2009

Team Workload: A Construct Worth Pursuing?

Gregory J. Funke
Benjamin A. Knott
Scott M. Galster
Rebecca D. Brown

Follow this and additional works at: https://corescholar.libraries.wright.edu/isap_2009

Part of the Other Psychiatry and Psychology Commons

Repository Citation
https://corescholar.libraries.wright.edu/isap_2009/45

This Article is brought to you for free and open access by the International Symposium on Aviation Psychology at CORE Scholar. It has been accepted for inclusion in International Symposium on Aviation Psychology - 2009 by an authorized administrator of CORE Scholar. For more information, please contact corescholar@www.libraries.wright.edu, library-corescholar@wright.edu.
TEAM WORKLOAD: A CONSTRUCT WORTH PURSUING?

Gregory J. Funke1, Benjamin A. Knott2, Scott M. Galster3, and Rebecca D. Brown4
1Consortium Research Fellows Program, Alexandria, VA, USA
2Air Force Research Laboratory, Wright-Patterson Air Force Base, OH, USA
3Aptima Inc., Fairborn, OH, USA
4Ball Aerospace & Technologies Corp., Boulder, CO, USA

To date, little research has been conducted on the workload experienced by teams. Attempts to evaluate team workload have typically relied upon validated measures of individual workload, but this approach may not adequately capture the drivers of team workload. Data from three previous research experiments were reexamined using hierarchical multiple regression in the present analysis. Each of the experiments assessed individual and team workload, though the measures employed differed in each. The goal was to investigate whether addition of team workload measures improved prediction of team performance after variance associated with experimental manipulations and individual workload had been removed. Results indicated that inclusion of team workload measures produced an increase of 5-15% in the variance explained. This suggests that workload associated with team processes is not adequately reflected in measures of individual workload. Researchers investigating workload in team settings are recommended to consider inclusion of a team workload measure in their experiments.

The mental workload experienced by an individual operator performing a task has received a significant degree of scientific inquiry, particularly within the last half century (Tsang & Vidulich, 2006). The result has been a proliferation of theories, methods, and metrics designed to evaluate individual workload. Workload assessment continues to remain a vital area of human performance research, providing understanding and prediction of human performance in complex systems (Parasuraman, Sheridan, & Wickens, 2008).

By contrast, however, research on team workload has been virtually overlooked (Bowers & Jentsch, 2005), even though teams have been acknowledged as an integral part of the military, industrial, medical, and public service sectors (Bowers, Braun, & Morgan, 1997). A comprehensive, validated theoretical framework for the construct ‘team workload,’ which includes a description of its relation to individual workload, has not yet been articulated. While some interesting work has been initiated (e.g., Bowers et al., 1997), progress in the area has been quite slow.

One potential explanation for these circumstances is that attempts to assess team workload have been rooted in measures of individual workload. Most research in the area has been conducted using existing individual workload measures, most frequently the NASA Task Load Index (TLX; Hart & Staveland, 1988), which have been modified by alteration of the instruction set, item set, or both to make them more applicable to teams (e.g., Bowers & Jentsch, 2005). This approach, however, is not without its difficulties.

First, consideration must be given to how the data from a team of individuals should be combined and interpreted (Bowers et al., 1997). Several different methods have been advanced, including the averaged workload of all team members, the lowest workload value obtained, and the highest workload value obtained (Bowers & Jentsch, 2005), though it should be noted that there are no theoretically-compelling reasons to adopt any of these. Second, it is perhaps unreasonable to presume that the psychometric properties of these measures, determined at the level of the individual, will be unchanged by transitioning to a team level of analysis. Lastly, several researchers have reported finding no differences in participant workload ratings using a modified team-TLX in response to manipulations of task demands (e.g., Bowers, Urban, & Morgan, 1992; Thornton, Braun, Bowers, & Morgan, 1992; Uban, Bowers, Monday, & Morgan, 1995), leading Bowers, Braun, and Morgan (1997) to suggest that individual workload measures may not be sensitive enough to the sources of team demands to adequately capture team workload.

Given the complexities of measurement and the slow progress of research in the area, a reasonable conclusion may be to question the utility of a team workload construct. If team workload is difficult to assess and provides little additional explanatory power beyond what is already provided by measures of individual workload, then perhaps it is defensible to focus research efforts elsewhere. Conversely, if individual measures of workload are insensitive to
The goal of the current experiment was to examine the additional explanatory power measures of team workload may afford beyond that provided by individual measures. To explore this issue, data-sets from several previous experiments were assembled for further analysis. These data were tested using hierarchical multiple regression to determine the degree of additional variance explained by team workload, beyond that accounted for by individual workload, as it was regressed onto a measure of team performance. It was initially hypothesized that the inclusion of a team workload factor in the analysis would significantly increase the variance explained by the regression model.

Method

Data sets from three previous experiments were assembled for further analysis. These experiments featured several communalities. First, each included a complex task which required participants to work together as a team in order to meet scenario goals. The simulated environment utilized in each experiment was Aptima, Inc.’s Distributed Dynamic Decision-making (DDD) simulation (MacMillan, Entin, Hess, & Paley, 2004). DDD provides a scriptable, low-to-moderate fidelity, team-in-the-loop simulated environment. DDD has successfully been used to simulate team command and control tasks and to study realistic and complex team processes in a variety of military and civilian research projects (MacMillan et al., 2004).

In each experiment, task feedback was provided to participants following a trial in the form of a ‘team score’ which reflected how well the team had achieved the goals of the scenario. This score was scaled so it could range from 0-100; a score of 0 indicated that the team had not met any of the goals of the scenario, and a score of 100 indicated that the team had met all goals perfectly. Team scores were presented to participants after all team members had completed the individual and team workload measures.

Secondly, across experiments, participants completed the same measure of individual workload, the NASA-TLX (Hart & Staveland, 1988), a standard measure of workload that is widely used in human performance research (Wickens & Hollands, 2000). The NASA-TLX provides a global index of task workload on a scale of 0 to 100 and identifies the relative contributions of six sources of workload: mental demand, temporal demand, physical demand, performance, effort, and frustration. In all experiments, participants completed the TLX immediately following a trial.

Thirdly, each experiment included a different, measure of team workload. It should be noted that the focus of these experiments was not team workload and that the included measures were of peripheral interest to the research questions investigated in each. Participants completed a team workload measure immediately following each trial.

A brief description of the experimental task and the team workload measure employed in each data set are included below.

Data Set 1: Schwartz (2008)

Schwartz (2008) examined the relationship between team decision-making structure (centralized, decentralized) and learning (trials) on team performance in a scenario designed to simulate an air strike against hostile forces. Participants completed the experiment in teams of five. Each of the ten participants completed the TLX and team workload measure 23 times in each (decision-making) condition.

Team workload measure. Utilizing a process similar to that proposed by Dickinson and McIntyre (1997), Schwartz (2008) assessed team workload using the Consensus-TLX (C-TLX), a consensus-scored version of the NASA-TLX. Team members worked collaboratively to assign a score reflecting the team’s overall perceived workload on each of the six dimension of the NASA-TLX. Participants were told to discuss each trial in conjunction with the six rating scales and to assign a single value to each scale that best reflected the team’s workload. Scales of the C-TLX were scored in the same fashion as those of the NASA-TLX (i.e., 0-100).
Data Set 2: Funke, Bennett, Nelson, & Galster (2007)

Funke et al. (2007) investigated the effects of UCAV-control (direct, supervisory) and team-collaboration (standard, whiteboard) conditions on team performance in a simulated suppression of enemy air defense (SEAD) mission. Participants were assigned to one of six team positions; positions differed from each other in their role in the scenario and in the simulated capabilities of their assets. Each of the six participants completed the TLX and team workload measure 12 times in each condition.

Team workload measure. Funke et al. (2007) assessed team workload using the Modified-TLX (M-TLX; Pharmer, Cropper, McKneely, & Williams, 2004). The M-TLX combines the standard six rating scales of the NASA-TLX with an additional five items; the supplementary items addressed sources of team workload and included communication demand, monitoring demand, control demand, coordination demand, and leadership demand. The team workload items of the M-TLX are scored in the same fashion as the NASA-TLX (i.e., item ratings may range from 0 to 100) and a global team workload score is calculated by computing the mean of those five items.

Data Set 3: Funke, Russell, Knott, & Miller (2009)

Funke et al. (2009) examined the impact of task demand (high, low) and collaborative tool availability (with and without access to a resource display, and with and without access to collaborative tools) on team performance in a simulated air defense task. Participants completed the experiment in teams of five. Each of the 105 participants completed the TLX and team workload measure twice in each experimental condition.

Team workload measure. Funke et al. (2009) utilized the Team Workload Assessment Scale (TWAS; Galster & Knott, 2007), a new, unvalidated measure specifically created to assess team workload. The TWAS provides a global index of team workload ranging from 0 to 100 and identifies the relative contributions of ten sources of team workload: temporal demand, physical demand, mental demand, task engagement, coordination demand, task difficulty, control demand, communication demand, team focus, and environmental interference.

Results

Correlations

Table 1 displays the mean team performance score, mean individual (NASA-TLX) and team workload ratings, and the correlations between them for each data set. Examination of the correlations suggests that, across data sets, individual and team workload measures were moderately correlated with team performance such that increases in workload were associated with decrements in team performance. The observed correlations also indicate that individual and team workload ratings were quite similar, and raises concerns about the relative contributions of each to predictions of team performance.

One potential explanation is that the measures of team workload included in this analysis do not provide additional information beyond what is provided by measures of individual workload (i.e., individual and team measures of workload are essentially equivalent, and capture the same variance). Alternatively, individual and team workload may strongly covary because comparable processes moderate ratings of each. To address these

<table>
<thead>
<tr>
<th>Source</th>
<th>Team Workload Measure</th>
<th>Team Score (Mean)</th>
<th>NASA-TLX Workload (Mean)</th>
<th>Team Workload (Mean)</th>
<th>Team Score &amp; NASA-TLX</th>
<th>Team Score &amp; Team Workload</th>
<th>NASA-TLX &amp; Team Workload</th>
</tr>
</thead>
<tbody>
<tr>
<td>Schwartz (2008)</td>
<td>C-TLX</td>
<td>49.67</td>
<td>53.87</td>
<td>58.81</td>
<td>.00</td>
<td>-.23*</td>
<td>.83*</td>
</tr>
<tr>
<td>Funke et al. (2007)</td>
<td>M-TLX</td>
<td>70.73</td>
<td>49.99</td>
<td>58.35</td>
<td>-.43*</td>
<td>-.61*</td>
<td>-.67*</td>
</tr>
<tr>
<td>Funke et al. (2009)</td>
<td>TWAS</td>
<td>91.39</td>
<td>48.66</td>
<td>31.32</td>
<td>-.32*</td>
<td>-.35*</td>
<td>.83*</td>
</tr>
</tbody>
</table>

Note. NASA-TLX = NASA-Task Load Index (Hart & Staveland, 1988); C-TLX = Consensus-TLX (Schwartz, 2008); M-TLX = Modified-TLX (Pharmer, Cropper, McKneely, & Williams, 2004); TWAS = Team Workload Assessment Scale (Galster & Knott, 2007). * p < .05.
possibilities, separate hierarchical multiple regression analyses were conducted for each of the three data sets included in this experiment.

Hierarchical Multiple Regression

The goal of these analyses was to examine the incremental increases in variance accounted for by the regression models after variance associated with experimental manipulations and individual workload had been removed. Analyses were conducted in four steps; team score served as the criterion in each analysis. For these regressions, effect-coded vectors for the team, trial, and experimental conditions of each data set were created using the method recommended by Pedhazur (1997). Interaction terms were constructed as product vectors of those task variables. A linear global individual workload term was computed from the mean workload ratings of all team members across the six subscales of the NASA-TLX. Similarly, a linear global team workload term was constructed as the mean workload ratings of all team members across the subscales of the team workload measure employed in each data set (i.e., the C-TLX, M-TLX, and TWAS).

In each regression analysis, the first step entered consisted of the effect coded vectors of the team, trial, task variables, and task variable interactions. These variables were entered first to initially partition variance associated with the experimental manipulations in each data set. The second and third steps entered were the linear individual workload and team workload terms, respectively. This order of entry provided a test of the increase in variance accounted for by the team workload measure after variance associated with individual workload had been removed. The final step in each regression was the entry of all subscales of the NASA-TLX and the relevant team workload measure. The purpose of the final step was to further clarify the potential drivers of individual and team workload in each sample.

Summaries of the results for each regression equation are displayed in Table 2†. All subsequently reported regression coefficients (β’s) are standardized. Regression coefficients related to experimental manipulations (i.e., those related to team, trial, and experimental conditions) are not reported here as they generally coincide with previously published results. Interested readers are directed to those publications for further information.

Data Set 1: Schwartz (2008). The results of the hierarchical multiple regression analysis indicated that inclusion of the individual and team workload vectors (steps 2 and 3) did not significantly increase the variance explained by the model. However, addition of the NASA-TLX and the C-TLX workload subscales (step 4) substantially increased the variance accounted for. Statistically significant predictors in the final model included the C-TLX performance subscale (β = -.538, \( p < .05 \)) and a trend toward the C-TLX effort subscale (β = -.333, \( p < .10 \)). In this experiment, C-TLX subscale scores were more predictive of team scores than were global individual and team workload

Table 2. Summary statistics for hierarchical multiple regression analyses of task variables (team, trial, experimental conditions), individual workload, team workload, and workload subscales onto team score.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( R^2 )</td>
<td>( \Delta R^2 )</td>
<td>df for ( \Delta F )</td>
<td>( F ) for ( \Delta )</td>
</tr>
<tr>
<td>1</td>
<td>Task variables</td>
<td>.65</td>
<td>.65</td>
<td>46, 43</td>
</tr>
<tr>
<td>2</td>
<td>Task variables</td>
<td>.67</td>
<td>.02</td>
<td>1, 42</td>
</tr>
<tr>
<td></td>
<td>Ind. workload</td>
<td>.59</td>
<td>.01</td>
<td>1, 31</td>
</tr>
<tr>
<td>3</td>
<td>Task variables</td>
<td>.69</td>
<td>.02</td>
<td>1, 41</td>
</tr>
<tr>
<td></td>
<td>Ind. workload</td>
<td>.64</td>
<td>.05</td>
<td>1, 30</td>
</tr>
<tr>
<td></td>
<td>Team workload</td>
<td>.70</td>
<td>.02</td>
<td>1, 291</td>
</tr>
<tr>
<td>4</td>
<td>Task variables</td>
<td>.85</td>
<td>.16</td>
<td>10, 31</td>
</tr>
<tr>
<td></td>
<td>Ind. workload</td>
<td>.74</td>
<td>.10</td>
<td>9, 21</td>
</tr>
<tr>
<td></td>
<td>Team workload</td>
<td>.77</td>
<td>.07</td>
<td>14, 277</td>
</tr>
<tr>
<td></td>
<td>Workload subscales</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. * \( p < .05 \).

†Copies of team workload measures and complete summaries of all regression equations may be obtained by contacting the lead author. Email: Gregory.Funke@wpafb.af.mil
estimates. Overall, increases in consensus-scored estimates of team performance and effort were associated with poor team performance.

Data Set 2: Funke et al. (2007). Results of the regression analysis indicated that inclusion of the individual workload vector did not significantly improve the variance explained by the model, but addition of the team workload vector did (albeit the gain was relatively modest). The last step of the regression analysis, which entered the NASA-TLX and M-TLX subscales, did not significantly increase the variance accounted for. The only statistically significant predictor in the final regression model (excluding those associated with task variables) was the global team workload vector ($\beta = -0.360, p < .05$). As demands on team processes increased, team performance generally decreased.

Data Set 3: Funke et al. (2009). In this data set, each step of the multiple regression contributed significantly to the variance accounted for in the model. Inclusion of the team workload and workload subscale vectors (Steps 3 and 4) resulted in modest increases in the variance explained. Statistically significant predictors in the final regression model included the NASA-TLX mental demand subscale ($\beta = -0.402, p < .05$), and the TWAS task engagement and environmental interference subscales ($\beta$s = -0.176 and -0.319, respectively, both $p < .05$). Generally, team performance decreased as the mental demands placed on individual team members increased. Team performance was also negatively impacted by team-wide difficulties maintaining task engagement and interference related to the task environment.

Discussion

The purpose of the current research was to explore the potential utility of ‘team workload’ as an initial stage in the development of more comprehensive theoretical and empirical models of the construct. This was achieved by reanalyzing data from three experiments using hierarchical multiple regression to examine the incremental increases in variance associated with team performance explained in the regression models by the inclusion of measures of individual and team workload. Broadly, the results of the regression analyses support the value of including measures of team workload in team research, though the experimental ‘gains,’ in terms of variance explained, may be small to moderate.

The results of the multiple regression analyses indicated that, in the experiments sampled, the averaged individual and team workload vectors did not contribute to the prediction of team performance in a uniform fashion. Across the experiments sampled, dissimilar and non-overlapping patterns of statistically significant predictors emerged. This is somewhat unsurprising given the substantial differences in manipulated variables, experimental tasks, and team workload measures originally employed in each data set. It is also worth mentioning again that none of the experiments included in this research were focused on assessment of team workload. Nonetheless, the observed regression results do support the utility of team workload assessment. Across data sets, the global team workload rating (and frequently, the associated subscale ratings) consistently emerged as a significant predictor of team performance.

Measures of Team Workload

The results of the current study do not definitively answer questions concerning the appropriateness of using individual workload measures to assess team workload. The results do support suggestions by Bowers et al. (1997) that measures such as the NASA-TLX may be insufficiently sensitive to sources of team workload. As such, researchers investigating workload in team settings may wish to consider inclusion of a team workload measure in their experiments.

However, it is worth reiterating that a validated measure of team workload has not yet emerged from the research community (Bowers & Jentsch, 2005). The team workload measures sampled in the current study were selected because of their availability, rather than because of their sound psychometric properties. These measures showcase the potential of a validated team workload measure for understanding team performance; they do not provide a means to circumvent the research required for this endeavor. In developing a metric of team workload, the key question researchers must answer is one of construct validity: can ‘team workload’ be adequately assessed using a more sensitive measure of individual workload, which is then aggregated across team members, or is team
workload a new construct which is related to, but distinct from, individual workload and which requires a new measure to address.

Conclusions

The current study reaffirms the need for a validated theory and measure of team workload. Significant research and expertise are still required to resolve the difficult theoretical issues surrounding the construct ‘team workload.’ In addition, further research concerning the issues of sensitivity and appropriateness of individual workload measures for team workload assessment are certainly warranted.

Acknowledgements

The authors would like to thank Daniel Schwartz for generously allowing the use of his dissertation data in the present study.

References


