*Neuroergonomics and Cognitive Engineering,* Vol. 167, 2025, 20–32
https://doi.org/10.54941/ahfe1006124

# A Cognitive Efficiency Approach to Assessing Workload-Performance Trade-Offs in Human-Autonomy Teams

**T'kara Mullins and Stephen M. Fiore**

University of Central Florida, Florida

## ABSTRACT

Team environments increasingly integrate autonomous technologies to reduce workload and enhance performance. However, traditional metrics may overlook indirect benefits, particularly cognitive efficiency—the balance between workload and performance. This study examined how autonomous decision support systems (ADSS) affected cognitive efficiency and team communication in a high-fidelity simulated combat environment. Twenty-eight military personnel completed ten missions, alternating between conditions with and without ADSS that provided task-switching cues and flexible task allocation. Results showed that missions using ADSS produced higher cognitive efficiency, indicating improved performance relative to workload, though benefits varied by role and team. Teams also demonstrated enhanced communication patterns with ADSS, using more insight-related language, and maintaining a more positive tone. These findings highlight cognitive efficiency as a valuable metric for evaluating autonomy in complex environments and reveal how autonomous systems can enhance team effectiveness through improved team processes.

**Keywords:** Human-autonomy teams, Team workload, Cognitive load, Cognitive efficiency, Team dynamics

## INTRODUCTION

The integration of autonomous technologies into complex team environments is often promoted as a means to improve performance by way of reducing cognitive workload. This rationale is grounded in capacity theories of cognition, particularly Limited Capacity Theory (Kahneman, 1973) and Multiple Resource Theory (Wickens, 1984), which posit that humans possess finite cognitive resources that, when depleted, lead to performance decrements. The general assumption, therefore, is that autonomous systems should enhance performance by reducing cognitive workload, and that increased workload or lack of direct performance improvements indicates suboptimal implementation. However, this view oversimplifies the dynamic relationship between workload and performance, particularly in complex team environments where task demands are interdependent and distributed between individuals.

In certain contexts, high workload levels may not only be acceptable, but necessary, especially when managing complex, interdependent tasks. The critical factor is not the absolute level of workload, but rather the relationship between workload and performance outcomes. We argue that a system that increases cognitive workload is not necessarily suboptimal if that increased load translates to meaningful performance improvements. Rather, the concern lies in the scenario in which cognitive demands increase without corresponding performance benefits. This perspective suggests that effective system evaluation must move beyond simple workload reduction metrics to consider the broader workload-performance relationship.

The relationship between workload and performance becomes even more pronounced in team environments, where changes in individual workload and performance must be interpreted within the team context. The performance of individual team members contributes to overall team outcomes in both explicit and implicit ways. While individual contributions are often measured through task-specific performance, team-level outcomes are fundamentally shaped by team processes (Kozlowski & Klein, 2000), which are the interdependent activities that teams engage in to achieve a shared goal (Marks et al., 2001), such as communication. This distinction highlights the potential for autonomous systems to influence team effectiveness not merely by enhancing individual task performance, but by impacting team processes. By offloading routine tasks, autonomous systems can free up cognitive resources that can be reallocated to higher order team processes such as coordination, decision-making, and problem-solving.

For instance, an autonomous target recognition system may increase a gunner's workload without directly improving their accuracy, but aided recognition may free up cognitive resources otherwise allocated to continuous monitoring, leading to enhanced team effectiveness because the gunner may subsequently contribute to better situational awareness via improved communication. In this way, the value of autonomy lies not only in its ability to enhance individual capabilities but also in its ability to elevate the quality of team processes, thereby driving superior team outcomes. As such, the evaluation of systems in team environments requires particular attention to the ways in which the outputs from individuals combine via team processes to produce collective outcomes.

The present study examines the relationship between workload and performance in the context of an autonomous team decision support system designed to enhance collective performance through improved information sharing and coordination. To capture the complex interaction between individual and team outcomes, we evaluated *team cognitive efficiency* (Johnston et al., 2013). This was based upon the approach developed by Fiore Scielzo, Jentsch, and Howard (2006), labeled cognitive efficiency, to study training effectiveness in the context of workload reductions during testing. As noted by Fiore and colleagues, this measure is conceptually similar to instructional efficiency (Paas & Van Merrienboer, 1993), which used standardized scores of workload and performance and has been used to generally study cognitive load (Paas & Tuovinen 2004; Paas, Tuovinen, Tabbers, & Van Gerven, 2003).

Although research had called for development of workload measures at a team level (e.g., Funke et al., 2012; Bedwell et al., 2014), the study on team cognitive efficiency, reported by Johnston et al. (2013), was one of the first to empirically demonstrate measures of workload in a collaborative setting. Specifically, this is a metric that quantifies the trade-off between individual-level workload against team-level performance outcomes. Team cognitive efficiency was originally developed to evaluate the effectiveness of static decision support systems (Johnston et al., 2013). Original findings demonstrated that teams with decision support technology showed positive efficiency scores (better performance relative to increased workload), while control teams showed negative scores (increased workload without increased performance). Building on this foundation, we adopt team cognitive efficiency to evaluate an autonomous decision support system (ADSS). Where Johnston et al. examined how static information displays affected workload-performance trade-offs, we investigate how dynamic task recommendations and allocation capabilities influenced team effectiveness.

By using team cognitive efficiency, we aim to assess whether the team-level performance gains realized by using the ADSS outweighed the individual levels of cognitive workload induced by the system. We hypothesize the following:

### H1: Cognitive Efficiency Scores Will be Higher for Missions That Used the ADSS as Compared to Those That Did not

Johnston et al. (2013) also found that cognitive efficiency scores varied by role, with some roles benefiting more from decision support systems than others. As detailed in the Method section below, each participant was assigned to a unique role with role-specific responsibilities. We anticipate that some roles may not benefit as much from the ADSS, particularly those with fewer opportunities for multitasking or task reallocation. Thus, we propose:

### H2: Cognitive Efficiency Scores Will Vary Between Roles

Finally, in line with capacity theories and those emphasizing the importance of team processes, we hypothesize that improvements in cognitive efficiency will be accompanied by changes to team processes, particularly communication. Specifically, we hypothesize that:

### H3: Team Communication Processes Will Vary Between Technology Use Conditions

This study contributes to the growing body of research on autonomy in team environments by shifting the focus from individual workload reduction to the broader workload-performance relationship at the team level. By examining how autonomous systems influence team processes and outcomes, we aim to provide actionable insights for evaluating and designing systems that enhance both individual and collective effectiveness.

## METHOD

This work is part of a larger project conducted by the Army Research Laboratory. We use a subset of data here. All procedures were approved by the Army Research Laboratory's Institutional Review Board (IRB), which served as the primary IRB on record, with a signed authorization agreement from a collaborating university IRB located in the western United States. Written informed consent was obtained from all participants.

### Participants

Twenty-eight military personnel were recruited from U.S. Army soldiers. Participation was voluntary and uncompensated. The sample was divided into two teams of 14 that participated in separate two-week blocks. Each team followed the same procedure.

### Procedure

Each team participated in 10 simulated combat missions, at a rate of two missions per day over five consecutive days. Their objective was to navigate a high-fidelity outdoor terrain while engaging and neutralizing enemy forces. The starting points, endpoints, and configurations of enemy forces varied to ensure diversity in task demands and challenges.

The teams were divided into two squads of seven members each. Each squad operated a manned combat vehicle (MCV) and two robotic combat vehicles (RCVs). Within each squad, roles were randomly assigned and remained consistent throughout the experiment: three participants served as drivers, three as gunners, and one as section commander.

Role-specific responsibilities were designed to reflect real-world operational dynamics. MCV drivers controlled their vehicles using a steering yoke and pedals, whereas RCV drivers used a touchscreen map interface to place waypoints, enabling the RCVs to navigate autonomously while avoiding obstacles. Gunners engaged enemy forces, supported by an autonomous target recognition system that highlighted opposing forces on their display. Additionally, gunners used a touchscreen map to mark critical areas of interest, such as bombed buildings or strategic locations. Commanders monitored team status (e.g., vehicle health, task assignments) and directed squad actions through predefined 'plays.' These plays included movement formations for advancing toward objectives and battle drills for engaging enemy forces.

In half of the missions, teams were provided with ADSS designed to enhance task coordination and situational awareness. These systems generated real-time cues to alert team members to events where task switches or play calls might be advantageous. Participants could then use a flexible tasking tool to dynamically reallocate tasks (e.g., switching from driving an RCV to providing situational awareness for another vehicle) as needed. The effectiveness of these ADSS in improving team performance is the primary focus of this evaluation.

## Measures

***Demographics*** Participants were recruited from a pool of U.S. Army soldiers cleared for extracurricular exercises, making prior familiarity among them likely. To account for potential impacts on team coordination, participants rated their familiarity with each teammate on a Likert scale (15). Additionally, participants reported their overall video game experience, number of hours spent playing video games per week, and the relative frequency of playing, as these factors could influence their baseline ability to operate the simulation platform.

***Workload*** Participants completed a battery of questionnaires at the end of each mission. We focus on only a subset in the current study. Perceived cognitive workload was captured using the NASA-TLX (Hart, & Staveland, 1988). In line with the methods in Johnston et al. (2013), we removed item 5 from the NASA-TLX, the item assessing *prediction* of performance. This was done because, as they noted, it "is conceptually distinct from measures of workload traditionally used in cognitive load theory" (p. 257), the theory upon which team cognitive efficiency is based (see Kalyuga, Chandler, Sweller, 1999).

***Team performance*** Since delays in any individual role's performance directly impacted the team's overall mission completion time. We chose total mission duration as a comprehensive metric that reflected both individual role performance and the team's ability to coordinate effectively across their interdependent tasks.

***Communication*** Team members wore individual lapel microphones that captured their speech throughout each mission. Audio recordings were transcribed using WhisperX (Bain et al., 2023), which also identified individual speakers. The transcripts were then cleaned to retain only the primary speaker's speech. We analyzed these transcripts using Linguistic Inquiry and Word Count (LIWC, 2022) to quantify three key categories:

*Insight*. According to Kozlowski & Ilgen (2006) effective teams engage in reflective communication that helps them adapt to dynamic tasks. Therefore, an increase in insight-related language (e.g., "Personally, I don't think that's a good idea here") may suggest that the ADSS facilitates deeper cognitive processing and problem-solving, which are critical for team coordination in complex tasks.

*Perception*. An increase in perceptual words (e.g., "They are on your left") could indicate heightened situational awareness. This would provide evidence that the ADSS allowed team members to redistribute cognitive resources to share information important for overall team coordination.

*Tone*. Positive tone during complex problem solving has been associated with positive team emergent states such as higher team morale, trust, and psychological safety (Edmondon, 1999). Higher positive tone during technology use may indicate greater cognitive bandwidth for social interaction, as teams with sufficient cognitive resources can engage in supportive communication and maintain positive team dynamics. Conversely, teams under high cognitive load often exhibit reduced social communication and more negative tone. Therefore, positive tone can serve as an indirect indicator of cognitive resource availability and team cohesion.

***Team cognitive efficiency*** The team cognitive efficiency score was calculated by taking the difference between the standardized workload and performance scores. Scores can be positive or negative. Positive results indicate comparatively better team performance in proportion to individual reported workload, while negative scores indicate worse team performance compared to workload, allowing us to examine how cognitive workload at the individual level manifests in collective outcomes. This shift in perspective illuminates the value of workload reduction as an independent benefit, while acknowledging its potential to enable conditions that support sustained team effectiveness.

## RESULTS

All analyses were conducted in R (R Core Team, 2021). We started by investigating demographics. For each variable, we conducted a Welch's two sample $t$-test to determine whether the mean responses were significantly different between the two teams. There were significant differences between the two teams in terms of their prior familiarity with each other ($M1 = 2.62$, $M2 = 1.84$, $t(19.15) = 2.65$, $p < .01$) but not in terms of their overall gaming experience ($t(25.60) = 0$, $p = 1$), hours spent playing games per week ($t(25.98) = -0.31$, $p = .76$), or gaming frequency ($t(23.88) = 0.37$, $p = .71$). Because there was a significant difference of familiarity between the two teams, we isolated team as a factor within the model building process to investigate its relative contribution to cognitive efficiency. Specifically, because research shows that team familiarity has an effect on processes and outcomes (e.g., Gruenfeld et al., 1996; Smith-Jentsch et al., 2009), it was included in the hypothesis testing. But the other, non-significant gaming-related demographic variables, were not.

To investigate the relationship between team cognitive efficiency and the use of the ADSS, the results were evaluated through progressive model building to understand the relative contribution of each factor. In each model, team cognitive efficiency was the dependent variable and the unique participant IDs were entered as a random effect to control for the dependence of observations. All models were calculated using R's lme4 package (Bates et al., 2015). All variables were scaled to allow for easier model interpretation, and so beta ($\beta$) coefficients represent the estimated change in the dependent variable for every one standard deviation change in the independent variable (see Table 1 for complete results). The assumptions for mixed-effects modeling were met. Team cognitive efficiency scores were normally distributed (Shapiro-Wilk, $p < .05$), and predictors showed no concerning multicollinearity (all VIFs < 2.0).

Our primary variable of interest, the presence of the autonomous decision support system (Tech), was entered as the sole fixed effect in Model 1 (Table 1). Next, team (one or two) and then role (commander, driver, or gunner) were entered separately to investigate the impacts of the nested structure of the experiment (Models 2 and 3, respectively). Team and role were then entered together to control their combined effects (Model 4). Lastly, role and team were treated as interactions to investigate their unique relationships with team cognitive efficiency (Model 5).

**Table 1:** Hierarchical linear mixed effects models predicting team cognitive efficiency.

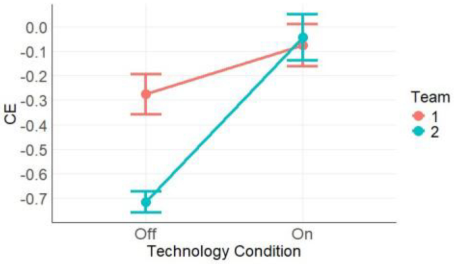
| Predictors | Model 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| (Intercept) | −.58*** | −.46* | −.29 | −.16 | .04 |
| Tech[On] | .54*** | .54*** | .55*** | .54*** | .06 |
| Team | . | −.24 | . | −.24 | −.45* |
| Role | . | . | −.59** | −.59** | −.62** |
| Tech[On] x Team | . | . | . | . | .64*** |
| Tech[On] x Role | . | . | . | . | .11 |
| Marginal $R^2$ | .071 | .100 | .199 | .224 | .245 |

*Note:* Asterisks indicate significance; *** at $p$ <.001, ** at $p$ <.01, and * at $p$ <.05.

The base model (Model 1) showed a significant positive effect of the ADSS ($\beta = .54$, p <.001), indicating that when teams used the system, they performed more efficiently relative to their reported workload (Figure 1).

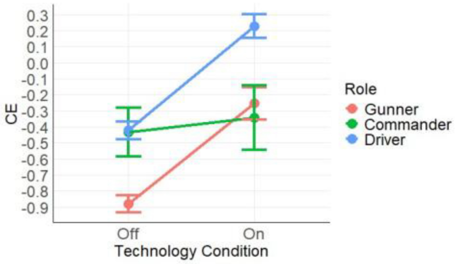
The main effect of technology use remained stable when team differences were introduced as a *fixed factor* in Model 2 ($\beta = .54$, $p$ <.001), reinforcing its robustness. The non-significant team effect ($\beta = -.24$, $p = .334$), and minimal increase in explained variance (56.1 % to 57.6%), suggests team differences had little impact on cognitive efficiency when controlling for technology. Thus, although Figure 1 shows that the less familiar team benefited more, the system's overall benefit was consistent across teams when team-specific factors were not considered.

Model 3 introduced role as a predictor and maintained the significant effect of technology ($\beta = .55$, $p$ <.001), while also revealing a significant effect of role ($\beta = -.59$, $p$ <.01). To investigate the impact of role, we recreated Model 3 by setting the reference category to each role to obtain pairwise comparisons. For each iteration, there was a significant positive effect of technology ($\beta = .47$, $p$ <.0001). Compared to gunners, drivers showed significantly higher cognitive efficiency scores ($\beta = .54$, $p = .011$), while commanders were not significantly different ($p = .38$, $p = .215$). Importantly, there was a non-significant interaction between technology and the role of driver ($\beta = -.07$, $p = .555$). As can be seen with their similar slopes (Figure 2), both drivers and gunners similarly benefited from the ADSS. In contrast, commanders showed a significantly different response pattern ($\beta = -.47$, $p = .013$), with a much smaller benefit from technology. These findings suggest that benefits of the ADSS were unique and consistent between drivers and gunners, but not commanders. Despite this, the overall explained variance decreased slightly (56.7%), indicating role differences added limited predictive value.




*Note.* CE = cognitive efficiency. Team two improvement in CE scores when using the ADSS

**Figure 1:** Interaction between ADSS and team on CE.




*Note.* CE = cognitive efficiency. Drivers and showed greater gunners showed the greatest benefit in CE when using ADSS.

**Figure 2:** Interaction between ADSS and role on CE.

This pattern of results persisted in Model 4 when both role and team were included. That is, there was a significant main effect of technology use ($\beta = .54$, $p < .001$), non-significant effects associated with team ($\beta = -.24$, $p = .334$), and significant effects associated with role ($\beta = -.59$, $p < .01$). The explained variance increased slightly to 57.9%, suggesting that both role and team contributed to the model, but team was the stronger predictor. This reinforces the idea that role-specific factors and team familiarity play a part in shaping cognitive efficiency.

However, the final model (Model 5) revealed a more nuanced relationship between these variables. While the main effect of technology became nonsignificant ($\beta = .11$, $p = .342$), the main effects of team ($\beta = -.45$, $p < .05$) and role ($\beta = -.62$, $p < .01$) remained significant. The interaction between technology use and role indicated that the ADSS's impact was consistent across roles in both teams. However, the significant interaction between technology use and team ($\beta = .64$, $p < .001$) revealed that the benefits of the ADSS varied substantially between teams. While both teams showed improved team cognitive efficiency with the technology, team two demonstrated markedly greater gains, improving from −.70 to .00, compared to team one's more modest improvement from −.27 to −.07 (Figure 1). The explained variance also increased from 57.9% in Model 4 to 59.2%, indicating that the interaction term added meaningful explanatory power to the model. Simply put, these results indicate that while the technology benefited both teams, its impact was stronger for team two; that is, the effect of the ADSS was greater when the team had lower familiarity with each other.

Next, we examined how technology conditions influenced different aspects of communication using separate linear regressions. For each analysis, we used the specific communication variable as the outcome, technology condition as the predictor and team as a fixed factor (see Table 2 for complete results). All variables were scaled to allow for easier model interpretation.

The analyses revealed several significant effects of technology use on communication patterns. Participants used more insightful words ($\beta = .29$, $p < .001$) and had a higher positive tone ($\beta = .21, p < .001$) during technology conditions. However, the use of perceptual words remained consistent across conditions ($\beta = -.04\ p = .68$). Notably, none of the communication variables showed significant team differences, as evidenced by non-significant effects of team ($p > .05$). This indicates that the technology's impact on communication patterns was consistent across both teams.

**Table 2:** Effects of ADSS condition and team on communication variables.

| | Insight | | Tone | | Perception | |
|---|---|---|---|---|---|---|
| **Predictor** | ***β*** | **p** | ***β*** | **p** | ***β*** | **p** |
| Tech (On) | .29*** | <.001 | .23*** | <.001 | .13 | .06 |
| Team | .33 | .62 | .18 | .56 | .04 | .67 |
| | | | | | | |
| Marginal $R^2$ | .43 | | .25 | | .32 | |
| Adjusted $R^2$ | .37 | | .19 | | .26 | |

*Note:* Asterisks indicate significance; *** at $p$ <.001, ** at $p$ <.01, and * at $p$ <.05.

## DISCUSSION

Taking the five hierarchical models together, we conclude that hypothesis one was supported. The ADSS improved team performance outcomes above and beyond the increases to individual workload. This conclusion is supported by the significant main effect of technology in Models 1–4, which consistently showed that when teams used the system, they achieved higher cognitive efficiency scores.

We also conclude that hypothesis two was supported. The ADSS was not as beneficial for commanders, as indicated by the significant negative effects of role in Models 3–5 and the subsequent pairwise comparisons in step 3. This finding highlights the importance of role-specific considerations when designing and implementing such systems. In this context, commanders likely did not benefit from the ADSS to the same degree because, while the AI delivered suggestions that changed how they gathered information, they still needed to critically evaluate these suggestions before distributing them to the team. Importantly, this finding reinforces our earlier point that examining individual changes may not be appropriate in team contexts. The commander was able to facilitate better coordination amongst team members without diminishing their own efficiency, which is a critical insight for system evaluation. While traditional evaluations might focus solely on individual improvement, our results demonstrate that the value of an ADSS should be measured by its impact on the entire team's functioning. Even though the commander's personal efficiency metrics did not improve, the team's performance as a whole did, suggesting the team effectively leveraged the system to enhance team coordination and performance. This emphasizes the need for evaluation frameworks that capture both individual and collective outcomes when assessing technological interventions in team environments.

Additionally, although not hypothesized, the ADSS was more beneficial for team two, as indicated by the significant Tech[On] × team interaction in Model 5. The main effect of technology being non-significant in this model does not negate its overall benefits. Rather, the team that had lower familiarity among teammates (team two), had much lower cognitive efficiency scores when not using the automated decision support. But, when team two had the ADSS, their scores markedly improved and were near equal to team one. Because team one had members significantly more familiar with each other, this suggests that the ADSS was able to overcome coordination challenges brought on by a lack of teammate familiarity. The greater benefit observed by team two underscores the importance of contextual factors in shaping the success of autonomous systems. It also highlights that leveraging autonomous systems in a given context may not be one-size-fit-all, with some teams disproportionately benefiting. This aligns with recent work on Human-AI teams showing that teams with lower collaborative potential performed better when paired with an AI coach, but teams with high collaborative potential did not (Bendell et al., 2024).

Lastly, we conclude that hypothesis three was supported. Team communication processes were different between technology conditions. The findings that teams exhibited more insight-related language and a more positive tone when they used the ADSS suggests improvement in team communication and cognitive processes (e.g., Mathieu et al., 2000; Stout et al., 2000). The increase in insight-related language indicates that the ADSS facilitated deeper cognitive engagement and reflective thinking, which are critical for developing shared mental models and enhancing team reflexivity (Kozlowski & Ilgen, 2006). This aligns with prior research showing that teams that engage in insight-oriented communication are better equipped to analyze complex situations, adapt strategies, and coordinate effectively (Mathieu et al., 2000). The more positive tone observed in team communication further highlights the role of the ADSS in fostering a supportive and collaborative environment. Positive communication is closely linked to psychological safety and trust (Edmondson, 1999), which are essential for open dialogue and effective coordination, particularly in highstakes or dynamic tasks. Together, these findings highlight the ADSS's potential to impact team communication by promoting cognitive depth and positivity, which are foundational to effective team coordination and performance.

The findings from this study offer several actionable insights for practitioners and researchers aiming to implement or evaluate ADSS's in team settings. Understanding the relationship between workload and performance can help stakeholders make informed decisions about autonomous technology implementation. Specifically, our findings suggest that measuring success solely through immediate performance gains may overlook crucial benefits in cognitive resource management, team dynamics, and team effectiveness.

The ADSS demonstrated clear benefits for team performance, particularly in enhancing cognitive efficiency and bridging performance gaps between teams. This suggests that such systems can be valuable tools for improving team outcomes, especially in contexts where performance disparities exist. However, the role-specific challenges faced by commanders highlight the importance of designing adaptable systems that account for the unique demands of different roles. Practitioners should consider tailoring ADSS functionalities to align with role-specific tasks and workflows, ensuring that all team members can fully leverage the technology.

Furthermore, the observed improvements in team communication reinforce the ADSS's potential to foster more effective and cohesive team dynamics. When assessing the impact of similar technologies, it is crucial to evaluate not only performance outcomes but also team processes, as these are key indicators of team coordination and functionality. By focusing on both performance and communication, organizations can better understand how autonomous systems influence team processes and identify opportunities for optimization.

## ACKNOWLEDGMENT

The research reported herein was sponsored by the U.S. Army Research Laboratory and was accomplished under Cooperative Agreement W911NF21-2-0279, and by the Air Force Office of Scientific Research (AFOSR) under Grant Number FA9550-22-1-015, both awarded to the second author. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies or positions, either expressed or implied, of the U.S. Army Research Laboratory, the U.S. Air Force Office of Scientific Research, the U.S. Government, or the University of Central Florida. The Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein.

## REFERENCES

Bain, M., Huh, J., Han, T., & Zisserman, A. (2023). WhisperX: Time-Accurate speech transcription of long-form audio, *INTERSPEECH 2023*.

Bates D, Mächler M, Bolker B, Walker S (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*,67(1), 1–48. https://doi.or g/1.18637/jss.v067.i01

Bedwell, W. L., Salas, E., Funke, G. J., & Knott, B. A. (2014). Team workload: A multilevel perspective. *Organizational Psychology Review*, *4*(2), 99–123.

Bendell, R., Williams, J., Fiore, S. M. & Jentsch, F. (2024). Individual and team profiling to support theory of mind in artificial social intelligence. *Sci Rep* 14, 12635. https://doi.org/10.1038/s41598-024-63122-8

Boyd, R. L., Ashokkumar, A., Seraj, S., & Pennebaker, J. W. (2022). The development and psychometric properties of LIWC-22. Austin, TX: University of Texas at Austin. https://www.liwc.app

Edmondson, A. (1999). Psychological safety and learning behavior in work teams. *Administrative Science Quarterly*, 44(2), 350–383. https://doi.org/1.2307/2666999

Fiore, S. M., Scielzo, S., Jentsch, F., & Howard, M. L. (2006). Understanding performance and cognitive efficiency when training for X-ray security screening. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (pp. 2610–2614). Los Angeles, CA: SAGE Publications.

Funke, G. J., Knott, B. A., Salas, E., Pavlas, D., & Strang, A. J. (2012). Conceptualization and measurement of team workload: A critical need. *Human Factors*, *54*(1), 36–51.

Gruenfeld, D. H., Mannix, E. A., Williams, K. Y., & Neale, M. A. (1996). Group composition and decision making: How member familiarity and information distribution affect process and performance. *Organizational Behavior and Human Decision Processes*, 67(1), 1–15.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), *Human mental workload* (pp. 139–183). North-Holland. https://doi.org/1.1016/S0166-4115(08)62386-9

Johnston, J., Fiore, S., Paris, C. R., & Smith, C. A. (2013). Application of cognitive load theory to develop a measure of team cognitive efficiency. *Military Psychology, 25*, 252–265.

Kahneman, D. (1973). Attention and effort. Englewood Cliffs, NJ: Prentice-Hall.

Kalyuga, S., Chandler, P., & Sweller, J. (1999). Managing split-attention and redundancy in multimedia instruction. *Applied Cognitive Psychology*, 13(4), 351–371.

Kozlowski, S. W. J., & Klein, K. J. (2000). A multilevel approach to theory and research in organizations: Contextual, temporal, and emergent Processes. In K. J. Klein, & S. W. J. Kozlowski (Eds.), *Multilevel Theory, Research, and Methods in Organizations* (pp. 3–90). Jossey-Bass.

Kozlowski, S. W. J., & Ilgen, D. R. (2006). Enhancing the effectiveness of work teams and teams. *Psychological Science in the Public Interest, 7*(3), 77–124. https://doi.org/1.1111/j.1529–1006.2006.0003.x

Marks, M. A., Mathieu, J. E., & Zaccaro, S. J. (2001). A temporally based framework and taxonomy of team processes. *The Academy of Management Review*, *26*(3), 356–376. https://doi.org/1.2307/259182.

Mathieu, J. E., Heffner, T. S., Goodwin, G. F., Salas, E., & Cannon-Bowers, J. A. (2000). The influence of shared mental models on team process and performance. *Journal of Applied Psychology, 85*(2), 273–283.

Paas, F. & Van Merriënboer, J. (1993). The efficiency of instructional conditions: an approach to combine mental effort and performance measures. *Human Factors, 35*, 737–743.

Paas, F., & Tuovinen, J. (2004). Exploring multidimensional approaches to the efficiency of instructional conditions. *Instructional Science, 32,* 133–152.

Paas, F., Tuovinen, J., Tabbers, H., & Van Gerven, P. W. M. (2003). Cognitive load measurement as a means to advance Cognitive Load Theory. *Educational Psychologist, 38*, 63–71.

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www. R-project.org/

Smith-Jentsch, K. A., Kraiger, K., Cannon-Bowers, J. A., & Salas, E. (2009). Do familiar teammates request and accept more backup? Transactive memory in air traffic control. *Human Factors, 51*(2), 181–192.

Stout, R. J., Cannon-Bowers, J. A., Salas, E., & Milanovich, D. M. (1999). Planning, shared mental models, and coordinated performance: An empirical link is established. *Human Factors*, *41*(1), 61–71.

Wickens, C. D. (1984). Processing resources in attention. In R. Parasuraman & D. R. Davies (Eds.), Varieties of attention (pp. 63–102). Orlando: Academic Press.