2009

What Can a Multidimensional Measure of Stress Tell Us About Team Collaborative Tools?

Gregory J. Funke

Sheldon M. Russell

Benjamin A. Knott

Brent T. Miller

WHAT CAN A MULTIDIMENSIONAL MEASURE OF STRESS TELL US ABOUT TEAM COLLABORATIVE TOOLS?

Gregory J. Funke[1], Sheldon M. Russell[1], Benjamin A. Knott[2], and Brent T. Miller[2]

[1]Consortium Research Fellows Program, Alexandria, VA, USA
[2]Air Force Research Laboratory, Wright-Patterson Air Force Base, OH, USA

A substantial body of research literature concerning the effects of collaborative tools on team performance has been generated, but the research has not considered subjective workload and stress associated with tool usage. The current experiment represents an initial, exploratory attempt to characterize the relationship between usage of collaborative tools, mental workload, and the subjective experience of stress. The NASA-TLX and the DSSQ-S were used to assess the workload and stress experienced by participants completing a simulated team command and control task. Task demands and collaborative tool availability were experimentally manipulated. Analysis of the data revealed that participants experienced increases in stress and workload with high task demands which were alleviated by the availability of collaborative tools under certain conditions. The results of this experiment demonstrate the complex relationships between collaborative technologies, workload, and stress.

Collaborative technologies, such as email and instant messaging (IM), are becoming vital tools for military organizations (e.g., Heacox, Moore, Morrison, & Yturralde, 2004). The availability of these tools has dramatically altered the ways in which personnel can communicate and collaborate, allowing organizations to shift from collocated teams to teams that may be geographically and temporally disbursed.

Within the military it has been suggested that collaborative technologies will enable a degree of command decentralization resulting in greater flexibility and adaptivity of forces (e.g., Alberts & Hayes, 2003). However, research indicates that the relationship between collaborative tools and distributed-team performance is complex, in that team task has consistently emerged as an important moderator of the influence collaborative technologies exert on performance (Hertel, Geister, & Konradt, 2005; Martins, Gilson, & Maynard, 2004).

One aspect of the collaborative tool literature that has not been considered yet is their relationship to operator workload and stress. While some research has been conducted examining job stress in fields that rely on collaborative technologies (such as call center workers; Zapf, Isic, Bechtold, & Blau, 2003), research examining the subjective workload and stress associated with tool usage has not yet been initiated, though several researchers have suggested that team affect and mood deserve greater attention from team researchers (Cohen & Bailey, 1997; Mathieu, Maynard, Rapp, & Gilson, 2008).

Modern theories of stress and workload are similar in that they posit that each can be viewed as an interaction between external demands and an individual's cognitive and behavioral responses to those demands (e.g., Gopher & Donchin, 1986; Lazarus, & Folkman, 1984). While workload and stress are considered to be separate theoretical constructs, they may influence performance through similar mechanisms. Attentional resource theories (e.g., Norman, & Bobrow, 1975) suggest that information processing and task performance are dependent on the availability of system resources. Such theories typically propose that system resources exist in a fixed quantity and that resources act as an energizer for information processing. It has also been suggested that subjective workload may represent the proportion of resources required to meet the demands of a task (e.g., Welford, 1978). As task demands increase, more resources are required for task performance and workload increases.

The effects of stress on performance may also be dependent on resource availability either by reducing the amount of resources available for task performance, or because some resources are diverted to processing stressful stimuli (Matthews, Davies, Westerman, & Stammers, 2000). In support of this viewpoint, various stressors, including noise, subjective tiredness, heat, anxiety and prolonged work, have been shown to impair performance most reliably when a task is attentionally demanding (Matthews et al., 2000).

This suggests a possible synergistic relationship between workload, stress, and collaborative technologies. To the extent that collaborative tools reduce operator stress, they may also be expected to reduce operator vulnerability to high workload, and vice versa. The purpose of the current experiment was to explore the influence of several collaborative technologies on subjective workload and stress in a simulated air defense task. Technologies included in this experiment were instant messaging, a virtual whiteboard, and a graphical data display. These technologies were selected because they are consistent with long-term military acquisition goals, and because they conform to anticipated future military capabilities (Sloan, 2008).

Method

Participants

Seventy men and 35 women, drawn from local universities and from a temporary work agency, were fiscally compensated for their participation. Participants were between the ages of 18 and 30 ($M = 21.94$, $SD = 3.16$), and completed the experiment in five-person teams, yielding a total of 21 experimental teams.

Experimental Design

A 3 × 2 × 2 × 2 mixed design was employed in this experiment. Team position was a between-participants factor with three levels (weapons director, strike operator, tanker operator). Within-participants factors included two levels of task demand (low, high), two levels of team communication (standard, enhanced), and two levels of data-display (tabular, graphical). Each team completed 2 trials in each experimental condition, for a total of 16 trials in each experimental session.

Materials

Questionnaires. Operator workload was assessed using the NASA Task Load Index (TLX; Hart & Staveland, 1988), which participants completed immediately following each trial. Subjective stress state was examined using the short version of the Dundee Stress State Questionnaire (DSSQ-S; Matthews, Emo, & Funke, 2005), an experimentally validated measure designed to assess multiple transient state factors associated with stress, arousal, and fatigue. DSSQ-S subscale scores are distributed with a mean of 0 and standard deviation of 1, so that the computed scores for a sample represent deviations from that sample's baseline values in standard deviation units. Participants in this experiment completed the measure immediately before beginning the experiment, and following each two-trial task demand block.

Apparatus. Five-person teams worked together to complete a simulated air defense command and control (C2) task. This task has been used in several previous experiments examining collaborative tool usage in military settings and has been demonstrated to be sensitive to experimental manipulations (e.g., Finomore, Knott, Nelson, Galster, & Bolia, 2007). Participants were randomly assigned to one of three team positions; positions differed in their roles and capabilities. The scenario required two weapons directors (WDs), two strike operators, and one tanker operator. Within the simulation, the WDs' roles were to match friendly fighters with appropriate enemy targets, schedule fighters for refueling and resupply, and communicate their plans with other team members. The role of the strike and tanker operators was to maneuver team assets as instructed, to engage enemy targets, and to provide pertinent information to teammates concerning asset resources.

The asset information available to team members was dictated by the data-display condition of that trial. In the tabular display condition, only strike and tanker operators had access to asset weapon and fuel status, presented in a digital format. WDs, therefore, had to rely on teammates for resource updates.

In the graphical display condition, asset fuel status was displayed in an analog format, and this display was available to all team members. In addition, the graphical display conveyed supplemental information to team members in that its associated asset fuel gauges changed to an amber color when fuel reserves were low, and it featured a black bar which indicated the minimum reserve fuel required to rendezvous with a tanker asset. Examples of both display types are presented in Figure 1.

The number of enemy targets present in each scenario was determined by the task demand condition of that trial. In the low and high demand conditions, 24 or 36 enemy targets, respectively, entered the simulation during the trial. At the conclusion of each trial, participants received a 'team score' based on three performance factors: a) prevention of enemy incursions, b) preservation of team assets, and c) protection of friendly ground forces.

Team communication. Communication between teammates in this experiment was manipulated through the team communication factor. In the standard communication condition, participants could communicate orally using a radio headset. All five team members communicated using the same radio channel to approximate the saturated communications experienced in many 'real world' military environments.

In the enhanced communication condition, participants could communicate using the radio or using two collaborative tools: instant messaging (IM) and a virtual whiteboard. The virtual whiteboard allowed a graphical annotated of participants' tactical displays to be distributed between teammates. This allowed participants to communicate spatial and tactical information (such as routes, enemy locations, etc.) without forcing them to divide their attention across multiple displays (Figure 2).

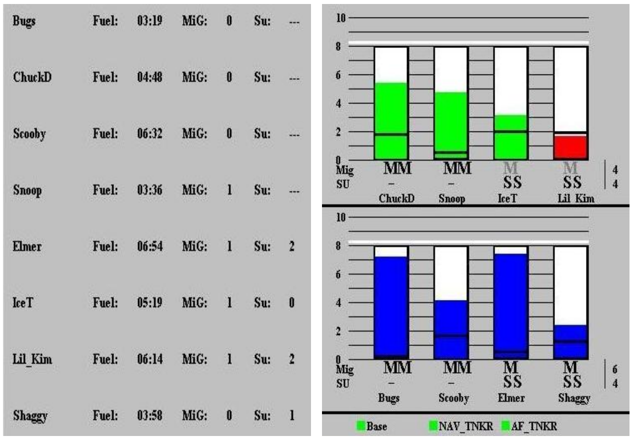

Figure 1. Tabular (left) and graphical (right) data displays. Both displays included information concerning remaining fuel and weapons of team assets.

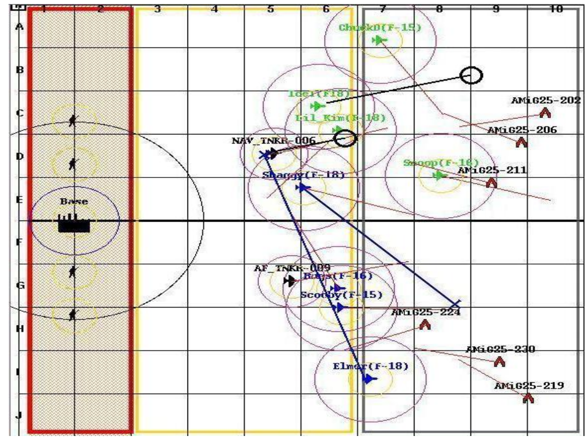

Figure 2. An image from the tactical display. Participant created whiteboard marks (blue and black lines) in the image indicate asset and target route information.

Procedure

The duration of the experiment was approximately 16 hours, conducted across two 8-hour session. The first session was devoted to training and the second to experimental data collection. In addition to 13 practice trials, participants received written and oral instructions detailing the C2 task, the team's goals, the roles and responsibilities of each team position, and the use of the collaborative tools. Participants were instructed on how to complete the DSSQ-S and the NASA-TLX. The experimental schedule of conditions was counterbalanced across teams to control order effects. After completing all experimental trials, participants were asked to complete a post-task debriefing form designed to elicit their impressions of the experimental factors and the C2 simulation. All experimental trials were ten minutes in duration.

Results

A full and detailed accounting of the results of this experiment is beyond the scope of this manuscript. As such, this section is focused chiefly on participants' subjective workload and stress responses to the experimental manipulations.

Team Communication

Following completion of the experimental data collection, audio recordings, instant messenger logs, and DRAW logs of the communications between teammates were compiled and examined. When the tools were available, teams sent, on average, 3.61 IM and 71.44 DRAW messages per trial. As a manipulation check, the mean number messages sent with each collaborative tool were tested against a value of zero using one sample *t*-tests to establish that teams were, in fact, using them. The results of these analyses indicated that participants were communicating at a rate greater than zero using IM, $t\ (20) = 5.94$, $p < .05$, and DRAW marks, $t\ (20) = 5.94$.

Workload

To test the effects of the experimental manipulations on participants' evaluation of task workload, the mean of each participant's TLX ratings in each condition was calculated. Mean TLX ratings for each experimental condition are presented in Table 1. Mean workload ratings were tested for statistically significant differences between conditions by means of a 3 (team position) × 2 (task demand) × 2 (team communication) × 2 (data-display) mixed-model ANOVA. The results of the analysis indicated statistically significant main effects of task demand, $F\ (1, 102) = 27.91$, $p < .05$, and data-display conditions, $F\ (1, 102) = 8.91$, $p < .05$, and statistically significant interactions between team position and team communication conditions, $F\ (2, 102) = 5.66$, $p < .05$, and between task demand and team communication conditions, $F\ (1, 102) = 4.85$, $p < .05$. No other sources of variance in the analysis were significant (all $p > .05$). Overall, participants rated their workload as higher in the high task demand condition compared to the low demand condition, and as higher when using the tabular display compared to the graphical display.

Table 1. *Mean NASA-TLX and DSSQ-S subscale change scores in each experimental condition.*

Team Position	Standard Communication				Enhanced Communication			
	Tabular Display		Graphical Display		Tabular Display		Graphical Display	
	Low	High	Low	High	Low	High	Low	High
				NASA-TLX Workload				
WD	55.14 (2.81)	57.62 (2.92)	51.25 (2.62)	54.65 (2.84)	50.57 (2.71)	54.11 (2.69)	48.08 (2.91)	54.29 (2.59)
Strike	46.76 (3.14)	48.63 (3.18)	44.68 (2.91)	45.94 (3.19)	49.24 (2.80)	52.54 (3.04)	46.18 (2.90)	50.53 (2.80)
Tanker	47.82 (3.41)	48.00 (3.48)	46.17 (3.53)	48.59 (3.34)	45.99 (3.46)	48.85 (3.55)	45.67 (3.89)	48.06 (3.56)
Mean	50.32 (1.84)	52.10 (1.90)	47.61 (1.73)	49.96 (1.86)	49.12 (1.70)	52.43 (1.76)	46.84 (1.80)	51.54 (1.68)
				DSSQ-S Task Engagement				
WD	-.25 (.24)	-.30 (.21)	-.09 (.24)	-.21 (.21)	-.44 (.26)	-.43 (.27)	-.02 (.20)	-.42 (.22)
Strike	.20 (.15)	-.06 (.18)	.17 (.16)	.08 (.17)	.14 (.13)	-.01 (.17)	.20 (.14)	.17 (.13)
Tanker	-.32 (.24)	-.12 (.25)	-.06 (.29)	-.02 (.23)	-.03 (.26)	-.30 (.37)	.11 (.32)	-.08 (.28)
Mean	-.09 (.13)	-.17 (.12)	.02 (.13)	-.05 (.12)	-.13 (.13)	-.23 (.15)	.09 (.11)	-.12 (.12)
				DSSQ-S Distress				
WD	.44 (.18)	.75 (.19)	.12 (.16)	.66 (.21)	.18 (.15)	.57 (.19)	.39 (.20)	.74 (.20)
Strike	.69 (.24)	1.18 (.28)	.35 (.18)	.46 (.19)	.87 (.25)	1.00 (.28)	.47 (.23)	.71 (.22)
Tanker	.19 (.15)	.41 (.25)	-.18 (.15)	-.23 (.17)	-.16 (.18)	-.04 (.26)	.00 (.18)	.69 (.24)
Mean	.49 (.13)	.85 (.14)	.15 (.10)	.40 (.12)	.39 (.13)	.62 (.15)	.34 (.13)	.72 (.13)
				DSSQ-S Worry				
WD	-.46 (.14)	-.41 (.13)	-.48 (.13)	-.61 (.13)	-.61 (.14)	-.47 (.12)	-.67 (.14)	-.54 (.12)
Strike	-.36 (.13)	-.41 (.13)	-.42 (.13)	-.46 (.13)	-.24 (.17)	-.32 (.15)	-.40 (.11)	-.34 (.13)
Tanker	-.25 (.13)	-.15 (.11)	-.27 (.13)	-.50 (.13)	-.54 (.20)	-.49 (.19)	-.41 (.18)	-.48 (.19)
Mean	-.38 (.08)	-.36 (.08)	-.42 (.08)	-.53 (.08)	-.45 (.10)	-.41 (.08)	-.51 (.08)	-.45 (.08)

Note. Values in parentheses are standard errors.

Follow-up post hoc paired sample *t*-tests for the team position × team communication interaction revealed no statistically significant differences between the team positions for either of the communication conditions (all comparisons $p > .05$). In these, and all subsequently reported post hoc analyses, the Dunn-Sidak alpha correction was employed to control Type-I error rates (Kirk, 1995). However, a trend within the data suggested that WDs rated their workload as slightly higher in the enhanced communication condition, and strike operators rated their workload as slightly lower in the same condition (both $p < .10$).

Post hoc paired sample t-tests investigating the task demand × team communication condition interaction indicated statistically significant differences between the low and high demand conditions in each team communication condition, t (104) = -3.60 and -5.63, respectively, $p < .05$. However the mean difference between the low and high task demand conditions was greater in the enhanced communication condition compared to the standard communication condition (i.e., participants' estimates of workload in the enhanced condition were lower in the low demand condition and higher in the high demand condition than those of the standard communication condition).

Stress State

Post-experiment, mean DSSQ-S subscale change scores were computed for each participant in each condition. In addition, a mean post-task subscale score was calculated for each participant as an index of participants' post-experiment state. However, due to a technical error, DSSQ-S data for three teams could not be recovered for analysis. Consequently, all subsequently reported analyses concerning team communications are based on data drawn from the remaining 18 participant teams.

Overall, the mean post-experiment scores indicated that participants' ratings of task engagement were largely unchanged (M = -.08, SD = 1.03), distress increased slightly (M = .50, SD = .95), and worry decreased slightly (M = -.44, SD = .67). Correlations between pre-task and post-experiment DSSQ-S ratings were .59, .60, and .78 for task engagement, distress, and worry, respectively (all $p < .05$), suggesting that participants' mood states were relatively stable from pre- to post-experiment.

Mean DSSQ-S change scores for each subscale are presented in Table 1. Subscale change scores were tested for statistically significant differences between experimental conditions by means of separate 3 (team position) × 2 (task demand) × 2 (team communication) × 2 (data-display) mixed-model ANOVAs.

Task engagement. The results of the task engagement analysis revealed statistically significant main effects for the task demand, $F(1, 87) = 7.64$, $p < .05$, and data-display factors, $F(1, 87) = 5.24$, $p < .05$. No other sources of variance in the analysis were significant (all $p > .05$). Participants rated their engagement as lower in the high demand condition compared to the low demand condition, and as lower in the tabular data-display condition compared to the graphical condition. Overall, participants were more engaged when the task was less demanding and when they had access to the graphical data-display.

Distress. For the distress subscale, the results of the ANOVA indicated a statistically significant main effect of task demand, $F(1, 87) = 21.03$, $p < .05$, and statistically significant interactions between team position and communication conditions, $F(2, 87) = 4.75$, $p < .05$, team communication and data-display conditions, $F(1, 87) = 8.91$, $p < .05$, and a three-way interaction between team position, task demand, and data-display condition, $F(2, 87) = 3.67$, $p < .05$. No other sources of variance in the analysis were statistically significant.

Follow-up post hoc paired sample *t*-tests for the team position × team communication interaction revealed no statistically significant differences between the team positions for either of the communication conditions (all comparisons $p > .05$). However, a trend within the data suggested that strike operators rated their distress as slightly higher in the enhanced communication condition ($p < .10$).

Post hoc paired sample *t*-test analyses of the team communication × data-display interaction indicated that participants rated their distress as higher when using the tabular data-display as compared to the graphical display, but only in the standard communication condition. No distress differences were observed between data-display conditions in the enhanced communication condition ($p > .05$).

To further explore the team position × task demand × data-display interaction, separate post hoc 3 (team position) × 2 (task demand) repeated measures ANOVAs were computed for each data-display condition. For the graphical data-display, the results of the analysis indicated that ratings of distress varied by task demand, $F(1, 87) = 17.78$, $p < .05$. Participants rated their distress as higher in the high task demand condition compared to the low condition when using the graphical display.

The results for the tabular data-display were more complex, in that a statistically significant main effect of task demand, $F(1, 87) = 7.09$, $p < .05$, and a statistically significant team position × task demand interaction, $F(2, 87) = 4.98$, $p < .05$, were identified. Subsequent post hoc paired sample *t*-tests indicated that, in the tabular display condition, WDs rated their distress as significantly higher in the high task demand condition compared to the low demand condition. No such differences were detected for strike and taker operators.

Worry. The results of the analysis for the worry subscale indicated a statistically significant main effect for data-display condition, $F(1, 87) = 5.17$, $p < .05$, and statistically significant interactions between task demand and team communication conditions, $F(1, 87) = 4.34$, $p < .05$, and between task demand, team communication, and data-display conditions, $F(1, 87) = 4.28$, $p < .05$.

To continue examination of the three-way interaction, separate 2 (task demand) × 2 (team communication) repeated measures ANOVAs were computed for each of the data-display conditions. For the tabular display condition, no statistically significant differences between conditions were detected (all $p > .05$). For the graphical display, however, the analysis indicated a statistically significant task demand × team communication interaction, $F(1, 89) = 6.34$, $p < .05$. Follow-up post hoc paired sample *t*-tests indicated that in the enhanced communication condition, participants did not rate their worry differentially between task demand conditions. Conversely, in the standard communication condition, participants rated their worry as lower in the high demand condition compared to the low demand condition.

Discussion

The purpose of this experiment was to provide a preliminary attempt to characterize the relationship between collaborative tools and subjective workload and stress. In general, the results of this experiment suggest that collaborative tools and technologies may be both a significant source of, and solution to, operator workload and stress.

Participants' workload ratings were higher in the high demand condition and when using the tabular display. Workload ratings were also influenced by the collaborative tools available to participants, but their effects were moderated by the team position and task demand factors.

Effects of the experimental manipulations on subjective stress response were more nuanced than anticipated. Overall, task engagement and worry decreased, and distress increased from pre- to post-experiment. The observed decrement in task engagement was exacerbated by high task demands and the tabular data-display condition, but was not changed by collaborative tool availability. Distress was further increased by high task demands, but the strength of this effect was dependent on team position, team communication, and data-display conditions. Worry decreased differentially depending on task demand, team communication, and data-display conditions.

Collaborative tools. Access to additional collaborative tools had relatively weak effects on subjective workload and stress in this experiment. Though there were some suggestions of incremental differences in workload and stress experience based on team position and collaborative tool availability, the magnitude of these effects was mostly negligible. This indicates that collaborative tool usage, as implemented in this experiment, does not exert any additional 'costs' in terms of workload or stress (though see below). However, these results do not indicate that organizations should be unconcerned about workload and stress associated with collaborative technologies; collaborative tools may still be a significant source of workload and stress for users for a variety of reasons (e.g., because of poor interface design, inadequate training, laborious implementation, etc.).

Data-display types. Access to the graphical data display decreased subjective estimates of workload and stress compared to the tabular display in this experiment. It is reasonable to assume that while some degree of benefit was provided by the reduction in communication required during a trial (i.e., that relating to WDs and operators exchanging asset weapon and fuel information), some of the observed benefit of the graphical display should also be attributed to its enhanced functionality, which provided WDs with salient cues concerning fuel management. This, in turn, may have allowed WDs to more efficiently allocate team assets to enemy targets, resulting in improved team scores.

The relationship between data display type and team position was also reflected in subjective distress and worry ratings, but this relationship was moderated by task demand and team communication conditions. An interesting aspect of these results is in the complexity of the interactions observed between the experimentally manipulated factors. The results do not 'add up' to a singular representation describing the relationship between subjective stress and the experimental factors. Instead, they illustrate that, under varying circumstances, some team members may be benefitted by the availability of collaborative technologies while others are simultaneously unchanged (or hindered) by exactly the same tools.

This suggests that teams may be better served by *adaptive* collaborative technologies, which may be tailored according to the needs and circumstances of individual team members (Baldwin, 2003). By allowing team members (or an automated decision aid) to flexibly and dynamically alter the functionality of these tools, it may be possible to maximize team performance while minimizing associated negative outcomes such as subjective workload and stress. Determining the nature and behavior of such tools is likely to be a fruitful area of future research.

References

Albert, D.S., Hayes, R.E, 2003. Power to the Edge. CCRP Publication Series, Washington, DC.

Baldwin, C.L. (2003). Neuroergonomics of mental workload: New insights from the convergence of brain and behaviour in ergonomics research. *Theoretical Issues in Ergonomic Science, 4*, 132-141.

Cohen, S.G., & Bailey, D.E. (1997). What makes teams work: Group effectiveness research from the shop floor to the executive suite. *Journal of Management*, *23,* 239-290.

Finomore, V.S., Knott, B.A., Nelson, W.T., Galster, S.M., & Bolia, R.S. (2007). The effects of multimodal collaboration technology on subjective workload profiles of tactical air battle management teams. *Proceedings of the International Symposium on Aviation Psychology, 14*, 190-196.

Gopher, D., & Donchin, E. (1986). Workload: An examination of the concept. In K.R. Boff, L. Kaufman, & J.P. Thomas (Eds.), *Handbook of human perception and human performance: Vol. 2. Cognitive processes and performance* (pp. 41-1 to 41-49). New York, NY: Wiley.

Heacox, N.J., Moore, R.A., Morrison, J.G., Yturralde, R.F. (2004, June). Real-time online communications: 'Chat' use in navy operations. In *Proceedings of the Command and Control Research and Technology Symposium*, San Diego, CA.

Hertel, G., Geister, S., & Konradt, U. (2005). Managing virtual teams: A review of current empirical research. *Human Resource Management Review, 15*, 69-95.

Kirk, R.E. (1995). *Experimental design: Procedures for the behavioral sciences* (3rd ed.). Pacific Grove, CA: Brooks/Cole.

Lazarus, R. S., & Folkman, S. (1984). *Stress, appraisal and coping*. New York: Springer.

Martins, L.L., Gilson, L.L., & Maynard, M.T. (2004). Virtual teams: What do we know and where do we go from here? *Journal of Management, 30*, 805-835.

Mathieu, J., Maynard, M.T., Rapp, T., & Gilson, L. (2008). Team effectiveness 1997-2007: A review of recent advancements and a glimpse into the future. *Journal of Management, 34,* 410-476.

Matthews, G., Davies, D.R., Westerman, S.J., & Stammers, R.B. (2000). *Human performance: Cognition, stress and individual differences*. London: Psychology Press.

Matthews, G., Emo, A.K., & Funke, G.J. (2005, July). *A short version of the Dundee stress state questionnaire*. Poster presented at the annual meeting of the International Society for the Study of Individual Differences, Adelaide, Australia.

Norman, D.A., & Bobrow, D.G. (1975). On data-limited and resource-limited processes. *Cognitive Psychology, 7,* 44–64.

Sloan, D. (2008, July 28). *Boeing to demo net-centric upgrade on AWACS aircraft at empire challenge*. Retrieved February 27, 2009, from the Boeing Company web site: http://www. boeing.com/news/releases/2008/q3/080708a_nr.html

Welford, A.T. (1978). Mental work-load as a function of demand, capacity, strategy and skill. *Ergonomics, 21,* 151–167.

Zapf, D., Isic, A., Bechtold, M. & Blau, P. (2003). What is typical for call centre jobs? Job characteristics, and service interactions in different call centres. *European Journal of Work and Organizational Psychology*, *12*, 311-340.