Teamwork Assessment and Development: Methodological challenges and solutions

Leonie V. Webster, Abby E. Laishley, Carole Deighton
Defence Science & Technology Laboratory (Dstl)
Salisbury, UK
lwebster@dstl.gov.uk, aelaishley@dstl.gov.uk, cdeighton@dstl.gov.uk

ABSTRACT

Teams have long been the object of scientific enquiry given the central role they play in complex, safety critical, innovative and impactful work. Understanding how the most effective teams function is complex and multi-faceted. Whilst the science of teams has established a broad and deep knowledge base, there is an overreliance on theoretical models that do not account for the dynamic nature of teams (Ramos-Villagrasa et al., 2018). Moreover, there is a need for the science to evolve to understand teams in a new era; one characterised by the explosion of novel technologies likely to change the way teams interact, and the means by which these interactions can be measured (Benishek & Lazzara, 2019).

Many current methods of teamwork measurement are static (measuring only at a single point in time), thus are not reflective of the dynamic and changing nature of teams. Furthermore, the literature is reliant on subjective, self-report data, or the use of observer-raters who may disrupt natural team functioning. Therefore, this paper will present three promising methods of capturing meaningful data related to team behaviours utilising technological approaches: 1) real-time communication data, 2) social network analysis (SNA), and 3) wearables and sociometric badges.

Each of these are discussed in turn, identifying applied sciences related challenges such as usability, validity, and the analysis and interpretation of large amounts of data, before potential solutions to these challenges are offered. For each technology explored, reference is made to contemporary studies, commissioned by UK MOD (e.g., Roberts et al., 2019a; Pleva et al., 2021; Myers et al., 2021[1]), to support the discussion. The paper is rounded off by drawing insights to inform further research opportunities.

ABOUT THE AUTHORS

Dr Leonie Webster is a Cognitive Psychologist specialising in maximising cognitive human performance at the Defence Science and Technology Laboratory within the United Kingdom Ministry of Defence. Leonie's principal fields of research interest lie in teamwork, leadership, and communication in high performance environments, having worked in both Defence and elite sport researching and applying these concepts.

Dr Abby Laishley is a Cognitive Psychologist specialising in learning and development at the Defence Science and Technology Laboratory within the United Kingdom Ministry of Defence. She currently leads research activities that explore learning ecosystems and data enabled training, and evaluates the efficacy, effectiveness, and value of training and education across the Defence Enterprise.

Dr Carole Deighton is an Applied Psychologist at the Defence Science and Technology Laboratory within the United Kingdom Ministry of Defence specialising in competence retention 'Skill Fade' at the individual, team and collective levels. She has held technical leadership roles across academia, defence, industry and consultancy, primarily in the areas of training and education and human factors / integration.

[1] Text and diagrams that relate to work undertaken within these studies are reproduced with kind permission of the University of Southampton, LiMETOOLS, the University of Chichester, and Cervus Defence & Security Ltd.

Teamwork Assessment and Development: Methodological challenges and solutions

Leonie V. Webster, Abby E. Laishley, Carole Deighton
Defence Science & Technology Laboratory (Dstl)
Salisbury, UK
lwebster@dstl.gov.uk, aelaishley@dstl.gov.uk, cdeighton@dstl.gov.uk

INTRODUCTION

Teams are fundamental to the success of military missions; required in tasks ranging from securing locations, operating equipment, to providing strategic direction and managing large military operations (Goodwin et al., 2018). In order to ensure that teams operate most effectively it is important to draw on the burgeoning field of team science to understand how best to manage, improve and sustain teamwork and team performance. Fundamental to the capacity to train and improve teamwork skills is the ability to measure the constructs of greatest significance (Salas et al., 2009).

Despite the science of teams and teamwork progressing considerably over the last century, the knowledge and understanding procured from this research is applied inconsistently across the military. Military teams face the added challenge of working in dramatically changing environments, organisations and systems (Kolbe & Boos, 2019); changes that are the result of the increasingly technology-centric nature of work (Walker et al., 2017). In this context there is a need to integrate emerging technologies into the measurement and development of teams.

In regards to the measurement of teamwork, self-report instruments, in particular, predominate the literature. It is widely accepted that teams are dynamic entities, yet empirical team research has used questionnaires and surveys to examine team states at static moments in a team's lifecycle (Cronin et al., 2011). Static survey measurement fails to capture how team processes emerge and change over time, thus any evaluation of teamwork based on such measures fails to accurately represent the status of the team. Although there are certain instances where self-report measures are appropriate (e.g., assessing team members' subjective perspective of team functioning for the purpose of reflection), such measures require aggregation of individual level responses to team-level characteristics. The use of mean-based variables derived from aggregation often oversimplify group-based phenomena and result in biased and unequivocal findings (e.g., Dineen et al., 2007). In addition, individuals may rate themselves and their team more favorably that an observer might (Marlow et al., 2018a). As such, overreliance on self-report measures risks inaccurate conclusions.

Observational measures circumvent some of the limitations of self-report measures, yet are also subject to rater biases (Kahneman et al., 2021) and are likely to interfere with the natural functioning of a team. In order to keep pace with the increasingly technology-centric nature of work, and capture the dynamic nature of team functioning, organisations need to engage in more labour-intensive methods of measurement (i.e., laborious due to the required data processing and analysis; Kolbe & Boos, 2019). For example, researchers have attempted to assess teams in naturalistic settings through the creation of synthetic worlds and computational simulations to measure team performance without the interference associated with observations (Chapman & Colegrove, 2013; Kozlowski et al., 2016). Aligned with this is an exploration of unobtrusive approaches to measure team performance through alternative data sources. The real-time data streams available to researchers include: 1) behaviours, 2) words, and 3) physiological responses (Luciano et al., 2018). This paper presents three unobtrusive methods of measuring team-related variables that aim to capture one or more of these data streams (communication data, social network analysis (SNA), and wearable sensors), providing contextual examples to demonstrate how these techniques have been applied in military based scenarios and simulations. These case studies do not necessarily reflect best practice per se, rather they do serve to highlight some of the opportunities and challenges faced when adopting such measurement approaches.

REAL-TIME COMMUNICATION DATA

Communication is regarded as one of the most critical team behaviours (Marlow et al., 2018b). That is, teams are dependent on effective communication for successful coordination, cognition, and performance outcomes (Cooke et al., 2013). Words are an important, dynamic data stream in the study of teams. Thus, communication lends itself well to unobtrusive measurement, as it can be automatically and continuously recorded during team tasks (Stanton &

Roberts, 2019). Although this measurement approach is not new, the advent of modern technology has enabled transcription and coding of communication data to progress from being painstakingly performed by humans, to being performed (in an increasingly reliable manner) by computers (i.e., computer aided text analysis; Yilmaz, 2016).

The analysis of communication can pertain to the physical properties of speech (i.e., frequency, duration, volume), content (what is being said), or sequential flow of information exchange between team members (Kiekel et al., 2002). Research has failed to demonstrate a consistent relationship between the frequency of communication and team outcomes (Marlow et al., 2018b), which suggests that analysis of the content and flow of communication might provide greater insight into team dynamics.

Team communication processes have also been studied to understand the concept of team cognition: the way in which the team thinks, remembers, makes complex decisions, and solves problems (Cooke et al., 2013). Cooke and colleagues suggest that social interaction processes constitute a critically important aspect of cognition, thus put forward the theory of Interactive Team Cognition, which can be monitored by focusing on real-time communication links between members (Cooke et al., 2007; 2013). Research has found that interactionist-based measures such as communication and coordination are better predictors of team performance, or team skill retention, than are aggregates of the component-based measures (Cooke et al., 2007; Gorman et al., 2006).

Technological advancements allow researchers to analyse large streams of communication data in a multitude of different ways. However, the use of this data to make conclusions about teamwork generally requires content-driven analysis, which involves human pre-processing (Klonek et al., 2019). Although progress is on the horizon for automatic or machine learning for obtaining team measures from communication data (e.g., Bonito et al., 2018), researchers must recognise that these techniques still need further development, with an emphasis on improving reliability. It is also recommended that rigor is increased by linking communication data with more traditional, validated measurement approaches (Khaleghzadegan et al., 2020). Whilst there is certainly a degree of messiness inherent in these measurement approaches, it is an exciting era for team-based research.

Case Study 1

The case study referenced in this section was designed to test the feasibility of a novel, immersive and personalised learning progression tool for Defence. Whilst specific measures of teamwork were not captured as part of the research, the case study provides a good example of relevant simulation and learning environments within which the technologically advanced measures discussed could be readily applied. Communication data (both verbal and written) was captured throughout the simulation, therefore the scenario described could be readily applied to future study of team-related properties through communication analysis.

LiMETOOLS Ltd, on behalf of Dstl, designed and trialled a video-based, online role-playing training exercise (Talya 2025) that simulated a real life future crisis management scenario (Pleva et al., 2021). Talya 2025 was designed to assess military decision making, leadership and communication, where measures of team performance could be inferred from real-time communication, information analysis and agile planning evaluation. The remote exercise progressed against the clock over a one-day workshop with five team members, playing out an accelerated exercise narrative of a power struggle between two fictional bordering geographical regions. A political coup triggered urban warfare, requiring the team to anticipate and mange street battles, resource sabotage, civilian hostages, cyber hacking and mobility blockades. Team members received an overwhelming amount of conflicting intelligence to test information analysis, with information being presented in transmedia methods such as video news desk reports, written updates from commanding organisations, video and written social media drama, audio calls and text messages. Each team member acted within a prescribed military role and assumed leadership of a team discussion and decision outcome at least once during the exercise. Communication took place through a shared whiteboard space and an ongoing audio call with written text chat. Throughout the exercise, teams were asked to make decisions on their course of action towards resolving conflict and a successful evacuation. The output of team decisions was automatically assessed within the online platform based on a numerical score of correct answers.

An experimenter facilitated the exercise narrative and captured observational data relating to individual and team performance by completing a structured monitoring report during and after decision points. This enabled a record of team communication from the text and audio chat, and presented an opportunity to assess leadership and influence behaviours in communication. Real time observations were captured with a structured focus on team member interactions, levels of engagement, possible team conflict, and the way in which information was sourced and disseminated. There were some limitations to this method, as it was somewhat labour intensive and did not address some of the aforementioned limitations inherent in observational data. However, the observations took place in a computational simulation with minimal interference from the observers, and with the intention of triangulating these data with future analysis of communication data.

Whilst further work is required to understand whether analysis of communication data procured through the scenario could provide insight into team functioning, feedback from participants suggested there is an appetite for experiential and collaborative learning techniques in defence. Such approaches provide the ideal context from which to measure and develop teamwork in an experiential manner, and it is hoped that Talya 2025 is exploited in future research efforts. However, having the ability to tailor bespoke content for organisations, schools or Service is a necessity.

SOCIAL NETWORK ANALYSIS

The seminal meta-analysis by LePine et al. (2008) firmly established the significant, positive association between teamwork processes (i.e., goal specification, monitoring and coordination) and team performance, cohesion, and potency. In concluding this research, LePine et al. suggested that team process data may be available through non-traditional means, advocating the use of recorded forms of member interactions to infer teamwork processes. One such non-traditional method that is increasingly being utilised in the study of team processes is SNA. SNA is a set of methodological techniques that aim to describe and explore patterns that emerge in the social interactions between individuals and groups (Wasserman & Faust, 1994). As such, social networks provide a unique insight into the fabric that binds team members to one another (O'Neill & Salas, 2018). Where traditional measurement approaches assess shared team properties, SNA is concerned with patterns, distributions, and variability (Crawford & LePine, 2013).

SNA has previously been applied to the study of team-related constructs to examine the effect of leadership networks (e.g., Carter et al., 2015), the effect of individual stress on communication network dynamics (e.g., Kalish et al., 2015), the network structures associated with interpersonal trust (Ferrin et al., 2006), and the effects of team conflict on network structure (Park et al., 2016). A benefit of the network analysis approach is that it enables the calculation of a wealth of network metrics, affording empirical investigation of how independent variables (e.g. a teamwork intervention) affect the nature of the network.

Case Study 2

A piece of research undertaken by the University of Southampton on behalf of Dstl sought to apply SNA to the measurement of teamwork by assessing whether the social networks of teams changed following delivery of teamwork training (Roberts et al., 2019a). The research was conducted within a submarine control room simulator based upon an operational Royal Navy submarine. The simulator consisted of five networked workstations that were running a naval warfare simulation game developed by Sonalysts Combat Simulations (Dangerous Waters©). The workstations accommodated two sonar operators (SOP1 & SOP2), two target motion analysts (TMA1 & TMA2) and one officer of the watch (the coordinator and designated team leader). An overview of roles and main duties can be found in Table 1. Twenty, newly formed teams of five participants were recruited. Each participant was randomly allocated a role within the team, and received training specific to this role before the team came together to complete a submarine command and control simulation (T1). Half of the teams then received a multi-method teamwork training intervention (i.e., education, simulation, and debrief) based on the Big Five model of teamwork (Salas et al., 2005), whilst the other half completed a series of individual tasks before all teams completed a second submarine simulation (T2). Communication data, used to compute the social networks, were captured throughout both simulations.

Table 1: Overview of Command Team Roles and Main Duties

Role	Overview of main duties
SOP	Sonar operators monitored the sonar arrays to detect potential contacts, used narrowband acoustic data to determine the classification of contacts and provide an estimate of speed.
TMA	Target motion analysts received bearing information on contacts automatically once designation completed by SOPs. This information was used to generate contact solutions (predict the behaviour of contacts) by analysing patterns of bearing cuts. The TMA operators manipulated a 'ruler' which represented the historical path of the contact in order to plot the

	estimated behaviour of all contacts designated. Solutions for contacts required bearing, course, speed, and range information to be relayed from SOPs.
Coordinator	The coordinator was responsible for directing submarine activity through interpretation of the tactical picture displayed on the map on their interface. The tactical picture map displayed all contacts and solutions once entered by the relevant operators. The coordinator was responsible for quality checking and ensuring that pertinent information was transferred appropriately.

To investigate whether various types of communication could be representative of teamwork processes, a unique, custom-built computer interface was designed for participants to interact with when communicating with team members (see Figure 1). Communication requests were submitted through this interface, in turn prompting the respondent to accept or reject the communication request. Acceptance would result in the entry of all invited members into a networked communications forum, within which they were required to communicate via headsets.

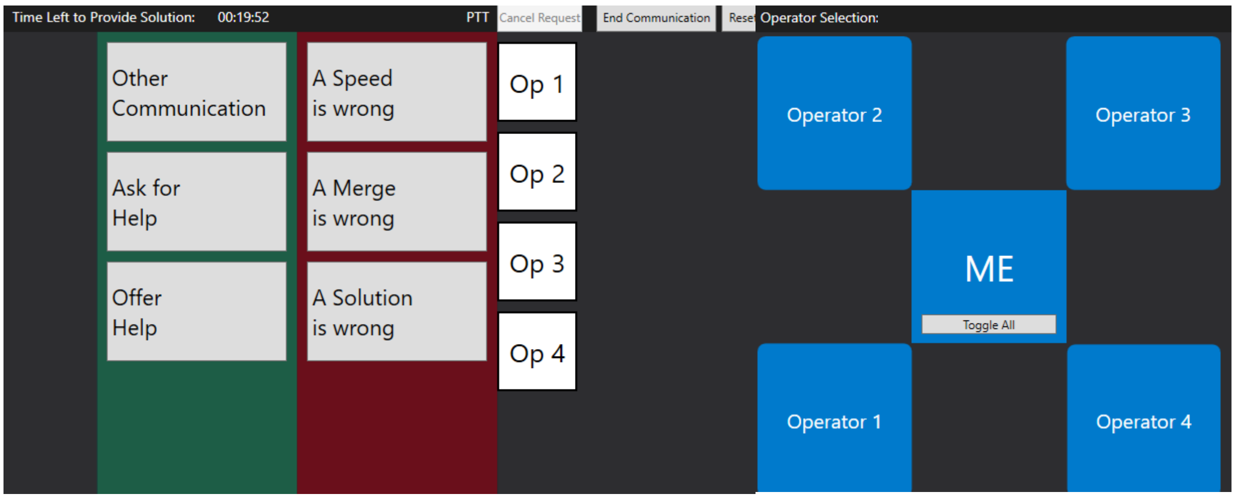

Figure 1: Example of communications interface used to collect embedded measures

It was reasoned that 'Ask for help' could be representative of affective teamwork processes such as mutual trust, 'Offer help' representative of behavioural processes such as backup behaviour, and communicating with all operators may infer cognitive processes such as the building of shared mental models. SNA was therefore performed by creating different networks for each of the different types of communication. Table 2 provides definitions and examples of the various metrics that are analysed as part of this process.

The construction of networks requires the generation of static adjacency matrices[2] derived from the initiations of communications between operators using the panel below. To generate the social networks, all instances of communication initiation between operators were automatically logged and a frequency count of communications between operators was compiled in adjacency matrices for each team across each simulation. A number of metrics can be calculated to facilitate understanding of network composition in terms of overall structure (global metrics) and the individual nodes contained in the network (nodal metrics). Further explanation of these metrics can be found in Table 2. Statistical analyses were performed to test for significant differences between networks at T1 and T2 and between experimental and control teams.

Table 2: Definitions of Global and Nodal Social Network Metrics (Roberts et al., 2019b)

Metric	**Definition**	**Example**
	Global	
Nodes	Number of entities in a network.	Each node represents an operator in the network (i.e., a team member).

[2] An adjacency matrix represents all potential combinations of directed and weighted communications between agents (operators). Rows and columns represent agents (operators), and the presence of a communication is given by a numerical value.

Edges	Number of pairs of connected entities.	A communication exchange between team members is an edge.
Density	Number of relations observed represented as a fraction of the total relations possible.	If the coordinator communicates with all team members the density of that network would be greater than if they communicated with only a single member.
Cohesion	Number of reciprocal connections in network divided by the maximum possible connections.	If all communications between team members were reciprocal (i.e., information relayed by one team member is confirmed by the other) the network would have a high level of cohesion.
	Nodal	
Emission	Number of links going from each node in the network.	Each communication from the coordinator to any other team member counts as one emission.
Reception	Number of links going to each node in the network.	Each time the coordinator receives communication from any other team member counts as one reception.
Sociometric	Number of emissions and receptions relative to number of nodes in the network.	If the coordinator communicates most with all of the team and receives the greatest amount of communications back they will have the highest sociometric status.
Betweenness	Number of times a node lies on the shortest path between other nodes.	If the coordinator acts as a broker of information between SOP1 and TMA1, 'betweenness' identifies how many times the coordinator 'bridges' the SOP1 and TMA1 nodes.

Key Findings

Figures 2-4 depict the change in the networks of both control (left) and experimental (right) teams from T1 to T2 for each of the different types of communication: overall communication, help requests, and offers of help. Dashed lines indicate a reduction in average frequency of emissions and/or receptions between operators, whereas solid lines indicate an increase. The thicker the line, the greater the magnitude of change. Overall, visual inspection of these figures suggests that social networks of teams changed as a result of the teamwork intervention. In Figure 2, the overall communication, it appears that the total number of communications between SOP and TMA operators increased more in teams that received the intervention than those who did not (i.e., those in the control group that worked on individual tasks). As alluded to in Table 1, successful completion of the simulation relied on these operators exchanging critical information. Communication from coordinator to TMAs appeared to decrease in control teams, where this was not the case with experimental teams.

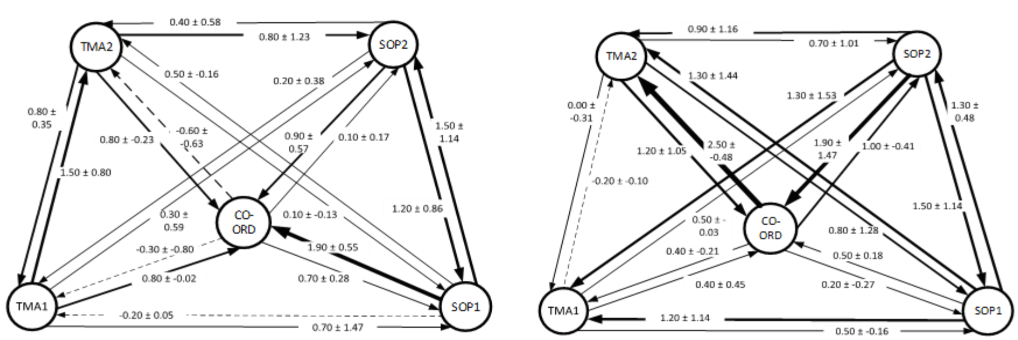

Figure 2a. Control teams mean difference in communication from T1 to T2

Figure 2b. Experimental teams mean difference in communication from T1 to T2

Figure 2. Means and standard deviations of difference between Time 1 and Time 2 emissions and receptions between operator for total communication requests

In the 'asking for help' networks, the metric of cohesion in experimental teams increased significantly more from T1 to T2 than it did in control teams. Essentially, members in teams that received the teamwork intervention were more likely to reciprocate requests for help. Figure 3 reveals the reciprocal requests between the TMAs and SOPs of experimental teams, whereas in control teams operators performing the same role seemed to be asking each other for help (i.e., SOP to SOP). The help request button on the communications interface was used as a proxy measure of mutual trust, in accordance with the definition of trust as a "willingness to accept vulnerability based on positive expectations of a specific other" (Fulmer & Gelfand, 2012, p. 1174). The act of asking for help arguably requires the acceptance of vulnerability in demonstrating a lack of knowledge, understanding, or inability to perform ones' own tasks without assistance. This finding represents the process of reciprocal trust, whereby when a trustee realises a trustor has taken a risk in trusting them, they tend to be motivated to behave in a correspondingly trustworthy manner (Serva et al., 2005). Furthermore, teams that received the intervention made more direct requests for help, and utilised a greater proportion of the network capacity for help requests at T2 (evident in a greater increase in edges and density). It is reasonable to surmise that the teamwork intervention had a positive impact on the trust of experimental teams.

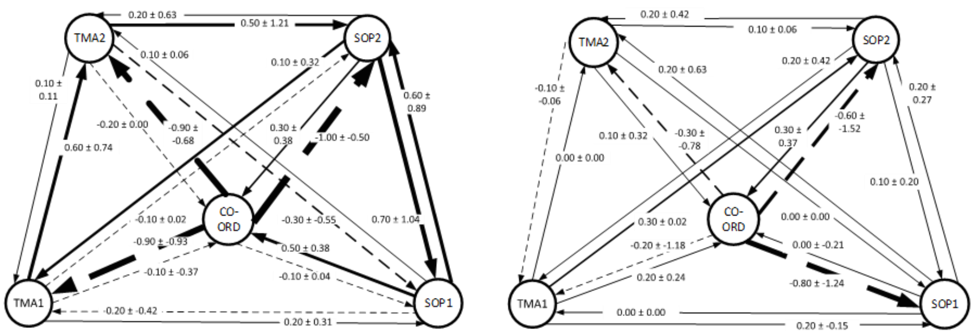

Figure 3a. Control teams mean difference in communication from T1 to T2

Figure 3b. Experimental teams mean difference in communication from T1 to T2

Figure 3. Means and standard deviations of difference between Time 1 and Time 2 emissions and receptions between operators for 'ask for help' initiations[1].

Analysis of the 'offer help' networks revealed that experimental teams with the teamwork training intervention had higher emission, receptions and sociometric status than the control teams at T2. Figure 4 reveals that experimental teams increased offers of help more than control teams, evidenced by the solid, bold lines. This is particularly pronounced for the communication between SOP1 and the TMA operators, potentially indicating greater sharing of information necessary for TMA operators to complete their taskwork. Conversely, in control teams SOP1 appeared to reduce offers of help from T1 to T2, as did the coordinator. This suggests more effective patterns of communication in the experimental teams. 'Offer help' was used as a proxy measure of back up behaviour; the ability to anticipate team members' needs and shift and balance workload (Porter et al., 2003). These results suggest that the teamwork intervention was successful in increasing back up behaviour in the experimental teams at T2.

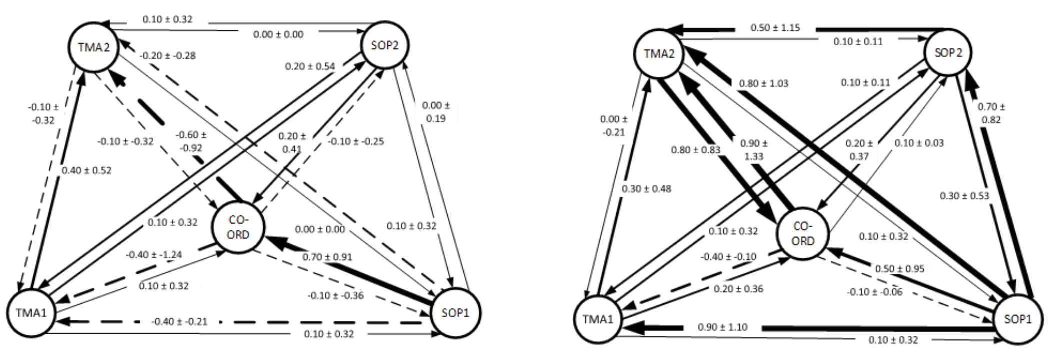

Figure 4a. Control teams mean difference in communication from T1 to T2

Figure 4b. Experimental teams mean difference in communication from T1 to T2

Figure 4. Means and standard deviations of difference between Time 1 and Time 2 emissions and receptions between operators for 'offer help' initiations[1].

Overall, the social networks of team communication changed as a result of the teamwork intervention. Requests for help became more reciprocal, and the frequency of offering help increased. More direct communication, i.e., asking for information/help from the most appropriate members, was evident in experimental teams post-intervention. This indicates that there was more targeted, task-relevant communication in experimental teams. These results provide tentative evidence that the understanding of teamwork can be enhanced through investigation of how teams utilise teamwork behaviours, how they interact with one another, and what information is exchanged and with whom. Teams trained in teamwork demonstrated greater trust and back up behaviour as inferred by the social networks, which intimates that this approach could be suitable to employ in teamwork development research in the future. SNA offers significant benefits as it is semi-automated and provides more objective data than 'traditional' measures of teamwork. An alternative, future, approach to analysing such data could be through the application of epistemic network analysis (ENA; Shaffer et al., 2009). ENA provides a means of quantifying qualitative data by combining principles of both SNA and discourse analysis. Analysis of team communication data using ENA could be employed to identify differences between high and low-performing teams (e.g., Sullivan et al., 2017).

However, a number of challenges relating to the use of SNA within this case study arose. Specifically, the custom interface designed may not have reflected the communication preferences of the operator. Many communications were missed due to technical failures or unwillingness of the operator to correctly interact with the communications panel. The requirement to conduct all communication through the technical interface may have constrained communication and prohibited more organic, informal exchanges. Therefore, collection of data needs to be integrated into normal ways of working. In addition, the time and expertise required to analyse social networks remains a significant obstacle. The capture of task or information networks (as opposed to just social networks) involves laborious transcription processes, and the computation of social networks requires lengthy transformation of data using bespoke pieces of software. Whilst this may be practicable for trained researchers, this is not the case for military personnel. Therefore, SNA needs to become more user-friendly before it can be readily applied in the field. The communication social networks of teams have been analysed increasingly over the last ten years, so it is likely that the technology will continue to advance. Wearable technologies have also been used, but have been more challenging to validate. They could, however, offer a lot of potential beyond the description of communication flow that is capture in SNA (i.e., with the ability to capture and process data faster without the need for neurolinguistic processing technology).

WEARABLES

Recent and ongoing technological advancements in mobile computing and wearable sensors provide the opportunity to collect objective, high-resolution data related to social interactions over extended periods. These advancements address many of the limitations of traditional measurement approaches (Chaffin et al., 2017). Wearable devices have largely been applied to the measurement of physiological parameters (e.g., Friedl, 2018), with a host of commercial wearables available that measure conventional physical signs such as heart rate, body temperature and activity patterns. The sociometric badge, on the other hand, is a specific wearable sensor that records the environmental context of the device-bearing person (Chaffin et al.). Sociometric badges gather inherently interpersonal information such physical proximity, face to face positioning, body movement and posture, as well as using microphones to measure verbal activity (Olguin et al., 2009). Wearable sensors have the advantage of collecting real-time, objective data from participants in a way that is unobtrusive and should not interfere with natural teams functioning. They allow for the application of team level analyses that are more organic and reliable than self-report and observation (Cheung et al., 2017). However, these sensors, in themselves, are of limited value to the study and improvement of teams (Sawka & Friedl, 2017). The true value in wearable monitoring systems lies in the algorithms that convert data into useful and actionable information for optimising performance (Piwek et al, 2016).

Wearables have been applied to examine affect and team cohesion in simulated space exploration missions (Zhang et al., 2018), cooperation (Taylor, 2013), communication in productive and creative teams (Pentland, 2012), social and task-related exchanges (Matusik et al., 2019), social networks (Wu et al., 2008), and emergent leaders (Chaffin et al., 2017). Additionally, Zhang et al. (2018) used sociometric badges to understand which behaviour features extracted from wearable sensors are linked to perceived cohesiveness – the shared attraction that drives team members to stay together and want to work together (Casey-Campbell & Martens, 2009). The study found that group perception of cohesion was influenced by mirroring features that were extracted from the sociometric data. This refers to the similarity of two individuals' behavioural patterns over time. Overall, Zhang et al. demonstrated that group-level aggregation of behaviour features extracted from wearables can be effective in assessing group cohesion.

Planned Study 3

The application of data analytics in professional sport is commonplace, and has been for a period of time. It is, therefore, surprising that the application of analytics to assess the performance of teams has not yet been readily applied in the military. Some wearable devices may be used by certain organisations to monitor fitness, but the true opportunities offered by the data collected through wearable devices has yet to be fully exploited. Therefore, a research study currently underway at Dstl, conducted by the University of Chichester and Cervus Defence and Security Ltd. (Myers et al., 2021), aims to use the data procured from wearable devices to provide information on performance to Commanders. The aim of this research is to investigate whether a range of physiological, cognitive and psychophysiological parameters measured via wearable devices can be predictive of individual and team performance during a field exercise. The project will collect a range of baseline data a month prior to a premier British Army patrolling event, immediately prior to, during and post the event, as well as 5-10 days following the exercise. The measures will include observational ratings provided by military directing staff, self-report questions, measures recorded by worn devices, and stress biomarkers using non-invasive methods. To assimilate all of these data, the intention is to produce visualisations driven by descriptive and statistical analyses.

Some of the physiological data captured through wearables could also be used to infer teamwork metrics (e.g., markers of stress and psychological pressure). Stress can result in reduced communication in teams (Sexton & Helmreich, 2000), impair collective cognitive functioning (Wallenius et al., 2004), and reduce information sharing (Wetzel et al., 2006). Heart rate (HR), heart rate variability (HRV) and galvanic skin responses are all stress-related measures that can be collected from wearable sensors (Mozos et al., 2017). By measuring collective stress responses in a team environment it is possible to investigate the team dynamics across the course of stress episodes. In addition, more recent research suggests that physiological measurement could address the call for more dynamic and unobtrusive measures of team processes and performance (Funke et al., 2012; Kozlowski & Chao, 2018). The continuous assessment of individual members' physiological states whilst performing as a team is referred to as team physiological dynamics (TPD; Kazi et al., 2021). Although TPD measures variables within each physiological subsystem (both central and peripheral nervous systems), it is measurement of the cardiovascular system that lends itself particularly well to the use of wearable sensors.

The wearables in the planned research will capture the volume and intensity of physical activity, sleep duration, HR, HRV, estimated distance, speed and activity type, total distance, time spent at various speeds, and posture. A saliva sample will be taken to measure the stress biomarkers of cortisol and dehydroepiandrosterone sulfate (DHEA-S). Subject matter experts will rate the performance of individuals and teams across a range of pre-determined criteria. Further detail on the measurement approach is provided in Table 3 (NB. The table does not list all the measures that will be collected (i.e. physiological measures), but are those of relevance to studying teamwork and team performance). Statistical analysis will examine relationships between the predictor variables (i.e., outputs from the wearables and other measurement tools) and team performance, as well as the teamwork dimensions as assessed through observer ratings. It is anticipated that the results will indicate which of these variables can most accurately predict team performance and teamwork.

Table 3: Measurement Schedule for Patrol Exercise.

Session	Measures
Baseline	Self-report background questionnaire 24-h Heart rate (HR/HRV)* Cognitive tasks Resilience questionnaire Saliva sample
Pre-Patrol	HR/HRV* Cognitive tasks Situational awareness task Inventory Cognitive and Somatic Anxiety Ratings of individual & team challenge & threat
Patrol	HR/HRV* Physical Activity* GPS speed & distance* Ratings of individual & team challenge & threat Ratings of individual & team performance
End-Patrol	Cognitive tasks Situational awareness task
Post-Patrol	Self-report assessment of teamwork dimensions The Team Emergency Assessment Measure

	Team Communication Visual Analogue Anxiety Scale Differentiated Transformational Leadership Inventory

* Denotes measures collected by wearable devices.

The use of wearable sensors in team research is evidently a promising avenue of measurement that has the potential to capture and reflect the dynamic nature of teams through continuous monitoring. However, many would argue that sensor-based measurement is best placed *augmenting* rather than replacing existing approaches, as there are still many challenges, and much to learn. First, the privacy and security of personal data generated by consumer wearables remains problematic and needs to be addressed through regulatory frameworks and ethical/legal guidelines that are not yet in place. Second, it is important to ensure participant compliance in wearing the sensor throughout the research period by making sure not to add significant weight or require regular battery recharging. Reduced size, weight and power is critical to soldier acceptability and tactical usability (Friedl, 2018). Third, the reliability and validity of wearable devices is somewhat concerning. Commercial devices, in particular, provide no empirical evidence to support the effectiveness of their products. Indeed, Endedijk et al. (2018) observed that the malfunctioning of one sensor obstructs the computation of all team interaction dynamics which makes unreliable technology problematic. There is a need, when working with such measures, to conduct pre-studies to determine the reliability and usability of the measure. Due diligence is required by the researcher to select the most valid, reliable, and appropriate device for the intended research. Chaffin et al. (2017) recommend that researchers evaluate whether sensor type, attachment location, and mode (e.g., lanyard, wristband, etc.) align well with the source and nature of the behavioural signal to be captured. Finally, researchers also need to have specific expertise in data cleaning and dealing with noise. Therefore, although wearable technology is considered to be less labour intensive than traditional methods of measurement, this is not yet necessarily a reality that can be implemented with validity/reliability in everyday teamwork environments.

DISCUSSION

Military teamwork is highly dynamic and emergent. This paper reinforces that these characteristics cannot be assessed comprehensively using traditional static approaches alone and that the exploration of unobtrusive ways to measure teamwork performance is required. Technological advancement affords the opportunity for the development and implementation of new approaches and three studies examining the capture of real time communication data, SNA and the use of wearable technologies have been outlined in this paper.

The adoption of unobtrusive approaches by the military, alongside other subjective and observational-based measures, has potential to enhance the outcomes and value to be gained from future team and collective training events conducted across a blend of live and virtual training environments. Realising this potential requires investigations that take into consideration, not only the scientific validity/reliability of such approaches, but also a broader set of people, ethical/legal, doctrinal and infrastructure related considerations. For example, the ethical collection, analysis and use of personal data captured using wearable technologies; and the resourcing, training and education of personnel who are responsible for the planning and evaluation of teamwork training and exercise events. This is a significant undertaking and will require a focused and collaborative effort across militaries, industry and academia.

KEY TAKEAWAYS

1) Challenges in the measurement of teamwork and team performance are evident, to a greater or lesser extent, for *all* types of measurement approaches. Many of these challenges can be overcome, or managed by applying a multi-method approach, utilising data from various, complementary sources (Dubrow et al., 2017).
2) When selecting teamwork measures rigorous consideration should be given, not only to the properties of the measures (e.g. validity, reliability, sensitivity, diagnosticity), but also to their proposed application environment or context (e.g. "live" versus "synthetic" environment).
3) Consideration should be given to the *depth* of evidence that is realistically required to inform decision-making (such as decisions relating to the readiness of a ship's company to "go to sea"), and the *return on investment* of applying particular types and combinations of measures.
4) Emerging technology-based approaches to the measurement of teamwork and team performance, particularly involving the collection, analysis, management and storage of biological data and information will require the employment of specialists that are suitably qualified, competent and experienced. The availability of such

personnel will be an important consideration in the successful use of wearables to assess teamwork within a military context.

5) The science to date highlights the opportunities that new technological advances could bring to the domain of teamwork. Despite inherent challenges with some of these, there is a degree of optimism that traditional barriers to researching teamwork effectively in the past could be overcome in the future.

REFERENCES

Benishek, L. E., & Lazzara, E. H. (2019). Teams in a new era: Some considerations and implications. *Frontiers in Psychology*, *10*, 1006. Doi: https://doi.org/10.3389/fpsyg.2019.01006

Bonito, J., & Keyton, J. (2018). Introduction to machine learning. In E. Brauner, M. Boos, & M. Kolbe (Eds.), *The Cambridge handbook of group interaction analysis* (pp. 387–404). Cambridge, England: Cambridge University Press. doi:10.1017/9781316286302.020

Carter, D. R., DeChurch, L. A., Braun, M. T., & Contractor, N. S. (2015). Social network approaches to leadership: An integrative conceptual review. *Journal of Applied Psychology*, *100*(3), 597- 622. Doi: https://doi.org/10.1037/a0038922

Casey-Campbell, M., & Martens, M. L. (2009). Sticking it all together: A critical assessment of the group cohesion–performance literature. *International Journal of Management Reviews*, *11*(2), 223-246. Doi: https://doi.org/10.1111/j.1468-2370.2008.00239.x

Chaffin, D., Heidl, R., Hollenbeck, J. R., Howe, M., Yu, A., Voorhees, C., & Calantone, R. (2017). The promise and perils of wearable sensors in organizational research. *Organizational Research Methods*, *20*(1), 3-31. Doi: https://doi.org/10.1177/1094428115617004

Chapman, R., & Colegrove, C. (2013). Transforming operational training in the Combat Air Forces. *Military Psychology*, *25*(3), 177-190. Doi: https://doi.org/10.1037/h0095980

Cheung, H. K., Hebl, M., King, E. B., Markell, H., Moreno, C., & Nittrouer, C. (2017). Back to the future: Methodologies that capture real people in the real world. *Social Psychological and Personality Science*, *8*(5), 564-572. Doi: https://doi.org/10.1177/1948550617709113

Cooke, N. J., Gorman, J. C., Duran, J. L., & Taylor, A. R. (2007). Team cognition in experienced command-and-control teams. *Journal of Experimental Psychology: Applied*, *13*, 146–157. Doi: https://doi.org/10.1037/1076-898X.13.3.146

Cooke, N. J., Gorman, J. C., Myers, C. W., & Duran, J. L. (2013). Interactive team cognition. *Cognitive Science*, *37*(2), 255-285. Doi: https://doi.org/10.1111/cogs.12009

Crawford, E. R., & LePine, J. A. (2013). A configural theory of team processes: Accounting for the structure of taskwork and teamwork. *Academy of Management Review*, *38*(1), 32-48. Doi: https://doi.org/10.5465/amr.2011.0206

Cronin, M. A., Weingart, L. R., & Todorova, G. (2011). Dynamics in groups: Are we there yet? *Academy of Management Annals*, *5*(1), 571-612. Doi: https://doi.org/10.5465/19416520.2011.590297

Dineen, B. R., Noe, R. A., Shaw, J. D., Duffy, M. K., & Wiethoff, C. (2007). Level and dispersion of satisfaction in teams: Using foci and social context to explain the satisfaction-absenteeism relationship. *Academy of Management Journal*, *50*(3), 623-643. Doi: https://doi.org/10.5465/amj.2007.25525987

Dubrow, S., Dobbins, C., Bannan, B., Zaccaro, S., Peixoto, N., Puohit, H., Rana, M., Au, M. & Segall, J. (2017). Using IoT sensors to enhance simulation and training in multiteam systems. In *The Interservice/Industry Training, Simulation and Education Conference (I/ITSEC) Published Proceedings (pp. 1-10).*

Endedijk, M., Hoogeboom, M., Groenier, M., de Laat, S., & Van Sas, J. (2018). Using Sensor Technology to Capture the Structure and Content of Team Interactions in Medical Emergency Teams during Stressful Moments. *Frontline Learning Research*, *6*(3), 123-147. Doi: https://doi.org/10.14786/flr.v6i3.353

Ferrin, D. L., Dirks, K. T., & Shah, P. P. (2006). Direct and indirect effects of third-party relationships on interpersonal trust. *Journal of Applied Psychology, 91*(4), 870–883. https://doi.org/10.1037/0021-9010.91.4.870

Friedl, K. E. (2018). Military applications of soldier physiological monitoring. *Journal of Science and Medicine in Sport*, *21*(11), 1147-1153. Doi: https://doi.org/10.1016/j.jsams.2018.06.004

Fulmer, C. A., & Gelfand, M. J. (2012). At what level (and in whom) we trust: Trust across multiple organizational levels. *Journal of Management, 38*(4), 1167-1230. Doi: https://doi.org/10.1177/0149206312439327

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*, 36–51. Doi: https://doi.org/10.1177/0018720811427901

Goodwin, G. F., Blacksmith, N., & Coats, M. R. (2018). The science of teams in the military: Contributions from over 60 years of research. *American Psychologist*, *73*(4), 322- 333. Doi: https://doi.org/10.1037/amp0000259

Gorman, J. C., Cooke, N. J., & Winner, J. L. (2006). Measuring team situation awareness in decentralized command and control systems. *Ergonomics, 49*, 1312–1325.
Kahneman, D., Sibony, O., & Sunstein, C.R. (2021). *Noise: A Flaw in Human Judgement*. Harper Collins: Glasgow.
Kalish, Y., Luria, G., Toker, S., & Westman, M. (2015). Till stress do us part: On the interplay between perceived stress and communication network dynamics. *Journal of Applied Psychology*, *100*(6), 1737- 1751.
Kazi, S., Khaleghzadegan, S., Dinh, J. V., Shelhamer, M. J., Sapirstein, A., Goeddel, L. A., ... & Rosen, M. A. (2019). Team physiological dynamics: A critical review. *Human Factors*, *63*(1). Doi: https://doi.org/10.1177/0018720819874160
Khaleghzadegan, S., Kazi, S., & Rosen, M. A. (2020). Unobtrusive Measurement of Team Cognition: A Review and Event-Based Approach to Measurement Design. *Contemporary Research*, 95-113.
Kiekel, P. A., Cooke, N. J., Foltz, P. W., Gorman, J. C., & Martin, M. J. (2002, September). Some promising results of communication-based automatic measures of team cognition. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 46, No. 3, pp. 298-302). Sage CA: Los Angeles, CA: SAGE Publications. Doi: https://doi.org/10.1177/154193120204600318
Klonek, F., Gerpott, F. H., Lehmann-Willenbrock, N., & Parker, S. K. (2019). Time to go wild: How to conceptualize and measure process dynamics in real teams with high-resolution. *Organizational Psychology Review*, *9*(4), 245-275. Doi: https://doi.org/10.1177/2041386619886674
Kolbe, M., & Boos, M. (2019). Laborious but elaborate: The benefits of really studying team dynamics. *Frontiers in psychology*, *10*, 1478. Doi: https://doi.org/10.3389/fpsyg.2019.01478
Kozlowski, S. W. J., & Chao, G. T. (2018). Unpacking team process dynamics and emergent phenomena: Challenges, conceptual advances, and innovative methods. *American Psychologist*, *73*(4), 576–592. Doi: https://doi.org/10.1037/amp0000245
LePine, J. A., Piccolo, R. F., Jackson, C. L., Mathieu, J. E., & Saul, J. R. (2008). A meta-analysis of teamwork processes: tests of a multidimensional model and relationships with team effectiveness criteria. *Personnel Psychology*, *61*(2), 273-307. Doi: https://doi.org/10.1111/j.1744-6570.2008.00114.x
Luciano, M. M., Mathieu, J. E., Park, S., & Tannenbaum, S. I. (2018). A fitting approach to construct and measurement alignment: The role of big data in advancing dynamic theories. *Organizational Research Methods*, *21*(3), 592-632. Doi: https://doi.org/10.1177/1094428117728372
Marlow, S., Bisbey, T., Lacerenza, C., & Salas, E. (2018a). Performance measures for health care teams: A review. *Small Group Research*, *49*(3), 306-356. Doi: https://doi.org/10.1177/1046496417748196
Marlow, S. L., Lacerenza, C. N., Paoletti, J., Burke, C. S., & Salas, E. (2018b). Does team communication represent a one-size-fits-all approach?: A meta-analysis of team communication and performance. *Organizational Behavior and Human Decision Processes*, *144*, 145-170. Doi: https://doi.org/10.1016/j.obhdp.2017.08.001
Matusik, J. G., Heidl, R., Hollenbeck, J. R., Yu, A., Lee, H. W., & Howe, M. (2019). Wearable bluetooth sensors for capturing relational variables and temporal variability in relationships: A construct validation study. *Journal of Applied Psychology*, *104*(3), 357-387. Doi: https://doi.org/10.1037/apl0000334.
Mozos, O. M., Sandulescu, V., Andrews, S., Ellis, D., Bellotto, N., Dobrescu, R., & Ferrandez, J. M. (2017). Stress detection using wearable physiological and sociometric sensors. *International journal of neural systems*, *27*(02), 1650041. Doi: https://doi.org/10.1142/S0129065716500416
Myers, S., Blacker, S., Greenlees, I., Runswick, O., Little, P., & Rolfs, P. (2021). *HS 1.027 Measurement of physiological, psychological and psychophysiological factors to predict military task performance*. O-HSSRC-1.027. Yeovil, UK: BAE Systems
Olguın, D. O., Gloor, P. A., & Pentland, A. S. (2009, September). Capturing individual and group behavior with wearable sensors. In *Proceedings of the 2009 AAAI spring symposium on human behavior modeling, SSS* (Vol. 9). Doi: https://www.aaai.org/Papers/Symposia/Spring/2009/SS-09-04/SS09-04-012.pdf
O'Neill, T. A., & Salas, E. (2018). Creating high performance teamwork in organizations. *Human Resource Management Review*, *28*(4), 325-331. Doi: https://doi.org/10.1016/j.hrmr.2017.09.001
Park, S., & Yeom, E. Y. (2016). Experience of the role conflict of nurses in long-term care hospitals. *The Journal of the Korea Contents Association*, *16*(7), 368-381. Doi: https://doi.org/10.5392/JKCA.2016.16.07.368
Pentland, A. S. (2012). The new science of building great teams. *Harvard business review*, *90*(4), 60-69.
Piwek, L., Ellis, D. A., Andrews, S., & Joinson, A. (2016). The rise of consumer health wearables: promises and barriers. *PLoS medicine*, *13*(2), e1001953. Doi: https://doi.org/10.1371/journal.pmed.1001953
Pleva, M., Dale, J., & Castelli, N. (2021). *HS 1.007 Defence Learning Ecosystems – ICE Talya 2025 Final Report*. Yeovil, UK: BAE Systems.

Porter, C. O. L. H., Hollenbeck, J. R., Ilgen, D. R., Ellis, A. P. J., West, B. J., & Moon, H. (2003). Backup behaviors in teams: The role of personality and legitimacy of need. *Journal of Applied Psychology*, *88*, 391-403. Doi: http://dx.doi.org/10.1037/0021-9010.88.3.391

Ramos-Villagrasa, P. J., Marques-Quinteiro, P., Navarro, J., & Rico, R. (2018). Teams as complex adaptive systems: Reviewing 17 years of research. *Small Group Research*, *49*(2), 135-176. Doi: https://doi.org/10.1177/1046496417713849

Roberts, A. P., Stanton, N. A., Fay, D. T., & Pope, K. A. (2019b). The effects of team co-location and reduced crewing on team communication characteristics. *Applied Ergonomics*, *81*, 102875. Doi: https://doi.org/10.1016/j.apergo.2019.102875

Roberts, A., Webster, L., Pope, K., Fay, D., Twist, S., & Fenton, C. (2019a). TIN2.108 *Collective Team Training Intervention Measurement and Evaluation (C-TTIME). Final Technical Report*. O-DHCSTC_I2_T_T2_108_005 Yeovil, UK: BAE Systems.

Salas, E., Rosen, M. A., Held, J. D., & Weissmuller, J. J. (2009). Performance measurement in simulation-based training: A review and best practices. *Simulation & Gaming*, *40*(3), 328-376. Doi: https://doi.org/10.1177/1046878108326734

Salas, E., Sims, D. E., & Burke, C. S. (2005). Is there a "big five" in teamwork? *Small Group Research*, *36*(5), 555-599. Doi: https://doi.org/10.1177/1046496405277134

Sawka, M. N., & Friedl, K. E. (2017). Emerging Wearable Physiological Monitoring Technologies and Decision Aids for Health and Performance. *Journal of Applied Physiology*, *124*(2), 430-431. Doi: https://doi.org/10.1152/japplphysiol.00964.2017

Shaffer, D. W., Hatfield, D., Svarovsky, G., Nash, P., Nulty, A., Bagley, E., Frank, K., Rupp, A., & Mislevy, R. (2009). Epistemic network analysis: A prototype for 21st century assessment of learning. *International Journal of Learning and Media*, *1*(2), 33–53. Doi: 10.1162/ijlm.2009.0013.

Serva, M. A., Fuller, M. A., & Mayer, R. C. (2005). The reciprocal nature of trust: A longitudinal study of interacting teams. *Journal of Organizational Behavior: The International Journal of Industrial, Occupational and Organizational Psychology and Behavior*, *26*(6), 625-648. Doi: https://doi.org/10.1002/job.331

Sexton, J. B., & Helmreich, R. L. (2000). Analyzing cockpit communications: the links between language, performance, error, and workload. *Human Performance in Extreme Environments*, *5*(1), 63-68. Doi: 10.7771/2327-2937.1007

Stanton, N. A., & Stanton, N. A., & Roberts, A. P. (2019). Better together? Investigating new control room configurations and reduced crew size in submarine command and control. *Ergonomics*, 1-17. Doi: https://doi.org/10.1080/00140139.2019.1654137

Sullivan, S., Warner-Hillard, C., Eagan, B., Thompson, R. J., Ruis, A. R., Haines, K., ... & Jung, H. S. (2018). Using epistemic network analysis to identify targets for educational interventions in trauma team communication. *Surgery*, *163*(4), 938-943. Doi: https://doi.org/10.1016/j.surg.2017.11.009

Taylor, P. J., (June, 2013). How technology is revolutionizing our understanding of human cooperation (Inaugural lecture). Twente University Press. Downloaded on 9 June, 2021, from https://www.utwente.nl/en/academic-ceremonies/inaugurallectures/booklets-inaugural-lectures/2007-2014/Oratieboekje_Taylor.pdf

Walker, G. H., Salmon, P. M., Bedinger, M. and Stanton, N. A. (2017) Quantum ergonomics: shifting the paradigm of the systems agenda. Ergonomics, 60 (2) 157-166. Doi: https://doi.org/10.1080/00140139.2016.1231840

Wallenius, C., Larsson, G., & Johansson, C. R. (2004). Military observers' reactions and performance when facing danger. *Military Psychology*, *16*(4), 211-229. Doi: https://doi.org/10.1207/s15327876mp1604_1

Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications* (Vol. 8). Cambridge UK: Cambridge University Press.

Wetzel, C. M., Kneebone, R. L., Woloshynowych, M., Nestel, D., Moorthy, K., Kidd, J., & Darzi, A. (2006). The effects of stress on surgical performance. *The American Journal of Surgery*, *191*(1), 5-10. Doi: https://doi.org/10.1016/j.amjsurg.2005.08.034

Wu, L., Waber, B., Aral, S., Brynjolfsson, E., & Pentland, A. (2008). Mining face-to-face interaction networks using sociometric badges: Predicting productivity in an IT configuration task. Retrieved from: SSRN. http://dx.doi.org/10.2139/ssrn.1130251

Yilmaz, G. (2016). What you do and how you speak matter: Behavioral and linguistic determinants of performance in virtual teams. *Journal of Language and Social Psychology*, *35*(1), 76-97. Doi: https://doi.org/10.1177/0261927X15575772

Zhang, Y., Olenick, J., Chang, C. H., Kozlowski, S. W., & Hung, H. (2018). TeamSense: assessing personal affect and group cohesion in small teams through dyadic interaction and behavior analysis with wearable sensors. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, *2*(3), 1-22.