MILITARY PSYCHOLOGY, 8(3), 235–245

Analyzing Team Performance: In the Eye of the Beholder?

David P. Baker
InterScience America
Leesburg, Virginia

Eduardo Salas
Naval Air Warfare Center, Training Systems Division
Orlando, Florida

The effect of team-member experience was examined in relation to perceptions of importance of teamwork behaviors during an analysis of team performance. Importance ratings were collected from military aircrews (i.e., two-member teams) from three distinct types of aircraft. Results indicated that when assessing team behavior importance, less experienced team members weighted difficulty of performing team behaviors more heavily, whereas more experienced team members weighted time spent performing team behaviors more heavily. Implications for analyzing team performance, with respect to the process of conducting such analyses and the type of information collected, are discussed.

Military teams have received, as a topic of interest, an increasing amount of attention in recent years (e.g., Driskell & Salas, 1991; Dyer, 1984; Foushee, 1984, Salas, Bowers, & Cannon-Bowers, 1995). Salas, Dickinson, Converse, and Tannenbaum (1992), for instance, noted that teamwork has become a critical element of many military organizations. These researchers suggested that the emphasis on teams and teamwork is a result of the complexity of modern military tasks, which often exceed the capabilities of a single individual. As the demands for teamwork increase, the corresponding need for effective strategies for understanding team performance increases as well.

Requests for reprints should be sent to Eduardo Salas, Naval Air Warfare Center, Training Systems Division, 12350 Research Parkway, Orlando, FL 32826.

A vast proportion of the research on teams has focused on military command and control teams and aircrews (Foushee, 1984; Glickman et al., 1987; Morgan, Glickman, Woodward, Blaiwes, & Salas, 1986; Oser, McCallum, Salas, & Morgan, 1990; Stout, Cannon-Bowers, Salas, & Morgan, 1990). These efforts attempted to establish underlying constructs and behaviors that influence team performance. For example, Morgan et al. (1986) hypothesized that there were two general categories of behavior that could be distinguished in a team: a taskwork track and a teamwork track. In this view, taskwork consists of behaviors that are performed by individual team members and are critical to the execution of individual team member functions, whereas teamwork consists of behaviors that are related to team member interaction and are necessary to establish coordination among individual team members in order to achieve team goals.

In a comprehensive examination of the teamwork track, Morgan et al. (1986), studying Naval Gunfire Support teams, identified a number of critical team behaviors that were believed to be organized around seven skill dimensions (i.e., giving suggestions or criticism, cooperation, communication, team spirit and morale, adaptability, coordination, and acceptance of suggestions or criticism). Since then, these skill dimensions and their associated behaviors have been generally confirmed. First, Oser et al. (1990) showed that several of these behaviors were related to team functioning and task outcomes. Next, Stout et al. (1990) demonstrated that team process variables, similar to those delineated by Morgan et al., directly influenced team performance. Last, McIntyre and Salas (1995) found team behavior to be fairly consistent across different task types.

Much of the research on military teams has relied on task analysis methodologies to analyze team performance. Typically, task analysis produces a description of the work performed on the job. This description may outline the tasks performed; the knowledges, skills, and abilities necessary to perform those tasks; or both (Goldstein, 1986). With respect to military teams, task analysis has been a useful tool in delineating behaviors that define effective team performance (behaviors are similar to tasks in the sense that they describe the process by which teamwork is achieved and maintained). For example, in the investigation of teamwork by Morgan et al. (1986), the critical incident technique was employed to extract initial lists of critical team behaviors. Likewise, Prince and Salas (1993) employed the critical incident methodology to identify team behaviors associated with effective performance in military crews. Last, Stout, Prince, Baker, Bergondy, and Salas (1992) employed a series of task analysis rating scales (e.g., difficulty of learning, criticality of error, overall importance, etc.) to determine the relative importance of various team behaviors associated with effective team performance for several naval aviation teams.

Although task analysis techniques have been employed in a number of civilian and military settings, the validity of various task analysis approaches for analyzing

team performance has come into question. As Bowers, Morgan, Salas, and Prince (1993) pointed out, there is currently little research on the effectiveness of various methodologies for understanding teamwork, and questions remain regarding the generalizability of individual task analysis strategies to the team level. For example, Bowers et al. (1993) found that expert team members could provide valid data on teamwork at a general dimension level but had difficulty providing accurate information about the specific team behaviors that defined these dimensions. More recently, Bowers, Baker, and Salas (1994) found that metrics used for establishing individual task importance did not generalize to the team level. These researchers called for new metrics for establishing team behavior importance and recommended a metric that consisted of a weighted combination of task criticality and importance for training when trying to identify critical team training needs.

Other researchers suggested that characteristics of the rater may have an impact on team performance analysis results (Baker, Salas, & Prince, 1991; Conley & Sackett, 1987; Landy & Vasey, 1991; Levine & Dickey, 1990; Sanchez, 1990; Sanchez & Levine, 1989). For example, Sanchez, in a study of individual-level tasks, suggested that subject matter expert perceptions of a task may not simply be a function of objective task characteristics. Idiosyncratic factors associated with a rater may have a significant effect on the degree to which a rater perceives the task to be important. Sanchez tested this hypothesis by having job incumbents from two jobs independently evaluate their respective task inventories on three dimensions (i.e., difficulty of learning, time spent performing the task, and overall task importance). For one of the two jobs studied, the results showed that time spent performing the task was more salient in evaluating overall task importance as job experience increased, whereas the importance of difficulty of learning the task decreased with job experience.

Other researchers also found an effect for experience on ratings of individual tasks. Ford, Smith, Sego, and Quinones (1992) found that Air Force Ground Equipment Mechanics, whose self-efficacy and breadth of experience (i.e., number of tasks an individual has performed) increased as a function of job tenure, showed an increase in training emphasis ratings. Ford et al. concluded that these results have significant implications for the practice of training needs assessment because they indicate that different experience levels are likely to produce different ratings.

As noted previously, the research cited suggests that characteristics of the rater may affect perceptions of team behavior importance. This is particularly important in the context of studying teams, because team research, especially research that focused on military teams, relied heavily on subject matter expert testimony and ratings with respect to critical team behaviors that define team performance (Foushee, 1984; Glickman et al., 1987; Morgan et al., 1986; Oser et al., 1990; Stout et al., 1990). However, these data may be unreliable because the results noted earlier imply that team members may define successful team performance differently

depending on the experience level a team member has with a team behavior. Such an effect would have significant implications for both studying teamwork and the approach by which team performance is analyzed.

The purpose of this study was to investigate the effects of experience on perceptions of team behavior importance. The effects of experience have been documented for individual tasks (Ford et al., 1992; Sanchez, 1990), but research has yet to investigate the extent to which team member experience affects perceptions of team behavior importance. Baker et al. (1991) pointed out that individual task analysis methodologies and research may not readily generalize to teams due to the critical interactional requirements of teamwork, which must be assessed. Therefore, this research replicated and extended results for the effects of experience on perceptions of importance of individual-level tasks (Sanchez, 1990) to the team level.

In this investigation, ratings were collected from experienced and inexperienced team members from three distinct military aviation communities on behaviors that comprise effective team performance. Military aviation teams have been studied extensively in the area of teamwork (Oser et al., 1990; Prince & Salas, 1993; Stout et al., 1990; Stout et al., 1992), and the teams studied in this investigation were deemed to conform to the characteristics (e.g., interdependency among team members, complex communication channels, etc.) proposed to define a team (Dyer, 1984; Morgan et al., 1986; Salas et al., 1992). Furthermore, collecting data on three distinct types of teams allowed us to determine the extent to which effects of experience are stable across different teams and team behaviors.

Based on Sanchez (1990), it was hypothesized that experience would moderate the extent to which team members weighted different team behavior characteristics when making judgments of overall team behavior importance. More specifically, we formulated two hypotheses:

H1: Low-experience team members would weight difficulty of learning a team behavior more heavily than would high-experience team members when making judgments regarding overall team behavior importance.

H2: High-experience team members would weight time spent performing a team behavior more heavily than would low-experience team members when making judgments regarding overall team behavior importance.

In addition, we collected data on other task characteristics such as criticality of error, difficulty, and importance for training. These ratings are often collected during task and training needs analysis (Bowers et al., 1994; Goldstein, 1986; Levine, 1983; Sanchez & Levine, 1989), and by collecting these ratings, we were able to examine more thoroughly the effects of experience on rater perceptions of team behavior importance.

METHOD

Participants

Military pilots from three aviation communities (i.e., training, fixed-wing attack, and cargo helicopter) filled out questionnaires that included behaviors associated with achieving effective teamwork in the cockpit. In each case, pilots fly their respective aircraft in two-member teams (i.e., aircrews), and questionnaires were developed to reflect the specific teamwork requirements associated with each aircrew. Respondents included 38 training community pilots with an average experience level of 5.56 years, 20 fixed-wing attack pilots with an average experience level of 6.75 years, and 46 cargo helicopter pilots with an average experience level of 7.60 years. Furthermore, in the training community, 19 respondents were instructor pilots, and 19 respondents were student pilots.

Teamwork Inventories

Teamwork inventories for each type of aviation team were developed in the same fashion. Initially, behaviors were identified through a review of the aircrew coordination and team training literatures (see Prince & Salas, 1993, for a detailed description). This procedure resulted in a list of more than 40 generic team behaviors that were believed to be associated with effective teamwork in the cockpit. This list was then tailored to each air community through a series of critical incident interviews (Flanagan, 1954) with pilots from each aircraft. Behaviors were deleted, rewritten, and added as deemed appropriate. This procedure resulted in 28 behaviors associated with team performance in the training community, 56 behaviors associated with team performance in the fixed-wing attack community, and 42 behaviors associated with team performance in the cargo helicopter community. In general, inventories included such team behaviors as "cross check information sources for agreement/confirmation," "provide assistance to other crew members as needed," and "inform crew members of mission progress."

Team Behavior Dimensions Measured

Five team behavior dimensions commonly used in task and team performance analysis were rated in this investigation (Baker et al., 1991; Bowers et al., 1994; Levine, 1983; Sanchez & Levine, 1989):

1. *Criticality of Error:* The criticality of error associated with performing the behavior incorrectly.
2. *Difficulty:* The difficulty of performing the behavior.
3. *Time Spent:* The time spent performing the behavior.

4. *Difficulty of Learning:* The difficulty of learning how to perform the behavior correctly.
5. *Importance for Training:* The degree to which the behavior should be addressed through training.

In addition, a global rating of overall behavioral importance (i.e., *Overall Importance:* the importance of a team behavior for effective team performance relative to all the other behaviors) was collected for each behavior. All five behavioral dimensions and overall behavioral importance were evaluated using 7-point relative rating scales, ranging from 1 *(low)* to 7 *(high).*

Procedure

For each team, the team performance inventory was distributed to team members along with the six rating scales and a description of each scale. Team members completed the rating process individually by rating each team behavior in their team performance inventory (i.e., training, fixed-wing attack, or cargo helicopter squadron) on the six rating scales. Data were collected individually, rather than by teams, in order to assess each individual team member's perception of team behavior importance.

RESULTS

Data were analyzed in a similar fashion for each team member. First, correlations were calculated individually for each team member between each of the five behavioral dimensions (i.e., criticality of error, difficulty, time spent, difficulty of learning, and importance for training) and the overall importance rating. These correlations provided an indication of the weight each team member placed on each behavioral dimension in rating the overall importance of a team behavior. Next, these correlations were transformed to Fisher *z* scores and arrayed as vectors for each team member. This procedure resulted in five transformed correlations for each rater. The transformed correlations Criticality of Error × Overall Importance (CRTXOAR), Difficulty × Overall Importance (DIFXOAR), Time Spent × Overall Importance (RTSXOAR), Difficulty of Learning × Overall Importance (LRNXOAR), Importance for Training × Overall Importance (TRNXOAR) served as the dependent variables in the analysis. Job experience served as the independent variable across analyses. For the training teams, job experience was operationalized as whether a team member was an instructor pilot or a student pilot. For the fixed-wing attack and cargo helicopter teams, job experience was operationalized as the number of months a team member had been a pilot. Team members were then assigned into

high- and low-experience groups on the basis of a median split performed on the experience data from each squadron.

A multivariate analysis of variance was performed to test the effects of experience on the five transformed correlations: CRTXOAR, DIFXOAR, RTSXOAR, LRNXOAR, and TRNXOAR for each air platform. For both the training and cargo helicopter teams, experience was found to have a significant effect on perceptions of team behavior importance, Wilks's $\Lambda = 0.45$, $F(5, 32) = 7.93$, $p < .05$, and Wilks's $\Lambda = 0.71$, $F(5, 40) = 3.30$, $p < .05$, respectively. Results for fixed-wing attack teams were nonsignificant.

Univariate analyses were performed on each dependent measure to explore further the effects of experience in the training and cargo helicopter teams. With respect to the training teams, this analysis resulted in a significant effect for experience on the following dependent variables: CRTXOAR, $F(1, 36) = 31.90$, $p < .01$; DIFXOAR, $F(1, 36) = 16.38$, $p < .01$; TRNXOAR, $F(1, 36) = 35.35$, $p < .01$; and LRNXOAR, $F(1, 36) = 7.55$, $p < .01$. Inspection of the means (refer to Table 1) showed that low-experience team members (i.e., student pilots) weighted the dimensions of criticality of error, difficulty, importance for training, and difficulty of learning more heavily than did high-experience team members (i.e., instructor pilots) when evaluating overall team behavior importance. With respect to the cargo helicopter teams, this analysis resulted in a significant effect for experience on the following dependent variables: DIFXOAR, $F(1, 44) = 7.03$, $p < .01$; and RTSXOAR, $F(1, 44) = 4.33$, $p < .05$. Inspection of these means (refer to Table 1) showed that

TABLE 1
Mean Transformed Correlations (z Scores) as a Function of Job Experience for the Three Aviation Teams

Squadron	*CRTXOAR*	*DIFXOAR*	*RTSXOAR*	*TRNXOAR*	*LRNXOAR*
Training[a]					
Students	$.77_a$	$.61_b$	.35	$.77_c$	$.31_d$
Instructors	$.04_a$	$.16_b$	.30	$.04_c$	$-.08_d$
Fixed-wing attack[b]					
Low experience	.85	.33	.38	.55	.49
High experience	.87	.20	.40	.58	.52
Cargo helicopter[c]					
Low experience	.79	$.21_e$	$.21_f$	.58	.16
High experience	.75	$.01_e$	$.32_f$	.58	.19

Note. CRTXOAR = Criticality of Error × Overall Importance; DIFXOAR = Difficulty × Overall Importance; RTSXOAR = Time Spent × Overall Importance; TRNXOAR = Importance for Training × Overall Importance; LRNXOAR = Difficulty of Learning × Overall Importance. Transformed correlations (i.e., z scores) with the same subscript letters (a, b, c, etc.) indicate a significant difference between those z scores at $p < .05$. For example, the z scores for the training squadron of $.77_a$ and $.04_a$ for the dependent variable CRTXOAR were found to be significantly different, $p < .05$.

[a] $n = 38$. [b] $n = 20$. [c] $n = 46$.

when evaluating overall team behavior importance, low-experience team members weighted the dimension of difficulty more heavily than did high-experience team members, whereas high-experience team members weighted the dimension of time spent more heavily than did low-experience team members.

DISCUSSION

This study examined the effects of experience on the degree to which various team behavior characteristics were weighted in a team member's perception of team behavior importance. In general, the results supported the hypothesis that experience would moderate the extent to which team members weighted different team behavior characteristics when making judgments of overall team behavior importance. For the training teams, low-experience team members were found to place a higher emphasis relative to high-experience team members on the criticality of error, difficulty, importance to train, and difficulty of learning a team behavior. For the cargo helicopter teams, low-experience team members were found to place a greater emphasis on difficulty of performing a team behavior, whereas high-experience team members were found to place a greater emphasis on the time spent performing a team behavior. Results for fixed-wing attack teams were nonsignificant. However, the fixed-wing attack analyses were based on the smallest sample size (N = 20), and inspection of the means in Table 1 suggests trends that are consistent with results for the cargo helicopter teams for the team behavior characteristics of difficulty and time spent. In summary, the results of this study suggest that team members may define the characteristics of teamwork differently as a function of individual differences and team experience. In other words, when conducting an analysis of team performance, the importance of teamwork and its corresponding behavioral characteristics are truly a function of the eye of the beholder.

Although the results of this study were inconsistent across the three aviation teams, one trend did emerge. That is, for all three aviation platforms, low-experience team members placed a greater emphasis on difficulty of performing a team behavior relative to high-experience team members when rendering judgments regarding the overall importance of a team behavior (refer to Table 1). Therefore, at least in aviation teams, difficulty of performing a team behavior may be a very salient characteristic for inexperienced team members when defining effective team performance, as opposed to difficulty of learning (i.e., as posited by the first hypothesis). These results are different from the results of Sanchez (1990) and lend support to the notion that results from research on individual-level tasks may not generalize to the team level (Baker et. al., 1991; Bowers et al., 1994). Furthermore, the behaviors that define effective team performance may change as team members gain team task experience (i.e., for cargo helicopter teams, emphasis on time spent

performing the team behavior increased with team member experience, which provided support for the second hypothesis). These results lend additional evidence to the proposition that teams evolve and mature over time (McIntyre & Salas, 1995). Furthermore, these results provide evidence regarding the complexity of team behavior, which has been highlighted by a number of researchers (Baker & Salas, 1992; Dyer, 1984; Salas et al., 1992).

The preceding results have practical implications for the process of conducting military team performance analysis and for the type of information included in team training programs that seek to establish and maintain modern military effectiveness. With respect to analyzing team performance, these results suggest that team-member experience may moderate the nature of information collected and the extent to which team behaviors are deemed important. Inexperienced team members may identify one set of behaviors that are important for effective team performance, and experienced team members may identify another. Therefore, researchers should sample team members from a variety of experience levels to ensure that all critical team behaviors are included in the team performance analysis. Based on these data and other research results (Ford et al., 1992; Sanchez, 1990), it may be prudent to include sampling requirements for subject matter experts in future team performance analysis guidelines.

With respect to team training, these results suggest that different team behaviors should be the target of training, depending on the experience level of the trainees. For example, initial training for inexperienced team members should focus on those behaviors that are difficult to perform. As team members gain experience and master difficult team behaviors, recurrency training should target team behaviors that occur frequently. In other words, team training should be designed and structured around the maturity of the team.

The research presented here provides some interesting findings on the effects of experience. Future research should continue to examine the impact of individual characteristics on perceptions of team effectiveness and should identify specific policies team members use to define effective teamwork. Such research would contribute to understanding the shared requirements team members have for effective team performance, as well as how these criteria change as the team matures.

ACKNOWLEDGMENTS

An earlier version of this article was presented at the annual meeting of the Society for Industrial and Organizational Psychology, Nashville, TN, April 1991. The views expressed in this article are those of the authors and do not reflect the official position of the Department of the Navy, Department of Defense, or the U.S. Government.

REFERENCES

Baker, D. P., & Salas, E. (1992). Principles for measuring teamwork skills. *Human Factors, 34,* 469–475.

Baker, D. P., Salas, E., & Prince, C. (1991, April). *Team task importance: Implications for conducting team task analysis.* Paper presented at the sixth annual meeting of the Society for Industrial and Organizational Psychology, St. Louis, MO.

Bowers, C. A., Baker, D. P., & Salas, E. (1994). The importance of teamwork in the cockpit: The utility of job/task analysis indices for training design. *Military Psychology, 6,* 205–214.

Bowers, C. A., Morgan, B. B., Salas, E., & Prince, C. (1993). Assessment of coordination demand for aircrew coordination training. *Military Psychology, 5,* 95–112.

Conley, P. R., & Sackett, P. (1987). Effects of using high versus low performing job incumbents as sources of job analysis information. *Journal of Applied Psychology, 72,* 434–437.

Driskell, J. E., & Salas, E. (1991). Group decision making under stress. *Journal of Applied Psychology, 76,* 473–478.

Dyer, J. L. (1984). Team research and training: A state of the art review. In F. A. Muckler (Ed.), *Human factors review: 1984* (pp. 285–323). Santa Monica, CA: Human Factors Society.

Flanagan, J. C. (1954). The critical incident technique. *Psychological Bulletin, 51,* 327–358.

Ford, J. K., Smith, E. M., Sego, D. J., & Quinones, M. A. (1992, April). *The impact of individual and task experience factors on training emphasis ratings.* Paper presented at the seventh annual meeting of the Society for Industrial and Organizational Psychology, Montreal, Canada.

Foushee, H. C. (1984). Dyads and tryads at 35,000 feet: Factors affecting group process and aircrew performance. *American Psychologist, 39,* 885–893.

Glickman, A. S., Zimmer, S., Montero, R. C., Guerette, P. J., Campbell, W. J., Morgan, B. B., & Salas, E., (1987). *The evolution of teamwork skills: An empirical assessment with implications for training* (Tech. Rep. No. 87–016). Orlando, FL: Naval Training Systems Center, Human Factors Division.

Goldstein, I. (1986). *Training in organizations: Needs assessment, development, and evaluation* (2nd ed.). Monterey, CA: Brooks/Cole.

Landy, F. L., & Vasey, J. (1991). Job analysis: The composition of SME samples. *Personnel Psychology, 34,* 27–50.

Levine, E. L. (1983). *Everything you always wanted to know about job analysis.* Tampa, FL: Mariner.

Levine, E. L., & Dickey, M. T. (1990, August). *Measuring task importance: A replication and extension.* Paper presented at the 98th annual meeting of the American Psychological Association, Boston, MA.

McIntyre, R. M., & Salas, E. (1995). Team performance in complex environments: What we have learned so far. In R. Guzzo & E. Salas (Eds.), *Team effectiveness and decision making in organizations* (pp. 9–45). San Francisco: Jossey-Bass.

Morgan, B. B., Glickman, A. S., Woodward, E. A., Blaiwes, A. S., & Salas, E. (1986). *Measurement of team behaviors in a Navy environment* (Tech. Rep. No. NTSC TR–86–014). Orlando, FL: Naval Training Systems Center.

Oser, R. L., McCallum, G. A., Salas, E., & Morgan, B. B., (1990). *Toward a definition of teamwork: An analysis of critical team behaviors* (Tech. Rep. No. 89–004). Orlando, FL: Naval Training Systems Center, Human Factors Division.

Prince, C., & Salas, E. (1993). Training and research for teamwork in the military aircrew. In E. L. Weiner, B. G. Kanki, & R. L. Helmreich (Eds.), *Cockpit resource management* (pp. 337–366). Orlando, FL: Academic.

Salas, E., Bowers, C. A., & Cannon-Bowers, J. A. (1995). Military team research: 10 years of progress. *Military Psychology, 7,* 55–75.

Salas, E., Dickinson, T. L., Converse, S., & Tannenbaum, S. I. (1992). Toward an understanding of team performance and training. In R. W. Swezey & E. Salas (Eds.), *Teams: Their training and performance* (pp. 3–29). Norwood, NJ: Ablex.

Sanchez, J. I. (1990, May). *The effects of job experience on judgments of task importance.* Paper presented at the fifth annual meeting for the Society for Industrial and Organizational Psychology, Miami, FL.

Sanchez, J. I., & Levine, E. L. (1989). Determining important tasks within jobs: A policy capturing approach. *Journal of Applied Psychology, 74,* 336–342.

Stout, R., Cannon-Bowers, J. A., Salas, E., & Morgan, B. B. (1990). Does crew coordination impact performance? *Proceedings of the 34th Annual Meeting of the Human Factors Society,* 1382–1386.

Stout, R., Prince, C., Baker, D. P., Bergondy, M. L., & Salas, E. (1992). Aircrew coordination: What does it take? *Proceedings of the 13th Psychology in the Department of Defense Symposium,* 133–137.