ARI-MS-MR 2024-05
Military Series
Military Report

February 2024

Evaluation of the Squad Threat Assessment Tool

Drew A. Leins
Joshua Shireman
Aptima, Inc.

Scott Flanagan
Sophia Speira, LLC

Christopher L. Vowels
U.S. Army Research Institute

United States Army Research Institute for the Behavioral and Social Sciences

United States Army Research Institute
for the Behavioral and Social Sciences

A Directorate of the Department of the Army
Deputy Chief of Staff, G1

Authorized and approved:

Scott B. Shadrick, Ph.D.
Acting Director

Research accomplished under contract
for the Department of the Army by

Aptima, Inc.

Research Unit Chief
Dr. Brian T. Crabb (Fort Cavazos Research Unit)

Team Leader
Dr. Christopher L. Vowels

Technical Reviewers

Dr. Cassie A. Berry, U.S. Army Research Institute
Dr. Nicole A. Morales, U.S. Army Research Institute

DISPOSITION

This Technical Report has been submitted to
the Defense Technical Information Center (DTIC).

REPORT DOCUMENTATION PAGE

1. REPORT DATE (Month Year)	2. REPORT TYPE	3. DATES COVERED (Month Year)	
		START DATE	END DATE
February 2024	Final	August 2019	November 2022

4. TITLE AND SUBTITLE
Evaluation of the Squad Threat Assessment Tool

5a. CONTRACT NUMBER	5b. GRANT NUMBER	5c. COOPERATIVE AGREEMENT NUMBER
W911NF-19-F-0030		

5d. PROGRAM ELEMENT NUMBER	5e. PROJECT NUMBER	5f. TASK NUMBER	5g. WORK UNIT NUMBER
633007	A792	205	

6. AUTHOR(S)
Leins, Drew A., Shireman, Joshua, Flanagan, Scott, Vowels, Christopher L.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Aptima, Inc.
12 Gill Street, Suite 1400
Woburn, MA 01801

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)	11. SPONSOR/MONITOR'S REPORT NUMBER(S)
U.S. Army Research Institute for the Behavioral and Social Sciences 6000 6th Street (Bldg. 1464 / Mail Stop: 5610) Fort Belvoir, Virginia 22060-5610	ARI	Military Report 2024-05

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES
ARI Research POC: Dr. Christopher L. Vowels, Fort Cavazos Research Unit

14. ABSTRACT
The U.S. Army Research Institute for the Behavioral and Social Sciences (ARI) Fort Cavazos Research Unit seeks to improve measures of collective performance, to support underlying skill development. In previous focus groups, Platoon Sergeants and other leaders reported needing: (1) a method for evaluating the quality of decision making within a training exercise, (2) a more specific metric that supplements the current “Go vs. No-Go” performance metric, and (3) guidance on how to convey performance requirements to Soldiers. In response, testing and refining was conducted for a tool developed to assess the cognitive skills underlying a range of Army collective tasks that involve threat detection. Two previously developed versions of the Squad Threat Assessment Tool (STAT) were examined within the context of a relevant Army training exercise. The more functional version was refined to optimize its validity and usability. The refined tool will allow Army leaders to better evaluate and remediate collective performance on tasks involving threat detection and assessment.

15. SUBJECT TERMS
Threat detection, Training, Evaluation, After Action Review

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES
a. REPORT	b. ABSTRACT	C. THIS PAGE		
Unclassified	Unclassified	Unclassified	Unclassified Unlimited	62

19a. NAME OF RESPONSIBLE PERSON	19b. PHONE NUMBER (Include area code)
Brian T. Crabb	254-383-1132

ACKNOWLEDGMENT

We would like to thank the instructors and students of the Dismounted Counter-Improvised Explosive Device (IED) Tactics Master Trainer Course for their time and efforts in supporting this research.

EVALUATION OF THE SQUAD THREAT ASSESSMENT TOOL

EXECUTIVE SUMMARY

Research Requirement:

The U.S. Army Research Institute for the Behavioral and Social Sciences (ARI) Fort Cavazos Research Unit is engaged in research that is designed to improve the way the U.S. Army measures collective performance. The long-term goal of that research is to develop valid measures of collective performance that correspond to underlying skill development. Here, we report on the development and evaluation of a tool—the Squad Threat Assessment Tool (STAT)—for assessing cognitive skills that underlie a range of Army collective tasks that involve threat detection.

To support training observation and evaluation, we developed the STAT to (1) outline the behaviors critical to threat detection, (2) offer a more granular scale for rating Soldiers' performance on those critical behaviors, (3) enable efficient and complete documentation of observations, and (4) inform comprehensive and specific feedback for Soldiers. The goal of this research was to test and refine the STAT based on user feedback. We developed and refined a digital mobile application version of the STAT, which we implemented and evaluated in the context of a counter-IED tactics training course. This evaluation included measures of users' perceptions of the utility and usability of the STAT.

Procedure:

We situated the research in the Dismounted Counter-Improvised Explosive Device Tactics Master Trainer (DCT-MT) course which trains students to assess and mitigate IED threats and to incorporate countering improvised explosive devices (C-IED tactics, techniques, and procedures (TTPs) into unit training. The course culminates in a set of three situational training exercises (STXs), which provided the specific context for our evaluation of the STAT. Two course instructors each used the application to record their observations during one STX and then completed two questionnaires following its use. Students who were assigned leadership roles for each STX answered questions about the quality of the feedback they received from the instructors.

The version of the STAT application that instructors implemented during STXs was adapted from two paper-based versions. We developed the digital STAT because of limitations inherent to paper-based notetaking while observing STXs, for example, limited space to record notes and rigid organization of notes. To overcome these (and other) limitations, we developed a digital version of the STAT that could be accessed on a mobile device and would enable users to take comprehensive notes, move fluidly across observations, and store and access observations easily for use in after action reviews (AARs).

Findings:

Feedback on the utility and usability of the STAT application was positive (overall mean rating across items was 3.68, SD = 0.64, on a 5-point scale on which 5 was the most positive rating). Thirteen of seventeen items yielded positive ratings from instructors, implying that the application as tested was largely useful and usable. In addition, feedback on the impact of the STAT application on instructors' ability to observe and provide feedback for training exercises was overwhelmingly positive (overall mean rating across items was 4.09, SD = 0.26). Trainees' ratings of the feedback they received (both with and without the support of the STAT) were overwhelmingly positive.

Utilization and Dissemination of Findings:

Instructors who used the STAT suggested that they would use a refined version in their course. They also indicated that the STAT may be a suitable candidate for other specialized training courses in which threat detection is trained. In particular, courses such as the Advanced Situational Awareness course, the Scout Leader course, and the Military Advisor Training Academy (MATA) could benefit from slightly modified versions of the STAT. To assess the STAT's relevance to courses other than the DCT-MT course, interviews and focus groups could be conducted with a sample of relevant leaders, trainers, and stakeholders. In addition, demonstrations or brief trainings with potential end users could help identify potential stakeholders' interest in using the STAT.

CONTENTS

Page

APPENDICES

LIST OF FIGURES

EVALUATION OF THE SQUAD THREAT ASSESSMENT TOOL

Introduction

The U.S. Army Research Institute for the Behavioral and Social Sciences (ARI) Fort Cavazos Research Unit is engaged in research that is designed to improve the way the U.S. Army measures collective performance. The long-term goal of that research is to develop valid measures of collective performance that correspond to underlying skill development. Here, we report on the development and evaluation of a tool for assessing some of the cognitive skills, such as hypothesis generation and testing, underlying a range of Army collective tasks that involve threat detection.

For this research and development effort, we contextualized threat detection in terms of dismounted tactics for countering improvised explosive devices (C-IED). Specifically, the research was conducted with instructors who trained Soldiers in C-IED tactics and then evaluated those Soldiers' ability to detect and mitigate mock improvised explosive devices (IEDs) in situational training exercises (STXs). The value of training C-IED skills cannot be understated. The IEDs are an enduring threat. They are relatively cheap and easy to manufacture and emplace; they can be used by all forces across nearly any operational scenario; and, critically, they are effective across tactical, operational, and strategic levels (Benson, 2012). Consequently, they will continue to be favored by technically inferior adversaries in conflict with superior forces.

The tool we evaluated in the C-IED context—the Squad Threat Assessment Tool (STAT)—was developed to enable instructors, observers, coaches, and trainers (OCTs) to more efficiently and comprehensively document their observations of individual and collective behaviors that lead to successful threat detection and to provide feedback about those behaviors during after action reviews (AARs). Specifically, the STAT was developed to support observation and documentation of the behaviors that reflect the cognitive processes critical to problem solving in threat detection, for example, cue perception and interpretation or assessment. The ability to provide feedback on the processes (rather than just the outcomes) of problem solving should enhance instructors' and OCTs' efforts to improve individual and collective performance on tasks that involve threat detection.

Threat detection can be a complex cognitive task. It involves attending to and interpreting a potentially large number of environmental cues and generating hypotheses to explain those cues (Zimmerman, et al., 2014). Across a variety of decision contexts, decision makers tend to generate hypotheses with little conscious effort (Thomas, et al., 2008). For example, when decision makers identify cues in their environment, they may quickly match those cues to details in their memory for previously experienced situations. This is essentially a recognition process, which serves to aid decision makers in interpreting current situations and identifying candidate courses of action based on past similar situations (e.g., the Recognition-Primed Decision model; Klein, 1993, 1997). Recognition-based decision making enables decision makers to choose a course of action and predict an outcome swiftly and confidently.

However, successfully employing this type of hypothesis generation generally requires two conditions: familiar environments and experienced decision makers. When these conditions are not met, when environments are unfamiliar or decision makers are inexperienced, recognition may not occur, and decision makers must either generate new hypotheses or entertain existing hypotheses that may or may not fit the current situation. These alternatives may result in slower, less certain, and less accurate decision making. Indeed, we observed this in experiments in which groups of Soldiers worked collectively to identify threats in familiar and unfamiliar decision environments (Leins, et al., 2017; Leins, et al., 2018a). When environments were familiar (e.g., identifying potential threats in an operational context), groups of experienced Soldiers were more likely to generate (or modify) hypotheses with optimal timing—that is, they formally identified a new hypothesis, or they modified an existing hypothesis when highly informative cues were introduced to the decision environment. By contrast, when the decision environment was unfamiliar (e.g., a medical triage scenario), groups of Soldiers (who were inexperienced in medical triage) did not generate or modify their hypothesis with optimal timing; rather, they tended to delay hypothesis generation until they had viewed and assessed all possible information in the decision environment, regardless of when high-value information was introduced. These groups of Soldiers generated hypotheses with different efficiency, likely as a function of differences in recognition across decision environments. Whereas Soldiers could recognize and differentiate between informative and uninformative cues in familiar environments, they could not similarly differentiate between informative and uninformative cues in unfamiliar environments. That inability to differentiate between cues required Soldiers to evaluate a larger set of cues, possibly evaluate those cues more effortfully, and take longer to generate hypotheses.

The efficiency afforded by recognition is characteristic of cognitive processing via heuristics (or mental shortcuts). A heuristic is a strategy for ignoring some information while using other information to reduce the complexity of decision making so that it can occur quicker or more efficiently (Gigerenzer & Gaissmeier, 2011). In the context of threat detection, heuristics can be used to reduce the number of environmental cues needed to decide threat risk. Threat detection in an operational environment is a prime candidate for leveraging heuristics, as it is often characterized by constraints on knowledge and time. Soldiers are not always familiar with the environments they operate in, nor do they often have time to deliberate upon the meaning of environmental cues or the best course of action in an uncertain situation. When the cost of slow, effortful deliberation appears prohibitive (e.g., life threatening), it will be sacrificed for quicker, more efficient decision processes. Hence, it is critical to train Soldiers how to (1) recognize informative cues in operational environments, for example, when a culvert or a roadside feature is more or less likely to pose a threat (Zimmerman, et al., 2013) and (2) engage heuristics appropriately when they do not recognize cues in operational environments.

Training Soldiers how to recognize threat-relevant cues is relatively straightforward. Instructors can expose Soldiers to relevant and irrelevant cues repeatedly, observe Soldiers' performance in identifying and responding to those cues, and provide feedback on that performance. Over time, Soldiers should learn how best to identify, sort, and prioritize relevant threat cues. Training Soldiers how to think their way through threat detection when they do not recognize cues is slightly more complex and requires another layer of instructional observation. It requires instructors to observe and assess the behaviors that correspond to decision making

processes as those processes occur. For example, it requires instructors to (1) identify the quantity and quality of cues that Soldiers attend to when assessing threats, (2) evaluate Soldiers' interpretations of those cues (e.g., as manifested in intrasquad communications), and (3) assess the courses of action that Soldiers propose to mitigate perceived threats. These requirements can be challenging, not least of all because it involves observing and documenting a lot of activity. Moreover, the cognitive processes and outcomes of threat detection can be influenced by multiple dynamic factors, including the cognitive characteristics of decision makers, the number of decision makers involved, features of the decision environment, and interactions between decision makers and the environment (e.g., see Baumann & DeSteno, 2010).

As Soldiers gain experience and develop proficiency in threat detection, their cognitive processes inherent in detecting threats will likely evolve (and vice versa: as Soldiers mature their cognitive skills for detecting threats, they will likely develop proficiency in threat detection). Hence, instructors may benefit from guidance regarding how to observe, assess, and provide feedback on the cognitive components of threat detection; and they may well benefit from a tool that helps them document and manage the numerous observations required to effectively assess cognition in threat detection.

To identify what guidance or resources would most benefit instructors who train threat detection skills, the findings of the previously described research—specifically, that a familiar context and early access to valuable information allowed groups of Soldiers with domain experience to successfully engage heuristics and generate hypotheses efficiently—were presented to focus groups comprising Army leaders and trainers (Leins, et al., 2018b). These focus groups were asked to consider (1) their current approaches to (and needs for) training and evaluating collective decision making skills in threat detection and (2) their strategies for (and potential hurdles to) implementing activities for training and evaluating collective decision making in threat detection.

Leaders in the focus groups reported that they want their Soldiers to be good decision makers, but that they do not train decision making per se. Hence, they would benefit from training that emphasizes decision making processes that are critical to completing identified tasks. They reported needing training that offers leaders guidance on what to look for when evaluating decision making and how to convey performance requirements to their Soldiers. In addition, leaders reported needing a method for evaluating the quality of decision making within a training exercise as well as a more specific metric that supplements the "Go versus No-Go" metric.

Leaders also indicated that training and evaluation of decision making in threat assessment should be implemented at the squad or team level and could lend itself to opportunity training or even be incorporated into existing training (e.g., in situation training, field training, or live fire exercises). Leaders identified several benefits of implementing a method or tool for evaluating behaviors associated with decision making in threat assessment, including a clarification of standards for new leaders and the potential to develop a more nuanced and valid metric of performance than the current "Go versus No-Go" metric. Leaders suggested that specific feedback on performance, including more detailed performance measures, would offer

Soldiers a better understanding of the skills underlying performance and thus a better chance of improving performance.

Given the feedback offered by leaders along with the data from previous research (e.g., Leins et al., 2018a), we developed a resource that (1) outlines the behaviors critical to threat detection, (2) offers a more granular scale for rating Soldiers' performance on those critical behaviors, (3) can be used to record observations more efficiently and completely during training exercises, and (4) enables instructors and OCTs to provide comprehensive and specific feedback to Soldiers. We began with two slightly different versions of a paper-based STAT. Both versions influenced the final, digital version. Figure 1 presents one of the paper-based versions (see Appendix A – Paper-based Versions of the STAT, for the other version).

Figure 1

One paper-based version of the STAT

Squad Threat Assessment							**Leader:**		
Exercise:							**Date:**		
Rating Scale:	**0 (Unsatisfactory), 1 (Minimum Standard), 2 (Standard), 3 (Above Standard), 4 (Exceeds Standard)**								
	Threats Identified		**Time taken to Identify Threat**		**Communicated Assessment with Team**		**Quality of Assessment**		
Soldier	**Raw**	**Rating**	**Raw**	**Rating**	**Raw**	**Rating**	**Raw**	**Rating**	**Comments:**
Group Assessment:									

The conceptual workflow for the paper-based version of the STAT involved the following: an observer (instructor or OCT) enters contextual data for the exercise they would observe (the grey cells at the top of the form in Figure 1), enters an identifier for each Soldier to be observed (the yellow cells in the first column of the form), and then enters observation data for each evaluated Soldier or for the group. Observers could record raw data (e.g., how many threats were observed, the latency to observe a threat, or a Soldier's assessment of a threat) and then assign that observation a rating according to the scale presented in the third row of the form. Finally, observers could enter comments associated with an observation. All these data would then be available to observers for generating and providing feedback during an AAR.

Research Question

The goal of this research was to test and refine the STAT, a tool for capturing observations of the behaviors that underlie hypothesis generation and collective decision making in tasks involving threat detection. We sought to conduct this work in the context of the current home-station training environment. Initially, we planned multiple research activities that required in-person participation, the first of which would be focus groups for the purpose of evaluating whether the STAT was relevant, useful, and usable for instructors and OCTs who train threat detection skills. However, shortly after we began those activities, the Corona Virus Disease (COVID-19) pandemic emerged, and we were forced to modify our plan. Because of the risk to Soldier health posed by in-person testing—specifically, the risk of transmitting COVID-19—we postponed those activities and instead focused on gathering data virtually and developing a digital version of the STAT. Despite being unable to conduct in-person research activities for most of the duration of the project, we were able to develop and refine (as described in the following Method section) a version of the STAT that we implemented and evaluated in the context of a home-station training course. This evaluation included measures of users' perceptions of the utility and usability of a version of the STAT compatible with mobile devices. The evaluation also included trainees' perceptions of the value of the feedback they received when instructors used the STAT.

Method

Research Design and Context

The research occurred in the context of the Dismounted Counter-Improvised Explosive Device Tactics Master Trainer (DCT-MT) course (Dismounted C-IED Tactics Master Trainer/9E-F59/950-F38). This course trains students to (1) assess IED threats, (2) recommend enablers to mitigate IED threats, (3), incorporate C-IED tactics, techniques, and procedures (TTPs) into unit training, and (4) conduct a company level C-IED training program. The DCT-MT course is a 10-day course implemented over two weeks, culminating in a set of three situational training exercises (STXs). These STXs provided the specific context for our evaluation of the STAT.

The three STXs that conclude the course are each executed at one of two training stages: walk or run. At each of these stages, exercises are characterized by different levels of complexity, difficulty, and instructor involvement. At the walk stage, instructors introduce challenges that require students to solve multiple problems using skills and TTPs that they learned during classroom training. At specified times during the exercise, instructors may pause the exercise to review student actions and remediate those actions as needed. At the run stage, instructors present realistic scenarios with components that are similar to what students might experience during deployment. During the run stage, instructors do not intervene while the exercise is in progress; rather, they provide cumulative feedback at the end of the exercise. For these walk- and run-stage exercises, students in the DCT-MT receive a Battalion/Company Operations Order (OPORD) followed by Troop Leading Procedures (TLPs). Upon completion of TLPs, students execute a platoon-level mission that incorporates IED threats. Instructors evaluate students on their ability to mitigate IED threats and accomplish the overall mission.

We sought to test the utility and usability of the STAT app in the context of the STXs described above. The research team visited the DCT-MT course twice—once to demonstrate the STAT application and elicit feedback on its face validity (i.e., its content and general workflow), and a second time to support instructors' implementation of the application during STXs. In between visits, we revised the application per instructors' feedback (we describe this and other STAT application development activities in the Materials section).

To evaluate the STAT application's utility and usability, we asked two instructors each to use the application to record their observations during one STX and then complete two questionnaires following its use (see Appendix B). We also asked students, particularly those in leadership roles during the STXs, to answer questions about the quality of the feedback they received from the instructors (see Appendix B). Some students received feedback from an instructor who had used the STAT, whereas other students received feedback from an instructor who had not used the STAT.

Participants

Two instructors participated as observers/trainers. Each held the rank of Staff Sergeant (SSG), occupied the duty position of instructor, and had at least 24 months' experience in that duty position. They self-rated as having intermediate or expert knowledge in the domain of threat detection, and they each reported training Soldiers in threat detection at a frequency of more than 4 times per year. Each instructor completed three questionnaires: STAT utility, STAT usability, and demographics.

Thirty-two students attended the DCT-MT course during the observed STXs. Of those 32 students, 11 participated as leaders during one of the observed STXs. These 11 students all held the rank of Second Lieutenant (2LT). Each of these 11 students completed a feedback quality questionnaire.

Materials

This section briefly describes the STAT application and the activities that the research team engaged in to design and develop it. Additional details regarding the application can be found in the unpublished user guide (Aptima, 2022).

STAT Design. The version of the STAT application that instructors implemented during STXs initially began as a paper-based tool. The original research plan involved evaluating and then enhancing that version's utility and usability. However, because we were prevented from conducting in-person research activities with course personnel (because of the public health risks associated with COVID-19) but still wanted to advance the tool, we opted to adapt the paper-based version to provide the foundation for the workflow of a digital, mobile application. To begin designing the mobile application, the research team held internal design sessions with user-experience experts to better understand design principles and how the initial, paper-based STAT could be improved. These sessions yielded critical questions that became the items in our utility

and usability questionnaires (Appendix B), and they prompted the research team to consider the following four critical aspects of observing and evaluating student performance during STXs:

1. The paper-based version offered limited space for writing notes, yet the observations and assessments we anticipated instructors would require substantial notetaking. Hence, we determined that a paper-based version might be too constraining and that a more flexible application that did not confine notes to a fixed space would be optimal.
2. The quick pace of STXs might impose time constraints on notetaking that could lead to incomplete or illegible handwritten notes. We determined that a version that enabled quick and legible notetaking would be optimal.
3. The combination of limited notetaking space and a prominent section for quantifying observations (i.e., rating performance quantitatively) might have overestimated the importance and ease of rating behavior quantitatively. Some behaviors may be more difficult to rate on a quantitative scale than others. When real-time quantitative ratings are too difficult, cognitively demanding, or inefficient, notetaking modes should enable users to record observations comprehensively so that users can make post-hoc quantitative ratings if necessary.
4. As users observe larger numbers of students, it becomes cumbersome to manage the multiple sheets of paper required to document observations for those students. Hence, we determined that a version that allowed users to record and access observations more easily for multiple students (e.g., squads or multiple squads) would be optimal.

These insights led us to consider a digital version of the STAT that could be accessed on a mobile device. A mobile, digital version would enable users to take comprehensive notes, move fluidly across observations, and store and access observations easily for use in after action reviews (AARs).

STAT Development. To identify the best application platform and workflow to enable the functionality that would best support instructors and OCTs, the research team conducted an analysis of alternatives. We held multiple requirements meetings to explore the potential of developing a digital, mobile version of the STAT. We evaluated multiple existing platforms and applications—for example, SPOTLITE® (MacMillan, et al., 2013) and the Enlisted-Leader Attribute Requirements Application (E-LARA; Ingurgio, et al., 2020)—each of which allows users to record contemporaneous or post-hoc assessments of students on a variety of measures. However, those platforms were developed for more complex workflows to which we could not adapt the STAT workflow; consequently, it would have required a prohibitive level of effort to recode those applications to host the STAT workflow.

Instead, we opted to pursue a lightweight application using the Ionic framework (https://ionicframework.com/docs). Ionic is an open-source toolkit for building mobile applications using web technologies, such as Hypertext Markup Language (HTML), Cascading Style Sheets (CSS), and JavaScript. It allows for the development of simple, customizable applications that run on iPhone operating system (iOS) and Android operating systems. To inform the coding of the application workflow, we took inspiration from an existing Aptima software, the Aircrew Performance Measurement and Proficiency System (A-PuMPS; Bunn, et al., 2020). A critical feature of A-PuMPS is that it enables users to input large amounts of notes

that can be easily accessed and sorted for use during an AAR. We developed an initial Ionic-based version of the STAT and then refined it iteratively via internal testing, as described in the following paragraphs.

Enhancements to the original paper-based workflow included options to populate an observation session with event (training) context data (or metadata) to enable instructors to track performance across training scenarios. Additional enhancements included options to enter metadata for squad composition and threat stimuli (i.e., the threat cues, such as trip wires or buried IEDs placed in an STX environment), as well as the capability to save and retrieve all data for use during an AAR. We demonstrated this initial digital version of the STAT to a subject matter expert—a retired Master Sergeant and recent senior analyst and consultant in support of the Joint Improvised Explosives Device Defeat Organization (JIEDDO) and the U.S. Army Asymmetric Warfare Group (AWG). This demonstration resulted in updates to options for documenting event context data (metadata), for recording assessment data for collective performance, for tagging observations with relevant key words, and for organizing observation data for use in an AAR. Following these updates, the research team completed development of a functioning instance of the mobile-based STAT application. The application enabled users to do the following (see Appendix C for screen captures of these features):

1. Record contextual data.
2. Populate a standard fire team whose individual team members are selectable for recording member-specific data (e.g., whether the automatic rifleman identified a threat, how they assessed the threat risk, and what course of action they recommended in response).
3. Populate a list of threats or cues for a given STX environment.
4. Pair a team member and a specific threat cue for categorizing an observation (e.g., the automatic rifleman identified a cue as relevant to a vehicle-borne IED).
5. Record observation-specific key words (e.g., the urgency or risk level of a threat cue), as well as general notes (e.g., in text boxes for recording open narratives).
6. Access observational data, organized by Soldier, by threat, and on a timeline.

The research team then met telephonically with the managers and instructors of the DCT-MT course to introduce the application and discuss its potential use and benefits. The course managers and instructors reviewed a slide deck of screen captures of the STAT and annotated that slide deck with requests for additional content (e.g., key words or terms) and functions. The research team subsequently revised the application based on this feedback, which consisted primarily of requests to add course-specific terms to make notetaking more efficient, and to modify how users label and select the units of observation (e.g., identification of Soldiers and the equipment assigned to them for an STX). Following this revision, the research team visited the DCT-MT course to demonstrate the STAT application. During this meeting, instructors practiced using the application outside the context of an STX (e.g., during a mock table-top exercise without students) and provided additional feedback and recommendations. One noteworthy sentiment shared by multiple instructors was that the STAT application would have value outside the context of the DCT-MT course. These instructors suggested that other courses for training specialized skills (for example, courses administered by the Military Advisor Training Academy

[MATA]), would benefit from slightly modified versions of the STAT. For the DCT-MT course, however, instructors recommended the following modifications:

- Tap one button to add a full squad (prepopulated with a squad leader and two fire teams) as a unit of observation (in addition to individual fire teams and Soldiers);
- Enable the assignment of equipment to individual Soldiers to allow for flexibility in Soldier-equipment pairings;
- Enable scrolling of all screens to accommodate several soldiers and threats in each observed exercise; and
- Enable export of data to pdf files.

The research team modified the STAT accordingly and produced the version that was tested in STXs, as described in the following section. The application was installed on two tablets that use the Android operating system. Instructors used these tablets to access the STAT during the field test (the STXs). Thus, the final set of materials for the field test included the mobile STAT application; utility, usability, and demographic questionnaires (for instructors); and a feedback quality questionnaire (for students).

Procedure

Two instructors each implemented the STAT application during one of two full-day (6-8 hours) STXs that comprised planning, briefing (OPORD), execution, and feedback (AAR). One instructor used the STAT during the STX walk stage; the other instructor used it during the run stage. Instructors used the STAT application to record notes as they observed OPORD and execution phases; and they used the STAT to access notes to provide feedback to students during the AAR phases. During each execution phase of the walk- and run-stages, one instructor observed three squads. Each of the instructors provided utility and usability feedback to the research team during and after each exercise. They provided qualitative feedback during unstructured and semi-structured interviews with researchers at the conclusion of each STX phase. At the conclusion of the STXs, instructors also provided quantitative feedback to the research team via formal questionnaire, and gave final thoughts for improvements to the STAT.

Trainees who were assigned leadership roles during observed exercises also provided quantitative feedback to researchers via a formal questionnaire at the conclusion of all training exercises. Some of these trainees received feedback for performance that was observed without the benefit of the STAT application. After all participants (instructors and trainees) completed their questionnaires, they were debriefed.

Analyses

Qualitative data (i.e., instructor feedback and recommendations provided via unstructured and semi-structured interviews) were recorded and translated into formal recommendations for refining the STAT. Quantitative data were aggregated and analyzed with simple descriptive statistics (i.e., means and standard deviations; see Appendix D for the table of quantitative data). An overall mean was calculated for all the items on each questionnaire after adjusting for reverse-coded items (i.e., after applying a correction of 5 - n, where n = the participants' mean rating).

Results

Quantitative Data

In general, feedback on the utility and usability of the STAT application was positive (overall mean rating across items was 3.68, *SD* = 0.64) on a 5-point scale, with 5 representing the most positive rating possible. Thirteen of seventeen items yielded positive ratings from instructors, implying that the application as tested was largely useful and usable. However, one item yielded negative ratings: At times, I lost track of where I was in the tool. Both instructors indicated that they had lost track of their location in the tool (see also qualitative feedback, below). Other items that failed to yield positive ratings included the following (average rating in parentheses; "*" indicates a consensus rating between instructors):

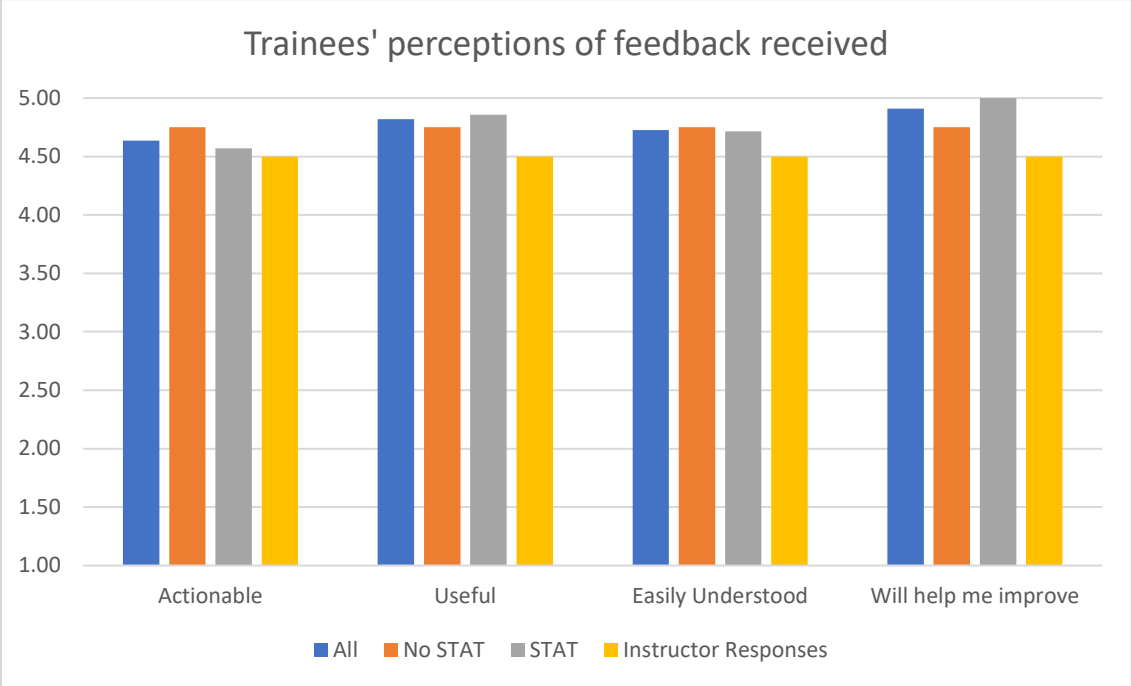
- It was easy to tell if I had completed an assessment (Neither agree nor disagree).
- I would need help/support to use the tool (Neither agree nor disagree)*.
- When I made a mistake, it was easy to undo my error (Neither agree nor disagree)*.

In general, feedback on the impact of the STAT application on instructors' ability to observe and provide feedback for training exercises was overwhelmingly positive (overall mean rating across items was 4.09, *SD* = 0.26), using the same 5-point scale.

Figure 2 presents the quantitative data provided by trainees. There are data included from seven trainees who received feedback from instructors who used the STAT during STX observations, and from four trainees who received feedback from instructors without the STAT application. Figure 2 also includes instructors' ratings of the feedback they provided to trainees (the final four items in the table of ratings in Appendix D).

Figure 2

Quantitative data provided by trainees



Note. The X-axis item labels represent the following statements:
Actionable: "The feedback that I received is actionable."
Useful: "The feedback that I received is useful."
Easily Understood: "I was easily able to understand this feedback."
Will help me improve: "This feedback would help me improve my performance."

Qualitative data

Specific observations made by instructors about the STAT application during and after implementation can be found in Appendix E. The essence of instructors' comments is that the STAT offered benefits beyond what paper-and-pencil notetaking offers, that it enabled them to provide personalized feedback to Soldiers, and that it forced them to attend to elements of the STX to which they typically might not attend. One major critique of the STAT was that its full complement of features seemed overwhelming and was more than the instructors needed.

Researchers observed that one instructor used only the notes section in the application—that is, he did not use the STAT to categorize his observations by trainee action (e.g., identify threat, interpret threat cues, recommend course of action). When asked why that was, the instructor responded that he needed the application to quickly capture observations and that if time allowed, he would use the other features. Moreover, he reported that if he had only one group to observe, he would have more time to use the other features. When asked to identify the most frustrating aspect of using the STAT, this instructor reported that he was frustrated by losing track of where his notes were stored in the application (see Appendix D for additional comments provided by instructors).

Discussion

In general, feedback on the utility and usability of the STAT application was positive. Thirteen of seventeen general utility/usability questionnaire items yielded positive ratings from instructors, implying that the application as tested was largely useful and usable. However, the following item yielded negative ratings: "At times, I lost track of where I was in the tool." Both instructors indicated that they had lost track of their location in the tool. This aligned with their observation that the tool had too many features beyond the three features they need—namely, the ability to document (1) what the threat was, (2) who saw it, and (3) what they did. Other items that failed to yield positive ratings included the ease with which instructors could tell whether they had completed an assessment (rated as, "neither agree nor disagree"), whether they would need help or support to use the tool (rated as, "neither agree nor disagree"), and whether it was easy to correct an error (rated as, "neither agree nor disagree").

It is worth noting, however, that both instructors commented that their ability to navigate the STAT, including maintaining awareness of where they were in the application (i.e., in what function or feature), would improve with practice. Both instructors accessed the test version of the STAT only one day prior to implementing it in an STX and thus were still somewhat unfamiliar with all its features and functions. Despite the challenges experienced by instructors, they both indicated a willingness to use the tool again. Both instructors gave the tool high ratings for its potential to support them in providing high quality feedback to trainees. In general, instructors perceived that the feedback informed by the STAT was actionable, useful, understandable, and helpful in improving Soldier performance.

Critically, both instructors noted that the STAT offered benefits that pencil-and-paper notetaking does not readily offer. For example, timestamping observations for specific Soldiers enabled instructors during an AAR to easily recount critical actions made by each Soldier observed throughout the STX. Tailored and specific feedback is often more actionable for individuals than is general feedback aimed at the group level; therefore, tailored, specific feedback often leads to better learning and subsequent performance outcomes (Kluger & DeNisi, 1996). In addition, being able to segment, label, and organize observations for each phase of an STX prompted instructors to pay better attention to more elements of each phase than they typically would if they were recording notes in a notebook. This enabled instructors to capture more comprehensive evaluation notes and transform those notes into feedback that otherwise would not have been given to trainees. Finally, being able to record and retrieve observation notes specific to a squad was useful, particularly as each instructor monitored up to three squads in an STX. If they were relegated to recording observations in a notebook, they would have to either organize those notes into squad-specific sections a priori and navigate through the notebook during the STX or else keep a running log of notes, identify each note with squad-specific label, and then search for those labels in the notebook during an AAR. With the STAT, however, they could simply tap a Soldier identifier, type a note, and the application would timestamp the note and organize it in whichever way was most useful to the instructor (e.g., by squad, by time of observation, or by threat type).

The STAT also showed potential for use beyond recording observations and informing feedback for an AAR. Instructors commented on its potential to support their administrative

tasks. Specifically, instructors hold their own internal AARs to identify opportunities to improve their instruction and the course. Moreover, they must report their instructional activities to leadership. Instructors noted that the STAT would support both of those activities, by providing access to detailed notes that can inform reflections on their own instructional performance and which can be organized into storyboard format for reporting up to leadership.

Of course, the STAT presents with limitations, some of which instructors identified and translated into recommendations for improvement. The research team modified the STAT to accommodate some of those recommendations. For example, we moved text-based data to a more prominent position in the AAR section so that users can access them quicker. We also enabled better labeling and organization of phase-specific data to give users more options to segment a workflow a priori and then to be able to access and re-order those data in the AAR section. Finally, we modified how squads are set up and edited to allow easier addition and deletion of individual squad members.

In previous research, we explored how individual and small groups of Soldiers make decisions regarding potential threats (Leins, et al., 2017) and have explored similar kinds of decisions made in familiar and unfamiliar contexts (Leins et al., 2018a). The Army anticipates the future battlefield will require small teams of Soldiers to be self-reliant and operate in a highly contested operational environment (Department of Army, 2022). As a result, small teams will face both familiar and unfamiliar threats across multiple domains to include IEDs. The present research extended the previous findings where we focused on the Soldier or trainee to the instructor who must train Soldiers on detecting IEDs during dismounted operations. In this regard, the research conducted to date provides a good foundation to further explore facets that impact both students and instructors, especially as small teams prepare for future operations.

Methodological Limitations

In addition to limitations posed by the STAT application, there were some limitations to the research methods we used. First, and most critical, we were able to engage only three course instructors for initial feedback on the STAT application, and we were able to engage only two instructors for the field test of the STAT application. Thus, even though the two instructors who tested the STAT are the instructors who will ultimately use it, we do not assume that their experience and perspectives are representative of all potential end users. Moreover, these two instructors observed only one STX each. Consequently, we were unable to capture data across repeated exercises or observations and hence we could not analyze for the impact of the STAT on trainee performance. If the STAT is to be considered for further transition (e.g., to other courses or training contexts), it should be tested with a larger sample of users and across repeated exercises.

In the context of the test reported here, it is important to note that instructors had limited time to become familiar with the STAT application. Ideally, instructors would have had multiple practice sessions to learn the different features and functions of the application, described in the user's guide. As they noted, if they had more time and familiarity with the STAT, they may have been able to avoid some challenges with using it, such as issues with navigating its multiple functions.

Recommendations and Future Directions

Despite the limitations of the research and the STAT itself, some recommendations for implementation and further testing can be considered. First, as noted by the instructors who tested the STAT, it may be a suitable candidate for other specialized training courses in which threat detection is trained. In particular, courses such as the Advanced Situational Awareness course, the Scout Leader course, and the Military Advisor Training Academy (MATA) could benefit from slightly modified versions of the STAT. To assess the STAT's relevance to courses other than the DCT-MT course, interviews and focus groups could be conducted with a sample of relevant leaders, trainers, and stakeholders. In addition, demonstrations or brief trainings with potential end users could help identify potential stakeholders' interest in using the STAT.

Although we were unable to empirically determine the impact of the STAT on instructor and trainee performance, we were able to take an important first step in evaluating the impact of transitioning from recording observations by hand in a notebook to recording observations digitally in a mobile application. Through this research effort, we began to understand the benefits and detriments of making that change. One unanticipated result was that instructors reported not needing all functions and features of the STAT. They reported needing to capture only three high-level categories of data (what was the threat, who identified, and what did they do in response). However, previous research implicates that evaluating more specific behavioral and cognitive components of threat detection can help support improvement in threat detection performance (Leins et al., 2018b). Thus, future research is needed to reconcile these two competing sentiments and to evaluate whether specific features of the STAT application (e.g., measuring latency to identify a threat, rating accuracy of cue interpretation, assessing proficiency in communicating assessments or courses of action to teams) will in fact support improvements in performance.

Acronym List

Acronym:	Definition:
AAR	After Action Review
A-PuMPS	Aircrew Performance Measurement and Proficiency System
AWG	U.S. Army Asymmetric Warfare Group
C-IED	Countering Improvised Explosive Devices
COVID-19	Coronavirus Disease 2019
CSS	Cascading Style Sheets
DCT-MT	Dismounted Counter-Improvised Explosive Device Tactics Master Trainer
E-LARA	Enlisted-Leader Attribute Requirements Application
HTML	Hypertext Markup Language
IED	Improvised Explosive Device
iOS	iPhone Operating System
JIEDDO	Joint Improvised Explosives Device Defeat Organization
MATA	Military Advisor Training Academy
OCT	Observer, Coach, Trainer
OPORD	Operation Order
SSG	Staff Sergeant
STAT	Squad Threat Assessment Tool
STX	Situational Training Exercise
T&EO	Training and Evaluation Outlines
TLP	Troop Leading Procedures
TTP	Tactics, Techniques, and Procedures

References

Aptima. (2022). *Squad Threat Assessment Tool – User Guide*. Aptima, Inc.

Baumann, J., & DeSteno, D. (2010). Emotion guided threat detection: Expecting guns where there are none. *Journal of Personality and Social Psychology*, *99*(4), 595-610.

Benson, C. M. (2012). *Detecting improvised explosive devices: Enduring threat requires enduring solutions*. Army War College. https://apps.dtic.mil/sti/citations/ADA560881

Bunn, M., Ditch, P., & Fegley, B. D. (2020). Aircrew performance measurement and proficiency: A need for better operational capabilities. In S. Nazir, T. Ahram, & W. Karwowski (Eds.), *Advances in human factors in training, education, and learning sciences. AHFE 2020. Advances in Intelligent Systems and Computing* (Vol. 1211, pp. 255-261). Springer, Cham. https://doi.org/10.1007/978-3-030-50896-8_37

Department of the Army (2021). *Training* (Field Manual 7-0). Author.

Department of the Army. (2022). *Operations* (Field Manual 3-0). Author.

Dismounted C-IED Tactics Master Trainer/9E-F59/950-F38. https://www.benning.army.mil/armor/316thcav/DCT-MT/

Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic decision making. *Annual Review of Psychology, 62*, 451-482.

Ingurgio, V. J., Ratwani, K. L., Nargi, B., Flanagan, S., Diedrich, F. J., & Toumbeva, T. H. (2020). *Tools for Assessing and Tracking Leadership Attributes and Competencies.* (Research Report 2030). U. S. Army Research Institute for the Behavioral and Social Sciences.

Klein, G. A. (1993). *A recognition-primed decision (RPD) model of rapid decision making*. Ablex Publishing Corporation.

Klein, G. (1997). The recognition-primed decision (RPD) model: Looking back, looking forward. In C. E. Zsambok & G. Klein (Eds.), *Naturalistic Decision Making* (pp. 285-292). Lawrence Erlbaum Associates, Inc.

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: a historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin, 119(2)*, 254-284. https://doi.org/10.1037/0033-2909.119.2.254

Leins, D. A., Leonard, J., Masters, G., & Vowels, C. L. (2018a). *The Influence of Expertise and Decision Environment on Collective Hypothesis Generation* (Research Report 2018). U. S. Army Research Institute for the Behavioral and Social Sciences.

Leins, D. A., Leonard, J., Masters, G., & Vowels, C. L. (2018b). *Implementing Measures of Individual and Collective Hypothesis Generation: A User's Guide* (Technical Report 1364). U. S. Army Research Institute for the Behavioral and Social Sciences.

Leins, D. A., Leonard, J., Zimmerman, L. A., Minchew, M. L., & Vowels, C. L. (2017). *Decision Environment and Heuristics in Individual and Collective Hypothesis Generation* (Research Report 2013). U. S. Army Research Institute for the Behavioral and Social Sciences.

MacMillan, J., Entin, E. B., Morley, R., & Bennett W. Jr., (2013). Measuring team performance in complex and dynamic military environments: The SPOTLITE method. *Military Psychology*, *25*(3), 266-279.

Thomas, R. P., Dougherty, M. R., Sprenger, A. M., & Harbison, J. I. (2008). Diagnostic hypothesis generation and human judgment. *Psychological Review, 115*(1), 155-185.

Zimmerman, L. A., Leins, D. A., Marcon, J., Pearlman, B., Singer, J.T., Mueller, R., & Vowels, C.L. (2013). *Improving Visual Threat Detection: Research to Validate the Threat Detection Skills Trainer* (Technical Report 1329). VA: U. S. Army Research Institute for the Behavioral and Social Sciences.

Zimmerman, L. A., Mueller, S. T., Grover, J., & Vowels, C. L. (2014). *Determining the requisite components of visual threat detection to improve operational performance* (Technical Report 1340). U.S. Army Research Institute for the Behavioral and Social Sciences.

Appendix A

Paper-based Versions of the STAT

STAT – Version A

Squad Threat Assessment						Leader:	
Exercise / Task:				Location / Lane:		Date:	
Rating Scale: 0 (Unsatisfactory), 1 (Minimum Standard), 2 (Standard), 3 (Above Standard), 4 (Exceeds Standard)							
Iteration:	Threats Identified			Situational Assessment (Number each assessment)	Time taken to communicate assessment	Quality of assessment	Feedback from Team
Soldier	Raw	Time	Rating				
STEP 1.					STEP 2.		
					Leader's Assessment	Leader's Original COA	

STEP 3.					STEP 4.	
Soldier	Alternative Situational Assessment	Time to commo to leader	Quality of Assessment	Feedback from Team	(If applicable) Change/update in Leader's COA	Result / Outcome of COA

Additional Notes:

STAT – Version B

Squad Threat Assessment							Leader:		
Exercise:							Date:		
Rating Scale:	0 (Unsatisfactory), 1 (Minimum Standard), 2 (Standard), 3 (Above Standard), 4 (Exceeds Standard)								
	Threats Identified		Time taken to Identify Threat		Communicated Assessment with Team		Quality of Assessment		
Soldier	Raw	Rating	Raw	Rating	Raw	Rating	Raw	Rating	Comments:
Group Assessment:									

Appendix B

Questionnaires

Demographics Questionnaire (for Course Instructors)

1. **Current rank?** _____
2. **Time in that rank?** _____ months
3. **Current Duty Position?** _____ (e.g., Gunner, Squad Leader, Company Commander)
4. **Time in that Duty Position?** _____ months
5. **What is your knowledge level in the domain of threat detection?**

   ☐ Novice

   ☐ Intermediate

   ☐ Expert
6. **How often do you train Soldiers on critical threat-detection behaviors?**

   ☐ Less than once a year

   ☐ Once a year

   ☐ 2-4 times a year

   ☐ More than 4 times a year
7. **List up to five training exercises you have led or evaluated that involved a component of threat detection:**

   1._____

   2._____.

   3._____

   4._____

   5._____

Utility and Usability Questionnaire: General STAT Assessment (for Course Instructors to complete after the practical exercise)

Date: _____ Event/Task: _____

Select the extent to which you agree or disagree with each statement below in using the STAT tool.

Statement	Strongly Disagree (1)	Disagree (2)	Neither Agree nor Disagree (3)	Agree (4)	Strongly Agree (5)	NA – I did not use this feature
1. I knew exactly where to go in the tool to make an assessment						
2. It was easy to tell if I had completed an assessment						
3. At times, I lost track of where I was in the tool						
4. I sometimes found it hard to remember how to use the tool						
5. It was easy to navigate to different performance areas within the tool						
6. I was able to make assessments quickly, even when using the tool for the first time						
7. The layout of the tool allowed me to capture data effectively						
8. The layout of the tool was well organized						
9. There was too much content in the tool						
10. The tool was easy to use						
11. I would need help/support to use the tool						
12. In general, I would find the tool helpful when making an assessment						
13. In general, I would find the tool helpful when providing feedback						
14. In general, I would find the tool helpful for improving Soldier performance						
15. The information presented in the instructions for using the tool helped me make accurate assessments						
16. The information presented in the instructions for using the tool helped me generate actionable feedback						
17. If I made a mistake, it was easy to undo my error						

Training Observation Questionnaire (for Course Instructors to complete after training observation)

Date: _____ Team: _____ Event/Task: _____

Considering the training observation, you just completed, indicate the extent to which you agree or disagree with each of the following statements.

Statement	Strongly Disagree (1)	Disagree (2)	Neither Agree nor Disagree (3)	Agree (4)	Strongly Agree (5)	NA – I did not use this feature
1. This training observation was mentally demanding						
2. This training observation was physically demanding						
3. I did not have enough time to observe and assess student performance						
4. I was successful in accomplishing what I was asked to do						
5. I had to work hard to observe and assess student performance						
6. I was effective in **observing** Soldiers' identification of threat-relevant or irrelevant cues						
7. I was effective in **observing** Soldiers' accuracy in interpreting threat-relevant or irrelevant cues						
8. I was effective in **observing** Soldiers' proficiency in communicating threat-relevant or irrelevant assessments to their teams						
9. I was effective in **assessing** Soldiers' identification of threat-relevant or irrelevant cues						
10. I was effective in **assessing** Soldiers' accuracy in interpreting threat-relevant or irrelevant cues						
11. I was effective in **assessing** Soldiers' proficiency in communicating threat-relevant or irrelevant assessments to their teams						
12. I was effective in conveying relevant feedback to Soldiers						
13. My assessment was helpful for improving Soldier performance						

16. Please provide some example feedback that you gave to the team that just completed this training exercise:

Think about the example feedback you provided in the box above when responding to the following statements:

Statement	Strongly Disagree (1)	Disagree (2)	Neither Agree nor Disagree (3)	Agree (4)	Strongly Agree (5)	NA
17. The feedback that I provided is actionable						
18. The feedback that I provided is useful for Soldiers						
19. Soldiers were easily able to understand this feedback						
20. This feedback will help Soldiers improve their performance						

Trainee Questionnaire (for Trainees to complete after completing the training exercise)

Date: _____ Team: _____ Event/Task: _____

Think about the feedback provided to you during the exercise you just completed when responding to the following statements. Indicate the extent to which you agree or disagree with each statement below:

Statement	Strongly Disagree (1)	Disagree (2)	Neither Agree nor Disagree (3)	Agree (4)	Strongly Agree (5)	NA
1. The feedback that I received is actionable						
2. The feedback that I received is useful						
3. I was easily able to understand this feedback						
4. This feedback would help me improve my performance						

Supplemental Trainer Questionnaire for Debrief (to be asked of Course Instructors after training exercise is complete)

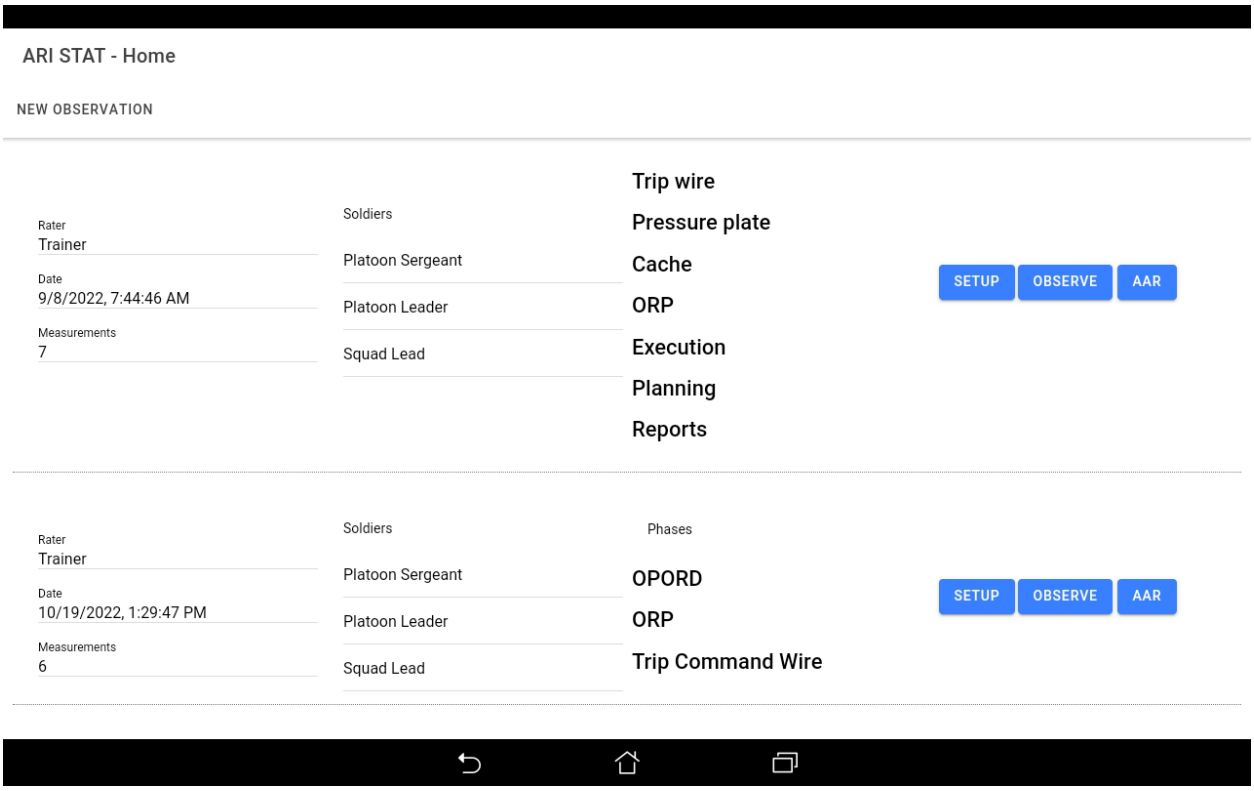
The following questions can be raised and discussed during debriefing:

Question Target	Question
Perceived Utility and Usability	1. What were your impressions of the layout of the tool?
	2. How does using the tool compare to other assessment tools you may use for training exercises?
	3. Are there any aspects/features of the tool you would not want to use?
	4. What features of the tool did you find most helpful?
	5. Would this tool help you learn to be a better observer? A better assessor? A better trainer? If so, how?
Suggested Improvements/ Enhancements	6. What challenges with assessing threat detection do you experience that are not addressed by the tool?
	7. How could the tool be changed to improve ease of use?
	8. How could the tool be modified to help you better assess Soldiers?
	9. How else can we improve the overall tool and content?
Use	10. Would you use the tool? In what situations/training exercises?
	11. Would you recommend the tool to others?

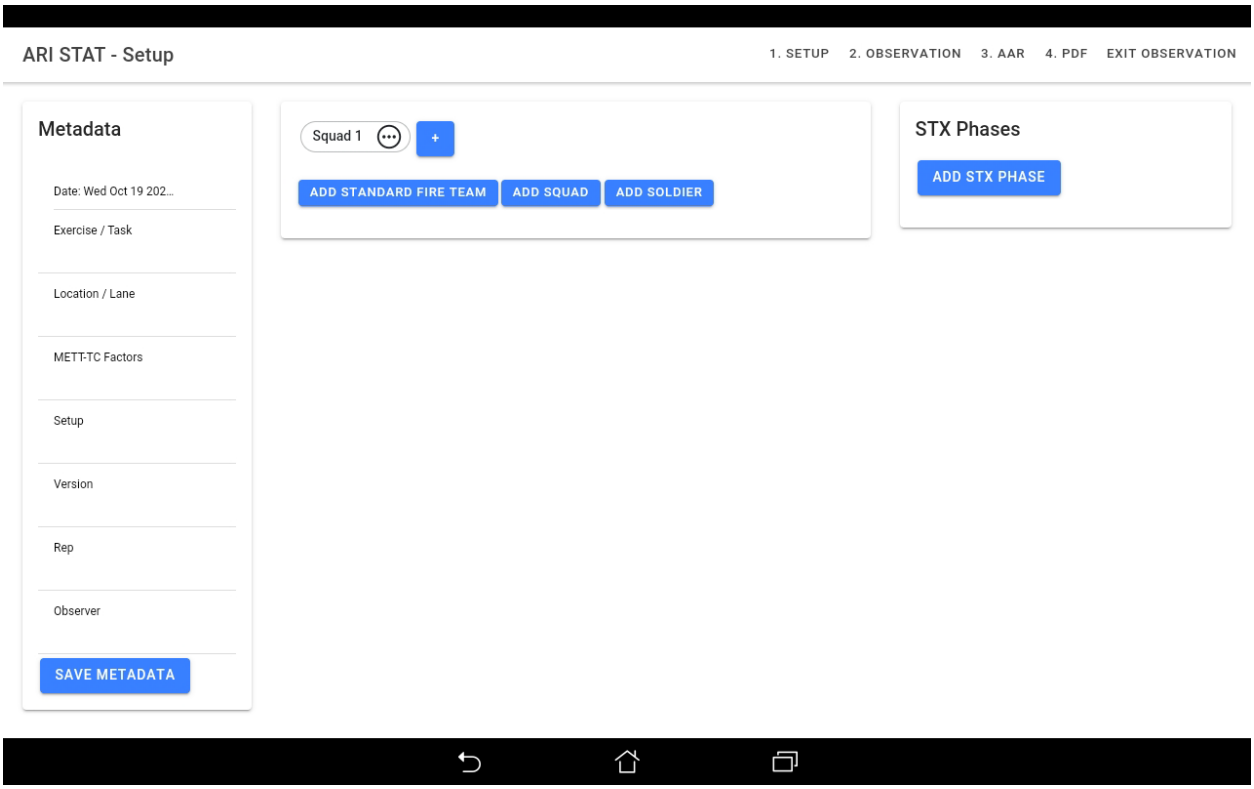
Appendix C

Screen captures of the STAT

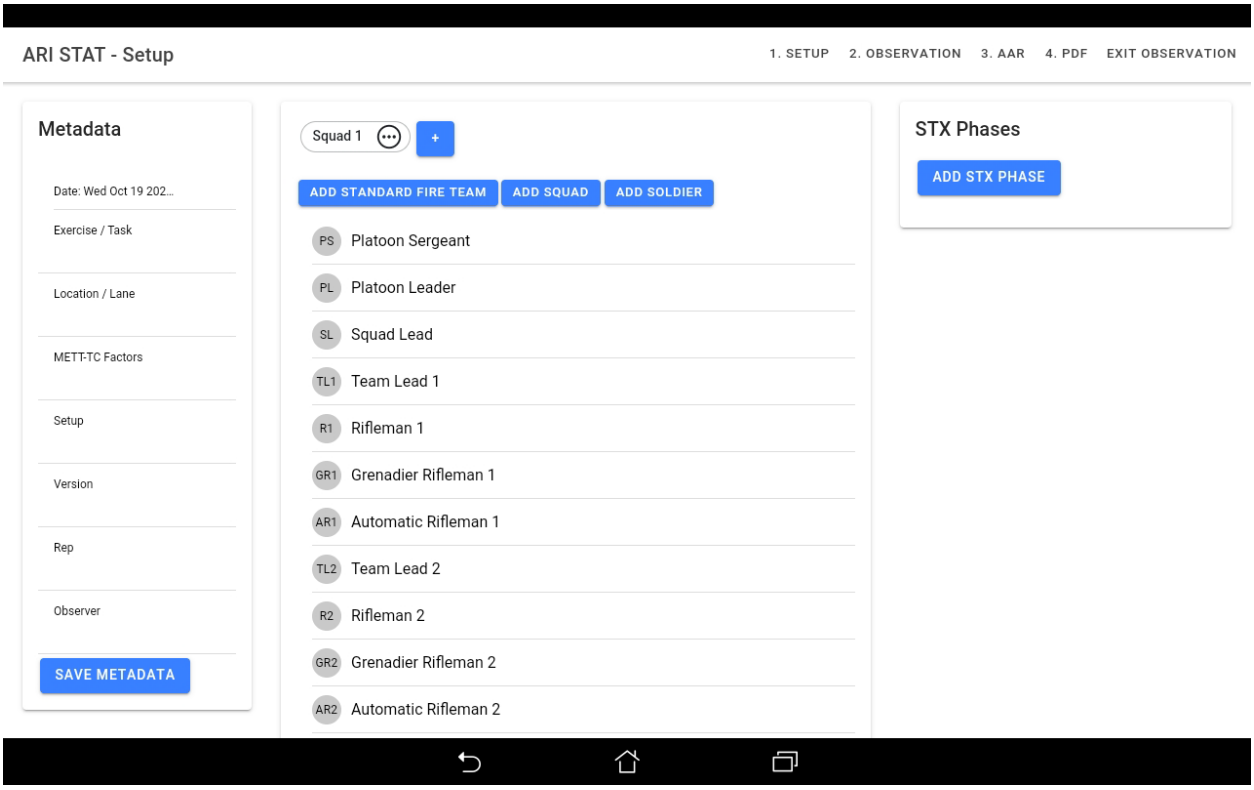
1. Should these bullets have punctuation? Example of the STAT home page. Here, users can (1) review and edit data from previous observations, by tapping on "SETUP," "OBSERVE," or "AAR" in an existing observation block, or (2) begin a new observation of an exercise by selecting "NEW OBSERVATION."



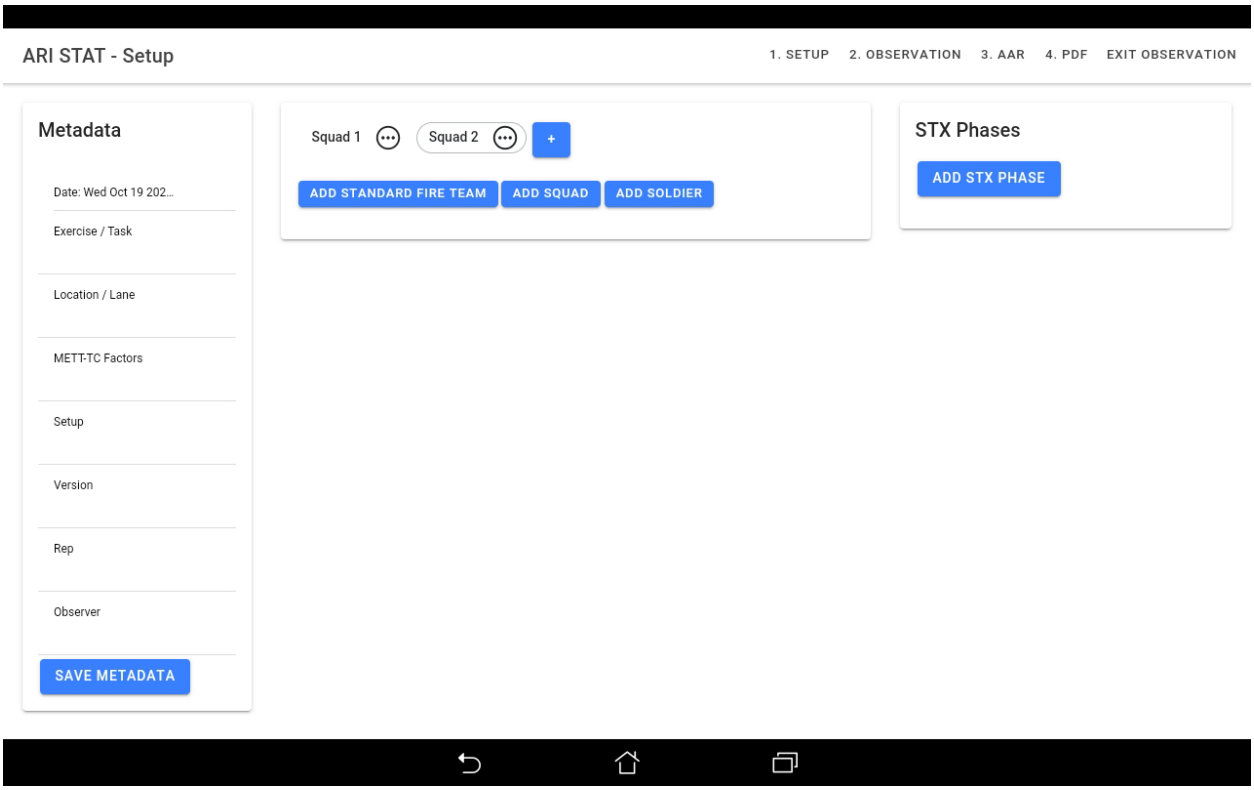
2. Example of a Setup page. Here, users can input exercise metadata (left column on screen), Soldier metadata (center column), and STX phase metadata (right column). All sections can be selected and edited by tapping on the label (e.g., “Exercise / Task” in Metadata or “Squad 1”). This page is used to enter all the data required to describe the context of the exercise to be observed.



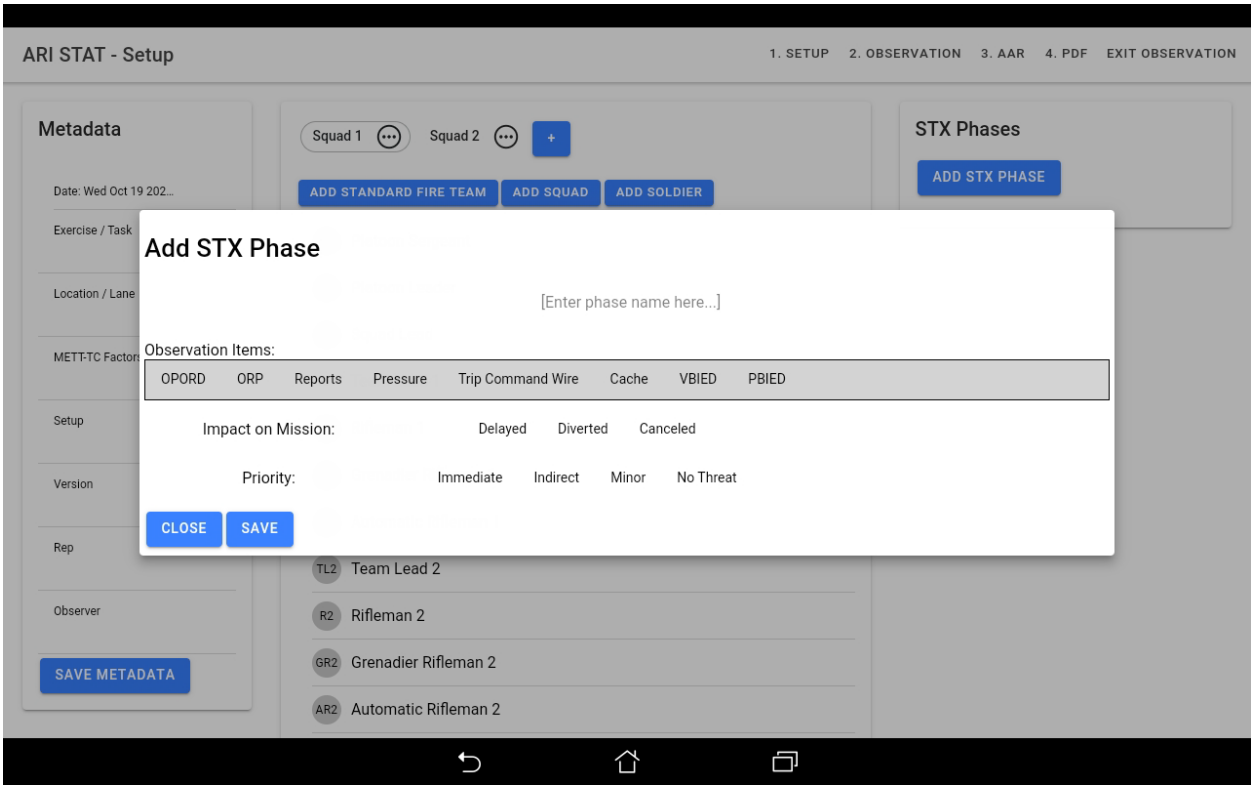
3. Example of Squad metadata. Users can tap on “ADD SQUAD” to populate a squad comprising a Platoon Sergeant, Platoon Leader, Squad Leader, and two complete fire teams. Users can also add induvial fire teams by selecting ADD STANDARD FIRE TEAM or add individual Soldiers by selecting ADD SOLDIER. These squads, fire teams, and Soldiers will be added to the enumerated squad selected at the top of the section (e.g., “Squad 1”). Users can also add and label new squads by selecting the “+” button next to the enumerated squad labels.



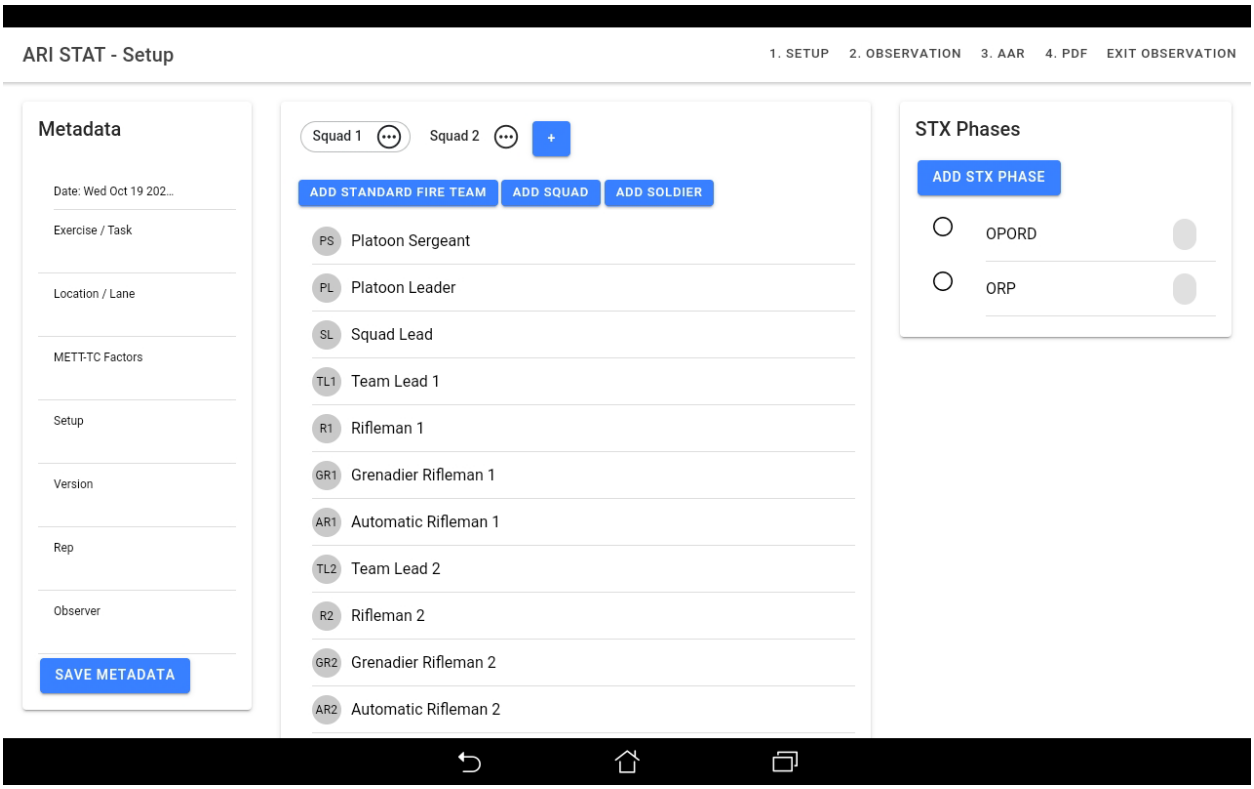
4. Example of a second enumerated squad added to the exercise.



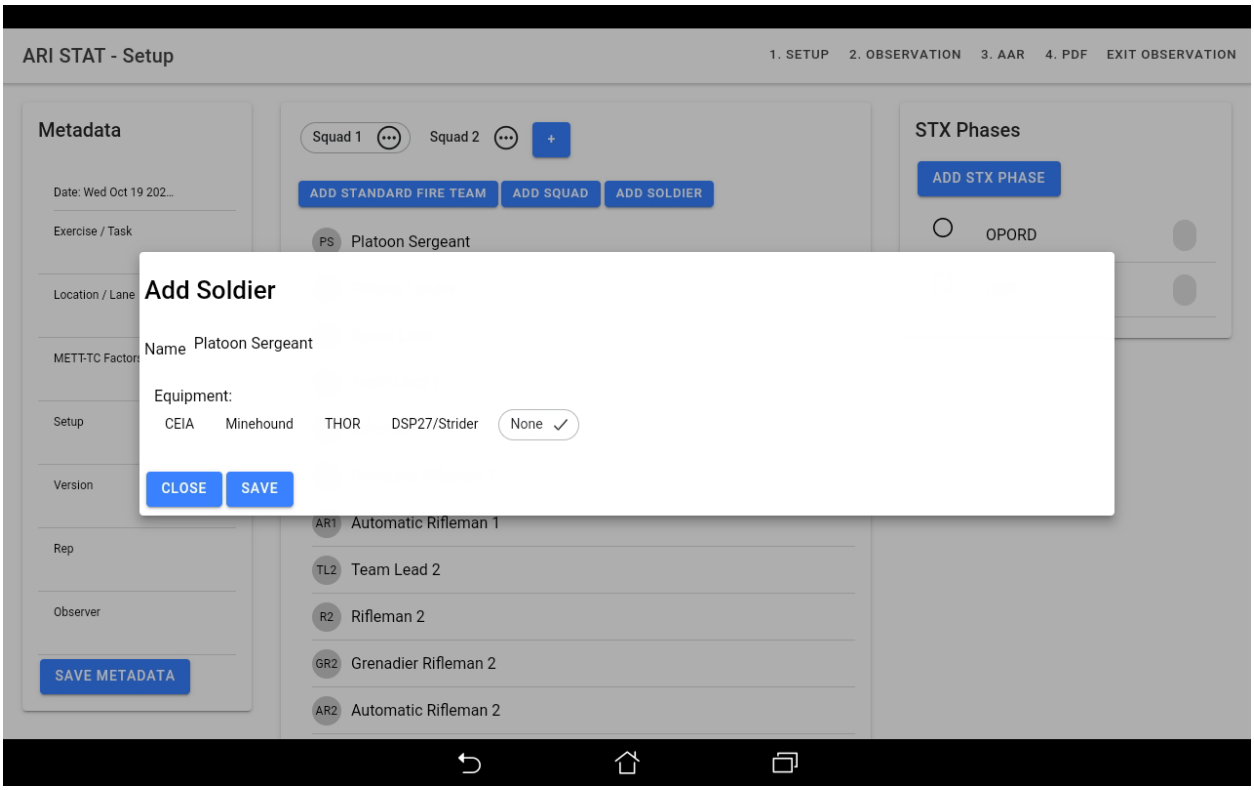
5. STX Phases. Users can add phases and components of the exercise to be observed by tapping on “ADD STX PHASE.” This will generate a pop-up box in which users can select from prepopulated phases or components (e.g., “OPORD,” “ORP,” “Pressure,” and “Cache”). Users can also enter and label a unique component by tapping on “Enter phase name here…” and typing in the name of the component. Users can also select phase/component metadata pertaining to threats implanted in a training lane, for example, a threat’s impact on the mission and its priority status.



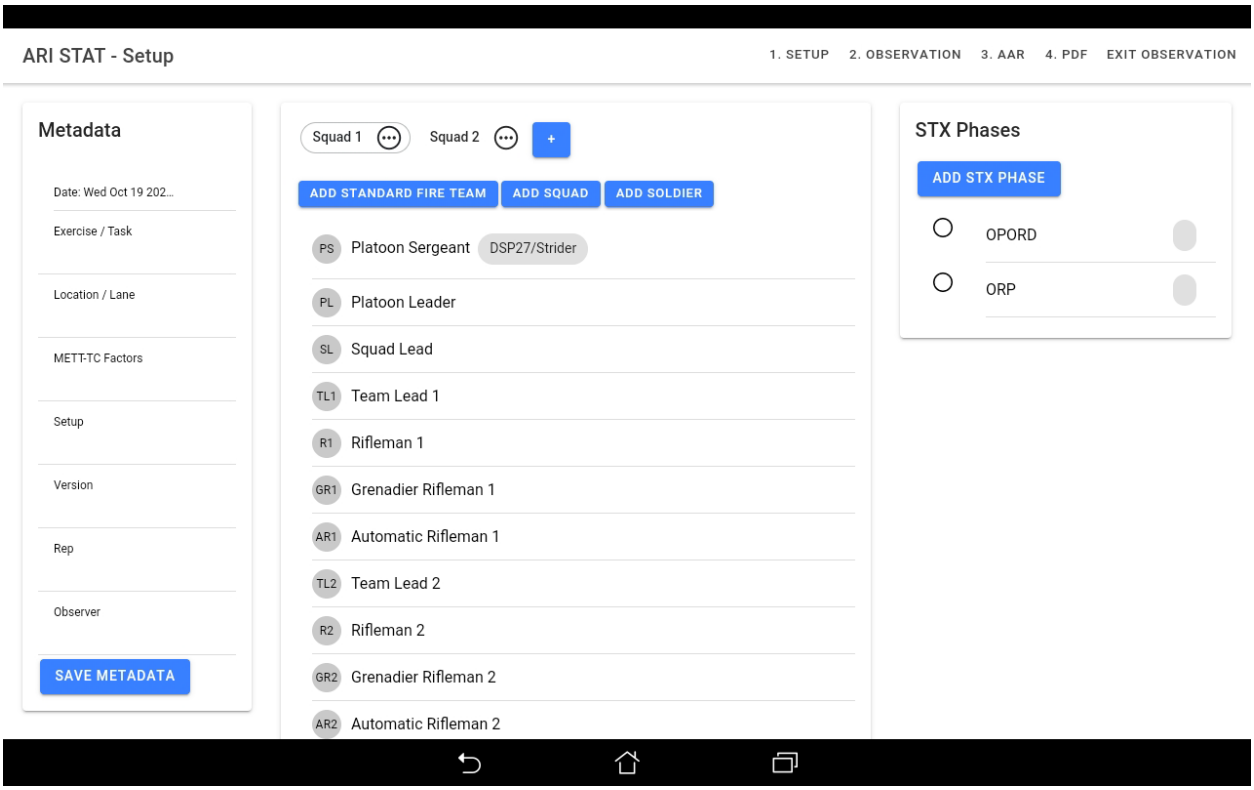
6. Example of a Setup screen with some squad and STX phase data entered. Users can edit these data by long-pressing the text of an item. For example, to edit the metadata for the Platoon Sergeant in Squad 1, the user would long press the label “Platoon Sergeant” in the center section. This generates a pop-up box in which users can edit the metadata associated with that Soldier (see also next screen capture).



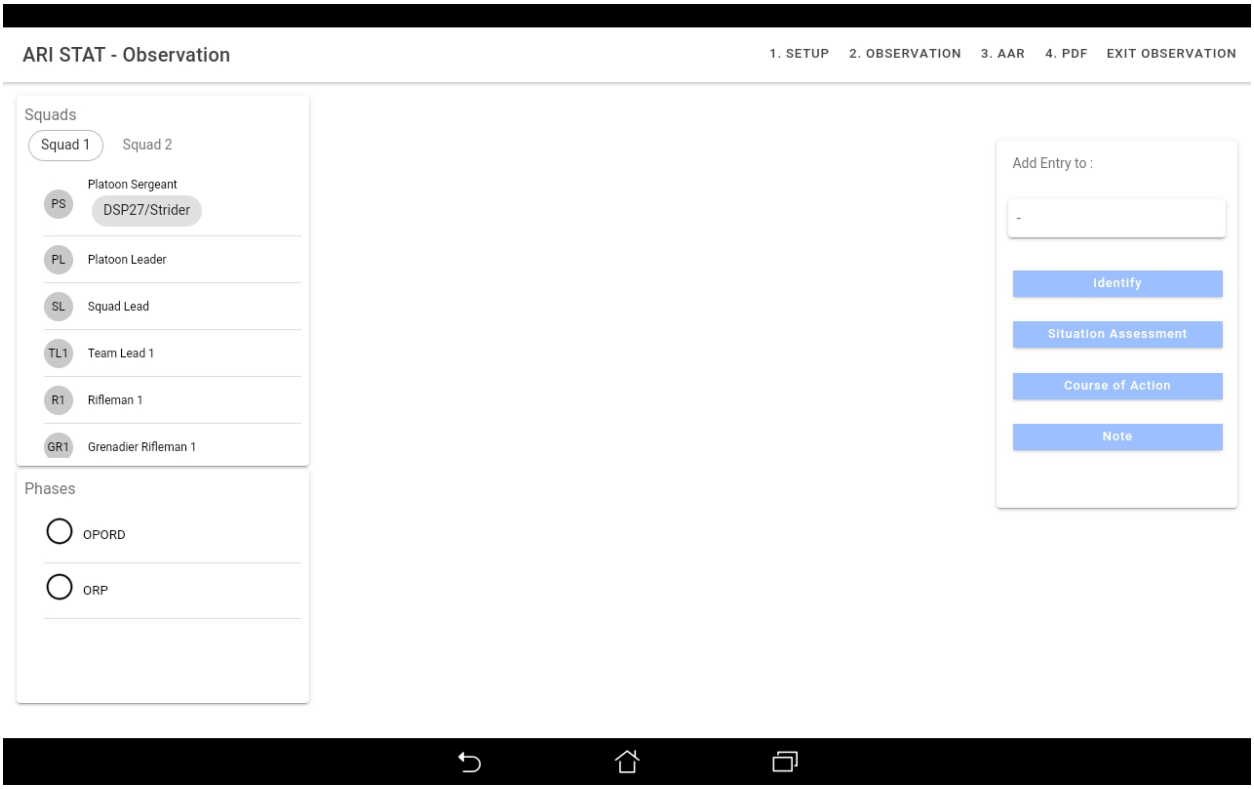
7. Example of pop-up box for editing Soldier metadata. Users can rename a Soldier or assign equipment to a Soldier using this pop-up box. To assign a piece of equipment, users would tap on the label for the desired equipment (e.g., “CEIA,” “Minehound”).



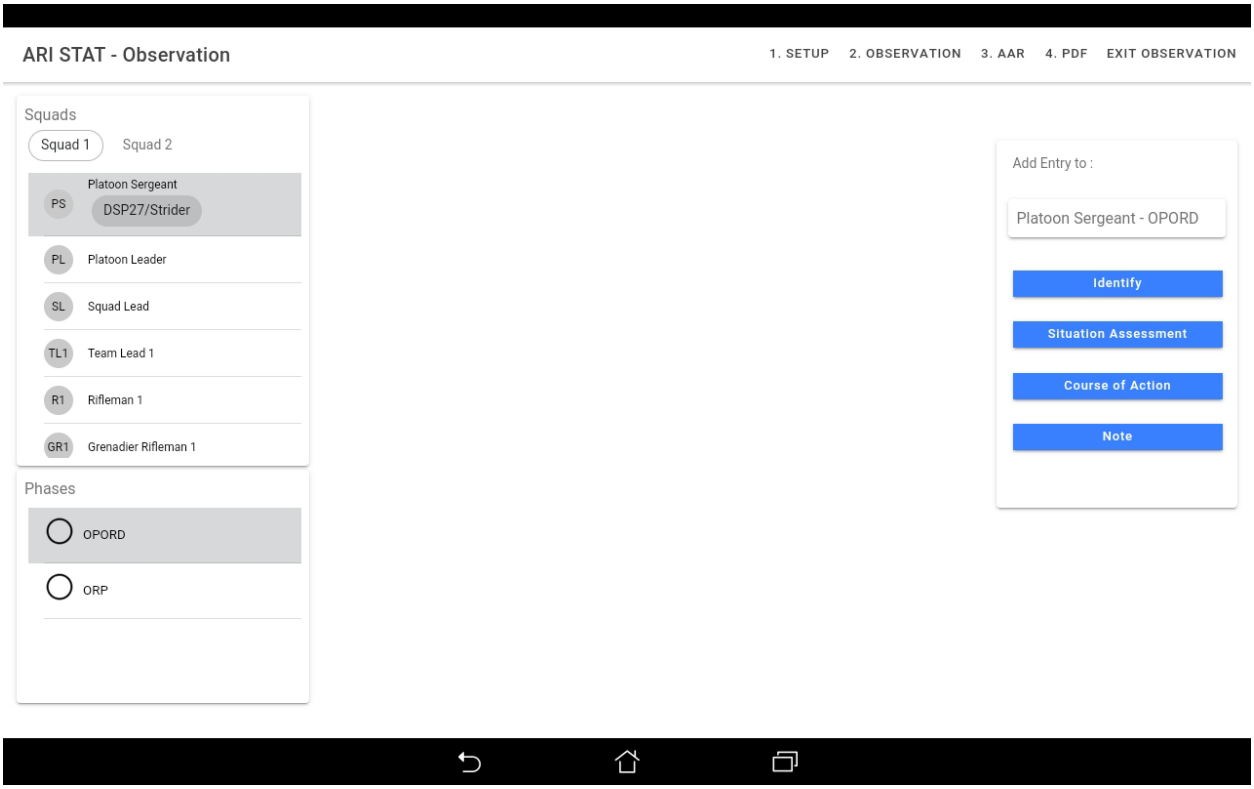
8. Example Setup screen showing equipment paired with a Soldier. After all metadata are entered, users tap on “2. OBSERVATION” in upper right portion of the screen. This advances the application to the observation section (see next screen capture).



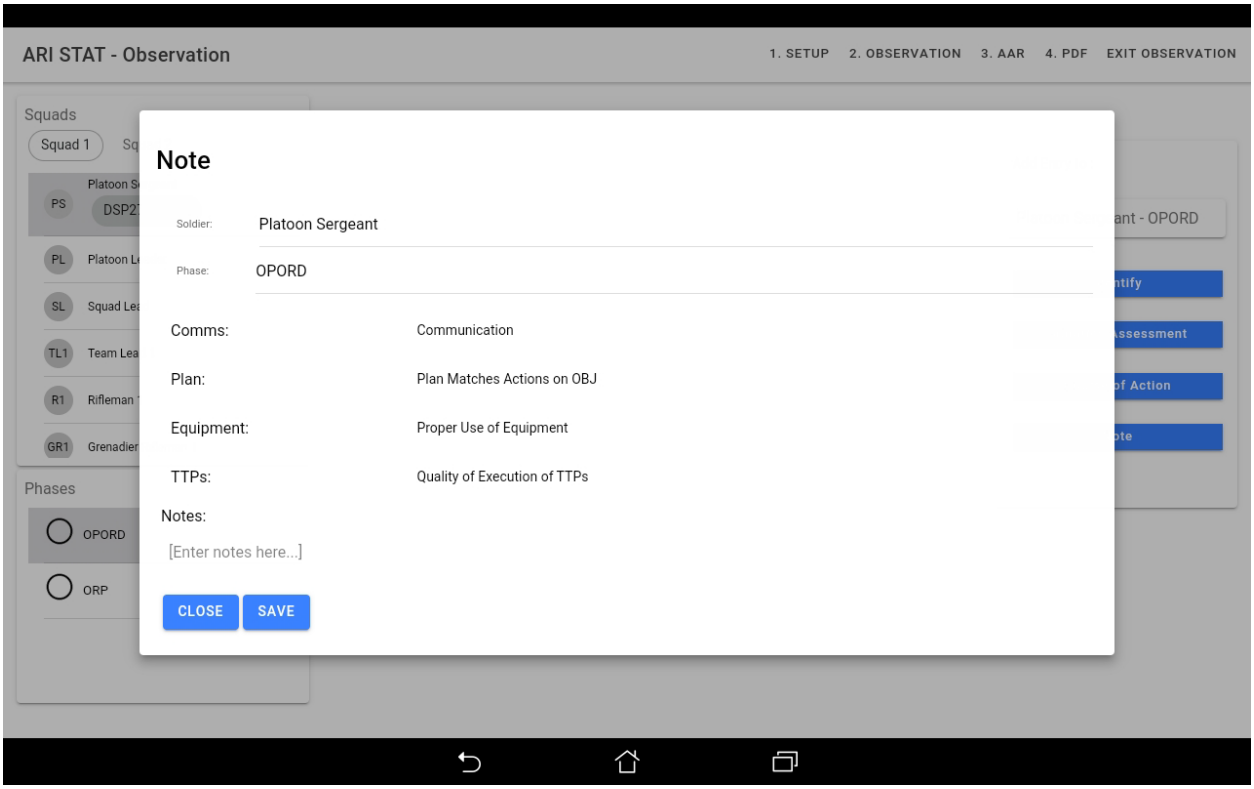
9. Example of Observation screen. Here, users select the units of observation for which they want to enter data. They select a Soldier by tapping on the desired Soldier label under the "Squads" section; they select a phase by tapping on the desired phase label under the "Phases" section. When one Soldier and one phase are selected, the blue buttons on the right of the screen become active and users can select from among them to enter observation data (see next screen capture).



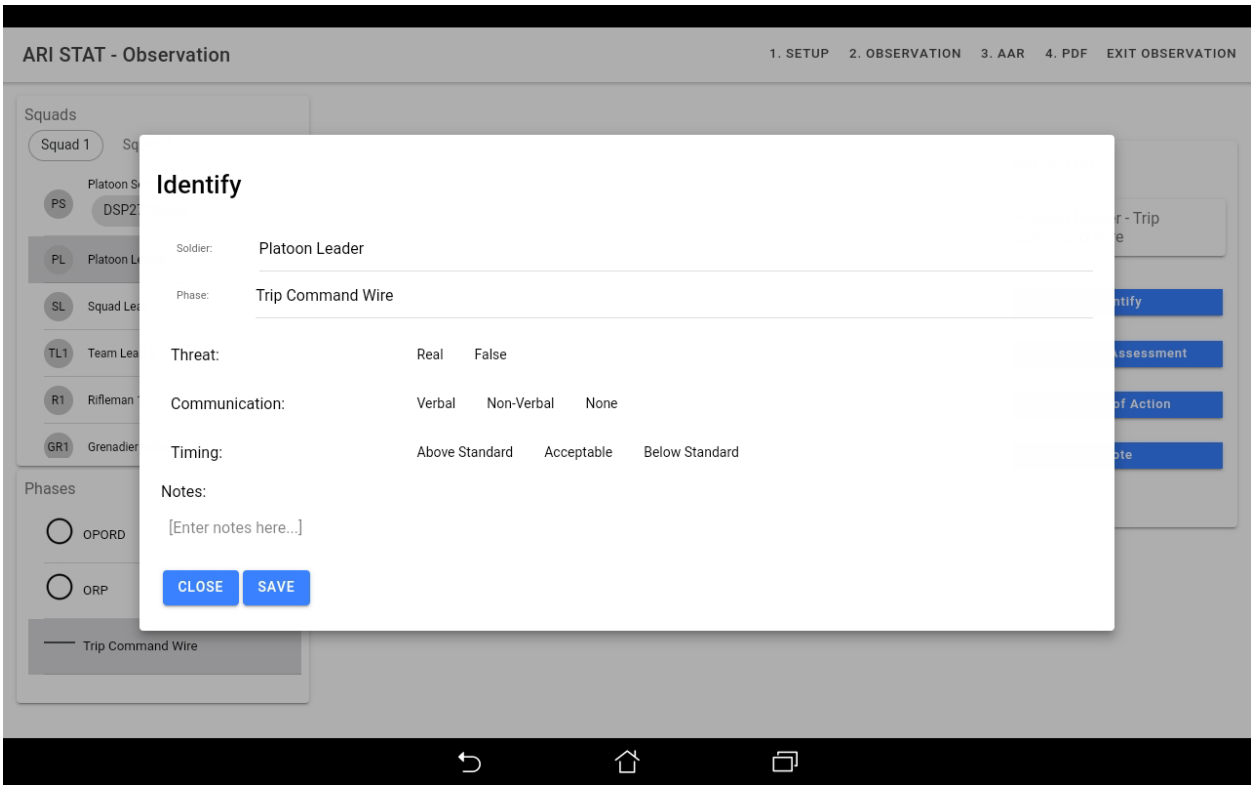
10. Example of Observation screen with active buttons for entering data. In this example, the user has selected the Platoon Sergeant and the OPORD as units of observation. These selections are indicated by grey highlights on the labels on the left of the screen and by the pairing of those labels in a box above the blue observation buttons on the right of the screen. When a user taps one of those blue observation buttons, it generates a pop-up box in which the user can enter data (see next screen capture).



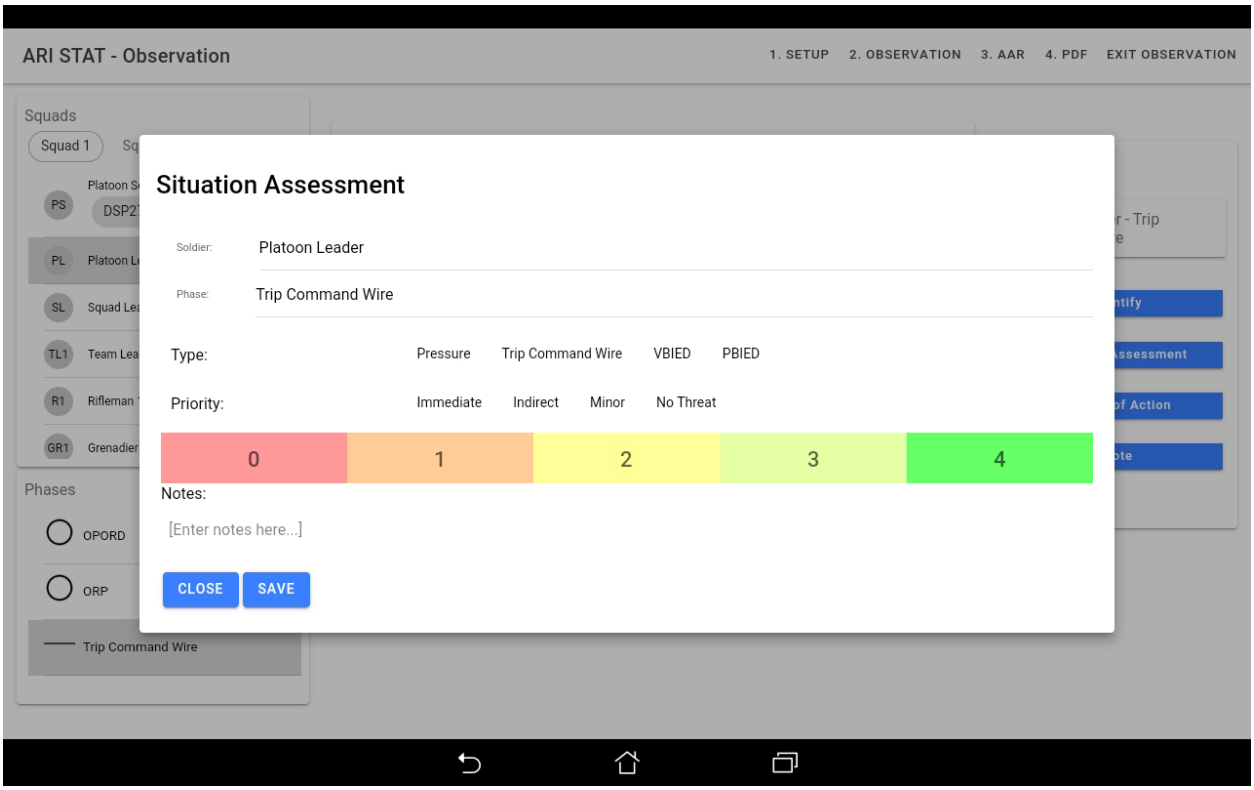
11. Example of pop-up box for entering observation data using the “Note” function from among the blue observation buttons. Here, users can tap on keywords (e.g., “Communication,” “Plan Matches Actions on OBJ”) that will appear with an observation entry. Users can enter open-narrative notes in the “Notes” section at the bottom of the screen (“Enter notes here…”).



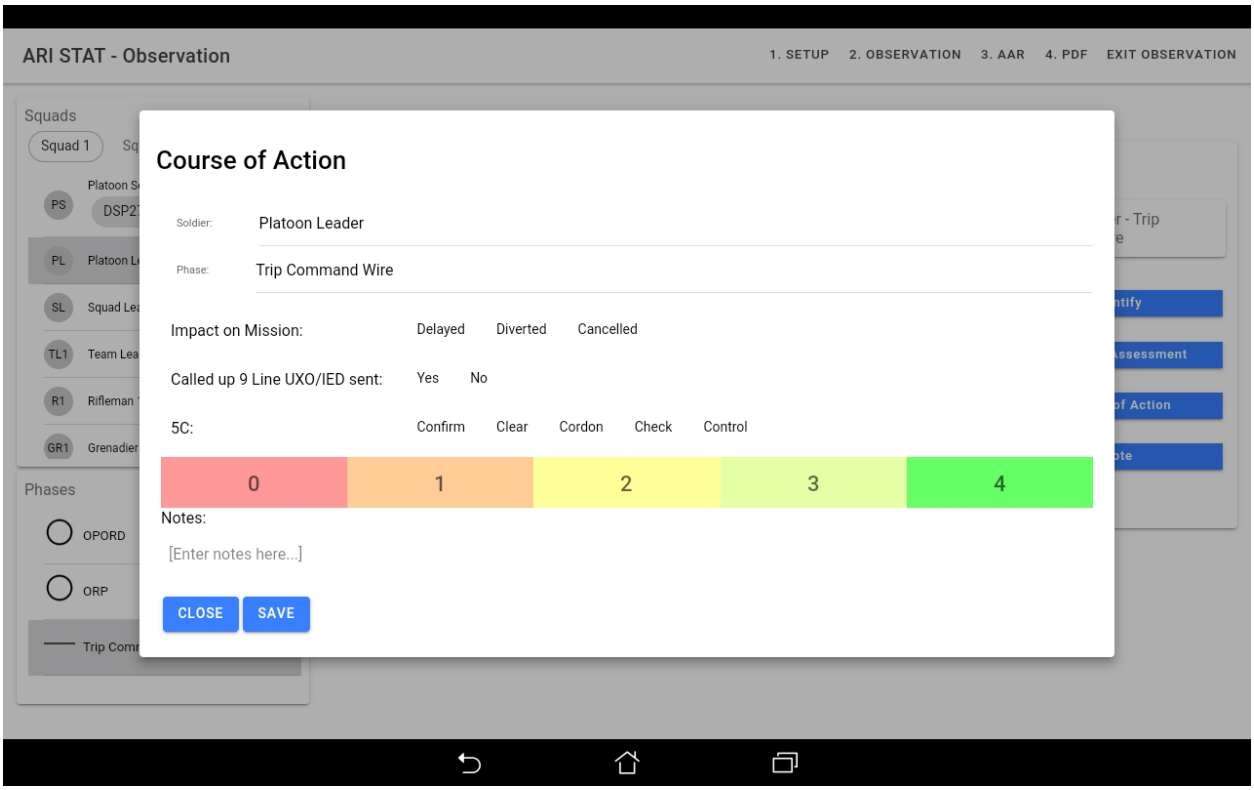
12. Example of pop-up box for entering observation data using the “Identify” function from among the blue observation buttons. Here, users can enter keywords to associate with Soldier’s identification of a threat (e.g., for “Threat” users can choose among “Real” vs. “False;” for “Communication,” users can choose among “Verbal,” vs. “Non-Verbal,” vs. “None”). Users can also add notes describing their observation.



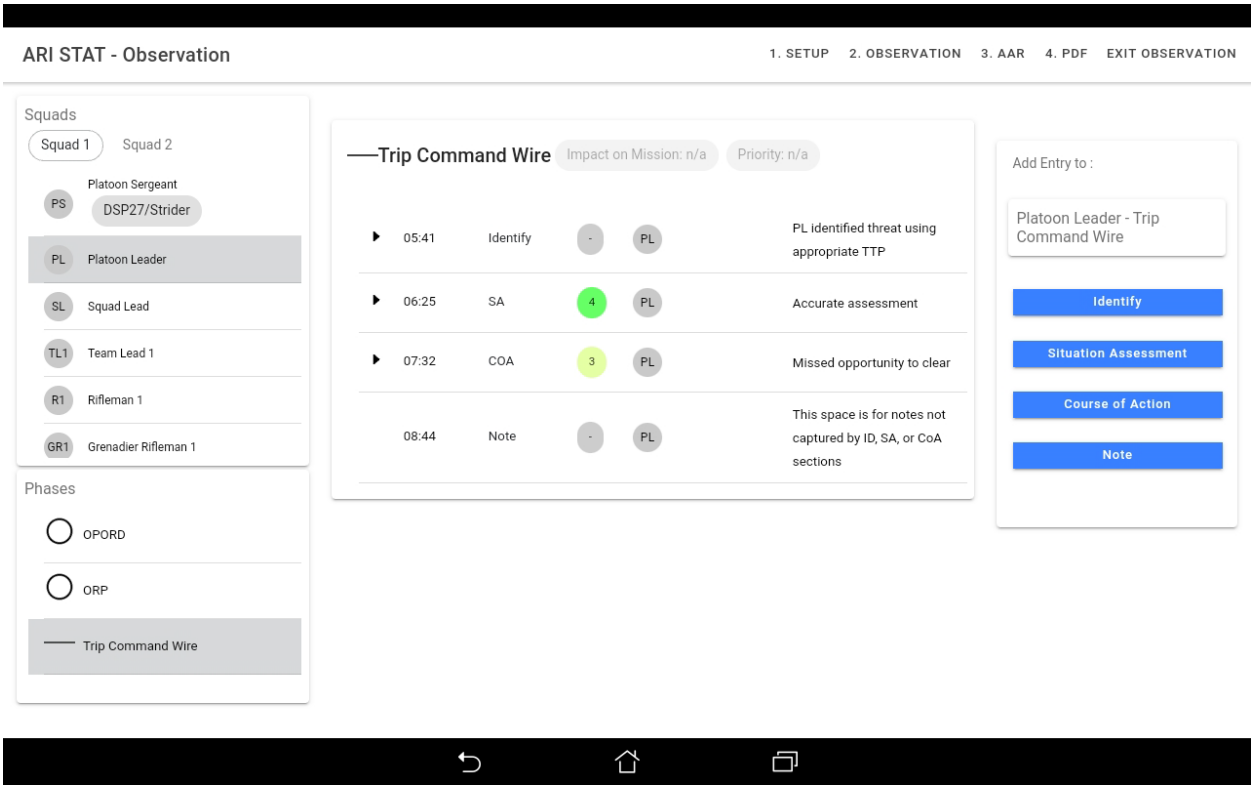
13. Example of pop-up box for entering observation data using the “Situation Assessment” function from among the blue observation buttons. Here, users can enter keywords relevant to the Soldier’s assessment of the threat (e.g., what type, what priority level). Users can also rate the quality of the assessment by tapping on a colored box of a 5-point scale. The points on the scale are not labeled, to allow flexibility in how it is used; however, the user guide (Aptima, 2020, p.14) recommends the following labels: 0 (Unsatisfactory), 1 (Minimum Standard), 2 (Standard), 3 (Above Standard), 4 (Exceeds Standard). Finally, users can enter open-narrative notes describing the Soldier’s situation assessment.



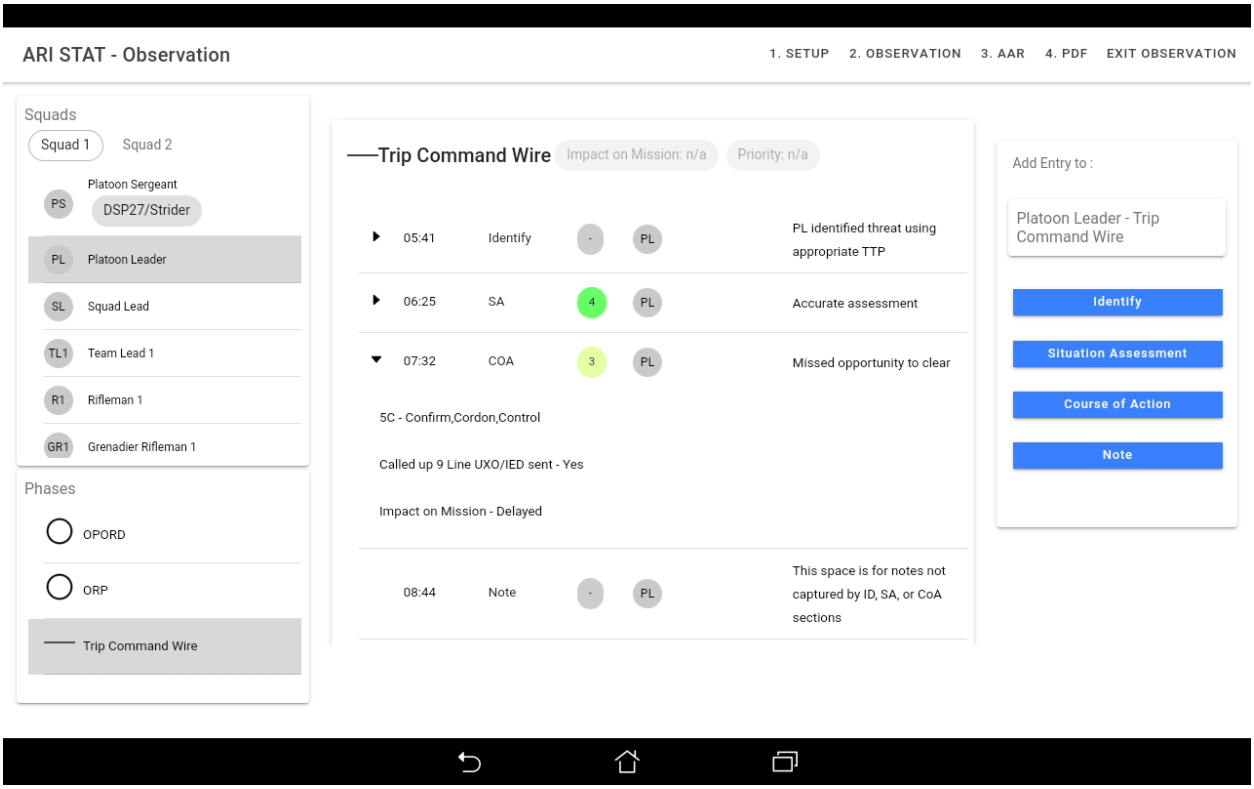
14. Example of pop-up box for entering observation data using the “Course of Action” function from among the blue observation buttons. Here, users can enter keywords relevant to the course of action identified by the Soldier, for example, whether the mission was delayed, diverted, or canceled; and whether the Soldier satisfied all 5Cs of counter-IED (confirm, clear, call, cordon, control). Users can also rate the quality of the course of action on a 5-point scale and enter open-narrative notes.



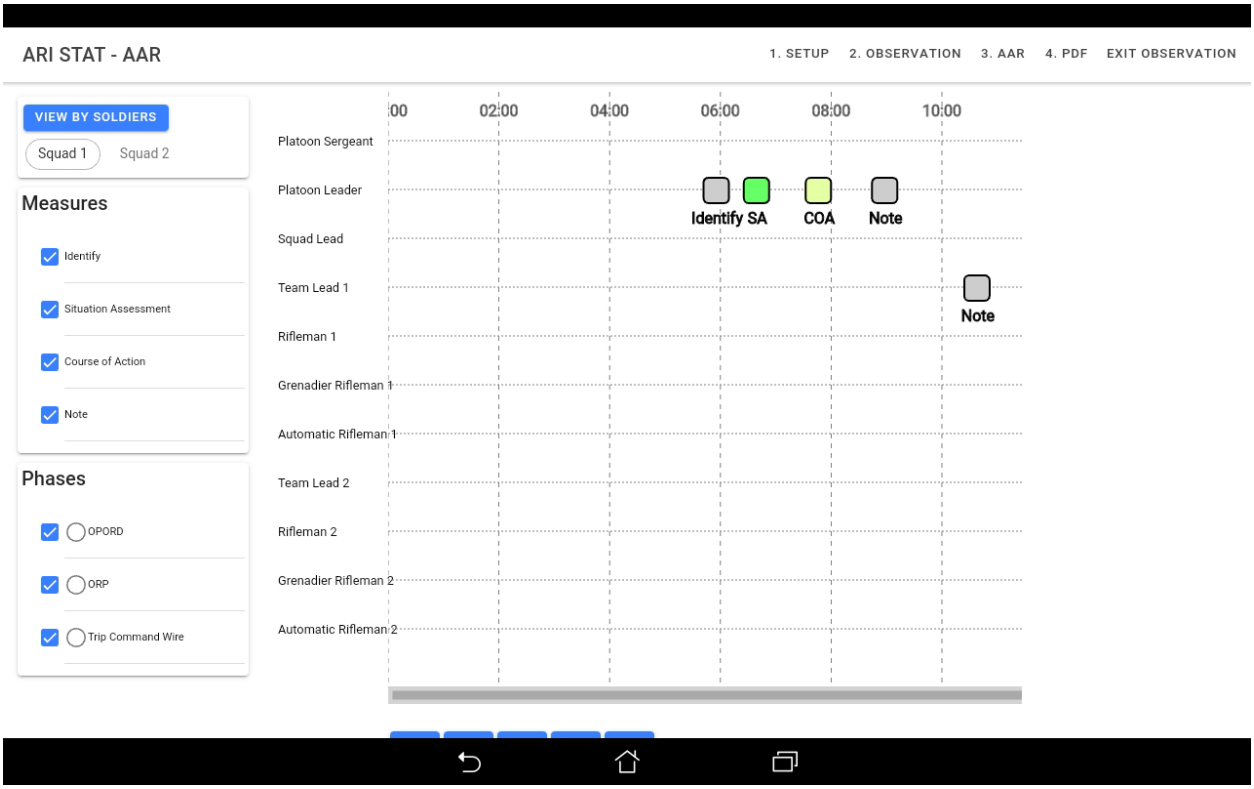
15. Example of an Observation screen with observations entered. Here users can view the observation data they entered. These data are organized by STX phase/component and by time within each component. Each entry includes the time it was entered (where time 0 = the start of the exercise), the type of observation (e.g., Identify, SA), the rating (if relevant), the Soldier identifier (PL, SL, etc.), and any open-narrative notes associated with the observation. Users can tap on the arrow (▶) next to each entry to view the keywords associated with that entry (see next screen capture).



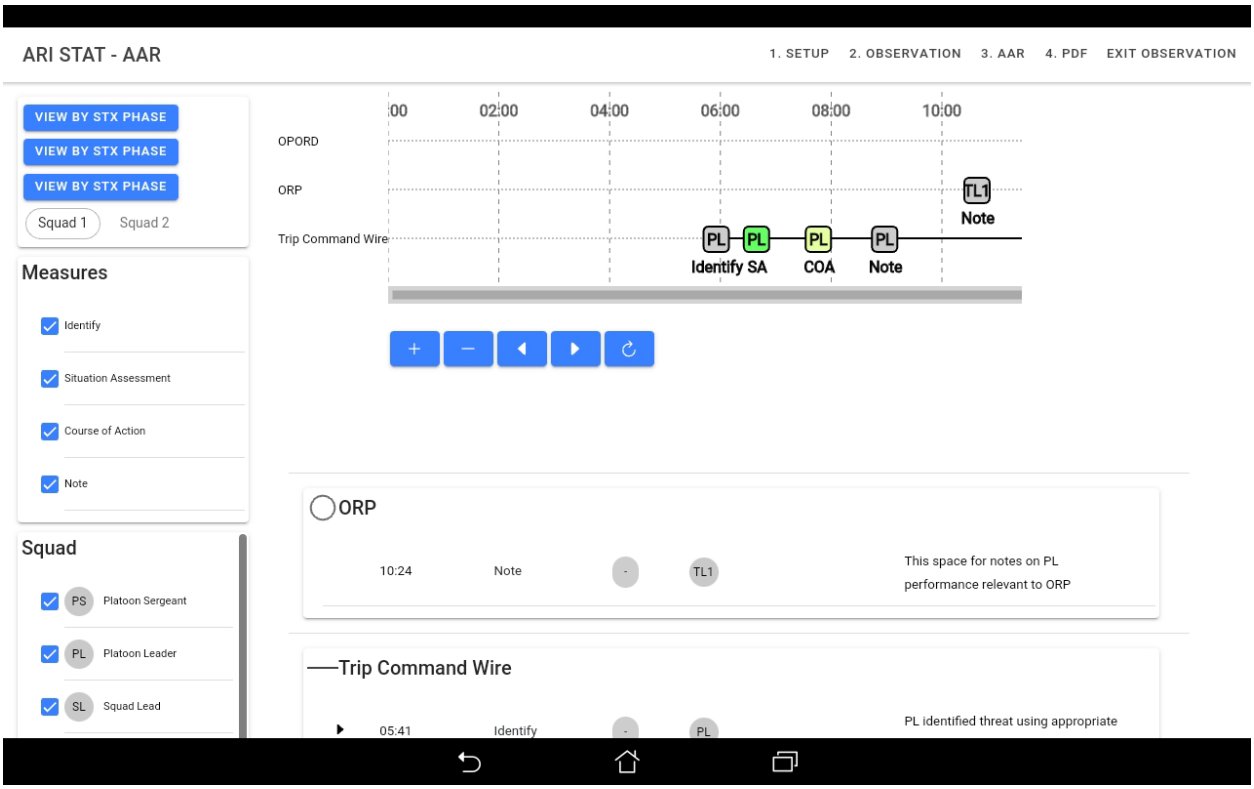
16. Example of Observation screen with observation data and one entry expanded to view associated keywords.



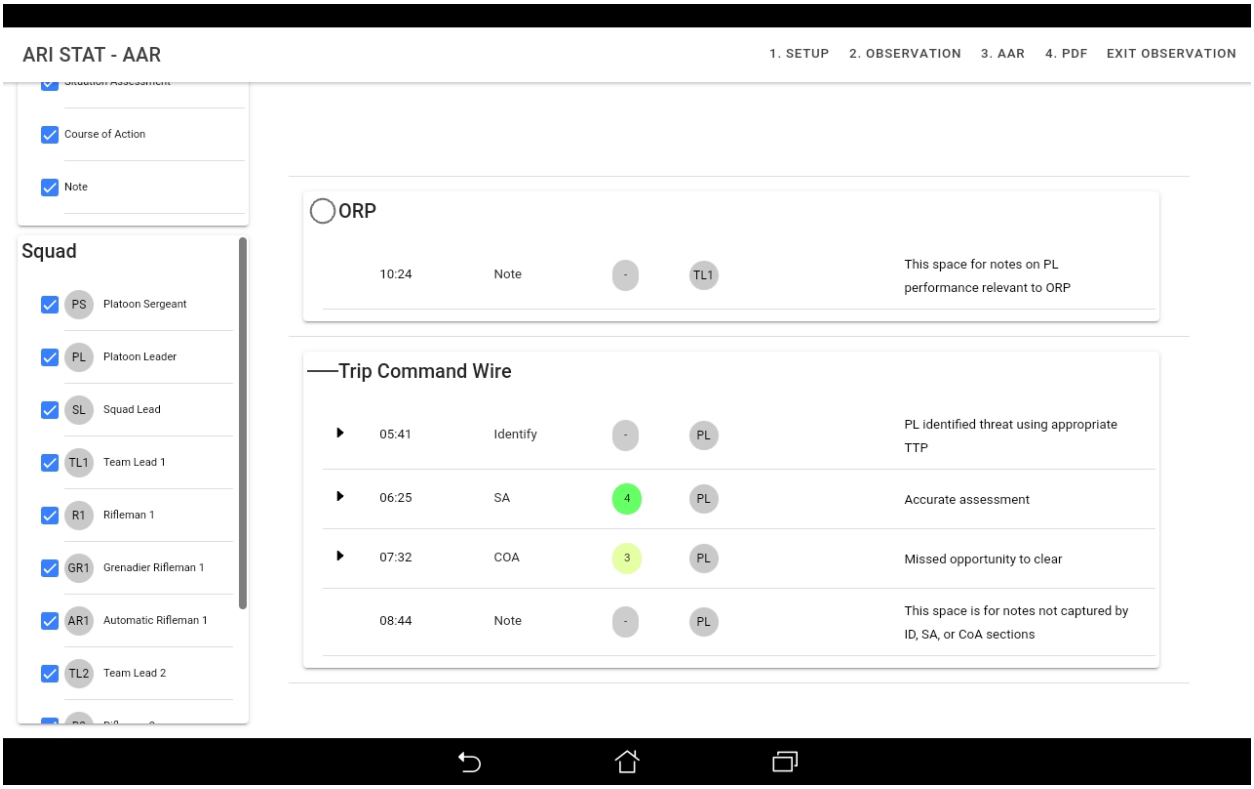
17. Example of AAR screen. After users have completed entering data, they can tap on “3. AAR” at the top right of the screen to advance to the AAR section of the application. Here, users can access all the data they entered and organize it in different ways. This section presents a timeline that presents an icon representing each data entry. It also presents check-mark (radio-style) buttons for viewing and hiding the data associated with each Soldier and Phase. The blue button at the top left of the screen (“VIEW BY SOLDIER” in this example) indicates the current organization of the data. Users can tap this button to change the organization (e.g., from “VIEW BY SOLDIER” to “VIEW BY PHASE”). This will change the way data are presented on the timeline (see next screen capture). Note that the data are presented by squad. Users can tap on the Squad labels at the top left of the screen to select data for each observed squad. Note also that this screen is scrollable. Users can drag the screen upward to reveal additional data frames (i.e., data presented in text mode vs. timeline mode).



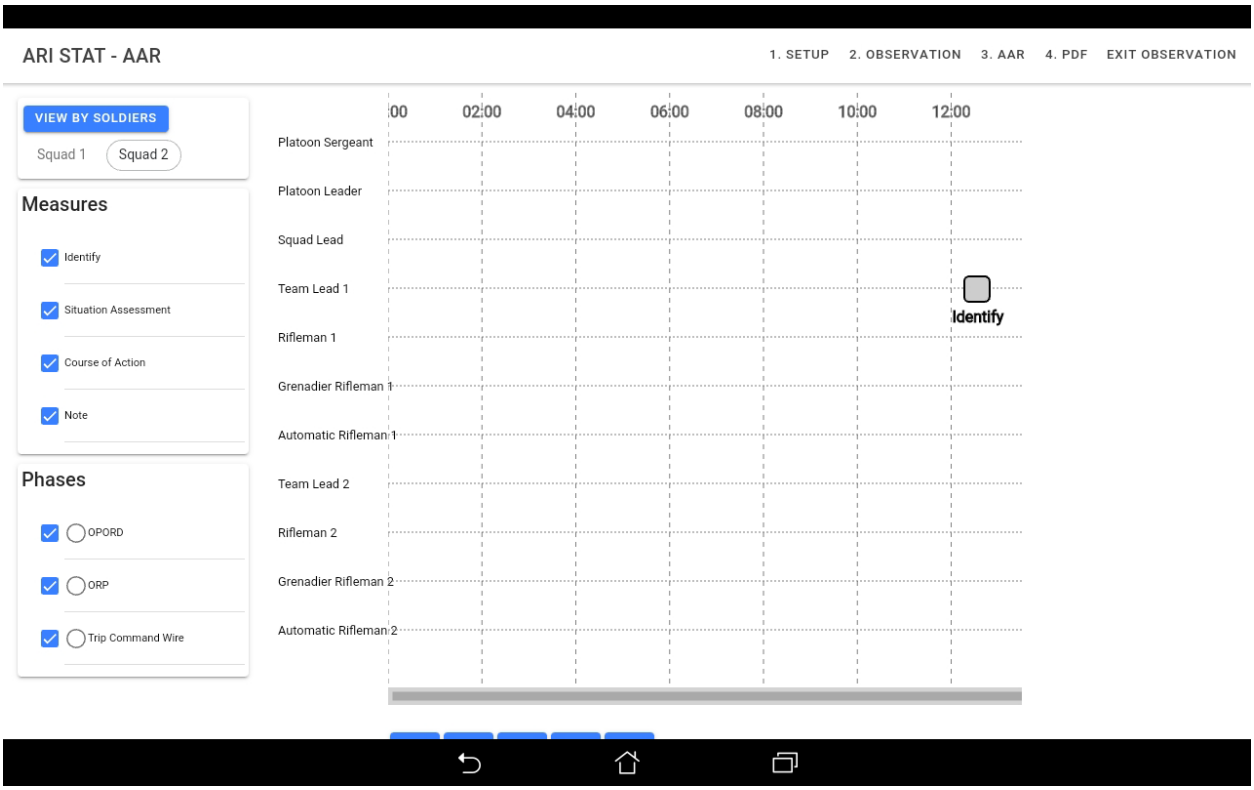
18. Example of AAR screen with data organized by Phase (vs. by Soldier). In this example, blue buttons ("+," "-," and various arrows) are visible below the timeline. These buttons allow a user to navigate across the timeline if it extends outside the limits of the viewable window. Note also, that this example shows some of the data presented in text mode below the timeline.



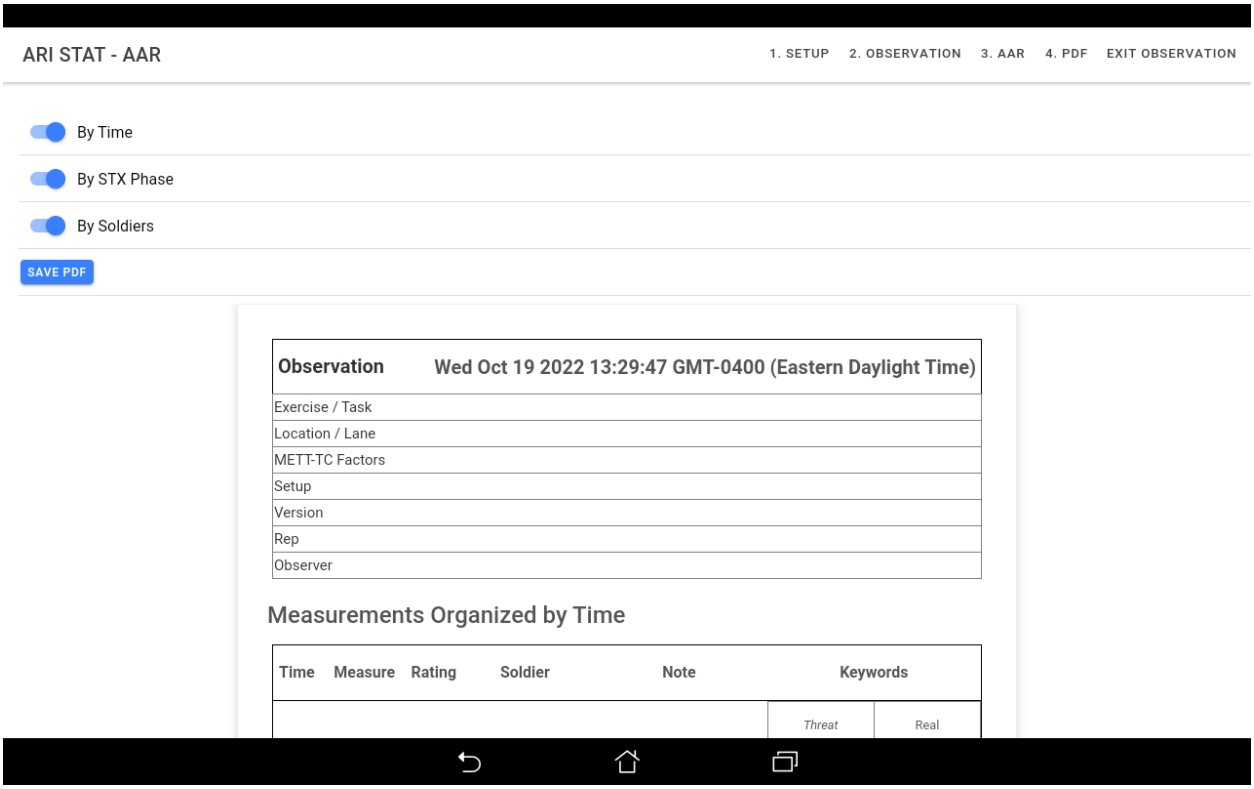
19. Example of observation data in text mode. Here, users can view data organized by threat or by Soldier. These data entries are expandable to allow users to view associated keywords.



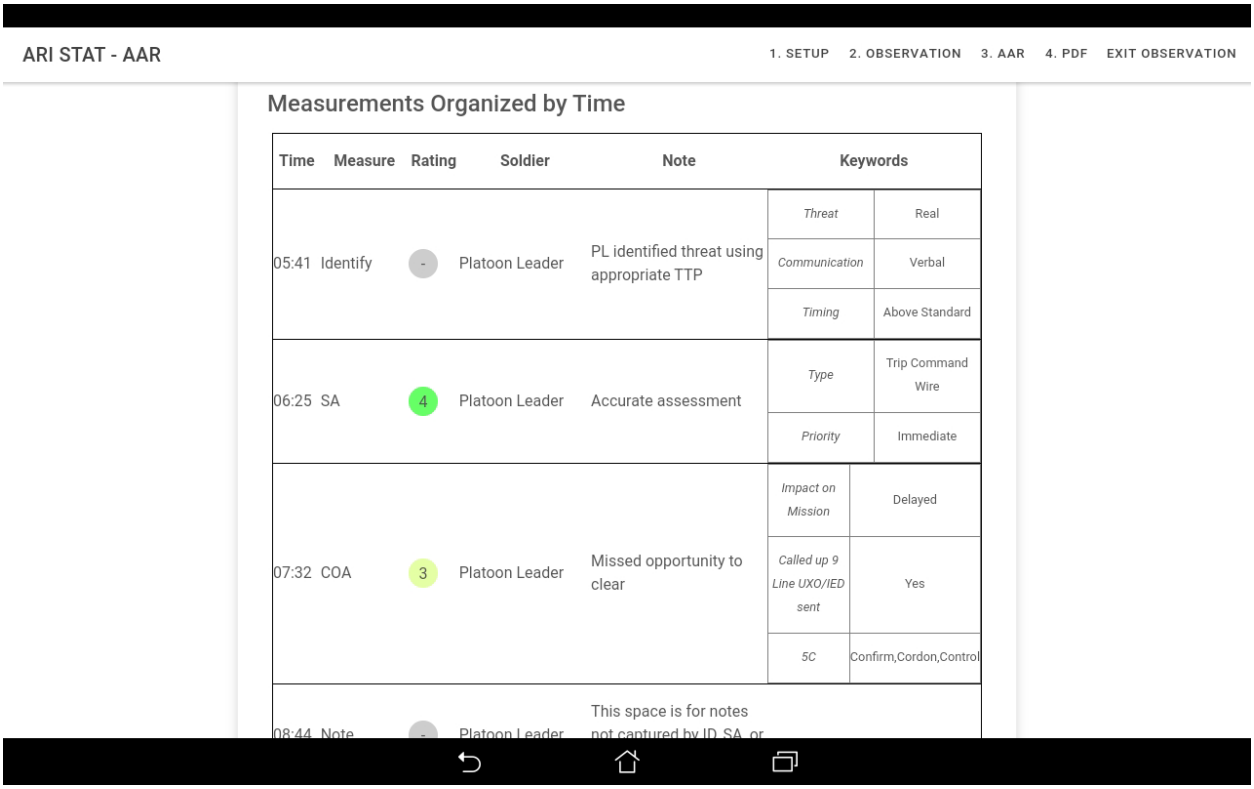
20. Example of AAR screen presenting data for Squad 2.



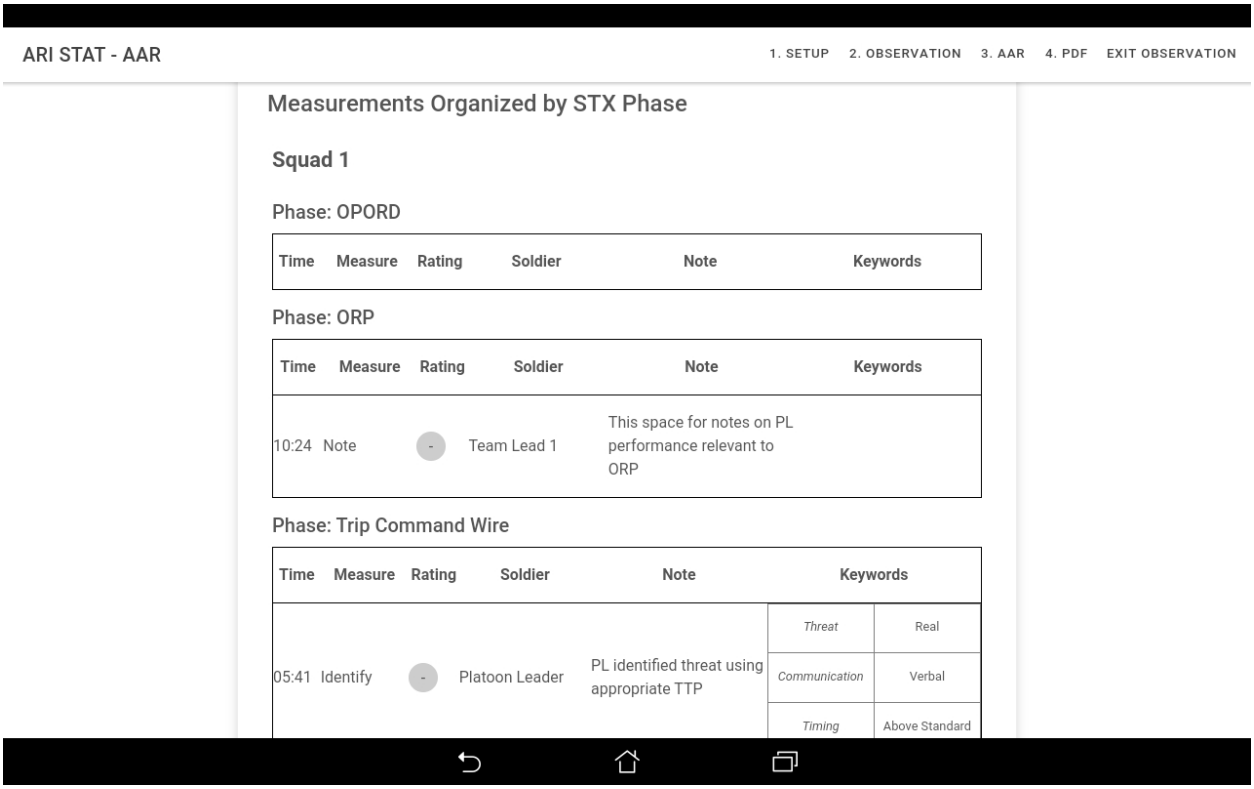
21. Example of PDF view of observation data. Users can view and export data in PDF form by tapping on “4. PDF” at the top right of the screen. This enables users to view only the text of their data (not the timeline). The PDF view allows users to organize and view data by time, phase, and Soldier, by selecting or deselecting the blue toggle switches at the top left of the screen. Regardless of how the data are organized, the PDF view always displays the exercise metadata at the top of the page.



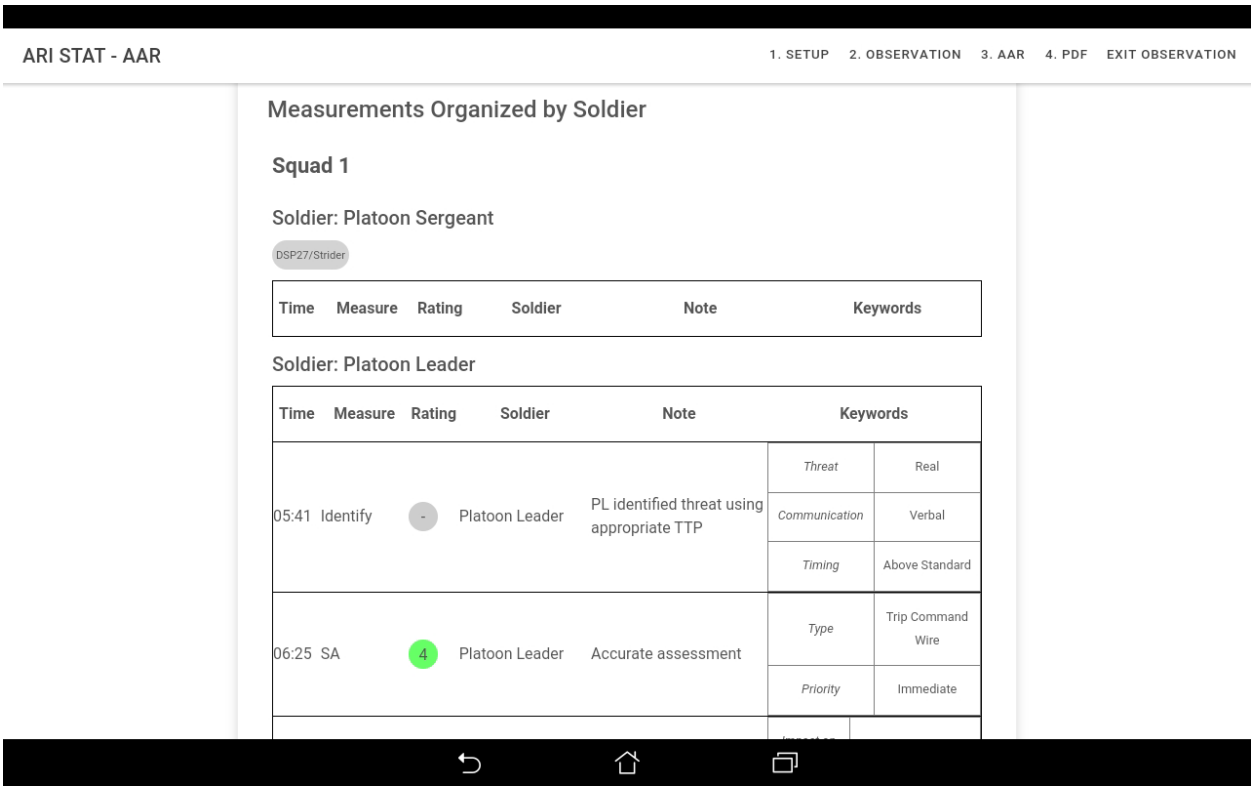
22. Example of data in PDF view, organized by time.



23. Example of data in PDF view, organized by Phase.



24. Example of data in PDF view, organized by Soldier



Appendix D

Quantitative data

The following table presents the quantitative data provided by instructors. Ratings were made using a 5-point scale (1 = strongly disagree; 5 = strongly agree).

Table D-1

Quantitative Data

Items general to STAT utility and usability	**Mean Rating**	**SD**
I knew exactly where to go in the tool to make an assessment	4	0
It was easy to tell if I had completed an assessment	3	1
At times, I lost track of where I was in the tool	4	0
I sometimes found it hard to remember how to use the tool	2.5	0.5
It was easy to navigate to different performance areas within the tool	4	0
I was able to make assessments quickly, even when using the tool for the first time	3.5	0.5
The layout of the tool allowed me to capture data effectively	3.5	0.5
The layout of the tool was well organized	4	0
There was too much content in the tool	2	0
The tool was easy to use	3.5	0.5
I would need help/support to use the tool	3	0
In general, I would find the tool helpful when making an assessment	4	0
In general, I would find the tool helpful when providing feedback	4.5	0.5
In general, I would find the tool helpful for improving Soldier performance	4.5	0.5
The information presented in the instructions for using the tool helped me make accurate assessments	4	0
The information presented in the instructions for using the tool helped me generate actionable feedback	4.5	0.5
When I made a mistake, it was easy to undo my error	3	0
Items specific to implementing the STAT in an STX	**Mean**	**SD**
This training observation was mentally demanding	2	0
This training observation was physically demanding	2	0
I did not have enough time to observe and assess student performance	2.5	0.5
I was successful in accomplishing what I was asked to do	4	0
I had to work hard to observe and assess student performance	2	0
I was effective in **observing** Soldiers' identification of threat-relevant or irrelevant cues	4	0
I was effective in **observing** Soldiers' accuracy in interpreting threat-relevant or irrelevant cues	4	0

I was effective in **observing** Soldiers' proficiency in communicating threat-relevant or irrelevant assessments to their teams	4	0
I was effective in **assessing** Soldiers' identification of threat-relevant or irrelevant cues	4	0
I was effective in **assessing** Soldiers' accuracy in interpreting threat-relevant or irrelevant cues	4	0
I was effective in **assessing** Soldiers' proficiency in communicating threat-relevant or irrelevant assessments to their teams	4	0
I was effective in conveying relevant feedback to Soldiers	4	0
My assessment was helpful for improving Soldier performance	4	0
The feedback that I provided is actionable	4.5	0.5
The feedback that I provided is useful for Soldiers	4.5	0.5
Soldiers were easily able to understand this feedback	4.5	0.5
This feedback will help Soldiers improve their performance	4.5	0.5

Appendix E

Additional feedback provided by instructors

Other instructor comments included the following observations:

- Making tabs (i.e., setting up observation phases) in advance let them attend better to specific elements (e.g., ORP). It forced them to attend to certain events and provide feedback on them.
- The STAT made it possible to call out students by name and provide feedback on their specific actions. This would not have been possible without the tool.
- Identifying threats, Squad tabs (separate labels for separate squads), and notes were most useful.
- The ability to timestamp and capture detailed notes are improvements over pen-and-paper notebooks.
- The STAT enables instructors to keep track of "leadership looks" (who has participated and who has not) and ensure that those who need those opportunities receive them.
- They would use the STAT again if it is made more user friendly.
- Would love to be able to use the tool early in the course.
- They would also use the STAT for internal AARs, to report to leadership.
- However, the number of tabs and features seemed overwhelming. They need only three sections/buttons to capture the following:

    1. What was the threat?
    2. Who saw it?
    3. What did they do?

Instructors requested or recommended the following:

- In the Set-up section:
    - Develop a library of missions to auto-populate the set-up page, to make the set-up process more efficient.
    - Ability to set up squads by name.
    - Use a tap function instead of slide function to open editing options for any given field on the set-up screen (e.g., Soldiers, Threats).
    - Ability to tailor the composition of units in a way that is easier than using the existing "add Soldier" function. For example, select a Soldier, select equipment, then add the pair to a squad.
    - Ability to delete Soldiers from squads in batches.
    - Add other items to "Threat" category in set up (e.g., planning, or other objectives).
- In the Observation section
    - Ability to take and integrate photos. This would help with storyboarding.
    - An autosave feature for notes, so if a user accidentally taps out of a screen, they do not lose their input.

    - Ability to identify and label different stages of training. For example, add or replace a metadata factor to include an item that identifies whether the observation is for planning, execution, or reports.
- In the AAR section:
    - Move the text (e.g., notes) of the observations to the top of the screen and put the timeline below, as the text-format data are more useful than the timeline.
    - Ability to organize by time and to tap and open a high-level category (e.g., OPORD or ORP).
    - A screen for viewing all notes (not just separated by squad).