C E ⊕ Center for Effective Organizations

UNDERSTANDING GROUP-EFFICACY: AN
EMPIRICAL TEST OF MULTIPLE
ASSESSMENT METHODS

CRISTINA B. GIBSON
Center for Effective Organizations
Marshall School of Business
University of Southern California

AMY E. RANDELL
Graduate School of Management
University of California, Irvine

P. CHRISTOPHER EARLEY
Kelly School of Business
Indiana University

October 1998

Understanding Group-Efficacy: An Empirical Test of Multiple Assessment Methods

Cristina B. Gibson
Center for Effective Organizations
Marshall School of Business
University of Southern California

Amy E. Randell
Graduate School of Management
University of California, Irvine

P. Christopher Earley
Kelly School of Business
Indiana University

October 1998

Understanding Group-efficacy:

An Empirical Test of Multiple Assessment Methods

Cristina B. Gibson
Center for Effective Organizations
Marshall School of Business
University of Southern California

Amy E. Randel
Graduate School of Management
University of California, Irvine

P. Christopher Earley
Kelley School of Business
Indiana University

Abstract

Methods of assessing group-efficacy were examined using a multi-party role play negotiation. Group-efficacy is defined as a group's perceived capability to perform. Three methods of measuring group-efficacy were compared: (1) group potency (Guzzo, Yost, Campbell, & Shea, 1993), (2) an aggregation of group members' estimates, and (3) group discussion. These methods were used to assess group efficacy regarding several group performance characteristics. Analyses were conducted to assess various validity and measurement issues. All methods and types demonstrated construct validity; however, predictive validity was modest and varied across the measures. The group discussion method predicted group outcomes better than the potency scale or aggregate method. As expected, generalized efficacy measures predicted general performance indicators and task-specific measures predicted task-specific performance.

Understanding Group-efficacy: An Empirical Test of Multiple Assessment Methods

Although increasing emphasis is being placed on team-based activities, an incomplete understanding of such activities exists. Only recently have researchers begun to understand many of the processes underlying task-performing teams (Ancona & Caldwell, 1992; George, 1990; Gersick, 1988, 1989; Guzzo, Yost, Campbell & Shea, 1993; Hackman, 1990; McGrath, 1984). Given the focus on team performance and competitiveness, it is quite important for us to understand group performance as a competitive tool.

Recent theorizing has suggested that a group's collective sense of capability may be an important determinant of its effectiveness (Bandura, 1986; Guzzo et al. 1993; Shamir, 1990). At the individual level of analysis, motivation research has established that self-efficacy is strongly related to performance. Self-efficacy refers to a person's belief in his or her capability to perform a task (Bandura, 1997; Wood & Bandura, 1989). The relationship between self-efficacy and individual performance is quite robust; indeed, self-efficacy has been related to such diverse outcomes as problem solving (Cervone & Peake, 1986), learning and achievement (Campbell & Hackett, 1986), adaptability to new technology (Hill, Smith, & Mann, 1987), and training effectiveness (Earley, 1994; Gist, 1987).

Although his focus was on self-efficacy, Bandura (1982; 1986; 1997) was among the first to suggest that performance beliefs exist at other levels of analysis. He suggested that the strength of groups and organizations lies in people's sense of group-efficacy that they can solve their problems and improve their lives through concerted effort (Bandura, 1986). Other researchers (Campion, Medsker, & Higgs, 1993; Gibson, 1995; Guzzo et al. 1993; Parker, 1994; Porter, 1992; Sayles, 1958; Zander & Medow, 1963) have since investigated the notion that group beliefs about capabilities influence performance. In the Campion et al. (1993) analysis, for example, three group effectiveness measures for 80 work groups were correlated with nineteen different work group characteristics drawn from the existing literature on group effectiveness. Of the nineteen characteristics, the strongest predictor of effectiveness was the measure of group performance beliefs. Furthermore, the extent to which the

groups possessed confidence in their ability was the only characteristic that demonstrated a statistically significant relationship with all three effectiveness criteria.

Although group performance beliefs appear to be a strong predictor of performance, researchers who have investigated such beliefs do not agree as to how the construct should be operationalized and measured. A number of significant issues exist including the extent to which group performance beliefs are: (1) an aggregated versus collective phenomenon and, (2) best operationalized as general versus task-specific beliefs. This study addresses these issues by first reviewing the various ways in which the construct has been operationalized and then empirically testing the competing approaches as to their predictive power for determining group performance.

Aggregated Versus Collective Phenomenon

Defining an attribute of a group, such as a group belief, raises certain requirements that are not applicable to the more common exercise of defining attributes of individuals. Four such requirements are: (1) the construct must reflect the group as a whole, rather than the individual members as separate units, (2) group members should agree with regard to the construct, (3) the construct must discriminate among groups, and (4) the origin of the construct must reflect the processes of interaction that occur within the group (Bar-Tal, 1990).

Several approaches to group performance beliefs that have appeared in the literature address one or more of these requirements, but none fulfills them all. One approach involves assessing individual perceptions as a representation of the group belief (e.g., Crockner & Luhtanen, 1990; Earley, 1993; Guzzo, Yost, Campbell & Shea, 1993; Shamir, 1990). For example, Guzzo and his colleagues (Guzzo, 1986; Guzzo et al. 1993; Shea & Guzzo, 1987ab) have developed the construct of "group potency" which captures efficacy beliefs at the group level. Guzzo and his colleagues define group potency as a belief in a group about its general effectiveness across multiple tasks. Potency is viewed as a social-psychological phenomenon that is linked to performance in a reciprocal relationship. Although they define group potency as a "group belief," Guzzo et al. (1993) utilize an approach that aggregates individuals'

assessments of group capability. They first ask individual members within a group to rate the group's belief using an eight-item scale in which the group is the referent (e.g., "This group has confidence in itself"). Next, they calculate the arithmetic mean of these responses to represent the group belief. In order to assess the degree of agreement within the group, they then calculate intra-member agreement. And finally, they assess the degree to which groups differ by calculating between-group variance.

Research using aggregation fulfills the first and second requirements suggested by Bar-Tal (1990) (the construct reflects the group as a whole and members must agree on the construct), but does not necessarily fulfill the third or fourth requirements (the construct must differentiate well between groups and must reflect processes that occur within the group). Indeed, based upon their findings, Guzzo et al. (1993) suggest that the most important task that lies ahead for research concerning group beliefs is to establish the validity of alternative approaches to measuring such beliefs, and further, to relate the approaches to performance criteria.

Both Guzzo et al. (1993) and Gist (1987) suggest that a different avenue is to assess group-efficacy using a group discussion procedure in which members provide a single response as a group. Using this technique, a group is presented with a rating scale to use in forming a single response to a question about its sense of efficacy with regard to a given task. The single response is obtained through free format discussion within the group. This approach eliminates the calculation of statistical indicators of inter-member agreement and was the method of choice for Zander and Meadow (1963) in their classic analysis of group aspirations.

Perhaps even more pertinent are the theoretical issues that are addressed through the group discussion approach to group performance beliefs. A group discussion approach addresses all of the requirements for group constructs set forth by Bar-Tal (1990). First, by using the group discussion approach, the construct represents the group as a whole. Second, recent research indicates that the group discussion approach results in significant agreement within the group and strong between-group

differences (Gibson, 1995). Finally, use of the group discussion approach results in a construct that appears to truly reflect the processes of interaction within the group.

That this last requirement is fulfilled by using a group discussion approach is suggested if one imagines the process by which group beliefs are formed. During the formation process, group discussion allows members to explore what each contributes to the group in terms of experience, skills and knowledge. In arriving at an efficacy estimate through group discussion, members are likely to share feedback they have received about their performance on similar tasks in the past. Performance evaluation and feedback have been shown to be important sources of the information used to arrive at efficacy estimates (Bandura & Cervone, 1986; Cervone, Jiwani & Wood, 1991). The group discussion approach to performance beliefs appears useful because it may better capture the group's collective history. Although individual members of a group come and go, there remains within the group a socially constructed, shared history as evidenced by group norms, routines, and patterns of interaction. Aggregating each person's own assessment may not accurately reflect the collective history that characterizes the group in comparison with other groups. In contrast, a group discussion approach better reflects the patterned interactions among group members that constitute the collective history of the group. At the same time, the group discussion approach may have limitations. Research investigating biases in group decision-making (e.g., risky shift), suggests that group interaction during the process of arriving at an efficacy estimate may accentuate a group's efficacy to a point that is unrealistic (Bandura, 1997). We need more investigation of the group discussion approach to measuring group-efficacy in order to better understand whether estimates arrived at through this approach demonstrate construct validity.

General Versus Task-specific Beliefs

Another unresolved issue is whether group performance beliefs are best represented as general beliefs concerning group effectiveness or as task-specific beliefs. For example, in a group of salespeople, a general performance belief would include the group's estimate of capability to retain

clients, to locate new clientele, to increase profit margins, and to increase overall sales. However, a task-specific group performance belief refers to only one specific group task. Such a view would imply, for example, that a group of salespeople have a performance belief regarding retaining clients, a separate belief with regard to their capability to locate new clientele, and another that pertains to increasing profit margin.

Zander and Medow's (1963) operationalization of group performance beliefs represents an extreme in terms of task specificity. Using their approach, participants estimate the group's ability to perform a *specific trial* of a specific task. Although this makes sense in a laboratory experiment, such an approach may not be readily generalizable to field settings. In most organizations for example, tasks are broadly defined (e.g., to "redesign personnel policies" or "develop a control mechanism for distribution") and are not easily separable into specific trials (e.g. to delineate specific "trials" on a task such as "design personnel policies" would be difficult, if not impossible). Instead, a range of efforts and activities are often required over an extended period of time in order to accomplish a task objective.

At the same time, defining the task too broadly can reduce the strength of the relationship between the belief and actual performance. Perhaps, when a task is broadly defined, the capabilities required for successful completion are less clear than when a task is narrowly defined. This same point has been made in the literature on self-efficacy, expectancy theory and self-esteem. The majority of the research on self- efficacy has defined efficacy beliefs as particular to specific tasks. Saks (1992), for example, argued that efficacy measures should be based on job analysis information to assure the alignment of perceptions with actual job requirements.

After reviewing the literature and working with the group-efficacy construct in a number of studies, we suggest that rather than referring to a specific trial of a specific task, group-efficacy is best conceptualized as a collective belief in the capability of the group to meet a task objective. Furthermore, we propose that this belief is best measured by asking group members to discuss among themselves their capability to perform the task objective. We propose that when arrived at through a group discussion

procedure, the group-efficacy belief will be a more accurate predictor of group capability than will be aggregated individual perceptions.

Domain of the Investigation

In this study, we empirically test three assessment methods and three types of group-efficacy as suggested by the literature. In terms of assessment, the first method uses Guzzo et al.'s potency scale. Individual members within each group rate their perception of their group's belief using an eight-item scale in which the group is the referent (e.g., "This group has confidence in itself"), and an arithmetic mean of these responses is calculated for each group. Doing so assumes that the "average" response from the group accurately captures a group-level construct (Guzzo et al. 1993). By averaging individual responses, individual variation within the group is presumed negligible (Klein, Dansereau, & Hall, 1994).

The second method of measuring group-efficacy was used by Earley (1993) and is based on similar assumptions. The measure uses a scale format developed at the individual level (Locke et al. 1994), but with a group referent. Rather than completing items referring to general competencies, individual members rate their perception of their group's capability for five specific levels of performance. Ratings on these five levels are averaged for each individual, and then individuals' scores are averaged across the group. Again, averaging assumes that the "average" response from the group accurately captures the group's perceived capability. We refer to this method as the aggregate method.

Finally, for the third method, a group is provided with a single response form and instructed to arrive at a group response based on group discussion and interaction (Gibson, 1995; Lindsley, Brass & Thomas, 1995; Zander & Medow, 1963). The group's perceptions regarding the group's capability for five specific levels of performance are recorded. No instructions are given regarding particular combinatorial rules (e.g., consensus, majority rules); groups are free to enact their own process for arriving at the group response. This approach assumes that the perception of capability is a holistic group characteristic. We refer to this method as the group discussion method.

In order to address the general versus task-specific debate, we compared three types of group-efficacy for a group-based negotiation. The first type of efficacy is task-specific, referring to a group's belief regarding the time needed to complete the negotiation. The second type of efficacy is also task-specific, referring to a group's belief regarding the degree of agreement that will exist within the group. And finally, the third type of efficacy is general, referring to a group's generalized belief in its ability to produce an effective solution. Our analyses compared several components of construct validity for the three methods and three types of group-efficacy. We investigated potential order effects by systematically varying the order in which the different measures were administered, discriminant and convergent validity, stability of the measures over time, factor structures, reliabilities and scale compositions, and the predictive validity of the various measures for group outcomes.

Method

Participants

The participants in our study were 166 male and female managers participating in 5-6 person study groups as part of their executive MBA program. Prior to participating in the study, the groups had been working together for five weeks in a number of intensive ways including a ten-day residential program, case analysis and presentation, and group homework exercises. Further, these groups were considered permanent for the duration of the multi-year executive MBA program. Throughout their program, groups were responsible for completing multiple group projects, preparing for class presentations, and studying for exams across all required courses. As a part of their program orientation, members of each group were encouraged to serve as a support network for one another and to socialize together outside of the program. Exit interviews with managers indicated that there was a high degree of interaction, commitment, and interdependence within the groups. In summary, the groups in our study had a history and members anticipated a subsequent future. In this regard, the groups had strong personal significance for group members.

Design and Task

The task used in this study was a multi-party role-play negotiation called the "Porsche Exercise" developed by Leonard Greenhalgh (1984) at Dartmouth College. We chose this role play because it provides a rich basis for group interaction and personally relevant roles for participants. Furthermore, effective performance on this exercise requires a combination of process efficiency and innovative decision making procedures in order to arrive at a negotiated solution. Research suggests that such a task is well-suited for studies of efficacy, because it requires a variety of complex cognitive skills and the optimal solution is multi-faceted (Bandura, 1997).

Performance on this task was assessed by: (1) the time it took the group to reach a final solution, and (2) the degree of intra-group agreement with the final solution, and (3) the perceived general effectiveness of the final solution by group members. Time and member agreement are critical objective outcomes that are becoming increasingly important in task-performing teams. The content of much of the work being carried out by teams is shifting from tasks that are physical in nature to tasks that are cognitive and data driven. Mohrman, Cohen and Mohrman (1995) refer to this phenomenon as a shift toward "knowledge work." As this trend gains momentum, we have an important need to understand indicators of team effectiveness such as time spent reaching solutions and the degree of agreement or support for solutions. Effectiveness in this regard will no doubt represent real gains in competitive advantage for teams in the future. In contrast to time and agreement, our third measure assessing general effectiveness is a broad-based indicator of team member perceptions of group processes and group outcomes. Efficacy was assessed for each of these three components of performance.

The Porsche role play task was a normally scheduled classroom activity. It was introduced as a realistic simulation of actual situations managers may confront at work; as such managers were told to take their roles seriously and behave as they would if their performance was being assessed by a supervisor or CEO. Although the managers were aware that the activity would not to be graded, they became highly engaged and demonstrated strong identification with their individual roles.

During the exercise, each group simulated a company task force consisting of six vice presidents from the following departments within Porsche of America: Finance, Marketing, Sales, Research and Development, Legal Affairs, or Production (Note 1). The simulation materials are extensive with each member receiving seven pages of general background information along with an additional 4-5 pages of role specific information. Within a maximum of 90 minutes, the task force is responsible for recommending a product strategy for the following year. Group recommendations involve the production quotas, body styles, and performance options for several lines of Porsche cars including: Porsche 911 (Carrera, Cabriolet, Coupe, and Targa), Porsche 928S, and Porsche 944. The materials are structured in such a way that there are natural conflicts of desired quotas as well as potential multiple coalitions that may form. Further, several enacted roles can participate in more than one coalition leading to a dynamic shift of power as the role play unfolds.

The study consisted of a 2 (method) x 3 (type) x 2 (trials) design. The first factor (method) referred to the method of assessing group-efficacy (aggregate or group discussion). The second factor (type) referred to three types of group-efficacy (efficacy regarding time to complete the task, efficacy regarding intra-group agreement, and efficacy regarding general effectiveness). The third factor (trials) referred to the timing of the efficacy measure (either pre-performance or mid-performance). These factors were fully crossed (an aggregate scale and a group discussion scale were completed for each of the three types of efficacy at each time trial). We also measured potency (the third method of assessing group-efficacy) at each time trial. As discussed above, the potency scale was developed by Guzzo et. al. (1993) to measure only beliefs regarding general effectiveness; therefore, it was impossible to use this method to assess other types of efficacy (efficacy regarding time or efficacy regarding intra-group agreement). This prevented us from using a factorial design in which all three methods are crossed with all three types at each trial.

Managers in one half of the groups (79 subjects) completed the scales in the following order: (1) potency scale, (2) aggregate method, and (3) group discussion method. The other half of the groups (87

subjects) completed the scales in the reverse order so that we could assess possible order effects. Groups were asked to use each scale at the beginning of the simulation (pre-performance trial) and again after 30 minutes of interaction (mid-performance trial). The timing of these trials was selected based on previous research indicating that efficacy beliefs formed prior to task performance can be different than those formed afterwards. However, assessing efficacy at the end of the exercise, when outcomes such as time to reach agreement and degree of agreement are known, would not be appropriate.

Measures

Independent variables. For the potency scale, we used Guzzo et. al.'s (1993) eight-item scale. This scales contains items such as, "My group has confidence in itself," "My group believes it can become unusually good at producing high-quality work," and "My group expects to have power around here." (See the Appendix for a complete list of items). Group members individually completed the eight items using a ten point scale (1=To no extent, 3=To a limited extent, 5=To some extent, 7=To a considerable extent, and 10=To a great extent). Each individual's score was computed by averaging across the eight items. Then the individual scores were averaged across group members to arrive at a group potency score for each group.

For both the aggregate and group discussion methods, we used variations of Bandura's traditional scale format for assessing efficacy (Bandura, 1986; Locke, Frederick, Lee, & Bobko, 1984) in which respondents are asked to state their level of certainty (0% to 100%) that several levels of performance can be achieved. Three types of group-efficacy were measured using this technique. The first type of group-efficacy (efficacy regarding time) referred to the amount of time it would take the group to complete their role play (e.g., 30-44 minutes, 45-59 minutes, 60-74 minutes, 75-89 minutes, and 90-104 minutes). The second type of group-efficacy (efficacy regarding intra-group agreement) referred to expectations regarding the number of members who would agree with the final solution by the end of the role play (e.g., 6 out of 6, 5 out of 6, 4 out of 6, 3 out of 6, or 2 out of 6 members would agree with the final solution). And finally, the third type of group-efficacy (efficacy regarding general

effectiveness) assessed the groups' expectations concerning the effectiveness of their final solution as rated on a ten-point scale by a panel of experts who would review their solutions (e.g., their certainty of getting at least a 9-10, 7-8, 5-6, 3-4, or 1-2 rating for the effectiveness of their final solution).

When utilizing the aggregate method, group members individually evaluated how certain they were that their group could achieve the five levels of performance for each of the three performance components (time, intra-group agreement, and general effectiveness) using the three scales in the Appendix. The arithmetic mean across levels of performance and across individual group members' responses was then computed and used to represent the group's level of efficacy for the three components. When utilizing the group discussion method, the same three types of efficacy were assessed; however, the group was provided with a single copy of the three efficacy scales and was instructed to come up with a "group response" that represented their group's expectations for the five levels of performance on each scale. Certainty scores across the five performance levels were averaged and this score was used to represent the group's level of efficacy for each of the three types of efficacy (time, agreement, and general efficacy). A comparison of the reliabilities, factor structures and inter-correlations of the various efficacy scales are presented in the results section.

Dependent variables. Both task-specific and broad-based indicators of group performance were assessed after completion of the simulation. The first task-specific performance indicator was the time it took each group to complete the negotiation. This indicator had a maximum value of 90 minutes (imposed by the experimenter). The second task-specific performance indicator was the degree of intra-group agreement expressed by the group members. For this measure, each group was asked to write down its final solution. Below the solution, each member was asked to provide their code number and initial their approval or disapproval of the solution. Agreement was scored as the percentage of group members indicating their agreement with the final written solution. A high degree of variance was obtained on this indicator across groups, ranging from two to six members agreeing with the final solution.

The two remaining indicators were relatively broadly defined. Process effectiveness was assessed by asking group members to rate their groups' ability to: use time productively, value efficiency, prioritize, and exert effort toward the task. These personal assessments were made using a five-point Likert scale. The reliability of this scale (Cronbach's Alpha) was .73. A principal component analysis demonstrated that all items loaded on a single factor having an eigenvalue of 2.24 and accounting for 55.9% of the variance. Factor loadings ranged from .45 to .70. Based on these analyses, a "process score" was created for each group by first taking the mean of each individual's responses across the four items and then aggregating this score across the group members.

A fourth performance indicator measured the perceived effectiveness of the groups' solution. This indicator was assessed by asking group members to rate the extent to which their group's solution was creative, practical, interesting, satisfactory, and effective using a five-point Likert scale. The reliability of this scale (Cronbach's Alpha) was .85. A principal component analysis demonstrated that all items loaded on a single factor having an eigenvalue of 3.10 and accounting for 61.9% of the variance. Factor loadings ranged from .49 to .71. Based on these analyses, a "solution score" was created for each group by first taking the mean of each individual's response across the five items and then aggregating this score across the group members.

Procedure

Participants reported to their classrooms in groups of 20-50 people as in a normal class meeting. After briefly describing the nature of the role play, participants were instructed to gather into their study groups and sit in a circle formation. Given the number of groups in each session (4-10), the role play was conducted in two separate classrooms with approximately 2-5 groups in each room after the general instructions had been provided to all of the participants.

Next, the participants were provided with a general information packet detailing the basics of the role play, information about Porsche of America, and their general purpose in the role play (to act as an advisory panel task force for the president of the company). While they were reading the introductory

material, group members were randomly assigned to one of the six roles in the simulation. Subjects were informed that the entire length of the exercise was approximately two hours with 90 minutes for group discussion and interaction, unless a group solution was reached in less time. Further, they were told that they would complete a short questionnaire followed by a 30-minute discussion period, complete a second questionnaire, and be provided with an additional 60 minutes of discussion time. It was emphasized that they would have this 90 minutes if needed, but that there was no minimum time for completing the role play. Students were informed of the realism of time constraints, given that time-related efficiency is fast becoming a means of gaining a competitive edge in teams.

After reading the general information and role-specific materials, participants were provided with the group-efficacy questionnaires. One half of the groups completed the potency scale, the aggregate method and then the group discussion method (sequence A), while the other half of the groups completed the scales in the reverse sequence (sequence B). In addition, the three different types of group-efficacy (agreement efficacy, time efficacy, and general efficacy) were presented in two different orders crossed with two of the methods in a block randomized fashion across groups within experimental sessions (order 1: general effectiveness efficacy, time efficacy, agreement efficacy; order 2: agreement efficacy, time efficacy, general effectiveness efficacy). After indicating their level of time efficacy, agreement efficacy, and general efficacy using one method (aggregate or group discussion), each group was asked to indicate their level of time efficacy, agreement efficacy, and general efficacy using the second method (the method not used first). As mentioned previously, the potency measure was also completed during each trial (in sequence A groups, the potency scale was completely immediately first; in sequence B groups, the potency scale was completed last); however, the potency scale could not be crossed with the three types of efficacy because the potency items measure only perceptions regarding general effectiveness. In summary, the counter-balancing procedure resulted in four sequence-order combinations. The objectives of this procedure were threefold: 1) to assess potential method sequence

effects, 2) to assess potential type order effects, and 3) to test potential biases associated with the within-groups design.

Next, the experimenter instructed the managers to begin their first 30-minute group discussion and negotiation period, and to notify her if they reached an agreement. After 30 minutes, the experimenter instructed the managers to stop working and they were provided with efficacy questionnaires identical to those they had completed prior to the first discussion period. Upon completing the questionnaires, they were provided with a 15-minute rest break during which they were instructed not to discuss their ongoing negotiations (the experimenter monitored this break in order to ensure the instruction was followed). After the break, the groups were reformed and asked to continue their discussions for up to a maximum of 60 additional minutes. They were told to inform the experimenter as soon as they reached a solution that they were willing to formally write down. It was emphasized, however, that the solution need not reflect a total consensus -- the group had complete discretion concerning what it considered to be the appropriate level of agreement that should be reached before the solution was written down.

As each group signaled to the experimenter that they were done, the time it took them to reach the solution was recorded and they were provided with a form on which to write the content of their group solution. This form included a line for each member to initial signifying that they had read the final solution and indicating whether or not they each agreed with the solution. If a group did not complete the negotiation within 90 minutes, they were stopped by the experimenter and instructed to write down the solution that currently existed and indicate their personal agreement/disagreement with it. After completing the written solution from, each manager was given a final questionnaire assessing the effectiveness of their group's process and solution. Finally, the managers congregated in a single classroom and, as a class, discussed the nature of the exercise, group discussions, and group beliefs. Prior to the exercise, the managers had not received any materials concerning efficacy or group

dynamics, and the role play was used as the basis for introducing these topics for the remainder of the class session.

Results

Overview

Before discussing the results of our analyses, we would like to emphasize that our within-groups design represents a very conservative comparison of the various measures. We took several precautions to ensure that our tests of comparative effects were *fair*, as defined by Cooper and Richardson (1986). These authors argued that in order to conduct fair comparisons, competing theories, factors or variables must be operationalized, manipulated, or measured with *equivalent strength* (1986: 179). By equivalent strength, Cooper and Richardson mean two things: (1) procedural equivalence; and (2) distributional equivalence.

Procedural equivalence is demonstrated when each variable is operationalized with equivalent fidelity and care. We ensured that all three methods and types were given equal opportunity to demonstrate validity by: (1) counter-balancing the sequence in which the measures were administered, (2) counterbalancing the order in which types were administered and; (3) by selecting a variety of different outcomes variables in order to test predictive validity. According to Cooper and Richardson (1986: 183), an omnibus check on procedural equivalence is to ask whether or no one or more of the variables have been given a procedural advantage. The aggregate method of measuring group-efficacy had a slight advantage regarding tests of predictive validity because several of the outcome variables were measured using an aggregation method. Recall, however, that we proposed that the group discussion method of measuring group-efficacy would be most robust so this potential bias makes a test of our predictions somewhat more conservative with regard to the group discussion approach.

The second component of fair comparisons, distributional equivalence, can be demonstrated by showing comparable variance for each variable being compared (Cooper & Richardson, 1986). Expanded variables (with greater variance) are advantaged; compressed variables (with lesser variance)

are disadvantaged. All of the group-efficacy scales (when revised, as will be discussed further below) demonstrated a similar level of variance. The only measure with a slightly lower variance was the group discussion measure of general group-efficacy, thus again providing a slight *disadvantage* for the measure we proposed would be most robust.

We were faced with a difficult decision regarding whether to analyze the data at the individual or group levels. Several of the measures involve measuring individual perceptions; only the group discussion measure captures group perceptions. Aggregation of individual perceptions is a controversial issue. Several recommendations have emerged in the literature (Goodman et al. 1987; James, 1982; Campion, Medsker & Higgs, 1993; Klein, Dansereau, & Hall, 1994). We chose to run our analyses at the group level. Our rationale for doing so is conceptual given that the level of our theory is the group (Klein, Dansereau, & Hall, 1994). Thus, the sample size for our analyses is 30 groups representing the sample of 166 group members (Note 2).

Investigating Potential Order Effects

The first component of our analyses was to conduct a series of analyses of variance (ANOVA) in order to determine whether or not differences attributable to method sequence and type order occurred. Specifically, we tested for statistically significant differences between the block randomized groups, and no such differences were obtained. Furthermore, there were greater differences between methods of assessing efficacy than within methods across the two sequences (sequence A = potency, aggregated, then group discussion; sequence B = group discussion, aggregated then potency). For example, regardless of sequence, there were greater differences between aggregated efficacy scores and group discussion scores than between aggregated efficacy scores obtained using sequence A and aggregated efficacy scores obtained using sequence B. For subsequent analyses, we combined the two method sequences and the two type orders for sake of simplicity.

Discriminant and Convergent Validity

The descriptive statistics for the method by type combinations across the two trials and the dependent variables are listed in Table 1 along with reliability estimates and inter-scale correlations. Using the framework developed by Campbell and Fiske (1959), the inter-scale correlations were examined for evidence of convergent and discriminant validity.

Insert Table 1 about here

First, the matrix provides evidence of both discriminant and convergent validity with regard to efficacy *types* (general efficacy versus time efficacy versus agreement efficacy). Discriminant validity of efficacy types is evidenced by the fact that the reliabilities for each scale are generally higher than the correlations among different measures within trials (e.g., the .80 reliability of the trail 1 aggregate generalized efficacy is higher than the .64 correlation of this measure with trial 1 aggregate time-based efficacy; the .81 reliability of the trial 1 aggregate time-based efficacy is higher than the .52 correlation of this measure with trial 1 aggregate % agreement efficacy; etc.). Convergent validity of efficacy types is evidenced by the fact that the highest correlations in the matrix are on the convergent validity diagonal for type (e.g., trial 1 aggregate generalized efficacy is correlated .74 with trial 1 discussion generalized efficacy; trial 1 aggregate time efficacy is correlated .84 with trial 1 discussion time efficacy; trial 1 aggregate % agreement efficacy is correlated .85 with trial 1 discussion % agreement efficacy, etc.).

The matrix also provides evidence, albeit weaker, for the discriminant and convergent validity of different *methods* of measuring efficacy (i.e., the potency scale versus aggregate method versus discussion method). Correlations comparing the same methods are generally higher than correlations comparing different methods (e.g., the .83 correlation between trial 1 discussion generalized efficacy and trial 2 discussion generalized efficacy is higher than the .63 correlation between trial 1 aggregate generalized efficacy with trail 2 discussion generalized efficacy).

Stability Across Time

The matrix also suggests that the measures were stable across the two time trials (i.e., pre-performance trial compared to mid-performance trial). Strong correlations were obtained between trial 1 and trial 2 measures when scales measuring the same type and using the same method are compared (e.g., trial 1 discussion generalized efficacy is correlated .83 with trial 2 discussion generalized efficacy; trial 1 discussion % agreement efficacy is correlated .70 with trial 2 discussion % agreement efficacy).

Based on these comparisons, two sets of preliminary analyses were run, first using the trial 1 measures (pre-performance) and then the using the trial 2 measures (mid-performance). The same general pattern of results was obtained (i.e., similar factor structures and inter-correlations); however, as expected, the trial 2 results were somewhat cleaner and stronger in magnitude. The researchers' observations of the groups suggest that this was because the subjects had more opportunities to calibrate as time passed. Based on these preliminary analyses, we chose to focus on the trial 2 measures in subsequent analyses.

Factor Structure, Reliability, and Scale Composition

Next, we conducted an exploratory factor analysis that included all of the items for the trial 2 measures. Specifically, this included the eight items on the potency scale, the five items on the aggregate measure of time efficacy (with each item representing an increasing level of difficulty), the five items on the aggregate measure of agreement efficacy, the five items on the aggregate measure of general effectiveness efficacy, the five items on the discussion measure of time efficacy, the five items on the discussion measure of agreement efficacy, and the five items on the discussion measure of general effectiveness efficacy. Because we were interested in the underlying structure of each scale and our focus here was on specific items, rather than aggregates, we ran these analyses on the individual level data (n=166 persons) before aggregating across scales or across groups.

The potency items loaded on one factor. The other factors that emerged represented levels of difficulty, as opposed to method or type. For example, the items on each scale that measured efficacy for the easiest level of performance loaded on a single factor (e.g., items measuring respondents' certainty
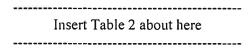
that they will complete the role play in 90–104 minutes out of a maximum of 120 minutes, their certainty that 2 out of 6 members will agree with the solution, and their certainty that their group will receive a 1-2 rating out of 10 on general effectiveness all loaded on the same factor). In their investigation of the self-efficacy scale, Locke, Frederick, Lee and Bobko (1984) found that scales composed of items representing the moderate levels of performance were the most robust measures. This led them to conclude, "self-efficacy strength estimates are most valid when they pertain to goals or levels of performance that are neither within the reach of all subjects nor within the reach of none" (1984: 248). Based on these results, and the precedent set in the self-efficacy literature, we re-constructed the aggregate scales and the discussion scales using the three items representing moderate levels of performance for each type of efficacy (i.e., items measuring respondents' certainty that they will complete the role play in 45-59 minutes, 60–64 minutes or 75–89 minutes; their certainty that 3 out of 6, 4 out of 6, or 5 out of 6 members will agree with the solution; and their certainty that their group will receive a rating of 3–4, 5-6 or 7-8 on general effectiveness).

Test of Scale Types

Reviewing results obtained in the preliminary analyses, we constructed seven scales to correspond with our theoretically derived scales: (1) a potency scale; (2) an aggregate general effectiveness efficacy scale; (3) aggregate time efficacy, (4) aggregate agreement efficacy; (5) discussion general effectiveness efficacy, (6) discussion time efficacy, and (7) discussion agreement efficacy. For all seven scales, reliabilities (Cronbach's Alpha) ranged between .72 and .94. For each scale, all items loaded on a single factor. In each analysis, the eigenvalue for the single factor was greater than 2.00 and accounted for 71.0% of the variance.

For subsequent analyses, we calculated group-level scores for all of the scales and our analyses reflect a sample size of 30 groups. The descriptive statistics for the revised scales and the dependent variables are listed in Table 2 along with reliability estimates and inter-scale correlations. Several findings should be noted regarding the correlations in this matrix. First, twenty out of the twenty-one

inter-measure correlations between the revised measures of group-efficacy were statistically significant ($p < .05$). The strongest inter-measure correlations were for those measures assessing the same type of group-efficacy (e.g., aggregate time group-efficacy was correlated .90 with discussion time group-efficacy; aggregate agreement group-efficacy was correlated .86 with discussion agreement group-efficacy). Inter-measure correlations comparing different types of group-efficacy were generally not as high. This demonstrates convergent validity of measures assessing the same type of group-efficacy and discriminant validity of measures assessing different types of group-efficacy. Strong inter-measure correlations were also obtained between measures utilizing the same methods (e.g., aggregate general group-efficacy was correlated .71 with aggregate time group-efficacy; discussion general group-efficacy was correlated .63 with discussion agreement group-efficacy), demonstrating convergent validity of measures utilizing the same method of measuring group-efficacy.

Insert Table 2 about here

Predictive Validity

Several of the correlations between the group-efficacy measures and the performance indicators are also noteworthy (see Table 2). First, all of the efficacy measures were negatively correlated with the dependent variable "minutes to complete agreement." This is reasonable because it suggests that when efficacy regarding time was high, groups took less time in completing the negotiation (i.e., they were more efficient). In fact, the single strongest performance-related correlation in the matrix was between discussion time group-efficacy and the time it took the groups to complete an agreement ($r = -.56$, $p < .01$). The correlation between aggregated time group-efficacy and time to complete an agreement was also strong ($r = -.47$, $p < .05$). A second observation regarding the matrix in Table 2 is that, as proposed, efficacy measures using the discussion method demonstrated relationships of greater magnitude with the performance indicators than did efficacy measures using the aggregate method although these differences were not significant.

A third observation regarding the matrix, is that specific types of group-efficacy were highly correlated with specific performance indicators, and general types were highly correlated with general indicators. For example, a strong positive correlation was obtained between discussion time group-efficacy and time it took to complete the negotiation ($r = -.56$, $p < .01$). This indicates that when group-efficacy regarding time was high, groups took less time to complete the negotiation. Potency, a highly generalized measure of group beliefs, was significantly correlated with the general indicator of process effectiveness ($r = .38$, $p < .05$).

Based on these results, additional analyses were conducted in order to compare the predictive validity of the various efficacy measures currently used for group research. We conducted four sets of regression analyses (see Table 3), each of which predicted a different performance indicator (time to complete an agreement, percentage of agreement with the final solution among members, effectiveness of group process, and effectiveness of group solution). Within each set, three equations were run, with each equation testing the predictive validity of a different type of group-efficacy (time efficacy, agreement efficacy, general efficacy). Again, it should be emphasized that the analyses were conducted at the group level; therefore, the sample size for each set of regressions was thirty groups and the analyses represent very conservative tests of the models.

Insert Table 3 about here

The first set of regressions predicted minutes to complete an agreement. In the first equation, minutes to complete an agreement was regressed on discussion time efficacy, aggregate time efficacy, and potency. In the second equation, minutes to complete an agreement was regressed on discussion agreement efficacy, aggregate agreement efficacy, and potency. In the third equation, minutes to complete an agreement was regressed on discussion general efficacy, aggregate general efficacy, and potency. The highest R^2 value was obtained in the first equation by regressing minutes on the time efficacy measures ($R^2 = .32$, $p < .05$). Within this equation, only the measure using the discussion method had a significant beta weight (beta = $-.69$, $t = -1.93$, $p < .06$).

The second set of regressions predicted percent agreement among members. Again, three equations were run, each involving a different type of efficacy. The highest R^2 value was obtained in the equation regressing percent agreement on the agreement efficacy measures (R^2 = .11, ns). In this equation, the measure using the discussion method had the largest beta weight (beta = .47, t=1.27, ns). The third set of regressions predicted effectiveness of the group process. The highest R^2 value was obtained in the equation regressing group process effectiveness on the agreement efficacy measures (R^2 = .23, p < .10). Generally, the measure using the general potency scale was the strongest predictor of general group process effectiveness. The fourth set of regressions predicted perceived effectiveness of the group solution. None of the R^2 values obtained across the three equations were significant.

Discussion

The focus of our study has been the empirical assessment of several methods of measuring group-efficacy. While there is a call for team-based work, it is unclear which factors underlie such group activities. With regard to performance, group-efficacy is posited as a significant predictor of group performance as is the analogous case at the individual level of analysis (Bandura, 1986; Shamir, 1990). Our findings support relationships between group-efficacy and assessments of group process and group outcomes. We examined three general methods for assessing group-efficacy, namely, the potency scale, aggregating group members' personal estimates of their group's capability, and a collective judgment formed by the group after discussion for sequential performance periods. In addition, we looked at several different types of group-efficacy, including efficacy concerning time to complete an agreement, efficacy regarding intra-group agreement, and generalized efficacy.

Our findings provide evidence regarding each of the various components of construct validity across the three methods and three types of group-efficacy. Also, our findings suggest that the original eight-item potency scale developed by Guzzo and his colleagues is unidimensional and reliable. Findings regarding the aggregation method and the discussion method indicated that the original scales needed to be revised. Based on traditional measures of self-efficacy, the original versions of both of

these measures required respondents (either individuals or groups) to rate their efficacy for several levels of performance varying in difficulty. As was the case with previous assessments of self-efficacy (e.g., see Locke et. al., 1984), our findings indicate that measures of group-efficacy should be composed of items which represent moderate levels of difficulty (levels of performance which not all respondents can achieve, but at least some respondents can achieve). When the aggregate and group discussion measures of group-efficacy were revised to consist of only moderate items, scales using either method demonstrated unidimensionality and high reliability.

Perhaps the most intriguing findings were obtained in our final set of analyses comparing the predictive validity of the various measures of group-efficacy. Our results suggest that assessments of group-efficacy using group discussion methods are better predictors of group outcomes than assessments using aggregation methods but that netiher approach was particularly powerful. The assessments using the group discussion method explained more variance in the time it took groups to reach agreement and in the level of intra-group agreement.

In addition, our analyses compared the predictive validity of different types of group-efficacy. In comparing specific versus general measures of group-efficacy, we found that specific group-efficacy measures best predicted specific group outcomes and general group-efficacy measures best predicted general group outcomes. For example, measures of task specific time-based efficacy were the strongest predictors of the actual time it took the groups to reach a solution in the negotiation; the very general group potency measure developed by Guzzo and his colleagues was the strongest predictor of general group process effectiveness. Based on these findings, we recommend that researchers select the type of group-efficacy which best reflects the specific group outcomes of interest in a given context.

A potential criticism of our design is that a within-groups approach can bolster the correlation among measures. As mentioned earlier, to test this potential bias, efficacy estimates for groups that utilized the aggregate method prior to the group discussion method in trial 1 (sequence A) were compared with groups that initially used the group discussion method followed by the aggregate method

in trial 1 (sequence B). We found greater differences between methods than within methods. Aggregated scores were different from group discussion scores, but aggregated scores obtained using Sequence A were not significantly different from aggregated scores obtained using Sequence B. Additionally, scores obtained in the two sequences show similar relations to our various performance indicators. Thus, our results do not merely reflect a group-imposed norm on individual's personal evaluations. As in most investigations of self-efficacy, the within-subjects design does not appear to have biased our results.

While our measures show modest promise for predicting multiple facets of group performance, analysis of alternative measures of group-efficacy may still be warranted. Bar-Tal (1990) suggests that group beliefs may be manifested in various group products. They may be inferred, for instance, from various verbal outputs, such as interactions during meetings. They may also be inferred from formally presented group products such as documents or presentations. In other words, a group may publicize their group beliefs. Thus, it is important to note that formal statements of group beliefs may not match what actually constitutes group beliefs. The formal presentation may instead describe ideal beliefs, may represent minority opinion that is not a part of the generally held belief, or may present beliefs about which there is disagreement within a group (Bar-Tal, 1990).

In terms of application, we realize that sample size and degree of administrative flexibility are important considerations. Aggregating individual responses may be more convenient for some tasks and certain work settings for which direct interaction is not practical or possible (e.g., virtual teams operating in different countries). However, it should be noted that with careful administrative planning, the group discussion approach has been utilized successfully in field settings involving samples as large as 180 employees consisting of 5 to 15 person groups (Gibson, 1995).

Group-efficacy measures that rely on group discussion methods appear to be somewhat more robust than aggregation methods, particularly when task specific outcomes such as time efficiency or degree of agreement among members is of critical interest, although these differences are not terribly

strong. When resources permit, we recommend using the discussion method. Second, a potency scale also appears to be a convenient and valid method of measuring general group-efficacy and should be especially useful when general group effectiveness is of greatest interest.

Reference Notes

1. The size of the study groups varied somewhat due to the university's assignment practices (giving priority to geographic proximity for group members) for group membership. Groups ranged from five to six members (16 groups had 6 members and 14 groups had 5 members). If a group only had five members, the role of vice president of legal affairs was dropped from the role play. Any group having fewer than five members (due to absence on the day of the study) were assigned the task of observing the various groups and helping the researchers. Five-person groups and six-person groups were equally distributed across experimental conditions.

2. We conducted parallel analyses using individual scores and found comparable results. While the magnitude of some findings was stronger at the individual level (with more degrees of freedom), the patterns of results generally did not differ across levels.

References

Ancona, D.G. & Caldwell, D.F. (1992). Demography and design: Predictors of new product team

performance. *Organization Science*, **3**, 321-341.

Bandura, A. (1982). Self-efficacy mechanism in human agency. *American Psychologist*, 37, 22-147.

Bandura, A. (1986). *The social foundations of thought and action.* Englewood Cliffs, NJ: Prentice-Hall.

Bandura, A. (1997). *Self-efficacy: The exercise of control.* New York, NY: W.H. Freeman &

Company.

Bandura, A., & Cervone, D. (1986). Differential engagement of self-reactive influences in cognitive

motivation. *Organizational Behavior and Human Decision Processes*, **38**, 92-113.

Bar-Tal, D. (1990). *Group beliefs.* New York: Springer-Verlag.

Campbell, D.T., & Fiske, D.W. (1959). Convergent and discriminant validation by the multitrait-

multimethod matrix. *Psychological Bulletin*, **56**, 81-105,

Campbell, N.K. & Hackett, G. (1986). The effects of mathematics task performance on math self-

efficacy and task interest. *Journal of Vocational Behavior*, **28**, 149-162.

Campion, M.A., Medsker, G.J., & Higgs, A.C. (1993). Relations between work group characteristics

and effectiveness: Implications for designing effective work groups. *Personnel Psychology*, **46**,

823-850.

Cervone, D., & Peake, P.K. (1986). Anchoring, efficacy, and action: The influence of judgmental

heuristics on self-efficacy judgments and behavior. *Journal of Personality and Social

Psychology*, **50**, 492-501.

Cervone, D., Jiwani, N., & Wood, R. (1991). Goal setting and the differential influence of self-

regulatory processes on complex decision-making performance. *Journal of Personality and Social

Psychology*, **61**, 257-266.

Cooper, W.H., & Richardson, A.J. (1986). Unfair comparisons. *Journal of Applied Psychology*, **71**,

179-184.

Earley, P.C. (1993). East meets West meets Mideast: Further explorations of collectivistic and individualistic work groups. *Academy of Management Journal*, **36**, 319-348.

Earley, P.C. (1994). Explorations of the individual and the collective self: Sources of self-efficacy across cultures. *Administrative Science Quarterly*, **39**, 89-117.

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

Gersick, C.J.G. (1988). Marking time: Predictable transition in task groups. *Academy of Management Journal*, **32**, 224-309.

Gersick, C.J.G. (1989). Time and transition in work teams: Toward a new model of group development. *Academy of Management Journal*, **31**, 9-41.

Gibson, C.B. (1995). *Determinants and consequences of group-efficacy beliefs in work organizations in the U.S., Hong Kong, and Indonesia.* Unpublished Doctoral Dissertation, University of California, Irvine.

Gist, M.E. (1987). Self-efficacy: Implications for organizational behavior and human resource management. *Academy of Management Review*, **12**, 472-485.

Goodman, P.S., Ravlin, E.C., & Schminke, M. (1987). Understanding groups in organizations. In B.M. Staw & L.L. Cummings (Eds.), Research in organizational behavior (Vol. 9, pp. 121-173). Greenwich, CT: JAI Press.

Greenhalgh, L. (1984). *Strategic marketing decisions at Porsche of America: The Carrera task-force simulation.* Dartmouth College, Hanover, New Hampshire.

Guzzo, R.A. (1986). Group decision making and group effectiveness in organizations. In P.S. Goodman (Ed.), *Designing effective work groups.* San Francisco, CA: Jossey-Bass.

Guzzo, R.A., Yost, P.R., Campbell, R.J., & Shea, G.P. (1993). Potency in groups: Articulating a construct. *British Journal of Social Psychology*, **32**, 87-106.

Hackman, R.J. (1990). *Groups that work (and those that don't).* San Francisco, CA: Jossey-Bass.

Hill, T., Smith, N.D., & Mann, M.F. (1987). Role of efficacy expectations in predicting the decision to use advance technologies. *Journal of Applied Psychology*, **72**, 307-314.

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

Klein, K.J., Dansereau, F., & Hall, R.J. (1994). Levels issues in theory development, data collection and analysis. *Academy of Management Review*, 19(2): 195-229.

Lindsley, D.H., Brass, D.J., & Thomas, J.B. (1995). Efficacy-performance spirals: A multilevel perspective. *Academy of Management Review*, **20**, 645-678.

Little, B.L. & Madigan, R.M. (1994). *Motivation in work teams: A qualitative and quantitative exploration of the construct of collective efficacy.* Paper presented at the 1994 Academy of Management Meeting, Dallas, TX.

Locke, E.A., Frederick, E., Lee, C., & Bobko, P. (1984). Effect of self-efficacy, goals, and task strategies on task performance. *Journal of Applied Psychology*, **69**, 241-251.

McGrath, J.E. (1984). *Groups: Interaction and performance.* Englewood Cliffs, N.J.: Prentice-Hall.

Mohrman, A.M., Cohen, S.G., & Mohrman, S.A. (1995). *Designing team based organizations.* San Francisco, CA: Jossey-Bass.

Neale, M.A., & Northcraft, G.B. (1991). Behavioral negotiation theory: A framework for conceptualizing dyadic bargaining. In L.L.Cummings & B.M.Staw (Eds.), *Research in organizational behavior, Vol. 13.* Greenwich, CT.: JAI Press.

Parker, L.E. (1994). Working together: Perceived self- and collective-efficacy at the workplace. *Journal of Applied Social Psychology*, **24**, 43-59.

Porter, G. (1992). *Collective efficacy and esteem: Measurement and study of the "self" attributes at a group level.* Unpublished doctoral dissertation, Ohio State University.

Saks, A.M. (1992). *A self-efficacy theory of organizational socialization and work adjustment.* Paper presented at the 1992 Academy of Management Meeting, Las Vegas, NV.

Sayles, L.R. (1958). *The behavior of industrial work groups*. New York: Wiley.

Shamir, B. (1990). Calculations, values, and identities: The sources of collectivistic work motivation. *Human Relations*, **43**, 313-332.

Shea, G.P., & Guzzo, R.A. (1987a). Groups as human resources. In K.M. Rowland & G.R. Ferris (Eds.) *Research in personnel and human resource management*, Vol. 5. Greenwich, CT: JAI Press.

Shea, G.P., & Guzzo, R.A.(1987b). Group effectiveness: What really matters? *Sloan Management Review*, **28**, 25-31.

Wood, R., & Bandura, A. (1989). Social cognitive theory of organizational management. *Academy of Management Review*, **14**, 361-383.

Zander, A.W. & Medow, H. (1963). Individual and group aspiration. *Human Relations*, **16**, 89-105.

Table 1

Descriptive Statistics and Correlations for Group-efficacy and Performance Scales

	M	SD	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)
														Pearson Correlations						
(1) T1 Aggregate General Group-efficacy	78.3	10.88	(80)	64	83	66	42	50	74	46	61	63	40	28	50	39	-10	-16	25	-08
(2) T1 Aggregate Time Group-efficacy	72.20	11.44		(81)	52	55	64	36	50	84	46	44	59	23	32	22	-25	-10	08	-10
(3) T1 Aggregate Agreement Group-efficacy	75.28	12.94			(84)	61	35	62	76	35	85	62	36	48	48	39	02	06	30	-01
(4) T2 Aggregate General Group-efficacy	75.66	14.29				(83)	70	80	78	42	64	83	53	68	32	48	-07	15	34	07
(5) T2 Aggregate Time Group-efficacy	61.39	15.87					(82)	62	51	60	57	50	91	47	30	37	-50	22	42	20
(6) T2 Aggregate Agreement Group-efficacy	70.57	17.54						(83)	63	28	76	68	51	86	18	40	-18	23	49	18
(7) T1 Discussion General Group-efficacy	82.01	13.51							(77)	47	75	83	51	56	38	38	-16	02	31	09
(8) T1Discussion Time Group-efficacy	72.65	17.44								(85)	40	41	62	21	25	22	-44	-14	12	-10
(9) T1Discussion Agreement Group-efficacy	76.35	17.76									(82)	68	59	70	36	35	-25	16	49	20
(10) T2 Discussion General Group-efficacy	80.80	14.05										(75)	50	68	27	40	-29	14	35	15
(11) T2 Discussion Time Group-efficacy	64.13	18.61											(75)	38	26	31	-61	23	39	22
(12) T2 Discussion Agreement Group-efficacy	75.07	19.61												(78)	13	32	-18	31	47	18
(13) T1 Potency	7.65	1.41													(91)	78	-06	-06	14	09
(14) T2 Potency	7.66	1.43														(94)	-11	11	25	18
(15) Minutes to Complete Agreement	69.66	13.55															--	-16	-53	-36
(16) % Agreement Among Group Members	94.31	14.10																--	36	54
(17) Effectiveness of Group Process	4.19	.38																	(74)	60
(18) Effectiveness of Group Solution	4.01	.43																		(85)

Note: Decimals omitted; correlations greater than .38 or less than -.38 are significant at p<.05; scale reliabilities are shown in parentheses; n=30 groups.

Table 2

Descriptive Statistics, Reliabilities and Correlations for Group-efficacy and Performance Measures Revised Scales (Includes Trial 2 and Moderate Difficulty Items Only)

	M	SD	Pearsons Correlations (1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
(1) Aggregate General Efficacy	80.75	15.80	(85)	71	82	82	51	69	68	-04	10	32	05
(2) Aggregate Time Efficacy	65.44	18.82		(83)	58	55	90	42	56	-47	21	41	18
(3) Aggregate Agreement Efficacy	74.49	19.78			(84)	65	43	86	61	-18	21	46	18
(4) Discussion General Efficacy	87.12	14.68				(77)	47	63	66	-30	05	35	16
(5) Discussion Time Efficacy	68.90	21.60					(72)	29	51	-56	23	39	22
(6) Discussion Agreement Efficacy	79.10	20.85						(75)	52	-13	30	43	21
(7) Potency	7.70	.89							(94)	-19	20	38	09
(8) Time in Minutes	69.66	13.56								--	-16	-53	-36
(9) % Intra-Group Agreement	94.32	14.11									--	36	54
(10) Process Score	4.19	.38										(73)	60
(11) Outcome Score	4.01	.43											(85)

Note: Decimals omitted; correlations greater than .32 or less than -.32 are significant at $p<.10$; correlations greater than .38 or less than -.38 are significant at $p<.05$; scale reliabilities are shown in parentheses; n=30 groups.

Table 3

Results of Multiple Regression Analysis Comparing Predictive Validity of Measures

Predictor	Minutes to Complete Agreement		% Agreement Among Members		Effectiveness of Group Process		Effectiveness of Group Solution	
	Beta	t for Beta	Beta	t for Beta	Beta	t for Beta	Beta	t for Beta
Time Efficacy/Potency								
Discussion Method	-.69	-1.93†	.17	.39	.09	.23	.32	.75
Aggregate Method	.09	.26	.00	.00	.22	.53	-.10	-.23
Potency	.09	.47	.11	.49	.21	1.01	-.01	-.06
	$R^2 = .32$*		$R^2 = .06$		$R^2 = .20$		$R^2 = .05$	
Agreement Efficacy/Potency								
Discussion Method	.09	.23	.47	1.27	.17	.49	.22	.59
Aggregate Method	-.18	-.45	-.26	-.65	.21	.59	.00	.10
Potency	-.13	-.52	.11	.48	.16	.74	-.03	-.11
	$R^2 = .05$		$R^2 = .11$		$R^2 = .23$†		$R^2 = .05$	
General Efficacy/Potency								
Discussion Method	-.82	-2.50*	-.19	-.56	.18	.55	.34	.98
Aggregate Method	.72	2.21*	.08	.21	-.01	-.02	-.26	-.72
Potency	-.13	-.51	.27	1.03	.26	1.04	.05	.17
	$R^2 = .24$†		$R^2 = .05$		$R^2 = .16$		$R^2 = .04$	

Note: † $p < .10$ * $p < .05$; n=30 groups.

APPENDIX:

<u>MEASURES USED IN STUDY</u>

POTENCY MEASURE

GROUP NAME _____

PARTICIPANT ID _____

Please answer the items below using a scale from 1-10 on which 1=To no extent, 3=To a limited extent, 5=To some extent, 7=To a considerable extent, and 10=To a great extent.

1. My group has confidence in itself.	
2. My group believes it can become unusually good at producing high-quality work.	
3. My group expects to be known as a high-performing team.	
4. My group feels it can solve any problem it encounters.	
5. My group believes it can be very productive.	
6. My group can get a lot done when it works hard.	
7. No task is too tough for my group.	
8. My group expects to have a lot of influence around here.	

DISCUSSION MEASURE OF GENERAL GROUP-EFFICACY

GROUP NAME_____

This group's effectiveness during the Porsche exercise is being observed and will be rated on a ten-point scale. This scale is defined below:

1 = Very ineffective; no clear solution will result from the group discussions
5 = Average effectiveness; collaboration on some issues and disagreement on others
10 = Very effective; mutually shared consensus as to an appropriate solution

When the researcher tells you to begin, talk among yourselves and try to estimate whether the group will be able to achieve each level of effectiveness. Have one person in the group record the group's response in the chart. We only need one form for each group.

Please indicate in column A how certain the group is that it can achieve each level of effectiveness. For example, 0% means that the group cannot achieve that level, 50% means that the group may or may not achieve at least that level, and 100% means that the group is certain that it can achieve at least that level of effectiveness. In addition, in column B, please indicate "Y" for yes or "N" for no for whether or not the group can achieve at least this level of effectiveness

	COLUMN A (0% TO 100% CERTAINTY)	COLUMN B Y = YES N = NO
Our group will receive a rating of 9 - 10 for general effectiveness.		
Our group will receive a rating of 7 - 8 for general effectiveness.		
Our group will receive a rating of 5 - 6 for general effectiveness.		
Our group will receive a rating of 3 - 4 for general effectiveness.		
Our group will receive a rating of 1 - 2 for general effectiveness.		

Note: The aggregate measure of general group-efficacy was identical to this measure with two exceptions: (1) in the instructions, "the group" was replaced with "your group" and (2) in the items, "Our group" was replaced with "My group." When completing the aggregate measure, each manager filled out a form individually and responses were aggregated at the group level.

DISCUSSION MEASURE OF TIME-BASED GROUP-EFFICACY

GROUP NAME_____

One person in this group will be in charge of keeping track of how long it takes the group to reach a solution during the Porsche exercise. We are interested in the group's expectations of how long it will take to reach a solution during the exercise.

When the instructor tells you to begin, talk among yourselves and try to estimate how long it will take the group to reach a decision during the Porsche exercise. Have one person in the group record the group's response in the chart. We only need one form per group.

Please indicate in column A how certain the group is that it can reach a decision in the specified number of minutes. For example, 0% means that the group cannot reach a decision in that number of minutes, 50% means that the group may or may not reach a decision in at least that number of minutes, and 100% means that the group is certain that it can reach a decision in at least that number of minutes. In addition, in column B please indicate "Y" for yes or "N" for no for whether or not the group can reach a decision in that number of minutes.

	COLUMN A (0% TO 100% CERTAINTY)	COLUMN B Y = YES N = NO
Our group will reach a solution in the Porsche exercise in 30 - 44 minutes.		
Our group will reach a solution in the Porsche exercise in 45 - 59 minutes.		
Our group will reach a solution in the Porsche exercise in 60 - 74 minutes.		
Our group will reach a solution in the Porsche exercise in 75 - 89 minutes.		
Our group will reach a solution in the Porsche exercise in 90 - 104 minutes.		

Note: The aggregate measure of time-based group-efficacy was identical to this measure with two exceptions: (1) in the instructions, "the group" was replaced with "your group" and (2) in the items, "Our group" was replaced with "My group." When completing the aggregate measure, each manager filled out a form individually and responses were aggregated at the group level.