A Tale of Two Users: Designing Data Visualizations for Researchers and Trainers

Michael King, Julian Abich IV, Eric Sikorski
Quantum Improvements Consulting
Orlando, FL
mking@quantumimprovements.net
jabich@quantumimprovements.net
esikorski@quantumimprovements.net

Clifford Hancock, Meghan O'Donovan
Combat Capabilities Development Command
Natick, MA
clifford.l.hancock4.civ@army.mil
meghan.p.odonovan.civ@army.mil

Gregory Goodwin
Combat Capabilities Development Command
Orlando, FL
gregory.a.goodwin6.civ@army.mi

ABSTRACT

Data visualizations are vital for trainers and instructors as they simplify complex information and help with decision-making. It is essential to tailor the design of these visualizations based on the specific needs and characteristics of the intended audience to maximize their effectiveness in training contexts. As part of the U.S. Army Combat Capabilities Development Command (DEVCOM) Soldier Center's Small Unit Performance Analytics (SUPRA) program, we developed processes and tools for building transcripts from audio files and then conducted classification and analysis of those communications. We experimented with a variety of data visualizations to communicate this information to a research-focused audience. Through further investigation, it was revealed that the data could also benefit military battle drill trainers. To transition the data into useful visualizations for trainers, it was important to understand their needs and limitations when evaluating battle drills. Observer Controller/Trainers (OC/Ts) have limited opportunity to observe all communications, possibly missing key information exchanges. Soldier-worn sensors can augment OC/T training evaluations, but it is critical that the data be presented in a way that facilitates assessment and feedback in order to maximize positive training outcomes. In this paper, utilizing the data collected, we propose approaches to presenting tactical communication data in ways that facilitate evaluation and feedback. These include evaluating team processes like supporting behavior and leadership, as well as individual military task execution. We also explore ways to integrate communication data with other squad-based performance measures so that the communication data can provide the proper context for their interpretation and vice-versa. We observed that there is no single solution for visualizing tactical communication data. Instead, we recommend a suite of visualization tools and methods should be used to support OC/T evaluations and assessments of small-unit training.

ABOUT THE AUTHORS

Michael King, Ph.D., is a Research Psychologist II at Quantum Improvements Consulting (QIC). While at QIC, Michael has led various training and human performance improvement projects, including a major initiative studying team communication and performance for the U.S. Army. Michael's research interests at QIC center on human performance, training, and the predictors of success in team-based performance contexts. Michael holds a Ph.D. and M.S. in Experimental Psychology from Case Western Reserve University.

Julian Abich IV, Ph.D., is a Senior Human Factors Engineer at Quantum Improvements Consulting. He holds a Ph.D. in Modeling and Simulation with a specialization in Human Factors, a B.S. in Psychology, and two certificates in Design for Usability and Instructional Design for Simulations from the University of Central Florida. He has expertise in applying human factors and ergonomics (HF/E) principles, modeling and simulation (M&S) approaches, and instructional design strategies to assess, predict, and improve human performance. He has extensive experience conducting usability & user experience evaluations across various human-computer interaction and training

applications. Continuing to take a user-centered approach, his work focuses on designing, applying, and evaluating the use of immersive technologies to improve job performance.

Eric Sikorski, Ph.D., is the Director of Programs and Research at Quantum Improvements Consulting, where he leads a team of Research Psychologists and Human Factors Engineers. His work at QIC includes staff development, strategic planning for the division, and managing various field studies, such as evaluating cutting-edge immersive training technology and team communication for multiple DoD customers. Eric holds a Ph.D. and M.S. in Instructional Systems from Florida State University and an M.A. in Human Factors and Applied Cognition from George Mason University.

Clifford Hancock, M.S., is a biomechanical engineering researcher with the US Army Combat Capabilities Development Command – Soldier Center (Natick, MA, USA). He received his M.S. degree in Biomedical Engineering from the University of Texas at Arlington in 2013.

Meghan O'Donovan, M.S., is a principal investigator with the US Army Combat Capabilities Development Command – Soldier Center (Natick, MA, USA) leading the Small Unit Performance Analytics program (SUPRA). She holds a M.S. degree in Biomedical Engineering from the University of Rochester.

Gregory Goodwin, Ph.D., is a Senior Research Scientist at the Combat Capabilities Development Command – Soldier Center in Orlando, FL. For the last fifteen years, he has worked for the Army researching ways to improve training methods and technologies. He holds a Ph.D. in Psychology from Binghamton University and an M.A. in Psychology from Wake Forest University.

A Tale of Two Users: Designing Data Visualizations for Researchers and Trainers

Michael King, Julian Abich IV, Eric Sikorski
Quantum Improvements Consulting
Orlando, FL
mking@quantumimprovements.net
jabich@quantumimprovements.net
esikorski@quantumimprovements.net

Clifford Hancock, Meghan O'Donovan
Combat Capabilities Development Command
Natick, MA
clifford.l.hancock4.civ@army.mil
meghan.p.odonovan.civ@army.mil

Gregory Goodwin
Combat Capabilities Development Command
Orlando, FL
gregory.a.goodwin6.civ@army.mi

INTRODUCTION

The Army Operating Concept recognizes that decentralized operations require adaptive leaders, cohesive teams, and resilient Soldiers (U.S. Army TRADOC, 2014). Central to all of these is the ability of small units to communicate effectively during military operations. When maneuvering as a squad, it is vital that all Soldiers within that squad have a shared understanding of each other's intent and can perform proficiently as a team. For instance, squad members must communicate essential information about the enemy, squad movement, and events such as weapon reloading or malfunction. They must do so clearly and concisely, typically in a noisy environment with limited time. Similarly, squad leaders must be able to communicate with the team members they lead. Breaks or failures in this communication have potentially dire consequences in the field. To that end, the United States Army is investigating the relationship between team communication and performance within the context of Army battle drills.

Battle drills are training exercises that enable combat units to rehearse doctrinally accurate tasks like "enter and clear a room" or "conduct a squad assault." They are typically rehearsed until they become second nature to Soldiers, so that they can be executed correctly under stressful, combat conditions (Department of the Army, 2016). When training to execute battle drills, leaders practice making decisions quickly and issuing effective brief orders. Squads practice communication, coordination, and supporting behavior in an environment that closely approximates combat. By recording and analyzing squad communication, much can be learned about the relationship between this communication and performance in simulated combat operations.

As part of the Small Unit Performance Analytics (SUPRA) initiative funded by the U.S. Army Combat Capabilities Development Command (DEVCOM), we collected and analyzed squad communication data during two battle drills. We recorded communication data from 18 dismounted infantry squads executing multiple iterations of Battle Drill 2A (BD2A). To better understand the impact of communication on performance, we created measures and metrics that capture differences in squad communication. Squad performance was assessed by expert Observer Controller-Trainers (OC/Ts), so we were able to compare communications as a function of squad performance. In the course of this work, we realized that there is a range of methods needed for summarizing and communicating research findings involving the relationship between communication and squad performance. No single method is sufficient. Toward this goal, we created several types of data visualizations to effectively communicate our findings and inform other researchers and stakeholders what these findings were in a concise and accessible way.

During the SUPRA study, we also learned that there is a need to provide visualizations of squad communication to aid OC/Ts in their training evaluations. OC/Ts play a critical role in assessing small team performance in the U.S. Army. During training exercises, OC/Ts closely monitor and evaluate squad performance. They use standardized evaluation criteria organized in training and evaluation outlines (T&EO) issued by the U.S. Army. One of the essential roles of OC/Ts is providing constructive feedback to the squads they observe during after-action reviews (AARs).

Assessing squad performance can be challenging because squads can be spread across a large area (e.g., Army situational training exercise (STX) lanes) when executing certain battle drills, making it difficult for an OC/T to hear and see all actions and communications. Therefore, OC/Ts may benefit from having access to recorded communication data. However, interpreting raw communication data is challenging, arduous and time-consuming. Therefore, it is critical that the data be presented in a way that facilitates efficient assessment and feedback.

Creating visualizations for researchers is different than creating visualizations for practitioners and trainers. For researchers, data visualizations provide an opportunity to display findings in ways that go beyond statistics and written explanations. Data visualizations make it easier for researchers to explain their analyses in a clear and compelling way, helping other researchers with different areas of expertise understand the outcomes of a study. Good data visualizations should emphasize key findings, tell a story about the data, and make it easier to identify patterns, trends, and relationships from complex analyses.

For practitioners, who generally do not have a background in statistical analysis, these types of visualizations serve a different purpose than is needed by an OC/T evaluating drill performance. OC/Ts need data visualizations that facilitate intuitive interpretation. Data visualizations should help learners and trainers see the relationship between communication and performance. Providing the appropriate data visualizations to trainers and trainees can be challenging. Hence, a user-centered design (UCD) approach was implemented to prioritize the needs, preferences, and goals of the end-user throughout the visualization design process.

In this paper, we discuss the technical challenges associated with collecting communication data in a live training setting. We then share the various types of visualizations created from communication data from SUPRA, and how our approach for these visualizations changed based on the particular end-user. For research and analysis purposes, we created visualizations that provided insight into the relationship between communication and performance. For practitioners, we took a UCD approach and created visualizations that to inform OC/Ts and facilitate assessment and feedback in order to maximize positive training outcomes.

METHODS

Our team gathered squad communication data via audio recorders and lapel microphones worn by each Soldier in a squad. Effective communication within a squad is essential in demanding work contexts, such as those in which performance is time-sensitive, the stakes are high, and an active opposition is involved. As a result, it is essential to get a deeper understanding of how teams interact and how variations in communication may influence performance. However, collecting this valuable communication data is difficult, especially in a complex field study environment where noise intrusion is ever-present. Below we discuss the data collection solution, how communication measures were conceived, and challenges associated with the SUPRA study methods.

Methodology for Capturing Communications

In preparation for the SUPRA data collections, our team sought to develop, research, and test data collection procedures to accurately capture squad communications. This process included searching for and testing recording devices to capture audio, researching speech-to-text solutions to transcribe audio files, and developing and testing data collection procedures to carry out data collection activities succinctly.

In the search for an effective data collection tool methodology, we aimed to select equipment that could produce high-quality communication data, isolating who was speaking, what they were saying, and minimizing background noise. We researched various microphones and audio recording devices, eventually selecting a recorder and lapel microphone combination. The research team conducted extensive testing in the office and simulated field conditions. The result of that testing confirmed that the recorder microphone combination was sufficient for producing quality audio data while not impeding Soldier range of motion.

Once we were able to collect high-quality audio recordings, we needed a process to transform the audio into data. Transcribing the verbal communications from the audio files manually was both costly and inefficient. At the time, there was not an off-the-shelf software solution that could accurately and reliably transcribe the speech of multiple people speaking at once in a noisy field environment, especially when communications involved the use of military terminology. We were able to identify an effective speech-to-text (STT) software that improved data processing time by providing an initial transcription for each speaker with timestamps, from which human transcribers could add to and correct as needed.

We developed a data processing pipeline to go from raw audio files to quantitative data. Audio files were trimmed, cleaned, and processed through speech-to-text software. Speech-to-text outputs were organized into a spreadsheet based on squad position and time, and then confirmed for accuracy by a human listener. Once each line of dialogue for each squad member was verified for accuracy, transcriptions were combined to create one flow of communication for the entire squad. From this completed transcription document, communication measures were then scored by two independent researchers.

Measuring Communication

Communications for the squad executing BD2A were analyzed using a five-factor team communication framework, adapted from the conceptual framework for teamwork measurement from Dickinson and McIntyre (1997). The constructs were (i) Communication Quality, which referred to the quality, frequency, and exchange of information, (ii) Coordination, which pertained to the management and timing of activities, (iii) Cooperation, which related to the assistance and support of others, (iv) Leadership, which entailed the provision of directions and assertiveness behind decisions, and (v) Monitoring, which related to the awareness of ongoing processes and team observation (Jeffcott & Mackenzie, 2008).

Most measures pertained to tactical and procedural tasks to be completed in BD2A drill completions, according to U.S. Army doctrine. For example, Squad Leaders must communicate the enemy's distance, direction, and description (also known as the enemy's "DDD") to the entire squad upon enemy contact. We treated this procedural communication as an independent measure, where Squad Leaders could receive between 0 and 3 points for every "D" they communicated or failed to communicate. This measure was categorized as a Leadership construct measure. Non-procedural communication measures were also utilized for this study. For example, it is well-known in team communication literature that closed-loop communication, especially in high-stress environments, is critical for team performance (Härgestam et al., 2013; Diaz et al., 2017; El-Shafy et al., 2018). We measured the rate of a squad's closed-loop communication by first identifying utterances that opened a loop by a communication sender and then identifying if each of these open loops was eventually closed by a communication receiver. We then calculated the rate at which this occurred. This measure was categorized as a Communication Quality construct measure.

DATA VISUALIZATIONS FOR RESEARCH AND ANALYSIS

In human performance research, creating effective data visualizations is critical for conveying complex research findings to fellow scientists. Rather than simply presenting statistical analyses in tables and written text, data visualizations help to facilitate a concrete understanding of research outcomes, enabling researchers to communicate their analyses in a more accessible and compelling manner. Additionally, well-crafted visualizations enhance the transparency of analyses, providing a visual narrative that supplements traditional statistical reporting. While the measures and metrics developed for this study effectively extract meaning from squad communication from raw audio files, to communicate our findings to both researchers and peers and to enhance our analysis of squad communication, we created data visualizations. Here, we report examples of how we did this: first, by constructing simple communication transcriptions, and second, by conducting social network analysis.

Communication Transcriptions

Squad transcriptions (see an example in Figure 1) from the battle drill iterations give a "play-by-play" of the drill events and how each squad member communicated and reacted to drill events in their own words. The transcripts revealed trends and patterns that further explained performance differences among squads. The drill can be broken down by stage, problems during each stage can be identified, and the squad's response to these problems can be analyzed. Breaking down transcriptions can expose where squads have communication breakdowns and how this impacts their performance. Transcriptions provide a detailed textual representation of squad interactions at various points in the drill, allowing researchers to understand the context for actions, identify patterns in teamwork and leadership dynamics, and even conduct sentiment analysis to assess emotional dynamics. Researchers can uncover recurring themes, assess emotional dynamics, and gain insights into the context surrounding communication events. Comparative analysis between different squads or within a squad at different time points enables the identification of behavioral differences, contributing to a deeper understanding of team dynamics and change over time.

Additionally, integrating raw transcriptions with other data sources facilitates a holistic analysis, which has been useful for case study approaches involving the identification of critical incidents and their causes. For example, integrating communication transcripts with firing data helped our team to uncover the root cause of a fratricide incident (detected by multiple integrated laser engagement system, a.k.a. MILES) among one of the higher performing squads in this study (King, Sikorski & Goodwin, 2023) (see Figure 1). The firing data indicated that a fratricide incident had occurred, and the transcription data revealed a communication breakdown between individual squad members stemming from a non-functioning radio (see Figure 1).

Position	Timestamp	Comms
SL	[00:10:23	Bravo be advised Alpha flanking New York New York. How copy over.
SL	[00:10:31	He can't hear me.
BTL	[00:10:31	[NAME] pick up rate of fire
SL	[00:10:32	Stay down, stay put. Keep fire on them
ATL	[00:10:35	Keep shooting at them!
ASAW	[00:10:38	I see him!

Figure 1. Sample transcripts from a BD2A event (King, Sikorski & Goodwin, 2023).

Examining the communication flow in transcriptions among the members of a squad can provide important additional information about how a squad performed and why some events occurred during the drill. Squad communications can provide valuable insight into actions and decisions, especially when incidents occur. This type of analysis is crucial for researchers to understand root-causes of events, and to better understand team dynamics as context for the results of more traditional statistical analysis.

Position	Timestamp	Comms
BTL	[00:05:55	Push out.
ATL	[00:06:56	Hey Contact 2 oclock.
ATL	[00:07:01	70 meters.
ASAW	[00:07:02	12 o'clock, 70 meters.
SL	[00:07:04	12 o'clock, 70 meters.
ATL	[00:07:04	Hey [NAME] pull up.
ASAW	[00:07:06	Come on.
ATL	[00:07:07	Yes. Pull up. Pull up. Get online right here.
ATL	[00:07:15	Alright, rapid 3 rapid 3 start it off.
SL	[00:07:17	Hey where they at?
AGRN	[00:07:17	Rapid 3 rapid 3.
ATL	[00:07:18	Hey 2-- 3 o'clock 70 meters, up on that hill.
SL	[00:07:19	Directly in front of us?
ASAW	[00:07:20	Malfunction.
BTL	[00:07:23	Hey get ready to move [NAME].
SL	[00:07:24	Alright, I see them.
ATL	[00:07:24	Yes alright alright fix it.
SL	[00:07:25	Aye.
AGRN	[00:07:25	Malfunction.
BTL	[00:07:25	Get ready to move.
ATL	[00:07:37	Alright, pick it back up [NAME] pick it back up.
SL	[00:07:42	Two, two.
ASAW	[00:07:46	Come on man.

Figure 2. Sample transcriptions from a BD2A event.

Interpreting raw communication transcripts is challenging and time-consuming (see Figure 2). You may have to sift through hundreds of lines of communication for analysis unless you know the specific timing of a critical event.

Examining communication around a significant event is valuable for understanding squad reactions and thought processes pertaining to that event, but this type of visualization lacks meaningful purpose in the absence of such context. From a researcher perspective, these types of visualizations may be of great use, but for a practitioner or trainer, they are not as accessible and easy to interpret. OC/Ts conducting squad evaluations, for example, do not always have the time or resources to meticulously sift through communication transcripts to uncover patterns and context for events that occurred.

Word Clouds

From a transcript, it is possible to build some low-level summary visualizations of the words spoken by the team, like word frequency either of the entire squad, team, or per individual. With a little bit of classification, one can also identify "filler words" or "non-essential" words. One way to visually represent these types of parametric statistics is in the form of a word cloud.

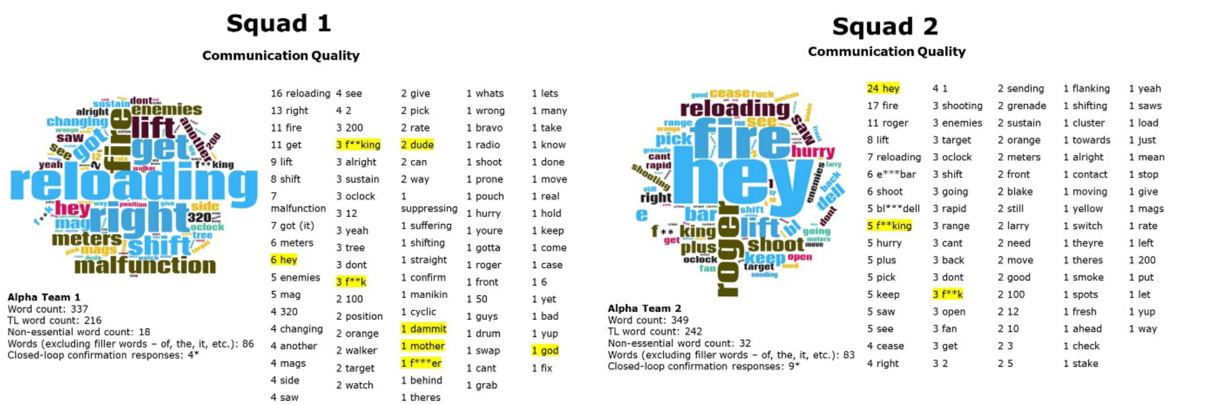

Figure 3 Word Cloud representations of two different support by fire teams that are from different squads. Note the color codes and character size in the word clouds denote the frequency range of the words.

One advantage of word cloud representations is that they make it easy to see which words are used most frequently. Looking at the two examples in figure 3, you can also see differences across the two teams. The color codes in these word clouds are related to the frequency ranges of the words, but they could also be color coded by other categories like filler words, whether they are part of a closed loop communication, etc. A disadvantage is that word clouds don't always provide an understanding of context in which specific words are spoken. For example, the reason that "hey" was the most frequently used word in the team from squad 2 is not evident from the image. Neither is it clear why "reloading" was spoken most frequently by the team from squad 1. For example, we don't know how many individuals said "reloading" or whether this was a good or bad thing. Despite these problems, word clouds do reflect patterns or trends in behavior at the team level and it is conceivable that some patterns might relate to some performance features. For trainers though, it is harder to see how a word cloud could be used to identify specific issues in a single training iteration.

Social Network Analysis

The Interaction Model of communication views communication as an interactive, two-way process in which participants generate meaning by sending messages and receiving feedback within physical and psychological contexts (Schramm, 1997). One way to summarize and visualize team communication consistent with this model is through a Social Network Analysis (SNA). An SNA is an analytical tool that allows a researcher to identify patterns of social relations among many actors with visual models and objective metrics that are grounded in scientific theory (Wasserman and Faust, 1994).

The SNAs conducted by our team consist of two key metrics, along with a model of each squad's closed-loop communication flow for the entirety of each drill.

- **Network Density**: measures how interconnected or "dense" a group or community is based on relationships or connections between its members. This helps us understand how closely connected or isolated members of a squad are. A high density means everyone is well-connected, while low density suggests some people might be more isolated from others in the network.

- **Network Centralization**: A measure to help us understand how much control or influence is concentrated within a single or a few individuals in a network. It helps determine whether a few individuals have a lot of control or influence over a group's interactions (centralized) or if influence is spread more evenly among many people (decentralized).

These metrics provide some insight into a squad's style of communication for a specific performance, but the SNA model visualization allows researchers to see what this actually looks like in a physical communication structure. These visualizations typically appear as a graphical representation of nodes (representing individuals Soldiers) and edges (representing closed-loop communication interactions) between them (see Figures 4 and 5). The visualization is structured to illustrate the relationships and communication patterns within a squad. Nodes are depicted as points, and edges as lines connecting these points. The size of the edges represent the frequency of interactions between two squad members. Clusters or groups of nodes with dense connections may suggest subgroups or closely connected Soldiers, often appearing among Alpha and Bravo teams, while isolated nodes or sparse connections may indicate areas of less frequent communication. Overall, visualization provides a clear and intuitive representation of the social structure, helping researchers and analysts identify key actors, patterns, and dynamics within the network.

One way to analyze these visualizations is by comparing one squad's SNA over time, or between drill completions. As squads gain experience and have repeated exposure to BD2A together, the SNA gain shows how they grow as a team. For example, in Figure 4, we can see how a squad's social network changed from day one to day three of completing BD2A.

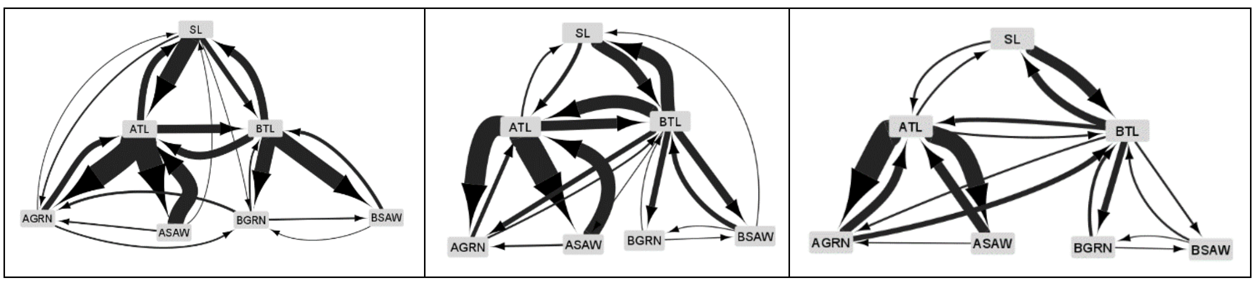

Figure 4. SNA figures from one squad over time (three drill completions).

Another way to analyze SNA visualizations is to compare between different squads (see Figure 5). In the example shown here, we can see stark differences between the models of the two squads. The squad on the left has a high frequency of communications from the Alpha Team Leader (ATL) to their team, and the Bravo Team Leader (BTL) to their team. Additionally, we see that the only two individuals who communicate to the Squad Leader (SL) are these two team leaders. The squad on the right's social network is drastically different. Several individuals communicate with the SL, and Soldiers from different teams (i.e., Alpha and Bravo) communicate with one another. For the squad on the left, only the leaders communicate with one another. These comparisons of social network visualizations better explain differences between squads beyond a linear transcript.

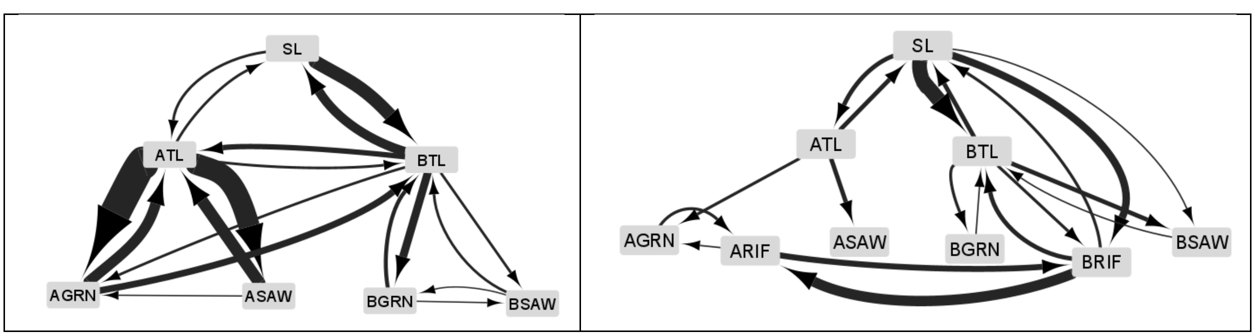

Figure 5. SNA figures comparing two squads.

While SNA is a powerful tool for researchers to better understand relationships and interactions within a communication network, there are shortcomings for use with OC/T practitioners. SNA provides a static snapshot of a network at a specific period of time. It may not capture the dynamic nature of relationships and interactions that evolve over time. Changes in the network structure may be easily overlooked. Additionally, SNA primarily focuses on structural patterns and may not provide insights into the qualitative aspects of relationships. It lacks context regarding the nature and quality of interactions, emotions, or the content of communication between nodes. While network density and centralization provide some data-backed context, drawing conclusions based on SNA figures requires a great deal of subjectivity. An interpretation of the figure may change from one OC/T to another. Misinterpretation of the visualizations can lead to incorrect conclusions about the nature of relationships or the overall network structure. OC/Ts may be able to identify patterns and correlations from an SNA, but these conclusions may fall short of establishing causation. Understanding why certain network patterns emerge requires additional context, and SNA figures may be misleading without this context.

DATA VISUALIZATIONS FOR TRAINING ASSESSMENT

During the SUPRA study, our team was able to get a first-hand understanding of how OC/Ts conduct their squad evaluations. These squad evaluations consist of subjective judgment ratings on a number of expert-identified performance competencies. These judgements are coupled with assessments of BD2A steps outlined in Army doctrine (O'Donovan et al., 2023). However, conducting these evaluations relies heavily on what the OC/Ts are able to observe and remember over the entirety of the drill. In SUPRA, there were limited opportunities for OC/Ts to incorporate sensor data into their evaluations. Thus, these ratings are subjected to bias and error in several areas.

- **Subjectivity**: Observer ratings are inherently subjective and may be influenced by the observer's personal biases, perspectives, and preferences. Different observers may interpret the same performance differently, leading to inconsistencies in evaluations.
- **Incomplete Information**: Observers might not have access to all relevant information about the performance. Certain aspects of the drill may not be observable during the evaluation, leading to an incomplete understanding of the squad's performance.
- **Limited Reliability**: Observer ratings may lack reliability over time and across different contexts. Changes in observer mood, attention, or understanding can impact the consistency of ratings, making it difficult to rely on them as a stable measure of performance.
- **Inefficiency**: Collecting observer ratings can be time-consuming and resource-intensive. In situations where real-time or continuous feedback is needed, relying solely on observer ratings may be impractical.

These biases are not unique to OC/Ts conducting BD2A squad evaluations and are nearly impossible to overcome for any type of observer evaluation. To gain a comprehensive understanding of squad performance, OC/T ratings should be complemented by other data sources, such as objective metrics stemming from the sensor data collected during SUPRA events. Relying solely on observer ratings may limit the richness and depth of the performance assessment.

Communication data from Soldier-worn sensors, in particular, can augment OC/T evaluations. However, data visualizations used for research and analysis, such as transcription logs and SNA models, do not encapsulate communication and performance in an easy to understand, user-friendly way. Additionally, these visualizations have the potential to impose a burden on OC/Ts, as they may require spending time and effort deciphering complex models and sifting through rows of data, thereby increasing the workload for OC/Ts. It is critical that the data be presented in a way that facilitates assessment and feedback, and saves OC/Ts time and effort, in order to maximize positive training outcomes. To address this gap, our team took a User-Centered Design approach to develop data visualizations that capture squad communication during BD2A.

UCD Approach

UCD prioritizes the needs and preferences of the end users throughout the entire design process. The principles of UCD (Gulliksen et al., 2003) guide designers in creating visualizations or tools that are intuitive, efficient, and user-friendly. Taking a UCD approach towards creating data visualizations allowed us to adapt our data and visualizations for research and analysis and to OC/Ts for practical purposes. This involves prioritizing the needs, preferences, and goals of the end user throughout the visualization design process. In the context of Army OC/Ts, where squad communication plays a crucial role in performance, UCD ensures that the data visualizations effectively communicate insights and support decision-making. Adhering to common principles of UCD (see Table 2), we created visualizations that not only meet OC/Ts needs but also contribute to positive user experiences and overall satisfaction.

OC/T Visualizations

Toward the goal of creating effective data visualizations for OC/Ts using a UCD approach, our team met with several U.S. Army OC/Ts to establish what information they currently use to evaluate performance during BD2A and solicit ideas for how additional data may help. After mocking up prototype visualizations with varying amounts of data and different data sources, our team conducted semi-structured interviews with OC/Ts to solicit feedback on the mockups to determine which would be most useful to them and how to improve them for future use. The majority of the sample (N = 20) consisted of individuals at the Sergeant First Class rank (n = 13), with the rest of the sample consisting of Staff Sergeants (n = 7).

Communication Data Visualizations

In our team's discussions with OC/Ts, we identified several areas where data visualizations would help their evaluations. The two most prominent areas were communication and GPS positioning data. For GPS positioning data, OC/Ts expressed that they would like to be able to see GPS locations and movements of Soldiers with timestamp trackers, visualizations of their movement throughout the drill, and aerial shots of Soldier movement from one point to another during the drill. For communication data, OC/Ts unanimously expressed that communication transcripts of the entire drill with timestamps would be helpful. To better provide OC/Ts with impactful visualizations of communication data, we decided to display transcripts on a GPS map showing Soldier movements, so that OC/Ts have the appropriate context for communications at any given time (see Figures 6 and 7). See Table 1 for a description of the communication visualization features.

Table 1. Communication visualization features.

Feature	Justification
Scrolling transcriptions	• OC/Ts expressed a need to be able to see what was being said at any time during the drill. • A scrolling log over time, shown with GPS movements, allows the OC/T to have context with communication and not be inundated with a long list of transcriptions.
The ability to play audio	• Audio of actual communications allow OC/Ts to better understand the emotional dynamics within a squad, as well as more context for how utterances were communicated (i.e., not loud enough for the whole squad to hear). • Highlighting text as the words are being spoken allow OC/Ts to easily follow along with the drill communications.
Word bubble option	• This was designed to give the OC/Ts options for how they preferred to view communications (word bubbles or scrolling log).
Ability to toggle speaker	• This allows OC/Ts further customization and options for how they want to see communications. • OC/Ts expressed a stronger need to see what leadership positions were communicating vs non-leadership positions.

Figure 6. OC/T communication visualization showing a word bubble feature vs scrolling log of communication.

Figure 7. OC/T communication visualization showing the option to toggle different speakers.

Alignment with UCD Principles

By adhering to common UCD principles, we were better able to tailor effective data visualization solutions for OC/Ts. Without bringing OC/Ts into the loop to better understand their needs and gather feedback on prototype visualizations, we would be providing practitioners with visualizations that were only useful for research and analysis purposes. See Table 2 for details on how our development process aligned with a UCD approach.

Table 2. Common UCD principles and their alignment with our data visualizations.

UCD Principle	Explanation	Alignment
User Involvement	Involve users throughout the design process. Engage them in activities such as user research, testing, and feedback sessions to understand their needs, preferences, and behaviors.	• Interviewed OC/Ts to understand how they currently evaluate squads and what their pain points are
Iterative Design	Embrace an iterative design process. Create prototypes or versions of the design and gather user feedback. Use this feedback to refine and improve the design in successive iterations.	• Several prototypes were developed and tested with usability researchers and OC/Ts for feedback based on their expertise
Holistic Design	Consider the entire user experience, including interactions, visual design, information architecture, and other aspects. Aim for a holistic design that addresses all elements to create a seamless and cohesive user experience.	• Designed features with the user in mind, creating the ability to toggle speakers and shooters, and turn on and off certain features depending on the user need at various points in the drill
Clear and Consistent Communication	Ensure clear and consistent communication between the product or system and the user. Use language, symbols, and interactions that are familiar and easily understood by the target audience.	• We incorporated symbols that are cohesive with what is used in the Army, and what OC/Ts currently use when evaluating squads • We incorporated features that highlight behavior that OC/Ts currently evaluate, but currently requires a higher level of effort on their part
Flexibility and Efficiency	Design for flexibility and efficiency. Allow users to accomplish tasks with minimal steps and provide options for customization to accommodate varying user needs and preferences.	• Provided features that allow for customization to accommodate differences in OC/T needs
User-centric Metrics	Define and measure success based on user-centric metrics. This could include task success rates, user satisfaction, and efficiency metrics. Use these metrics to gauge the impact of the design on the user experience.	• Used metrics when interviewing OC/Ts to help determine how the visualizations would help them for different types of tasks (evaluation vs AAR) • Used metrics when interviewing OC/Ts to help determine difficulty level of their current processes

DISCUSSION

Verbal communication is essential for quality team performance during battle drills like BD2A (Abd El-Shafy et a., 2018, Jeffcott & Mackenzie, 2008). This makes accurate summaries of squad communications important for both research and for routine assessment of squads during training. We learned that visualizations of these communications for researchers and trainers are not necessarily interchangeable. This is probably because researchers are more interested in understanding repeatable trends that occur across a range of squads (e.g., they are looking for the ways in which expert squads perform differently from novice squads). On the other hand, OC/Ts may focus on trying to understand the performance of a single squad during a specific training event. They want to be able to dissect the causes and consequences of a single incident. It is perhaps not surprising then, that visualizations that help reveal trends across many squads may not be particularly good at illuminating the actions of a single squad at a single point in time. It is not to say that data summaries for these two purposes have no connection. For instance, if research can identify repeatable patterns of behaviors that reflects levels of performance and a pattern occurs in a single instance of training, then a notification of this pattern might be useful to an OC/T or other trainer. Conversely, as OC/Ts gain access to more detailed communication data, they can provide researchers with greater insight into what constitutes good or bad performance as reflected in these data sets. This will in turn inform the hypotheses that researchers might look to explore in larger data sets.

It is also possible that as trainers have access to larger databases, they may desire to become more proficient at providing feedback to individual units regarding how their performance compares to historical trends. Units will likely want to know how their performance compares to that of squads in different types of units, like elite special forces units or even other units in their own division or brigade. Such comparisons could be facilitated by the kinds of visualizations currently used by researchers to identify group trends.

It is clear that investment in technologies that can rapidly transcribe and classify communication data will support both researchers and trainers, and will therefore have a significant impact on both training of and research on team performance. Though we have developed solutions to many of the challenges associated with collecting voice communications from a squad during live training of BD2A, it is not automated to the degree that would make it feasible in a training environment. Units typically conduct an AAR shortly after completing the exercise when details are still fresh in everyone's mind. This means that they would need communication data ready within 10-15 minutes of the conclusion of the training exercise. It currently takes us about 6 hours to collect and fully process the communication data from a single exercise by a single squad. This includes building an accurate transcript from each recorder with the support of speech-to-text software and then clipping the overlapping voice comms and attributing them to respective individuals. Transcripts are then placed on a timeline on which key events are also tagged, and each communication is classified, and key metrics are scored.

To fully implement an automated pipeline that would enable the kinds of visualizations presented in Figures 5 and 6, we would also need to automatically compile GPS location data, helmet and weapon inertial measurement unit data, and MILES data. Of all of these, communication data is currently the most challenging to process rapidly enough to meet AAR requirements. Further research is needed to develop efficient approaches to process and translate communication data in a way that can rapidly be provided to users in a visually meaningful way.

ACKNOWLEDGMENTS

This research was sponsored by the U.S. Army Combat Capabilities Development Command - Soldier Center (DEVCOM). The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of CCDC-SC or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation hereon.

AUTHORSHIP STATEMENT

All authors were involved in research planning, data collection, data analysis, report writing and editing.

REFERENCES

Abd El-Shafy, I., Delgado, J., Akerman, M., Bullaro, F., Christopherson, N. A., & Prince, J. M. (2018). Closed-loop communication improves task completion in pediatric trauma resuscitation. *Journal of Surgical Education*, *75*(1), 58-64.

Diaz, M. C. G., & Dawson, K. (2020). Impact of simulation-based closed-loop communication training on medical errors in a pediatric emergency department. *American Journal of Medical Quality*, *35*(6), 474-478.

Dickinson, T. L., & Mclntyre, R. M. (1997). A conceptual framework for teamwork measurement. In M. T. Brannick, E. Salas, & C. Prince (Eds.), *Team Performance and Measurement: Theory, Methods, and Applications*, 19-43. Mahwah, NJ: Lawrence Erlbaum Associates.

Gulliksen, J., Göransson, B., Boivie, I., Blomkvist, S., Persson, J., & Cajander, Å. (2003). Key principles for user-centred systems design. *Behaviour and Information Technology*, *22*(6), 397-409.

Härgestam, M., Lindkvist, M., Brulin, C., Jacobsson, M., & Hultin, M. (2013). Communication in interdisciplinary teams: exploring closed-loop communication during in situ trauma team training. *BMJ Open*, *3*(10).

Jeffcott, S. A., & Mackenzie, C. F. (2008). Measuring team performance in healthcare: review of research and implications for patient safety. *Journal of Critical Care*, *23*(2), 188-196.

King, M., Sikorski, E., & Goodwin, G. (2023). Diagnosing critical incidents during army battle drills. *Proceedings of the 14th MODSIM World*, Norfolk, VA.

O'Donovan, M. P., Hancock, C. L., Coyne, M. E., Racicot, K., & Goodwin, G. A. (2023). Assessing the impact of dismounted infantry small unit proficiency on quantitative measures of collective military performance part 1: Recommended test methodologies.

Schramm, W. (1997). The beginnings of communication study in America: A personal memoir. (S. Chaffee & E. Rogers, Eds.). Thousand Oaks, CA: Sage Publications.

Wasserman, S and Faust, K. (1994). Social network analysis: Methods and applications (Cambridge University).