Research Report 1972

Validation and Evaluation of Army Aviation Collective Performance Measures

Martin L. Bink
U.S. Army Research Institute

Courtney Dean and Jeanine Ayers
Aptima Incorporated

Troy Zeidman
Imprimis Incorporated

January 2014

United States Army Research Institute for the Behavioral and Social Sciences

U.S. Army Research Institute
for the Behavioral and Social Sciences

Department of the Army
Deputy Chief of Staff, G1

Authorized and approved for distribution:

MICHELLE SAMS, Ph.D.
Director

Research accomplished under contract
for the Department of the Army by

Aptima Incorporated

Technical review by

M. Glenn Cobb, U.S. Army Research Institute
Randall Spain, U.S. Army Research Institute

NOTICES

DISTRIBUTION: Primary distribution of this Research Report has been made by ARI. Address all correspondence concerning distribution of reports to: U.S. Army Research Institute for the Behavioral and Social Sciences, ATTN: DAPE-ARI-ZXM, 6000 6th Street, Bldg 1464 / Mail Stop 5610, Fort Belvoir, VA 22060-5610

FINAL DISPOSITION: Destroy this Research report when it is no longer needed. Do not return it to the U.S. Army Research Institute for the Behavioral and Social Sciences.

NOTE: The findings in this Research Report are not to be construed as an official Department of the Army position, unless so designated by other authorized documents.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
January 2014	Final	August 2011 – July 2012

4. TITLE AND SUBTITLE

Validation and Evaluation of Army Aviation Collective Performance Measures

5a. CONTRACT NUMBER
W5J9CQ-11-D-0004

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
622785

6. AUTHOR(S

Martin L. Bink;
Courtney Dean and Jeanine Ayers;
Troy Zeidman

5d. PROJECT NUMBER
A790

5e. TASK NUMBER
225

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

U.S. Army Research Institute
for the Behavioral and Social Sciences
6000 6th Street (Building 1464 / Mail Stop 5610)
Fort Belvoir, VA 22060-5610

Aptima, Inc.
12 Gill Street
Suite 1400
Woburn, MA 01801

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U. S. Army Research Institute
for the Behavioral & Social Sciences
6000 6th Street (Building 1464 / Mail Stop 5610)
Fort Belvoir, VA 22060-5610

10. SPONSOR/MONITOR'S ACRONYM(S)
ARI

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
Research Report 1972

12. DISTRIBUTION/AVAILABILITY STATEMENT: Distribution Statement A: Approved for public release; distribution unlimited.

13. SUPPLEMENTARY NOTES
Contracting Officer's Representative and Subject Matter Expert: Martin L. Bink @ ARI, Fort Benning

14. ABSTRACT

Simulation-based Aviation Training Exercises are critical for preparing U.S. Army Combat Aviation Brigades for deployment. However, while offering the opportunity to practice mission segments at the unit level, the effectiveness of this training remains unclear due to a need for objective assessments focused on observable team behavior. Unit Commanders and trainers need tools for measuring collective task performance in order to understand performance gains, facilitate feedback, and guide the learning of aviation tactical teams. To address this challenge, a set of aviation team performance measures were developed, data were collected to validate these measures, and strategies were created to facilitate application of the measures to collective training events. The measures used behaviorally-based observations to assess performance of aviation tactical teams. The measures were evaluated at multiple training events to assess overall utility. Data were collected on inter-rater reliability and on agreement between the measures and overall mission performance. Results provided evidence of both acceptable reliability and validity for the measures. Moreover, requirements were developed for electronic data collection tools that can be used by unit Commanders and trainers to assess team performance at collective training exercises.

15. SUBJECT TERMS
Aviation collective performance, Measurement validation, Simulation training, Army aviation

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
Unclassified	Unclassified	Unclassified	Unlimited Unclassified	74	Dorothy Young 703-545-2316

Research Report 1972

Validation and Evaluation of Army Aviation Collective Performance Measures

Martin L. Bink
U.S. Army Research Institute

Courtney Dean and Jeanine Ayers
Aptima Incorporated

Troy Zeidman
Imprimis Incorporated

Fort Benning Research Unit
Scott E. Graham, Chief

U.S. Army Research Institute for the Behavioral and Social Sciences
6000 6th Street, Bldg 1464
Fort Belvoir, VA 22060

January 2014

ACKNOWLEDGEMENT

The authors thank COL Stephen Seitz and COL Christopher Sullivan, current and previous Directors of Simulation, for their support of this project. The authors also thank LTC Michael Hansen, MAJ Michael Stachour, Mr. Kevin Hotel, and many others in the Directorate of Simulation who provided input, feedback, and coordination throughout the execution of this project. Special recognition goes to the Army aviators, simulation experts, and engineers, who served as workshop participants, for their dedication and commitment to improving Army training. Their input was of exceptional quality and was key to the success of this effort. This research effort would not have been possible without the high-quality contributions of all members of the technical team: Kerri Chik, Andy Chang, and Melinda Seibert. John Stewart of the U. S. Army Research Institute provided aviation-training expertise and critical insights throughout. Finally, the authors thank Glenn Cobb, Randy Spain, and John Stewart for thoughtful input on previous drafts of this report.

VALIDATION AND EVALUATION OF ARMY AVIATION
COLLECTIVE PERFORMANCE MEASURES

EXECUTIVE SUMMARY

Research Requirement:

Previous Army Research Institute efforts identified approximately 115 aviation collective tasks that would be performed in a "typical" scout-reconnaissance mission (Seibert, Diedrich, Stewart, Bink, & Zeidman, 2011). The goal of the current effort was to determine the empirical psychometrics of the measures by evaluating inter-rater reliability and criterion-related validity. After demonstrating acceptable inter-rater reliability, criterion-related validity was explored to determine if the measures related to performance outcomes in aviation tactical missions. The ultimate goal of the analyses was to determine the usefulness of the measures and to inform revisions to the measures and scale anchors as appropriate.

Procedure:

Reliability and validity data were obtained during two separate Aviation Training Exercise events conducted at Fort Rucker, AL. A total of 21 missions across two different units were simultaneously rated by two or more experienced aviators using the developed measures. Inter-rater reliability was estimated from these ratings. Mission-success metrics were also obtained from each mission and used as indicators of criterion-related validity. In addition to reliability and validity data, end-user feedback was obtained with surveys and interviews to understand the usefulness and utility of the measures and the measurement tools.

Findings:

The overall inter-rater reliability of the measures was considered "substantial" when ratings were within one point on the rating scales. However, five measures had unacceptable levels of reliability and were removed from the final set of measures. The criterion-related validity was acceptable, though not extremely high. The imprecision of the criteria (i.e., broad mission outcomes) likely limited the estimation of validity. Raters found the measures to be generally useful. Together, the empirical validation and ratings of usefulness supported a final set of 96 aviation collective performance measures. In addition, feedback indicated that additional interface functionality (e.g., touch screen interface and voice or video notes), additional feedback displays, and the ability to "down-select" measures would be helpful in future versions of the measurement tool.

Utilization and Dissemination of Findings:

The results from the analyses reported here were presented in briefings to the U.S. Army Aviation Center of Excellence Director of Simulation and to Project Manager – Unmanned Aviation Systems. The results were also presented at the International Symposium on Aviation Psychology in 2013. The usability and utility feedback gained in this project were used as requirements for developing a tablet-based observer measurement tool.

VALIDATION AND EVALUATION OF ARMY AVIATION
COLLECTIVE PERFORMANCE MEASURES

CONTENTS

Page

APPENDICES

TABLES

VALIDATION AND EVALUATION OF ARMY AVIATION COLLECTIVE PERFORMANCE MEASURES

Introduction

Measurement of training performance is essential for providing feedback and adapting training (e.g., Bransford, Brown, & Cocking, 2000; Ericsson, Krampe, & Tesch-Romer, 1993; Hawley, 1984; Snow & Swanson, 1992). In the case of aviation collective training, performance metrics historically have been difficult to define. For example, the assessment of performance on broad mission segments does not provide enough detail on specific collective skills (Cross, Dohme, & Howse, 1998). There is also debate for collective-performance assessments about the relative importance of measuring individual skills versus team (i.e., collective) skills and about appropriateness of outcome metrics versus process metrics (Dwyer & Salas, 2000; Turnage, Houser, & Hofmann, 1990). The issues of how and what to measure in aviation collective performance are especially relevant to simulation training. Diminishing training resources (e.g., maintenance costs, fuel cost, and access to suitable training areas) necessitate increased utilization of simulation for aviation collective training both at home station and at brigade-level mission-readiness exercises. In order to address the gap in aviation collective performance measurement, recent U.S. Army Research Institute (ARI) research (a) identified the dimensions that differentiate high-performing aviation teams from low-performing aviation teams and (b) developed measures to assess aviation collective tasks in the context of simulation-based training (Seibert, Diedrich, Stewart, Bink, & Zeidman, 2011).

The ARI measures of aviation collective performance identified approximately 115 collective tasks that would be performed in a "typical" scout-reconnaissance mission (Seibert, et al., 2011). All measures were designed to differentiate high-performing teams from low-performing teams and were intended to be used by trainers at home station or training centers such as the U. S. Army Aviation Warfighting Simulation Center (AWSC). Each of the measures listed a collective task and provided a 5-point behaviorally-anchored scale for ratings of each task. The next steps of the development process for the measures are to (a) empirically validate the discriminability (i.e., low-performing versus high-performing) of the measures, (b) evaluate the effectiveness of the measures in providing training feedback, (c) identify the most efficient format for the measures, and (d) accordingly revise the measures for implementation. The current report documents the psychometric analyses (i.e., estimates of reliability and validity) and utility analyses for these recently developed measures of aviation collective performance.

Summary of the Development of Measures of Aviation Collective Performance

The ARI aviation collective performance measures (Seibert, et al., 2011) were constructed using the Competency-based Measures for Performance ASsessment Systems (COMPASS^{SM}; MacMillan, Entin, Morley, & Bennett Jr., 2013) approach. COMPASS is a methodology for developing performance measures that combines experiential knowledge of subject matter experts (SMEs) with established psychometric practices. A set of three SME-based workshops took place over the course of five months that moved from the identification of key observable behaviors to the construction of performance measures. The first and third

workshops were group interviews while the second workshop consisted of individual or small group interviews. A total of 27 SMEs participated across all workshops including 3 SMEs who participated in all three workshops. SME expertise ranged from military aviators to simulation training experts and software engineers.

In the first step of measure development, the phases of the attack/reconnaissance mission were deconstructed into observable behaviors, or performance indicators (PIs), that allow an expert to determine whether an individual or team was performing well or poorly. The resulting PIs and relevant missions/tasks provided a solid basis on which to develop benchmarked measures that were less sensitive to subjective biases and more reliable over repeated sessions. In the second step, SME-provided input was crafted into specific performance measures associated with each PI in order to create performance measures with appropriate behaviorally-based rating scales (i.e., 5-point Likert-type scales). To obtain exemplar behavior information, SMEs were asked to describe and identify explicit behaviors that were representative of good, average, and poor performance. Altogether, 130 candidate observer-based performance measures were developed. Figure 1 provides an example for the performance measure *Request Clearance of Fires from Ground Commander*.

Does the flight request clearance of fires from Ground Commander?				
1	2	3	4	5
Flight does not request clearance of fires		Flight considers ROE; establishes friendly/enemy positions; requests clearance of fires; not ready to effect the target while going through this process		Flight considers ROE; establishes friendly/enemy positions; requests clearance of fires; anticipates clearance and sets up shot during this process

Figure 1. Example Performance Measure - Request Clearance of Fires from Ground Commander. *ROE* stands for rules of engagement.

Throughout the measure development process, care was taken to ensure that measures were operationally relevant, thorough, and appropriately worded using domain language and terminology. The third step of the development process was a final check to ensure the quality and face validity of the performance measures. A group of SMEs reviewed the full set of measures and was asked to revise each measure to ensure the measures could be understood and accepted by a wide range of potential users. During this workshop, each performance measure was reviewed with respect to the following criteria:

- *Relevance:* Does the measure assess an aspect of performance that is important for mission readiness?
- *Observability:* Does the measure assess a behavior that is truly observable?
- *Question wording:* Does the measure make sense to other SMEs?

- *Scale type:* Is the scale used appropriate for differentiating behavior?
- *Scale wording:* Do the behavioral anchors make sense to other SMEs?

As appropriate and based on SME input, modifications were made to the measures, resulting in a final list of 115 observer-based performance measures for assessing the performance of an aviation collective team performing an attack/reconnaissance mission.

Technical Objectives

In order to determine the empirical psychometrics of the measures, estimates of inter-rater reliability and criterion-related validity were obtained. These analyses also informed revisions to the measures. In addition to analyses of reliability and validity, end-user reaction feedback was obtained after first-hand use of measurement tools to understand how the measures and measurement tools were perceived. Several individual interviews and focus group discussions were conducted regarding the measures themselves, particularly how the measures could be combined to create a useful performance report and how the measures can best be implemented to facilitate ease of use in real-time training events. During these interactions, trainers, unit leaders, and other senior aviators were asked to describe how/if they could picture the ARI measures being used and applied in training events like the aviation training exercise (ATX) as well as home-station collective training. Finally, revisions were made to the measures as a result of the empirical analyses and the end-user feedback.

Reliability and Validity Analyses

In general, reliability analyses demonstrate that measures are consistent across time and/or between raters. Once acceptable reliability has been achieved, validity analyses demonstrate that measures apply to intended constructs and/or that they predict anticipated outcomes. In the current effort, *inter-rater reliability* was evaluated as the intended use of the measures required that different raters use the scale similarly. After establishing acceptable reliability, *criterion-related validity* was explored to determine if the measures related to team outcomes in aviation tactical missions. The ultimate goal of the reliability and validity analyses was to inform revisions to the measures and scale anchors as appropriate.

Reliability and validity data were obtained in a naturalistic setting during two separate ATX events conducted at Fort Rucker, AL. The ATX is the primary mission-readiness exercise for a Combat Aviation Brigade (CAB) before deployment. ATX utilizes the simulation capabilities of the AWSC to place CAB aircrews and battlestaff in a common virtual environment. The aircrews fly networked cockpit simulators that can be reconfigured to represent the Army's four currently operational combat helicopters (AH-64D/E Apache, CH-47D/F Chinook, OH-58D/F Kiowa Warrior, and UH-60 A/L/M Blackhawk). Aircrew performance is currently evaluated by Observer-Controllers-Trainers (OC/Ts) who watch real-time video and audio feeds as the missions are flown.

Method

Participants

Expert raters represented a combination of current and former OH-58, UH-60, and AH-64 pilots with recent deployment experience in the current theater of operations (i.e., Southwest Asia). The primary rater across every exercise was a member of the research team and former combat-experienced Army aviator. Active-duty raters were OC/Ts at the respective ATXs. These OC/Ts came from various aviation units and were selected because of recent deployments to the specific area of operations in which the ATX missions took place. Active-duty raters included grades of Chief Warrant Officer 2, Chief Warrant Officer 3, Lieutenant, Captain, and Major. Fifteen active-duty pilots served as raters for a total of 16 raters. Because the primary duty of OC/Ts is to train aircrews during ATX, participation in the research represented an additional task for these individuals.

Materials

Each of the 115 ARI aviation collective performance measures (Siebert, et al., 2011) were implemented in electronic format on a ruggedized laptop. The implementation of the measures was supported by the SPOTLITE application (Jackson, et al., 2008; MacMillan, et al., in press; Wiese, Nungesser, Marceau, Puglisi, & Frost, 2007). This measurement tool allowed raters to evaluate collective performance in real time. Mission-success metrics (see Appendix A) were designed to serve as validity criteria for ARI aviation collective performance measures. The mission-success metrics were composed of nine objective mission outcomes based on Aircrew Training Manuals, Mission Essential Task Lists, Army Training and Evaluation Programs, and other training documentation for collective training. Examples of mission-success metrics include number of targets destroyed, number of friendly aircraft lost, and instances of fratricide. Raters made their evaluations in real time in the AWSC "Stealth" observation room. The Stealth room includes various feeds including radio communications, "God's-eye-view" video, and in-cockpit visual systems with which raters were able to monitor the flight teams' behaviors.

Procedure

Data were obtained during two separate ATX events conducted between late 2011 and early 2012. Complete missions were rated from pre-launch to landing/mission completion. No interruptions or interactions with the flight teams occurred between the raters and the flight teams. Additionally, no injects, or changes to the scenario were introduced by the expert observers. During each mission, each flight was composed of either two OH-58 aircraft or two AH-64 aircraft, and each flight included an Air Mission Commander, a Pilot-in-Command, and two pilots. No formal process for selection of pilots was applied. Pilots were assigned to missions by unit leaders based on unit priorities and the exercise's mission set. The identities and qualifications of the pilots flying missions were not made available to the research team.

A total of 21 Attack Weapons Team or Scout Weapons Team missions across three different units were observed. Missions ranged from convoy escort to deliberate operations (e.g., Air Assault) and flight times lasted between 100 and 120 minutes each. Missions required coordination with other aircraft, ground forces, and tactical operations centers. Of the 21 missions, 15 were simultaneously rated by two or more raters. Three of those 15 featured three separate raters. The remaining six missions were rated by one rater. Mission-success metrics were obtained from 21 missions and focused on more objective collective outcomes of the mission (e.g., mission accomplishment, achievement of objectives, number of targets destroyed, aircraft lost). While raters evaluated flight team performance in real time, mission-success metrics were completed at the end of missions by the same raters. Both collective-performance ratings and mission-success metrics were recorded using the computer-based measurement tool. Given these data, inter-rater reliability was evaluated on the 15 missions with multiple raters while criterion-oriented validity was evaluated on all 21 missions.

Results and Discussion

Inter-rater reliability

While inter-rater reliability is a standard approach for demonstrating that raters use measures and scale anchors similarly, evaluations of other measure properties such as percent agreement can be insightful tests of the reliability of ratings (Howell, 1997). Further, percent agreement as computed in this research can help identify measures that were especially problematic for raters to agree upon – an important step for revising as well as down-selecting the large measures set to a manageable number of the best performing and most useful items. As a result, inter-rater agreement was first assessed and then supplemented with an inter-rater reliability analysis.

Inter-rater agreement was established using a percent agreement method based on the range of ratings for each measure across the raters (e.g., both raters within one rating point). Raters indicated the level of performance observed for each measure on a 9-pt scale (i.e., response options ranged from 1 – 5 with half point intervals). For each measure that was rated by two or more raters for the same event, comparisons were made between the recorded ratings. Agreement was calculated as the net difference between the values supplied by the two raters. Several categories of agreement were established:

1. *Absolute agreement*. Both raters provided the exact same rating (e.g., 4) resulting in a net difference of 0.
2. *Strong agreement*. The absolute difference between ratings was 0.5 or 1.
3. *Some agreement*. The absolute difference between ratings was 1.5 or 2.
4. *No agreement*. Raters differed substantially in their ratings (e.g. more than 2-point differences).

For each level of agreement, *percent agreement* was calculated by dividing the observed agreement counts by the total number of possible observations.

As Table 1 shows, when aggregated across all rated missions, raters achieved a 72% agreement within 1-point on the measurement scales. Put differently, if one rater gave a rating of five, for example, the other rater(s) was likely to give a rating of at least four in 72% of the occasions. Thus, in this example, both raters agreed that behavior was well above average.

Table 1.
Percent Rater Agreement by Level of Agreement.

Agreement Level	Number of Paired Ratings	Percent Agreement	Cumulative Percent
0	160	29%	29%
0.5	88	16%	45%
1	145	27%	72%
1.5	45	8%	80%
2	65	12%	92%
>2	41	8%	100%

Inter-rater reliability was estimated using Cohen's Kappa (κ). Kappa is a generally conservative measure of inter-rater agreement that estimates exact agreement between two raters and that accounts for chance agreement (Cohen, 1960; Fleiss, 1981). Interpreting the significance of Kappa is based on degrees of confidence at different value intervals rather than on *p* values (Fleiss 1981; Landis & Koch, 1977). Kappa values ranging from 0.01 – 0.20 are regarded as *Slight Agreement*. Values between 0.21-0.40 are considered *Fair Agreement*. *Moderate Agreement* is achieved when values range from 0.41-0.60. *Substantial Agreement* corresponds to values between 0.61 – 0.80, and values above 0.81 are considered *Almost Perfect Agreement*. Negative values are interpreted as having *Poor Agreement*.

In the interest of exploration, Kappas based on both exact agreement and agreements within one point were examined in the present analysis. Because of the high rate of agreement within one point across items, it made sense to analyze the level of agreement between raters at or within one point on the rating scale. Exact Kappa suggested *slight agreement* among raters ($\kappa = 0.13$). However, agreement was *substantial* when Kappa was computed for agreements within one point ($\kappa = 0.66$).

Overall, the inter-rater reliability analyses suggested that the ARI aviation collective performance measures were generally interpreted similarly by different raters. However, these results also suggested that some measures were less reliable than others. Further examination

assessed which specific measures tended to have lower or higher levels of agreement. To accomplish this, a set of criteria were developed based on the limited size and distribution of the data set that attempted to capture both frequency of use and agreement given the nature of the data. More specifically, the most reliable individual measures were identified using the following criterion: a minimum of 10 instances where two or more pilots rated the item for the same mission (slightly less than 50% of total possible), and rating agreement at or within 1-point in 80% of the observations. In contrast, the least reliable measures followed a less stringent criterion: a minimum of eight paired observations with disagreement equaling or exceeding 1.5 on at least 45% of observations. These criteria, although somewhat arbitrary, permitted the sorting and identification of measures that should be considered for revision and/or removal given the nature of the data collected. Furthermore, it is important to note that small changes in these criteria (e.g., number of instances required) did not substantially impact subsequent analyses. Table 2 identifies the eight measures with highest agreement, and Table 3 identifies the five measures with lowest agreement.

Table 2.
Measures with Highest Levels of Agreement.

Measure	Number of Rated Pairs	Agreement
Does the flight monitor ground channels?	10	80%
Does the flight follow appropriate communication protocol?	12	83%
Does the flight receive the SITREP from ground?	10	90%
Does the flight confirm location of friendlies verbally?	11	82%
Does the flight use the appropriate sensors to search for targets?	10	80%
Does the flight use the correct terms to announce target in sight?	11	82%
Does the wingman confirm target detection?	10	90%
Does the wingman use correct terms to confirm target?	10	90%

Note: SITREP = Situation report.

Table 3.
Measures with Lowest Levels of Agreement.

Measure	Number of Rated Pairs	Agreement
Does the flight communicate location of the friendlies to the ground Tactical Operations Center?	8	50%
Does the flight identify the location of friendlies using all sources available?	10	50%
Does the flight work with the ground to establish task and purpose for their mission?	9	56%
Does the flight discuss applicable changes to the tactical mission?	11	45%
Does the flight consider ground Commander's intent?	8	50%

Criterion-related Validity

While this analysis does not directly speak to criterion validity because of the lack of an existing standard for collective task performance, it served as a way to identify consistency and relation between ARI measures and objective outcomes. For the criterion-related validity analysis, the five least reliable measures were omitted. Mean ratings across measures for each rater were compared to the mean ratings across mission-success metrics. A scatterplot of the resulting rating pairs (Figure 2) illustrates the nature of the relation between ARI aviation collective measures and mission-success metrics. As Figure 2 shows, there was a positive relation between the measures and the objective outcomes ($r = 0.48$, n = 32, $p < 0.01$).

Conclusions

Taken as a whole, these data provided initial evidence that the recently developed ARI aviation collective performance measures are, in general, reliable and correlate with objective mission outcomes. These analyses also provided insights on how to revise the measures by identifying subsets of measures that were most and least reliable. While the findings on reliability and validation were promising, the results did not, however, provide definitive evidence for the validity of the ARI aviation collective performance measures. There were two primary contributing factors to the lack of conclusiveness from the current results.

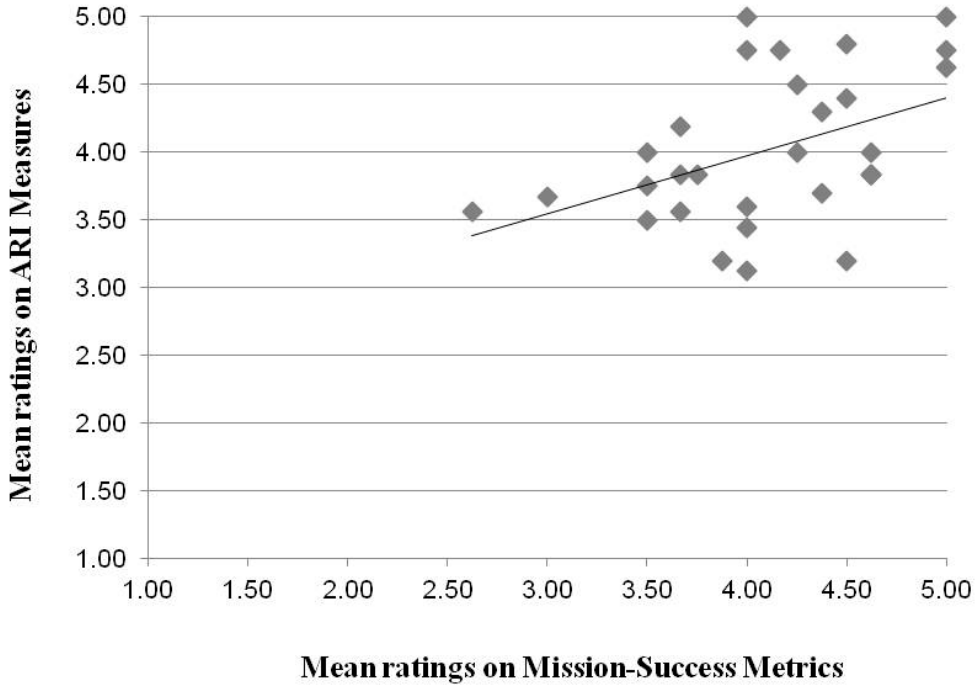


Figure 2. Scatterplot of Mean Ratings for ARI Aviation Collective Performance Measures and Mean Ratings of Mission-Success Metrics.

First, the lack of a controlled observation environment likely led to the instability of ratings. On the one hand, data collection at ATX enabled exploration of the use of the measures in an actual training setting by actual trainers (i.e., OC/Ts) thereby enhancing applicability of the measures to the intended training setting (i.e., ecological validity). On the other hand, given time constraints and demands on OC/Ts at ATX, this environment also limited the ability to extensively train observers and to engineer scenario events to explicitly explore reliability and validity as might be done in a laboratory setting. The current effort generally did not afford the opportunity to train raters more than five minutes prior to mission start, and it was not uncommon for unit leaders to re-direct rater attention to a particular task or mission event, thereby creating variance in when and how measures were recorded. Considering the many uncontrollable environmental factors present during the ratings, these results were quite promising. Future evaluation of these measures should include more extensive rater training on the measurement tool and scales prior to testing as well as more control over rater attention and focus.

Second, the lack of existing measures of aviation collective performance and of sufficient outcome metrics of collective performance made the estimation of validity difficult. That is, if there are no existing definitive benchmarks of aviation collective performance, then validating new measures of aviation collective performance against a criterion is nearly impossible. In the current research, an attempt was made to define the best criteria as possible. However, it has been suggested that broad mission segments, such as used here for criteria, are not the best

indicators of collective performance because they lack details about pilot interaction, decision making, critical thinking skills, and team actions (Cross, et al., 1998). So, even though the current results showed only moderate correlation between the new measures and the criterion metrics, it may be the case that the criterion was less indicative of performance rather than the lack of precision of the measures. It should be noted that the content validity of the measures was carefully demonstrated by the relation to doctrine and the reliance on subject-matter expertise in the development of the measures (Seibert, et al., 2011). It should also be noted that because of the lack of clear criterion, few Army tests are ever validated (Turnage, et al., 1990). Clearly, additional research will be needed to further support the validity of these measures and to address the validity of the measures in other contexts. However, support for validity in this initial analysis can guide further validity analyses with other criterion metrics and/or analyses of construct validity through comparisons with other measures of team performance.

Usability and Utility Analyses

The goal of this set of analyses was to provide evidence that the set of measures could be utilized as a viable asset to training. Whereas the reliability and validity analyses addressed the empirical and conceptual properties of the measures, the usability and utility analyses addressed the practical properties of the measures and the software tool used to collect the measures. There were three central issues addressed in present analyses. First, the analyses attempted to determine the usefulness of the measures for training. Second, the analyses attempted to determine how the volume of measures (i.e., over 100 individual measures) could be best managed to provide effective feedback without overwhelming raters. Third, the analyses attempted to determine the degree of usability of the software tool and to gather additional requirements for a hand-held tool to implement the measures. Ultimately, the usability and utility analyses were intended to support the validation data in demonstrating the value of the measures to the Army.

Method

Participants, Materials, and Procedure

An iterative series of focus groups and individual interviews with Army Aviation SMEs from two different Army installations were conducted within the continental United States at different stages in the deployment cycle process. Overall, participants' backgrounds varied by role within a CAB (e.g. Battalion Commander, Instructor Pilot, Company Commander, Rated Pilot, Military Intelligence) as well as by platform (OH-58D, AH-64D, UH-60L/M), and by grade (Chief Warrant Officer 3 to Colonel). While the majority of SMEs were experienced active duty rated pilots, the variation in background and experience provided a variety of perspectives on the utility of performance measures and the ease of use of the measures. In general, participants were provided with the ARI aviation collective performance measures (Seibert et al., 2011) either as a printed hard-copy document or implemented in the software tool previously described. Participants were asked to rate the performance of flight teams during

simulation-based training events using the measures. Following completion of this task, participants were asked to complete surveys or to participate in a discussion about their experiences using the measurement tool and their overall impressions of the measures.

Utility Focus Group and Interviews

Nine senior pilots and leaders participated in individual and group structured interviews during three different ATXs. In these structured interviews, participants were presented a paper-based set of measures and asked to provide feedback on the perceived utility of the measure set at events like ATX and usability feedback on the initial measurement tool. A two-part interview protocol was followed, which featured specific questions addressing the training utility of the measures and the ease of use of the assessment system to collective training exercises (see Appendices B & C).

Usability Surveys

Seven participants completed a post assessment survey following use of the performance measures. The surveys were distributed to participants from two different CABs during ATXs. The survey was developed to gather feedback from users on the usability of the measurement system. The survey asked respondents to rate different dimensions of utility for the measures and rate the usefulness of the measurement tool software (see Appendix D). The response scale for each item had five options ranging from "Not at all" (1) and "Very Much" (5). Participants were also provided an opportunity to indicate any additional comments on the measures and tool during this time.

Usability Focus Groups and Interviews

Two separate half-day usability focus groups were conducted with four and two participants, respectively. An interview with an additional representative of the CAB was held via telephone after the focus groups. The focus groups were conducted during the CABs participation in a collective training event at home station. At a later ATX, 12 pilots from the same CAB participated in individual and group interviews. To start the workshops, participants were presented a slide deck that outlined the functionality of the current measurement tool and feature ideas for a revised tool (see Appendix E). In addition, the interview questions in Appendix B were again used during this workshop. The structured interview approach was applied to guide the discussion and ensure an unbiased assessment of the utility of the collective performance measures and the usability of the measurement tool. Usability questions were designed to determine how easy it was to use the observer measures to collect feedback on collective performance. Additional questions were designed to determine how adaptable the measures are for various training requirements. That is, input was given on how to logically reduce the number measures used for different training requirements. During each focus group, participant comments were documented and displayed in real time so that participants could refer to the results throughout the interview.

Results and Discussion

Utility Focus Group and Interviews

From participants' responses, a list of key themes was compiled, and the data were organized according to utility. Several key themes appeared frequently and, as expected, there was a considerable amount of consistency in comments both within and between groups of participants. The six most frequently mentioned themes are represented in Table 4. Participants indicated support for inclusion of a formative performance measurement system in training events. Overall, participants communicated a preference for qualitative feedback over quantitative feedback. However, being able to track performance across time and units (e.g., trend analysis) was expressed as a necessary training capability. Many participants felt that current methods of after-action reviews (AARs) are effective, but most participants agreed that additional performance measures would enhance the quality of AAR feedback and help to identify shortcomings and opportunities for improvement. Additional thematic analysis identified three core areas where participants felt performance measurement and feedback were most useful: (a) assessing crews on their communication and team tactics, (b) assessing how crews execute standard operating procedures and other tactical processes, and (c) assessing how effectively crews use their systems and sensors. These core areas corresponded with training objectives that were later identified as a basis for customizing the measurement tool interface. Finally, the ability to track and analyze trends in units and training over time was mentioned in three cases as being desired.

Table 4.
Response Frequencies for Utility Themes Compiled from Focus Groups and Interviews.

Utility Theme	Response Frequencies
Performance Measurement	24
After Action Review	17
Crew Coordination/Team Tactics	11
Standard Operating Procedure and Processes	11
Systems and Sensor Usage	3
Trend analysis and tools	3

Usability Surveys

The survey results suggested that the measures and measurement tool supported effective performance assessment. Table 5 shows the response frequencies for the Likert-scaled items. All responses were at least a '3' ("Somewhat") and the modal response was a '5' ("Very Much"). In addition to the items listed in Table 5, Item 4 and Item 9 asked "Yes/No" questions with open-ended options for pilots to detail their answers. Item 4 asked for any specific measures that were confusing, out of place, or inappropriate. The majority of participants (57%) responded "Yes" and provided a description of one or more issues. This feedback was incorporated into measure revision. Item 9 asked if pilots would "use a measurement tool like this in the future" if given the opportunity. All of the participants (100%) responded "Yes" and several provided comments that communicated satisfaction with the performance measures. One interesting result from this survey was that Item 5, though only answered by two participants, received a rating of '5' from both. This suggested the measurement tool was effective. However, the low response rate makes these data difficult to interpret and suggests some caution.

Table 5.
Response Frequencies for Usability Survey Likert-scaled Items.

Item	Response				
	1	*2*	*3*	*4*	*5*
1. How useful were the measures for assessing Soldier performance during the exercise?	0	0	0	3	4
2. How well did the answer scales match the questions?	0	0	0	2	5
3. How well did the measure questions match the mission events unfolding?	0	0	1	4	2
5. How easy was it to provide ratings using the software?	0	0	0	0	2
6. How easy was it to navigate through the mission phases in the "tree" on the left?	0	0	0	1	2
7. How easy was it to match the questions with the events unfolding in the mission?	0	0	1	3	1
8. Overall, how useful was the device for assessing Soldiers conducting aviation missions?	0	0	0	2	3

Usability Focus Groups and Interviews

A list of key usability themes was compiled from participants' responses. Key usability themes were defined as frequent responses across participants. The key themes fell into four categories: Information Volume, Ratings Interface, AAR, Desired User Interface Functions and Features. The top 10 key usability themes are presented in Table 6 and organized by category. The table identifies each theme and the frequency of mentions across participants.

Table 6.
Response Frequencies of Key Usability Themes.

Theme	Response Frequency
Information Volume	10
Measure Customization	8
Trainee Data	2
Ratings Interface	17
Desired New Features	9
Add Attachments to Measures	6
Measure Tree	2
After Action Review	11
Training Objectives	5
Trending Tools	3
Visual Representations of Data	3
General User Interface Functions and Features	14
Touch Controls	8
Human Factors Issues	6

The key usability themes were used to revise the measures. To address issues raised for the Information Volume themes, the idea of incorporating filters to downselect the measures was offered. More specifically, it was determined that sets of measures could be selected based on (a) mission type, (b) mission phase, (c) training objective, or (d) role (e.g., air-mission commander). Accordingly, only measures in the selected category would be presented and would serve to focus the types of performance measures made. Feedback also indicated the need to be able to provide more specific student data for the purposes of tracking and trending.

Much of the feedback for the Ratings Interface themes focused on flagging key measures to ensure they were completed. This feedback also indicated the need to be able to add attachments such as pictures, video, and voice notes to the individual measures. Generally, participants were pleased with the navigation within the software tool and the procedures for the management of within-mission measures. Key themes for AAR capabilities indicated a need for measures to be linked back to specific events and for the results to be displayed in a meaningful way that could facilitate timely, formative feedback. Participants offered many suggestions on how to display those results. An additional capability that was identified was a trending capability that would allow trainers and leaders to compare performance both across flight teams and within flight teams over time.

The primary User Interface Functions and Features usability themes centered on a lack of features (e.g., voice, video memos, measure results/AAR). Participants were also consistently concerned about the awkwardness of switching between the stylus and keyboard to provide inputs on the laptop implementation of the software tool. Feedback suggested that a smaller and lighter touchscreen tablets would be a more appropriate platform to implement the measurement tool. Additional feedback indicated how such a touchscreen tool should operate.

Revisions

Final revision of the measures was largely based on the SME input received during focus groups and interviews. Several measures were identified as being either redundant or rarely observed in normal operations. Some of the measures with low inter-rater reliability were retained because SMEs believed the content reflected mission-critical behaviors. It was difficult to determine how to modify the low-reliability measures to increase rater agreement, so the measures were retained without revision. In addition, the rating anchors for one measure were revised. SMEs indicated that the anchors contained language that was too leading. That is, the original anchors only implied an incomplete mission outcome and left no rating option to indicate the mission was completed. The final set of validated measures contained 96 items once the revisions were made and superfluous measures removed (see Appendix F).

General Discussion

The primary objective of this research effort was to develop a reliable, valid, and useful set of measures to assist trainers and leaders in assessing aviation collective performance. Using these measures, it is anticipated that trainers and leaders will be better able to review performance, identify strengths and weaknesses, and provide consistent behaviorally-based feedback to improve the performance of aviation teams. Here, the focus was on collective tasks critical to performing typical scout and attack missions. More generally, beyond ATX, these measurement tools could be useful in preparing for and conducting assessments in a variety of collective training events (e.g., at home station).

The research effort reported here resulted in the construction of 96 revised measures focusing on key skills for aviation collective tasks. Initial data concerning reliability, validity, utility, and usability were collected and led to the refinement of these measures. These data

provided evidence that the measures are in general reliable and indicated an acceptable association between the measures and available metrics of mission outcomes. It should be noted that while the findings on reliability and validity were limited and preliminary, these analyses provided data on the subsets of measures that were most and least reliable, which enabled measure revision and refinement. In addition, information was collected on the refinements to best enable future use of the measures.

The reliability and validity results were not unequivocal. The high levels of inter-rater reliability were contingent on ratings being within one point of each other. Although absolute agreement would be preferred, the use of agreement within one point is not without precedent (e.g., Chouinard & Margolese, 2005). The levels of reliability were likely influenced by the fact that at least one rater was unfamiliar with the measures during each rating event. That is, for each rating event, one rater was a member of the research team who had familiarity both with the nature of the measures and with the design of the measurement tool and the other rater was from the training exercise personnel who was using the measures and measurement tool for the first time. The lack of familiarity with the measures and the additional duties of the training personnel may have caused some inattention to the measures and led to some discrepancies in ratings. Of course, familiarity is only one source of variation in ratings that may (or may not) influence reliability (e.g., Murphy & De Shon, 2000).

An additional issue for the reliability of the measures and for the criterion-related validity of the measures was the lack of clear and objective mission-success metrics. The mission-success metrics used only provided gross estimates of team performance as compared to the more fine-grained collective performance measures. This difference in specificity between the measures and criteria can result in *underestimates* of both content and criterion-related validity estimates (Hogan & Holland, 2003). Unfortunately, more precise mission-success metrics are not available for Army aviation collective performance (see Cross, et al., 1998). The implication of the lack of appropriate mission-success criteria is that any estimate of criterion-related validity would be somewhat imprecise and sub-optimal. Accordingly, the accuracy of the measures of performance may be substantially more robust than indicated by estimates of validity reported here (i.e., $r = 0.48$ between measures and mission-success criteria).

Regardless of the robustness of the estimates of validity presented here, criterion-related validity is only one part of the holistic view of validity. As previously indicated, construct validity of these measures may not be possible, but evidence already exists for the content validity of the measures. As part of the original development of the measures, the content of each measure was vetted by SMEs for mission criticality and training criticality (Siebert et al., 2011). SMEs indicated that the measures were accurate for aviation collective tasks and contained critical performance metrics for training. Clearly, additional support for the validity of the measures would be helpful. However, it may be the case that the developed measures of performance will serve as better benchmarks for aviation collective tasks in subsequent research than existing mission-success metrics because of the finer level of specificity of the developed measures.

The results from the analyses reported here were presented in briefings to the U.S. Army Aviation Center of Excellence Director of Simulation and to the Project Manager – Unmanned Aviation Systems. The results were also presented at the International Symposium on Aviation Psychology (Bink, Seibert, Dean, Stewart, & Zeidman, 2013). The usability and utility feedback gained in this project were used to inform requirements for developing a tablet-based observer measurement tool.

References

Bink, M. L., Seibert, M., Dean, C., Stewart, J. E., & Zeidman, T. (2013). Development and validation of measures for army aviation collective training. *Proceedings of the 17th International Symposium on Aviation Psychology.* Dayton, OH.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (2000). *How people learn: Brain, mind, experience, & school*. Washington, DC: National Academy Press.

Chouinard, G., & Margolese, H. C. (2005). Manual for the Extrapyramidal Symptom Rating Scale (ESRS). *Schizophrenia Research, 76,* 247 – 265.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20,* 37- 46.

Cross, K.D., Dohme, J.A., & Howse, W.R. (1998). *Observations about defining collective training requirements: A White Paper prepared in support of the ARMS program*. (Technical Report 1075). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (DTIC No. ADA349437).

Dwyer, D. J., & Salas, E. (2000). Principles of performance measurement for ensuring aircrew training effectiveness. In H. F. O'Neill & D. H. Andrews (Eds.), *Aircrew training and assessment* (pp. 223-244). Mahwah, NJ: Lawrence Erlbaum Associates.

Ericsson, K. A., Krampe, R. T., & Tesch-Romer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review, 100,* 363 – 406.

Fleiss, J. L. (1981). *Statistical methods for rates and proportions*. (2nd ed). New York: Wiley.

Hawley, J. K. (1984). Some considerations in the design and implementation of a training device performance assessment capability, *Proceedings of the Human Factors Society 28th Annual Meeting*, pp.201 – 205.

Hogan, J., & Holland, B. (2003). Using theory to evaluate personality and job-performance relations: A socioanalytic perspective. *Journal of Applied Psychology, 88,* 100 – 112.

Howell, D. C. (1997). *Statistical methods for psychology (4th ed.).* Belmont, CA: Duxbury Press.

Jackson, C., Woods, H., Durkee, K., O'Malley, T., Diedrich, F., Aten, T., Lawrence, D., & Ayers, J. (2008). Tools for assessment of operator contribution to system performance. *Proceedings of the Undersea Human Systems Integration Symposium*. Bremerton, WA.

Landis, J. R. & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33,* 159–174.

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

Murphy, K. R., & De Shon, R. (2000). Interrater correlations do not estimate the reliability of job performance ratings. *Personnel Psychology, 53,* 873 – 900.

Seibert, M. K., Diedrich, F. J., Stewart, J. E., Bink, M. L., & Zeidman, T. (2011). *Developing performance measures for Army aviation collective training*. (Research Report 1943). Arlington, VA. U.S. Army Research Institute for the Behavioral and Social Sciences. (DTIC No. ADA544425).

Snow, R. E., & Swanson, J. (1992). Instructional Psychology: Aptitude, adaptation, and assessment. *Annual Review of Psychology, 43,* 583 – 626.

Turnage, J. J., Houser, T. L., & Hofmann, D. A. (1990). *Assessment of performance measurement methodologies for collective military training*. (Research Note 90-126). Alexandria, VA. U.S. Army Research Institute for the Behavioral and Social Sciences. (DTIC No. ADA227971).

Wiese, E., Nungesser, R., Marceau, R., Puglisi, M., & Frost, B. (2007). Assessing trainee performance in field and simulation-based training: Development and pilot study results. *Proceedings of the 27th Annual Interservice/Industry Training, Simulation and Education Conference,* Orlando, FL

APPENDIX A

MISSION-SUCCESS METRICS

OC/OT Observer Number: _____ Today's Date: _____

TEAM EVALUATING

BN: _____ CO: _____ Launch Time: _____z Type of Mission: _____

For the team indicated above, please provide ratings for the following questions:

1. Did the flight accomplish its mission?

☐ N/A

☐ N/O

1	2	3	4	5
No mission objectives were met		Flight met enough objectives to accomplish the mission		Flight met all mission objectives and met commander's intent

2. Did the flight have the desired effect on the target?

☐ N/A

☐ N/O

1	2	3	4	5
Flight missed target or had no effect on target		Flight hit target but did not have desired effect (e.g., disabled, but not destroyed)		Flight hit target with desired effect (e.g., observed, disabled, or destroyed)

3. Were there instances of collateral damage during the mission?

☐ N/A

☐ N/O

1	2	3	4	5
Actions resulted in collateral damage (civilian deaths, one or more buildings leveled)		Actions resulted in minimal collateral damage (no civilian deaths, no buildings leveled)		Actions resulted in no collateral damage

4. How many enemy targets were ENGAGED? _____
5. How many enemy targets were DESTROYED? _____
6. How many TOTAL friendly aircraft were lost? _____
    a. Specify number lost to enemy fire: _____
    b. Specify number lost to fratricide, or accidents? _____

7. How does this team compare to other teams you have observed performing a similar mission?

☐ N/A

☐ N/O

1	2	3	4	5
This team's performance was in line with the BOTTOM third of teams I have observed		This team's performance was in line with the MIDDLE third of teams I have observed		This team's performance was in line with the TOP third of teams I have observed

FLIGHT TEAM COMPARISON MEASURE

OC/OT Observer Number: _____ Today's Date: _____

TEAM EVALUATING

BN: _____ CO: _____ Launch Time: _____z Type of Mission: _____

For the team indicated above, please provide ratings for the following questions:

1a. If you observed other teams performing a similar mission, list them in the table below specifying them by date and launch time. Then, please rank order the teams you observed where a rank of 1 indicates the best team you observed. Please use each rank number only once, as demonstrated in the example.

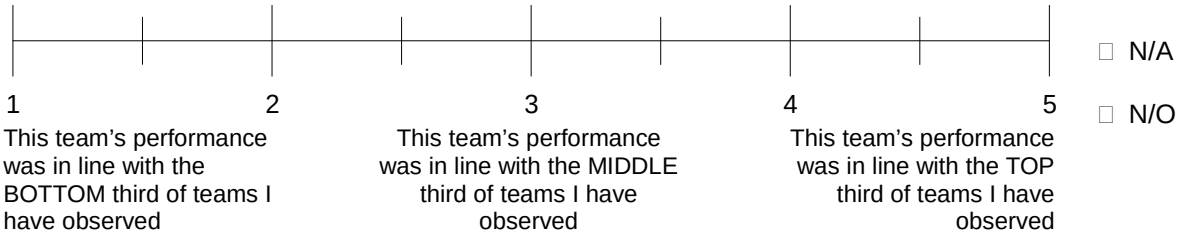
Example Ranking

Date of Exercise	Launch Time	Rank (1 = best team)
5 APR 2011	1500 z	4
5 APR 2011	1200 z	5
6 APR 2011	1500 z	1
7 APR 2011	1830 z	3
7 APR 2011	1200 z	2

(o 1b.)

		Rank (1 = best team)
	z	
	z	
	z	
	z	
	z	
	z	
	z	

1b. If no other teams performed the same mission during this ATX, how does this team compare to past teams performing a similar mission?



1 — This team's performance was in line with the BOTTOM third of teams I have observed

3 — This team's performance was in line with the MIDDLE third of teams I have observed

5 — This team's performance was in line with the TOP third of teams I have observed

APPENDIX B

UTILITY INTERVIEW PROTOCOL

Suggestion Questions for in-person interviews:

1. Overall, how useful was the *device* for assessing Soldiers conducting aviation missions?
2. How useful were the *measures* for assessing Soldier performance during the exercise?
    a. Are there ways the measures could be improved?
3. If given the opportunity, would you use a device such as this in the future to assess your Soldiers?
    a. Why/Why not?
4. What did you like about the measures? The technology?
5. What did you dislike about the measures? The technology?
6. How well did the measure questions match the mission events unfolding?
    a. Are there ways this could be improved?
7. Were any specific questions confusing, out of place or inappropriate for the mission exercise? (yes/no)
    a. If yes, which one(s)?

b. Why were they confusing? (Examples: wording was confusing, measure was irrelevant to the mission, the scale didn't match the question, the measure was out of place – should have been grouped with other measures)

8. How easy was it to provide ratings using the software?
    a. What would make it easier?
9. How easy was it to navigate through the mission phases in the "tree" on the left?
    a. What would make it easier?
10. How easy was it to match the questions with the events unfolding in the mission?
    a. What would make it easier?

APPENDIX C

USEFULNESS INTERVIEW PROTOCOL

SUGGESTED QUESTIONS:

1. What types of flight team performance feedback would be most useful to pilots following collective task exercises?
    a. How should this feedback be formatted/presented?
2. What do your pilots/you as a pilot desire as a product of ATX exercises?
3. What format does feedback at ATX currently take?
    a. Is this format satisfactory? Why/why not?
    b. How would you like information presented in the future?
4. What qualitative (e.g., descriptive) mission performance feedback is currently provided at ATX?
5. What quantitative mission performance feedback is currently provided at ATX?
6. What type, format, and content of feedback from ATX would most benefit flight teams as they prepare for deployment?
7. If you could have any feedback from your performance in ATX (flight team level performance), what would you want to know?
    a. What would help you the most as you prepare for your deployment?
    b. What elements of mission performance
8. Who should be the target audience(s) for this type of feedback?
    a. Who should provide this type of feedback to them?/From whom should the feedback come?
    b. Is there different feedback that could/should be provided to different audiences?
9. As a CO CDR, BD CDR, or IP and you were to receive a 'take-home package' about flight team performance during ATX, what would you want it to tell you?
    a. How would you use it to support preparation for deployment?

APPENDIX D

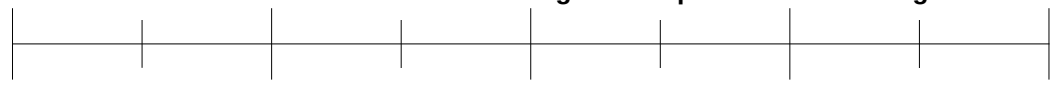
USABILITY SURVEY

Date:_____ Your Role (IP, OC/OT, etc.) _____ TF under review:_____ Mission:__ _____

REACTION TO ASSESSMENT SYSTEM:

The following is a brief set of questions pertaining to the assessment system you just used. Please respond to each item based on your experience using the questions and the technology.

1. **How useful were the *measures* for assessing Soldier performance during the exercise?**



1 2 3 4 5

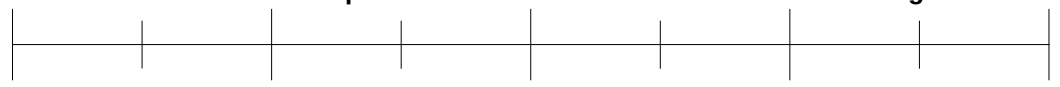
Not at all Somewhat Very much

2. **How well did the answer scales match the questions?**

1 2 3 4 5

Not at all Somewhat Very much

3. **How well did the measure questions match the mission events unfolding?**



1 2 3 4 5

Not at all Somewhat Very much

4. **Were any specific questions confusing, out of place or inappropriate for the mission exercise?**

- ☐ Yes
- ☐ No

b. If yes, which one(s)?

_____

_____

_____

_____

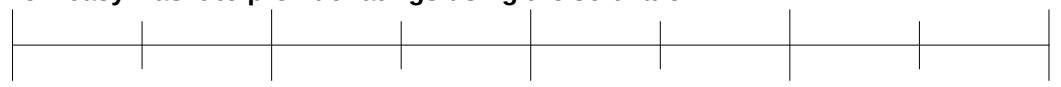
_____

_____

c. Why?

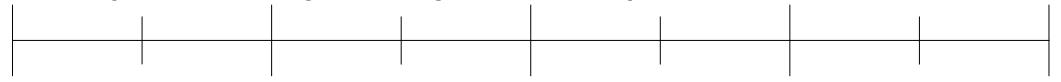
- ☐ Wording was confusing
- ☐ Measure was irrelevant to the mission
- ☐ The scale didn't match the question
- ☐ The measure was out of place – should have been grouped in other part of the mission
- ☐ Air/Ground scheme of maneuver
- ☐ Other (specify):

5. **How easy was it to provide ratings using the *software*?**



1 2 3 4 5
Not at all Somewhat Very much

6. **How easy was it to navigate through the mission phases in the "tree" on the left?**



1 2 3 4 5
Not at all Somewhat Very much

7. **How easy was it to match the questions with the events unfolding in the mission?**

1 2 3 4 5
Not at all Somewhat Very much

8. **Overall, how useful was the device for assessing Soldiers conducting aviation missions?**

1 2 3 4 5
Not at all Somewhat Very much

9. **If given the opportunity, would you use a measurement tool like this in the future to assess your Soldiers?**
   - ☐ Yes
   - ☐ No

   a. Why or Why not?

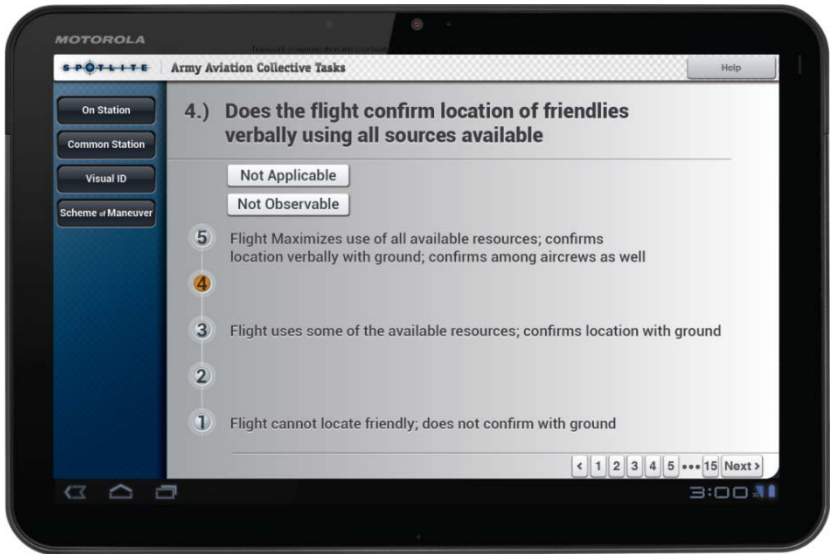
_____
_____
_____
_____
_____
_____
_____
_____
_____
_____

Thank you for your time and assistance

APPENDIX E

MOCK-UPS OF REVISED MEASUREMENT TOOL


MOTOROLA
SPOTLITE
Army Aviation Collective Tasks
Help
Assessments
Create New Assessment
Observer Number:
Observer Unit:
Observer Email:
Observer Phone:
Today's Date:
Mission Start Time:
BN or TF:
CO:
ATX Mission Number:
Mission Type:
Phase:
Position:
Training Objective:
Air Assault
Area Recon
Convoy Security
Deliberate Ops (Cordon and Knock, Air Assault, Grnd Unit Suppo...
Select Type
OK
Cancel
3:00


MOTOROLA
SPOTLITE
Army Aviation Collective Tasks
Help
On Station
Common Station
Visual ID
Scheme of Maneuver
4.) Does the flight confirm location of friendlies verbally using all sources available
Not Applicable
Not Observable
5 Flight Maximizes use of all available resources; confirms location verbally with ground; confirms among aircrews as well
4
3 Flight uses some of the available resources; confirms location with ground
2
1 Flight cannot locate friendly; does not confirm with ground
‹ 1 2 3 4 5 ••• 15 Next ›
3:00


MOTOROLA
SPOTLITE
Army Aviation Collective Tasks
Help
Assessments
Timeline
Training Objectives
3
3
Enroute
On-Station
Target Acquisition
Close
AMC Development
Developing newer pilots' familiarity/skills in this role.
3.5/5
5.0 Is the flight utilizing communication in accordance with the Signal Operations Instructions (SOI) and SOP?
4.0 Does the flight react appropriately to the threat update?
4.0 Does the flight coordinate team tactics internally?
Done
Next
3:00

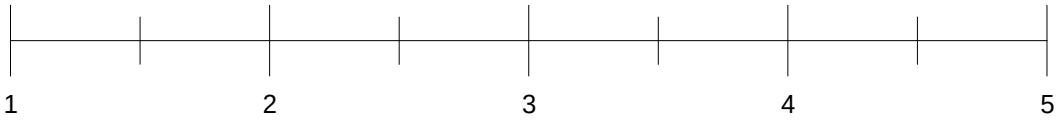
APPENDIX F

REVISED AVIATION COLLECTIVE PERFORMANCE MEASURES

Mission Planning

1.1 Ops Summary

1. Does the flight incorporate the elements of the Operation Summary in their pre-mission planning?



☐ N/A
☐ N/O

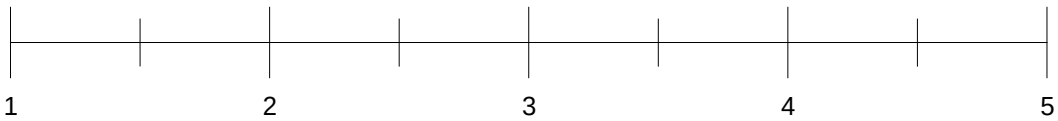
1	2	3	4	5
Flight does not incorporate all information from the brief		Flight incorporates some information from the brief; plan is not fully developed		Flight incorporates all information from the brief; plan is fully developed

   a. If applicable, which required elements were missed?

   - ☐ Fires
   - ☐ Airspace
   - ☐ Signal (Call signs, grids, frequency)
   - ☐ Weather
   - ☐ Air/Ground scheme of maneuver
   - ☐ Timelines
   - ☐ ISR platforms
   - ☐ Last 12-24 hrs
   - ☐ Last 24-72 hrs
   - ☐ Intel analysis of Enemy Course of Action in area of operation (e.g. most likely, most dangerous)
   - ☐ Refine and Update PIR (BOLO)
   - ☐ Terrain Analysis
   - ☐ Other (specify):

1.3.1 Flight Team Brief

2. Does the flight include the required information in the mission brief?



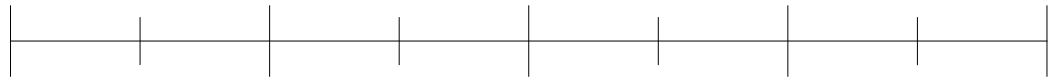
☐ N/A
☐ N/O

1	2	3	4	5
Flight does not brief all required information		Flight briefs all required information, but not all accurate or timely		Flight briefs and discusses all required information in an accurate and timely manner

a. If applicable, which required elements were missed?

- ☐ Fires
- ☐ Airspace
- ☐ Signal (Call signs, grids, frequency)
- ☐ Weather
- ☐ Air/Ground scheme of maneuver
- ☐ Timelines
- ☐ ISR platforms
- ☐ Last 12-24 hrs
- ☐ Last 24-72 hrs
- ☐ Intel analysis of Enemy Course of Action in area of operation (e.g. most likely, most dangerous)
- ☐ Refine and Update PIR (BOLO)
- ☐ Terrain Analysis
- ☐ IMC Breakout
- ☐ Clearance of Fires
- ☐ Other (specify):

3. Does the flight discuss and designate roles for the mission?



☐ N/A
☐ N/O

1 — AMC does not assign roles
2
3 — AMC assigns roles per SOP
4
5 — AMC assigns and flight discusses roles per SOP

1.3.3 Follow SOP

4. Does the aircrew follow the aircrew brief checklist in accordance with SOP?

☐ Yes ☐ N/A
☐ No ☐ N/O

a. If no, what was missed?

- ☐ Mission Overview and Flight Plan
- ☐ Crew actions, duties, and responsibilities
- ☐ Emergency Actions and Downed Aircraft Procedures

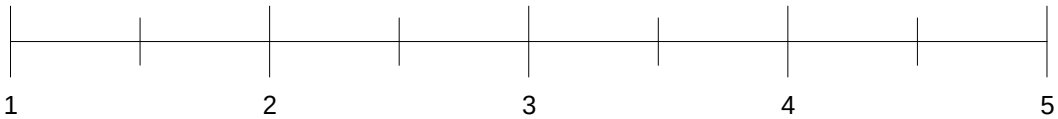
- Downed Aircraft Procedures
- Analysis of the Aircraft (logbook, maintenance, PPC)
- SPINS
- Fighter Management and Risk Mitigation
- Other (specify)

2.1 Airspace Deconfliction

5. Does the flight develop the appropriate flight deconfliction measures?
   - Yes
   - No
   - N/A
   - N/O

2.2.3 Mission Intel

6. Were mission intel products supplied/requested?



- N/A
- N/O

1	2	3	4	5
No requests for any updates were made		Products requests submitted; products generated independently using organic resources		Products received from support unit as well as updates throughout planning cycle

a. If yes, what updates were reviewed?
   - UAS live feeds
   - Imagery of the Target
   - Imagery of the Area (terrain, man-made objects)_
   - Descriptions of the Target
   - Other (specify):

2.2.4 Friendly Situation

7. Does the flight ensure it is aware of changes to the friendly situation?

- N/A
- N/O

1	2	3	4	5
Flight does not receive or ask for updates		Flight receives updates on friendly situation		Flight proactively requests updates to maintain situational awareness

2.2.5 Call Signs & Freq

8. Does the flight verify they have to call signs and frequencies for the mission?

☐ N/A
☐ N/O

1	2	3	4	5
Flight relies on previously developed comm card		Flight verifies Call Signs and Freqs with TOC		Flight verifies with TOC, checks against current comm card, and ensures all team members have correct info

2.2.7 Grid Locations

9. Does the flight ensure accurate grid location for friendlies?

☐ N/A
☐ N/O

1	2	3	4	5
Flight does not obtain accurate grid locations		Flight obtains grid location information		Flight obtains grid locations, enters into aircraft systems, and verifies all aircrews have the most recent information

2.2.8 Threat Update

10. Does the flight request a threat update?

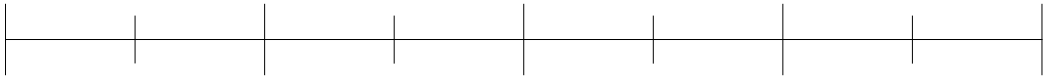
☐ Yes ☐ N/A
☐ No ☐ N/O

11. If required, does the flight request additional information based on content of threat update?

☐ N/A
☐ N/O

1	2	3	4	5
Flight does not request additional information		Flight requests some additional information		Flight requests action to obtain all information required to fill SA

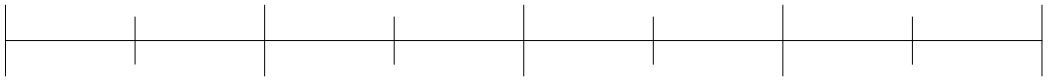
12. Does the flight change their plan based on updates to the threat/enemy that affect their mission/safety?



☐ N/A
☐ N/O

1	2	3	4	5
Flight executes pre-planned mission regardless of threat updates		Flight briefly discusses adjustments to their plan; decides to create hasty plan in flight		Flight creates a thorough plan, prior to launch, to mitigate threat risk

13. Does the flight develop and/or adjust their mission plan according to information provided in the pre-mission brief and WARNO?



☐ N/A
☐ N/O

1	2	3	4	5
Flight does not incorporate all information from the brief		Flight incorporates some information from the brief; plan is not fully developed		Flight incorporates all information from the brief; plan is fully developed

a. If applicable, which required elements were missed?

- ☐ Fires
- ☐ Airspace
- ☐ Signal (Call signs, grids, frequency)
- ☐ Weather
- ☐ Air/Ground scheme of maneuver
- ☐ Timelines
- ☐ ISR platforms
- ☐ Last 12-24 hrs
- ☐ Last 24-48 hrs
- ☐ Intel analysis of Enemy Course of Action in area of operation (e.g. most likely, most dangerous)
- ☐ Refine and Update PIR (BOLO)
- ☐ Terrain Analysis
- ☐ Other (specify):

Final Mission Brief

14. Does the Flight adjust their plan according to differences between WARNO and Final Mission Brief

☐ N/A
☐ N/O

1	2	3	4	5
Flight does not review current plan for changes		Flight acknowledges changes and discusses impact on their plan		Flight acknowledges changes and makes formal changes to their plan

3.1 Report changes

15. Does the AMC request mission updates from TOC prior to launch?

☐ N/A
☐ N/O

1	2	3	4	5
AMC does not request changes		AMC requests update		AMC request update; verifies update with wingman

3.2 SITREP

16. Does the AMC request SITREP from all appropriate resources prior to launch?

☐ N/A
☐ N/O

1	2	3	4	5
AMC does not check in with resources		AMC checks in with some resources (e.g. UAS, ground, BAE)		AMC checks in with all resources (e.g. UAS, ground, BAE)

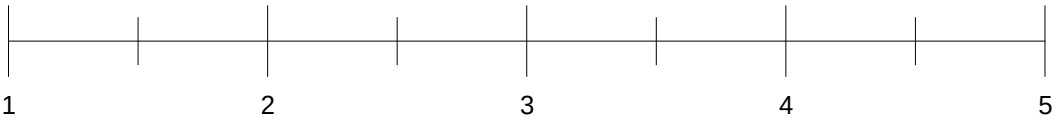
2 Enroute

Launch

17. Did the Flight Team launch on Time?

☐ Yes ☐ N/A
☐ No ☐ N/O

a. If no, was the delay communicated to the TOC?



□ N/A
□ N/O

1	2	3	4	5
No communication of delay to the TOC		Communicates delay to TOC; no indicates of new launch time		Immediately alerts the TOC will not make launch time; explains reason for delay; recommends new launch time; provides updates as necessary

3.4 Call Off to TOC

18. Does the aircrew commander successfully call off to Battalion TOC?

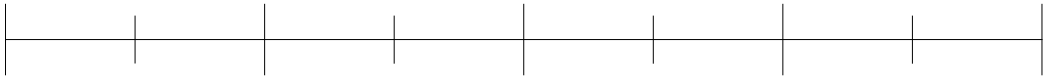
□ Yes □ N/A
□ No □ N/O

a. Does the flight conduct battle checks (WAILR-M)?

□ Yes □ N/A
□ No □ N/O

4.1 Deconflict airspace

19. Does the flight deconflict the airspace?



□ N/A
□ N/O

1	2	3	4	5
Flight does not use information sources to anticipate events; does not make radio calls		Flight uses information sources to anticipate events but does not push information to rest of team		Flight uses information sources to anticipate events and proactively pushes information to rest of team

4.2 Monitor Updates

20. Does the flight monitor air to air radio communication?

□ N/A
□ N/O

1	2	3	4	5
Flight does not monitor or acknowledge air to air communications		Flight monitors and disseminates air to air communications		Flight monitors air to air communications; discusses and addresses impacts to mission

21. Does the flight monitor ground channels?

☐ N/A
☐ N/O

1	2	3	4	5
Flight does not monitor updates		Flight monitors and disseminates updates		Flight monitors updates; discusses and addresses impacts to mission

4.3 Coordinate Team Tactics

22. Does the flight select holding area?

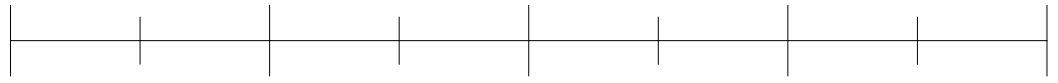
☐ N/A
☐ N/O

1	2	3	4	5
Flight selects area without considering tactical implications (e.g., too small, indefensible, cannot communicate)		Flight selects an area; area is suitable but not optimal		Flight selects an area; considers size, comms, defensibility, etc. to identify optimal location

a. If applicable, what tactical implications were missed?

- ☐ Concealment
- ☐ Obstacles
- ☐ Key terrain
- ☐ Approach and departure directions
- ☐ 360° Security
- ☐ Other (specify):

23. Does the flight select loiter area?



☐ N/A
☐ N/O

1	2	3	4	5
Flight selects area without considering tactical implications (e.g., too small, indefensible, cannot communicate)		Flight selects an area; area is suitable but not optimal		Flight selects an area; considers size, comms, defensibility, etc. to identify optimal location

a. If applicable, what tactical implications were missed?

- □ Size
- □ Suitable location
- □ Communication availability
- □ Altitude for loiter
- □ Pattern of loiter
- □ Time to target
- □ Other (specify):

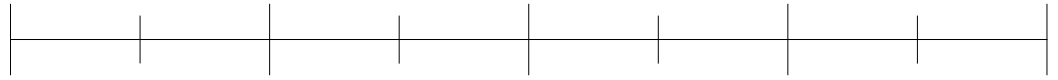
24. Does the flight delegate and coordinate flight related duties (e.g., communication) in response to as changes in the current situation occur?

□ N/A
□ N/O

1	2	3	4	5
Flight does not have a plan; confusion over assigned tasks		Flight continues with assigned flight duties; adapts to current situation		Flight reassigns duties to suit situation

4.4 Adherence to SOP

25. Does the flight adhere to requirements given in the mission briefs?



□ N/A
□ N/O

1	2	3	4	5
Flight does not follow team brief; does not communicate deviation from SOP		Flight follows team brief and adapts within its constraints		Flight understands team brief and deviates when necessary; communicates deviation for flight internal SA

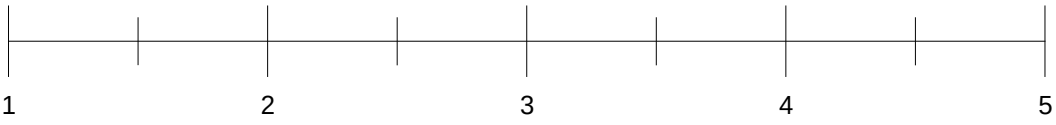
26. If a deviation was required, did the flight appropriately deviate from the mission brief?

□ N/A
□ N/O

1	2	3	4	5
Flight did not recognize need for deviation; did not deviate from the team brief		Flight recognized need for deviation; did not appropriately deviate from the team brief		Flight recognized need for deviation; deviated appropriately from the team brief

4.4.5 Tactics

27. Does the flight develop appropriate tactics if there is misalignment between SOP and situation?



☐ N/A
☐ N/O

1	2	3	4	5
Flight does not recognize the misalignment; makes no changes to SOP or tactics		Flight recognizes misalignment; tries to fit the SOP to the situation		Flight recognizes misalignment; appropriately modifies the tactics based on the environment

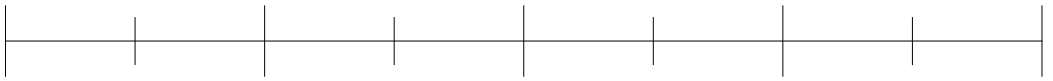
4.4.1 Formation

28. Does the flight continue to discuss vertical and lateral displacement, tactics, and protection of aircraft while in flight?

☐ N/A
☐ N/O

1	2	3	4	5
No detailed plan or discussion		Defines all measures but does not discuss adjustments		Defines all measures and discusses required adjustments

29. Does the flight adhere to the flight formation as briefed?



☐ N/A
☐ N/O

1	2	3	4	5
Flight maintains loose formation; no discussion on the formation or changes to the formation		Flight sets formation based on briefed tactics		Flight sets formation based on briefed tactics; constantly updates formation based on current situation

4.4.2 Flight Duties

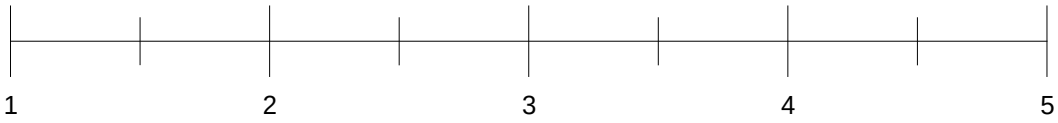
30. Does the flight adhere to the flight duties required for the mission?

☐ N/A
☐ N/O

1	2	3	4	5
Flight does not adhere to flight duties		Flight adheres to flight duties; no further coordination or backup		Flight adheres to flight duties; adjusts to mitigate task overload and based on current situation

4.4.3 Communication Protocol

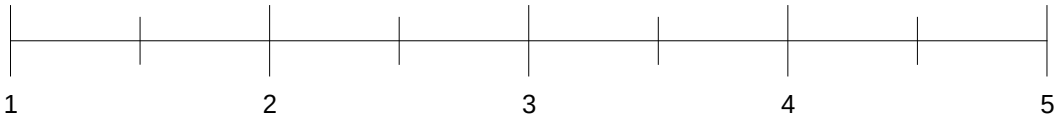
31. Does the flight follow appropriate communication protocol?



☐ N/A
☐ N/O

1	2	3	4	5
Flight uses some available communication tools (e.g., data, voice); does not use pro-words or brevity codes as required; stepovers occur		Flight uses some available communication tools (e.g., data, voice); uses pro-words and brevity codes		Flight uses appropriate communication tools (e.g., data, voice); uses pro-words and brevity codes correctly; messages contain appropriate level of detail

32. Does the flight use proper theater aircrew procedures guide (APG) throughout flight?

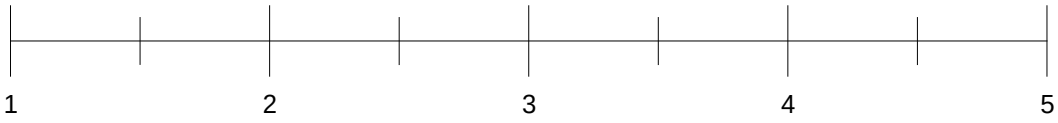


☐ N/A
☐ N/O

1	2	3	4	5
Flight does not follow the proper APG communications guidance		Flight follows the proper APG communications guidance		Flight follows the proper APG communication guidance; adjust flight path based on updates

4.5.1 Check-in with Ground

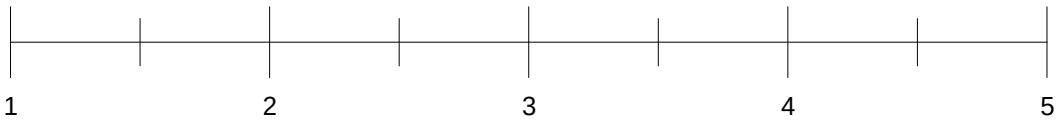
33. When does the flight make the check-in call to Ground?



☐ N/A
☐ N/O

1	2	3	4	5
On arrival		Before they arrive in the AO		Immediately upon radio communication range

34. Does the flight provide the required information as they check-in with ground?



☐ N/A
☐ N/O

1	2	3	4	5
Flight does not know proper format; misses multiple items		Flight checks in with all appropriate items; check-in not done in correct order		Flight checks in with all appropriate items; check-in done in correct order

a. Which items were missed?

- □ Call sign
- □ Type and Number of Aircraft
- □ Type and Number of Weapons System Available
- □ Station Time
- □ Request SITREP

b. Did the ground acknowledge aircraft?

□ Yes □ N/A
□ No □ N/O

c. Did the ground send SITREP to flight?

□ Yes □ N/A
□ No □ N/O

4.5.2 Receive SITREP

35. Does the flight receive the SITREP from Ground?

□ N/A
□ N/O

1	2	3	4	5
Flight does not receive SITREP; does not follow up with request		Flight receives SITREP		Flight receives SITREP; verifies completeness and requests information pertinent to mission

4.5.2.4 Obtain UAS feed

36. If UAS feed was available, when does the flight request it?

□ N/A
□ N/O

1	2	3	4	5
Flight does not request		Flight requests, but not as soon as possible		Flight requests as soon as possible in mission thread

3 On Station

5.1 Arrive On Station

37. Does the flight arrive on-station on time?

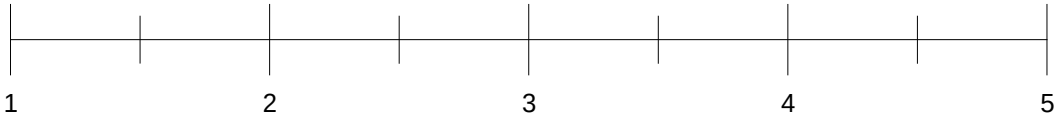
☐ Yes ☐ N/A
☐ No ☐ N/O

38. Does the flight communicate on-station arrival to the supported Battlespace owner?

☐ Yes ☐ N/A
☐ No ☐ N/O

4.3.2 Deconfliction Measures

39. Does the flight verify the airspace is clear or free from obstacles (e.g. helicopters, fixed wing, UAS, artillery)?



☐ N/A
☐ N/O

1 2 3 4 5

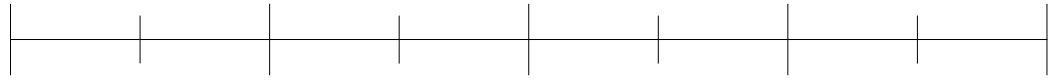
Flight does not verify airspace deconfliction

Flight reviews ROZ information

Flight reviews ROZ information; makes final call to verify prior to entering airspace

5.2 Location of Friendlies

40. Does the flight confirm location of friendlies using all sources (Visually, BFT, Sensors) available?



☐ N/A
☐ N/O

1 2 3 4 5

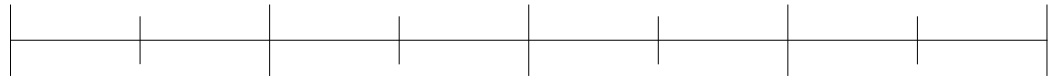
Flight cannot locate friendlies; does not confirm with ground

Flight uses some of the available resources; confirms location with ground

Flight maximizes use of all available resources; confirms location verbally with ground

5.3 Develop Plan/Scheme of Maneuver

41. Does the flight work with the ground to establish task and purpose for their mission?



☐ N/A
☐ N/O

1 2 3 4 5

Flight does not communicate with ground to establish task and purpose

Flight communicates with ground to establish task and purpose

Flight collaborates with ground to meet commander's intent

5.3.2 Clearance of Fires Authority.

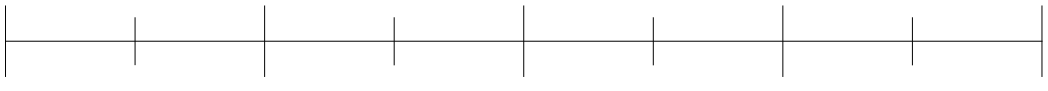
42. Does the flight establish who has clearance of fires authority?

☐ Yes ☐ N/A

☐ No ☐ N/O

5.3.3 Shooter Duties

43. Do the aircrews coordinate and designate shooter duties within the flight?



☐ N/A

☐ N/O

1	2	3	4	5
Aircrews do not have an established plan for designating shooter; still discussing roles.		Aircrews have a plan; assigns duties according to plan		Aircrews have a plan; continuously updates plan and assigns or reassigns duties based on updates to the situation

5.3.4 Discuss Plan Within

44. Does the flight discuss applicable changes to the tactical mission?

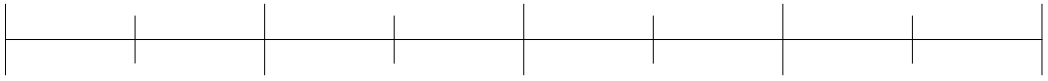
☐ N/A

☐ N/O

1	2	3	4	5
Flight does not discuss		Flight discusses; considers changes if required but tries to force initial plan		Flight discusses; makes changes if required

5.3.5 Recommend COA to Ground

45. Does the flight recommend course of action to Ground Commander?



☐ N/A

☐ N/O

1	2	3	4	5
Flight does not recommend course of action; moves forward with own plan		Flight recommends course of action before discussing with ground		Flight establishes course of action based on discussion with ground over execution of task and purpose

5.4 Provide Security Per SOP

46. Does the flight maintain security posture based on MET-TC throughout mission?

☐ N/A
☐ N/O

1	2	3	4	5
Flight does not establish or follow security plan		Flight maintains security using assigned duties		Flight maintains security using assigned duties; constant communication between aircraft to maintain a high level of security

5.5 Develop the Situation

47. Does the flight continue to develop the situation with ground?

☐ N/A
☐ N/O

1	2	3	4	5
Flight continues the mission using only the information provided in check-in with ground and their recommended COA		Flight continues to observes situation and passes information to ground		Flight continues to observe the situation; constant dialog with ground to develop situation

5.5.1 ISR Data

48. If ISR (e.g., CAS, UAS) data is available, does the flight communicate information to ground forces?

☐ N/A
☐ N/O

1	2	3	4	5
Flight does not offer information to ground		Flight shares some information with ground when asked		Flight pushes all relevant and available information to ground

49. When AMC needs to maneuver UAS (non-MUM) for mission accomplishment, does the AMC establish UAS control authority?

☐ N/A
☐ N/O

1	2	3	4	5
Flight does not establish relationship		Flight establishes relationship		Flight establishes relationship; works to develop the situation

5.6 Pattern of Life

50. Does the flight communicate observed differences in pattern of life?

☐ N/A
☐ N/O

1	2	3	4	5
Flight doesn't recognize or communicate differences		Flight recognizes and communicates differences		Flight recognizes and communicates change; discusses courses of action based on differences

4 Target Acquisition

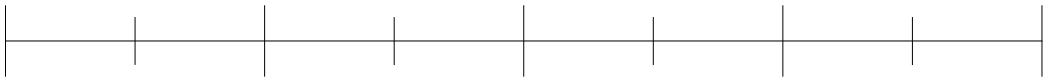
6.1 Locate Target

51. Does the flight communicate with ground to locate the target?

☐ N/A
☐ N/O

1	2	3	4	5
Flight uses last SITREP and begins search; does not coordinate further with ground		Flight asks ground for target location		Flight maintains dialogue with ground to identify targets

52. Does the flight incorporate an ISR plan?



☐ N/A
☐ N/O

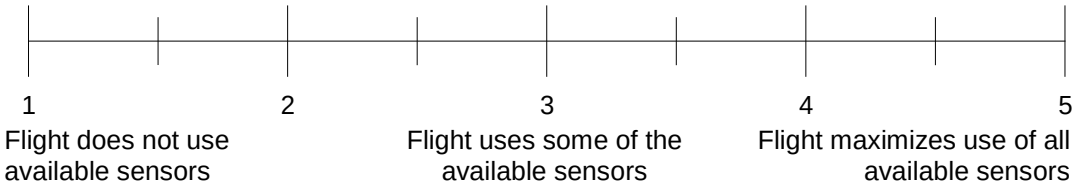
1	2	3	4	5
Flight does not use all assets available; does not distribute areas of observation		Flight coordinates use of sensors and available assets; distributes areas of observation		Flight coordinates use of sensors and available assets; distributes areas of observation; factors in METT-TC

53. Does the flight actively search for the target?

☐ N/A
☐ N/O

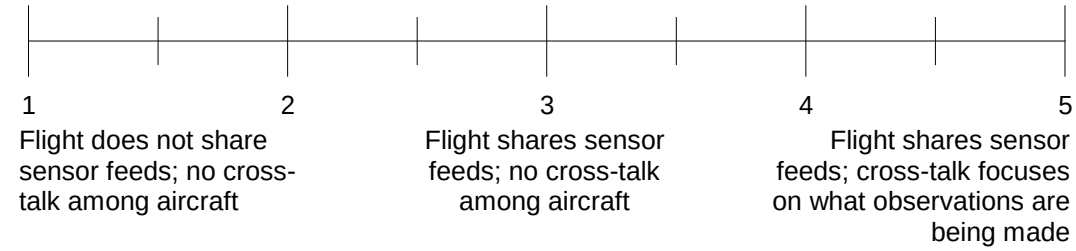
1	2	3	4	5
Flight does not have a plan in place for search; does not distribute areas of observation		Flight executes an ISR plan; does not effectively distribute areas of observation		Flight executes an ISR plan that is adapted to search environment

54. Does the flight use the appropriate sensors to search for targets?



☐ N/A
☐ N/O

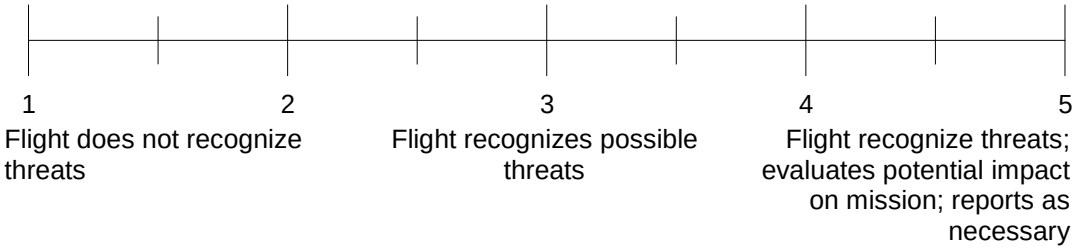
55. Does the flight share sensor feeds among aircrews?



☐ N/A
☐ N/O

6.2.4 Recognize threats

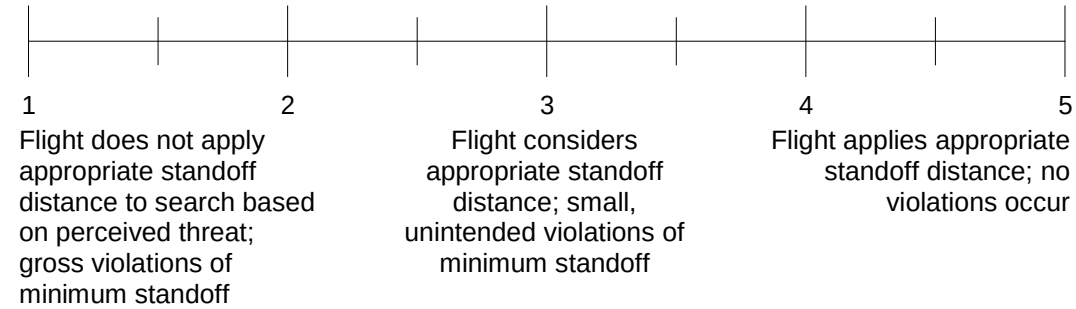
56. Does the flight effectively recognize threats?



☐ N/A
☐ N/O

6.2.5 Standoff Distance

57. Does the flight utilize an appropriate standoff distance?



☐ N/A
☐ N/O

6.3 Announce Target in Sight

Does the flight use correct pro-words during target identification?

1 2 3 4 5

Flight does not use correct terms; not clear, concise, descriptive

Flight uses correct terms with a few errors; clear, and concise

Flight uses correct terms; clear, concise, and descriptive

☐ N/A
☐ N/O

58. Does the aircrew announce (within the cockpit) that the target is in sight?

1 2 3 4 5

Aircrew does not communicate detection of target to other aircraft

Aircrew announces detection

Aircrew announces detection; pushes target information

☐ N/A
☐ N/O

59. Does the flight use correct pro-words to announce target in sight to the team?

1 2 3 4 5

Flight does not use correct terms; not clear, concise, descriptive

Flight uses correct terms with a few errors; clear, and concise

Flight uses correct terms; clear, concise, and descriptive

☐ N/A
☐ N/O

6.3.1 Wingman Confirm

60. Does the wingman confirm target detection?

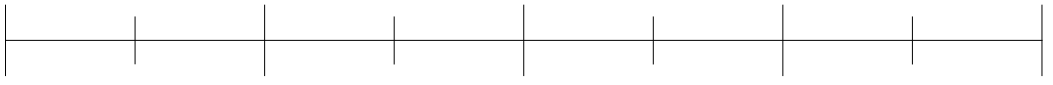
1 2 3 4 5

Wingman does not acknowledge detection of the target

Wingman acknowledges detection of the target

Wingman establishes PID

☐ N/A
☐ N/O



1 2 3 4 5

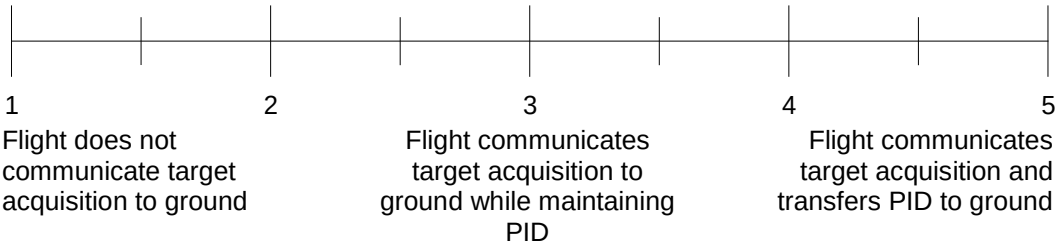
Wingman does not use correct terms; not clear, concise, descriptive

Wingman uses correct terms with a few errors; clear, and concise

Wingman uses correct terms; clear, concise, and descriptive

☐ N/A
☐ N/O

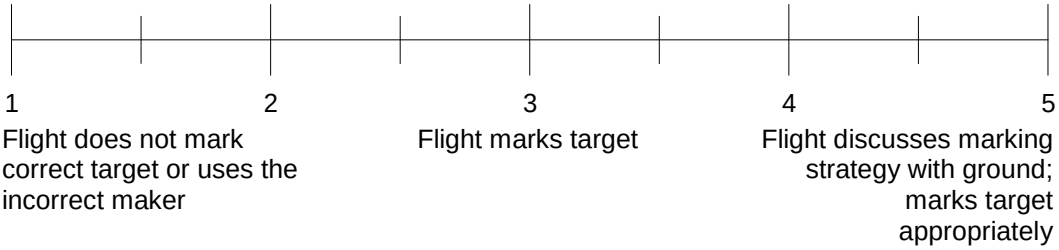
61. Does the flight confirm target acquisition to ground?



☐ N/A
☐ N/O

6.5 Confirm Target

62. Does the flight mark the target to confirm its location?

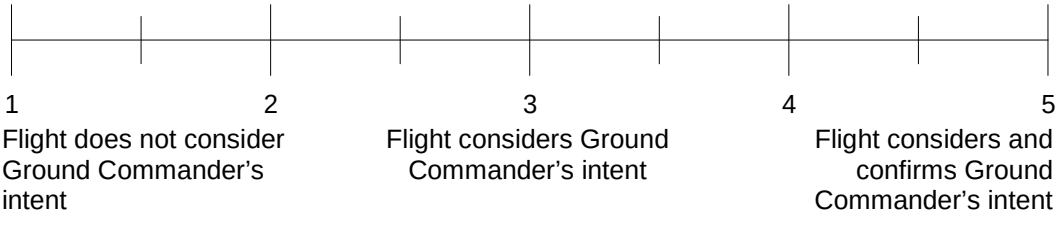


☐ N/A
☐ N/O

ROE/Clearance of Fires

7.1 Confirm Ground Commander's Intent

63. Does the flight consider Ground Commander's intent?



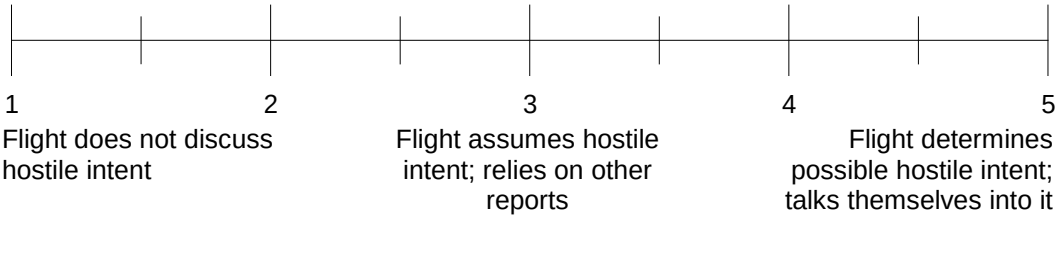
☐ N/A
☐ N/O

Confirm hostile intent

64. Does the flight confirm hostile intent prior to applying lethal force?

☐ Yes ☐ N/A
☐ No ☐ N/O

a. If no, why not?



☐ N/A
☐ N/O

7.2 Discuss Lethal/Nonlethal COAs

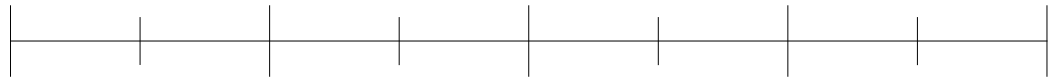
65. Does the flight discuss lethal and nonlethal COAs with Ground Commander?

☐ N/A
☐ N/O

1	2	3	4	5
Flight does not know commander's intent; acts without considering commander's intent		Flight applies commander's intent		Flight applies commander's intent; communicates alternative COAs based on aerial perspective

7.3 Discuss proportionality

66. Does the flight discuss proportionality?



☐ N/A
☐ N/O

1	2	3	4	5
Flight does not discuss proportionality		Flight discusses proportionality		Flight uses proportionality considerations to select the most appropriate weapon

7.3.2 Weapon Choice

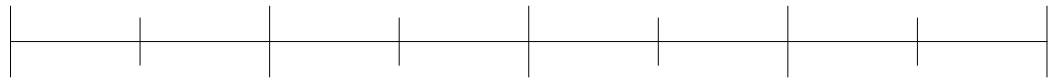
67. Does the flight choose the right weapon and coordinate fires?

☐ N/A
☐ N/O

1	2	3	4	5
Flight misappropriates weapons; does not coordinate fires within the flight		Flight selects appropriate weapons system; coordinate fires within the flight		Flight selects appropriate weapons system; coordinates fires within flight and with ground

7.3.3 Engagement Scheme of Maneuver

68. Does the flight coordinate an engagement scheme of maneuver?



☐ N/A
☐ N/O

1	2	3	4	5
Flight does not coordinate a scheme of maneuver		Flight coordinates a scheme of maneuver applicable for the situation		Flight coordinates a scheme of maneuver optimized for the situation

7.4 Discuss collateral damage

69. Does the flight consider collateral damage?

☐ N/A
☐ N/O

1	2	3	4	5
Flight does not consider collateral damage		Flight discusses collateral damage		Flight uses collateral damage considerations to select method of engagement and weapon system

7.5 Shoot/Don't Shoot

70. Does the flight make an appropriate shoot/don't shoot decision (e.g. considers commander's intent; hostile intent; collateral damage)?

☐ N/A
☐ N/O

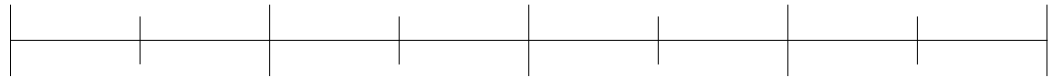
1	2	3	4	5
Flight does not consider critical variables before making shoot/don't shoot decision		Flight considers critical variables before making shoot/don't shoot decision		Flight considers critical variables before making shoot/don't shoot decision; develops alterative COAs if don't shoot is determined

71. Does the flight communicate shoot/don't shoot decision to ground?

☐ N/A
☐ N/O

1	2	3	4	5
Flight does not communicate decision to ground		Flight communicates decision to ground; does not provide alternative COAs		Flight communicates decision to ground; provides alternative COAs

72. Does the flight continue to observe the target after don't shoot decision is made?



☐ N/A
☐ N/O

1	2	3	4	5
Flight goes off station after don't shoot decision is made		Flight remains on station and continues to observe target		Flight remains on station and continues to develop the situation

Clearance of Fires

8.1 Request Clearance of Fires

73. Does the flight request clearance of fires from Ground Commander?

☐ N/A
☐ N/O

1	2	3	4	5
Flight does not request clearance of fires		Flight considers ROE; establishes friendly/enemy positions; requests clearance of fires; not ready to effect the target while going through this process		Flight considers ROE; establishes friendly/enemy positions; requests clearance of fires; anticipates clearance and sets up shot during this process

8.1.1.1 Cleared Hot

74. Does the flight receive acknowledgement of clearance of fires from ground prior to engagement?

☐ Yes ☐ N/A
☐ No ☐ N/O

75. Does the AMC communicate weapons release clearance within the flight?

☐ N/A
☐ N/O

1	2	3	4	5
AMC does not communicate weapons release clearance to the rest of the flight		AMC communicates cleared hot, but does not confirm acknowledgement from rest of the flight		AMC communicates and confirms cleared hot from rest of the flight

9 Employ Weapon System

9.1 Fire weapon based on Plan.

76. Does the flight establish inbound heading and formation in accordance with briefed tactics?

☐ N/A
☐ N/O

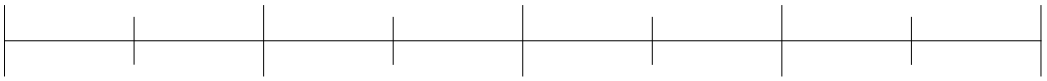
1	2	3	4	5
Flight does not communicate, does not follow briefed inbound heading and formation		Flight communicates follows briefed inbound heading and formation		Flight communicates and follows briefed inbound heading and formation; trail effectively covers lead

77. Does the wingman provide overwatch and cover?

☐ N/A
☐ N/O

1	2	3	4	5
Wingman does not provide overwatch and cover; fixates on target		Wingman provides overwatch and cover; uses appropriate resources to aid in overwatch		Wingman provides overwatch and cover; uses appropriate resources to aid in overwatch; communicates developing situation with lead aircraft

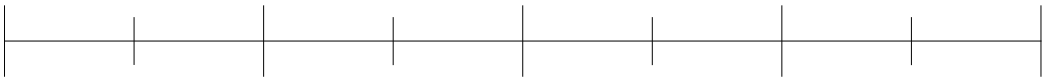
78. Does the flight apply appropriate weapons engagement technique based on threat environment (METT-TC)?



☐ N/A
☐ N/O

1	2	3	4	5
Flight does not apply appropriate technique; target not prosecuted		Flight does not apply appropriate technique; target still prosecuted		Flight applies appropriate technique; target prosecuted

79. Does the flight communicate appropriately throughout the engagement ?

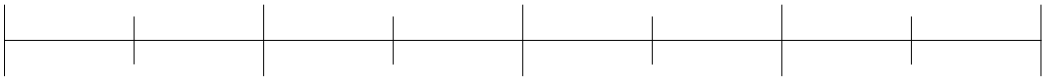


☐ N/A
☐ N/O

1	2	3	4	5
Aircrews employ inappropriate radio chatter		Aircrews employ clear, concise, and timely radio calls		Aircrews employ clear, concise, and timely radio calls; acknowledgements made in a timely manner

9.2 Weapon Effects

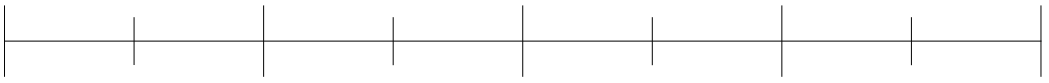
80. Does the flight determine effects of weapons and meeting of engagement objectives?



☐ N/A
☐ N/O

1	2	3	4	5
Flight does not determine weapons effects; does not communicate effects		Flight determines effects of weapons; communicates within flight		Flight determines effects of weapons; communicates within flight; determines next COA

81. Does the flight communicate weapons effect to ground?



☐ N/A
☐ N/O

1	2	3	4	5
Flight does not communicate weapons effect		Flight communicates weapons effect; not clear and descriptive		Flight communicates weapons effect; clear, concise, and descriptive

9.3 Health State of Aircraft

82. If aircrews took fire or if aircraft is damaged or has a warning or caution light, do the aircrews choose an appropriate course of action?

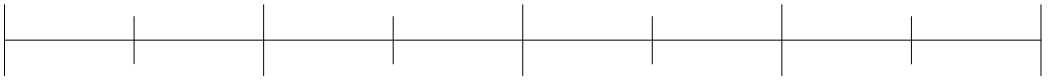
☐ N/A
☐ N/O

1	2	3	4	5
Aircrews do not consider health status of aircraft		Aircrews check warning and caution lights; determine go/no-go		Aircrews check warning and caution lights; perform visual inspections of others' aircrafts; discuss observed damage locations; determine go/no-go

a. Which aircrews did not determine health status?

	Present	Checked health
Lead aircraft	☐	☐
Trail aircraft	☐	☐
Other (text box)	☐	☐
Other (text box)	☐	☐

83. If no-go, do the aircrews choose an appropriate course of action?



☐ N/A
☐ N/O

1	2	3	4	5
Aircrews react to no-go decision without consideration for threat, other aircrews, etc.; do not coordinate with others		Aircrews determine proper course of action in response to no-go decision; coordinate with ground and higher aviation		Aircrews follow predetermined contingency plan; coordinate with ground and higher aviation

BDA & Follow-on Mission

10.1 Give BDA to Ground Commander

84. Does the flight conduct a battle damage assessment?

☐ N/A
☐ N/O

1	2	3	4	5
Flight does not conduct BDA; assumes target is destroyed without verification		Flight evaluates target; reports BDA to ground and command elements		Flight evaluates target; proactively reports complete BDA in proper protocol pushes/reports BDA to ground and appropriate command elements

a. Which required elements were missed?

- ☐ Supported unit (ground)
- ☐ Air Element TOC
- ☐ Other (specify):

b. What BDA items were missed?

- ☐ Sending to right place
- ☐ Alpha, Call sign of observing source
- ☐ Bravo, location of target
- ☐ Charlie, time strike started and ended
- ☐ Delta, percentage of target coverage
- ☐ Echo: itemized destruction

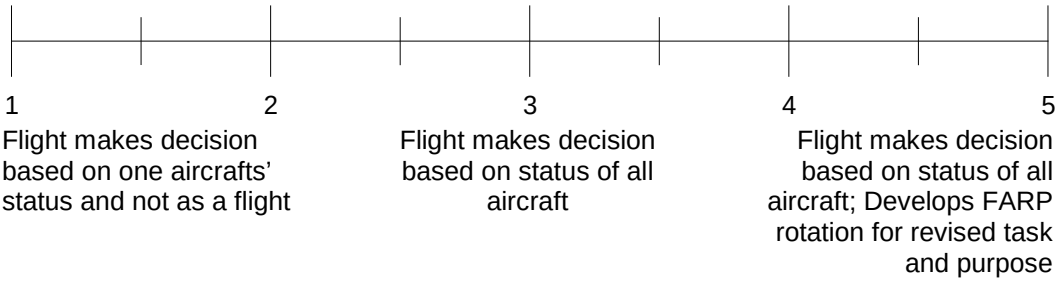
11.1 FARM (Fuel Ammo Rockets Missiles)

85. Do the aircrews discuss FARM (Fuel, Ammo, Rockets, Missiles)?

☐ N/A
☐ N/O

1	2	3	4	5
Aircrews do not determine FARM; assume appropriate stores		Aircrews make accurate assessment of FARM		Aircrews make accurate assessment of FARM; discuss capabilities based on FARM

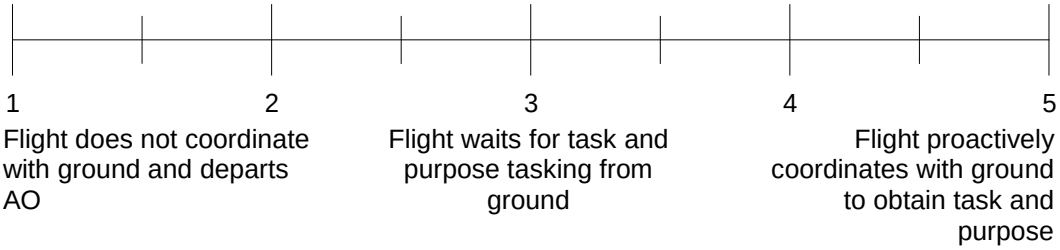
86. Does the flight advise ground of remaining mission time and capabilities based on FARM?



☐ N/A
☐ N/O

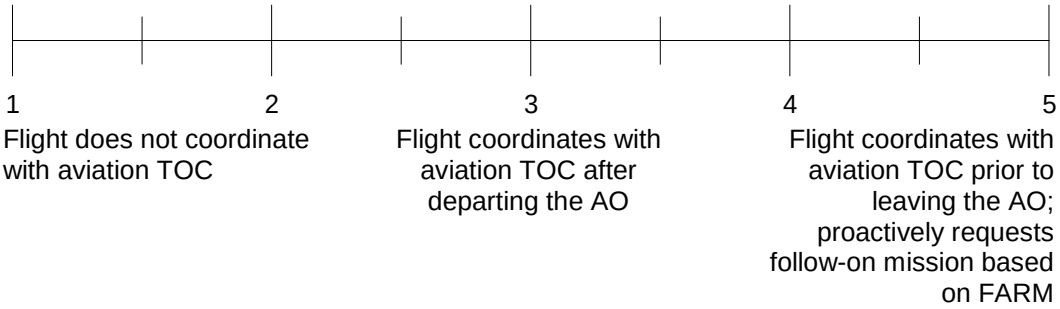
11.2 Obtain Next Mission

87. Does the flight coordinate with ground for follow-on tasking or mission complete?



☐ N/A
☐ N/O

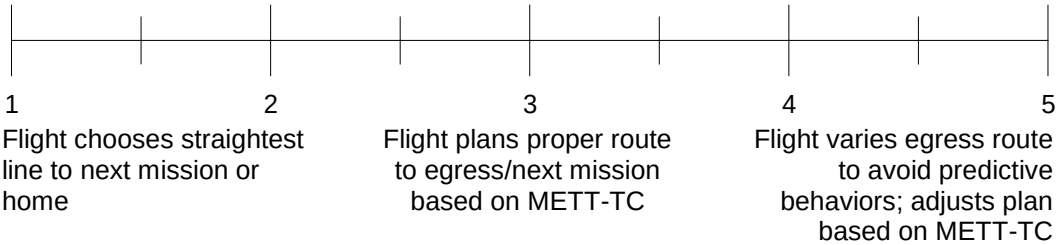
88. Does the flight coordinate with aviation TOC after being released from ground?



☐ N/A
☐ N/O

11.3 Egress Per Unit SOP and APG

89. Does the flight tactically egress from the AO?



☐ N/A
☐ N/O

Post Mission

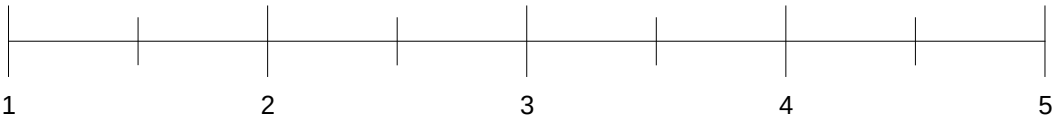
12.1 Post Flight Mission Tasks

90. Does the flight log down with aviation TOC?

□ Yes □ N/A

□ No □ N/O

91. Does the flight conduct post flight mission tasks per SOP?



1 2 3 4 5

Flight does not conduct post flight mission tasks

Flight follows tasks according to SOP as a group

Flight divides tasking among team members; follows tasks according to SOP

□ N/A

□ N/O

92. Do the aircrews conduct post flight maintenance and tie down?

1 2 3 4 5

Aircrews skip multiple steps; skips maintenance and tie down altogether

Aircrews conduct post flight maintenance and tie down

Aircrews conduct a thorough post flight maintenance and tie down; reports to maintenance personnel

□ N/A

□ N/O

12.2 Conduct Debrief

93. Does the flight conduct debrief in accordance with unit SOP?

1 2 3 4 5

Flight provides a limited debrief; numerous errors and omissions

Flight provides review of the mission and minimally debriefs

Flight provides clear, concise, and complete review; reports additional observations not related to the mission

□ N/A

□ N/O

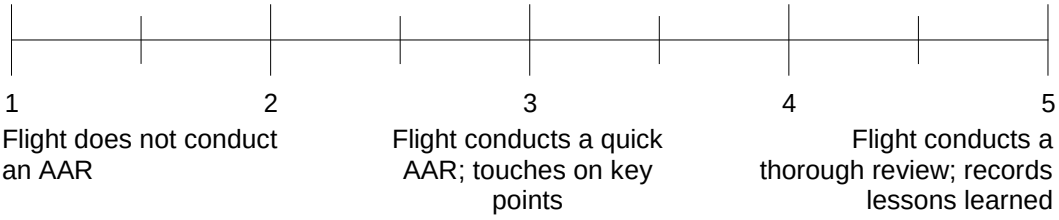
94. Does the flight provide input to the storyboard?

□ Yes □ N/A

□ No □ N/O

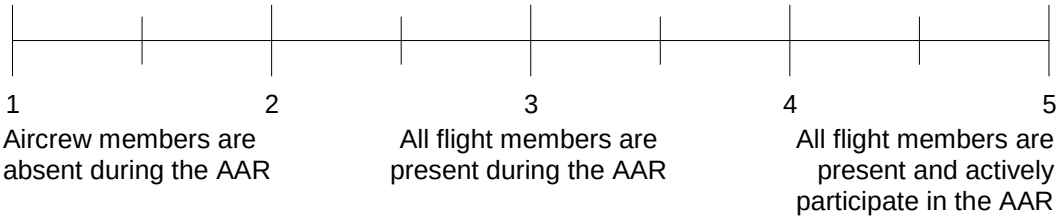
12.3 Conduct AAR

95. Does the flight conduct an AAR?



☐ N/A
☐ N/O

96. Do all crew members participate in the AAR?



☐ N/A
☐ N/O