Small Group Research

http://sgr.sagepub.com

When Member Homogeneity is Needed in Work Teams: A Meta-Analysis

Clint A. Bowers, James A. Pharmer and Eduardo Salas

Small Group Research 2000; 31; 305
DOI: 10.1177/104649640003100303

WHEN MEMBER HOMOGENEITY IS NEEDED IN WORK TEAMS
A Meta-Analysis

CLINT A. BOWERS
JAMES A. PHARMER
EDUARDO SALAS
University of Central Florida

A meta-analytic integration of 57 effect sizes from 13 studies (567 teams, 2,258 participants) was performed to determine if groups that are homogeneous with respect to gender, ability level, and personality achieve higher levels of performance than teams that are heterogeneous on these attributes. Although individual studies often show marked differences between homogeneous and heterogeneous groups, the results of this integration show the combined effect sizes of these studies to be small, though not significant, in favor of heterogeneous groups. It appears that the significant effects found in many of the included studies can be attributed to the type and difficulty of the task used in the investigation. Implications for team construction are discussed.

Team performance has been described as the outcome of dynamic processes reflected in coordination and communication processes that teams develop over time (Hackman, 1983). These process variables, in turn, are influenced by situational and organizational characteristics as well as the characteristics of the task (e.g., complexity, organization), the work (e.g., structure, norms), and the knowledge, skills, and abilities of the individual. Although team characteristics have also been posited to influence the coordination and communication variables that reflect team performance (Nieva, Fleischman, & Reick, 1978), there is a relative paucity of literature devoted to understanding how team characteristics intervene in team processes. This is surprising, given the changing demographics of the labor force within the United States over the

SMALL GROUP RESEARCH, Vol. 31 No. 3, June 2000 305-327

past several decades. These changing demographics account for, but are not limited to, increased gender diversity, cultural diversity, and age diversity (Jackson, May, & Whitney, 1995). Women, in particular, have steadily made strides into occupations that had previously been predominantly staffed by males (e.g., military and management positions). In fact, Jackson et al. (1995) suggested that by the year 2000, the labor force would be gender balanced.

Work teams are routinely assembled from individuals varying in knowledge, skills, abilities, personality, and attitude. Yet, it is still unclear whether teams composed of a variety of these attributes perform better than teams of homogeneous composition. The purpose of the current meta-analytic study is to investigate the effects of similarity (or dissimilarity) among team members on performance across the different attributes that determine the content of diversity in teams.

One of the challenges of performing a meta-analysis of this type is to determine variables on which team members may have been systematically varied. In this case, it was necessary to decide on a taxonomy that accurately reflects the types of individual attributes that differentiate team members in such a way as to affect overall team performance. Several useful taxonomies have been posited.

Jackson et al. (1995) characterize the individual attributes that reflect the content of diversity within a work team as either *readily detectable attributes* or *underlying attributes*. Readily detectable attributes are apparent after only brief exposure to an individual team member. These include such attributes as gender, ethnicity, and age. Underlying attributes are more difficult to determine from brief exposure and may include such characteristics as knowledge, skills, abilities, and personality characteristics. Furthermore, Jackson asserts that individual attributes may be categorized in two ways. Attributes that are closely associated with the objectives of the work team (e.g., knowledge, skills, abilities, and organizational tenure) are task related. Attributes that are likely to affect social relationships (e.g., gender, ethnicity) between work-team members are relations oriented. This taxonomy provides a useful framework for understanding the relationship between individual attributes, team diversity, and performance. However, it is important to note that specific attributes may not fall exclusively into any one category. For example, ethnicity may affect the relationships between particular team members and therefore be categorized as relations oriented. However, ethnicity may also be a determinant in the likelihood that an individual is promoted within an organization and therefore affect such attributes as organizational tenure and knowledge, skills, and abilities. However, from Jackson's taxonomic approach, it is clear that diversity may affect task performance directly by influencing the objectives of the work team. It may also affect performance indirectly by influencing the social interactions between team members. Research efforts over the past decade have begun to identify and understand the sequences and interactions that relate a team's structure to its decision output (Poole & Roth, 1989). These efforts may illuminate the influence of these interactions on the output of teams.

An earlier taxonomy introduced by Morgan & Lassiter (1992) distinguished the differences between team members in four broad categories: (a) biographical differences, (b) personality differences, (c) differences in abilities, and (d) leadership differences. For the purpose of categorizing team composition variables that may affect overall team performance, this taxonomy has at least two advantages over the categorization strategy put forth by Jackson and her colleagues (1995). First, individual attributes more easily fall into one of the four categories in the taxonomy. Second, there appears to be less room for experimenter bias in deciding where an individual attribute belongs. However, whereas a meta-analytic review including all four of these categories in the Morgan & Lassiter taxonomy would certainly have utility, it would be unwieldy and redundant to include leadership within the current integration. This is because many excellent meta-analytic reviews already exist within the literature (see Lord, de Vader, & Alliger, 1986; Dansereau, Yammarino, Markham, & Alluto, 1995; Gastil, 1994; Peters, Hartke, & Pohlman, 1985). Hence, this meta-analysis focuses on only three of the categories within the taxonomy. However, one of the goals of meta-analysis is to provide areas that require further study. Consequently, greater specificity is necessary within these three broad categories. A more detailed description of this further categorization follows.

Biographical differences. Of the numerous possible biographical differences (race, age, educational background, sociocultural background, etc.), the effect of gender composition on team performance was chosen for this meta-analysis for two reasons. First, the number of studies addressing the relationship between gender composition and team performance warrants a meta-analytic integration. Second, the number of studies addressing other biographical differences in work teams is relatively sparse (Baugh & Graen, 1997). This is perhaps due to the restricted age and socioeconomic ranges of the population pool from which many of these studies have drawn participants (i.e., introductory psychology students). Furthermore, gender differences are investigated within the category of biographical differences for the reasons previously discussed.

Wood (1987) provided an excellent integration of the literature on the effect of gender composition on team performance. Much of the literature she reviewed has shown significant differences in performance at the group level. In general, mixed-gender groups outperformed same-gender groups, although this relationship was not significant. Furthermore, all male groups outperformed all female groups. However, this literature is subject to several methodological concerns. For example, few studies have controlled for individual differences that might have been present prior to group interaction (i.e., personality differences, racial differences, etc.). The type of task that was used in a particular study may have been inherently more suited to one gender over another. For example, women may be more motivated by tasks that require more coordination between team members; on the other hand, men may be more motivated than women by tasks requiring little social interaction and more quantitative skills. Furthermore, the performance measure used may also benefit one gender over another. In her integration of the gender composition literature, Wood notes that men may be more likely to perform well on tasks where the dependent measure is the quantity of solutions, whereas women may excel at tasks where quality of solutions is the dependent measure. This difference may exist because men tend to be task oriented and women tend to be more process oriented.

Interestingly, the relationship between team composition and performance may be related to who is rating the effectiveness. In an investigation of cross-functional project teams, Baugh and Graen (1997) found that ratings of effectiveness by team members of teams that were heterogeneous with respect to either gender or racial composition were lower than ratings by external observers.

To further complicate the role of gender composition on team performance, there are a number of intervening variables that may affect the outcome of a gender-differences study. Researchers have found gender differences in such variables as task motivation (Kerr & MacCoun, 1985; Vancouver, Rubin, & Kerr, 1991), self-presentation (Swanson & Tjosvold, 1979), and leadership (Jacobson & Effertz, 1974).

In light of these findings, it was decided to include gender as a predictor variable in the current investigation because it was likely to be responsible for a significant portion of the variance in these studies. Furthermore, as mentioned previously, women are assuming a larger role in the workforce, which has important implications on the design of work teams.

Differences in initial ability. Steiner (1972) maintained that the potential productivity of a team is equal to the sum of the individual abilities within a group. This summation represents the maximum level of performance that a team can hope to achieve. The difference between the potential team performance and the actual team performance is considered *process loss*. According to Steiner, process loss is the likely result of performance decrement due to added coordination factors inherent in team tasks and decreased motivation among team members. However, it is often the case that teams perform beyond the additive abilities of individual members. This *process gain* could be the result of efficient workload sharing among team members or better error-capturing because team members are able to cross-check each other's work.

In a review of the team performance literature, Hill (1982) cites numerous examples of research suggesting that group processes could lead to process gain. Hill suggests that process gain can be attributed to two potential sources: (a) member capacity to learn (i.e., observational learning, incidental learning, social facilitation, etc.) and (b) cognitive stimulation (i.e., encouragement from other team members, etc.). Further evidence of process gain is provided in an investigation of 3-person tank crews. Tziner and Eden (1985) found that teams composed exclusively of high-ability individuals exhibited process gains in performance, teams composed exclusively of low-ability individuals exhibited process loss, and teams of heterogeneous ability composition showed additive levels of performance. These findings indicate that teams composed exclusively of low-ability team members should be avoided in complex tasks because of the likelihood of process loss (Morgan & Lassiter, 1992). Alternatively, maximum performance will be achieved when teams are composed exclusively of high-ability individuals. However, it remains unclear whether this effect is limited to complex tasks with high coordination demands or if it is likely to occur with all types of tasks. Furthermore, the dependent measure used in the Tziner & Eden study was ranked effectiveness as rated by crew commanders. Although these measures do provide a good means of comparing performance globally, it remains unclear which specific components of the task were most affected by the distribution of ability within teams.

Differences in personality. There is surprisingly little research on the effects of homogeneity of personality composition on team performance (Altman & Haythorne, 1967; Driskell, Hogan, & Salas, 1987; Hackman & Morris, 1975). Furthermore, research suggests that the relationship between personality variables and performance is mediated by several other intervening process variables. In a series of classic studies designed to investigate the relationship between team composition and creativity in dyads, Triandis, Hall, and Ewen (1965) concluded that dyads composed of members of equivalent abilities that are heterogeneous in attitude are more creative than dyads that are homogeneous. This suggests that in tasks that require high levels of creativity, teams composed of individuals with differing attitudes may perform at higher levels than teams with like-minded individuals. Self-efficacy also appears to mediate the relationship between personality and performance (Thoms, Moore, & Scott, 1996). Altman and Haythorne (1967) suggest that the relationship between personality and performance is mediated by compatibility between team members. Although some personality differences may produce decrements in this compatibility, this does not necessarily translate into decreases in performance. Likewise, personality similarities may produce better compatibility between team members, but this does not necessarily translate to higher levels of performance (Mullen & Copper, 1994).

The relationship between personality and performance also appears to be dependent on the type of task that is being performed and the personality dimensions that vary between team members. This again could be related to the intervening process variables between personality composition and performance output. Because of this interaction between compatibility and team performance, it is difficult to determine the relative main effects of homogeneity of group composition individually without taking the task into account.

One need only perform a cursory review of the personality literature to notice that there are literally hundreds of individual personality dimensions for which to investigate the effects of team composition on performance. An integration of each of these dimensions is beyond the scope of the present meta-analysis. As such, this integration will focus on the homogeneity of team personality composition as a whole, rather than focusing on more narrowly defined individual personality dimensions. Future studies should address the individual dimension in more detail.

THEORIES OF HOMOGENEITY EFFECTS

Tziner (1985) identifies two competing theories of social psychology to explain how performance may be affected by team composition. *Similarity theory* argues that homogeneous groups are likely to be more productive because of the mutual attraction shared by team members of similar demographics. Heterogeneous groups, on the other hand, are predicted to be less productive because of inherent tensions between team members. Superficially, this theory seems very plausible. Individuals of similar backgrounds and abilities often do appear to be attracted to one another. However, whether interpersonal attraction among team members translates to higher levels of performance is debatable. Likewise, the suggestion that differences between team members engender negative feelings toward one another and in turn abate performance is speculative at best. In fact, the second theory, *equity theory,* predicts that team performance is enhanced by the tension that arises between dissimilar individuals within a group. According to equity theory, individuals compare themselves with other members of the group. If the perceived ratio of personal input to reward appears inequitable, the team members may rely on several strategies to restore equity. If the individual perceives that other team members are being unfairly rewarded for less work, the individual may slow their own productivity or work to increase the productivity of others. If the individuals believe that they themselves are being unfairly rewarded, equity theory predicts that the individuals will increase output to restore effort-to-reward equity. Moreover, equity theory can be extended beyond effort to reward ratios. Perceived differences in status may serve to limit the degree to which team members are willing to communicate and interact with one another, as well as share resources equitably (Jackson et al., 1995). Certainly, these interpersonal dimensions may restrict the performance of diverse teams. For example, in work teams where status is defined by the expertise of the team member, those team members who lack credentials may be unwilling to put forth their own opinions or to request resources from other more qualified members. Alternatively, high-expertise team members may not give much weight to the opinions of others with low expertise, and they may be unwilling to equitably share resources because of the differential status.

Although both of these competing hypotheses predict performance based on team composition, their predictions lead to different conclusions. For example, similarity theory predicts that performance of homogeneous, high-ability groups will exceed the additive capacities of individual team members (i.e., process gain). Equity theory, on the other hand, predicts that the performance of the team would most likely be less than or equal to the summed abilities of individual team members because much of the potential productivity of the group is focused on interpersonal issues instead of the task. Similarity theory also predicts that those teams that are heterogeneous in ability will not perform as well as those teams with homogeneous ability composition. Equity theory predicts that performance in heterogeneous groups is dependent on the ratio of high performers to poor performers.

The literature on diversity in management teams also reaches mixed conclusions on how compositional variables may affect team process and outcome. Weirsema and Bantel (1992) have noted that homogeneity on demographic traits has been shown to lead to a shared language among individuals that enhances communication frequency and integration. Weirsema and Bantel have suggested that a team's demographic homogeneity allows a shared language to develop between individuals of like backgrounds. This shared language is likely to enhance communication and integration within the team. As a result, homogeneous teams would be more likely to perform at a high level on tasks that require coordinated activities between team members. However, similarity has also been shown to be related to a lack of openness to new sources of information (Bantel, 1994). Thus, demographically homogeneous teams would be predicted to perform poorly on tasks that require the team to seek and use a broad range of information.

Demographic heterogeneity, on the other hand, has been shown to be positively related to creativity in decision making (Bantel & Jackson, 1989; Hoffman & Maier, 1961). This is perhaps due to the diversity of experience and perspective brought to the team by individuals with dissimilar attributes. Furthermore, heterogeneous composition is positively related to planning openness (Bantel, 1994). However, in contrast to homogeneous teams, heterogeneous teams are likely to have poorer communication.

Suffice to say, varying levels of performance can be expected from teams composed of similar or dissimilar team members, depending on the type of task and level of communication or coordination necessary to perform it. We posit that in general, an integration of the literature will show that homogeneous teams attain higher levels of performance than heterogeneous teams for at least two reasons. First, an integration of the team cohesion literature (Mullen & Copper, 1994) revealed evidence of a small relationship between team cohesiveness and performance. Because operational definitions of cohesion often use measures of interpersonal attraction as a component of cohesion, the theorized interpersonal attraction between similar team members should produce at least a small performance gain. Second, it seems more probable that conflicts will arise between dissimilar individuals than similar individuals. These conflicts may interfere with performance in that the finite resources available for team coordination are being used for conflict resolution of interpersonal issues instead of actual productivity, especially in tasks that require smooth and timely interactions among team members (Morgan & Lassiter, 1992). As many team tasks performed in the laboratory require smooth coordination between team members, it is perhaps more likely that these tasks will favor homogeneous groups.

PREDICTORS OF THE EFFECTS OF HOMOGENEITY

There are a number of factors that may intervene in the relationship between team composition and performance. First, the size of the team has been implicated in such phenomena as social loafing and diffusion of responsibility. Clearly, perceptions of equity of the ratio between reward and effort must be affected by the size of the group as well. For example, in small teams, it is often obvious who is performing at high levels and who is not. As team size increases, these relationships become less and less obvious due to diffusion of responsibility (Latane’, 1981). Second, the type of task that the team performs is also critical to the understanding of the relationships between composition and performance. The studies that were integrated for this meta-analysis used tasks that can be considered to be cognitive (e.g., intelligence testing, “Tower of Hanoi” strategy game), performance (e.g., surveying tasks, target detection), or production (e.g., sewing tasks, typing tasks). These different types of tasks intrinsically require different levels of coordination and teamwork, and distinct differences in performance on these types of tasks can therefore be expected. Third, the difficulty of the task determines the resources that the team must use to perform it. A low-difficulty task will require individual team members to use fewer cognitive resources than a high-difficulty task. As such, team members will have more resources to devote to coordination and teamwork. Therefore, higher levels of performance are expected in low-difficulty tasks than in high-difficulty tasks. Fourth, the measure of performance itself influences how large a performance effect will be revealed in the particular study. For example, a team that has been instructed to manufacture as many products as possible within a specified time will probably produce a high quantity of low quality products. If the performance measure is the subjective quality of the product, then this team would be rated low. However, if the performance measure is quantity, then the team would receive a high performance score.

Although the relative importance of any of the intervening factors discussed above may be debated, ignoring these factors entirely would limit the generalizability of the meta-analysis. As such, these factors will also be included in the meta-analytic review to ensure external validity.

METHOD

PROCEDURE

A comprehensive search of the literature on team and small-group research was conducted using computer database searches (i.e., PsychLit, CARL UNcover, PsychFirst). The "invisible college" technique (Mullen, 1989) involves contacting individuals who are currently working in the field of interest to obtain studies that have not been published. One major concern in meta-analytic integration is that important studies may not be included because they did not show significant findings and were unfortunately not deemed worthy of publication. These types of studies often end up sitting in a file drawer despite the fact that their results could provide useful data in integrations such as the current investigation. This concern is often referred to as the file drawer problem. Although the invisible college technique was exhaustively performed by contacting authors and laboratories that often publish research on team performance, none of the unpublished studies that were located met the criteria of this study. From this comprehensive search, 13 articles yielding 57 hypothesis tests (567 teams, 2,258 participants) related to the effects of homogeneity of team composition on performance were found. Studies were included in the meta-analytic review if they met the following criteria: First, individual studies had to compare teams whose members were homogeneous with respect to ability, attitude, gender, or personality, to teams that were heterogeneous on these attributes. Second, the studies had to be empirical in nature. Third, the studies had to report relevant statistics (e.g., ANOVA, t test, p value). However, a number of studies reported only means and standard deviations of the groups. Where possible, the corresponding inferential statistics were gleaned from these descriptive values to be included in the meta-analytic database. The determination of which studies were to be included in the analysis and the ratings of the individual studies were made by one of the authors of the current investigation. The data from the four predictor variables were coded and analyzed as nominal data. Reliability of the ratings was established by comparing a subset of 20 hypothesis tests rated by the same individual more than a year earlier. This comparison yielded 85% test-retest reliability.

To illuminate the processes that mediate the effects of group composition on performance, four predictors were included for each of the hypothesis tests reported in this review. First, the type of independent variable for the study (ability, attitude, gender, or personality) was included as a predictor to determine how specific attributes contribute to performance differences related to team composition. Second, the type of measurement used as a performance measure, operationally defined as measures of performance quality, quantity, or accuracy, was included to determine how these measures are affected by the team structure. Third, the difficulty of the task (high, medium, or low), as judged by one of the researchers for this study, was included as a predictor to illuminate the interactions between team composition, performance, and increasing levels of difficulty. Task difficulty was operationally defined as high, medium, or low, depending on the level of stimulus uncertainty, processing demands, and response complexity within the task. Tasks such as business games were deemed high difficulty, whereas tasks such as story writing and puzzle solving were considered medium and low difficulty, respectively. When information regarding task difficulty was available in the articles that were reviewed, this information was taken into account when coding for task difficulty. This was done to try to ensure that the coding accurately reflected the difficulty that was experienced by the participants of the study. Unfortunately, few of the studies reviewed provided this information. Finally, the types of tasks were included as a predictor and were operationally defined as intellectual tasks (e.g., cognitive work), productive tasks (i.e., making a physical product), and performance (i.e., low cognitive demand, physical work) tasks.

For the purposes of this investigation, observations were assumed to be independent even though some studies had multiple observations drawn from the same subject pool. Although this assumption is indeed inaccurate, it allows for a scrutiny of individual measurements rather than combining a number of dependent variables into a grouping variable that is unrepresentative of any of the dependent measures within the individual study. The assumption of independence from studies using the same subject pools has been made in other similar meta-analyses (cf. Mullen et al., 1990; Sanders & Mullen, 1983; For a discussion, see Mullen, 1989).

RESULTS

ANALYSIS

Advanced BASIC Meta-analysis software (Mullen, 1989) was used to integrate the 57 hypothesis tests that met the criteria for this investigation to determine Fisher's *Z*, a measure of effect size weighted by sample size. Initially, combinations and diffuse comparisons of significance and effect sizes were performed to determine the general effects of team composition on performance. It was then necessary to perform focused comparisons of effect sizes for each predictor variable to determine the effect of each predictor variable on team performance. Finally, combinations and comparisons were made within levels of each predictor, using the predictor as a block.

The Fail Safe *N* is the largest number of studies that failed to reach $p = .05$ that one could assume to have failed to discover and still conclude that the combined result is significant (Mullen, 1989). Consequently, this number was calculated for each of the comparisons of effect size for individual predictors as a method of checking robustness.

General effects of homogeneity and heterogeneity. Table 1 shows the Fisher's *Z* and *R* values from the 57 hypothesis tests used for this meta-analysis. In addition, the values of the four predictor variables are included in this table. Although individual studies yielded effects for both homogeneous and heterogeneous composition, the combined effects of team composition on performance for these hypothesis tests, weighted by sample size, yielded a small and insignificant overall effect of team composition on performance in favor of heterogeneous groups, Fisher's $Z = 1.269$, $p = .89$. The Fail Safe *N* for this effect was calculated to be 31.98 studies.

Effects of compositional variable. A focused comparison of effect sizes from the individual investigations revealed a highly significant relationship between the dimension on which the team was varied (ability, personality, and gender) and team performance, Fisher's $Z = 7.292$, $p = .000$. Although the magnitude of effect on team performance significantly covaried with the dimension varied in the individual studies, none of the three compositional measures in this meta-analysis showed a significant effect of homogeneity. In fact, ability composition, personality composition, and gender composition all tended toward higher performance levels in hetero-

TABLE 1: Studies Included in Meta-Analysis

Study	*Statistic*	*Significance (Fisher's* Z*)*	*Effect Size (*r*)*	*CV*	*M*	*D*	*T*	*S*
Aamodt & Kimbrough (1982)	$t(22) = 2.20$ (–)	–2.06	–0.42	2	3	3	1	4
Clement & Schiereck (1973)	$t(3) = 4.22$ (+)	2.20	0.92	4	2	1	3	4
	$t(3) = 2.17$ (+)	2.17	0.78	4	2	1	3	4
Herschel (1994)	$p = 0.91$ (–)	–1.36	–0.10	4	2	1	1	5
	$p = .49$ (+)	2.05	0.00	4	2	1	1	4
Hoffman & Maier (1961)	$t(29) = 0.31$ (+)	–0.31	–0.06	2	1	2	1	4
	$t(31) = 0.74$ (–)	–0.73	–0.13	2	1	2	1	4
	$t(34) = 1.22$ (–)	–1.20	–0.20	2	1	2	1	4
	$t(59) = 0.19$ (–)	–0.19	–0.02	4	1	2	1	4
Hoffman, Harburg, & Maier (1962)	$Z = 1.19$ (–)	–1.19	–0.06	4	2	3	1	4
Hooper, Ward, Hannafin, & Clark (1989)	$F(1,69) = 0.53$ (+)	0.72	0.08	1	2	1	1	4
	$F(1,66) = 2.94$ (+)	1.69	0.20	1	2	1	1	4
Hooper & Hannafin (1991)	$t(123) = 0.69$ (+)	0.69	0.06	1	2	1	1	2
Jacobson & Effertz (1974)	$F(1, 20) = 0.09$ (–)	–0.29	–0.07	4	3	1	1	3
Swanson & Tjosvold (1979)	$t(112) = 1.64$ (+)	1.63	0.15	4	2	1	3	2
Terborg, Castore, & DeNinno (1976)	$F(1,190) = 0.25$ (+)	0.49	0.03	3	1	2	3	3.5
Valacich, Wheeler, Menneche, & Wachter (1995)	$t(6) = 3.63$ (+)	2.52	0.82	1	2	3	1	5
	$t(6) = 0.77$ (+)	0.71	0.29	1	2	3	1	6
	$t(6) = 0.13$ (–)	–0.13	–0.05	1	2	3	1	7
	$t(6) = 0.74$ (–)	–0.70	–0.29	1	2	3	1	8
	$t(6) = 1.52$ (–)	–1.33	–0.52	1	2	3	1	9
	$t(6) = 1.32$ (–)	–1.18	–0.47	1	2	3	1	10
	$t(6) = 3.75$ (+)	2.58	0.84	1	2	3	1	5
	$t(6) = 0.97$ (+)	0.89	0.37	1	2	3	1	6
	$t(6) = 0.28$ (+)	0.27	0.11	1	2	3	1	7
	$t(6) = 1.50$ (–)	–1.32	–0.52	1	2	3	1	8
	$t(6) = 1.64$ (–)	–1.43	–0.55	1	2	3	1	9
	$t(6) = 1.98$ (+)	1.66	0.62	1	2	3	1	10
	$t(6) = 4.71$ (+)	2.91	0.89	1	1	3	1	5
	$t(6) = 1.42$ (+)	1.26	0.50	1	1	3	1	6
	$t(6) = 0.15$ (–)	–0.14	–0.06	1	1	3	1	7
	$t(6) = 2.95$ (–)	–2.22	–0.77	1	1	3	1	8
	$t(6) = 2.33$ (–)	–1.88	–0.69	1	1	3	1	9
	$t(6) = 1.22$ (–)	–1.10	–0.44	1	1	3	1	10
	$t(6) = 4.21$ (+)	2.75	0.86	1	2	3	1	5
	$t(6) = 0.75$ (+)	0.70	0.29	1	2	3	1	6
	$t(6) = 0.16$ (–)	–0.15	–0.06	1	2	3	1	7
	$t(6) = 0.85$ (–)	–0.79	–0.32	1	2	3	1	8
	$t(6) = 1.75$ (–)	–1.51	–0.58	1	2	3	1	9
	$t(6) = 1.53$ (–)	–1.35	–0.53	1	2	3	1	10

(continued)

TABLE 1: Continued

Study	*Statistic*	*Significance (Fisher's* Z*)*	*Effect Size* (r)	*CV*	*M*	*D*	*T*	*S*
	$t(6) = 3.51$ (+)	2.48	0.81	1	2	3	1	5
	$t(6) = 0.98$ (+)	0.90	0.37	1	2	3	1	6
	$t(6) = 0.28$ (+)	0.27	0.11	1	2	3	1	7
	$t(6) = 1.51$ (–)	–1.33	–0.52	1	2	3	1	8
	$t(6) = 2.04$ (–)	–1.70	–0.64	1	2	3	1	9
	$t(6) = 1.95$ (–)	–1.64	–0.62	1	2	3	1	10
	$t(6) = 4.73$ (+)	2.92	0.88	1	1	3	1	5
	$t(6) = 1.42$ (+)	1.26	0.50	1	1	3	1	6
	$t(6) = 0.15$ (–)	–0.14	–0.06	1	1	3	1	7
	$t(6) = 2.95$ (–)	–2.22	–0.76	1	1	3	1	8
	$t(6) = 2.46$ (–)	–1.96	–0.71	1	1	3	1	9
	$t(6) = 1.23$ (–)	–1.11	–0.45	1	1	3	1	10
Vancouver et al. (1991)	$t(27) = 4.14$ (+)	3.61	0.62	4	2	1	3	2
	$t(31) = 3.30$ (+)	3.03	0.51	4	2	1	3	2
Volkema & Gorman (1998)	$t(24) = 0.39$ (+)	0.38	0.08	2	2	2	1	4
	$t(24) = .476$ (+)	0.476	0.05	2	1	2	1	4

NOTE: CV = Compositional Variable; M = Performance Measure; D = Task Difficulty; T = Task Type; S = Group Size. For Statistic, (+) = favors homogeneous group; (–) = favors heterogeneous group. For CV, 1 = Ability; 2 = Personality; 3 = Attitude; 4 = Gender. For M, 1 = Quality, 2 = Quantity, 3 = Accuracy. For D, 1 = Low, 2 = Medium, 3 = High. For T, 1 = Intellectual, 2 = Production, 3 = Performance.

geneous groups, although none were significant ($z = –1.13$, $p > .05$, Fail Safe $N = 37.7$; $z = –1.2$, $p > .05$, Fail Safe $N = 1.02$; $z = –0.38$, $p > .05$, Fail Safe $N = 16.92$, respectively). These results suggest that building teams homogeneously or heterogeneously based on any of the attributes noted above will not result in significant gains in team performance.

Effects of measurement type. Focused comparisons of the combined effect sizes from the individual investigations also revealed a highly significant covariation between the type of measurement used (quality, quantity, or accuracy) and team performance, Fisher's $Z = 3.038$, $p = .000$. However, the magnitude of the effect of homogeneity on team performance was not significant across quality ($z = –1.40$, $p > .05$, Fail Safe $N = 14.07$), quantity ($z = –0.31$, $p > .05$, Fail Safe $N = 33.91$) or accuracy ($z = –1.40$, $p > .05$, Fail Safe

N = 0.06) measures. This suggests that team composition does not significantly affect the magnitude of performance whether the performance measure is quality, quantity, or accuracy.

Effects of team size. Focused comparisons of individual hypothesis tests also showed highly significant effects of the size of the team on performance, Fisher's $Z = 2.96$, $p = .001$. Although specific team sizes showed some significant differences in performance, these differences between homogeneous teams appeared to follow no regular pattern. These results are most likely due to the small number of investigations that used teams composed of more than two participants.

Effects of task difficulty. Focused comparisons of individual hypothesis tests indicated a highly significant relationship between task difficulty (low, medium, high) and the team performance effect, Fisher's $Z = 7.88$, $p = .000$. Analyses revealed that in low difficulty tasks (i.e., low stimulus uncertainty, processing demands, and response complexity), homogeneous teams performed moderately better than heterogeneous teams (Fisher's $Z = 1.85$, $p < .05$, Fail Safe $N = 52.13$). This result suggests that, for low difficulty tasks, moderate gains in performance can be expected from teams in which individual team members are of like gender, attitude, ability, and personality. In high difficulty tasks, it appears that the opposite result may be true. Heterogeneous teams performed significantly better than homogeneous teams, Fisher's $Z = –2.37$, $p < .01$, Fail Safe $N = 30.54$.

Effects of task type. The type of task performed in the individual studies was shown by focused comparisons of effect sizes to have a significant relationship to the magnitude of the team performance measure, Fisher's $Z = 11.09$, $p = .000$. No significant effect of homogeneity on team performance was found for intellectual (i.e., cognitively demanding) tasks (Fisher's $Z = –2.21$, $p = 0.99$, Fail Safe $N = 44.24$). However, for performance tasks, homogeneous teams significantly outperformed heterogeneous teams ($z = 3.30$, $p < .01$, Fail Safe $N = 51.75$).

DISCUSSION

Although intuitively it may make sense to try to build teams of either homogeneous or heterogeneous composition, the results of this meta-analysis do not show any particular attribute to be superior for matching (or differentiating) team members. Ability, personality, and gender did not show effects of homogeneity. In fact, the results of this meta-analysis suggest that on the whole, the homogeneity of a group has only a small effect on task performance. However, these findings do not necessarily negate the findings of individual studies that have found that for specific tasks, homogeneity (or heterogeneity) of groups has effects on team performance. The results of the present integration indicate that the amount and type of data available to the team are factors that should be considered when deciding whether to compose a team of similar (or dissimilar) individuals. Tasks that have low stimulus uncertainty, processing demands, and response complexity are generally performed better by homogeneous groups. The fact that this effect only occurs on low-difficulty tasks partially supports similarity theory, which posits that homogeneous teams have higher levels of performance than heterogeneous teams because of inherent mutual attraction evoked by commonalties. However, if similarities do indeed produce higher levels of performance for homogeneous teams, it is unclear why there would not be higher performance on high-difficulty tasks as well.

The results of this meta-analytic investigation also partially support equity theory, which posits that interpersonal tensions arising from differential output to reward ratios between team members create higher levels of performance in more heterogeneous teams. Heterogeneous teams performed better than homogeneous teams on both high-difficulty and performance tasks. However, once again, this difference was not seen across all levels of difficulty.

The absence of clear support for either equity theory or similarity theory suggests that the effects of homogeneity on team performance are more complex than either of these theories implies. Moreover, more research is needed to illuminate the relationship between task difficulty, task type, and homogeneous composition. However, the data from this integration do suggest that homogeneous teams will benefit from tasks that (a) are well-defined, (b) require little integration of data, and (c) require simple responses. Tasks in which limited available data require a great deal of computation and complex responses may be better suited to teams with more diverse membership. This finding lends added support to much of the research on team composition. Complex tasks defined by limited data, by definition, would require higher levels of creativity to perform. Thus, the findings of Triandis and his colleagues (1965) that teams that are heterogeneous in attitude are more creative is supported by this integration. Furthermore, the fact that these complex tasks would also require individuals to seek out all available sources of information supports Bantel's (1994) findings of a positive relationship between heterogeneity and planning openness. Heterogeneous teams working on complex tasks with limited information must focus on a wider range of options to perform their task effectively.

One caveat should be mentioned in interpreting the results of this meta-analysis. The criteria for inclusion of articles into this integration, by nature, have excluded studies that aid in our understanding of the role of team composition on performance. In particular, the rule that research be empirical necessarily excludes survey data that has significantly contributed to our understanding of how individuals of varying backgrounds interact with one another in a team setting.

Finally, although this meta-analysis focused on the link between team input variables and performance output, it is abundantly clear that these variables are mediated by many process variables. Specifically, compositional variables appear to affect the ways and means in which individuals interact within a team setting. This interaction process determines the performance output of the team. This suggests that future integrations of team composition should address the link between work-team composition and team processes related to the communication and decision strategies within teams.

CONCLUSION

The purpose of this meta-analytic review of the team performance research from 1960 to the present was to examine whether a relationship exists between team homogeneity and performance. The integration of past research does indeed show a small relationship between team homogeneity and performance for certain types of tasks (performance tasks, high and low difficulty tasks). The second purpose of this integration was to determine which specific attributes, when matched, would produce superior performance. The integration showed no clear advantage of homogeneity or heterogeneity of one particular attribute over another. It appears that any advantages are dependent on the task. The final purpose of the current integration was to investigate whether the measure used, task difficulty, type of task, and team size were predictive of the effect of homogeneity on performance. Homogeneity of composition had little or no effect on team performance in production tasks and cognitive tasks, but showed moderate increases in effectiveness on performance tasks. Composing teams of similar individuals may improve performance on low-difficulty tasks and decrease performance on high-difficulty tasks. Future investigations should address the relationship of task difficulty and task type to team composition and performance.

REFERENCES

Aamodt, M., & Kimbrough, W. (1982). Effect of group heterogeneity on quality of task solutions. *Psychological Reports*, *50*(1), 171-174.

Altman, I., & Haythorne, W. W. (1967). The effects of social isolation and group composition on performance. *Human Relations*, *20*, 313-340.

Bantel, K. A. (1994). Strategic planning openness. *Group and Organization Management*, *19*(4), 406-424.

Bantel, K. A., & Jackson, S. E. (1989). Top management and innovations in banking: Does the composition of the top team make a difference? *Strategic Management Journal*, *10*, 107-124.

Baugh, S. G., & Graen, G. B. (1997). Effects of gender and racial composition on perceptions of team performance in cross-functional teams. *Group and Organization Management*, *22*(3), 366-383.

Clement, D., & Schiereck, J. (1973). Sex composition and group performance in a visual signal detection task. *Memory & Cognition*, *1*(3), 251-255.

Dansereau, F., Yammarino, F. J., Markham, S. E., & Alluto, J. A. (1995). Individualized leadership: A new multiple-level approach. *Leadership Quarterly*, *6*(3), 413-450.

Driskell, J. E., Hogan, R., Salas, E. (1987). Personality and group performance. In C. Hendrick (Ed.), *Group processes and intergroup relations. Review of personality and social psychology* (Vol. 9, pp. 91-112). Beverly Hills, CA: Sage.

Gastil, J. (1994). A meta-analytic review of the productivity and satisfaction of democratic and autocratic leadership. *Small Group Research*, *25*(3), 384-410.

Hackman, J. R. (1983). *A normative model of work team effectiveness* (Technical Report No. 2). New Haven, CT: Yale University.

Hackman, J. R., & Morris, C. G. (1975). Group tasks, group interaction processes, and group performance effectiveness: A review and proposed integration. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 8., pp. 45-99). New York: Academic Press.

Herschel, R. (1994). The impact of varying gender composition on group brainstorming performance in a GSS environment. *Computers in Human Behavior, 10*(2), 209-222.

Hill, G. W. (1982). Group versus individual performance: Are N+1 heads better than one? *Psychological Bulletin*, *91*(3), 517-539.

Hoffman, L., & Maier, N. (1961). Quality and acceptance of problem solutions by members of homogeneous and heterogeneous groups. *Journal of Abnormal & Social Psychology*, *62*(2), 401-407.

Hoffman, R., Harburg, E., & Maier, N. (1962). Differences and disagreement as factors in creative group problem solving. *Journal of Abnormal and Social Psychology*, *64*(2), 206-214.

Hooper, S., & Hannafin, M. (1991). The effects of group composition on achievement, interaction, and learning efficiency during computer-based cooperative instruction. *Educational Technology Research and Development*, *39*(3), 27-40.

Hooper, S., Ward, T., Hannafin, M., & Clark, H. (1989). The effects of aptitude composition on achievement during small group learning. *Journal of Computer-Based Instruction*, *16*(3), 102-109.

Jackson, S. E., May, K. E., & Whitney, K. (1995) Understanding the dynamics of diversity in decision-making teams. In R. A. Guzzo & E. Salas (Eds.), *Team effectiveness and decision making in organizations* (pp. 383-396). San Francisco: Jossey-Bass.

Jacobson, M., & Effertz, J. (1974). Sex roles and leadership perceptions of the leaders and the led. *Organizational Behavior and Human Performance*, *12*, 383-396.

Kerr, N., & MacCoun, R. (1985). Role expectations in social dilemmas: Sex roles and task motivation in groups. *Journal of Personality and Social Psychology*, *49*(6), 1547-1556.

Latane', B. (1981). The psychology of social impact. *American Psychologist*, *36*, 343-356.

Lord, R. G., de Vader, C. L., Alliger, G. M. (1986) A meta-analysis of the relation between personality traits and leadership perceptions: An application of validity generalization procedures. *Journal of Applied Psychology, 71*(3), 402-410.

Morgan, B. B., & Lassiter, D. L. (1992). Team composition and staffing. In R. W. Sweezey & E. Salas (Eds.), *Teams: Their training and performance* (pp. 75-100). Norwood, NJ: Ablex.

Mullen, B. (1989). Advanced BASIC meta-analysis [Computer software]. Hillsdale, NJ: Lawrence Erlbaum.

Mullen, B., & Copper, C. (1994). The relation between group cohesiveness and performance: An integration. *Psychological Bulletin*, *115*(2), 210-227.

Mullen, B., Copper, C., Cox, P., Fraser, C., Hu, L., Meisler, A., Smith, C., & Symons, C. (1990). Boundaries around group interaction: A meta-analytic integration of the effects of group size. *Journal of Social Psychology*, *13*(2), 271-283.

Nieva, V. F., Fleischman, E. A., & Reick, A. (1978). *Team dimensions: Their identity, their measurement, and their relationships* (Contract No. DAH19-78-C-0001). Washington, DC: Advanced Research Resources Organization.

Peters, L. H., Hartke, D. D., & Pohlman, J. T. (1985). Fiedler's contingency theory of leadership: An application of the meta-analysis procedures of Schmidt and Hunter. *Psychological Bulletin*, *97*(2), 274-285.

Poole, M. S., & Roth, J. (1989). Decision development in small groups V: Test of a contingency model. *Human Communication Research*, *15*(4), 549-589.

Sanders, G. S., & Mullen, B. (1983). Accuracy in the perceptions of consensus: Differential tendencies of people with majority and minority positions. *European Journal of Social Psychology*, *13*, 57-70.

Steiner, I. D. (1972). *Group processes and productivity.* San Diego, CA: Academic Press.

Swanson, M., & Tjosvold, D. (1979). The effects of unequal competence and sex on achievement and self-presentation. *Sex Roles*, *5*(3), 279-285.

Terborg, J. R., Castore, C., & DeNinno, J. A. (1976). A longitudinal field study of the impact of group performance and cohesion. *Journal of Personality and Social Psychology*, *34*(5), 782-790.

Thoms, P., Moore, K. S., & Scott, K. S. (1996). The relationship between self-efficacy for participating in self-managed work groups and the big five personality dimensions. *Journal of Organizational Behavior, 17*, 349-362.

Triandis, H. C., Hall, E. R., & Ewen, R. B. (1965). Member heterogeneity and dyadic creativity. *Human Relations*, *18*(1), 33-35.

Tziner, A. (1985). How team composition affects task performance: Some theoretical insights. *Psychological Reports*, *57*, 1111-1119.

Tziner, A., & Eden, D. (1985) Effects of crew composition on crew performance: Does the whole equal the sum of its parts? *Journal of Applied Psychology*, *70*(1), 85-93.

Valacich, J., Wheeler, B., Menneche, B., & Wachter, R. (1995). The effects of numerical and logical group size on computer-mediated idea generation. *Organizational Behavior and Human Decision Processes*, *62*(3), 318-329.

Vancouver, J., Rubin, B., & Kerr, N. (1991). Sex composition of groups and member motivation III: Motivational losses at a feminine task. *Basic and Applied Social Psychology*, *12*(2), 133-144.

Volkema, R. J., & Gorman, R. H. (1998). The influence of cognitive based group composition on decision-making process and outcome. *Journal of Management Studies*, *35*(1), 105-121.

Weirsema, W. F., & Bantel, K. A. (1992). Top management team demography and corporate strategic change. *Academy of Management Journal*, *35*(1), 91-121.

Wood, W. (1987). Meta-analytic review of sex differences in group performance. *Psychological Bulletin*, *102*(1), 53-71.

Clint A. Bowers is an associate professor of psychology at the University of Central Florida. He is the director of the university's Team Performance Laboratory and serves as the lead scientist of the FAA/NAWCTSD/UCF Partnership for Aviation Team Training Research.

James A. Pharmer is a doctoral student in the Human Factors program at the University of Central Florida. He is currently working at the Naval Air Warfare Center Training Systems Division.

Eduardo Salas is a professor of psychology at the University of Central Florida. He also holds an appointment as principal scientist for human factors research at the university's Institute for Simulation & Training. His research interests include team training and performance, training effectiveness and evaluation, crew resource management, and group dynamics.