . [is] a focal construct as opposed to merely a statistical prerequisite for aggregation” (Chan, 1998, p. 239). As Chan (1998, p. 239) noted, “dispersion is by definition a group-level characteristic . . . because it refers to the variability within a group and a variance statistic is indexing an attribute of a group as opposed to an attribute of any individual-level response.” Because dispersion models are relatively rare within the organizational literature and of central concern within our research, we provide several examples. Perhaps the most common example of a dispersion variable within the empirical organizational literature is demographic heterogeneity—a measure of the variability of unit members’ age, gender, race, tenure, educational background, and other characteristics. In the past decade, many studies (e.g., Chatman, Polzer, Barsade, & Neale, 1998; Harrison et al., 1998; Tsui, Egan, & O’Reilly, 1992) have examined the correlates of demographic heterogeneity. Researchers have more rarely examined the correlates of variability in unit members’ perceptions of their work environment. Bliese and Halverson (1998), however, hypothesized and found that the greater the variability in unit members’ perceptions of their unit’s leadership climate and peer relations, the lower group members’ average psychological well-being.

Dispersion constructs dot the theoretical literature, as well. Klein and House (1995), for example, explored the homogeneity of charisma, a dispersion construct, arguing that a leader may share charismatic relationships with all, some, or none of his or her followers. Hambrick (1994, pp. 202–203) examined top management team members’ paradigm homogeneity—the extent to which team members share “the same view of the world,” a common vocabulary, consensus on strategic means and ends, and shared perceptions of the environment, the organization’s strategic position, and its prospects. Other recent examples of dispersion include Waldman and Yammarino’s (1999) discussion of culture strength; Kostova’s (1999) discussion of the degree of compatibility between the values of a parent company and its subsidiaries; and Drazin, Glynn, and Kazanjian’s (1999) discussion of the extent to which diverse occupational groups share a common creative frame of reference; and Klein, Palmer, and Conn’s (2000) discussion of possible antecedents and consequences of variability among the individuals involved in interorganizational relationships.

Researchers and theorists have thus begun to explore the implications of within-group dispersion in a wide variety of individual characteristics (including demographic variables, perceptions, goals, and values). Intriguingly, a common thread unites studies and conceptual analyses of dispersion. Across topic areas, organizational scholars have proposed, and increasingly documented, that dispersion within a group—dispersion of demographic variables, skills, attitudes, perceptions, or values—is positively related to group creativity but negatively related to social integration within the group, cohesion, speed of decision making, and ease of decision implementation (e.g., Bliese & Halverson, 1998; Drazin et al., 1999; Hambrick, 1994; Klein & House, 1995; Kostova, 1999; Williams & O’Reilly, 1998). Of course, numerous factors moderate this relationship. The effects of diversity may depend on the culture of the organization, for example (Chatman et al., 1998). Further, some forms of diversity (e.g., diversity of visible characteristics such as gender) may yield different effects than other forms of diversity (e.g., diversity of less visible characteristics such as functional background; Pelled, 1996). Nevertheless, the basic and common thesis remains: Dispersion within a group is likely to enhance group creativity but diminish social integration within the group.

**Summary Comments: The Meaning of Agreement**

In group consensus and referent-shift composition models, within-group agreement is a prerequisite for the aggregation of the individual-level data to the group-level. In these models, within-group agreement is more than a statistical hurdle. It is an integral element in the definition of the group-level construct. In espousing a consensus model of compilation, a researcher proposes that the members of each group in his or her sample are homogeneous with respect to the construct of interest, although groups differ in the absolute level of the construct. The consensus researcher seeks to learn the correlates of between-group variability in the absolute level of the focal construct. In espousing a dispersion model of compilation, in contrast, a researcher suggests that the groups in his or her sample vary in their level of homogeneity. The disper-
sion researcher seeks to learn the correlates of between-group variability in the homogeneity of the focal construct. Nevertheless, both dispersion and consensus researchers attribute the same meaning to within-group agreement. In consensus models as in dispersion models, within-group agreement indicates the extent to which group members constitute a homogeneous group, united by their common perceptions, interests, behaviors, and/or other characteristics.

Antecedents of Agreement in Group Members’ Perceptions of the Work Environment

Organizational research and theory examining dispersion remain rare, as we have noted. In this study, we examined possible antecedents of agreement in group members’ perceptions of their work environment. We use the term perceptions of the work environment broadly to refer to employees’ observations, interpretations, and/or evaluations of their group’s working conditions, patterns of interaction, norms, climate, culture, outcomes, or external contacts and competitors. Measures of perceptions of the work environment are common in group-level, organization-level, and multilevel studies. Thus, for example, Schmitt, Sacco, Ramey, Ramey, and Chan (1999) used a measure of perceptions of the work environment—that is, teacher perceptions of school climate—as a key independent variable in their recent longitudinal study of children’s academic and social development. Shamir, Zakay, Breinin, and Popper (1998) assessed perceptions of the work environment in asking members of military units to rate their own unit’s discipline and unit culture. And, Tesluk and Mathieu (1999) assessed perceptions of the work environment in asking work crew members to rate their crew’s effectiveness, teamwork processes, self-management, and internal leadership.

Within-group agreement in perceptions of the work environment is a dispersion construct. As noted above, existing theory suggests that groups whose members vary a great deal in their perceptions of the work environment may be unable to “bridge their differences to formulate, much less implement, adaptive responses to the environment” (Hambrick, 1994, p. 206). Thus, for example, crew members who disagree about the quality of their teamwork processes may be unable to reach consensus regarding their crew’s need for training and other interventions. In a similar vein, manufacturing plant managers and supervisors who disagree about the innovativeness of and availability of slack financial resources in their plant—two aspects of the work environment that we assessed—may be unable to reach a consensus regarding the need for and feasibility of implementing strategic innovations in the plant. Groups whose members vary a great deal in their perceptions of the work environment may experience conflict among group members (Klein & House, 1995; Pelled, 1996), low cohesiveness (Harrison et al., 1998; Jackson, May, & Whitney, 1995), and high levels of stress and dissatisfaction (Bliese & Halverson, 1998; Jackson et al., 1995).

What causes group members to vary in their perceptions of the work environment? Existing theory and research, as we detail below, suggest that group members are most likely to vary in their perceptions of the work environment when group members are demographically diverse and when social interaction and work interdependence among group members are low. Further, the wording of the survey measures used to assess group members’ perceptions of the work environment may also influence the heterogeneity observed in group members’ survey responses.

Demographic Homogeneity

Organizational demography theory and research (e.g., Kanter, 1977; Pfeffer, 1983; Wiersema & Bird, 1993; Williams & O’Reilly, 1998) suggest that group members who are similar in their demographic characteristics are likely to share similar backgrounds, experiences, values, and beliefs. This similarity of backgrounds, experiences, values, and beliefs in turn fosters attraction and interaction among group members, further strengthening the homogeneity of group members’ experiences, values, and beliefs. Accordingly, the greater the demographic homogeneity of a group, the more likely group members are to perceive, interpret, and evaluate their work environment in a similar fashion. In contrast, demographic heterogeneity among group members may yield dissimilarity of values and beliefs, conflict, and turnover (e.g., Jackson et al., 1991; O’Reilly, Caldwell, & Barnett, 1989; Wiersema & Bird, 1993). Thus, we predicted the following:

Hypothesis 1: The greater demographic heterogeneity (with respect to age, education, pay, tenure, and gender) within a group, the greater the variability in group members’ perceptions of the work environment.

The proposed relationship between demographic homogeneity and perceptual agreement has not, to our knowledge, been previously tested.

Social Interaction and Work Interdependence

A large body of organizational theory and research (e.g., Berger & Luckmann, 1966; Weick, 1979) suggests that interaction among the members of a group fosters similarity among group members’ beliefs and perceptions. Walsh and Ungson (1991, p. 60) suggested that an organization itself is a “network of intersubjectively shared meanings that are sustained through the development and use of a common language and everyday social interaction.” In a similar vein, Schneider and Reichers (1983, p. 31) commented, “It is through social interactions that individuals in the work place come to have similar perceptions of that context.” Hambrick (1994) argued that high behavioral integration (e.g., information exchange, collaborative behavior) among the members of a top management team enhances perceptual agreement and speed of decision making. Social information processing theory (e.g., Salancik & Pfeffer, 1978; Thomas & Griffin, 1989) suggests that group members discuss their perceptions of work events and thereby reach a shared interpretation of the work environment.

The results of Rentsch’s (1990) study of workplace interaction and organizational meaning lend credence to these arguments. Using network analysis to operationalize friendship and workflow interactions and multidimensional scaling to operationalize similarity of meaning, Rentsch found that people involved in the same interaction groups attached similar meanings to organizational events, whereas people involved in different interaction groups attached different meanings to these events. Krackhardt and Kilduff (1990) found similar results and concluded that people who work together perceive work environments similarly. We thus predicted the following:
Hypothesis 2: The greater social interaction among the members of a group, the less the variability in group members’ perceptions of the work environment.

Hypothesis 3: The greater the work interdependence among the members of a group, the less the variability in group members’ perceptions of the work environment.

The Wording of the Survey Instrument

Several authors (e.g., Klein et al., 1994; Payne, Fineman, & Wall, 1976; Schneider, 1990; Schriesheim, 1979; Yammarino, 1990) have suggested that survey wording may influence the extent of homogeneity observed in group members’ ratings of the work environment. More specifically, survey items that contain wording directing respondents’ attention to their individual experiences (e.g., “I” or “my”) may encourage respondents to look within, disregarding their observations of others’ experiences. Conversely, survey items that contain wording directing respondents’ attention to the common experiences of the group (e.g., “we” or “employees here”) may encourage respondents to assume the shared perspective of the group in completing survey items. In Chan’s (1998) typology, survey items of the first type, if aggregated to the unit level, represent direct consensus composition. Survey items of the second type, if aggregated, represent referent-shift consensus composition. Chan’s typology suggests that a researcher’s choice of survey wording may determine the precise nature of the construct he or she has actually measured. For example, direct consensus items measuring climate for safety and referent-shift items measuring climate for safety may in fact measure two subtly different constructs. A finding that items differing only in their referent (“we” vs. “I”) differ also in observed within-group agreement may support Chan’s (1998) view.

A second aspect of survey wording may also influence observed within-group agreement. Survey items asking employees to describe objective characteristics of the work environment may encourage employees to assume a relatively factual, reportorial stance. Conversely, survey items asking employees to evaluate the subjective characteristics of their work environment may encourage employees to apply their personal values in interpreting work events. Thus, relatively objective items may engender more within-group agreement than do evaluative items. In the first case, respondents indicate whether relatively objective and neutral descriptions of the work environment are accurate or inaccurate. In the second case, respondents rate evaluative (positive or negative) descriptions of the work environment, in light of their own values, experiences, and expectations. Supporting these ideas, Glick (1985, p. 608) advised researchers studying organizational—rather than psychological—climate to ask “descriptive rather than affective questions” and to treat respondents “as key informants . . . describing organizational characteristics not as individual actors revealing their unique experiences.” We predicted the following:

Hypothesis 4: Survey items containing an individual referent yield greater variability in group members’ responses than do survey items using a group referent.

Hypothesis 5: Survey items requiring individuals to evaluate the subjective characteristics of their work environment yield greater variability in group members’ responses than do survey items requiring individuals to respond to objective descriptions of their work environment.

Further, we predicted that the two features of survey item wording would interact. Descriptive survey items may, as we have suggested, cause respondents to assume a more factual, reportorial, and objective stance than do evaluative survey items. If so, then participants’ responses to descriptive survey items with an individual referent may differ little from participants’ responses to descriptive survey items with a group referent (provided the items describe a feature of the work environment in fact shared by group members). Evaluative survey items, in contrast, may invite respondents to be more subjective and idiosyncratic in their responses. Accordingly, item referent may be more consequential for evaluative than descriptive survey items. Evaluative items with a group referent may direct respondents’ attention to group members’ shared views—their collective sense making. Evaluative items with an individual referent may cause a respondent to think only of his or her own views. Thus, we hypothesized the following:

Hypothesis 6: The effects of item referent on within-group variability are greater for evaluative survey items than for descriptive survey items. More specifically, use of a group referent (instead of an individual referent) fosters greater reductions in within-group variability in response to evaluative survey items than in response to descriptive survey items.

In summary, we tested likely antecedents of agreement in group members’ survey ratings of their work environment. Within-group agreement bears close study for several reasons. First, dispersion constructs are gaining new prominence in the organizational literature. Second, within-group homogeneity in employee perceptions may influence important group outcomes, including group cohesiveness, conflict, creativity, speed of decision making, and ease of decision implementation. And third, composition-based measures of the work environment lack construct validity in the absence of substantial within-group agreement. Very little research, however, has examined the antecedents of within-group agreement. The results of our research may inform a diverse array of scholars whose research and theories incorporate multilevel group consensus constructs, referent-shift consensus constructs, and/or dispersion constructs.

For the sake of clarity, we first present the method, results, and discussion of our tests of Hypotheses 1, 2, and 3 (regarding demographic heterogeneity, social interaction, and work interdependence). We then present the method, results, and discussion of our tests of Hypotheses 4, 5, and 6 (regarding survey wording). This format is unorthodox but, we believe, helpful given that (a) the two sets of independent variables are of a fundamentally different character (the first describes group characteristics, whereas the second set describes features of the survey wording); (b) we used a naturalistic, field design to test Hypotheses 1, 2, and 3 and an experimental design to test Hypotheses 4, 5, and 6; and (c) our constructs, measures, and results are quite complex. We conclude with a brief general discussion of our overall findings.
WITHIN-GROUP AGREEMENT

Method: Hypotheses 1, 2, and 3

Overview and Sample

Our data are drawn from a recent study of 42 manufacturing plants, within 35 companies, that have implemented the same computerized manufacturing resource planning (MRP) system (Klein, Conn, & Sorra, in press). Within each plant, we collected survey data from three groups: (a) the management group (managers and supervisors who directly or indirectly supervise MRP users); (b) the implementation team (employees responsible for planning MRP training and supporting MRP implementation and use); and (c) MRP users (production, purchasing, scheduling, customer service, distribution, and/or accounting personnel who used the MRP system to perform many of their day-to-day tasks). A key respondent (typically the plant manager or implementation team leader) at each plant identified the members of each of these groups.

To test Hypotheses 1-5, we used three work-environment measures that were completed by both the management group and the implementation team in each plant: (a) Plant Innovativeness; (b) Financial Resource Availability (General); and (c) Financial Resource Availability (for the MRP implementation). We excluded groups with fewer than 3 respondents. Our resulting sample for the tests of Hypotheses 1-5 consisted of 65 groups of managers or of implementation team members (n of individuals = 419). The average number of respondents per management group was 7.33 (SD = 8.34). The average number of respondents per implementation team was 5.45 (SD = 3.15). The average response rate for each group was 68%.

Measures

Demographic heterogeneity. We calculated the coefficient of variation (the standard deviation divided by the mean) to assess group heterogeneity in the four continuous demographic variables: age (measured in years), plant tenure (measured in years), pay (measured in eight increments: 1 = Less than $10,000 a year, 8 = More than $70,000 a year), and education (measured in eight increments: 1 = Some elementary school; grades 1-7, 8 = Graduate degree: Masters, Ph.D., J.D., M.D., etc.). We measured respondents’ ethnicity, but the ethnic composition of the groups showed too little between-group variability to justify the inclusion of demographic heterogeneity with respect to ethnicity as a predictor. The greater the coefficient of variation, the greater the variability in group members’ demographic composition. 1

To assess each group’s gender composition, we calculated Blau’s (1977) index of heterogeneity (used to calculate the heterogeneity of categorical variables). The index ranges from 0 (if all group members are male or all are female) to .5 (in the case of gender and other dichotomous variables, when group members are equally divided across the two categories). The larger Blau’s index, the greater the variability in the gender composition of the group (i.e., the more equal the representation of men and of women).

Social interaction. The 9-item Social Interaction scale asked group members to describe how frequently they interacted with each other as friends. Questions were worded to fit the respondents’ group. This measure included five questions, such as “How often do managers and supervisors [implementation team members] get together with other managers and supervisors [implementation team members] outside of work?” Respondents answered using a 5-point response scale (1 = not at all, 5 = very much). Respondents also indicated the agreement on four additional items using a 5-point response scale (1 = not true, 5 = true), such as “Managers and supervisors [implementation team members] here are good friends.”

Alpha for the scale was .85.

To test Hypothesis 4, we aggregated the Social Interaction scale to the group level. In Chan’s (1998) typology, Social Interaction is a referent-shift consensus measure. Accordingly, we assessed the extent of within-group agreement in social interaction prior to aggregating the measure. The mean $r_{wg}$ was .95 (not including one group’s $r_{wg}$, which, with a value over 1.00, was out of range). Out-of-range $r_{wg}$ scores occur, as James, Demaree, and Wolf (1984) noted, when the obtained variance within a group exceeds the expected variance. We also conducted a one-way analysis of variance (ANOVA), predicting social interaction scores on the basis of plant and group membership. The resulting $F^{2}$ was .37 ($p < .01$). These results suggest that the Social Interaction scale shows adequate within-group agreement and between-group differentiation to be aggregated to the group level.

Work interdependence. The 6-item measure of the Work Interdependence scale required group members to rate the extent to which they coordinated with and depended on each other to accomplish their work tasks. This measure used a 3-point response scale (1 = not at all, 5 = very much). Questions were worded to fit the respondents’ group. For example, two questions were “How much must managers and supervisors [implementation team members] coordinate their work activities with other managers and supervisors [implementation team members] to get their jobs done?” and “How much do managers and supervisors [implementation team members] here perform their work tasks as a team?” Alpha for the scale was .80.

To test Hypothesis 5, we aggregated the Work Interdependence scale to the group level—consistent, as above, with our conceptualization of Work Interdependence as a referent-shift consensus measure. The mean $r_{wg}$ was .88 (not including one group’s $r_{wg}$, which, with a negative value, was out of range) and the $F^{2}$ from a one-way ANOVA was .26 ($p < .01$).

Within-group variability in survey responses. To assess within-group variability, our dependent variable in Hypotheses 1-5, we calculated the average within-group standard deviation for each item of the three work environment survey scales: Plant Innovativeness, Financial Resource Availability (General), and Financial Resource Availability (MRP), respectively. To measure plant innovativeness, we developed a 9-item scale assessing respondents’ perceptions of their plant’s openness to change (e.g., “This plant can be described as flexible and adaptable to change”). Alpha for the scale was .83. To measure financial resource availability (general), we created a 5-item scale assessing respondents’ perceptions of the availability of slack financial resources for the plant (e.g., “This plant is under intense budget pressure to cut costs”). Alpha for this scale was .86. Finally, to measure financial resource availability (MRP), we used a 7-item scale assessing respondents’ perceptions of the availability of financial resources to support the implementation of MRP within the plant (e.g., “In this plant, money has been readily available to support activities related to the implementation of MRP”). Alpha for this scale was .86. Finally, to measure financial resource availability (MRP), we used a 7-item scale assessing respondents’ perceptions of the availability of financial resources to support the implementation of MRP within the plant (e.g., “In this plant, money has been readily available to support activities related to the implementation of MRP”). Alpha for this scale was .86. Each measure used a 5-point Likert response scale (1 = not true, 5 = true).

To calculate a group’s average standard deviation for each scale, we calculated the standard deviation of group members’ responses to each item of each scale. We then averaged the item standard deviations, yielding

1 As Allison (1978) recommended in his review of measures of within-group heterogeneity, the coefficient of variation is widely used in studies of demographic variability (e.g., Jackson et al., 1991; O’Reilly, Caldwell, & Barnett, 1989; Wiersema & Bird, 1993). Allison (1978) advocated use of the coefficient of variation because it, unlike the variance and standard deviation, provides a measure of relative, not absolute, variability. The coefficient of variation is not scale invariant. Allison (1978) provided the following example: Suppose three individuals have incomes of $5,000, $15,000, and $25,000. The differences among their incomes are, of course, substantial. However, if each individual receives an additional million dollars, the differences among the three individuals’ incomes become trivial. The standard deviation of the three individuals’ incomes, however, is identical in both cases; the coefficient of variation is not.
three standard deviation scores—for innovativeness, financial resource availability (general), and financial resource availability (MRP)—for each group in the sample. The larger a group’s average within-group standard deviation, the greater the variability in group members’ responses to the 

items of the survey scale.

Bliese and Halverson (1998) used a very similar procedure in their study of group consensus, arguing that assessing agreement in each item of a scale and then averaging the results leads to less range restriction than does assessing agreement in the scale scores. Bliese and Halverson (1998), however, calculated mean item variance, rather than the mean item standard deviation. Taking the square root of the variance (i.e., using the standard deviation rather than the variance) may normalize the distribution and thus yield more robust estimates. However, for our three work environments, the average group-level correlation between the average within-group standard deviation of the scale items and the average within-group variance of the scale items was .95. Not surprisingly, then, our results were essentially identical when we used the average within-group item standard deviation of the work environment scales and when we used the average within-group item variance of the same scales. Finally, we note that the average within-group standard deviation is strongly positively correlated with James et al.’s (1984) $r_{wg}$. The correlation between the two indices falls below .8 or .9 only when the value of $r_{wg}$ for a given group on a given scale falls outside the expected $r_{wg}$ range of 0 to 1.0.

**Analyses**

We performed correlations and regressions to assess the relationship between the independent variables (demographic heterogeneity, group tenure, work interdependence, and social interaction) and the dependent variables, the average within-group standard deviations of plant innovativeness, financial resource availability (general), and financial resource availability (MRP). Our unit of analysis was the group for Hypotheses 1–5 ($n = 65$).

**Results: Hypotheses 1, 2, and 3**

**Demographic Heterogeneity**

Table 1 shows the means, standard deviations, and correlations among all the variables used to test Hypotheses 1, 2, and 3. The results provide little support for Hypothesis 1. The heterogeneity of group members’ tenure is, as predicted, significantly positively related to variability in group members’ responses to the Plant Innovativeness scale ($r = .35$, $p < .01$). However, this is the only significant correlation of the 15 correlations tested (3 Measures of Within-Group Variability $\times$ 5 Measures of Demographic Heterogeneity) and, therefore, may be due to chance.

**Social Interaction and Work Interdependence**

The results in Table 1 provide substantial support for Hypotheses 2 and 3. As predicted, the greater group members’ reported social interaction, the smaller the variability in group members’ responses to both the Financial Resource Availability (MRP) scale ($r = -.28$, $p < .05$) and the Plant Innovativeness scale ($r = -.43$, $p < .001$). Similarly, the greater group members’ reported work interdependence, the smaller the variability in group members’ responses to the Financial Resource Availability (General) scale ($r = -.32$, $p < .05$), the Financial Resource Availability (MRP) scale ($r = -.52$, $p < .001$), and the Plant Innovativeness scale ($r = -.33$, $p < .01$). Thus, five of the six correlations testing Hypotheses 4 and 5 (3 Measures of Within-Group Variability $\times$ 2 Measures of Group Interaction) are significant and in the predicted direction. The remaining correlation was in the predicted direction but did not achieve significance.

To assess the simultaneous influence of the two predictors, we regressed the within-group standard deviation of financial resource availability (MRP) on work interdependence and on social interaction. The resulting equation is statistically significant, $F(2, 64) = 11.63$, $R^2 = .27$, $p < .001$. However, work interdependence is the only statistically significant predictor ($\beta$ for work interdependence $= -.51$, $p < .001$ and $\beta$ for social interaction $= -.01$, $p > .05$). In a regression of the within-group standard deviation of plant innovativeness on the two predictors, the overall equation is statistically significant, $F(2, 64) = 7.74$, $R^2 = .20$, $p < .001$. However, social interaction is the only statistically significant predictor ($\beta$ for social interaction $= -.36$, $p < .01$ and $\beta$ for work interdependence $= -.14$, $p > .05$). Given the substantial correlation between the aggregated measures of social interaction and work interdependence ($r = .46$, $p < .001$), it is perhaps not

---

**Table 1**

**Means, Standard Deviations, and Zero-Order Correlations for Hypotheses 1–5**

<table>
<thead>
<tr>
<th>Variable</th>
<th>M</th>
<th>SD</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Tenure variability</td>
<td>0.61</td>
<td>0.24</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Age variability</td>
<td>0.16</td>
<td>0.07</td>
<td>0.10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Pay variability</td>
<td>0.21</td>
<td>0.10</td>
<td>0.03</td>
<td>0.14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Education variability</td>
<td>0.17</td>
<td>0.08</td>
<td>0.11</td>
<td>0.04</td>
<td>0.11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Gender variability</td>
<td>0.24</td>
<td>0.21</td>
<td>0.03</td>
<td>0.08</td>
<td>0.26*</td>
<td>-.12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Group tenure</td>
<td>9.15</td>
<td>4.09</td>
<td>-.04</td>
<td>0.04</td>
<td>0.04</td>
<td>-0.2</td>
<td>.17</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Social interaction</td>
<td>3.18</td>
<td>0.40</td>
<td>0.05</td>
<td>0.21</td>
<td>0.03</td>
<td>-0.14</td>
<td>-.24</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Work interdependence</td>
<td>3.41</td>
<td>0.25</td>
<td>0.08</td>
<td>0.06</td>
<td>0.13</td>
<td>-0.14</td>
<td>.46***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. Mean SD of plant innovation</td>
<td>0.91</td>
<td>0.21</td>
<td>.35**</td>
<td>0.16</td>
<td>0.21</td>
<td>0.05</td>
<td>-0.05</td>
<td>.17</td>
<td>-0.43***</td>
<td>-.33**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. Mean SD of financial res. (Gen’l)</td>
<td>0.95</td>
<td>0.31</td>
<td>0.00</td>
<td>0.02</td>
<td>-0.13</td>
<td>-0.10</td>
<td>.20</td>
<td>.19</td>
<td>-0.10</td>
<td>-.32*</td>
<td>.09</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. Mean SD of financial res. (MRP)</td>
<td>1.04</td>
<td>0.33</td>
<td>0.07</td>
<td>0.00</td>
<td>-0.13</td>
<td>-0.02</td>
<td>-0.04</td>
<td>.20</td>
<td>-0.28*</td>
<td>-.52***</td>
<td>.40***</td>
<td>.45***</td>
<td></td>
</tr>
</tbody>
</table>

*Note. N = 65. res. = resource; Gen’l = general; MRP = manufacturing resource planning.  
* $p \leq .05$. ** $p \leq .01$. *** $p \leq .001$. 

---


surprising that only one of the predictors is statistically significant in the regression equations.

Discussion: Hypotheses 1, 2, and 3

The results provide minimal support for Hypothesis 1 (demographic heterogeneity) but substantial support for Hypotheses 2 and 3 (social interaction and work interdependence). Contrary to Hypothesis 1, we found very little evidence that demographic heterogeneity among the members of a group yields within-group variability in survey responses. Hypothesis 1, and much of the organizational demography literature, rests on the assumption that demographic homogeneity leads to group member attraction, cohesion, and interaction, which in turn engenders agreement among group members' values, attitudes, and actions. In our study, and contrary to prior research (e.g., O’Reilly et al., 1989), demographic heterogeneity was not significantly related to social interaction or work interdependence, nor was it related to within-group agreement on survey measures assessing perceptions of the work environment. Some recent theory and research (Harrison et al., 1998; Pelled, 1996) suggest that the effects of the demographic composition of a group may lessen over time, as group members interact and share more experiences. Unfortunately, we could not test whether group tenure moderated the relationship between demographic heterogeneity and within-group agreement as our data included no measures of group tenure (only measures of plant tenure).

In support of Hypotheses 2 and 3, social interaction was significantly negatively related to within-group variability in two of the three survey measures, and work interdependence was significantly negatively related to within-group variability in each of the three survey measures. Social interaction and work interdependence were significantly positively correlated, as might be expected given that both measures tap similar constructs, were completed by the same respondents, and were aggregated to the group level. The results suggest, however, that interaction—whether work-related, social, or both—among the members of group fosters similarity in group members' views, as evidenced in their survey responses. Alternatively, individuals may choose to socialize and work most closely with individuals with whom they share common perceptions. In short, the causal arrow may flow (a) from interaction to perceptual agreement, (b) from perceptual agreement to interaction, or (c) in both directions.

Method: Hypotheses 4, 5, and 6

Overview and Sample

To test Hypothesis 5 (survey item referent: individual or group), Hypothesis 6 (survey item focus: descriptive or evaluative), and Hypothesis 7 (the interaction of item referent and item focus), we conducted an experiment within our study of the implementation of MRP in manufacturing plants. More specifically, we created four sets of survey items (54 items in all) that varied in their referent (individual or group) and their focus (descriptive or evaluative): (a) survey items with an individual referent and a descriptive focus; (b) survey items with a group referent and a descriptive focus; (c) survey items with an individual referent and an evaluative focus; and (d) survey items with a group referent and an evaluative focus. We randomly assigned groups to receive one of the four sets of survey items. All the members of a given group within a given plant received and completed just one of the four sets of the items. Research participants were aware that we were studying the implementation of MRP; they were not aware of our study of the effects of survey wording on within-group agreement. Our sample for the tests of Hypotheses 4, 5, and 6 consisted of 86 groups of managers, or of implementation team members, or of MRP users (n of individual respondents = 690). Once again, groups with fewer than 3 respondents were eliminated from the sample.

Survey Items

The four sets of survey items that we created to test Hypotheses 4, 5, and 6 are listed in the Appendix. (The Appendix lists the items for the user group only.) Survey items pertain either to respondents' work or to respondents' supervisors. We created work-related and supervisor-related survey items simply because we hoped to test a variety of items. Work-related and supervisor-related survey items are, of course, very common within studies of numerous aspects of the work environment. In our analyses, we have treated survey-item content (work or supervisor) as a control variable, not a predictor variable.

The individual-referent and group-referent survey items are identical with the exception of the referent, of course, and necessary grammatical changes. We matched the group referent to the specific group of respondents. Thus, we used the group referent "managers and supervisors here" in surveys completed by the management group, the group referent "MRP implementation team members" for the implementation team, and the group referent "employees" for MRP users. The evaluative and descriptive items, by definition, are distinct. In creating items that we designated evaluative, we sought to write items that required the respondent to provide a subjective assessment of the quality of the respondent(s)' work or supervisor. In creating items that we designated descriptive, we sought to write items that required the respondent to provide a factual report (not an evaluation) of a readily observable aspect of the work or of the supervisor(s)' behavior. The items do not and were not intended to form meaningful scales. Thus, for example, the work-related, group-referent descriptive items capture diverse aspects of group members' work, not a single coherent dimension.

To contrast descriptive and evaluative survey items, we coded descriptive items -1 and evaluative items 1. To contrast individual and group referents, we coded items with an individual referent -1 and items with a group referent 1. We coded work-related survey items 1 and supervisor-related survey items -1.

Analyses

To test Hypotheses 4, 5, and 6, we conducted a hierarchical regression in which we regressed within-group variability (the average within-group standard deviation) on the following variables, in this order: (a) item content (work or supervisor), a control variable; (b) item referent (individual or group); (c) item focus (descriptive or evaluative); and (d) the interaction of item referent and item focus. Our unit of analysis was the survey item (n = 54 survey items). We explored the results in a series of post hoc analyses explained below.

---

2 We tested whether group members' average plant tenure had a main or moderating effect on within-group agreement in response to survey measures assessing perceptions of the work environment. Group members' average plant tenure was not significantly related to within-group agreement on the survey measures. Further, plant tenure did not show a clear and consistent moderating effect on the relationship between demographic heterogeneity and within-group agreement. Only 2 of the 15 tests of the moderating effects of plant tenure were significant. Further, the significant results were difficult to interpret, in one case supporting predictions based on recent theory and research, but in the other case countering these predictions.
Results: Hypotheses 4, 5, and 6

As shown in Table 2, the control variable, item content (work or supervisor), is not significantly related to within-group variability. Further, the results reported in Table 2 provide no support for Hypothesis 4; within-group variability does not vary significantly as a function of item referent (individual or group; \( \Delta R^2 = .00, p > .05 \)). Item focus (descriptive or evaluative) is, however, significantly related to the within-group variability of survey responses (\( \Delta R^2 = .28, p < .001 \)). These results counter Hypothesis 5: Within-group variability is significantly smaller for survey items with an evaluative focus than for items with a descriptive focus. Finally, the interaction of item referent (individual or group) and item focus (descriptive or evaluative) is statistically significant (\( \Delta R^2 = .11, p < .001 \)). We had expected the use of group rather than individual referent to lead to a greater reduction in within-group variability in response to evaluative items than in response to descriptive items. The results do not support this hypothesis. Rather, the results indicate that use of group referent reduces within-group variability in response to descriptive items but increases within-group variability in response to evaluative items. Together, the predictors explain 45% of the variability in within-group agreement in response to our survey items.

Post Hoc Analyses

To gain a better understanding of the unexpected effects of survey wording on within-group agreement, we conducted a series of exploratory, post hoc analyses. We began by examining survey items that yielded the smallest within-group standard deviations. The four items showing the smallest within-group variability are (a) "My work here is meaningless" (average within-group standard deviation = .41); (b) "The supervisor to whom I report is impolite to me" (average within-group standard deviation = .47); (c) "My work here is boring" (average within-group standard deviation = .61); and (d) "My work here is dissatisfying" (average within-group standard deviation = .63). (As a point of reference, consider the survey item "The supervisor to whom I report is good at motivating me." Its within-group standard deviation is one of the largest of our items: 1.89. The average within-group standard deviation across all of our items is .97.) Contrary to our hypotheses, but consistent with the results described above, the four survey items showing the least within-group variability are all evaluative with an individual referent. Further, the items are strongly negative or pejorative in tone, describing rare, socially undesirable feelings and events. We speculated that these items show little within-group variability because a majority of respondents disagreed with these items, marking these items not true or slightly true. Indeed, the average employee response on our 5-point scale to each of these four items was 1.7 or less.

We thus wondered whether, in our effort to write clearly descriptive and clearly evaluative items, we had written items varying also in social desirability. If so, the varying social desirability of our items may have obscured and confused the effects of item referent and item focus. To investigate this possibility, we asked 24 part-time MBA students who were blind to our hypotheses to rate the social desirability of each of our 54 items.5 More specifically, we adopted Holtgraves, Eck, and Laskey’s (1997) procedure, asking the 24 MBA students to indicate "How desirable is it for people to answer this item ‘True’?" for all 54 items (1 = extremely undesirable; 5 = neither desirable or undesirable; 9 = extremely desirable). The MBA students reported that responding True on the negatively toned, evaluative items was socially undesirable (see Table 3). Table 4 shows the correlations between item social desirability and the other variables used in our post hoc analyses.

We retested Hypotheses 4, 5, and 6, after first controlling for the social desirability of each item and for item content. The results of the hierarchical regression appear in Table 5. As we suspected, the greater an item’s social desirability, the greater the item’s within-group variability (\( \Delta R^2 = .30, p < .001 \)). That is, socially undesirable items yield greater within-group agreement than socially desirable items. As in our initial analysis, neither item content (\( \Delta R^2 = .00, p > .05 \)) nor item referent (\( \Delta R^2 = .00, p > .05 \)) is significantly related to within-group variability. Even after we have controlled for social desirability, item focus remains significantly related to within-group variability (\( \Delta R^2 = .21, p > .001 \)). That is, as in our initial analysis and contrary to our hypothesis, evaluative items show less within-group variability than do descriptive items. Finally, the interaction of item referent and item focus is statistically significant (\( \Delta R^2 = .10, p > .001 \)), indicating—as in our initial analysis—that the use of a group rather than individual referent fosters a decrease in within-group variability in response to descriptive items but an increase in within-group variability in response to evaluative items. Together, the predictors explain 63% of the variability in within-group agreement in response to our survey items.

In a final post hoc analysis, we examined the effects of survey wording on between-group variability in survey responses. When researchers gather individual-level survey data from group members to assess consensus constructs, they hope to find substantial within-group agreement and substantial between-group variability. In the absence of substantial within-group agreement, group consensus measures lack construct validity. But, in the absence of substantial between-group variability, the measures are unlikely to be significantly related to other measures varying between groups. Reviewing the results of original hypothesis-testing and post hoc analyses, we speculated that items low in social desirability would show not only low within-group variability but also low between-group variability.

To explore this issue, we performed one-way ANOVAs of each of the 54 items. The 54 \( \eta^2 \) values resulting from the 54 ANOVAs indicate the magnitude of the between-group variability in individuals’ responses to each of the 54 items. That is, each \( \eta^2 \) indicates the extent to which group membership explains variability in the individuals’ responses to a given item. We used \( \eta^2 \) from each one-way ANOVA as our new dependent variable in a hierarchical regression in which we examined the effects of survey wording on between-group variability in responses to the survey items. This analysis provides an initial answer to the following question: Why do some of our items generate greater between-group variability than do other items? Although our use of \( \eta^2 \) as a dependent variable may appear unusual, our analysis is analogous

3 Twelve of the students were women and 12 were men. The students were 25 to 33 years old (\( Mdn = 28, M = 28.43 \)). They had worked full time for 4 to 10 years (\( Mdn = 6, M = 6.26 \)).
to meta-analysis, hierarchical linear modeling, and other analyses (e.g., statistical process control) in which parameter estimates from prior analyses are used as dependent variables in subsequent analyses. Researchers use meta-analysis to identify the correlates of correlation coefficients and effect sizes obtained in previous studies (e.g., Cooper & Hedges, 1994; Hunter & Schmidt, 1990). Researchers use hierarchical linear modeling to identify the correlates of intercepts and slopes obtained in preceding steps of the analysis (Bryk & Raudenbush, 1992; Kreft & deLeeuw, 1998). In a similar vein, we used hierarchical regression to identify correlates of between-group variability ($\eta^2$) in the survey items.

The results appear in Table 6. Social desirability is not significantly related to between-group variability in responses to the item ($R^2 = .00, p > .05$). Item content is, however, significantly related to between-group variability ($\Delta R^2 = .16, p < .01$); items describing respondents' work yield greater between-group variability than do items describing respondents' supervisors. Further, item referent is significantly related to between-group variability ($\Delta R^2 = .10, p < .05$); items with a group referent show significantly greater between-group variability than do items with an individual referent. The effect of item focus (descriptive or evaluative) approaches statistical significance ($\Delta R^2 = .05, p = .06$); evaluative items generate less between-group variability than do descriptive items. Finally, the interaction of item focus and item referent approaches statistical significance, as well ($\Delta R^2 = .05, p = .07$). The interaction indicates that the use of group rather than individual referent generates a greater increase in between-group variability in response to descriptive survey items than in response to evaluative survey items. Overall, the predictors explained 36% of the variability in the extent to which our survey items varied between groups.

### Table 2

<table>
<thead>
<tr>
<th>Step</th>
<th>Variable</th>
<th>B</th>
<th>SE B</th>
<th>β</th>
<th>$\Delta R^2$</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Item content (work or supervisor)</td>
<td>.05</td>
<td>.03</td>
<td>.24</td>
<td>.06</td>
<td>.06</td>
</tr>
<tr>
<td>2</td>
<td>Item content (work or supervisor)</td>
<td>.05</td>
<td>.03</td>
<td>.24</td>
<td>.00</td>
<td>.06</td>
</tr>
<tr>
<td></td>
<td>Item referent (individual or group)</td>
<td>-.01</td>
<td>.03</td>
<td>-.04</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Item content (work or supervisor)</td>
<td>.04</td>
<td>.02</td>
<td>.22</td>
<td>.28***</td>
<td>.34***</td>
</tr>
<tr>
<td></td>
<td>Item referent (individual or group)</td>
<td>-.01</td>
<td>.02</td>
<td>-.04</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Item focus (descriptive or evaluative)</td>
<td>-.11</td>
<td>.02</td>
<td>-.53***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Item content (work or supervisor)</td>
<td>.04</td>
<td>.02</td>
<td>.22*</td>
<td>.11**</td>
<td>.45***</td>
</tr>
<tr>
<td></td>
<td>Item referent (individual or group)</td>
<td>-.01</td>
<td>.02</td>
<td>-.05</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Item focus (descriptive or evaluative)</td>
<td>-.11</td>
<td>.02</td>
<td>-.53***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Item Referent $\times$ Item Focus</td>
<td>.07</td>
<td>.02</td>
<td>.33**</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note. N = 54 items. The coding was as follows: for item referent, individual = -1, group = 1; for item focus, descriptive = -1, evaluative = 1; and for item content, work = -1, supervisor = 1. * $p \leq .05$. ** $p \leq .01$. *** $p \leq .001$.

### Table 3

<table>
<thead>
<tr>
<th>Item</th>
<th>Mean social desirability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Most socially desirable</td>
<td></td>
</tr>
<tr>
<td>1. The supervisor to whom I report praises me for excellent performance.</td>
<td>ID 7.96</td>
</tr>
<tr>
<td>2. My work here is enjoyable.</td>
<td>IE 7.38</td>
</tr>
<tr>
<td>3. My work here is motivating.</td>
<td>IE 7.17</td>
</tr>
<tr>
<td>4. The supervisor to whom I report meets with me to discuss my work goals.</td>
<td>ID 7.12</td>
</tr>
<tr>
<td>5. Employees' work here is motivating.</td>
<td>GE 7.08</td>
</tr>
<tr>
<td>6. My work here is rewarding.</td>
<td>IE 7.08</td>
</tr>
<tr>
<td>Least socially desirable</td>
<td></td>
</tr>
<tr>
<td>1. My work here is meaningless.</td>
<td>IE 2.04</td>
</tr>
<tr>
<td>2. Employees' work here is meaningless.</td>
<td>GE 2.17</td>
</tr>
<tr>
<td>3. My work here is dissatisfying.</td>
<td>IE 2.46</td>
</tr>
<tr>
<td>4. Employees' work here is irritating.</td>
<td>GE 2.67</td>
</tr>
<tr>
<td>5. My work here is boring.</td>
<td>IE 2.71</td>
</tr>
<tr>
<td>6. My work here is irritating.</td>
<td>IE 2.79</td>
</tr>
</tbody>
</table>

*Note. ID = individual, descriptive; GD = group, descriptive; IE = individual, evaluative; GE = group, evaluative.

### Discussion: Hypotheses 4, 5, and 6

The results of our initial and post hoc analyses reveal the complex and substantial effects of survey wording on within-group agreement and between-group variability in employee responses to survey items. In this section, we review the predictors of within-group agreement and of between-group variability and then comment on lessons learned and limitations.

Within-group agreement in response to survey items is a function of item content and item referent. Although we manipulated the social desirability of survey items inadvertently, we found that socially undesirable items—items with which, not surprisingly, a large proportion of respondents disagreed—yielded substantial within-group agreement. Contrary to hypothesis regarding item focus, our evaluative survey items yielded greater within-group agreement than did our descriptive survey items, even after we controlled for social desirability. Finally, item focus
Table 4

Means, Standard Deviations, and Zero-Order Correlations for Hypotheses 6–8 and Post Hoc Analyses

<table>
<thead>
<tr>
<th>Variable</th>
<th>M</th>
<th>SD</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Within-group variability</td>
<td>0.97</td>
<td>0.20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Between-group variability</td>
<td>0.23</td>
<td>0.10</td>
<td>.03</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Item content (job or supervisor)</td>
<td>-0.11</td>
<td>1.00</td>
<td>.24</td>
<td>- .38**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Item referent (individual or group)</td>
<td>0.00</td>
<td>1.00</td>
<td>-.04</td>
<td>.32*</td>
<td>.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Item focus (descriptive or evaluative)</td>
<td>0.04</td>
<td>1.00</td>
<td>-.54**</td>
<td>-.23</td>
<td>-.03</td>
<td>.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Social desirability</td>
<td>5.40</td>
<td>1.74</td>
<td>-.55**</td>
<td>-.01</td>
<td>.37**</td>
<td>- .02</td>
<td>- .16</td>
<td></td>
</tr>
</tbody>
</table>

Note. N = 54 items. The coding was as follows: for individual referent, individual = -1, group = 1; for item focus, descriptive = -1, evaluative = 1; and for item content, work = -1, supervisor = 1.

* p ≤ .05. ** p ≤ .01. *** p ≤ .001.

moderated the effect of item referent (individual or group) on within-group agreement. The use of a group (versus individual) referent reduced within-group variability in response to descriptive items but increased within-group variability in response to evaluative items.

Between-group variability in employee responses to the survey items was predicted, we found, by item content, item focus, and item referent. Item social desirability was not significantly related to between-group variability in employee survey responses. Item content was significantly related to between-group variability; work-related items yielded greater between-group variability than did supervisor-related items. We included item content as a control variable, making no hypotheses about its effects on within-group or between-group variability. Our results indicate, however, that respondents' perceptions of their work varied more substantially between groups than did respondents' perceptions of their supervisors. Item referent was a significant predictor of between-group variability; items with a group referent yielded greater between-group variability than did items with an individual referent. The effect of item focus approached statistical significance; descriptive survey items generated greater between-group variability than did evaluative survey items. Finally, item focus (evaluative or descriptive) had a nearly statistically significant moderating effect on the relationship of item referent and between-group variability. The use of a group (versus individual) referent increased between-group variability in response to descriptive items more than it increased between-group variability in response to evaluative survey items.

Table 5

Summary of Post Hoc Hierarchical Regression of Average Within-Group Variability on Survey Item Characteristics, Controlling for Social Desirability

<table>
<thead>
<tr>
<th>Variable</th>
<th>B</th>
<th>SE B</th>
<th>β</th>
<th>ΔR²</th>
<th>R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 1</td>
<td></td>
<td></td>
<td></td>
<td>.30***</td>
<td>.30***</td>
</tr>
<tr>
<td>Social desirability</td>
<td>.06</td>
<td>.01</td>
<td>.55***</td>
<td>.00</td>
<td>.31***</td>
</tr>
<tr>
<td>Step 2</td>
<td></td>
<td></td>
<td></td>
<td>.00</td>
<td>.31***</td>
</tr>
<tr>
<td>Social desirability</td>
<td>.06</td>
<td>.01</td>
<td>.54***</td>
<td>.00</td>
<td>.31***</td>
</tr>
<tr>
<td>Item content (work or supervisor)</td>
<td>-.01</td>
<td>.03</td>
<td>.04</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Step 3</td>
<td></td>
<td></td>
<td></td>
<td>.00</td>
<td>.31***</td>
</tr>
<tr>
<td>Social desirability</td>
<td>.06</td>
<td>.01</td>
<td>.54***</td>
<td>.00</td>
<td>.31***</td>
</tr>
<tr>
<td>Item content (work or supervisor)</td>
<td>-.01</td>
<td>.03</td>
<td>.04</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item referent (individual or group)</td>
<td>.01</td>
<td>.02</td>
<td>-.03</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Step 4</td>
<td></td>
<td></td>
<td></td>
<td>.21***</td>
<td>.51***</td>
</tr>
<tr>
<td>Social desirability</td>
<td>.05</td>
<td>.01</td>
<td>.46***</td>
<td>.01</td>
<td>.02</td>
</tr>
<tr>
<td>Item content (work or supervisor)</td>
<td>-.01</td>
<td>.02</td>
<td>-.03</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item referent (individual or group)</td>
<td>-.01</td>
<td>.02</td>
<td>-.03</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item focus (descriptive or evaluative)</td>
<td>-.09</td>
<td>.02</td>
<td>-.45**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Step 5</td>
<td></td>
<td></td>
<td></td>
<td>.10***</td>
<td>.63***</td>
</tr>
<tr>
<td>Social desirability</td>
<td>.05</td>
<td>.01</td>
<td>.45***</td>
<td>.01</td>
<td>.02</td>
</tr>
<tr>
<td>Item content (work or supervisor)</td>
<td>-.01</td>
<td>.02</td>
<td>-.04</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item referent (individual or group)</td>
<td>-.01</td>
<td>.02</td>
<td>-.04</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item focus (descriptive or evaluative)</td>
<td>-.05</td>
<td>.02</td>
<td>-.46**</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. N = 54 items. The coding was as follows: for individual referent, individual = -1, group = 1; for item focus, descriptive = -1, evaluative = 1; and for item content, work = -1, supervisor = 1.

*** p ≤ .001.
Table 6

Summary of Post Hoc Hierarchical Regression of Between-Group Variability on Survey Item Characteristics, Controlling for Social Desirability

<table>
<thead>
<tr>
<th>Variable</th>
<th>B</th>
<th>SE B</th>
<th>β</th>
<th>ΔR²</th>
<th>R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 1 Social desirability</td>
<td>.00</td>
<td>.01</td>
<td>-.01</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>Step 2 Social desirability</td>
<td>.01</td>
<td>.01</td>
<td>.15</td>
<td>.16**</td>
<td>.16*</td>
</tr>
<tr>
<td>Item content (work or supervisor)</td>
<td>-.05</td>
<td>.01</td>
<td>-.43**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Step 3 Social desirability</td>
<td>.01</td>
<td>.01</td>
<td>.16</td>
<td>.10*</td>
<td>.27**</td>
</tr>
<tr>
<td>Item content (work or supervisor)</td>
<td>-.05</td>
<td>.01</td>
<td>-.44**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item referent (individual or group)</td>
<td>.03</td>
<td>.01</td>
<td>.32*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Step 4 Social desirability</td>
<td>.01</td>
<td>.01</td>
<td>.12</td>
<td>.05†</td>
<td>.32***</td>
</tr>
<tr>
<td>Item content (work or supervisor)</td>
<td>-.04</td>
<td>.01</td>
<td>-.43**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item referent (individual or group)</td>
<td>.03</td>
<td>.01</td>
<td>.32**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item focus (descriptive or evaluative)</td>
<td>-.02</td>
<td>.01</td>
<td>-.23†</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Step 5 Social desirability</td>
<td>.01</td>
<td>.01</td>
<td>.12</td>
<td>.05†</td>
<td>.36***</td>
</tr>
<tr>
<td>Item content (work or supervisor)</td>
<td>-.05</td>
<td>.01</td>
<td>-.43**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item referent (individual or group)</td>
<td>.03</td>
<td>.01</td>
<td>.33**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item focus (descriptive or evaluative)</td>
<td>-.02</td>
<td>.01</td>
<td>-.23†</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item Referent X Item Focus</td>
<td>-.02</td>
<td>.01</td>
<td>-.22†</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. N = 54 items. The coding was as follows: for individual referent, individual = −1, group = 1; for item focus, descriptive = −1, evaluative = 1; and for item content, work = −1, supervisor = 1.

†p ≤ .05. *p ≤ .01. **p ≤ .01. ***p ≤ .001.

Our finding that the evaluative survey items engendered greater within-group agreement than the descriptive items did, even after controlling for social desirability, is surprising. Evaluative items would seem to permit—indeed, require—greater respondent subjectivity than descriptive survey items. The results of our post hoc analyses of the effects of survey wording on between-group variability are potentially instructive. These results suggest that although evaluative survey items may foster relatively high within-group agreement, they may also foster relatively low between-group variability—that is, relatively high agreement across an entire sample of groups.

We caution, however, that whereas our comparison of individual referent and group referent items was a clean comparison of identical items save for the referent, our comparison of evaluative items and descriptive items was not as clean. By definition, evaluative and descriptive items differ in their content. Further, we were surprised to discover how difficult it was to write purely descriptive survey items. In describing relatively positive or negative events (e.g., pay raises), descriptive items may take on an evaluative tone. And clearly descriptive items—such as “Employees here work in a building” (not one of our items)—may be so obvious that researchers find them useless and respondents find them silly. An additional consideration is the actual content of our descriptive items. The behaviors and events we specified in our descriptive items may in fact have varied within-groups more than we anticipated. For example, the members of a group may indeed vary substantially in the extent to which they talk or interact with others while doing their work. We suspect that many of the survey items used in organizational research fall midway between the endpoints of the evaluative—descriptive continuum. Climate items, for example, are likely to be somewhat descriptive and somewhat evaluative. Perhaps in future studies of the consequences of survey wording, researchers should not operationalize the descriptive—evaluative distinction as a dichotomy, as we did, but as a continuum.

Our analyses of the effects of item referent suggest, as previously proposed (e.g., Klein et al., 1994; Payne et al., 1976; Schneider, 1990) but not empirically tested, that the presence of a group referent may indeed encourage respondents to assume the shared perspective of the group when answering survey items—particularly descriptive survey items. Group referents may thus render survey items (again, particularly descriptive survey items) more likely to display the low within-group variability and high between-group variability expected of group-level consensus measures designed to tap group-level constructs. Should a researcher who seeks to measure a shared property of a unit thus use group referent rather than individual referents? The researcher’s choice must be based on theory—on the definition of the construct (Chan, 1998; Klein et al., 1994; Kozlowski & Klein, 2000). Indeed, our results provide some preliminary support for Chan’s (1998) suggestion that items differing only in referent (i.e., direct vs. referent shift consensus measures) may in fact assess closely related but nevertheless subtly different constructs.

Concluding Comments

During the past decade, as multilevel organizational scholarship has gained new prominence, increasing numbers of researchers and theorists have used group-level consensus and dispersion constructs within their work. Within-group agreement is a central consideration for researchers and theorists evoking consensus and dispersion constructs, yet within-group agreement itself has been the object of very little research. Indeed, we know of no other studies in which researchers have examined the antecedents of within-group agreement in perceptions of the work environment.
Consistent with social constructionism, social information processing theory, and related theory (e.g., Hambrick, 1994; Salancik & Pfeffer, 1978; Schneider & Reichers, 1983; Weick, 1979), our results suggest that social interaction and work interdependence among the members of a group fosters within-group agreement in perceptions of the work environment. Perhaps, however, within-group agreement predicts social interaction and work interdependence; group members who share common perceptions may enjoy interacting with each other and work particularly well together. Indeed, reciprocal causation seems likely. Our research is limited insofar as we cannot verify the precise causal relationship underlying the social interaction and work interdependence findings. In future research, longitudinal and experimental methods should be used to shed more light on the nature of the causal relationship.

Our demographic heterogeneity results, showing no significant relationship between these measures and within-group agreement, counter central but largely untested propositions within organizational demography theory. Inconsistencies are common within the organizational demography literature (Williams & O'Reilly, 1998), as the effects of demographic heterogeneity may vary as a function of the organizational context, the outcome variables of interest, and the specific demographic measures used. Given the growth of organizational demography research in recent years and the centrality of the proposition, within organizational demography theory, that demographic heterogeneity within a group yields within-group variability in perceptions, attitudes, and values, further study of the effects of demographic heterogeneity on within-group agreement is warranted.

Our results regarding survey wording highlight the complexity and considerable power of survey wording in fostering within-group agreement and between-group variability. Our results suggest that socially undesirable items may foster substantial within-group agreement precisely (if ironically) because most survey respondents disagree with these items. Further, our findings suggest that a group referent may engender between-group variability as well as within-group agreement (particularly in response to descriptive items). Clearly, we have only just begun to document the effects of survey wording on within-group agreement and between-group variability. Additional research is needed to examine the consequences of survey wording not only for within-group agreement and between-group variability but also for the predictive validity of measures varying in referent and focus.

We encourage researchers to be very precise in defining their group-level consensus constructs. Is the construct designed to capture group members' evaluations of a phenomenon or their descriptions of the phenomenon? Is the underlying composition model a direct consensus model or a referent shift model? The answers to these defining questions lay the groundwork for careful survey wording. In developing survey wording to operationalize a loosely defined construct, a researcher runs the risk of introducing considerable noise into his or her measures—noise that may influence the extent of within- and between-variability observed in the measures and thus the relationships between the focal measure and other measures.

Our findings, in sum, shed new light on the correlates of agreement among group members' perceptions of the work environment. Within-group agreement, as we have noted, is a key consideration in multilevel theory and research. Additional research is needed to replicate and extend the conclusions of our research. We have already noted the importance of longitudinal and experimental methods in verifying the antecedents of agreement. Also needed are studies of agreement among differing kinds of groups regarding a broad range of constructs. Further, we have not examined the effects of agreement among group members' perceptions of the work environment on group creativity, decision making, conflict, and morale. These consequences—much discussed, but little studied—merit further research.

References

WITHIN-GROUP AGREEMENT


(Appendix follows)
Appendix

Survey Items Used in Testing Hypotheses 3 and 4

**Individual-Referent, Descriptive Items for the User Group**

1. The work that I do in a typical day here is physically demanding.
2. To do my work, I typically talk on the telephone.
3. I typically do my work while sitting down.
4. I typically do not talk or interact with others much while doing my work.
5. I typically use a computer to perform my work tasks.
6. I typically perform the same tasks every work day.
7. I typically wear a uniform or protective gear on the job.
8. The supervisor to whom I report is informal in interactions with me.
9. The supervisor to whom I report sets specific quality standards for my work.
10. The supervisor to whom I report provides me with written performance feedback.
11. The supervisor to whom I report provides me with information about the financial performance of the company.
12. The supervisor to whom I report meets with me to discuss my work goals.
13. The supervisor to whom I report praises me for excellent performance.

**Group-Referent, Descriptive Items for the User Group**

14. The work that employees do in a typical day here is physically demanding.
15. To do their work, employees typically talk on the telephone.
16. Employees typically do their work while sitting down.
17. Employees typically do not talk or interact with others while doing their work.
18. Employees typically use a computer to perform their work tasks.
19. Employees typically perform the same tasks every work day.
20. Employees typically wear uniforms or protective gear on the job.
21. The typical supervisor to whom employees report is informal in their interactions with employees.
22. The typical supervisor to whom employees report sets specific quality standards for employees’ work.
23. The typical supervisor to whom employees report provides employees with written performance feedback.
24. The typical supervisor to whom employees report provides employees with information about the financial performance of the company.
25. The typical supervisor to whom employees report meets with individual employees to discuss their work goals.
26. The typical supervisor to whom employees report praises employees for excellent performance.

**Individual-Referent, Evaluative Items for the User Group**

27. My work here is enjoyable.
28. My work here is dissatisfying.
29. My work here is rewarding.
30. My work here is irritating.
31. My work here is motivating.
32. My work here is meaningless.
33. My work here is exciting.
34. My work here is boring.
35. The supervisor to whom I report is good at motivating me.
36. The supervisor to whom I report is cold towards me.
37. The supervisor to whom I report is supportive of me.
38. The supervisor to whom I report is effective in guiding me.
39. The supervisor to whom I report is impolite to me.
40. The supervisor to whom I report is sensitive to my needs.

**Group-Referent, Evaluative Items for the User Group**

41. Employees’ work here is enjoyable.
42. Employees’ work here is dissatisfying.
43. Employees’ work here is rewarding.
44. Employees’ work here is irritating.
45. Employees’ work here is motivating.
46. Employees’ work here is meaningless.
47. Employees’ work here is exciting.
48. Employees’ work here is boring.
49. The typical supervisor to whom employees report is good at motivating employees.
50. The typical supervisor to whom employees report is cold towards employees.
51. The typical supervisor to whom employees report is supportive of employees.
52. The typical supervisor to whom employees report is effective in guiding employees.
53. The typical supervisor to whom employees report is impolite to employees.
54. The typical supervisor to whom employees report is sensitive to employees’ needs.

Received May 18, 1998
Revision received January 13, 2000
Accepted January 18, 2000