MILITARY PSYCHOLOGY, *6*(4), 205–214

Measuring the Importance of Teamwork: The Reliability and Validity of Job/Task Analysis Indices for Team-Training Design

Clint A. Bowers, David P. Baker, and Eduardo Salas
Naval Air Warfare Center Training Systems Division
Orlando, Florida

Training interventions designed to improve coordination and communication in the cockpit increasingly emphasize the teaching of specific coordination skills. However, there is little guidance in the literature regarding the manner by which these skills should be selected. This investigation compared a variety of task-importance indices used previously with individual tasks in predicting the overall importance of team tasks. All of the indices demonstrated relatively poor reliability. Composite indices, including one newly derived index, demonstrated greater validity. The results are discussed in terms of implications for future research and for team-training design.

Modern military effectiveness is often dependent on the ability of individuals to coordinate their activities in order to perform as a team. Several researchers have emphasized the need to understand, and eventually to train, appropriate teamwork skills (e.g., Dyer, 1984; Salas, Dickinson, Converse, & Tannenbaum, 1992). This need is especially apparent in the aviation domain, which requires individuals to come together quickly as a team and perform under circumstances that are frequently stressful. Unfortunately, history has demonstrated that aircrewmen do not automatically combine to form good teams. Crew coordination failures are frequently cited as causal factors in military aviation accidents (Prince & Salas, in press). The history of mishaps due to faulty teamwork suggests that there is an urgent need to understand the nature of team processes in military aviation in order to develop effective team-training programs.

Requests for reprints should be sent to Eduardo Salas, Naval Air Warfare Center Training Systems Division, Code 262, 12350 Research Parkway, Orlando, FL 32826–3224.

Scientists have responded to the need for team training in civilian aviation by developing a variety of Cockpit Resource Management or Aircrew Coordination Training programs. In general, these programs attempt to facilitate crew coordination by improving crew members' attitudes toward coordination (Helmreich, Foushee, Benson, & Russini, 1985). Although it appears that these programs do enhance attitudes toward coordination (Helmreich, 1991), their effectiveness in improving flight performance is dependent on three key assumptions: (a) Crew members already possess all necessary teamwork skills within their behavioral repertoire, (b) crew members have the ability to utilize these skills to cope with task demands, and (c) failure to employ coordination skills in coping with flight demands is a function of insufficient motivation (i.e., negative attitudes).

Recently, investigators suggested that attitude-based programs, and the assumptions encompassed therein, might not be optimal for military crews (Prince, Chidester, Bowers, & Cannon-Bowers, 1992; Prince & Salas, 1993). Rather, these scientists suggested that crew coordination is best accomplished by training the specific skills that comprise coordination and by providing opportunities to practice these behaviors. Laboratory studies have supported the effectiveness of this type of skill-based training (e.g., Smith & Salas, 1991), but its application in field settings poses a number of practical problems for training development. For example, in a recent article, Bowers, Morgan, Salas, and Prince (1993) pointed out that there is currently no methodology to determine which coordination behaviors are required in coping with the demands of various flight tasks.

One approach to identifying important team tasks for training was described by Bowers et al. (1993), using a Coordination Demand Questionnaire designed to assess the degree to which each behavioral dimension is required in executing a sample of flight tasks. The results indicated that pilots' self-reports appeared to be a valid method of obtaining coordination demand data. However, although existing methods might be useful for assessing the general dimension-level requirements associated with various flight tasks, they are not effective in providing information about the specific coordination behaviors that should be incorporated into aircrew coordination training for any given platform. Because the development of skill-based coordination training is dependent on the accurate identification of crucial behaviors to be trained, and because there is a need to develop more effective methods to identify these skills, the purpose of this investigation was to identify a psychometrically sound index of team task importance to guide the selection of behaviors for training.

TASK-IMPORTANCE INDICES AND TEAMWORK

Estimates of task importance are required in any thorough job analysis that serves as the foundation for the design of training programs and perfor-

mance-appraisal systems (Levine, 1983; Levine, Sistrunk, McNutt, & Gael, 1986). Consequently, psychologists have endeavored to establish reliable and valid measures of task importance. For example, Sanchez and Levine (1989) conducted a policy-capturing study of four different jobs. Job incumbents rated their respective tasks on five dimensions—task difficulty, task criticality, relative time spent, difficulty of learning, and task responsibility—as well as on an overall criterion of task importance. The results indicated that a simple linear combination of two of these ratings (task criticality and difficulty of learning) resulted in the most psychometrically sound index of task importance. This pattern was replicated by Levine and Dickey (1990), using a different set of jobs.

Despite the interest in task analysis for the jobs that individuals perform, there is relatively little guidance in the literature regarding how best to conduct such an analysis for teams. Analyses of team tasks have thus far relied on task-importance indices developed for individual tasks (e.g., Levine & Baker, 1990). However, research conducted by Glickman et al. (1987) indicated that team performance includes two distinct dimensions of behavior: taskwork (i.e., behaviors required in the execution of individual subtasks) and teamwork (i.e., behaviors required for cooperative functioning). Task-analysis strategies at the individual level have, by their design, focused on taskwork with little consideration given to teamwork skills. In fact, there are few compelling data to suggest that existing strategies are effective in evaluating tasks that require teamwork. It is for this reason that articles in the team-performance literature have cautioned that valid measurement instruments for individual tasks are not necessarily valid tools for team tasks (e.g., Dyer, 1984). Task-importance indices seem especially prone to this type of inaccuracy. For example, the task-importance index suggested by Sanchez and Levine (1989) includes difficulty of learning in creating its estimate of task importance. However, teamwork tasks are typically not difficult to learn and may provide an extremely limited range of responses when rated on this dimension. Thus, the best estimate of the importance of teamwork tasks might require the derivation of a new variable especially suited to this purpose.

This investigation sought to evaluate the utility of task-importance indices typically used in analyses of individual-level tasks for assessing the importance of teamwork tasks in military aviation. Specifically, the research assessed the reliability and validity for five commonly used task-importance indices. Furthermore, a new index of team task importance was developed using a policy capturing technique similar to that described by Sanchez and Levine (1989). The psychometric data for these variables were evaluated for three military aircraft: the MH-53 cargo helicopter, the A-6 attack aircraft, and the F-14 fighter aircraft. These aircraft were chosen because they have extremely different flight characteristics and mission demands, thereby allowing the assessment of the degree to which the adequacy of these indices generalizes across team tasks.

METHOD

Subjects

A total of 113, active-duty military pilots served as subjects in this investigation (46 helicopter, 33 attack, and 34 fighter). Across aircraft, the average age of subjects was 28.2 years. The number of years of flight experience ranged from 1 to 16 years with a mean of 5.6 years.

Team Task Inventory

A team task inventory was developed for each of the three aircraft. Each inventory was tailored to the specific community via the following process:

1. Teamwork behaviors were identified from the team performance and aircrew coordination literature.
2. A number of teamwork behaviors were selected by the researchers as potentially important for military aviation. These included behaviors such as "Coordinates gathering of required information in an effective manner" and "Verbalizes plans for flight procedures and maneuvers."
3. The resulting inventory was reviewed by three pilots from each community to assess the appropriateness of each item for each platform.
4. Based on feedback from the pilots, items were modified or removed as necessary. This resulted in a total of 42 team tasks for the helicopter sample and 56 items for each of the fixed-wing samples (see Franz, Prince, Cannon-Bowers, & Salas, 1990, for a more thorough description of the task-selection process).

Task Dimensions Measured

Four task dimensions were rated in this investigation: importance to train (i.e., the benefit of dedicating training time to a task relative to all other tasks in the job), task criticality (i.e., the degree to which failure in the task causes negative consequences), task frequency (i.e., the number of times the task must be performed relative to other tasks within the job), and overall task importance. These dimensions were included because they have been used in previous studies of individual-level task importance and appear to have reasonable reliability (Sanchez & Levine, 1989). Additional data were collected to allow computation of two composite indices suggested by previous investigators (i.e., task difficulty for the Levine, 1983, index and difficulty of learning to compute the Sanchez & Levine, 1989, index). All of the task dimensions were rated using a 7-point, relative rating scale format where 1 represented the lowest end of the scale (i.e., not important, not critical, etc.) and 7 was the highest rating.

Procedure

The Team Task Inventory and a description of each rating scale were distributed to pilots during regularly scheduled squadron meetings. Pilots were asked to rate all the tasks, using each of the four rating scales. Completing the Team Task Inventory required approximately 45 min.

RESULTS

Five measures of task importance were calculated for each subject. These included (a) Levine's (1983) method of multiplying criticality of error and task difficulty and then adding relative time spent; (b) Sanchez and Levine's (1989) method of summing task criticality and difficulty of learning; (c) the new Team Task-Importance Index (TTII), which includes task criticality and importance to train; (d) task frequency (i.e., relative time spent); and (e) overall task importance.

Derivation of the New Composite

Linear models of the MH-53 helicopter sample were computed, using stepwise multiple regression. The helicopter sample was chosen because the larger sample size provided the greatest opportunity to maximize the statistical power of the regression. Importance to train, task criticality, task frequency, task difficulty, and difficulty of learning were included in the pool of possible predictors with overall task importance as the dependent variable. Interactions were not considered in the analysis because, historically, they have predicted very little unique variance (e.g., Sanchez & Levine, 1989; Valenzi & Andrews, 1973). The results revealed that only criticality and importance to train were significant predictors in the equation predicting overall importance. The resulting regression equation was used to compute a new composite variable for all three aircraft in an application of bootstrapping consistent with that described by Dawes (1971, 1982). The resulting composite score, the TTII, was computed by using the formula:

$$\text{TTII} = (.53 \times \text{Criticality}) + (.47 \times \text{Importance})$$

The results of the regression indicated that criticality explained 73% of the variance in overall importance. The addition of importance to train raised this total to 79%.

Table 1 presents the correlations among dimensions. These correlations were computed after determining the average task ratings across judges for each of the task dimensions and for each of the tasks. The intercorrelations were computed using the resulting dimension averages. It is interesting to

TABLE 1
Intercorrelations Among Task-Importance Indices (Across Aircraft)

Index	*Levine*	*Sanchez & Levine*	*Overall Importance*	*Importance to Train*	*Frequency*
Sanchez & Levine	.99**				
Overall Importance	−.19*	−.24*			
Importance to Train	.21*	.18	.79**		
Frequency	−.08	−.11	.60**	.67**	
Criticality	−.02	−.02	.57**	.60**	.34*

*$p = .05$. **$p = .01$.

note the large correlation between the original Levine (1983) and Sanchez and Levine (1989) indices. This might be due to the fact that each of these indices relies heavily on elements of task difficulty, which is typically rated as quite low for teamwork behaviors.

Interrater Agreement

Interrater agreement was then calculated for each of the task-importance indices by computing the correlation for each pair of judges, transforming these scores to Fisher's z scores, and computing the average across judges. These correlations were then compared on a pairwise basis using t tests for dependent correlations.

Table 2 presents the results of the interrater-agreement analysis. As can be seen in Table 2, there were no significant differences among the indices. The Levine (1983) index was among the indices with highest reliability coefficients for two of the three aircraft.

Validity

The validity of the task-importance measures was evaluated by comparing the convergence of each task-importance index with the average overall importance rating on each task and then testing the difference between these correlations using dependent t tests. This criterion was calculated by averaging the overall importance for judges, excluding the judge under consideration. This approach is consistent with Levine and Dickey's (1990) assertion that, in the absence of an ultimate criterion, average overall task importance should be used because it is assumed that the individual biases in rating importance are overcome by aggregating across judges.

Table 3 presents the results of the validity analysis. As illustrated by Table 3, frequency demonstrated significantly lower levels of reliability for two of the three aircraft. There were no significant differences among the remaining

TABLE 2
Estimates of Interrater Agreement for the Task-Importance Indices (Presented as Correlations)

Aircraft	*Frequency*	*Criticality*	*Overall*	*Sanchez & Levine*	*Levine*	*TTII*[d]
MH-53[a]	.28	.16	.22	.31	.33	.20
A-6[b]	.08	.14	.14	.14	.11	.13
F-14[c]	.15	.11	.12	.22	.31	.14

[a]$n = 46$. [b]$n = 33$. [c]$n = .34$. [d]TTII = Team Task-Importance Index.

TABLE 3
Correlations of Task-Importance Indices With Averaged Ratings of Overall Task Importance

Aircraft	*Frequency*	*Criticality*	*Overall*	*Sanchez & Levine*	*Levine*	*TTII*[d]
MH-53[a]	$.08_a$	$.29_b$	$.37_b$	$.30_b$	$.29_b$	$.38_b$
A-6_b	$.11_a$	$.16_a$	$.16_a$	$.15_a$	$.10_a$	$.21_a$
F-14_c	$.13_a$	$.19_{ab}$	$.30_b$	$.30_b$	$.29_b$	$.35_b$

Note. Within aircraft cells with identical subscripts are not significantly different ($p < .05$).
[a]$n = 46$. [b]$n = 33$. [c]$n = .34$. [d]TTII = Team Task-Importance Index.

indices. Of the composite indices, the TTII demonstrated the highest validity coefficients, but this difference did not achieve statistical significance.

DISCUSSION

The training of effective team processes is quickly becoming a requirement of military aviation. The ability to determine the specific behaviors that comprise teamwork is a prerequisite to developing effective team training. Identifying these skills will allow the development of specific skill-based training interventions, as well as the creation of practice opportunities (e.g., role playing, simulated flights) for these skills. In the absence of a behavioral understanding of teamwork, training will proceed in an idiosyncratic fashion.

This investigation attempted to assess the degree to which task-importance indices developed for use in individual tasks are appropriate for use in team tasks. In essence, this effort represents a replication and extension of Sanchez and Levine's (1989) effort. However, the present emphasis on team performance represents a new application of this method. The emphasis on team tasks is likely to have practical implications for military training. Training irrelevant team skills is not cost effective. However, failure to train important teamwork skills is likely to result in ineffective performance. In other words, this study empirically assessed the degree to which existing task-importance indices are useful in identifying behaviors for use in team-training applications.

As in Sanchez and Levine (1989), the present results indicate that there are relatively large correlations among the indices of task importance. Therefore, all the indices appear to measure some aspects of a shared construct. Also consistent with Sanchez and Levine, the present results indicate that job incumbents' ratings of a task's overall importance were best predicted by a combination of relatively few variables, one of which was task criticality. However, the present results indicate that task criticality is best augmented by estimates of importance to train when rating team tasks. This is consistent with our assertion that difficulty of training might not offer sufficient variance when rating team tasks. Furthermore, importance to train is likely to be more accessible to military pilots, because the absence of these behaviors is frequently associated with the causes of aviation mishaps.

In evaluating the psychometric qualities of these measures, one is struck by the generally low reliability coefficients associated with all the indices. This tendency serves to highlight the fact that ratings of team task importance are difficult for job incumbents to agree on. The validity data yielded a similar pattern. In contrast to the suggestion of Levine and Baker (1990), previous composite indices were not associated with highest convergence with average estimates of overall importance for the majority of aircraft. The TTII, which uses task criticality and importance to train, demonstrated the strongest validity. Again, however, the coefficients were generally very low.

In a larger context, the present results serve to support Dyer's (1984) concern regarding the use of measurement instruments derived from individuals in team-performance research. In the case of military aviation, none of the traditional indices of task importance yielded particularly strong reliability or validity. Furthermore, a new index designed to optimize the predictability of these indices for team tasks did not result in significantly stronger prediction of task importance. Thus, the present results point out the need to identify other task dimensions that incumbents might use in evaluating team task importance. Results of the regression analysis suggest that almost one fourth of the variance in overall importance was unexplained by these predictors. Thus, it might well be the case that dimensions specific to teamwork are required to accurately evaluate the importance of these tasks. Similarly, it might be useful to consider alternative criterion measures for teamwork. Although subjective ratings of overall performance have been used traditionally in studies of individuals, objective measures, such as communication frequency, might provide a more valid measure of teamwork. This line of research requires further study in order to arrive at an optimal method of recommending tasks to be the targets for training.

ACKNOWLEDGMENTS

The views expressed in this article are those of the authors and do not reflect the official position of the Department of the Navy, the Department of Defense, or the U.S. Government.

An earlier version of this article was presented at the 1991 meeting of the Society for Industrial/Organizational Psychology, St. Louis, MO.

Dr. Bowers conducted this research at the Naval Training Systems Center during the Summer Faculty Research Program sponsored by the American Society for Engineering Education. He has subsequently returned to the University of Central Florida. Dr. Baker is now at InterScience America, Sterling, VA.

REFERENCES

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

Dawes, R. M. (1971). A case study of graduate admissions: Application of three principles of human decision making. *American Psychologist, 26,* 180–188.

Dawes, R. M. (1982). The robust beauty of improper linear models in decision making. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 391–407). Cambridge, England: Cambridge University Press.

Dyer, J. (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.

Franz, T. M., Prince, C., Cannon-Bowers, J. A., & Salas, E. (1990). The identification of aircrew coordination skills. *Proceedings of the 12th Annual Department of Defense Symposium,* 97–101.

Glickman, A. S., Zimmer, S., Montero, R. C., Guerette, P. J., Campbell, W. J., Morgan, B. B., & Salas, E. (1987). *The evolution of team skills: An empirical assessment with implications for training* (Tech. Rep. NTSC 87-016). Arlington, VA: Office of Naval Research.

Helmreich, R. L. (1991). The long and short term impact of crew resource management and training. *Proceedings of the AIAA, NASA, FAA, and Human Factors Society Conference on Challenges in Aviation Human Factors: The National Plan,* 81–83.

Helmreich, R. L., Foushee, H. C., Benson, R., & Russini, W. (1985). Cockpit resource management: Exploring the attitude-performance linkage. *Aviation, Space, and Environmental Medicine, 57,* 1198–1200.

Hogan, J., Broach, D., & Salas, E. (1990). Development of a task information taxonomy for human performance systems. *Military Psychology, 2,* 1–19.

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

Levine, E. L., & Baker, C. V. (1990). *Team task analysis for training design: A procedural guide to the Multiphase Analysis Performance (MAP) system and a tryout of the methodology.* Orlando, FL: Naval Training Systems Center.

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

selected organizations: A description and evaluation. Paper presented at the meeting of the Southeastern Industrial/Organizational Psychology Association, Orlando, FL.

Prince, C., Chidester, T. R., Bowers, C., & Cannon-Bowers, J. (1992). Aircrew coordination: Achieving teamwork in the cockpit. In R. W. Swezey & E. Salas (Eds.), *Teams: Their training and performance* (pp. 329–355). Norwood, NJ: Ablex.

Prince, C., & Salas E. (1993). Training for teamwork in the military aircrew: Military CRM programs and research. In E. Wiener, B. Kanki, & R. Helmreich (Eds.), *Cockpit resource management.* New York: Academic.

Salas, E., Dickinson, T. L., Converse, S. A., & 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. 3–31). New York: Ablex.

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

Smith, K., & Salas, E. (1991, March). *Training assertiveness: Impact of active participation.* Paper presented at the meeting of the Southeastern Psychological Association, New Orleans, LA.

Valenzi, E., & Andrews, I. R. (1973). Individual differences in the decision process of employment interviewers. *Journal of Applied Psychology, 58,* 49–53.