Behav Res (2011) 43:522–536
DOI 10.3758/s13428-011-0070-3

Assessing group-level participation in fluid teams: Testing a new metric

Susannah B. F. Paletz · Christian D. Schunn

Published online: 17 March 2011

Abstract Participation is an important factor in team success. We propose a new metric of participation equality that provides an unbiased estimate across groups of different sizes and across those that change size over time. Using 11 h of transcribed utterances from informal, fluid, colocated workgroup meetings, we compared the associations of this metric with coded equality of participation and standard deviation. While coded participation and our metric had similar patterns of findings, standard deviation had a somewhat different pattern, suggesting that it might lead to incorrect assessments with fluid teams. Exploratory analyses suggest that, as compared with mixed-age/status groups, groups of younger faculty had more equal participation and that the presence of negative affect words was associated with more dominated participation. Future research can take advantage of this new metric to further theory on team processes in both face-to-face and distributed settings.

Keywords Participation · Teams · Metric · Heterogeneity · Measurement

The typical structures of teams are changing, as is the sophistication of methods involved in studying teams. Much research on teams presumes, methodologically if not conceptually, that each team is composed of the same number of people and that team size is fixed over time. And yet, teams can be fluid in several ways. First, the overall formal membership of the team can change, as when school boards gain and lose members according to local elections or when managers hire new engineers to join a growing division in a company. Using this perspective, researchers have examined the possible effects of newcomers in teams (e.g., Nemeth & Ormiston, 2007) and the socialization of new group members (e.g., Moreland & Levine, 2002).

A second type of group boundary fluidity involves permeability between one group and other groups, such that "at some times, a given group stands on its own and, at other times, it combines into a larger collective with another group" (Poole, Keyton, & Frey, 1999, p. 100). For example, an academic department functions as a whole in departmental faculty meetings but then may function as subgroups in program-specific meetings or in executive committee meetings of program chairs.

Third, team membership may change over the course of a specific, bounded meeting. Even when the greater work group has fairly stable day-to-day membership, who talks to whom, particularly in work groups, may change throughout the day. For example, faculty can arrive late to or leave early from faculty meetings. Particular conversations in online chat rooms or discussion groups can grow or lose members over time. This third type of fluidity typically takes place in the context of the other two types of fluid groups.

All three types of fluidity require measures of participation that balance for different team sizes. The third case is especially complicated for measurement: Participation is a construct of a conversation, or at least of multiple turns, whereas the third type of fluidity can happen at the turn level. We focus on the issue of measuring team participation across all three types of fluidity; outside of experimental lab settings, all three types of fluidity are likely to be common occurrences. Because the third type of fluidity is most problematic, we empirically examine the measurement issue in a context with high temporal fluidity.

S. B. F. Paletz (✉) · C. D. Schunn
University of Pittsburgh,
LRDC 816, 3939 O'Hara Street,
Pittsburgh, PA 15260, USA
e-mail: sbfpaletz@gmail.com

This article has two goals. The first is to present a new metric for assessing group-level participation equality/dominance across all three types of fluidity in workgroup meetings. The second goal is to see how the new metric fares when compared with more traditional measures and, in so doing, explore a set of plausible predictors and covariates for participation equality/dominance.

Throughout this article, we will refer to fluid versus stable workgroup *meetings* as fluid versus stable groups or teams. Fluid meetings within a larger workgroup fall into the definition of the group: They include two or more members who are interdependent because of the nature of their work, are embedded in a social system, affect others via their tasks, and are perceived by themselves and others as a social entity (cf. Guzzo & Dickson, 1996). Even if the specific impromptu meeting does not comprise the entirety of the greater team, individuals inside and outside of the group can identify the meeting as involving that team (e.g., members of an engineering design team had a loud discussion near someone's cubicle).

How common are meetings of fluid size? Formal workgroup meetings may still involve a set number of people sitting in a conference room, but with the implementation of radical colocation (e.g., open office floors) and virtual meeting tools, informal group meetings can and do attract a fluid number of interested parties. For example, a software designer might approach a programmer with a specific question, then another designer and a programmer whose workstations are nearby might suggest other ideas, with the team then approaching their manager to get his or her input. Designers at other locations can quickly be consulted via chat, phone, or video for just part of a meeting. These task-relevant, impromptu meetings are an important feature of how work is conducted today, particularly in settings designed for innovation (Hinds & Kiesler, 2002; Olson & Olson, 2000). Fluid team meetings are also likely important for overcoming boundaries due to multidisciplinarity (Schunn, Crowley, & Okada, 2002).

Researchers who examine team processes in depth need to have the tools to study these types of fluid, work-related meetings just as they study more formal, bounded meetings; they also need methods for analyzing formal, bounded meetings that contain aspects of informal meetings, such as fluid membership, when studied in the wild. In this article, we study one such dynamic environment. Our case entails not only a multidisciplinary group of experts, but also radical co-location, with its resulting fluid team meetings.

Participation is one critical element of team process that is especially difficult to study. Sufficient participation is considered central for multidisciplinary team innovation (see Paletz & Schunn, 2010, for a review). Electronic communication has been examined as a method of democratizing or more evenly distributing participation, with mixed results (e.g., Poole, Holmes, & DeSanctis, 1991; Siegel, Dubrovsky, Kiesler, & McGuire, 1986).

Group fluidity can cause problems for two measurement approaches: self-reports (who should be surveyed?) and turn counting (what is the denominator?). We focus on the challenges of turn counting. Even so, at a time when psychology too rarely measures actual behavior (Baumeister, Vohs, & Funder, 2007), it is particularly important to create a logistically simple metric for this variable that goes beyond self-report. For the purposes of this article, we examine team-level participation equality; specifically, we focus on the degree to which each participant talks for an equal amount (of time or words) or one individual dominates the conversation.

Participation

Social scientists have long studied participation in a variety of ways (Frey, Gouran, & Poole, 1999; Kiesler, Siegel, & McGuire, 1984; e.g., as interaction patterns, Bales, 1950). Participation is a complex construct, including not only interacting in a group setting (e.g., seeking information, suggesting goals, providing feedback, etc.) and providing information, but also premeeting preparation, distractions (e.g., on-time arrival and departure), and nonverbal behavior, among other aspects (Bailey, Helsel-DeWert, Thiele, & Ware, 1983). Communication researchers have often studied participation as a phenomenon for its own sake and have thus developed a wealth of theory and measures, including ones that assess participation in general (Frey et al., 1999).

Communication is necessarily relational, a dynamic interchange between conversants. Researchers have created coding schemes to quantify conversational interactions and reactions to what the previous person said (e.g., Millar & Rogers, 1976; Rogers & Farace, 1975). Some studies have examined the tendency for certain team members to dominate the conversation, whereas others have presented and tested computational models of participation and turn taking (e.g., Stasser & Vaughan, 1996). For example, the longer a conversational partner talked during a speaking turn, the more he or she was perceived as interpersonally dominant (e.g., Folger, 1980; Palmer, 1989). Other research has examined dyad-level dominance as pairs of messages, such as when one speaker makes a domineering message and the other speaker responds with a submissive statement (Courtright, Millar, & Rogers-Millar, 1979).

Within social and organizational psychology, participation—and equality of participation, in particular—has been examined as a key factor in team performance and decision making (De Dreu & West, 2001; Mesmer-Magnus & DeChurch, 2009; Paletz & Schunn, 2010; Stasser & Titus, 1985, 1987). A recent meta-analysis confirmed that information sharing positively predicts team performance (Mesmer-Magnus & DeChurch, 2009). For example, knowledge diversity is thought to be positively associated with team innovation via the team's having access to a broader range of perspectives, information, and opinions (Mannix & Neale, 2005; Nijstad & Stroebe, 2006; van Knippenberg & Schippers, 2007). These different perspectives must be communicated across the group. Participation is vital for minority dissent to lead to innovation (De Dreu & West, 2001). Confirming this line of reasoning, the sharing of *unique* information is more positively predictive of team performance than is the *breadth* of information shared (Mesmer-Magnus & DeChurch, 2009). Only via sufficient participation and information sharing, which are important for groups whose members hold unshared information, can a team take advantage of everyone's diverse knowledge (e.g., Stasser & Titus, 1985, 1987). Research on team performance in general has also suggested the importance of sufficient participation: In a study of four-person groups playing a complex team video game, the most successful groups had more equal participation than did the unsuccessful teams (Fischer, McDonnell, & Orasanu, 2007).

Measuring participation equality/dominance

Unlike studies in which each person's turn or each conversational utterance (i.e., clause or thought statement) is examined for its relational purpose, research examining the role of participation in team performance requires a group-level, block-of-time-level metric: That is, participation equality can be inherently a property of a group, not an individual, and of a segment of time, not an instance. Although some studies have examined individual participation rates (e.g., Straus, 1996), this team-level approach has a long history within studies of the benefits and limitations of electronic communication (e.g., Kiesler et al., 1984; Poole et al., 1991). Participation equality is a construct related but dissimilar to interpersonal dominance, which has been assessed in relational communication as perceptions of attempts to influence and persuade (e.g., Burgoon & Hale, 1987) and as a personality subfactor (such as in extraversion in the Big Five; John & Srivastava, 1999).

In addition, there is the issue of measuring equity rather than equality of communication. Equity—to each according to his/her need, from each according to his/her knowledge/ability—can be an important measure of sufficient information sharing (Kiesler et al., 1984). Perceived equity can be assessed using self-report round-robin surveys, where the participants assesses themselves and each other (e.g., Jarvenpaa, Rao, & Huber, 1988). But equity has a number of troubling aspects. First, it is difficult to measure objectively in real-world settings without first thoroughly assessing participants' background knowledge. In experiments, it is possible to measure equity: In the hidden profile research paradigm, researchers can manipulate the information each participant has and then determine the proportion of the time different kinds of information are discussed (e.g., Schulz-Hardt, Brodbeck, Mojzisch, Kerschreiter, & Frey, 2006). Second, in naturalistic settings, the problems are so complex that it is not clear in advance which team members' expertise will prove to be more germane to the current task. In some settings, with multiple possible solutions, it may never become clear which expertise was necessarily more relevant. It may be that a general practice of equal participation is a better policy than assuming in advance that particular team members have more to contribute. Thus, in settings where information is not experimentally manipulated, researchers rely on measuring *equality* of communication (see Kiesler et al., 1984; Zmud, Mejias, Reinig, & Martinez-Martinez, 2002).

Participation equality has often been examined as self-reported perceptions (Zmud et al., 2002). Self-report surveys and interviews may be useful in capturing perceptions, attitudes, and an understanding of the norms and relationships within a group (Poole et al., 1999; see, e.g., Berdahl & Craig, 1996; Burke & Chidambaram, 1995; Tyran, Dennis, Vogel, & Nunamaker, 1992). For example, Berdahl and Craig used round-robin self-report items to assess the perceived participation of each team member. Nevertheless, self-reported participation may not reflect actual behavior (Poole et al., 1999). Self-reported participation may be correlated only moderately with observer ratings of participation: Bailey et al. (1983) found, at best, a correlation of .34 between self-reported participation and dimensions of observer ratings of participation. Although perceptions of participation dominance are likely to have psychological correlates, more objective measures of participation equality/dominance are necessary at times. For example, when participation reflects on-task behavior (communication of information, problem solving, decision making), the actual participation, rather than the perceived participation, is likely to be more important to team functioning.

Communication researchers often measure interactions by coding observed behavior (Poole et al., 1999). One type of observation method involves using trained observers to watch teams at work and then either assess the interaction patterns (e.g., Bales, 1950; Fischer et al., 2007) or rate the participation of the team members on standardized scales (e.g., Bailey et al., 1983). These methods are useful if one has a way to embed observers or observe unobtrusively, or if the researcher has access to an audio–video record. Observer coding may be sensitive to nuance and context, but the observers must cover many team meetings to gain in-depth information about a particular team's dynamics. In addition, observers may have inherent biases regarding participation across different group sizes, and subjective measures (including assessments by trained observers) can involve considerable extra time to ensure reliability and validity of coding.

Another common method involves simply counting turns (or length of turns) for each member and then calculating participation from these easily coded counts. This turn count measure is especially practical when transcripts of speech exist or when the team communications are electronic. Even in live-coding or coding-from-video cases, it is relatively straightforward. Second Messenger is a recent technological advance in which each participant wears a noise-canceling microphone that automatically captures aggregate time-stamped segments of individual speech (DiMicco & Bender, 2007).

The most typical derivation of participation involves the standard deviation of number of remarks or word count over a set period of time (e.g., Fischer et al., 2007; Kiesler et al., 1984; Siegel et al., 1986; Zmud et al., 2002). Sometimes, researchers also control for the mean total participation units in a meeting (Jarvenpaa et al., 1988) or calculate the standard deviation of the proportion of speech acts each member makes (Poole et al., 1991). Another variation is the average *relative* standard deviation of group members' participation rates (Siegel et al., 1986). The larger the standard deviation, the more inequality exists in how much each individual spoke, and the more likely that one person dominated over the others. These standard deviation metrics do not offer a clear method for adjusting for group size when researchers wish to compare across different group sizes.

Another, less common metric that does adjust for different group sizes per meeting, albeit not varying over time, entails (1) counting the number of lines of text written or uttered by each participant and (2) determining the difference between the proportion of comments or lines that there would be if there were complete equality and (3) the observed distribution (Hiltz, Turoff, & Johnson, 1989). However, this measure and the standard deviation measures are predicated on group size being stable *within* a meeting. As was discussed earlier, in many real-world settings, the number of participants in a meeting fluctuates. For researchers interested in objective measures of participation equality/dominance, the fluid nature of meeting membership necessitates the development of a new measure. For these cases, we have created such as measure of participation, P_s.

The new metric

Intuitively, the measure is based on the number of people present at any given point in the transcript and gives credit for participation to each person in proportion to the number of people present. For example, a person is more saliently participative in a given utterance (i.e., clause or thought statement; Chi, 1997) when he or she speaks in a group of 10 than when he or she speaks in a group of 3. Similarly, people not speaking a given utterance are more saliently non-participative when they fail to speak in a group of 3 than when they fail to speak in a group of 10. Our measure adds up these relative participative /non participative moments over a block of time and then computes an aggregate, mean participation level for individuals in the group that is on the same 0 (*equal*) to 1 (*dominated*) scale for all group sizes.

Our metric can be used for any length of meeting (Paletz & Schunn, 2009). It also need not be limited to audio–video data or time/turn data captured using a tool such as Second Messenger: This metric can be applied to any type of written words, including chat room transcripts, online multiplayer game transcripts, Twitter threads, and comments on blogs or in news stories. We formulated this metric to be used on conversations that are separated into turns or speech acts, much as in cognitive psychology research in which each utterance is coded (Chi, 1997). The metric computation is easily implemented in spreadsheets of coded transcripts, codebooks of speech turns, or lines of written text (as in Hiltz et al., 1989). These utterances can be shorter than turns, and a person may speak several utterances consecutively, depending on the grain size of analysis. The reason for segmenting turns into separate utterances for participation coding is that long diatribes should count more than simple affirmations toward relative dominance of a conversation, although this can also be approximated with measures that weight participation in each turn by word counts or turn length. Equation 1 shows the computation of the overall metric, but we will present its creation step by step.[1]

$$P_s = \bar{n}^2 \bullet \frac{\sum_{i=1}^{N} \left| \sum_{K=1}^{M} f(n_k, i, K) \right|}{2(\bar{n}-1)\sum_{i=1}^{N} m_i}, \tag{1}$$

where

n_k	is the number of people present on utterance (line) K
M	is the maximum number of utterances spoken in the block, segment, or clip being studied
i	is the index of a particular person in the group
N	is the number of people ever present in the block, segment, or clip being studied
m_i	is the number of utterances (lines) when person i is present
$\bar{n}$	is the average number of people

[1] We have created a set of Excel spreadsheets that calculate this formula, which we are happy to share with interested readers.

present per utterance in the block, segment, or clip being studied ; and

$$f(n_k, i, K) = \begin{cases} \frac{n_k - 1}{n_k}, & \text{if } i \text{ is present and speaking utterance K} \\ \frac{-1}{n_k}, & \text{if } i \text{ is present and } \textit{silent} \text{ during utterance K} \\ 0, & \text{if } i \text{ is absent during utterance K.} \end{cases}$$

The first step in computing this metric is to assess the function for the three conditions detailed above for each utterance or line of text spoken. This aspect of the metric takes into account variable group size, weighting each person's utterance on the basis of how many people are present at that moment. The more people present, the heavier the person speaking is weighted, whereas those not speaking are penalized less. For example, if two people are present, the speaker gets the number1/2 and the other 1/2. If three are present, the speaker gets 2/3, and each silent party gets 1/3. Similarly, if six people are present, the speaker gets 5/6 for that utterance, and each silent person is given 1/6. Individuals not present during that part of the conversation (i.e., people who show up later in the time period studied) are given a zero (see Table 1). Assuming equal participation across utterances (e.g., a person in a group of six participates one sixth of the time), the measure sums to zero across lines. Also note that the measure sums to zero within a line: Participation credit given to the speaker equals the sum of the participation debits taken away from the nonspeakers.

In the second step, each person's speaking-or-not numbers given in the first step are summed across all utterances in the block of time analyzed (say, 25 utterances, about 1 min; see Table 1). The third step involves taking the absolute value of each speaker's sum across the utterances (see Table 1, last row): Equal participation is zero in the line formula, and an individual's participating below or above the equal participation rate contributes toward nonequal participation. The next element is to compute a weighted average across the participants (steps 4 and 5). This weighting takes into account an important issue: Not every individual is present for the entire block of utterances. If one individual shows up for a brief amount of time, makes a request, and leaves, that activity would skew the formula toward dominance, even if the majority of the block involved a fairly equal conversation between other individuals. The weighting is accomplished by steps 4 and 5. The fourth step computes a participation sum across all the individuals who are present during the block of time with their respective numbers created by the third step. In other words, the participants' absolute values, as created in step 3, are added together (e.g., in our example [see Table 1], add together 0.83, 0.17, 0.83, 2.50, 0.50, and 0.50 to get 5.33; see Table 2). This participation sum gives a number that is truly at the group and block-of-time level. Then the fifth step involves controlling for the number of lines for which each person is present. This is done via summing the total of the *number* of lines for which each person is present and dividing the result of step 4 by this number (see Eq. 1). In Table 1, persons 1 and 2 are each present for 9 utterances, person 3 for 7 utterances, and persons 4–6 for 3 utterances each, making the sum of each individual's number of utterances equal to 34. The number in step 4 (5.33) is divided by 34 to give 0.16 (Table 2). Finally, in order for the metric to range from zero to one, it needs to control for the maximum amount possible. Otherwise, larger groups would have a larger possible maximum as a result of the absolute value operation adding both below and above equal participation rates. This renorming to a zero–one scale is accomplished by the last step, step 6, which involves multiplying the number generated by the

Table 1 Steps 1–3, utterance-level assignments: one example

Utterance	Speaker	Person					
		1	2	3	4	5	6
Have you received the file yet?	1	1/2	1/2	0	0	0	0
No	2	1/2	1/2	0	0	0	0
Wait, which file?	2	1/3	2/3	1/3	0	0	0
Have you guys seen the file?	3	1/3	1/3	2/3	0	0	0
No	2	1/3	2/3	1/3	0	0	0
Yes	1	2/3	1/3	1/3	0	0	0
I just sent this great document,	4	1/6	1/6	1/6	5/6	1/6	1/6
did you guys get it?	4	1/6	1/6	1/6	5/6	1/6	1/6
It's about the new picture we just got.	4	1/6	1/6	1/6	5/6	1/6	1/6
Step 2: Sum across utterances		0.83	0.17	0.83	2.50	0.50	0.50
Step 3: Absolute value of step 2		0.83	0.17	0.83	2.50	0.50	0.50

Table 2 Steps 4–6: examples

Type of Group	Step 4	Step 5	Step 6: P_s (0, equal to 1, dominated)
Mixed, fluid (Table 1 example)	5.33	0.16	0.40
Dominated 2-person group	8.00	0.33	0.67
Dominated 3-person group	12.00	0.33	0.75
Dominated 6-person group	16.00	0.22	0.80
Dominated, fluid 2- to 4-person group	14.50	0.35	0.85
(Almost) equal 2-person group	2.00	0.08	0.17
Equal 3-person group	0.00	0.00	0.00
Equal 6-person group	0.00	0.00	0.00
(Almost) equal, fluid 2- to 4-person group	2.5	0.06	0.15

fifth step by the following formula (Eq. 2), which is embedded in Eq. 1:

$$\frac{\overline{n}^2}{2(\overline{n}-1)}. \tag{2}$$

As was noted above, n is the average number of people present per utterance in the block, segment, or clip being studied. This average is created by taking every utterance and counting how many people are present at that time, and then taking the average of that number across the block or segment that is being analyzed. Thus, the new measure, labeled P_s, ranges from zero to one, with zero for entirely equal participation and one for complete dominance of the discussion block by any single individual. The metric is at the level of the group for a particular block of time. Table 2 shows examples of what occurs in each of steps 4–6, including the example in Table 1. Other examples include mostly equal and strongly dominated groups of two, three, and six, as well as mostly equal and strongly dominated fluid groups of two to four.

Exploring participation equality/dominance in fluid teams

In this study, we set out to compare this new metric with more traditional measures of participation equality/dominance. These more traditional measures are (1) the standard deviation of words spoken during a block of time, controlling for block size, and (2) participation equality/dominance as judged by coders. The first measure, like our own, has the advantage of being an objective formula and, thus, is easy to implement with high reliability. We expect that even given a substantial proportion of blocks of time with fluid teams, P_s will be positively and strongly correlated with the traditional measures of participation equality/dominance. But because the traditional standard deviation metric ignores changing group sizes, the correlation between the new metric and the standard deviation measure will be lower under conditions of fluid teams.

> *H1a:* P_s will be positively correlated with the other participation equality/dominance measures, providing convergent validity.
>
> *H1b:* P_s will still be positively correlated with the other participation equality/dominance measures under conditions of fluid versus stable groups, but the correlation will be smaller.

In addition to examining its convergence with the other measures, we also wish to test several likely predictors of participation and compare these sets across the three measures, with each participation measure a dependent variable. As a first exploration of measuring participation in fluid team settings, our predictors are mainly face-valid factors.[2]

First, given that standard deviation does not take into account fluid versus stable teams, we expect fluidity versus stability of teams to have a significant effect on the standard deviation measure. In other words, the bias for the type of team should be measurable. It should not significantly impact coded participation or our new metric (P_s), because these measures, in different ways, take into account the natural fluidity of the team.

> *H2:* Fluid versus stable team meeting membership will be significantly confounded with participation as measured by standard deviation, but it will not be significantly associated with the new metric (P_s) or coded participation.

Theoretically, of the three metrics, standard deviation in the face of mixed stable/fluid teams should be the least accurate measure of participation equality/dominance. Coded judgments that take into account the fluid/stable

[2] It is worth noting that this is not a traditional multimethod–multitrait study, because we are not testing the construct validity of participation equality/dominance per se. We are simply testing this particular formula by comparing it with more traditional metrics.

team membership should be a more accurate measure of equality/dominance. We propose that our new metric is more objective and less time consuming to create (if written records already exist) than are coded judgments but should be equally valid. In this next section, we propose a set of specific hypotheses (4–6), which are tested for each of the three metrics separately in order to determine whether the pattern of results is similar between our new metric and the other two types. In essence, we are testing the relative utility of the three metrics in theory testing. Regardless of what occurs in terms of other specific hypotheses (hypotheses 4–6; see below), in general, we expect that the pattern of results for P_s will be similar to the pattern of results for coded participation equality/dominance but that both of them should be different from the pattern of results for standard deviation of words. This is, in a sense, a meta-hypothesis, but it is in line with the goals of the study.

> *H3:* Given all three participation equality measures, the pattern of results will be similar between the new metric (P_s) and coded participation but different for standard deviation. In other words, we expect the new metric and coded participation to be predicted by similar factors, whereas standard deviation's pattern of results will diverge.

The goal of testing the three hypotheses below is to show the empirical utility of the new measures. Thus, some provide only tentative theoretical value, whereas others are primarily replications to provide tests for H3. We select a range of variables that can influence participation but are not themselves measures of participation, thus providing another method of examining concurrent validity for our metric (hypothesis 3).

Team demographic composition: Gender and age/status A series of reviews suggest that surface-level diversity, such as social category differences, will have negative effects on team processes and performance (e.g., Joshi & Roh, 2009; Mannix & Neale, 2005; van Knippenberg & Schippers, 2007). This negative effect is explained using social identity and self-categorization theories (as well as similarity attraction theories), suggesting that demographic diversity can hurt social integration and cohesion via disrupting in-depth processing of task information (van Knippenberg, De Dreu, & Homan, 2004; see also Mannix & Neale, 2005; van Knippenberg & Schippers, 2007). A recent meta-analytic review of the literature suggested that gender diversity in teams is more likely to have negative effects on team performance in occupations where gender is not balanced (Joshi & Roh, 2009). Our data set is one of these cases. In addition, a separate meta-analysis argued that gender composition has an effect on talkativeness, such that men are likely to talk more than women in mixed-gender groups, as opposed to same-gender groups (Leaper & Ayres, 2007). If one assumes that group cohesion and performance are heightened by equal participation, these combined findings would suggest that mixed-gender groups will have less equal participation overall.

> *H4:* Mixed-gender groups will have less equal participation/more dominated participation than will same-gender groups.

Age diversity tends to have one of the weakest effects on performance (Mannix & Neale, 2005). Although Joshi and Roh (2009) warned against possible effects of age stereotyping on older workers, our particular data set might have the opposite effect. Our data set involves scientists from academia and government, such that older workers are more likely to be senior, tenured faculty and have higher status in general. The approach/inhibition theory of power suggests that power leads to disinhibition and attention to rewards (Keltner, Gruenfeld, & Anderson, 2003). Following this line of reasoning, Berdahl and Martorana (2006) found that high-power individuals were more likely than low-power individuals to feel that they openly expressed their opinions. Thus, we would expect that mixed-age groups would be more likely to be dominated, as compared with homogeneous groups.

> *H5:* Mixed-age groups will have less equal participation/more dominated participation, as compared with homogeneous old or younger groups.

Expressed affect It is possible that participation equality/dominance is associated with the type of affect expressed. Affect management is an important interpersonal process in teams (Marks, Mathieu, & Zaccaro, 2001). Managing emotions within the team entails controlling frustrations, boosting social cohesion and morale, and mitigating interpersonal problems between members (Marks et al., 2001). On the other hand, unsuccessful team building can increase negative affect (Marks et al., 2001). This framework implies that participation equality will be correlated with heightened positive affect. Indeed, in Fischer et al. (2007) study, the successful teams had more equal communication and expressed more positive affect, and unsuccessful teams were more likely to be defensive and insult each other. Although this hypothesis is inherently correlational, this particular hypothesis also provides support for the possible predictive validity of an association between participation dominance/equality and a type-of-group outcome measure.

> *H6:* Participation dominance will be positively associated with expressed negative affect and negatively associated with positive affect.

Method

Research context: The Mars Exploration Rover (MER) mission

For this study, we used audio–video data from one large case that included many smaller semiindependent teams. The overall team experienced large gains in efficiency over several months of activity (Tollinger, Schunn, & Vera, 2006). The case involved scientists on the Mars Exploration Rover (MER) mission, where two rovers were sent to opposite sides of Mars in order to dig, analyze, and drive to discover whether Mars had a history of liquid water. The Spirit rover arrived on Mars first, and the Opportunity rover landed about 3 weeks later. Different scientists worked on MER A (Spirit) and B (Opportunity), occasionally switching subteams. Within each rover team, the scientists were further grouped by discipline (e.g., geology, geochemistry, soil science, and atmospheric sciences). The social identity of the subteams was thus to the superordinate group (MER), to the specific rover subteam (A vs. B), and to the discipline subgroup. The scientists were colocated at the Jet Propulsion Laboratory in Pasadena for the first 90 days of the project, with each rover subteam on a different floor of the same building. During these first 90 days, most of the work of discovery and planning took place in face-to-face meetings in one of two large, open rooms, one for each rover team, where different discipline groups had clustered workstations and large shared screens. Some of the MER scientists' meetings were formal and had a stable number of participants within a meeting, but many meetings were informal, occurring naturally as scientists walked up to others' workstations and started conversations. The parallel science teams for each rover were, by their nature, both multidisciplinary and a hub of activity for members of disciplinary subteams with questions, suggestions, and concerns. In these impromptu groups, two individuals might be sitting at their workstations chatting when a third scientist approached them; a fourth might join, and then one of the original scientists might leave.

Participants, clips, and blocks

The greater MER science team had over 100 members during the first 90 days of the mission, almost all of which appeared at some point in the large video data set. This study involved 11 h and 25 min of informal conversation chosen on the basis of the audibleness of the conversations and, also, when the conversations naturally began and stopped. These conversations will be referred to as *clips*. The audio–video clips were transcribed into 12,336 utterances (thought statements or clauses). We coded whether each utterance was on-topic talk or not, to exclude conversations irrelevant to the MER mission (κ = .96). The analyses were conducted on the remaining 11,856 utterances (roughly 11 h) of on-topic talk. The remaining 114 clips were from 8 to 760 utterances long (M = 104, SD = 122).

Because the conversations during the clips flowed and changed in terms of both topic content and number of conversationalists, we segmented the data at the level of the block. One could have segmented blocks by group size at the moment (i.e., when new individuals came and went). However, some blocks could then be very short and others very long, with the latter greatly diluting the ability to track changing participation of those present against other changing variables. Furthermore, we specifically wished to examine whether group size influenced participation, rather than defining participation such that group size *must* influence participation. Rather than being created by coders, the blocks were created by taking sets of 25 utterances, each of which was roughly a minute long. In addition, the onset of a small subset of blocks was based on their proximity to analogies (a topic of another article; Paletz, Schunn, & Kim, 2011) and by their proximity to the beginning/end of the clip: Because clips varied by length, a block might have been smaller than 25 utterances long, particularly if it came at the end of a clip. Thus, utterances were nested inside blocks, which were nested inside clips. In total, there were 673 blocks (M = 17.6 utterances, SD = 9.5 utterances), but these analyses are limited to those with blocks of 5 or more utterances, or 549 blocks (M = 21.3 utterances, SD = 6.1 utterances). Due to their length, these smaller blocks offered unreliable estimates of all of our variables and, thus, were removed from analysis.

Overall, clips ranged from 2 to 10 total participants. Fifty-three percent (60) of the clips had fluid numbers of team members, where individuals would come into and go from the conversation. As an extreme example, one clip had as many as 10 participants and as few as 3 participants in the conversation. Participants could leave a conversation by physically leaving, but also by getting involved in another, untranscribed conversation or by being deeply engaged in something else (e.g., their work, computers, etc.). Scientists were counted as participating in the conversation if they were both present and attending to (listening, watching, etc.) the conversation. If they seemed not to be attending to the conversation but were present (i.e., working), but then asked a question that showed they had been listening, they were then included as being present during that earlier period. Individuals who joined the conversation by speaking from offscreen were counted as being present. At the relatively microscopic level of measurement, even 30% (163) of the block-level groups were fluid. That is, even within a minute of a conversation, a team member commonly arrived or left the conversation.

In terms of gender composition, 72% (393 of 549) of those blocks had only males present, and 28% (156) were a mix of males and females. We also visually estimated the ages of the scientists present as older faculty, younger faculty, and students (who were relatively rare). In this setting, age and status are highly correlated, unlike in some other academic areas, where students can be in their 50 s. On the basis of this coding, 28% of the blocks (155 of 549) had only young faculty and students present, 15% (80) had only older faculty present, and 56% (305) had a mix of older and younger individuals present (1.6% were missing due to an inability to estimate the age, usually due to someone's speaking out of view of the camera).

Measures

Participation equality/dominance Participation was assessed for each block of utterances in three ways: (1) through having two independent coders rate the equality/dominance in team participation that occurred in each block, (2) through taking the standard deviation of number of words spoken per person in a block, divided by the total number of words in that block to control for block size (similar to Jarvenpaa et al., 1988), and (3) via the new metric described above that took into account the constantly changing number of participations so as to have total equal participation (0) and total dominance by one person (1). For all three measures, higher numbers reflected more inequality (dominance) in team participation.

The two independent coders were undergraduates who were trained in subjectively assessing each block of transcribed text for its level of equality/dominance. They were told to use a 0 to 100 scale, where 100 indicated that one person dominated the entire conversation or only one person spoke during the entire block and 0 indicated that everyone present participated equally. For each utterance, the transcript file included not only the associated speaker number, but also who was present in the conversation for the utterance (e.g., speaker #1, speaker #3). The coders were encouraged to consider in their ratings how long in the block each person spoke in terms of both number of words and number of utterances, how many people were present at each utterance, and whether the number of potential speakers changed. They were also warned not to try to make up a formula but to use subjective judgments. The coders were blind both to this study's hypotheses and to how the new metric was calculated. The reliability for the coded participation was excellent: The intraclass correlation (ICC) from the two coders was .90 (95% confidence interval from .89 to .92). Differences between coders of 15 points or greater on the 0 to 100 scale were discussed with the first author until consensus was reached. The coders and the first author also explicitly discussed blocks with differences of less than 15 but where one of the coder's assessments was 0 or 100, because those end points represented absolute statements about equality and dominance. Differences of less than 15 where a rating was not 0 or 100 (e.g., coder A gave a 40 and coder B gave a 45) were simply averaged.

Team context covariates We also examined as covariates two context variables on the different participation metrics: the different rover subteams and the time during the first 90 days of the mission. Spirit had some serious technical problems soon after it landed (Squyres, 2005). Spirit's science mission team (MER A) discovered evidence for water toward the end of the 90 days, but due to where Opportunity had landed, and being able to take advantage of lessons from MER A, its science team (MER B) discovered evidence of liquid water much earlier. Testing for rover team and for early/late in the first 90 days of the mission was thus important both for determining generalizability and for accounting for these major differences in rover science team experience. At the clip level, 44% of the clips were from MER A and 65% were from before day 50 of the mission (vs. days 50–90).

Expressed affect To measure affect, we utilized Pennebaker, Booth, and Francis's (2007) Linguistic Inquiry and Word Count (LIWC), a computer program that identifies specific affect words in text (Pennebaker, Mehl, & Niederhoffer, 2003). The identified words include both positive and negative affect words, with negative affect words including anxiety, sadness, and anger words. The LIWC has been validated by comparison with human coding (see Pennebaker, Chung, Ireland, Gonzales, & Booth, 2010—specifically, their Table 1) and has been used successfully in other team studies that included the analysis of conversational transcripts (e.g., Fischer et al., 2007). For the present analyses, we dichotomized these variables so that they simply indicated the presence versus absence of positive or negative affect words in the blocks. Because 95% of the blocks had positive affect words, we limited our analysis to examining the presence of negative affect words (58% of blocks vs. 42% of blocks with no negative affect words).

Analyses

We examined the correlations between the different participation measures, using Spearman rho correlations at the block level.[3] Because each of the participation measures had significant clip-level variance tested via the base hierarchical linear model, we used hierarchical linear modeling (HLM) with blocks nested within clips to examine the effects of various factors on each of the

[3] Although these variables follow roughly normal curves, Shapiro–Wilks tests reveal that they are not strictly normal.

participation measures. Gender composition was a single variable of mixed-gender versus all-male groups. Age composition was two vectors, comparing homogeneous older faculty groups with mixed-age groups and homogeneous younger groups with mixed-age groups. Negative affect was either present or absent in a block. Random effects were fixed at zero if the parameter estimates were not significant. Test statistics of parameter estimates using robust standard errors were used, which adjust for non-normality and heteroscedasticity (Raudenbush & Bryk, 2002). In addition to controlling for science mission team (MER A vs. B) and time in mission (early or late), given the different sizes of the blocks, in each analysis we controlled for number of utterances per block, which was a significant variable for all three dependent variables.

Results

We tested the six hypotheses listed in the introduction. In addition, in creating the three different base hierarchical linear models, we ascertained the ICC reliability of the three measures as how participation equality/dominance is different between clips and the same within clips. This HLM-generated ICC is not the same as coder reliability but is a measure of clustering—in our case, between clips/ conversations. Our new metric shows almost as much clustering as standard deviation, and both were more sensitive to the multilevel nature of the data than was coded participation (P_s = .53, standard deviation metric = .57, and coded participation = .40).

H1: Correlation between participation measures First, we examined whether the three different participation measures correlated with each other. The correlations between the P_s metric and the coded participation and standard-deviation-derived measures were significant and strongly positive (see Table 3), as was predicted. These correlations provide convergent validation for the P_s metric.

In addition, we examined the correlations between the three participation metrics under two conditions: when the groups were fluid and when they were stable (i.e., no change in group membership during the block). When blocks of time involved groups that were stable, all three were highly correlated (see Fig. 1; n = 386; P_s and coded participation, r_s = .81, p < .001; P_s and standard deviation metric, r_s = .75, p < .001; and coded participation and standard deviation metric, r_s = .80, p < .001). When groups were fluid such that members came and went during the course of the conversation, P_s was still highly correlated with participation as judged by coders, but P_s and standard deviation of words were not correlated as strongly as before (n = 163; P_s and coded participation, r_s = .64, p < .001; P_s and standard deviation metric, r_s = .34, p < .001). Coded participation was not as highly correlated with standard deviation of words as it was with P_s, either (see Fig. 1; n = 163; r_s = .54, p < .001).

Hypotheses 2–6: Patterns of predictors of participation equality/dominance In order to test hypotheses 2–6, we conducted three parallel hierarchical linear models (Tables 4, 5, and 6). First, we tested the covariates. For all three participation measures, the number of utterances per block was significant. Mission rover team was not significant for any of the participation measures, but time in the mission was marginally significant for P_s and significant for the standard deviation measure, such that later conversations in the first 90 days tended to be dominated by a few individuals.

As was expected, whether a block had fluid versus stable team membership was a significant predictor/confound for the standard deviation participation measure, but not for the other two metrics (hypothesis 2; see Tables 4, 5, and 6). Fluid teams were more likely to be judged by the standard deviation measure as having more equal participation. This bias was not found for the other two participation measures.

Overall, the pattern of results between P_s and the coded participation measure were quite similar (hypothesis 3). Figure 2 shows the significant and marginally significant relationships for these HLM analyses across the three

Table 3 Correlations, means, and standard deviations of participation measures

Variable	*M (SD)*	Correlations (Spearman Rho)	
		1	2
1. P_s	.40 (.21)		
2. Standard deviation of words	.26 (.13)	.66**	
3. Coded participation	52.19 (21.39)	.76**	.74**

* p < .05, ** p < .001, N = 549

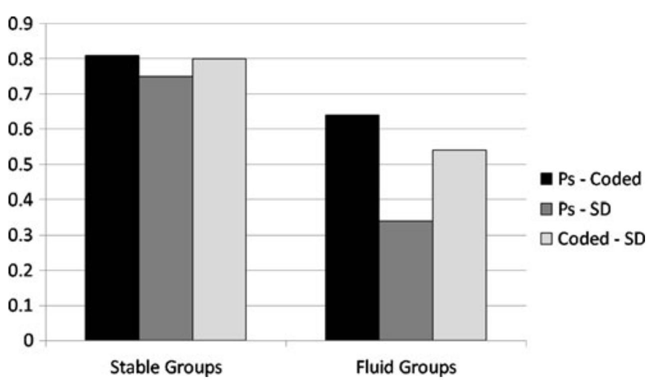

Fig. 1 Correlations between P_s, coded participation, and standard deviation (*SD*) of words by stable versus fluid groups

Table 4 Fixed effects model of fluid/stable team, team composition, and covariates on P_s

Independent Variable	γ Coefficient	*SE*	*t*	*df*	*p*
Covariates					
Number of utterances per block	0.007	0.002	4.45***	531	< .001
Mission science team (A vs. B)	0.007	0.031	0.25	110	.81
Early/later in first 90 days	0.054	0.029	1.85^{+}	110	.067
Predictors					
Fluid versus stable teams	0.027	0.019	1.43	531	.15
All males versus mixed-gender	0.002	0.025	0.08	531	.94
Homogeneous older versus mixed-age groups	0.042	0.048	0.88	531	.38
Homogeneous younger versus mixed-age groups	0.084	0.031	2.70**	531	.008
Presence of negative affect	0.040	0.017	2.39*	531	.017

$^{+}$ .10 < *p* < .05, **p* < .05, ***p* < .01, ****p* < .001

measures. There were two exceptions, but even these were not clearly different: The marginal effect for time during the first 90 days (early vs. late) for P_s was not significant for coded participation. Second, the significant effect for homogeneous younger teams for P_s was marginal (but at $p = .052$) for coded participation. These findings would suggest that the composition effect is small but real and that the effect for time during the mission is negligible. Standard deviation had a noticeably different pattern of results (see below and Fig. 2).

In terms of the specific predictors, mixed-gender groups were not significantly more dominated than all-male groups (hypothesis 4). There was a marginal effect for the standard deviation metric (see Table 6) such that mixed-gender groups were more likely to have equal participation, but it was not confirmed by the other two participation measures.

In terms of age/status composition (hypothesis 5), as was noted above, there was a small but significant (for P_s) and marginal (for coded participation) effect for the second age vector, such that homogeneous groups of young faculty/students had more equal participation than did mixed-status groups. This effect was not found for the standard deviation participation measure, even given that it controlled for group membership stability.

We found a consistent modest relationship across all three participation metrics, such that blocks with negative affect words were more likely to be dominated (i.e., have lower participation equality; hypothesis 6).

Discussion

Measuring participation The primary goal of this study was to propose and test a new metric for measuring participation equality/dominance in a real-world dynamic team: that of informal, task-relevant meetings where group members naturally come and go. Researchers wishing to understand team processes have good reasons to be interested in participation equality/dominance. Through participation, unique ideas can be shared for the greater good of a team. But real-world teams are at times fluid, sometimes even at the minute-by-minute level, and this fluidity creates significant measurement challenges. In particular, this is true of dynamic work environments, especially cross-disciplinary ones.

This study showed that our new metric (1) correlated highly with more traditional measures of participation equality/dominance in the more traditional stable case, (2) differentiates from the traditional syntactic approach in the

Table 5 Fixed effects model of fluid/stable team, team composition, and covariates on coded participation

Independent Variable	γ Coefficient	*SE*	*t*	*df*	*p*
Covariates					
Number of utterances per block	0.415	0.152	2.73**	531	.007
Mission science team (A vs. B)	1.478	2.876	0.51	110	.61
Early/later in first 90 days	1.859	2.896	0.64	110	.52
Predictors					
Fluid versus stable teams	1.470	1.797	0.82	531	.41
All males versus mixed-gender	1.098	2.328	0.47	531	.64
Homogeneous older versus mixed-age groups	1.352	4.866	0.28	531	.78
Homogeneous younger versus mixed-age groups	5.150	2.654	1.94^{+}	531	.052
Presence of negative affect	4.934	2.044	2.41*	531	.016

$^{+}$ .10 < *p* < .05, **p* < .05, ***p* < .01, ****p* < .001

Table 6 Fixed effects model of fluid/stable team, team composition, and covariates on standard deviation of participation

Independent Variable	γ Coefficient	*SE*	*t*	*df*	*p*
Covariates					
Number of utterances per block	0.005	< 0.001	4.93***	531	< .001
Mission science team (A vs. B)	0.030	0.023	1.33	110	.19
Early/later in first 90 days	0.041	0.188	2.17*	110	.032
Predictors					
Fluid versus stable teams	0.028	0.010	2.75**	531	.007
All males versus mixed-gender	0.028	0.017	1.71^{+}	531	.088
Homogeneous older versus mixed-age groups	0.035	0.026	1.32	531	.19
Homogeneous younger versus mixed-age groups	0.010	0.021	0.49	531	.63
Presence of negative affect	0.027	0.011	2.54*	531	.012

$^{+}$ .10 < p < .05, *p < .05, **p < .01, ***p < .001

fluid membership case, and (3) is consistent with human coded participation in the fluid membership case. As a similar indicator, whether the group was fluid or stable had a main effect on the standard deviation measure, but not the coded measure or our new metric. Finally, the pattern of results for the new metric most closely matched that for coded participation. From these results, we suggest that standard deviation is a poor measure of participation in data with a substantial frequency of fluid membership. In our case, 30% fluidity was enough to produce substantial differences in results. That is, researchers might obtain incorrect results if they relied on such a mathematically biased measure, even when controlling for fluid/stability status of the groups, as we did.

Although we considered it the most potentially accurate measure, coded participation had its own shortcomings that could be overcome by our new metric. Calculating P_s is less time consuming than coding for participation equality/dominance, assuming that transcripts or, at least, turn counts have already been obtained. Although calculating P_s is more time consuming if one includes taking turn counts or creating transcripts, transcripts and the like are useful to researchers for computing numerous other variables as well. This metric also works well in situations in which participant conversation is automatically recorded, such as in online settings. Furthermore, examining the base hierarchical linear models suggested that our new metric's HLM-generated ICC reliability was more sensitive to the

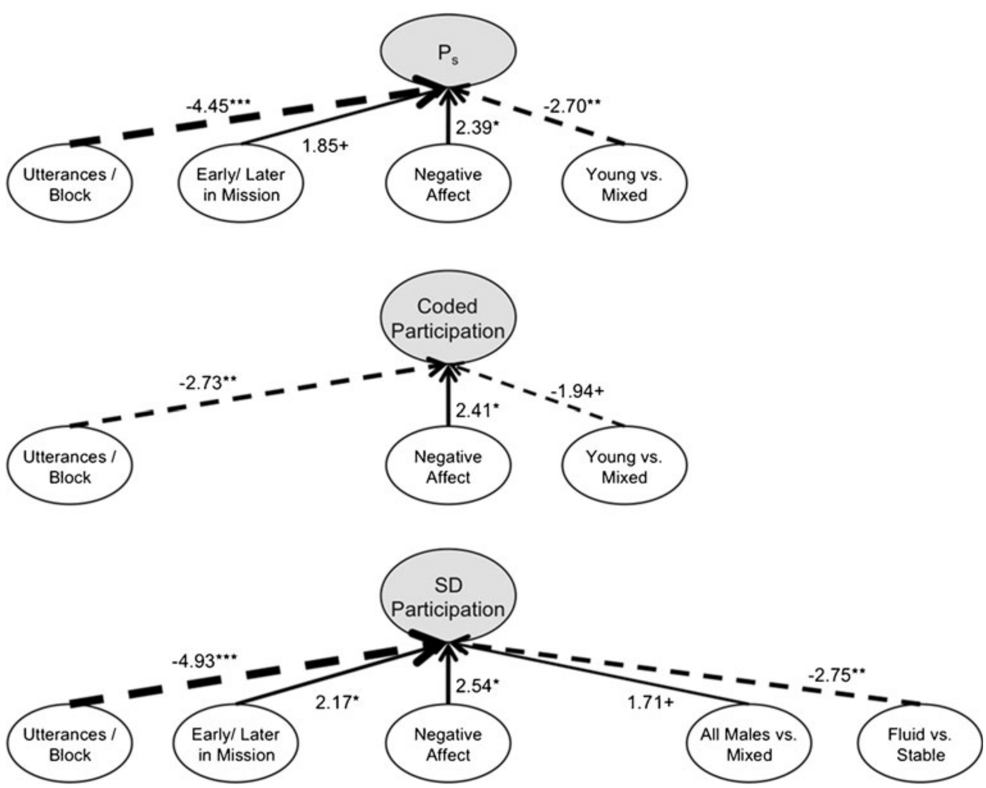

Fig. 2 Hierarchical linear modeling *t*-statistics for significant and marginal covariates and predictors of P_s, coded participation, and standard deviation (*SD*) of words. Dotted lines are negative relationships

multilevel nature of the data than was coded participation (.53, as compared with .40).

Factors that influence participation The secondary goal of this research was to explore factors that might influence participation equality and dominance. Having three different dependent measures complicates the interpretation of results. Yet there were clear, common patterns. The number of utterances per block was significantly associated with participation dominance. This covariate finding is intuitive: The longer the block, the less likely it was to be dominated by one person. Dominated participation was also more likely when negative affect words were spoken. Negative affect expression is a group outcome variable, thus suggesting the predictive validity of our metric but, in this case, not distinguishing between our new metric and the more traditional ones. Given that the dominating person was probably doing most of the talking, this implies a "ranting" effect, rather than displeasure expressed on the part of people who speak very little. We also saw a small but meaningful effect such that mixed-age/status groups were more likely to be dominated, as compared with homogeneous young faculty groups. It is interesting that a similar or opposite effect was not found for homogeneous older faculty groups, which also assumedly involved people of similar levels of status. In other words, homogeneous older faculty groups did not have significantly more equal participation than did mixed-age groups, but younger faculty groups did. This finding does not automatically mean that it was the older faculty who were dominating the mixed-status groups, simply that someone was dominating the conversation—be it younger faculty, older faculty, or a mix that depended on the particular block. Younger faculty in general, however, may be lower in power, as compared with older faculty, and so may be more inhibited and aware of turn taking (extrapolated from Keltner et al., 2003).

Further research using this metric as a starting place can unpack who is comfortable dominating under what conditions and what sort of information they bring to bear. Researchers could test this metric's predictive validity on a range of team outcome variables. For example, group cohesion would assumedly be related to participation equality. Also, as was noted earlier, when everyone's knowledge is important, teams diverse in background knowledge that have equal participation will have more successful outcomes. The relationship between team success and participation equality, or cohesion and participation equality, would assumedly be more accurately revealed in fluid groups if researchers utilized our metric.

Conclusion

Teams are increasingly composed of experts from multiple disciplines, both in the workplace and in academe. Funding agencies have recognized that solving complex problems often requires multidisciplinary teams, and universities are continuing to develop cross-disciplinary programs (Derry, Schunn, & Gernsbacher, 2005). Companies known for design innovation, like IDEO, thrive on teams composed of people with different types of expert roles, such as usability experts, designers, and others. Information sharing via participation is vital to these sorts of teams' success. This paper provides a methodological contribution, but by making this particular aspect of participation equality easier to measure in an objective fashion, it helps lay the groundwork for additional theoretical contributions regarding the role of participation equality. Further research on participation equality should test the use of this metric in other settings with dynamic membership of team conversations, such as agile software development teams, other science space missions, and design teams. This metric is not limited to task-relevant, work conversations: Those interested in participation equality in general could use this measure to examine conversations of people at leisure such as public places, in parks, or at private parties.

Furthermore, online settings have long been studied to assess their superiority to face-to-face groups with regard to participation equality (e.g., Burke & Chidambaram, 1995; Kiesler et al., 1984; Zmud et al., 2002). This metric could be used to capture participation equality/dominance more accurately in online settings such as course bulletin boards, chat rooms, and online gaming groups. Although, in face-to-face settings, it can be difficult to ascertain when participants are actively listening, this is even more problematic in online settings. Researchers interested in studying online discussion boards will need both technological and theoretical guidelines for determining when someone is participating, even if that person is late to the conversation. Joining a thread late and reading older posts does not count as participation during that earlier time, but it may be difficult to tell the difference between being present earlier and not participating until late versus joining the thread late, reading older comments, and then participating. Clearly, without additional data on timing of reading activities in asynchronous settings, this metric works best in synchronous online and face-to-face settings. In online gaming groups, it is clear when participants join, because they must log in and then converse and/or be in the "nearby" vicinity of their teammates. Future research on the antecedents and predictors of equal participation can also look carefully at the effects of status, age, gender, and prior information both on overall levels of participation equality and on who, specifically, is dominating a conversation.

Equal participation and information sharing are fundamental team processes. Even more important, they are necessary for multidisciplinary teams to take advantage of their background knowledge in order to be innovative. With this new metric, we can now measure participation in natural, fluid group conversations on or offline—the type that genuinely occurs in the changing world of work. The applications of this metric go beyond our study of scientists to any setting where researchers wish to measure participation in dynamic groups.

Acknowledgements Susannah Paletz and Christian Schunn are at the Learning Research and Development Center at the University of Pittsburgh. This research was supported in part by a National Science Foundation Grant #SBE-0830210 to the authors through the National Science Foundation (NSF)'s Science of Science and Innovation Policy Program. The opinions expressed are those of the authors and not necessarily those of NSF. The authors are grateful to Kerry Glassman, Justin Krill, Tsunhin Wong, and Carmela Rizzo for their assistance on this project and to Kevin Kim for answering questions about HLM analyses. An earlier version of parts of this paper was presented at the Atlanta Conference on Science and Innovation Policy, Atlanta, Georgia, 2009, and published as an IEEE conference proceedings report.

References

Bailey, D. B., Helsel-DeWert, M., Thiele, J. T., & Ware, W. B. (1983). Measuring individual participation on the interdisciplinary team. *American Journal of Mental Deficiency, 88*, 246–254.

Bales, R. F. (1950). *Interaction process analysis: A method for the study of small groups*. Cambridge, MA: Addison-Wesley.

Baumeister, R. F., Vohs, K. D., & Funder, D. C. (2007). Psychology as the science of self-reports and finger movements: Whatever happened to actual behavior? *Perspectives on Psychological Science, 2*, 396–403.

Berdahl, J. L., & Craig, K. M. (1996). Equality of participation and influence in groups: The effects of communication medium and sex composition. *Computer Supported Cooperative Work, 4*, 179–201.

Berdahl, J., & Martorana, P. (2006). Effects of power on emotion and expression during a controversial group discussion. *European Journal of Social Psychology, 36*, 497–509.

Burgoon, J. K., & Hale, J. L. (1987). Validation and measurement of the fundamental themes of relational communication. *Communication Monographs, 54*, 19–41.

Burke, K., & Chidambaram, L. (1995). Developmental differences between distributed and face-to-face groups in electronically supported meeting environments: An exploratory investigation. *Group Decision and Negotiation, 4*, 213–233.

Chi, M. (1997). Quantifying qualitative analyses of verbal data: A practical guide. *Journal of the Learning Sciences, 6*, 271–315.

Courtright, J. A., Millar, F. E., & Rogers-Millar, L. E. (1979). Domineeringness and dominance: Replication and expansion. *Communication Monographs, 46*, 179–192.

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

Derry, S. J., Schunn, C. D., & Gernsbacher, M. A. (Eds.). (2005). *Interdisciplinary collaboration: An emerging cognitive science*. Mahwah, NJ: Erlbaum.

DiMicco, J. M., & Bender, W. (2007, April). Group reactions to visual feedback tools. In *Proceedings of the second international conference on persuasive technology.* Stanford, CA.

Fischer, U., McDonnell, L., & Orasanu, J. (2007). Linguistic correlates of team performance: Toward a tool for monitoring team functions for space missions. *Aviation Space and Environmental Medicine, 78*, B86–B95.

Folger, J. P. (1980). The effects of vocal participation and questioning behavior on perceptions of dominance. *Social Behavior and Personality, 8*, 203–207.

Frey, L. R., Gouran, D. S., & Poole, M. S. (Eds.). (1999). *The handbook of group communication theory and research*. Thousand Oaks, CA: Sage.

Guzzo, R. A., & Dickson, M. W. (1996). Teams in organizations: Recent research on performance and effectiveness. *Annual Review of Psychology, 47*, 307–338.

Hiltz, S. R., Turoff, M., & Johnson, K. (1989). Experiments in group decision making: 3. Disinhibition, deindividuation, and group process in pen name and real name computer conferences. *Decision Support Systems, 5*, 217–232.

Hinds, P., & Kiesler, S. (Eds.). (2002). *Distributed work*. Cambridge, MA: MIT Press.

Jarvenpaa, S. L., Rao, V. S., & Huber, G. P. (1988). Computer support for meetings of groups working on unstructured problems: A field experiment. *MIS Quarterly, 12*, 645–666.

John, O., & Srivastava, S. (1999). The Big Five Trait taxonomy: History, measurement, and theoretical perspectives. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research* (2nd ed., pp. 102–139). New York: Guilford.

Joshi, A., & Roh, H. (2009). The role of context in work team diversity research: A meta-analytic review. *Academy of Management Journal, 52*, 599–627.

Keltner, D., Gruenfeld, D. H., & Anderson, C. (2003). Power, approach, and inhibition. *Psychological Review, 110*, 265–284.

Kiesler, S., Siegel, J., & McGuire, T. W. (1984). Social psychological aspects of computer-mediated communication. *The American Psychologist, 39*, 1123–1134.

Leaper, C., & Ayres, M. M. (2007). A meta-analytic review of gender variations in adults' language use: Talkativeness, affiliative speech, and assertive speech. *Personality and Social Psychological Review, 11*, 328–363.

Mannix, E., & Neale, M. A. (2005). What differences make a difference? The promise and reality of diverse teams in organizations. *Psychological Science in the Public Interest, 6*, 31–55.

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

Mesmer-Magnus, J. R., & DeChurch, L. A. (2009). Information sharing and team performance: A meta-analysis. *The Journal of Applied Psychology, 94*, 535–546.

Millar, F. E., & Rogers, L. E. (1976). A relational approach to interpersonal communication. In G. R. Miller (Ed.), *Explorations in interpersonal communication* (pp. 87–103). Beverly Hills, CA: Sage.

Moreland, R. L., & Levine, J. M. (2002). Socialization and trust in work groups. *Group Processes & Intergroup Relations, 5*, 185–201.

Nemeth, C. J., & Ormiston, M. (2007). Creative idea generation: Harmony versus stimulation. *European Journal of Social Psychology, 37*, 524–535.

Nijstad, B. A., & Stroebe, W. (2006). How the group affects the mind: A cognitive model of idea generation in groups. *Personality and Social Psychology Review, 10*, 186–213.

Olson, G. M., & Olson, J. S. (2000). Distance matters. *Human–Computer. Interaction, 15*, 139–178.

Paletz, S. B. F., & Schunn, C. D. (2009). A new metric for assessing group level participation in fluid teams. In S. E. Cozzens & P. Catalan (Eds.), *Proceedings of the Atlanta conference on science and innovation policy.* Los Alamitos, CA: IEEE. Retrieved from http://ieeexplore.ieee.org/xpl/RecentCon.jsp?punumber=5353037.

Paletz, S. B. F., & Schunn, C. (2010). A social-cognitive framework of multidisciplinary team innovation. *Topics in Cognitive Science, 2*, 73–95.

Paletz, S. B. F., Schunn, C. D., & Kim, K. H. (2011). *Unpacking problem-solving conversations: Some analogies spark conflict, other conflicts spark analogies.* Manuscript under review.

Palmer, M. T. (1989). Controlling conversations: Turns, topics, and interpersonal control. *Communication Monographs, 56*, 1–18.

Pennebaker, J. W., Booth, R. J., & Francis, M. E. (2007). LIWC2007 (2007 version) [Computer software]. Austin, TX: University of Texas. Purchased February 25, 2009. Retrieved from http://www.liwc.net/index.php

Pennebaker, J. W., Mehl, M. R., & Niederhoffer, K. G. (2003). Psychological aspects of natural language use: Our words, our selves. *Annual Review of Psychology, 54*, 547–577.

Pennebaker, J. W., Chung, C. K., Ireland, M., Gonzales, A., & Booth, R. J. (2010). *The LIWC2007 application*. Retrieved from http://www.liwc.net/liwcdescription.php

Poole, M. S., Holmes, M., & DeSanctis, G. (1991). Conflict management in a computer-supported meeting environment. *Management Science, 37*, 926–953.

Poole, M. S., Keyton, J., & Frey, L. R. (1999). Group communication methodology: Issues and considerations. In L. R. Frey, D. S. Gouran, & M. S. Poole (Eds.), *The handbook of group communication theory and research* (pp. 92–112). Thousand Oaks, CA: Sage.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Newbury Park, CA: Sage.

Rogers, L. E., & Farace, R. V. (1975). Analysis of relational communication in dyads: New measurement procedures. *Human Communication Research, 1*, 222–239.

Schulz-Hardt, S., Brodbeck, F. C., Mojzisch, A., Kerschreiter, R., & Frey, D. (2006). Group decision making in hidden profile situations: Dissent as a facilitator for decision quality. *Journal of Personality and Social Psychology, 91*, 1080–1093.

Schunn, C. D., Crowley, K., & Okada, T. (2002). What makes collaborations across a distance succeed? The case of the cognitive science community. In P. Hinds & S. Kiesler (Eds.), *Distributed work* (pp. 407–430). Cambridge, MA: MIT Press.

Siegel, J., Dubrovsky, V., Kiesler, S., & McGuire, T. (1986). Group processes in computer-mediated communication. *Organizational Behavior and Human Decision Processes, 37*, 157–187.

Squyres, S. (2005). *Roving mars: Spirit, opportunity, and the exploration of the red planet*. New York: Hyperion.

Stasser, G., & Titus, W. (1985). Pooling unshared information in group decision making: Biased information sampling during discussion. *Journal of Personality and Social Psychology, 48*, 1467–1478.

Stasser, G., & Titus, W. (1987). Effects of information load and percentage of shared information on the dissemination of unshared information during group discussion. *Journal of Personality and Social Psychology, 53*, 81–93.

Stasser, G., & Vaughan, S. I. (1996). Models of participation during face-to-face unstructured discussion. In E. H. Witte & J. H. Davis (Eds.), *Understanding group behavior: Consensual action by small groups, vol 1* (pp. 165–192). Mahwah, NJ: Erlbaum.

Straus, S. (1996). Getting a clue: The effects of communication media and information distribution on participation and performance in computer-mediated and face-to-face groups. *Small Group Research, 27*, 115–142.

Tollinger, I., Schunn, C. D., & Vera, A. H. (2006). What changes when a large team becomes more expert? Analyses of speedup in the Mars Exploration Rovers science planning process. In *Proceedings of the 28th Annual Conference of the Cognitive Science Society* (pp. 840–845). Mahwah, NJ: Erlbaum.

Tyran, C. K., Dennis, A. R., Vogel, D. R., & Nunamaker, T. F. (1992). The application of electronic meeting technology to support strategic management. *MIS Quarterly, 16*, 313–334.

van Knippenberg, D., & Schippers, M. C. (2007). Work group diversity. *Annual Review of Psychology, 58*, 515–541.

van Knippenberg, D., De Dreu, C. K. W., & Homan, A. C. (2004). Work group diversity and group performance: An integrated model and research agenda. *The Journal of Applied Psychology, 89*, 1008–1022.

Zmud, R., Mejias, R. Reinig, B., & Martinez-Martinez, I. (2002). *Participation equality: Measurement within collaborative electronic environments: A three country study* (Purdue CIBER working paper 2001/2002-005). Retrieved from http://www.krannert.purdue.edu/centers/ciber/publications/working-papers.html on June 15, 2010.