A Comprehensive Framework for Determining Measurement Types of Group-Level Construct*

Hyun Sun Chung(Corresponding Author)
Seoul National University
(chunghsunny@snu.ac.kr)
Hyunjee Hannah Kim(Co-Author)
Seoul National University
(hjeekim@snu.ac.kr)
Jihye Lee(Co-Author)
Seoul National University
(wisdom0405@snu.ac.kr)
Jaeun Lim(Co-Author)
Cornell University
(jl3828@cornell.edu)
Won-Woo Park(Co-Author)
Seoul National University
(wwpark@snu.ac.kr)

Although group-level studies are increasing, methods of measuring group-level constructs have remained quite simple because researchers assume compositional conceptualization of group-level constructs. Authors of many studies simply aggregate individual responses to team responses; however, simple aggregation can hinder research by creating significant barriers for conducting group-level studies and impeding our understanding of group microdynamics. That is, simple aggregation cannot represent group dynamics well became composition of individual responses might result in biased estimation. In this paper, we defined a comprehensive list of measurement methods that exist for measuring group-level constructs by reviewing 141 articles from organizational behavior literature published between 2012 and 2016. Based on the literature review, we show how much previous research on group-level studies has relied on skewed methodologies for measuring group-level constructs, and we urge using more diverse ways of measuring them.

Key words: Group-level, team, group dynamics, aggregation, measurement

Ⅰ. Introduction

In recent years, as groups have become increasingly responsible for organizational work (Sundstrom, 1999), notable advances in organizational behavior research have been made in both group-level and multi-level studies (Anderson, Potočnik, and Zhou, 2014). For instance, the most common keyword for the *Academy of Management Journal* during the period of 2000 to 2009 was “group processes

최초투고일: 2019. 2. 1 게재확정일: 2019. 3. 7
* 본 연구는 서울대학교 노사관계연구소 연구비 지원에 의하여 작성되었습니다. 본 연구에 대해 귀중한 조언을 주신 두 분의 심사자님께 감사드립니다.

/performance," which was found in 30% of 237 published articles (Morrison, 2010). In addition, based on manuscripts submitted to the *Academy of Management Journal* from 2007 to 2009, 24% of 437 organizational behavior (OB) papers featured keywords "groups" or "teams" (Morrison, 2010). Another recent trend is group-level studies occurring at higher frequency. For example, although comparison is limited to just one academic journal, only seven out of 86 articles (8.14%) published in the *Journal of Applied Psychology* in 2004 were group-level studies. In the same journal in 2014, 17 out of 91 articles (18.68%) were group-level studies.

Although group-level research is increasing in prominence, study of groups, particularly within quantitative analysis research, have assumed homogeneity of group member inputs or group structures (Dansereau, Yammarino, and Kohles, 1999; Henderson, Wayne, Shore, Bommer, and Tetrick, 2008; Klein and House, 1995) and provided a highly static explanations of groups by mostly aggregating team members' characteristics to the team-level. However, groups, as dynamic entities, evolve over time as individuals interact within them (Cronin, Weingart, and Todorova, 2011; Humphrey and Aime, 2014; Kozlowski and Klein, 2000). In examining group dynamics, Cronin et al. (2011) argued for the importance of a system dynamics perspective, which shows how group-level constructs evolve, based on interaction of cumulative feedback from individuals, emergent feedback from groups, and influence of exogenous variables on these groups. Additionally, group boundaries are often permeable, their memberships can be fluid, and their traits may be more or less shared by people who are members of multiple teams (Cronin et al., 2011). Given both the static and dynamic nature of groups, quantitative measuring of group-level constructs requires that measurement methods capture different characteristics of groups.

However, in reality, measuring group-level constructs has remained relatively simple, with most measurements assuming that group-level constructs are static and characterized by "chain-like unidirectional cause-effect relationships" (McGrath, Arrow, and Berdahl, 2000). Most group-level constructs have been measured either by aggregating individuals' responses in terms of how they view themselves or their groups (Van Mierlo, Vermunt, and Rutte, 2009) or by asking group leaders to assess group attributes. When aggregating individual responses, previous studies mostly used an average of individual responses and assumed that these averages represented a group's attributes, but the degree of members' divergence on certain attributes was largely ignored (Cole, Bedeian, Hirschfeld, and Vogel, 2011). To capture group attributes more comprehensively, researchers have introduced other methods, such as measuring group-level constructs by consensus based on group discussion. However, due to groupthink pressure and intense labor requirements, researchers seldom use a group consensus approach (Wallace, Edwards, Paul, Burk, Christian, and Eissa, 2013).

These limitations in measuring group-level constructs create significant barriers for conducting dynamic nature of group-level studies and impede our understanding of group dynamics. First, simple aggregation of individual responses at the group-level might not accurately represent group dynamics, which in turn leads to a poor fit between study design and measurement levels (Klein, Dansereau, and Hall, 1994). Second, composition of individual responses—without a good rationale for doing so—could result in biased estimates, particularly an over-proliferation of constructs (James, 1982; Meade and Eby, 2007). Finally, for some more established group-level constructs, such as group efficacy, researchers have reached a consensus in terms of appropriate methods for measuring those constructs. However, when it comes to emerging group-level constructs, researchers lack a framework for thinking about measuring them. This results in adoption of simple aggregation methods and leveraging of group leaders to examine group attributes. Most recently, a few attempted measuring dyadic social network relations among group members within the group to measure the dynamics nature of groups (Humphrey and Aime, 2014), but this approach created a debate over whether measuring dyadic relationships could be considered as measuring group-level phenomena. Currently, studies lack a comprehensive framework for measuring group-level constructs.

Hence, in this paper, we first, focus on defining a comprehensive list of measurement methods that exist for measuring group-level constructs in current organizational behavioral research. We suggest nine different measurement methods for assessing group-level constructs and offer pros and cons of each measurement method, mostly based on a review of previous research. Second, we examine the five most recent years (2012 - 2016) of empirical articles from major journals to show the current status of group-level construct measurement and highlight potential issues related to this topic. We specifically highlight how group-level constructs are measured using different methods based on different characteristics of studies or construct characteristics, team tenure, and team sample size, etc. We focus on reviewing the four major journals: *Academy of Management Journal, Journal of Applied Psychology, Journal of Organizational Behavior,* and *Organizational Behavior and Human Decision Processes*. Finally, we suggest factors to be considered in deciding measurement methods and urge future researchers to consider various ways of measuring group-level constructs.

We expect that this paper will contribute to group-level research. It will contribute to methodologies of group-level research by defining nine comprehensive measurement types based on bottom-up reviews of recent empirical papers. The most widely known measurement framework of group-level constructs is Chan's (1998) five models. Chan proposed a typology of five different composition models—additive, direct consensus, referent-shift consensus, dispersion,and process—that aim to be sublated by collecting and aggregating individual responses to measure group-level phenomena and encourage widened perspectives and approaches in collecting data. However, Chan's five models primarily focus on compositional aspects of group work. Thus, we expand Chan's models by defining nine different measurement methods for group-level constructs and shed light on the importance of using a variety models to capture group dynamics more accurately.

Ⅱ. Nine Measurement Methods for Group-Level Constructs

We developed a comprehensive framework for determining measurement methods of group-level constructs based on a review of theoretical papers on group-level constructs and empirical papers that include those constructs. Chan's (1998) five composition models are by far the most popular methods for measuring group-level constructs. Assuming that recent theoretical papers on group-level constructs have cited Chan's research, we reviewed 1,511 articles that cited Chan (1998)[1] and closely examined 15 that included either ways to measure group-level constructs or multi-level measurement methods. Chan's composition models clarify functional relationships among constructs that are conceptually identical but differ on an analysis level. Chan's typology includes five composition models: additive, direct consensus, referent-shift consensus, dispersion, and process. The additive model indicates that relationships between constructs is driven by a mere summation (or averaging) of lower-level values into upper-level constructs, regardless of variance. The direct consensus model considers the notion of intragroup variance by positing within-group agreements to justify aggregation. The referent-shift model is similar to the direct consensus model in terms of justification through within-group consensus, but it involves a conceptual distinction between lower-level (original) and higher-level constructs. According to the dispersion model, within-group variance itself is a representation of a higher-level construct. Finally, the process model reflects emergent

1) We selected 1,511 articles by counting all articles that cited Chan (1998) in a Google Scholar search in June 2015.

mechanisms of lower- to higher-level constructs.

Most scholars in organizational research have developed direct and referent-shift consensus composition models either by specifying measurement methods (e.g., van Mierlo et al., 2009) or by applying models to examples (e.g., Bashshur, Rupp, and Christopher, 2004; Wallace et al., 2013). For example, van Mierlo et al. (2009) specified five complementary steps to be used to examine the distinction between two models. Some scholars have adapted these models to specific constructs and examined which models are most appropriate for different contexts. For example, Wallace et al. (2013) tested which models were suitable for measuring organizational climate.

Some scholars gave attention to other models from Chan's (1998) work. The dispersion model has been further developed by a few scholars (e.g., Roberson, Sturman, and Simons, 2007; Cole et al., 2011). For example, Cole et al. (2011) showed how dispersion-composition models could be applied to capture variability among group members' collective judgments, illustrating a six-step sequential framework for applying dispersion-composition models using data from two independent field samples.

Cronin et al. (2011) also further developed Chan's models, suggesting new ways to classify measurement methods for group-level constructs depending on degree of dynamics. They based their categories on Kozlowski, Gully, Nason, and Smith's (1999) argument that "group constructs and phenomena are not static, but emerge upwards from individual to team-level and unfold via complex temporal dynamics" (p. 242). Depending on the dynamic profile of group-level property, Cronin et al. (2011) classified group-level constructs as contextual, cumulative, or emergent. Emergent constructs are highly dynamic states that change as group members' interaction patterns or attitudes toward interaction change. Although Marks et al. (2001) defined emergent states as "constructs that characterize properties of a team that are typically dynamic in nature and vary as a function of team context, input, processes, and outcomes" (p. 357) and emphasized differences between emergent state and team process, emergent constructs emphasize the end state, which includes a team process component as well as emergent state. Cumulative constructs are marginally dynamic; they change only gradually as membership changes or with passage of time. Contextual constructs are fixed, but can moderate other intra-team dynamics. Cronin et al. (2011) also tried to link these constructs with Chan's (1998) five categories, arguing that emergent constructs could be best measured with direct consensus and referent-shift methods, and that cumulative constructs could be best measured with additive and dispersion methods.

Incorporating recent empirical papers that include group-level constructs, we expand these previous models by developing nine measurement methods for assessing group-level constructs. Our nine methods are based on the following criteria: who measures (team members vs. team leaders vs. external experts or coders vs. researcher manipulation vs. archival), which target they measure (individuals, team), and how individual-level data is aggregated (mean, dispersion). With these criteria, twenty combinations can be created. However, this paper proposes nine measurement methods. The five of them are existing measurement methods which were proposed by Chan (1999) and Cronin et al. (2011) and four measurement methods are new analyzing approaches which are adopted in the recent literature.

Each dimension serves its purpose: The who dimension plays an important role because perspectives of team members and leaders may differ in team dynamics, and external experts may bring new perspectives lacking in internal dynamics. Also, researcher manipulation and archival data can be influential in different ways. Out of the three dimensions (who, which, and how), the who dimension can be the most subjective and can be influenced by internal dynamics.

Table 1 presents definitions and summaries of our nine measurement methods with these criteria. Our categories were based on previous empirical examples, including Chan's (1998) five composition models, which were helpful in summarizing previously used measurement methods, but they were not successful in suggesting new ways to measure group-level constructs. Our nine measurement methods include new ways to measure group-level constructs comprehensively.

2.1 Measurement Method 1: Aggregation of Individual Responses on Self

Measurement method 1 is probably the most familiar and popular form. It is similar to Chan's (1998) additive and direct consensus models. This method measures group-level constructs by combining individual's lower-level, self-reported scores with higher-level team indices. By combining lower-level individual responses, an average (arithmetic mean) is used. For example, when measuring team-level emotional skills, Troth, Jordan, Lawrence, and Tse (2012) questioned each team member using the Workplace Emotional Intelligence Profile to capture individual's self-assessments of their emotional awareness and emotional management skills within a team context. They then aggregated their responses for the team-level construct. Sample items included the following: "I can explain emotions I feel to team members" and "When I am frustrated with fellow team members, I can overcome my frustration."

In a majority of cases, within-group agreement indices (e.g., r_{wg} and ICC(1)) or between-group indices (e.g., ICC(2)) were provided to

〈Table 1〉 Nine measurement types for measuring group-level constructs

Measurement types		Description	Example	Chan's (1998)
		Type 1: Aggregation of Individual Responses on Self		
Rater Ratee(s) Aggregation	Members Members Mean	Group members respond to items related to individual-level constructs, then, these scores are aggregated to the group-level construct using mean score.	Team's emotional skill was assessed by averaging each group members' self-assessment of their emotional awareness and emotional management skills within a team context (Troth et al., 2012).	Additive or direct consensus models
		Type 2: Aggregation of Individual Responses on Team		
Rater Ratee(s) Aggregation	Members Group Mean	Group members respond to items related to group-level constructs, then, these scores are aggregated to the group-level construct using mean score.	Task conflict was assessed by individual team members' assessment on their team's level of conflict (Chun & Choi, 2014).	Referent-shift consensus model
		Type 3: Aggregation of Individual Responses with Dispersion		
Rater Ratee(s) Aggregation	Members/Leader Members/Group Standard deviation	Regardless of raters or ratees, all group-level aggregation indexes using standard deviation (or variance) scores.	Measure polychronicity diversity via the within-group standard deviation of individually assessed scores (Mohammed & Nadkarni, 2014).	Dispersion model
		Type 4: Assessment by Group Leaders		
Rater Ratee(s) Aggregation	Leader Members/Group Mean	Group leader responds to items related to individual-level constructs of each of the group members, then, these scores are aggregated to the group-level construct using mean score. / Group leader responds to items related to group-level constructs.	Ask group leaders to judge each group member's OCB using a 24-item scale, then, these assessments were aggregated to yield an overall, group-level OCB score (Cole, Carter, & Zhang, 2013) / Group innovation as process was assessed via a four-item measure of supervisor-rated group innovation (Gajendran & Joshi, 2012)	N/A
		Type 5: Assessment by External Experts or Coders		
Rater Ratee(s) Aggregation	Outsider(s) Members/Group N/A	Outside experts rate value of the group-level constructs.	Measuring group performance by external managers' ratings with three-item performance measure (Cole et al., 2013)	N/A
		Type 6: Manipulation by Researchers		
Rater Ratee(s) Aggregation	Researcher(s) Members/Group N/A	Researchers rate value of the group-level construct (mostly used in experimental studies).	An experimental study measured supervisor-focused interpersonal justice climate by researchers through manipulation (Stoverink, Umphress, Gardner, & Mincer, 2014).	N/A
		Type 7: Group Consensus Approach		
Rater Ratee(s) Aggregation	Group Group N/A	Group consensus approach: all group members collectively respond to items related to group-level constructs.		N/A
		Type 8: Integration of Group Leader and Group Members' Assessments		
Rater Ratee(s) Aggregation	Members/Leaders Group N/A	Each group leader and group member responds to the assessment on the team and integrates those ratings to come up with one single group-level rating.		N/A
		Type 9: Integration of Self and Other Group Members' Assessments		
Rater Ratee(s) Aggregation	Members Members/Group N/A	Integrates an individual's self-assessments with other group members' assessments of that individual to create a group-level construct. Since both subjective and objective perspectives are essential to gain a sound understanding of an issue, the results of self-assessments and observers' assessments should be compared. It creates opportunities to compare how individuals evaluate themselves and how others evaluate them.	Belbin developed a software called *Interplace*, which measures the level of teamwork by aggregating team roles of group members. Those measurements were calculated by combining results from self-perception inventory and observer assessments (Belbin, 1981, 1993).	N/A

justify aggregations. Chan (1998) emphasized that within-group agreement is the proper standard for aggregation in a direct consensus composition, as this model describes a group-level construct in terms of a consensus among individual members. However, as most individual-level assessment variables have intragroup correlation (ICC[1]) values greater than zero, Bliese, Chan, and Ployhart (2007) pointed out that these nonzero values are evidence of group effects. Some group properties show a simple aggregation of individual group members' attitudes, characteristics, or perceptions. Klein and Kozlowski (2000) referred to this as configural team properties, whereas Cronin et al. (2011) defined it as a cumulative construct. Using measurement method 1 is most appropriate when group-level constructs relate to a group's configural aspects. In other words, this measurement method is most suited to capturing preexisting attributes among individual group members rather than emerging group-level dynamics based on interactions of individual members. We will discuss this theme further in the recommendation section.

2.2 Measurement Method 2: Aggregation of Individual Responses on a Team

Measurement method 2 is similar to measurement method 1 in that both are based on aggregation of individual-level ratings. The only difference between the two measurement methods is the target of measurement: team or self. In measurement method 2, group-level constructs are computed using mean values of individual group members' assessments of a group. The *task conflict* measurement in Chun and Choi's research (2014) is a relevant example. The sample items in this study were adopted from Jehn and Mannix's study (2001) which included: "My team members experience conflict of ideas with others" and "My team members frequently have disagreements with others about tasks they are working on." Like measurement method 1, this type mostly provides within-group agreement indices. In addition, the task conflict model is comparable to Chan's (1998) referent-shift consensus model in terms of conceptual differences between an assessed group-level construct and an original individual-level construct. Wallace et al. (2013) suggested that referent-shift indices are cognitive measures, which makes them more appropriate for capturing effectiveness-related attributes. According to Klein and Kozlowski (2000), shared group properties emerge from shared experiences, perceptions, values, climates, and interpretations among group members. They suggest that when measuring shared group properties, researchers who gather data from individuals with shared norms must justify aggregation. Therefore, we regard our measurement method 2 as applicable for measuring constructs related to share group properties.

2.3 Measurement Method 3: Aggregation of Individual Responses with Dispersion

In measurement method 3, consensus is not a necessary condition for construct validity at a higher level of aggregation. Instead, the within-group variance index itself represents a single group-level construct. That is, within-group variance becomes the operationalization of a focal construct rather than an error variance. This measurement method is analogous to Chan's (1998) dispersion model. In this model, a rater(s) can be a group leader, team member, outsider, or anyone else who follows research design, whereas individuals being rated are focal team members.

Given that dispersion can capture variability, it can provide new perspectives and insights compared to the aggregation of mean values of individual-level responses. As Peeters et al. (2006) explained, dispersions capture distinct composition effects in group settings to examine internal group dynamics, which are not captured via the computation of mean values. Cole et al. (2011) claimed that dispersion indices are much more powerful than averaging when capturing variability. These indices also strengthen research validity. According to Cronin et al. (2011), Chan's dispersion model (1998) is appropriate for measuring cumulative constructs.

Furthermore, one can view dispersion from another perspective. Several studies in organizational behavior research have used dispersion indices to examine congruence (e.g., fit, similarity, match, and cohesion) between two constructs to predict an outcome. We can consider dispersion by adopting this perspective of concept-opposite congruence. Edwards (1994) codified several ways of operationalizing congruence by collapsing two or more measures into a single index and summarizing problems associated with the indices used. Dispersion indices can be divided into two groups, one consisting of bivariate dispersion indices—the most commonly used are algebraic, absolute, and squared difference—and another consisting of profile-similarity indices (D, D^2, etc.). However, these indices present some methodological problems because they are inherently ambiguous; confound effects of each component, and obscure information, which yields misleading results.

There are various approaches to dispersion measurement. For example, diversity-related constructs have often been measured with this method. For example, Mohammed and Nadkarni (2014) measured polychronicity diversity via a within-group standard deviation of individually assessed scores. Scholars have used several indices to capture dispersions and coefficients of variation (Colquitt, Noe, and Jackson, 2002). Roberson et al. (2007) suggested appropriate indices of dispersions, including standard deviation, average deviation, inter-rater agreement indices, and coefficient variation. They argued that standard deviation—the most common dispersion index—is useful for binary and nonbinary items and that it can be applied across populations with two or more participants. In addition, standard deviation is a relatively convenient index that is easy to interpret. However, an average deviation index, calculated as the average of absolute differences between each score and the overall mean, is better when a sample is not normally distributed because it is less susceptible to outlying scores or deviations from the norm (Stigler, 1973).

Therefore, measurement method 3 is appropriate for measuring cumulative constructs, such as diversity and individual differences or agreements, as it reflects intragroup variability, inter-individual difference, and congruence between constructs.

2.4 Measurement Method 4: Assessment by Group Leaders

Measurement method 4 refers to group leaders' assessments—either of a group itself or of each group member—followed by an aggregation to group level. The most representative example of a group leader's assessment of individual group members followed by aggregation is measurement of group-level organizational citizenship behavior (OCB). According to Cole et al. (2013), group-level OCB measurement is conducted by asking group leaders to judge each group member's OCB using a 24-item scale. These data are then aggregated to yield an overall group-level OCB score. However, assessment of a group itself by group leaders is more common. For example, Gajendran and Joshi (2012) measured team innovation by asking supervisors to rate their teams using measurements of four items from De Dreu and West (2001). Sample items included "This team often produces new services, methods, or procedures" and "Team often implements new ideas to improve quality of our products and services."

According to Klein and Kozlowski (2000), global team properties are "relatively objective, descriptive, and easily observable team characteristics" (p. 215). According to Cronin et al. (2011), contextual constructs are predetermined, regardless of interaction among group members. Thus, global team properties do not emerge from or combine with attributes of individual team members and, thus, they cannot be differentiated from personal characteristics or processes. Although there has been some criticism regarding whether group leaders should measure groups and their individual members due to overestimating, overvaluing, or overrating, measurement method 4 can be considered suitable for measuring this type of group property.

2.5 Measurement Method 5: Assessment by External Experts or Coders

Measurement method 5 refers to external experts or coders either assessing individuals within a group (followed by aggregation to group-level) or assessing a group itself. External experts or coders rely on archival data, survey data, or video recordings of group interaction to assess group-level constructs. They might use objective criteria for assessing a group (i.e., number of ideas created) or subjective judgments based on certain criteria (i.e., novelty of a group's ideas). Cole et al. (2013) measured team performance by asking external managers to rate groups based on a three-item performance measure. Sample items included "This team gets its work done very effectively" and "This team has performed its job well." By having external experts perform assessments, researchers can obtain objective group-level measurements. However, hiring and training external experts or coders can be costly and require a significant amount of effort.

2.6 Measurement Method 6: Manipulation by Researchers

Measurement method 6 involves researchers manipulating group-relevant constructs and studying effects caused by these manipulations. Researchers mostly use experimental studies to easily manipulate constructs. For instance, Stoverink et al. (2014) examined the relationship between a supervisor-focused, interpersonal justice climate and team cohesion. Therefore, by manipulating the justice climate, the researchers were able to determine its effect on team outcomes and other group dynamics. This manipulation was possible because a justice climate measure is a cumulative (Cronin et al., 2011) or configural (Klein and Kozlowski, 2000) construct that exists regardless of interactions between group members.

2.7 Measurement Method 7: Group Consensus Approach

As a way to measure group constructs, measurement method 7 involves observing the way groups collectively form consensus. The group consensus approach has merits in that a group can collectively formulate group attributes. Some researchers argue that this measurement can capture a group's emerging dynamics more accurately than other methods (Gibson, Randel, and Earley, 2000). Gibson et al. (2000) tested the effectiveness of different measurement methods in assessing group efficacy and concluded that group efficacy measured with the group consensus approach resulted in the highest validity. However, the group consensus approach needs to be used carefully, as it requires extensive effort, time, and resources. Additionally, the phenomenon of groupthink may occur: If one individual has strong opinions regarding an issue, that individual might influence others in a group or increase pressure to disagree with fellow members in a group, resulting in a biased group decision. Recently, due to these concerns, measurement method 7 has not been widely used.

2.8 Measurement Method 8: Integration of Group Leader and Group Members' Assessments

Measurement method 8 is an integration of measurement methods 2 and 4. Specifically, each group leader and group member responds to an assessment of his or her team and integrates the assessment results to come up with a single group-level assessment. Given that group dynamics emerge as group leaders and members interact (Cronin et al., 2011), assessing a group-level construct from only one party's point of view—either group leader or group members—only provides a partial assessment. Thus, understanding how both parties assess the group-level construct and integrate the assessments is imperative. To our knowledge, there are no published empirical studies that have integrated these two measurement methods.

2.9 Measurement Method 9: Integration of Self-Assessments and Observer Assessments

Measurement method 9 integrates an individual's self-assessment with observers' assessments of that individual to create a group-level construct. Because both subjective and objective perspectives are essential for gaining a sound understanding of an issue, results of self-assessments and observers' assessments should be combined. Researchers frequently adopt self-reporting assessments, but this technique can provide inaccurate information because it might deal with social desirability biases. This method, however, creates opportunities to compare how individuals evaluate themselves and how others evaluate them. Belbin's team role balance assessment uses this methodology (Belbin, 1981, 1993). Software called *Interplace* developed by Belbin Associates measures levels of teamwork (i.e., team role balance) by aggregating team roles of group members, which were calculated by combining results from a self-perception inventory and observer assessments (Belbin, 1981, 1993).

Litrico and Choi (2013) conducted research on assessment efficacy, measuring both team efficacy and reflected team efficacy. Although this particular study did not combine measures (due to their differences), a methodology that combines them can still be adopted in future studies.

III. Review of Recent Empirical Articles Measuring Group-Level Constructs

3.1 Literature Search and Selection Criteria

To better understand how group-level constructs are currently measured in empirical articles, we compiled a comprehensive database by retrieving relevant articles published between January 1, 2012, and December 31, 2016. We searched for studies that provided data on team and/or group-level constructs, regardless of the themes of the articles. We chose to target the four most relevant and popular journals in the organizational behavioral discipline: *Academy of Management Journal*, *Journal of Applied Psychology*, *Journal of Organizational Behavior*, and *Organizational Behavior and Human Decision Processes*. To obtain each article's full text, we performed a combined search of several electronic databases, including PsyArticles, Business Source Complete, ProQuest Central, and ScienceDirect. These electronic database searches resulted in 992 records.

We limited search results to organizational behavior-related studies. We did not include book chapters in the search because data presented in book chapters are often preliminary and tend to be included in empirical journal papers at a later date. After obtaining an initial pool of suitable articles, we reviewed them and checked that empirical studies described group-level constructs. In the screening process, we read full texts for details and then excluded papers based on our criteria. Such decisions were made in the following ways: (a) which theory was tested in a paper, (b) how a paper related to validity of team-level constructs, (c) whether a study included a meta-analysis, (d) whether a study used survey methods in its team-based sample and whether all variables were at the individual level, and (e) whether control variables were included in our databases regardless of levels. Meanwhile, we counted studies as separate cases if variables were collected at different times, if different sample sizes were used, and/or if distinct studies were combined in one published paper. For example, in Lehmann-Willenbrock and Allen's (2014) study, the authors collected the team performance variable at two different times and with different sample sizes. Therefore, we considered this variable independent and counted it twice.

After adopting these criteria, we identified 141 relevant articles with 650 group-level variables. Among relevant articles, 56 were published in the *Journal of Applied Psychology*, 36 in the *Academy of Management Journal*, 29 in the *Journal of Organizational Behavior*, and 20 were published in *Organizational Behavior and Human Decision Processes* (see Table 2 for more details).

〈Table 2〉 Description of the Current Study: The Number of Articles and Group-level Variables per Journal

Journal		2012	2013	2014	2015	2016	Total
AMJ	Articles	8	10	9	3	6	36
	Variables	37	59	30	19	27	172
JAP	Articles	6	9	17	14	10	56
	Variables	62	32	86	58	50	288
JOB	Articles	10	2	4	5	8	29
	Variables	43	13	16	10	18	104
OBHDP	Articles	4	6	2	4	4	20
	Variables	19	34	8	10	15	86
Total	Articles	28	27	32	26	28	141
	Variables	161	138	140	101	110	650

3.2 Coding Procedure and Results

To guide coding of group-level variables and information from samples, we developed a set of standardized procedures and decision rules. The coding scheme was developed a priori and was revised after pilot coding. All authors adopted this coding scheme when rating all of the collected studies. All authors reviewed each journal that was allocated to each person, respectively. Two of the authors then checked the coding. The authors discussed and recorded any problems or coding errors before conducting the analyses.

A large amount of information was recorded and coded for each article. The basic information of each article (author, year, title, and journal) was recorded. Then, each study was coded as one of eight categories of study type: experiment (E), observation (O), questionnaire/survey (S), interview (I), content analysis (C), archival study (A), other (e.g., an algorithm), and mixed (coded with all applicable types). As shown in Table 3, over half of the examined studies (91 out of 154) used survey design, and almost one-third (47 out of 154) were laboratory experiments. Only a few studies used other forms or more than one design type. Thus, studies were weighted toward particular types (survey and experiment), which implies that current group-level studies are limited to specific forms of study design. This could be problematic, given that surveys and experiments cannot capture all group dynamics. For instance, Hargadon and Bechky (2006) used an ethnographic observation study to examine the moment when collections of creative people become creative collectives. Capturing the moment in which collective creativity arises is almost impossible using a survey or experimental study.

To classify collected studies using nine

〈Table 3〉 Description of the Current Study: The Number of Articles per Study Type

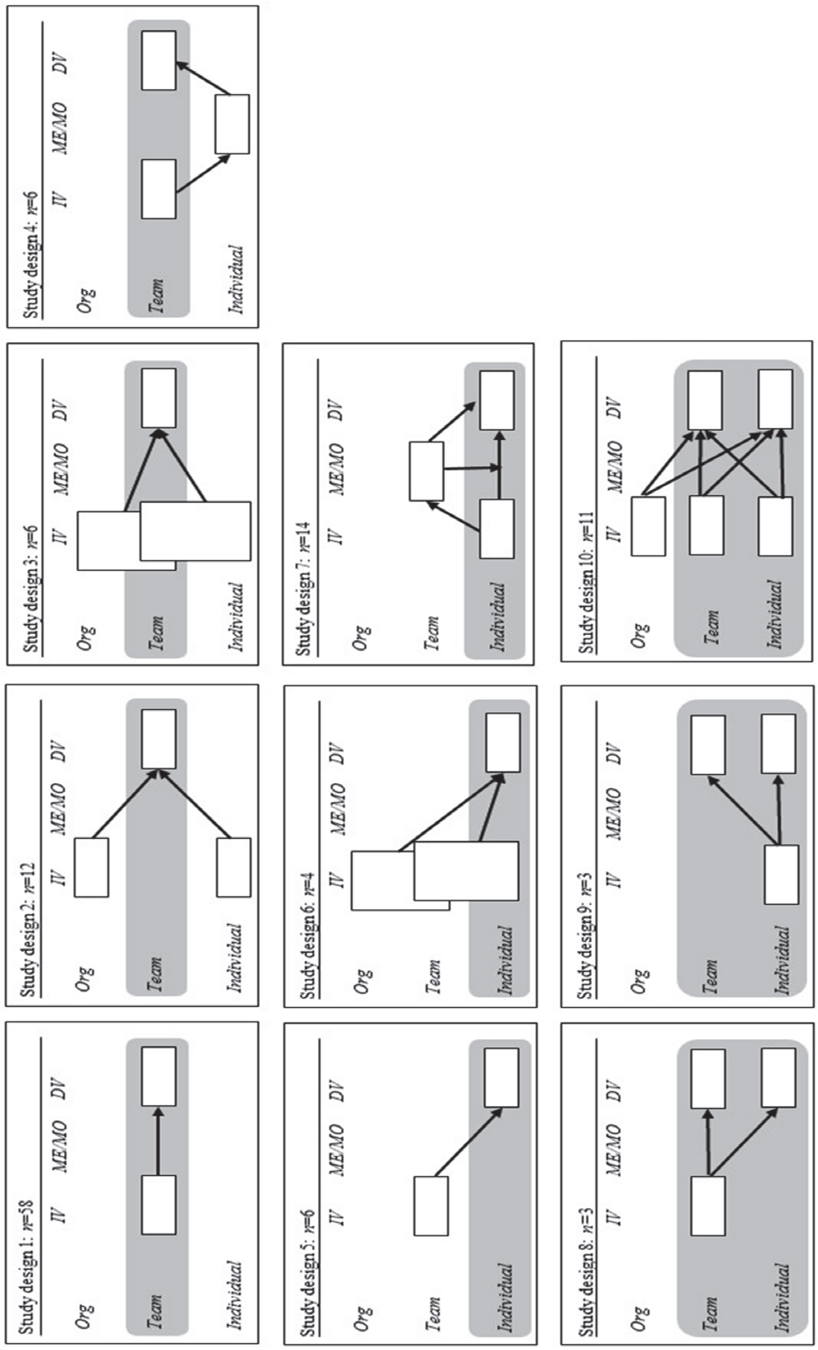
Study type	2012	2013	2014	2015	2016	Total
Survey	27	12	17	14	21	91
Experiment	10	11	9	7	10	47
Archival/Algorithm/Interview	1	4	1	3	0	9
Mixed	0	0	5	2	0	7
Total	38	27	32	26	31	154

measurement methods, group-level constructs were coded into one of nine categories. In our sample, 39.8% of variables were observed via measurement methods 1 and 2, which involve aggregations of individual assessments using mean scores; 32.8% were assessed by group leaders. Variables measured with assessments by outsiders or researchers represented only 7.6% and 12.5% of articles, respectively. Only 1.4% of articles used standard deviation scores for aggregations, and none of the variables employed a group consensus approach.

Study designs of each article were coded considering levels (e.g., individual, team, and organization) of independent variables (IVs), dependent variables (DVs), mediator, and moderator variables. As shown in Figure 1, 10 different study designs are clarified. For example, studies of team-level DVs could be reclassified into four different designs depending on whether their IVs are at team-level (study design 1), organizational/ individual level (study design 2), or cross levels (study design 3). When IVs and DVs are all at team-level but mediators or moderators are at other levels, an article could be classified

〈Table 4〉 Description of the Current study: The Number of Group-Level Variables per Measurement Type

Measurement Type	Number of variables	Ratio
1. Aggregation of individual responses on self	133	17.73%
2. Aggregation of individual responses on team	149	22.14%
3. Aggregation of individual responses with dispersion	13	1.43%
4. Assessment by group leaders	152	32.80%
5. Assessment by external experts or coders	61	7.64%
6. Manipulation by researchers	98	12.51%
7. Group consensus approach	0	0%
8. Integration of Group Leader and Group Members' Assessments	0	0%
9. Integration of Self and Other Group Members' Assessments	0	0%
Total	650	100.00%



〈Figure 1〉 Classification of Study Designs into a different study design (study design 4). Similarly, studies of individual-level DVs could be reclassified into three different designs depending on whether their IVs are at team-level (study design 5) or cross levels (study design 6). When IVs and DVs are all at individual levels, but mediators or moderators are at team-level, an article could be classified into a different study design (study design 7). Finally, studies of both individual-level and team-level DVs could be reclassified into three different designs depending on whether their IVs are at team-level (study design 8), individual level (study design 9), or cross levels (study design 10). When we considered the total number of articles for individual, team, and multi-level studies, 47.15% of articles that contained group-level constructs were designed solely at the team-level.

〈Table 5〉 Description of the Current Study: The Number of Group-Level Variables per Study Design

Study design	Number of variables	Ratio
1	401	64.2%
2	22	2.5%
3	49	7.6%
4	30	3.8%
5	39	5.4%
6	14	2.5%
7	20	3.9%
8	24	3.2%
9	8	1.3%
10	43	5.7%
Total	650	100.0%

Note. Specification of study designs are shown in Figure 1

We initially coded variables using their original names as defined by researchers. We categorized those variables according to their characteristics and contexts. For example, we created three high-level classifications—context, process, and outcome—depending on how variables were used. Hence, a single measure, such as group climate, could be coded as either context or outcome depending on how it was used. For example, most of the climate-related variables in our database were independent or moderator variables and were coded as "group structure-climate." However, the dependent variable *safety climate* was coded as "outcome-others." We then created a mid-level classification by specifying the context in terms of group characteristics (individual difference and others), group structure (team composition/diversity, leadership, and others), group climate, knowledge management/learning, and group strategy (task strategy and team strategy). We also specified outcomes in terms of four subcategories (creativity/innovation, performance, OCB, and others). Finally, we left the process as it was. The distribution of variables depended on their characteristics. Later, this mid-level classification was used to determine whether a team-level construct had emergent, cumulative, or contextual characteristics based on Cronin et al.'s (2011) classification.

〈Table 6〉 Description of the Variable Classification

Variable classification			Number of variables
Context	Group member characteristics	individual difference	47
		other group characteristics	66
	Group structure	group composition/diversity	55
		leadership	39
		other group structure	55
	Group context	group climate	31
		other contextual variables	11
	Group learning	KM/learning	31
	Group strategy	team strategy	54
Process		team process	71
Outcome		outcome-performance	88
		outcome-creativity/innovation	19
		outcome-OCB	15
		outcome-others	56
		others	12
		Total	650

3.3 Critical Issues of Current Measurements of Group-Level Constructs

In summary, when we integrates what we have found from reviewing recent research on group-level constructs, there are two critical issues on measurements of group-level constructs. First, as we defined measurement methods of group-level constructs and counted how many variables were measured by each method, and the results indicate that there are critically skewed usage of a few measurement methods. Aggregating individual responses on self or team comprises 39.87% of measurement methods of variables, followed up assessment of group leaders comprising 32.80%. On the other hand, measurement methods that might be effective in measuring the dynamics of teams, such as aggregation of individual responses with dispersion is measured only limitedly (1.43%). More objective way of measuring by external experts or coders were limited to 7.64%. Moreover, more sophisticated methods of measuring group level constructs by integrating two or more ways of assessments were limited to 5.75%.

Second, when researchers select measurement methods for measuring group-level constructs, they provide limited explanations of chosen methods. Team context or team structure, such as team tenure, team size, or degree of task interdependence of external parties, would affect the ways of measuring group-level constructs; however, in the reviewing process, there are only limited explanations in choosing certain methods. In particular, when researchers focused on emergent phenomena of groups, measurement methods using static ways of measuring group-level constructs, such as assessment of group leaders, could be misleading.

Ⅳ. Future Recommendation on Measurement Methods of Group-Level Constructs

Given that current measurements of group-level constructs are limited to usage of a few methods without the full consideration of group characteristics, this paper attempts to recommend several ways to consider in deciding measurement methods. In this paper, among various factors that affect determination of appropriate measurement methods for group-level constructs, we examine how characteristics of group variables, such as emergent, cumulative, and contextual characteristics; study designs; team tenure; or team size might affect determination of measurement methods.

4.1 Characteristics of Group Variables

With more established group-level constructs such as group climate or group efficacy, researchers have established consensus on the measurement method that should be used for given conditions. For instance, in order to measure group efficacy, the predictability of group efficacy does not depend on the measurement method. Whiteoak, Chalip, and Hort (2004) examined differences between aggregation of self-efficacy perceptions, aggregation of individual perceptions of team efficacy, and the group consensus approach; they found that in a low interdependence task, no differences could be found between three measurement methods. However, when it comes to more nascent group-level constructs such as group learning or group cohesion, no study has yet explained which measurement method is most appropriate. Moreover, no framework exists for how to determine an appropriate measurement method.

We suggest that a measurement method should match characteristics of group-level constructs. Many different researchers have written typologies that characterize the nature and properties of such constructs (Chan, 1998; Chen, Mathieu and Bliese, 2004; Kozlowski and Klein, 2000). Given that some group constructs are not static (e.g., group cohesion) but that others are static (e.g., task interdependence), we recommend that distinguishing group-level constructs using three categories—emergent, cumulative, and contextual— depending on group constructs' dynamic profiles (Cronin et al., 2011) could give an appropriate guideline for selecting measurement methods.

An emergent construct comes into existence based on the interaction among lower-level elements (Klein and Kozlowski, 2000). Since the concept of emergence often takes time to occur and can change over time (Klein and Kozlowski, 2000), this paper's measurement methods 1, 2, and 3, which involve aggregation either of self-reports to group-level, of group reports by group members, or of dispersion within individual ratings could be helpful to understand it since the process of emergence implies that a group-level construct is collectively created through interactions among group members (Cronin et al., 2011). Additionally, for capturing emergent aspects of group dynamics, measurement method 5— Assessment by External Experts or Coders — and measurement method 7— Group Consensus Approach- could be also appropriate. In Resick, Murase, Randell, and DeChurch's (2014) laboratory experiment study, external coders assess emergent collective leadership by coding recordings during a team task.

Cumulative constructs are those in which higher-level phenomena are based on stable individual properties (Cronin et al., 2011). Cumulative constructs come into existence when a collective boundary is drawn. Unlike emergent constructs, which are based on interaction of individuals in a group, cumulative constructs are based on preexisting objective attributes (e.g., age, tenure, and functional background) defined by a researcher, and it can be aggregate these attributes to team-level based on mean, dispersion, or more sophisticated mathematical functions. Thus, cumulative constructs should not be sensitive to individual interaction and should result from aggregation of stable individual attributes, assessment by group leaders (measurement method 4) and assessment by external experts or coders (measurement method 5) could be used.

Contextual constructs refer to those imposed on a team by external forces (Cronin et al., 2011). Contextual constructs might include organizational-level environmental conditions in which a team operates and it exist regardless of interaction between individual group members and a group—and that they are mostly created as a group is formed (as in team strategy)—the most relevant measurement method for these constructs could be either assessment by group leaders (measurement method 4) or assessment by external experts or coders (measurement method 5). In addition, researchers are able to manipulate these constructs (measurement method 6) if they want to identify the team contextual variable's effect on other variables.

4.2 Research Design

A measurement type should be determined with consideration of a research design. Even if the same variables are measured for similar purposes, different measurement types might be used based on a research design. For instance, when a research design is entirely at team-level, measuring team-level constructs at that level (e.g., aggregating team-level perceptions or aggregating team assessment by team leaders) will capture team dynamics better than simple aggregation of individual responses on self-perceptions. Future research can replicate this work using other references and examine the applicability of our suggestion that it is important to match measurement types with study designs and determine if any potential pitfalls may occur due to mismatching.

4.3 Group Tenure

Group tenure could help determine measurement types for group-level constructs. Emergent phenomena at group-level are known to take time to manifest at higher levels, and this time frame may be brief or lengthy depending on the phenomenon (Kozlowski, Chao, Grand, Braun, and Kuljanin, 2013). Allen and O'Neil (2015) examined 20 years of research conducted on groups and found that, for more established firms, measuring variables at the team-level was more appropriate for capturing stable team-level attributes; whereas, for newly formed teams, aggregating individual-level measurements demonstrated greater predictability. In our coding sample, although team tenure was less than 3 years, only 24 variables were measured by aggregating individual-level self-reports, compared to 34 variables measured at team-level. Whether use of certain measurement types biased these sample results should be explored further.

4.4 Other Considerations

First, in our coding sample, 30 of 650 variables used multiple measurement types to measure one team-level construct. This represents a promising direction because use of multiple measurement types could increase construct validity and reduce common method biases. For example, Goncalo and Duguid's (2012) experimental study measured group creativity using multiple methods, including an aggregation of team members' perceptions of team creativity (measurement method 2) and a quantitative and qualitative assessment by external coders (measurement method 5). External coders counted the number of non-repetitive ideas generated in each group and subjectively determined the selected ideas' novelty and usefulness. In addition, at the end of the team task, Goncalo and Duguid (2012) asked team members to rate their creative performances. By using multiple methods, Goncalo and Duguid (2012) were able to show that their findings were consistent across different measurement types, thus increasing the validity of their findings. To highlight the importance of this trend, it will be interesting to examine further examples of the use of multiple measurement types for one team-level construct.

Second, although this paper did not propose as a measurement method, leveraging social network analysis could be a promising approach to capture dynamic nature of groups. Teams are known as assemblies of interdependent relations (Humphrey and Aime, 2014). In order to measure inter-relations between dyads within the team, it could capture dynamic phenomenon occurring at the team-level. For instance, network density, calculated by proportion between potential connections and actual connections, is known to depict the pattern of emergent relationships within a unit well (Kozlowski et al., 1999), whereas, a similar construct, group cohesion, defined as the tendency for a group to be in unity while working towards a goal (Carron and Brawley, 2000), captures mostly a static aspect of group characteristics. By measuring all the combination of dyadic relations, not only it might capture more exact team dynamism but also it could leverage the dyadic information to asses team such as how any extra ordinal network relations affect the overall team dynamism.

Given the increased use of social network analysis, social network density could be used as a possible measure of group cohesion (e.g., Yang and Tang, 2004). Recent literature consider the definition of group cohesion in three ways: task commitment, interpersonal attraction, and group pride (Beal et al., 2003). However, existing studies often used the traditional cohesion construct which using an attitudinal instrument, and it relies on perceptions of members which are aggregated at the team-level. There are a few researchers who have noted the similarities or have investigated the differences. According to Beal et al. (2003), group cohesion has been defined as a way to measure the strength of the bonds within a group based on the premise that stronger bonds result in more cohesive groups. It appears as group cohesion and social network density are both defining the similar type of interaction and the strength of that interaction between group members.

Ⅴ. Conclusion

Although group-level studies are increasing in prominence, in this study we have shown how measurement of group-level constructs is limited to certain types, particularly an aggregation of individual self-responses, team responses, or an assessment by group leaders. Thus, we propose a comprehensive framework for determining which measurement methods to use to measure group-level constructs under certain conditions. By defining nine measurement methods and their subsequent guidelines for use—based on a literature review of current studies regarding group-level constructs—we specified in detail how characteristics of group-level constructs, such as emergent, cumulative, or contextual, affect determination of appropriate measurement methods. Moving forward, this study may contribute to the knowledge base of group-level studies, specifically measuring group-level constructs. We hope that this study encourages other researchers to conduct group-level studies with widened perspectives of measuring group-level constructs, which will result in a sound understanding of group-level organizational phenomena.

References

Allen, N. J., and O'Neill, T. A. (2015). The trajectory of emergence of shared group-level constructs. *Small Group Research*, 46, 352-390.

Anderson, N., Potočnik, K., and Zhou, J. (2014). Innovation and creativity in organizations: A state-of-the-science review, prospective commentary, and guiding framework. *Journal of Management*, 40, 1297-1333.

Bashshur, M., Rupp, D. E., and Christopher, J. (2004). Theoretically-based strategies for defining and measuring justice climate: Implications for multilevel research in organizational justice. In 64th annual meeting of the Academy of Management, New Orleans, Louisiana.

Beal, D. J., Cohen, R. R., Burke, M. J., and McLendon, C. L. (2003). Cohesion and performance in groups: A meta-analytic clarification of construct relations. *Journal of Applied Psychology*, 88(6), 989-1004.

Belbin, R. M. (1981). *Management teams: Why They Succeed or Fail*. Oxford: Butterworth Heinemann.

Belbin, R. M. (1993). Team Roles at Work. Oxford: Butterworth Heinemann.

Bliese, P. D., Chan, D., and Ployhart, R. E. (2007). Multilevel methods: Future directions in measurement, longitudinal analyses, and nonnormal outcomes. *Organizational Research Methods*, 10, 551-563.

Carron, A. V. and Brawley, L. R. (2000). "Cohesion: Conceptual and measurement issues." *Small Group Research*, 31(1), 89-106.

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

Chen, G., Mathieu, J. E., and Bliese, P. D. (2004). A framework for conducting multilevel construct validation. In F. J. Yammarino and F. Dansereau (Eds.), *Research in multilevel issues: Multilevel issues in organizational behavior and processes* (Vol. 3, pp. 273-303). Oxford: Elsevier.

Chun, J. S., and Choi, J. N. (2014). Members' needs, intragroup conflict, and group performance. *Journal of Applied Psychology*, 99, 437-450.

Cole, M. S., Bedeian, A. G., Hirschfeld, R. R., and Vogel, B. (2011). Dispersion-composition models in multilevel research: A data-analytic framework. *Organizational Research Methods*, 14, 718-734.

Cole, M. S., Carter, M. Z., and Zhang, Z. (2013). Leader-team congruence in power distance values and team effectiveness: The mediating role of procedural justice climate. *Journal of Applied Psychology*, 98, 962-973.

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

Cronin, M. A., Weingart, L. R., and Todorova, G. (2011). Dynamics in groups: Are we there yet? *The Academy of Management Annals*, 5, 571-612.

Dansereau, F., Yammarino, F. J., and Kohles, J. C. (1999). Multiple levels of analysis from a longitudinal perspective: Some implications for theory building. *Academy of Management Review*, 24(2), 346 - 357.

De Dreu, C. K. W., and West, M. A. (2001). Minority dissent and team innovation: The importance of participation in decision making. *Journal of Applied Psychology*, 86, 1191-1201.

Edwards, J. R. (1994). The study of congruence in organizational behavior research: Critique and a proposed alternative. *Organizational Behavior and Human Decision Processes*, 58, 51-100.

Gajendran, R. S., and Joshi, A. (2012). Innovation in globally distributed teams: The role of LMX, communication frequency, and member influence on team decisions. *Journal of Applied Psychology*, 97, 1252-1261.

Gibson, C. B., Randel, A. E., and Earley, P. C. (2000). Understanding group efficacy: An empirical test of multiple assessment methods. *Group and Organization Management*, 25, 67-97

Goncalo, J. A., and Duguid, M. M. (2012). Follow the crowd in a new direction: When conformity pressure facilitates group creativity (and when it does not). *Organizational Behavior and Human Decision Processes*, 118, 14-23.

Hargadon, A. B., and Bechky, B. A. (2006). When collections of creatives become creative collectives: A field study of problem solving at work. *Organization Science*, 17, 484-500.

Henderson, D. J., Wayne, S. J., Shore, L. M., Bommer, W. H., and Tetrick, L. E. (2008). Leader-member exchange, differentiation, and psychological contract fulfillment: A multilevel examination. *Journal of Applied Psychology*, 93(6), 1208-1219.

Humphrey, S. E., and Aime, F. (2014). Team microdynamics: Toward an organizing approach to teamwork. *The Academy of Management Annals*, 8(1), 443-503.

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

Jehn, K. A., and Mannix, E. A. (2001). The dynamic nature of conflict: A longitudinal study of intragroup conflict and group performance. *Academy of Management Journal*, 44, 238-251.

Klein, K. J., and House, R. J. (1995). On fire: Charismatic leadership and levels of analysis. *The Leadership Quarterly*, 6(2), 183-198.

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

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

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

Kozlowski, S. W., Chao, G. T., Grand, J. A., Braun, M. T., and Kuljanin, G. (2013). Advancing multilevel research design: Capturing the dynamics of emergence. *Organizational Research Methods*, 16, 581-615.

Kozlowski, S. W., Gully, S. M., Nason, E. R., and Smith, E. M. (1999). Developing adaptive teams: A theory of compilation and performance across levels and time. Pulakos (Eds.), (1999). *The changing nature of performance: Implications for staffing, motivation, and development*, 240-292.

Lehmann-Willenbrock, N., and Allen, J. A. (2014). How fun are your meetings? Investigating the relationship between humor patterns in team interactions and team performance. *Journal of Applied Psychology*, 99, 1278-1287.

Litrico, J. B., and Choi, J. N. (2013). A look in the mirror: Reflected efficacy beliefs in groups. *Small Group Research*, 44, 658 - 679.

Marks, M. A., Mathieu, J. E., and Zaccaro, S. J. (2001). A temporally based framework and taxonomy of team processes. *Academy of Management Review*, 26(3), 356-376.

McGrath, J. E., Arrow, H., and Berdahl, J. L. (2000). The study of groups: Past, present, and future. *Personality and Social Psychology Review*, 4, 95-105.

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

Mohammed, S., and Nadkarni, S. (2014). Are we all on the same temporal page? The moderating effects of temporal team cognition on the polychronicity diversity - team performance relationship. *Journal of Applied Psychology*, 99, 404-422.

Morrison, E. (2010). OB in AMJ: What is hot and what is not? *Academy of Management Journal*, 53(5), 932-936.

Peeters, M. A. G., van Tuijl, H. F. J. M., Rutte, C. G., and Reymen, I. M. M. J. (2006). Personality and team performance: A meta-analysis. *European Journal of Personality*, 20, 377-396.

Resick, C. J., Murase, T., Randall, K. R., and DeChurch, L. A. (2014). Information elaboration and team performance: Examining the psychological origins and environmental contingencies. *Organizational Behavior and Human Decision Processes*, 124, 165-176.

Roberson, Q. M., Sturman, M. C., and Simons, T. L. (2007). Does the measure of dispersion matter in multilevel research? A comparison of the relative performance of dispersion indexes. *Organizational Research Methods*, 10, 564-588.

Stigler, G. J. (1973). General economic conditions and national elections. *The American Economic Review*, 63, 160-167.

Stoverink, A. C., Umphress, E. E., Gardner, R. G., and Miner, K. N. (2014). Misery loves company: Team dissonance and the influence of supervisor-focused interpersonal justice climate on team cohesiveness. *Journal of Applied psychology*, 99, 1059-1073.

Sundstrom, E. D. (1999). Supporting work team effectiveness: Best management practices for fostering high performance. San Francisco, CA: Jossey-Bass.

Troth, A. C., Jordan, P. J., Lawrence, S. A., and Tse, H. H. M. (2012). A multilevel model of emotional skills, communication performance, and task performance in teams. *Journal of Organizational Behavior*, 33, 700-722.

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

Wallace, J. C., Edwards, B. D., Paul, J., Burke, M., Christian, M., and Eissa, G. (2013). Change the referent? A meta-analytic investigation of direct and referent-shift consensus models for organizational climate. *Journal of Management*, 83, 234-246.

Whiteoak, J. W., Chalip, L., and Hort, L. K. (2004). Assessing group efficacy comparing three methods of measurement. *Small Group Research*, 35, 158-173.

Yang, H. L., and Tang, J. H. (2004). Team structure and team performance in IS development: a social network perspective. *Information & Management*, 41(3), 335-349.

그룹수준의 변수측정 방법결정을 위한 종합적 체계

정현선* · 김현지** · 이지혜*** · 임자은**** · 박원우*****

요 약

조직관련 연구에서 그룹수준의 연구가 증가하고 있음에도 불구하고, 국내 · 외적으로 그룹수준의 변수를 측정함에 있어서는 비교적 단순한 방법론을 채택하여 진행하고 있다. 그 이유는 연구자들이 개인수준에서 측정한 것을 집산·종합한 것이 그룹수준의 변수를 의미한 것으로 가정하기 때문이다. 하지만 이러한 방법론은 그룹 내의 미세한 역학관계를 인지하지 못하는 등 연구의 이해도를 저하시킬 수 있을 뿐만 아니라 측정된 변수에 편향된 오류를 발생시킬 수 있다. 본 연구는 그룹수준의 변수측정을 위한 방법론을 9개로 분류한 종합적 체계를 제시하였다. 2012년부터 2016년까지 조직행동론분야의 대표적 국제학술지의 논문들을 중 그룹수준의 변수를 다룬 141편의 논문들의 검토하여 그룹수준의 변수측정 방법을 9가지 체계에 맞춰 분류 및 정리하였다. 그 결과, 다수의 그룹수준의 연구들이 편중된 방법론을 채택하여 연구를 진행한 것으로 확인되었기에, 본 연구는 보다 다양한 방법론을 채택하여 그룹수준의 변수를 측정할 것을 권하는 바이다.

주제어: 그룹수준, 팀, 측정방법

* 서울대학교 경영대학, 교신저자
** 서울대학교 경영대학, 공저자
*** 서울대학교 경영대학, 공저자
**** Cornell University, 공저자
***** 서울대학교 경영대학, 공저자

- 저자 정현선은 현재 서울대학교 경영대학 박사과정에 재학 중이다. 주요 연구 관심분야는 innovation, motivation, change management, organizational culture 등이다.
- 저자 김현지는 현재 서울대학교 경영대학 박사과정에 재학 중이다. 주요 연구 관심 분야는 creativity, innovation, team process, culture 등이다.
- 저자 이지혜는 현재 서울대학교 경영대학 박사과정에 재학 중이다. 주요 연구 관심 분야는 leadership, motivation, efficacy 등이다.
- 저자 임자은은 현재 미국 Cornell 대학 박사과정에 재학 중이다. 주요 연구 관심분야는 social network, economic sociology, organization 등이다.
- 저자 박원우는 서울대 경영대학에서 학사, 석사를 취득한 후, 미국정부의 Fulbright 장학금을 받고 Pittsburgh대학에서 수학하여 경영학(인사조직) 박사(1989년)를 취득하였다. Pittsburgh대학에서 조교수로 근무한 후 귀국하여 중앙대와 경희대 교수를 거쳐 1998년부터 서울대 교수로 근무하고 있다. 학계에선 한국경영학회 부회장, 한국인사조직학회 부회장, 한국윤리경영학회 회장 등으로 봉사하였으며, 응용과학자로서 다양한 영리/비영리조직의 발전을 지원하여 왔다. 주요 연구분야는 groupthink, empowerment, trust, efficacy, goal orientation, culture change, 및 happiness인데, 그간 130여 편의 국내외 학술논문과 16편의 단행본 도서를 출간하였고, 서울대학교 경영대학의 우수강의상을 수차례, 2018년엔 서울대학교 교육상을 수상하였다.