

A Systems Analysis of the Army's Tactical Evaluation Process

by

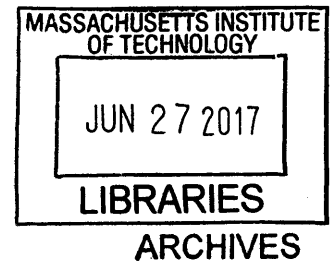
John D. Caddell

B.S. Systems Management United States Military Academy, 2008

SUBMITTED TO THE SYSTEM DESIGN AND MANAGEMENT PROGRAM IN PARTIAL  
FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF ENGINEERING IN SYSTEM DESIGN AND MANAGEMENT AT  
THE MASSACHUSETTS INSTITUTE OF TECHNOLOGY

[Master of Science] JUNE 2017



©2017 John D. Caddell. All rights reserved.

The author hereby grants to MIT permission to reproduce and to distribute publicly paper and electronic copies of this thesis document in whole or in part in any medium now known or hereafter created.

Signature Redacted

Signature of Author: \_\_\_\_\_  
System Design and Management Program  
May 5, 2017

Signature Redacted

Certified by: \_\_\_\_\_  
Bryan R. Moser  
Lead Instructor, Systems Design and Management Program  
Thesis Supervisor

Signature Redacted

Accepted by: \_\_\_\_\_  
Joan S. Rubin  
Executive Director, System Design and Management Program



# A Systems Analysis of the Army's Tactical Evaluation Process

by

John D. Caddell

Submitted to the System Design and Management Program on May 5, 2017 in Partial Fulfillment of the Requirements for the Degree of Master of Engineering in System Design and Management

## **Abstract**

In preparation to fight and win the nation's wars, the United States Army spends a lot of time and money on tactical training. While few doubt that the Army produces some of the world's finest tactical units, there is still room for improvement. The Army's current tactical evaluation process utilizes a series of subjective and uncalibrated metrics that describe the capability of military units. Leaders fail to see the value in the provided evaluation forms and formats and have begun developing alternative methods to evaluate and provide feedback to their units. The abandonment of a standard evaluation process has added to the inability to objectively or accurately compare units across formations and time. The current evaluation system depends almost entirely on one set of humans to monitor safety, performance, and provide feedback to the training unit for improvement. Humans in this system fatigue and fall prey to natural pressures to reduce standards and push units through training exercises regardless of undesirable performance. Additionally, internal pressure, organizational structure, and competition encourage units to exaggerate or inflate their tactical readiness reports. While this situation is widely recognized in the Army, the current evaluation system results in unreliable, and possibly inaccurate, outputs that are unusable for decision making. As a consequence of this system, the Army maintains very few ways to distinguish between the training levels and capability of tactical units. By examining the organizational structure, individual and unit incentives, and evaluation methods available for the Army to adjust, this document provides a method to compare and evaluate the possible alternatives.

Thesis Advisor: Bryan R. Moser  
Title: Lead Instructor, System Design and Management Program

**Disclaimer**

The views expressed are those of the author and do not reflect the official policy or position of the Army, Department of Defense, or the U.S. Government.

## **Acknowledgements**

I'd like to thank my advisor, Dr. Bryan Moser, for his patience and faith in my ability to address this topic in a meaningful way. His guidance and insistence that I "explore" before I narrowed my topic made this an enjoyable academic endeavor. I'm also grateful for my colleague from Draper, Chris Boger, who acted as a sounding board and azimuth check during this project. My thanks to my classmates and professors, but most notably to Joe Lessard. Joe never let me settle for good enough during my studies, and helped me rethink what it means to be a warrior.

Most importantly, I'd like to thank my wife, Jessica. I would never have had this opportunity without her persistent belief in me. Thank you for nurturing our family while we both pursued degrees. I couldn't ask for a better partner or mother for our children.

**Table of Contents**

- 1. Introduction.....10**
  - 1.1. Motivation..... 10**
  - 1.2. How to read this thesis ..... 10**
- 2. Research Question .....11**
  - 2.1. Primary Research Question and Constraints..... 11**
  - 2.2. Secondary Research Questions ..... 11**
- 3. Literary Review .....12**
  - 3.1. Overview ..... 12**
  - 3.2. Team Training and Team Building ..... 12**
  - 3.3. Training Approaches, Evaluations, and Measurement Characteristics ..... 14**
  - 3.4. Ethical Decision Making ..... 17**
  - 3.5. Literary Review Summary and Critique ..... 18**
- 4. Understanding the Enterprise Landscape .....20**
  - 4.1. The Organizational Structure ..... 20**
  - 4.2. Doctrine and the Training Management System..... 23**
  - 4.3. Army Training Principles and Overview ..... 24**
- 5. Stakeholder Analysis.....38**
  - 5.1. Brigade and Higher Army Headquarters ..... 41**
  - 5.2. Training Enabler Organizations ..... 42**
  - 5.3. Battalion ..... 42**
  - 5.4. Companies ..... 42**
  - 5.5. Value Flow Relationships and Impacts ..... 43**
- 6. Current Architecture .....46**
  - 6.1. TEO Analysis ..... 46**
  - 6.2. TEO Analysis Summary..... 55**
- 7. System Dynamics Modeling .....56**
  - 7.1. Overview ..... 56**
  - 7.2. Model Scope ..... 57**
  - 7.3. B1: Train the Task ..... 57**
  - 7.4. B2: More Effort ..... 58**
  - 7.5. B3: Learn to Train..... 58**
  - 7.6. R1: Evaluator Fatigue ..... 59**
  - 7.7. B4: Goal Erosion..... 59**
  - 7.8. Reference Modes of Interest..... 60**
  - 7.9. Modeling Moral Intent..... 64**
  - 7.10. Sensitivity Analysis..... 64**
- 8. Alternative Tactical Training Architectures.....66**
  - 8.1. Architecture Generation and Reasoning ..... 66**
  - 8.2. Levers and the Morphological Matrix..... 67**
  - 8.3. Sampling Possible Tactical Training Architectures ..... 70**
- 9. Tradespace Analysis.....71**
  - 9.1. Initial Simulations ..... 71**

# A Systems Analysis of the Army's Tactical Evaluation Process

9.2.	Refinement .....	74
10.	Limitations and Recommendations .....	83
10.1.	Limitations .....	83
10.2.	Recommendations .....	84
11.	Bibliography.....	88
12.	Annexes .....	91
12.1.	System Dynamics Base Model.....	91
12.2.	System Dynamics Moral Intensity .....	92
12.3.	System Dynamics Model Code .....	92

**Table of Figures**

Figure 1: Measurement Tools (Cannon-Bowers and Salas 1997) ..... 16

Figure 2: TPMs in Training (Cannon-Bowers and Salas 1997) ..... 17

Figure 3: Infantry Brigade Combat Team Task Organization ..... 21

Figure 4: Infantry Battalion Task Organization..... 22

Figure 5: Infantry Rifle Company Task Organization..... 23

Figure 6: Sustaining Proficiency in the Band of Excellence (“FM 7-0: Train to Win in a Complex World” 2016) ..... 25

Figure 7: Nesting of METL, Collective Tasks, and Individual Tasks..... 27

Figure 8: The 8 Step Training Model ..... 29

Figure 9: Objective Task Evaluation Criteria from a TEO..... 31

Figure 10: Stakeholder Value Network for Infantry Company ..... 40

Figure 11: Stakeholder Value Network for an Infantry Company (Simplified)..... 41

Figure 12: SVN Loop 1 - Inflated Company Reports..... 43

Figure 13: SVN Loop 2 - Inflated Battalion Reports..... 44

Figure 14: Conduct a Support By Fire - Supporting Collective Tasks ..... 47

Figure 15: TEO Analysis - Feedback Granularity ..... 48

Figure 16: TEO Analysis - Feedback Objectivity..... 49

Figure 17: TEO Analysis - Feedback Type ..... 50

Figure 18: TEO Analysis - Feedback Levels..... 51

Figure 19: TEO Analysis - Feedback Competency Focus..... 53

Figure 20: TEO Analysis - Feedback Stage of Execution ..... 53

Figure 21: TEO Analysis - Warfighting Functions..... 54

Figure 22: TEO Analysis - Measurement Intent..... 55

Figure 23: Balancing Loop ..... 56

Figure 24: Reinforcing Loop ..... 56

Figure 25: B1 - Train the Task..... 57

Figure 26: B2 - More Effort..... 58

Figure 27: B3 - Learn to Train..... 58

Figure 28: R1 - Evaluator Fatigue ..... 59

Figure 29: B4 - Goal Erosion..... 59



## A Systems Analysis of the Army's Tactical Evaluation Process

Figure 30: Effect of Evaluator Capability on Feedback Quality .....	60
Figure 31: Effect of Evaluator Experience on Capability of Evaluators .....	61
Figure 32: Effect of Time on Evaluator Capability .....	61
Figure 33: Effect of Training Hours on Evaluator Fatigue.....	61
Figure 34: Effect of Quality of Feedback on Perception Gap .....	62
Figure 35: Effect of Capability on Perception Gap .....	62
Figure 36: Effect of Time on Accepted Performance.....	63
Figure 37: Effect of Augmentation on Evaluator Capacity .....	63
Figure 38: Modeling Moral Intent .....	64
Figure 39: Capability vs Inflation - Tradespace 1 .....	71
Figure 40: Training Efficiency - Tradespace 2.....	72
Figure 41: Capability, Inflation, and Augmentation - Tradespace 3 .....	73
Figure 42: Capability, Inflation, and Feedback Format - Tradespace 4 .....	74
Figure 43: Task Variability and Assessment Objectivity. ....	78
Figure 44: Capability vs Inflation - Refined Options .....	79
Figure 45: Training Efficiency - Refined Options.....	80
Figure 46: Capability, Inflation, and Augmentation - Refined Options .....	81
Figure 47: Capability, Inflation, and Feedback Format - Refined Options .....	82
Figure 48: Recommend Tradespace.....	85

## **1. Introduction**

### 1.1. Motivation

While trying to interpret my education in systems thinking and analysis into the Army context, I became extremely interested in team work and the development of our combat forces. During this exploration, I realized that our current system of training management failed to adequately capture the current state of units, their development, or provide bench marking for comparison. I found it perplexing that we had a difficult time answering the question “Are we a more capable fighting force?”

As an infantryman, I've spent over eight years devoted to military training and the development of military teams. During my two company commands, I relished our time in the field, training soldiers, and building our unit's capability. I understand that at its core, training is a highly important and emotional topic for tactical leaders. I also understand that at the most fundamental level tactical leaders are unable to objectively compare and assess our units. The current system for evaluating tactical units in the US Army produces inflated readiness reports and isn't followed as prescribed. While this may initially seem a nuisance or simply a frustrating system to operate in, the consequences of this system can be quite serious. Units that are unprepared for certain missions may be accessed as fully prepared. Additionally, how do we make objective decisions on how to change or adjust our formations and doctrine? How can we change the Army's tactical evaluation process to provide a more transparent and effective system?

### 1.2. How to read this thesis

Army personnel or people significantly trained in US Army training management should skim/skip most of Chapter Four as it tries to describe the overall environment and influencers that a company might find during its training development.

For non-Army personnel, I've attempted to explain most major training processes to a level that allows you to understand the problem. Please remember that for most of these processes there is an entire manual or series of manuals that help the Army conduct it. After over eight years of service in the Army, I've learned that I don't know it all and probably never will.

## **2. Research Question**

### **2.1. Primary Research Question and Constraints**

After examining the origins and current state of the training architecture, a primary research question evolved. Because of the currently accepted capability and the dynamic nature of the tactical domain, the major concern remained the accurate reporting of training readiness. Within the context of this work, this accuracy was defined as the difference between the final reported capability level and the actual capability level. From this the question follows, how can the Army improve the accuracy of its training readiness reports?

Upon additional reflection it was decided that the question did in fact have additional constraints. Accuracy of reporting could not come at the detriment of the development of capability. Furthermore, while costs of intervention could not reasonably be estimated within the scope of this work, there were organizational constraints due to culture and structure. It would not be beneficial to pontificate on or develop architectures that sought to drastically change the fundamental nature of Army training. Such alternatives would be difficult to envision secondary and tertiary consequences and generally be considered untenable by the organization.

With these ideas in mind the research question was reframed to: How can the Army improve the accuracy of its training readiness reports while maintaining the development of capability and offering organizationally acceptable alternatives?

### **2.2. Secondary Research Questions**

The primary research questions led to secondary quandaries that generally involved these constraints. How is capability developed within the tactical domain? What levers are available to improve report accuracy? What are the limits of organizational acceptance? These secondary questions helped both frame the research and modeling as well as guide alternative selection and evaluation.

### **3. Literary Review**

#### 3.1. Overview

The training literature is a large body of knowledge centered on process and performance, in individual and team contexts. Because the vast nature of the literature, this review focuses on several meta-analyses which attempt to evaluate the body of knowledge to pull overarching trends. Efforts were made to select the most relevant material to infantry teams and from leading researchers within the military training domain. Team training draws the majority of its cases from commercial and military aviation, health care, academia, and sports. When searching for infantry team training specifically, few cases appeared. While the military was relatively well represented (9/82 studies (McEwan et al. 2017), 762/2650 military teams (Salas et al. 2008)), military aviation and navy command and control teams seem to account for the majority of the research. Only a handful of Army maneuver studies were found mentioned within one source (Michael T. Brannick et al. 1997). Within the team domain most reviews describe a lack of rigorous experimental design (Salas et al. 2008)(Klein et al. 2009)(McEwan et al. 2017)(Kleingeld, van Mierlo, and Arends 2011). This complaint seems to be the basis for several of the conflicting findings and inability to draw many definitive statements within the team training domain. The following review will attempt to capture some of the findings and state of understanding as pertaining to team training.

#### 3.2. Team Training and Team Building

The literature often divides first into two categories – team training and team building. Team training is a collection of team interventions that are aimed at directly improving the team's function through cross training, team coordination and adaption training, guided team self-correction or other similar treatments (Salas, Nichols, and Driskell 2007). Team building is a collection of team interventions (goal setting, interpersonal relations, problem solving, role clarification...) that seeks to increase participation or commitment to the team in an effort to increase team performance. Because the specific definitions of these interventions varies greatly throughout the literature (Salas et al. 1999)(Klein et al. 2009) it seems the methods of each are more easily agreed upon.

All reviews found that team training had at least some positive effects on team function. Team training appears to have the greatest effect on process outcomes but also has the ability to

influence performance outcomes. Salas et al. found moderate positive effects in all team function (Salas et al. 2008) and small to moderate effects on performance outcomes (Salas, Nichols, and Driskell 2007). When comparing the impact on team performance, taskwork, teamwork, and mixed content seemed to display similar magnitudes of effect. One explanation presented for the lack of superiority in mixed content was the possible limits on teams to either absorb, inculcate, or understand the entirety of the mixed content interventions. (Salas et al. 2008) This seems to suggest a learning limit regardless of quality of training intervention – a limit to training growth. However, teamwork training demonstrated “enhanced affective outcomes in comparison with taskwork training.” There was some evidence to suggest that team training that focused on improvements in coordination strategy offered the most performance improvement (Salas, Nichols, and Driskell 2007). While team training interventions showed “nearly equal impact on outcomes as process,” Salas et al suggest that improving process will positively influence performance outcomes. (Salas et al. 2008)

The other common approach to team improvement is team building. Team building has become increasingly popular but presented mixed reviews within the literature. Luc et al. found large significant impact of team building on performance. Interestingly, they found that goal setting had the largest impact from the team building interventions. Perhaps a ceiling issue, team building was found to be significantly more beneficial in individual sports than in team sports. The idea was that team sports often already spur much of the interpersonal skills and commitments that are the objective of team building exercises (Martin, Carron, and Burke 2009). Another review found no significant evidence for most team building interventions (goal setting, interpersonal relations, problem solving, role clarification) and their effect on performance. Role clarification was the only individual component of team building that had a significant, albeit weak effect on team performance. Interestingly enough they also found that these effects varied in regards to objective and subjective measurements of performance. Team building showed no significant effect on objective measures and only weak effect on subjective measures (Klein et al. 2009). This analysis also found diminishing effects of team building as team size increased. “Optimal benefit from a team-building intervention seems most likely to be obtained with relatively small teams” (Klein et al. 2009). Crew Resource Management (CRM) developed within the aviation community as a team building system. CRM traditionally focuses on interpersonal skills, such as communication, situational awareness, problem solving, decision making, leadership,

assertiveness and teamwork. Because of the success of CRM within the aviation community, other industries, most notably medical, adopted its use. While O'Dea et. Al. found significant and large effect of CRM training on knowledge and training behavior, they were unable to support claims of clinic care outcomes or the long term impact of the training (O'Dea, O'Connor, and Keogh 2014).

The literature contained several conflicting reports highlighting the team building method of group goal setting (Kleingeld, van Mierlo, and Arends 2011; Nahrgang et al. 2013; Salas et al. 2008; Martin, Carron, and Burke 2009). Though research on individual goal setting suggests specific learning goals positively impacts individual performance, in the team setting, it limits the scope of coordination and hurts performance when compared to generalized "do your best" learning goals (Nahrgang et al. 2013). Nahrgang et al explain further stating, "the negative effects of specific learning goals, relative to general "do your best" learning goals and specific performance goals, are particularly pronounced in complex tasks" (Nahrgang et al. 2013). In possible explanation, teams with general "do your best" learning goals had higher levels of team coordination relative to team with specific learning goals while accomplishing complex tasks (Nahrgang et al. 2013). In direct contradiction to analysis simply two years prior, Kleingeld et al found that specific difficult goals increased group performance and stated that "goal setting thus appears to be at least as effective at the group level as at the individual level (Kleingeld, van Mierlo, and Arends 2011)." They go on to explain that task complexity failed to moderate goal setting's effect on team performance, but submit that their data set for task complexity held relatively few examples and were primarily field studies (Kleingeld, van Mierlo, and Arends 2011). The sports domain found similar results, stating that goal setting produced large effects in performance improvements (Martin, Carron, and Burke 2009).

### 3.3. Training Approaches, Evaluations, and Measurement Characteristics

Throughout the literature, team process and performance are evaluated and utilized to provide feedback to teams. The type of feedback a team receives has significant impacts on their learning. Performance is by far the most common criterion that training and teamwork has been evaluated against (McEwan et al. 2017). Process evaluation decomposes into teamwork and taskwork. Salas explains that "taskwork involves the execution of core technical competencies within a given domain, teamwork refers to the range of interactive and interdependent behavioral

process among team members that convert team inputs into outcomes (Salas et al. 2008).” Taskwork training often falls in line with the ideas of Taylor’s scientific management and has demonstrated its applicability since the early 1900’s (Taylor 1911). Teamwork training has proven effective at improving both teamwork and team performance (McEwan et al. 2017). New teams benefited greater from teamwork training than intact teams in terms of teamwork improvement; however, intact teams better leverage teamwork training to improve team performance (McEwan et al. 2017). Salas et. al found similar conclusions, showing that while both intact and new team types benefited from team training, intact teams better utilized feedback into enhanced performance outcomes (Salas et al. 2008). The manner in which teamwork is measured, general/omnibus/composite or individual dimensions, does not seem to change the magnitude of effect on teamwork or team performance. Both methods of criterion yielded similar positive results (McEwan et al. 2017).

The literature contained evaluations of several different training approaches or methods. Some of the most common approaches include class instruction, workshop interactions, simulations, demonstration, and reviewing process/performance after performing the concerned task. The variance in task type, context, and complexity all has significant effects on the type of training approach selected. Often organizational restrictions and norms dictate the approach, so comparisons between approaches within the same task and condition are rare. However, all training methods were found to be effective at enhancing team performance (McEwan et al. 2017).

The considerable amount of teamwork literature revolves around the measurement of teamwork. In efforts to forward this work, several authors have contributed frameworks and recommendations for the measurement and assessment (Michael T. Brannick et al. 1997). Dickinson and McIntyre presented three separate measures to help with the construction of teamwork measurement: behavioral observation scale, behavioral summary scale, and behavioral event. Behavioral observation scales are a numerical scale that are used to rate the frequency of teamwork observations during a training event (i.e. almost never, sometimes, almost always...). On this form, an evaluator would rate several sub-components of particular type of teamwork. Behavioral summary scales are numerical scales that are used to rate a component of teamwork by the degree of teamwork displayed. This form utilizes only one rating for capture a component of teamwork. Behavioral event formats are checklists of critical events that have been crafted for a specific event and scenario (Dickinson and McIntyre 1997). Each format differs in the ease of use

## A Systems Analysis of the Army’s Tactical Evaluation Process

as well as the manner of implementation. For example, behavior event formats are meant to be completed while the team completes the event, while the other two formats should be filled out at the completion of the training exercise. Behavior event formats also require extensive prior training for evaluators to ensure they understand the critical events and standards for each (Dickinson and McIntyre 1997).

The measurement of training fell into three main types – third party, self-report, and objective. Third party measurement could be further decomposed into internal (supervisory) and external parties, but the literature reviewed analyzed these categories jointly. All types of measurement seemed to have positive effects on training outcomes; however, they varied in effectiveness. Teamwork was improved most by third party assessments and then self-reporting. McEwan also found that objective measures slightly edged out third party assessments in terms of effectiveness on team performance, but each showed moderately positive effects (McEwan et al. 2017). However, Salas found no statistical difference between the magnitude of effect between objective and subjective supervisory ratings (third-party), stating they found “a significant, small-to-moderate tendency for team training to lead to an increase in team performance ... for objective productivity measures of performance and for supervisory ratings of performance (Salas, Nichols, and Driskell 2007).” Interestingly, Bommer et al’s analysis suggests that objective and subjective measures cannot be used interchangeably; however, he caveats this with the suggestion that objective and subjective measures of the same level and construct might be more interchangeable (Bommer et al. 1995).

Cannon-Bower and Salas provide recommendations for selecting and using team performance measures (TPMs). They recommend that TPMs must include multiple levels of measurement (individual/team..), address process as well as outcome/performance, describe-evaluate-diagnose

Measurement Tools to Assess TPMs in Training

	TEAM	INDIVIDUAL
PROCESS	<ul style="list-style-type: none"> <li>• Observational Scales</li> <li>• Expert Ratings</li> <li>• Content Analysis</li> <li>• Protocol Analysis</li> </ul>	<ul style="list-style-type: none"> <li>• Decision Analysis</li> <li>• Policy Capturing</li> <li>• Protocol Analysis</li> <li>• Observational Scales</li> </ul>
OUTCOME	<ul style="list-style-type: none"> <li>• Observational Scales</li> <li>• Expert Ratings</li> <li>• Critical Incidents</li> <li>• Automated Performance Recording</li> </ul>	<ul style="list-style-type: none"> <li>• Automated Performance Recording</li> <li>• Critical Incident</li> <li>• Expert Rating</li> <li>• Archival Records</li> </ul>

Figure 1: Measurement Tools (Cannon-Bowers and Salas 1997)

performance, and provide a basis for remediation (Cannon-Bowers and Salas 1997). Their work outlines the importance of measurement as a tool to evaluate teams with the ultimate task of



identifying reasons for underperformance. They developed a simple framework to help practitioners develop both TPMs as well as measurement tools for training events. They recommend sampling from each section of their quadrant diagram for training measures and then identify the appropriate tools to measure them from their tool quadrant. This perhaps over simplified framework provides a good basis for any organization to access their measurement and evaluation methodology. Interestingly enough, Cannon-Bower and Salas conduct the majority of their research on military teams.

### Training Performance Measures in Training

	TEAM	INDIVIDUAL
PROCESS	<ul style="list-style-type: none"> <li>• Shared Task Models</li> <li>• Cue/Strategy Associations</li> <li>• Task Organization</li> <li>• Compensatory Behavior</li> <li>• Collective Efficacy</li> <li>• Dynamic Reallocation of Function</li> <li>• Task Interaction</li> </ul>	<ul style="list-style-type: none"> <li>• Assertiveness</li> <li>• Information Exchange</li> <li>• Task-Specific Role Responsibilities</li> <li>• Procedures for Task Accomplishment</li> <li>• Cue/Strategy Association</li> <li>• Mutual Performance Monitoring</li> <li>• Flexibility</li> </ul>
OUTCOME	<ul style="list-style-type: none"> <li>• Mission/Goal Accomplishment</li> <li>• Aggregate latency</li> <li>• Error Propagation</li> <li>• Aggregate Accuracy</li> </ul>	<ul style="list-style-type: none"> <li>• Accuracy</li> <li>• Latency</li> <li>• Errors</li> <li>• Safety</li> <li>• Timeliness</li> <li>• Decision Biases</li> </ul>

Figure 2: TPMs in Training (Cannon-Bowers and Salas 1997)

### 3.4. Ethical Decision Making

Maybe rightfully so, almost all literature reviewed on team training failed to address the possible issues or ethical dilemmas in reporting evaluations. Work remained primarily focused on capturing details and searching for better methods to measure and assess performance rather than improving an entire reporting and training development system. For this analysis one article Wong and Gerras discussed the surprising but pervasive ethical dilemmas and obstacles within the US Army (Wong and Gerras 2015). They outlined the organizational pressures, reporting norms, and typical behavior that often arise within the Army, as well as hypothesized about the possible reasons for an organizational culture so steeped in values to veer off course. They argued that “the conditions are set where subordinates and units are often forced to determine which requirements will actually be done to standard and which will only be reported as done to standard (Wong and Gerras 2015).” This analysis led to an examination of the components of ethical decision making, particularly within an organizational setting.

Ethical decision making has another large a separate field from team training. Interestingly, most literature on decision making within organizations usually referenced moral intensity and Thomas M. Jones (Craft 2013). Dr. Jones synthesized a new theoretical model from a base of literature and adapted them to be “issue-contingent (Jones 1991)” Jones’ model of moral intensity

## A Systems Analysis of the Army's Tactical Evaluation Process

contains six dimensions: magnitude of consequences, social consensus, probability of effect, temporal immediacy, proximity, and concentration of effect. Each of these terms is defined below according to Jones:

- **Magnitude of Consequences:** the sum of the harms (or benefits) done to victims (or beneficiaries) of the moral act in question
- **Social Consensus:** the degree of social agreement that a proposed act is evil (or good)
- **Probability of Effect:** the probability that the act in question will actually take place and the act in question will actually cause the harm (benefit) predicted
- **Temporal Immediacy:** the length of time between the present and the onset of consequences of the moral act in question (shorter length of time implies greater immediacy)
- **Proximity:** the feeling of nearness (social, cultural, psychological, or physical) that the moral agent has for victims (beneficiaries) of the evil (beneficial) act in question.
- **Concentration of Effect:** an inverse function of the number of people affected by an act of given magnitude

Craft's meta-analysis found 22 empirical studies that investigated moral intensity and its effects on ethical decision making. In general, moral intensity was found to be a good predictor or influencer of ethical behavior (Craft 2013).

### 3.5. Literary Review Summary and Critique

After reflecting on the body of knowledge available within the training and ethical decision making literature, there are few concrete truths that appear. Any hopes of finding definitive answers, or silver bullets, to address optimizing team training have been diminished. Teams are complex and training and evaluating them proves difficult. The takeaways for infantry training, as I see them, are simple and more guidelines than hard and fast rules.

- **Training works.** Team building that is focused on communication and role clarification seems most appropriate. Team training techniques follow much of the standard techniques utilized by infantry units.

- **Objective measures are great, but incomplete.** While objective measures can be extremely helpful when comparing teams or iterations without bias and at high fidelity, they often miss the overall picture. Subjective measures allow evaluators to capture those things we haven't figured out how to capture objectively, or can't. It seems that these measures should be used in conjunction, which objective measures utilized when available and appropriate. There may come a time where technology allows us to completely capture team performance in objective measures, but no current examples or literature suggest this has been achieved in any domain.
- **Performance and Process measures must be utilized together.** While performance measures might capture the end state or current stats of a training unit, they typically fail to provide the why or present clues to improve performance. Maybe just as important, completely focusing on process measures fails to identify the point of dimensioning returns, process bottlenecks, or the final end state. These measures must be utilized together to ensure that teams receive holistic and helpful feedback.
- **Who evaluates is important – Objective  $\geq$  Third Party  $>$  Self-Evaluation.** These relationships seem intuitive, though the objective and third party assessment should be considered carefully (see #2). Self-Evaluations also should not be discounted as they seem to be important for team building and perhaps buy-in to training overall.

The critique of the literature is simple; the vast nature of teams and tasks makes application tremendously difficult. It would be foolish, if not negligent, to take findings from literature in one domain and apply it blindly to another. Because of this dilemma, it seems that organizational self-study is the only viable option for the optimization of a training process. The infantry spends the majority of its time training individuals and teams, but fails to utilize these opportunities to conduct serious studies at scale or question existing best practices. As such, there is a dearth of literature for practitioners to draw from in the infantry domain.

#### **4. Understanding the Enterprise Landscape**

The mission of the United States Army is “to fight and win our Nation’s wars”(“United States Army: Organizational Information” 2017). Training is the primary mechanism that the Army uses to prepare for current and future missions as well as inform decision makers about current capabilities. In order to understand the challenges that the infantry company currently faces while training, it is important to grasp the external environment in which it exists. The following section describes this larger environment in terms of organizational structure, training doctrine, technology, and resources. The chapter concludes with a brief explanation of the identity of the infantry company, some guiding principles on training, and the motivations for change.

##### **4.1. The Organizational Structure**

The Army structures itself into standardized tiered formations to enable dissemination of information and allows for its members to both quickly understand and predict the organizational needs and processes. Units and formations are mission based organizations that are structured to accomplish a set of tasks. Units often contain several different specialized organizations or soldiers to accomplish subtasks that support the unit’s larger missions. In order to meet several operational needs the Army chose to augment the brigade task organization so that it could independently deploy and support itself. It is the “primary fighting headquarters of the US Army tactical fight.” (“FM 3-21.20: The Infantry Battalion” 2006)

The infantry brigade combat team (IBCT) centers around its infantry units but organically contains cavalry, field artillery, intelligence, signal, sustainment, and other unit types to support its operations. Figure 3 graphically depicts this organizational structure. A colonel commands an IBCT and retains responsibility for ensuring his unit is fully trained and prepared for future operations. He serves as direct supervisor/rater to all 6 battalion commanders and senior rater to all 35+ company commanders in his formation. Brigade commanders are responsible for training battalions and certifying companies. (“FM 3-96: Brigade Combat Team” 2015)

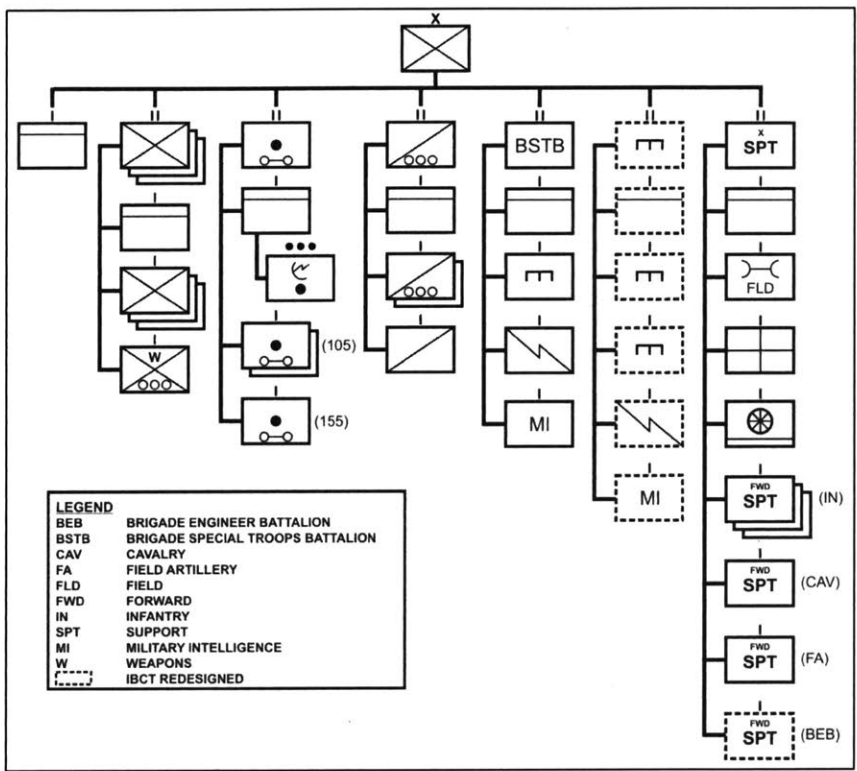


Figure 3: Infantry Brigade Combat Team Task Organization

The infantry battalion acts as the IBCT's main maneuver force and was designed for expedited employment in full spectrum operations. Each of the three infantry battalions consists of a headquarters and headquarters company, three rifle companies, and one weapons company. The battalion commander serves as the direct supervisor and rater to all five company commanders. Battalion commanders train company level tasks and certify platoons. ("FM 3-21.20: The Infantry Battalion" 2006)

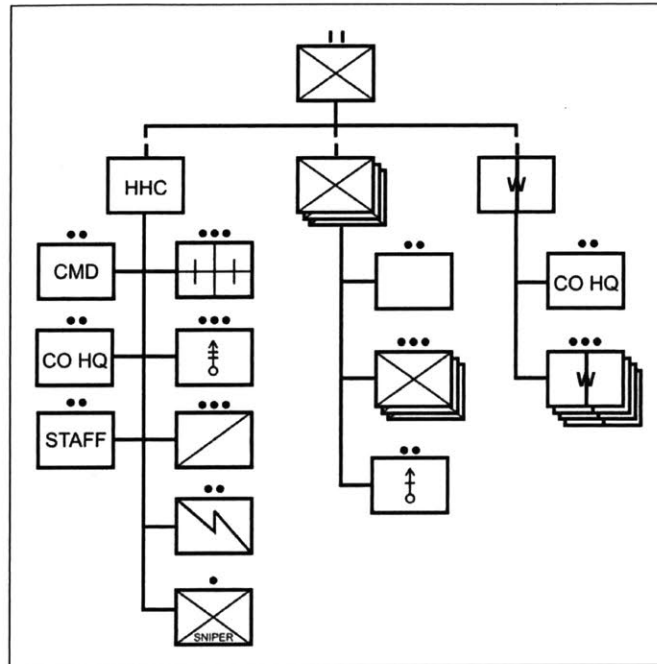


Figure 4: Infantry Battalion Task Organization

The infantry rifle company contains a company headquarters section, mortar section, and three rifle platoons. Each rifle platoon consists of a platoon headquarters, three rifle squads, and one weapons squad. Company commanders hold responsibility to for training their platoons and certifying all squads within their formation. (“FM 3-21.10: The Infantry Rifle Company” 2006)

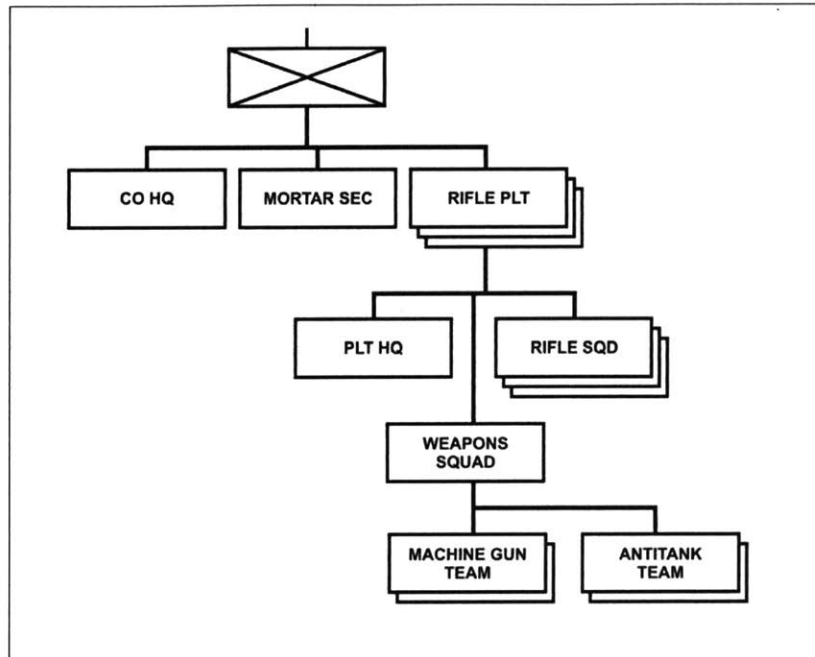


Figure 5: Infantry Rifle Company Task Organization

This dissection of the IBCT organization helps us understand the scale and complexity of the organization. Managing the training of over 10 different company types and 35 total companies creates a myriad of issues and frustrations. In order to facilitate the development of these organization's capability, the Army utilizes doctrine.

#### 4.2. Doctrine and the Training Management System

Doctrine is the body of knowledge that the Army uses as a baseline for operations and training. Doctrine "states what the Army is, what the Army does, how the Army does it, and where the Army is going." ("ADP 1: The Army" 2012) Doctrine is official, approved and distributed by the Army headquarters. Changes occur slowly and with deliberate thought and consideration. These documents set what personnel will make up a unit, the equipment they will utilize, how they maintain and train with the equipment, and how a commander will develop his staff to visual the battlefield.

*"Doctrine provides a means of conceptualizing campaigns and operations, as well as a detailed understanding of conditions, frictions, and uncertainties that make achieving the ideal difficult."* ("ADP 3-0: Operations" 2016)

If there is a task that the Army conducts, there is an associated set of doctrine to set a standard and provide guidance for its execution. The Army spends considerable time outlining the method and guidance for the conduct of its training. The doctrine that outlines this process is FM 7-0: Train to Win in a Complex World.

In effort to clarify and improve the Army's training doctrine, FM 7-0 was extensively reformed and published in October of 2016. While many of the core tenants and ideas of training remained the same, the methodology of unit assessment changed significantly. The following sections will present the overall key concepts covered in FM 7-0 as they provide the most current and projected methods and thoughts on training within the Army. The first part will focus on how the Army views training and its principles, followed by how a unit develops its training plan, and ends with the methods of training assessment.<sup>1</sup>

### 4.3. Army Training Principles and Overview

The Army begins its training doctrine with a set of principles to guide this process. The Army lists their training principles as ("FM 7-0: Train to Win in a Complex World" 2016):

- Train as you fight.
- Training is commander driven.
- Training is led by trained officers and noncommissioned officers (NCOs).
- Train to standard.
- Train using appropriate doctrine.
- Training is protected.
- Training is resourced.
- Train to sustain.
- Train to maintain.
- Training is multi-echelon and combined arms.

Unit commanders and staffs often cite or echo these principles while planning training. They are frequently tied back to lessons learned from previous training events and operations as

---

<sup>1</sup> Most of these changes found within FM 7-0 are the result of the OBJ-T pilot program. COL Hudak gave the author a brief that outlined the pilot and current state which greatly contributed to author's overall understanding of the newly developed system.



## A Systems Analysis of the Army's Tactical Evaluation Process

well as reasoning for future priorities. These principles set the basis for many of the training regulations, guidance, and norms found within the Army's training system.

Units routinely report their training status to the Army and their higher units. Often this occurs in a formal brief or report that informs decision makers on the current pace of capability development and status of available units. The Army standardizes its proficiency ratings into a simplistic scale that ranges from Trained (T) to Untrained (U) ("FM 7-0: Train to Win in a Complex World" 2016).

- T is fully trained (complete task proficiency).
- T- is trained (advanced task proficiency).
- P is practiced (basic task proficiency).
- P- is marginally practiced (limited task proficiency).
- U is untrained (cannot perform the task).

It is understood that training proficiency remains a temporary status and personnel loss, position changes/promotion, skill deterioration, and other factors degrade the level of training proficiency. The Army attempts to employ and train its units in such a manner as to ensure that a units training proficiency remains sufficiently high or in the "band of excellence."

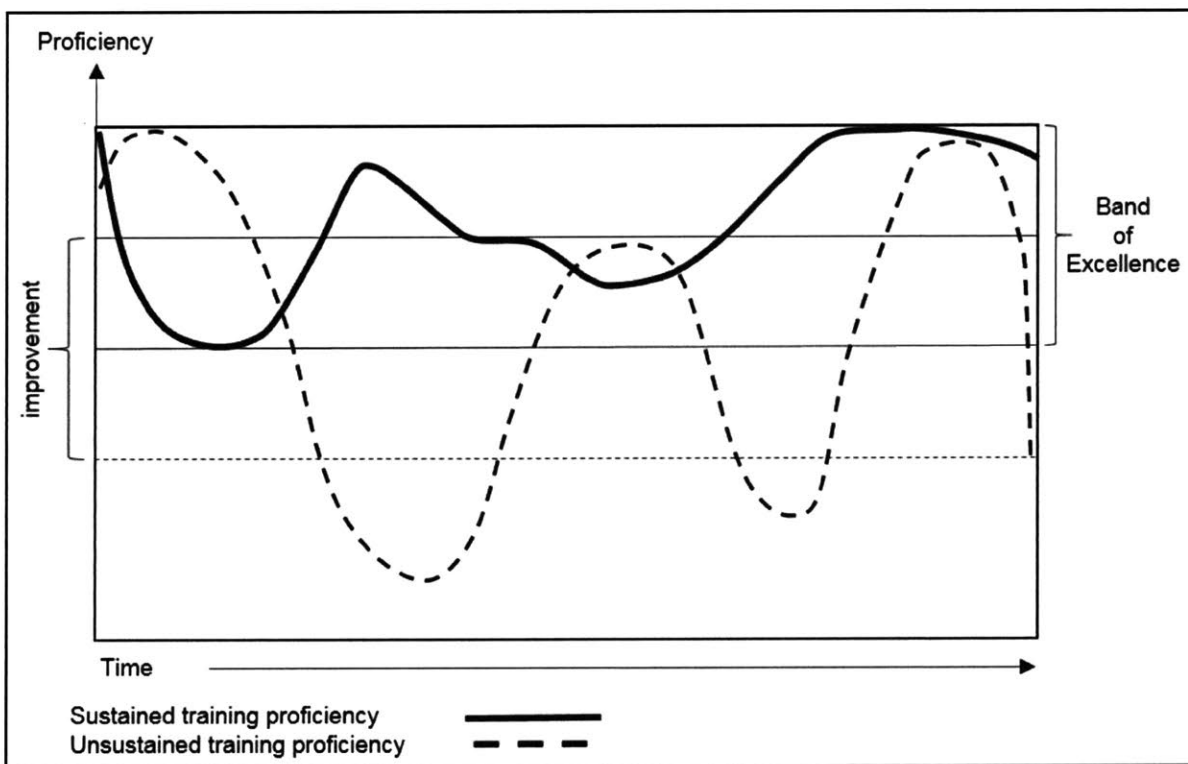


Figure 6: Sustaining Proficiency in the Band of Excellence ("FM 7-0: Train to Win in a Complex World" 2016)

Sustaining training proficiency within this “band of excellence” ensures that units are capable of quickly addressing training shortfalls and mobilizing with the full desired set of capabilities. Commanders internalize this concept and build their unit training plans to address these factors.

### 4.3.1. Unit Training Plans

The Army made a significant change to FM 7-0 when they standardized Mission Essential Task Lists (METL) for standard units at the company level and above. Prior to this change, units utilized their training guidance, assigned mission, and the higher-level unit's METL to develop their METL. This previous condition ultimately made it extremely difficult to assess like units across formations. For example, one infantry unit could have 7-10 tasks in their METL while their sister unit has only 3 similar but different tasks. This change ensures that all units of the same type and size will rate themselves against the same METL. While this may seem fairly arbitrary, the significance of this small change cannot be understated.

The METLs of a unit nests within its higher-level unit. In this way, higher level units can ensure that the independent training and focus of the subordinate units helps develop their capacity and ultimate METL assessments. This nesting is a concept that we find throughout Army doctrine (mission tasks, structure, training...). Figure 7 helps demonstrate this concept within a field artillery context. Individuals soldier skills build into small collect team tasks and feed larger team tasks, every subordinate units' training nesting within the parent organization's task.

## A Systems Analysis of the Army's Tactical Evaluation Process

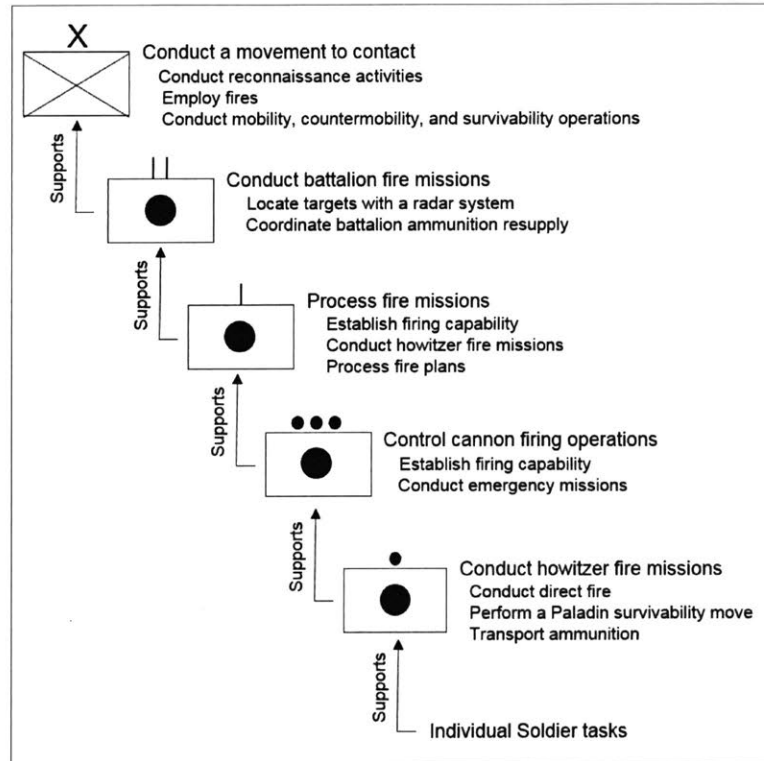


Figure 7: Nesting of METL, Collective Tasks, and Individual Tasks

Training guidance starts at the highest levels of the Army and terminates at the battalion level. Usually, division and higher develop and disseminate annual training guidance while brigades and below develop annual and quarterly training guidance. These documents often cite fairly standard requirements (regulations on weapons qualification, what commanders certify which units, ...) that rarely change, but often provide context for training priorities, methods, and upcoming culminating events. FM 7-0 describes a commander's training guidance and highlights the following topics as key points to expound upon ("FM 7-0: Train to Win in a Complex World" 2016):

- Unit's training focus, including its capabilities and mission
- Desired readiness level
- Long-range planning horizon
- Installation or command time management cycle
- External Evaluations (EXEVAL) dates and responsibilities
- CTC rotations
- Training environments in which to train

## A Systems Analysis of the Army's Tactical Evaluation Process

Once a unit has the prescribed METL and layers of training guidance, they can conduct a self-assessment on these tasks and start to develop a Unit Training Plan (UTP). The unit training plan is a long-term plan that addressed how a unit will develop and maintain a set of tactical capabilities through a series of training events, exercises, and assessments. Units typically develop this UTP to nest within their larger units UTP. It generally starts from the end of culmination of a large training event or at the beginning of a reset phase. For example, a company might map who they plan to train and progress from individual tasks, buddy team, team, squad, platoon, and company tasks. Units generally develop a stair step model that evolves from the decomposition of METL tasks into collective and individual tasks.

The first step in developing this UTP is to identify the collective and individual tasks that a unit feels best supports the training guidance and prescribed METL. METL tasks tend to be general and lack specific environmental, enemy, or context information. For example, Conduct an Attack in an Urban Area might be an assigned METL task. A unit must look to the training guidance to find direction or suggestions as to the context of that associated task. A unit might be preparing for operations in a low-intensity conflict with complex host nation support. This environment and situation would vary greatly in comparison to an initial invasion of an occupied territory. Units also look at the description of their higher's proposed EXEVAL and training events to guide their collective task selection. Each collective task comprises of a series of individual tasks. Many of the individual tasks overlap between collective tasks, with one or two specialized individual tasks differentiate the collective tasks.

Once a unit has identified its collective and individual tasks to train, it can evaluate the remaining training time available. This available training time receives a lot of attention. It is widely believed that company level units do not have enough time to train all the required tasks and missions given to them. A unit starts with the big blocks first – higher's dictated training events, EVEVALs, mandatory training events, and battle rhythm events. Commanders then map the command or installation time management cycle to the available time. These time management cycles prescribe training and tasking priorities and let units know when they will most likely be able to gain access to training resources to be tasked with additional duties (limited or no ability to train). With these final constraints, units begin working backwards from the culminating training events. Commanders develop a series of courses of action to brief to their higher-level unit and receive approval.

#### 4.3.2. Conducting Training

After companies get their UTPs approved they begin conducting training. The Army has a specific process that helps commanders develop, monitor, and track their training events. This process is called the 8 Step Training Model. This simplistic model helps leaders break down training requirements think about the training event in a more holistic fashion. A key highlight of this process is the After-Action Review (AAR). Upon completion of a training event or mission the training unit conducts a short meeting to discuss what happened, what was supposed to happen, and how they can improve. This short meeting is called the AAR. The AAR dominates Army training culture and is often thought of as the primary if not only vehicle to review performance and provide feedback in the training environment. The Army feels that this is so important that they have an entire manual to explain the best ways to conduct an effective AAR, Training Circular 25-20 A Leader's Guide to After Action Reviews. The AAR may or may not be the entirety of the evaluation process for an Army unit.

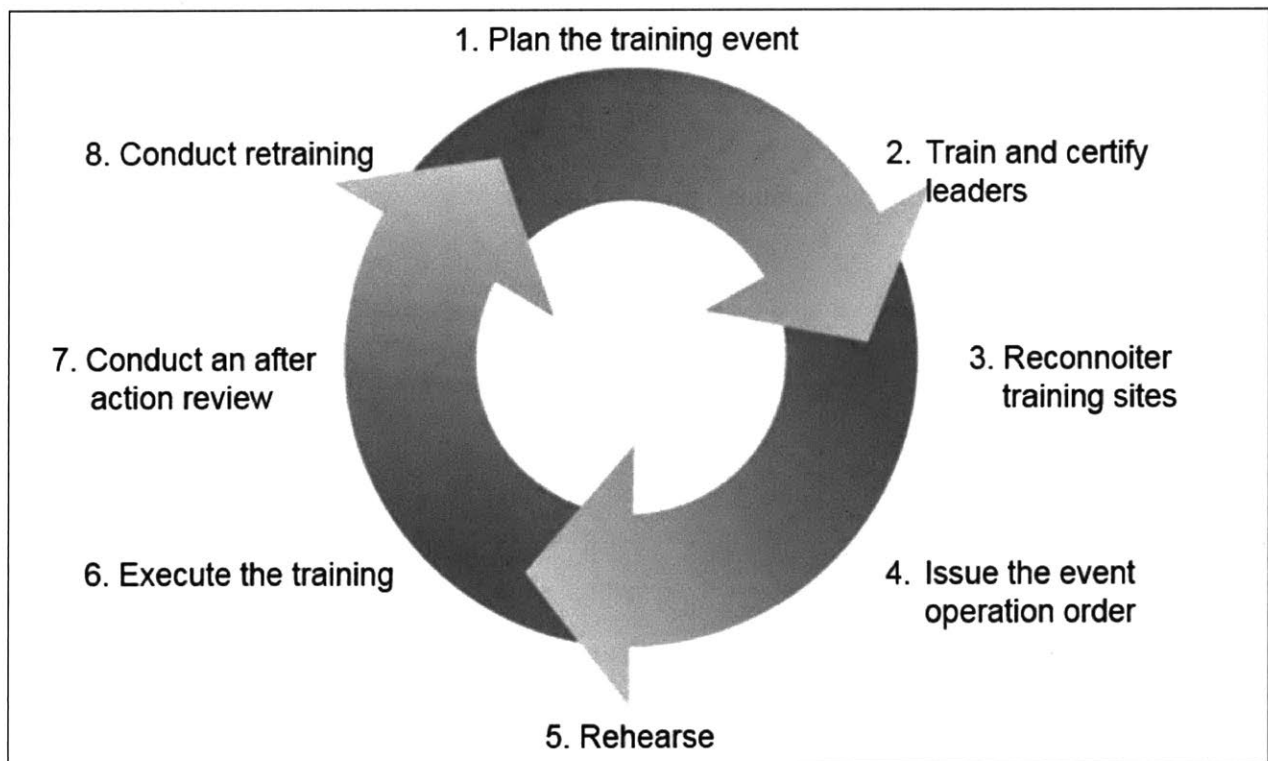


Figure 8: The 8 Step Training Model

## A Systems Analysis of the Army's Tactical Evaluation Process

The Army utilizes two methods of evaluation – internal and external evaluations. The propensity of training events find themselves subject to internal evaluations. Evaluators from within the organization provide feedback and assessments to the training unit and commander. External evaluations are reserved for large scale culminating events. These events typically involve an entire brigade. Often a unit will only receive external evaluation during a Combat Training Center (CTC) rotation. These events happen infrequently and require a significant amount of resources and coordination.

The Army uses the Training Evaluation Outline (TEO) to assess all tactical training. The TEO is an ordered list of steps to accomplish a particular task. The TEO contains individual and team tasks and spans planning, execution, and follow preparations.

After a unit has conducted training they transform their TEO assessments into updates of their training readiness ratings. Previously, the training readiness rating (T, P, U, ...) came from a subjective assessment of the commander and their thoughts on how capable their unit was at a particular METL task. With the latest revision, the Army outlined the training readiness ratings with associated training conditions and completion standards. This formalized a standard scale that units can use to compare their performance. This change also took it one step further and sought to incorporate manning and the training environment conditions into account. This makes the new training readiness rating a much more complete and standardized scale.

Plan and Prepare				Execute					Assess	
Operational Environment			Training Environment (L/V/C)	% Leaders present at training/authorized	% Present at training/authorized	External evaluation	Performance measures	Critical performance measures	Leader performance measures	Task assessment
SQD and PLT	CO and BN	BDE and above								
Dynamic (single threat)	Dynamic and complex (4 + OE variables and hybrid threat)	Dynamic and complex (all OE variables and hybrid threat)	Proponent establishes training environment standards	≥85%	≥80%	Yes	≥90% GO	All	≥90%	T
				75-84%			80-90% GO		80-89%	T-
Static (single threat)	Dynamic (single threat)	Dynamic and complex (all OE variables and single threat)		65-74%	75-79%	No	65-79% GO	<All	<80%	P
				60-64%			60-74%			51-64% GO
	Static (single threat)	Dynamic & complex (< all OE variables and single threat)		Day	<60%		<60%			<51% GO
BDE	brigade	OE		operational environment	T	fully trained				
BN	battalion	P	practiced	T-	trained					
C	constructive	P-	marginally practiced	U	untrained					
CO	company	PLT	platoon	V	virtual					
L	live	SQD	squad							

Note: The percentages used in this figure are for illustration only. See the collective task's published training and evaluation outline for the applicable percentages.

Figure 9: Objective Task Evaluation Criteria from a TEO

It is worth noting that while this new scale attempts to make a previously subjective assessment objective in nature, FM 7-0 caveats its use by stating that:

*“While the T&EO and task proficiency standards provide objective criteria for determining task proficiency, assessments allow leaders to take into account the subjective nature of training. Leaders’ assessments combine their professional*

## A Systems Analysis of the Army's Tactical Evaluation Process

*observations with other information to develop an overall assessment of the unit's ability to accomplish its mission. Final authority of a unit's assessment lies solely with the commander."*

With this final matrix, we see how TEOs drive individual and collective assessments of tasks and lead to training readiness ratings. This process attempts to standardize and simplify the highly complex task of evaluation combat training into a singular metric that provides a holistic representation of a unit's tactical capability.

### 4.3.3. Army Training Economy

While the Army has few competitors to achieve its core mission, it has leverage several outside stakeholders to augment its tactical development. It may not be helpful to outline the entire military industrial process, but the civilian agents that facilitate the direct development of tactical capability warrant mention. The Army has chosen to outsource large portions of the current training process. This has created a series of stakeholders that have interests in the methods and processes the Army uses. While it is not completely within the scope of this work, it is important to realize that beyond the normal obstacles that arise during that adjustment of a large bureaucratic process, any adjustment to this system will face scrutiny and challenges from these additional stakeholders.

The land and ranges that the Army utilizes for its tactical training is run by civilians. Range Control is the term universally used for this series of organizations that manage the development and operations of training land. They develop pieces of land to help Army units conduct specific types of training exercises and ensure for the proper maintenance of the land and training equipment. Targetry, structures, safety, intra-range coordination, and communications are just some of the portions of the training environment that Range Control oversees. Any changes to Army training directly affects the associated tasks of Range Control.

The Training Audiovisual Support Center (TASC) also supports Army training. They provide props, costumes, simulators, Multiple Laser Engagement System (MILES) gear, and many other things. Their main purpose is to enhance the training effects by providing a more realistic and immersive environment for the soldier.



### 4.3.4. Resourcing Training

Installations control the land that units utilize for training. Commands set priorities for land resources based off of unit type, mission timelines and priorities, and training cycles. Units often employ cycles to ensure everyone has an opportunity to conduct training on key training grounds. Installations generally allocate the land in a quarterly meeting based of the requests of the training units for pending training events. Most posts contain plenty of space to accommodate simple training exercises, but key training events often require special ranges due to safety constraints. It is not uncommon for a post to only have one or two pieces of land that can support a platoon live fire exercise. Units handle disputes and conflicts according to the priorities. All unclaimed land remains available for training as long as a unit receives Range Control's permission.

Units receive a yearly allocation of ammo based off of Standards in Training Commission: DA PAM 350-38 (STRAC). This document outlines a unit's ammo recommendations by type and event. For example, this document tells a unit how many bullets they are allowed to use by weapon type and number for a team live fire exercise. Units identify training events that require ammunition and forecast the needed ammo three months from the month of utilization. Units strictly adhere to ammo timelines and policies in regards to forecasting. Commander's sometimes find reasons to deviate from STRAC recommendations for ammo usage.

### 4.3.5. Technology Usage

The Army has the one of most technological advanced fighting forces in the world, but leverages little technology in its training efforts.

The Army created an online system, the Army Training Network (ATN), to house and provide resources for training doctrine. Junior officers receive an introduction into this system in their basic course. Within this system soldiers find the Combined Arms Training Strategy (CATS), resources on Unit Training Management (UTM), and tools to help identify collective and individual tasks and their TEOs. ATN still requires a solid grasp of the Army training system, but it functions fairly intuitively and provides all the necessary resources for a leader to find desired tasks and assessment standards. Previously, the Army relied on a publication system the issued hard copy manuals. By having all of these resources digitized and available online, soldiers can ensure they have the most up to date standards.

## A Systems Analysis of the Army's Tactical Evaluation Process

The ranges and training areas the Army uses generally have low or no augmentation. Traditional ranges consist of a human shaped plastic targetry with a controller that can raise and lower an array of targets. While weapons qualification ranges record the hits during a standardized iteration, most ranges fail to capture any data from a training iteration. For example, if an infantry platoon conducts a live fire exercise, there is currently no record of how many targets they shot, rounds they expended, or the duration of the event. Evaluators (people) conduct all event monitoring and data recording during this type of exercise. Generally, no data is collected. Units are able to request and use an improved MILES gear and tracking system. This system tends to be rarely utilized due to the high logistical costs of use and perceived lack of usefulness. MILES is the Army's version of laser tag, but results tend to mixed.

The Army has begun augmenting a few ranges. Most of these ranges are shoot houses, an enclosed area that is meant to simulate a small structure in urban combat. These ranges are outfitted with video recording, more realistic targets, and additional contextual props. These improved ranges are limited in the types of tasks they can train, throughput, and tend to be difficult to schedule due to demand.

The CTCs utilize MILES and other tracking sensors to help simulate their exercises and provide feedback. The system receives attention from external evaluators on a daily basis to ensure it remains functioning. Additionally, a team of civilians provides logistical and technical support to training units to maintain the sensorized aspect of a CTC rotation. MILES originated in the 1980s and received a few updates since implementation, but remains a fairly rudimentary tool. While the effort and organizational cost of using the MILES gear is substantial, this system provides a means to hold large scale exercises in a more realistic fashion. However, this would not be possible without the significant emphasis and effort of the CTC staff to ensure functionality and use.

### 4.3.6. Environment and Organizational Focus

The Army has been in a constant state of war since 2001 and maintained two endeavors since 2003. During the majority of this period, units and soldiers found themselves deploying for yearlong deployments after 12-18 month resets. The quick pace of reintegration, train up, and deployment had many effects on the organization. Besides the anticipated individual and organizational fatigue, the Army found that units adopted fairly standardized training regimens

and habits to fit within this new norm. Units received an excess of training supplies, but less time to train. Organizations adopted lessons learned from previous deployments, trained for extremely specific mission sets, and utilized a series of quickly escalating training exercises to prepare units. This relatively brief period of material glut and time constraints reduced the junior officer's competencies in traditional Army training management.

The mission of the US Army continues to expand. In the past 15 years, the Army deployed in support of several missions and across the spectrum of armed combat. An infantry unit could be preparing for Direct Action (high intensity conflict with a near-peer adversary), counter-insurgency operations (low intensity conflict), or humanitarian missions. The range of tasks and concerns of each of these missions limits an organizations ability to specialize or retain organizational competencies.

### 4.3.7. Personnel

The Army's strength lies ultimately in the quality of its soldiers and leaders. Units attempt to build organizational resiliency by teaching a soldier their job and the next level up. Additionally, when units lack adequate personnel, units fill the gaps in their formations by moving soldiers into levels of responsibility in which they may not hold the traditional rank. Units also rotate soldiers through different positions on a routine basis. It is uncommon for a soldier to hold the same position for more than a year. Likewise, officers tend to move jobs once they receive their annual officer evaluation report. The Army seeks to develop breadth in the larger organization by moving soldiers between posts every two to four years. This ensures that units share lessons learned and soldiers learn how to operate in different types of organizations (Airborne, Air Assault, Mechanized ...).

The constant movement of personnel and leaders makes training even more difficult. Once a unit trains a task, the trained team may not stay together for more than a few months. Leaders attempt to stagger the transition of key positions within an organization or team and maintain a sense of continuity, but this only mitigates the effect. Ultimately, this constant rotation of personnel leads to extremely experienced and diversified leaders and a fairly robust organization. It does however come at the expense of heavily specialized units and teams.

### 4.3.8. Internal Landscape

While outlining and describing the functions of the Army training system was necessary to better understand the context of this system, realizing how the organization internalizes this process and their culture provides the most insight.

First and foremost, the Army and its leaders view themselves as an organization of professionals and values. Soldiers are taught the “Army Values” as the bedrock of their military education and find themselves expected to uphold these values in their daily lives. At the core of most decisions and guidance lies an idea of what the Army standard is and how to support the organization's values.

The first tenant found within FM 7-0 is “Train to Win” (“FM 7-0: Train to Win in a Complex World” 2016) This concept, combined with the personnel combat experiences of leaders, pushes the organization to sometimes extreme measures. Seasoned soldiers understand the difficult complexities found in war and the ultimate consequences of untrained units and leaders. The personal losses and tribulations of previous deployments often become the motivational factor that ensures units feverishly train and critique their performance in pursuit of excellence. The lives of soldiers rely on a unit's ability to develop tactical capability and leaders own this.

A well-documented and heavily discussed issue is the quantity of tasks and trainings that the Army asks subordinate units to tackle. While studies commonly find deficits in training hours directed compared to training hours available, the most recent study found almost 20 months of tasks for annual training requirements (“No Time, Literally, for All Requirements | Association of the United States Army” 2017). Because this situation has endured for over a decade, it has had tremendous impact on the culture of the Army. A 2015 report, *Lying to Ourselves: Dishonesty in the Army Profession*, outlines this conditions impact on training reports clearly:

*“...it is literally impossible to execute to standard all that is required. At the same time, reporting non-compliance with the requirements is seldom a viable option. As a result, the conditions are set where subordinates and units are often forced to determine which requirements will actually be done to standard and which will only be reported as done to standard.” (Wong and Gerras 2015)*

## A Systems Analysis of the Army's Tactical Evaluation Process

The report details further how the Army's zero defect attitude and overwhelming requirements has created an organization that holds itself to very high moral standards, but has adjusted its language and norms to justify the inflation and at times complete falsification of reports. Wong and Gerras refer to this as ethical fading or the "psychological processes and influencing factors subtly neutralize the "ethics" from an ethical dilemma ethical fading allows Army officers to transform morally wrong behavior into socially acceptable conduct by dimming the glare and guilt of the ethical spotlight." They go on to describe how the Army has adapted terms like "prioritizing", "checking the box", "assuming risk", and others to change the language to fit the ethical ethos that the Army prides itself on. A significant portion of the Army's unreliable training readiness reporting has to do with this ethical fading.

The Army understands that they have room for improvement in the training management system. While the Army believes that training is vital to capability development and future decision making, they know that the reports of subordinate units do not accurately reflect their current training status. Leaders understand that TEOs are often used as references instead assessment tools and that unit readiness reports tend to be highly inflated. The Army has been thinking about this problem, conducting pilot programs, and investigating methods to improve the current situation. They recognize they need a better system to manage and understand unit training and readiness reports.

## 5. Stakeholder Analysis

The analysis of this work focuses on the infantry company and associated units, doctrine, methods, and processes that influence a company's development of tactical capability. This section identifies the major stakeholders of this system and the associated value flows in a stakeholder value network (SVN). SVNs allow for the graphic representation of different kinds of relationships between entities. While typically qualitative in nature, previous work has demonstrated the ability to transpose these relationships into quantitative metrics and investigation (Feng 2013; Feng et al. 2010). This stakeholder analysis will focus on the impact of these value flows and timing between the process interactions.

Labeled boxes signify entities within the system. The box represents an organization or groups of organizations within the system that hold a particular function. The box's color identifies the entity as the focal organization, market stakeholder, or non-market stakeholder. Market stakeholders are those organizations that benefit from or have direct interactions from the focal organization. Non-market stakeholders maintain influence or involvement in byproducts or resources within the system, but do not interface directly with the focal organization or have significant influence within the system. In this system, the other companies act as competition, utilizing similar resources and contending for rankings/ratings, but do not directly affect or influence the development of capability of the focal organization.

Flows are transactional relationships between organizations and are represented as arrows that connect the two entities. These relationships can be uni or bi-directional. The color of the arrow identifies the category of transaction as financial (red), political (purple), informational (black), and goods/services (blue). While only flows that related to the training of tactical units appear, all categories and seven different types of flows emerge in this system representation.

Funds (financial) denote the allocation of monies from one entity to another. Some designations originate from or are based off long term guidance or standards. An example of this might be the maintenance funds for particular units. These are allocated based off previous analysis from higher echelons that includes historical trends and projected usage. While the majority of funding for units falls into this category, they do tend to find "wiggle room" or discretionary funding to apply as desired. Processes to address one time needs and exceptions also fall into this financial category.

## A Systems Analysis of the Army's Tactical Evaluation Process

Training Resources (goods/services) symbolize all the goods that a unit might receive to facilitate training. These include items like ammo, fuel, land, simulators, sensors, chemlights, etc. Ammo and fuel, like monetary funds, often mirror long term guidance and institutional standards. However, similar to money, these resources tend to have flexibility depending on the relationship and need of the training unit. For example, a company that expresses a significant need for additional rifle ammunition to complete a training event might be allowed to exceed their allotment if their commander warrants it.

Capability (goods/services) represents the expression of tactical capability through either external evaluation or mission accomplishment. Capability can only be actualized upon the implementation of a unit. Planning and the allocation of missions often depends on the expectation of capability through the interpretation of past performance and the unit's training reports.

Reports (informational) represent the formal and informal training reports that units use to depict current training status. The formal training reports happen in Quarterly/Annual Training Briefs and Unit Status Reporting (Monthly). Informal reports vary between units but might be in the form of weekly situation reports or training storyboards. Understandably, reports are not the only method a commander uses to assess the training status of his subordinate units. However, commanders are limited in their ability to personally oversee the majority of training events and tend to rely on reports and spot checks. Units tend to respond to these spot checks by placing their best performing units on display to ensure the unit looks at its best. This mitigates the effectiveness of spot checks as a monitoring task and places additional reliance on a units training reports.

Missions (goods/services) are tasked or assigned and vary in difficulty and importance. The Army formalizes this concept when it subdivides larger missions into decisive and shaping operations ("ADP 3-0: Operations" 2016; "FM 3-96: Brigade Combat Team" 2015). The "best" units typically receive the hardest or most important missions, usually labeled the decisive operation. The other units perform supporting roles, labeled shaping operations, that enable the decisive operation. Performance on missions helps commanders understand the capability of their units and heavily influences the associated rating a subordinate commander might receive. For example, a commander that receives a difficult/important mission and performs exceptionally would expect high praises and an associated high rating when compared to their peers. A commander that receives easy/less important missions and performs well may not expect high

## A Systems Analysis of the Army's Tactical Evaluation Process

marks on their rating. The relative importance and performance of a unit's mission matters when comparing units and rating their commanders.

Rating/Ranking (political) flows signify the rating relationship between the unit commanders. Unit commanders rank their subordinate commanders amongst themselves. Commanders are also rated by their first line supervisor's rater. This second tier of rating (senior rating) has significant impacts on an officer's career, traditionally considered more impactful than rating from the first line supervisor. The best officers are rated numerically (i.e. 1 of 4) while the others receive percentile (top 50%) or simply qualitative remarks. Ratings are based off several different factors, but heavily depend on the tactical performance of a unit.

The initial SVN, figure 10, makes graphical analysis difficult due to the several flows. Entities interact heavily and at different layers. It is important to visualize the system in this fashion to better appreciate the true complexity and relationships that exist. The similar connections between units and their higher-level commands demonstrates the competition that often emerges from this system. However, not all value flows spur competition. Guidance, regulations, and to a certain degree even funds have specific allocations that rarely fluctuate and remain independent of other units.

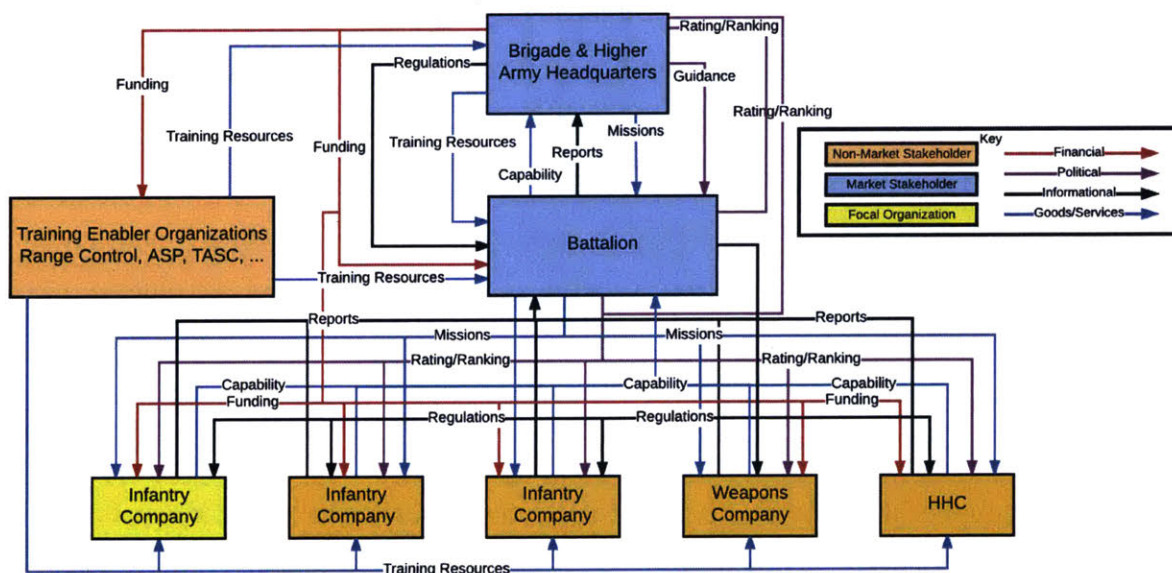


Figure 10: Stakeholder Value Network for Infantry Company

Additionally, the timing of these exchanges makes a significant difference. At the most basic level, higher level entities provide training guidance, funds, and resources to subordinate units. These units train and provide reports that update their training status. Higher level units



## A Systems Analysis of the Army's Tactical Evaluation Process

allocate missions based off of their interpretation of these reports. Subordinate units perform these mission, express their capability through performance the given mission, and get rated on their performance. The timing and cycle of these transactional relationships has tremendous impacts.

In an effort to make the relationships more apparent and identify common relationships, figure 11 simplifies the flows by combining entities with common relationships (companies). This simplified version of the SVN makes identification of the flows and relationships easier to recognize.

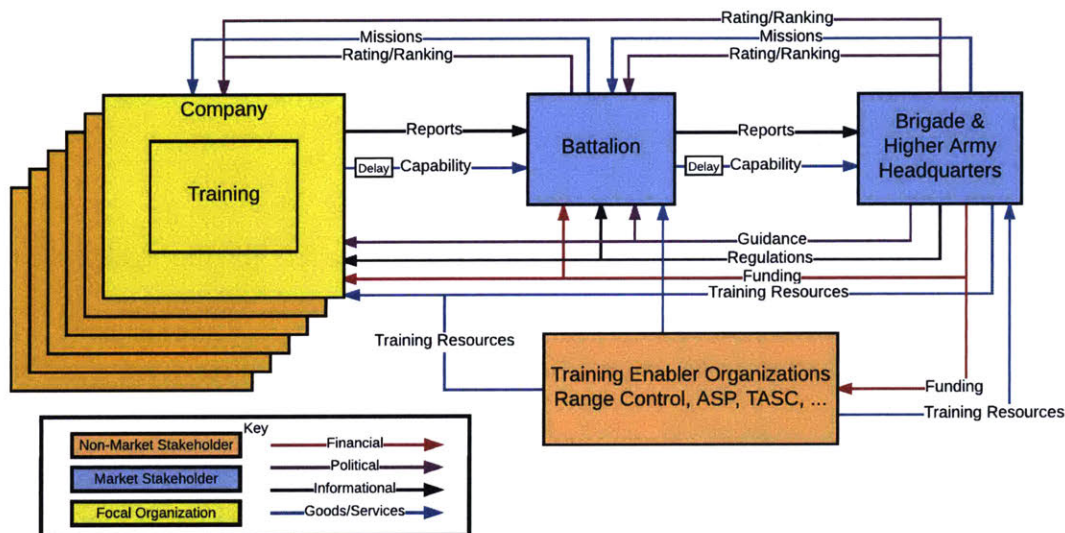


Figure 11: Stakeholder Value Network for an Infantry Company (Simplified). Companies are in direct competition for resources and reports. These reports are sent to their Battalion who provides Missions for the Company to execute. Based off their performance on these Missions, the Companies demonstrate Capability. This Capability is utilized to in-turn provide the unit with Rating/Ranking in comparison to the competition. Because Battalions are also competing, and they utilize Company reports for their reports, they are incentivized to inflate or at least maintain the Company's inflated measure (even if they know its inflated). The inability to objectively quantify training performance, self-reporting, and the organizational incentive structure results in systemic inflated reports.

To better understand this diagram, each unit will be briefly described in terms of its organizational relationships and value flows with other entities. This analytic breakdown will begin with the higher echelons and fringe organizations before analyzing the focal unit (company) and culminating in a discussion on the impacts and takeaways of this examination.

### 5.1. Brigade and Higher Army Headquarters

The Brigade and Higher Army Headquarters unit represent the echelons of units that have a significant effect on an infantry company, but do not interact directly or only through proxies.

## A Systems Analysis of the Army's Tactical Evaluation Process

The brigade acts as the intermediary for most outside agencies/units and battalions/companies. This entity sets the larger end state and training objectives for battalions and companies. This is done primarily through the development of training guidance and the approval and support of training plans. The brigade commander utilizes his staff to orchestrate the coordination of resources and guidance to ensure that lower level units are capable of completing the directed missions. This entity provides guidance, regulations, funding, and missions. Brigade commanders rate battalion commanders and senior rate company commanders. The brigade receives reports and tactical capability from battalions. The higher Army headquarters section of this entity provides funding to the Training Enabler Organizations.

### 5.2. Training Enabler Organizations

The Training Enabler Organizations entity represents all the mainly civilian run organizations that augment the army's training development during home station training. These primarily include range control, the Ammo Supply Point (ASP), and TASC. Each of them provide specialized training resources to units. They receive funding from the higher echelons of the Army. These organizations are typically the only supplier of their respective resource and follow a set of bureaucratic procedures when administering their duties. This is not necessarily a negative reflection of their organization or process, but an important characterization of their function. They follow their rules, the Army's regulations, adapt slowly to changes, and rarely approve exceptions.

### 5.3. Battalion

The battalion is the smallest Army unit that maintains a staff and produces training guidance. At this level, all the higher resources and guidance combine to produce specific and actionable direction for companies. Battalions give training reports to brigade as well as capability. They give training guidance, regulations, funding, training resources, and missions to companies. Battalion commanders are rated by the brigade commander and rate companies based in part off their reports and expressions of capability through mission execution.

### 5.4. Companies

The company is the smallest unit in the Army with a commander and a METL. All companies receive similar value flows from their higher echelons. They receive training guidance,

regulations, funding, training resources, and missions from their battalion. Brigades also give them training guidance and resources. Three of the five companies within the battalion (infantry companies) have identical structure and function and therefore compete for the same missions. The other two companies (Weapons Company and HHC) have more specialized functions and tend to receive missions more tailored to their specific function. Most training is planned, conducted, and assessed at the company level. At this level the propensity of effort focuses on establishing the foundations of tactical proficiency. Battalion commanders rate company commanders on their unit's performance while brigade commanders senior rate the companies.

### 5.5. Value Flow Relationships and Impacts

While the individual flows and relationships have great importance and help bring a better understanding to the system, the coupling and loops of these flows offers a greater appreciation for the true challenges present. Two of these loops warrant a more detailed discussion. The first loop involves the relationship between companies and battalions. The second loop adds an additionally layer by describing an extension of this interaction between battalions and brigade.

The first loop, shown in figure 12, flows from the companies' training reports (1), to missions (2), to capability (3), and ends with rating/ranking (4). By examining the end state, rating/ranking, the implications of this value loop become apparent.

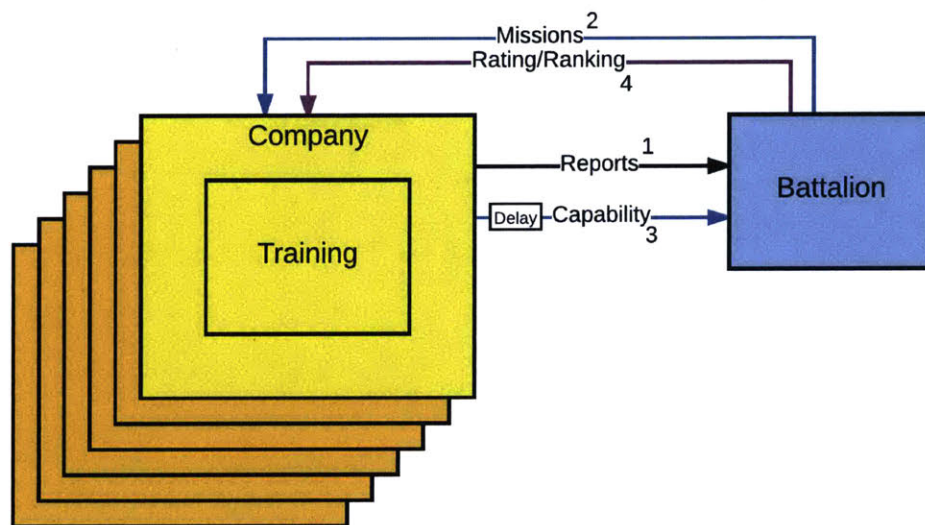


Figure 12: SVN Loop 1 - Inflated Company Reports

Companies understand that they compete against all other companies in their battalion for rating/ranking (4). This rating comes primarily from the unit's performance or expressed capability

## A Systems Analysis of the Army's Tactical Evaluation Process

(3) and the unit's reports (1). Capability expression primarily happens when a unit receives a mission (2) to perform. Missions tend to vary in importance and difficulty and are assigned based off perceived training status and reports. Because of this circular relationship, units are incentivized to inflate their training status reports (1) to compete for the "best" missions and have the best opportunity to boost their eventual rating. While this might explain some of understood report inflation, the battalion's interactions with higher adds another layer of complication.

Because battalion level training happens infrequently, battalion training readiness reports often become an interpretation of the aggregated company reports. This dynamic, exacerbated by the senior rating of company commanders by brigade commanders, creates an organizational obstacle to accurate reporting. Companies report their training readiness to the battalion. Battalion understanding that company reports feed the battalion reports and the implications of battalion reports to brigade, influences the company commanders to report inflated scores. Battalion collects the companies inflated scores (1) and aggregates them to make the battalion training readiness report (2) for brigade. These reports help drive mission allocation (3) and the eventual expression of tactical capability (4). At the end of the respective commander's rating period they receive their rating (5), heavily influence by their tactical performance or capability (4).

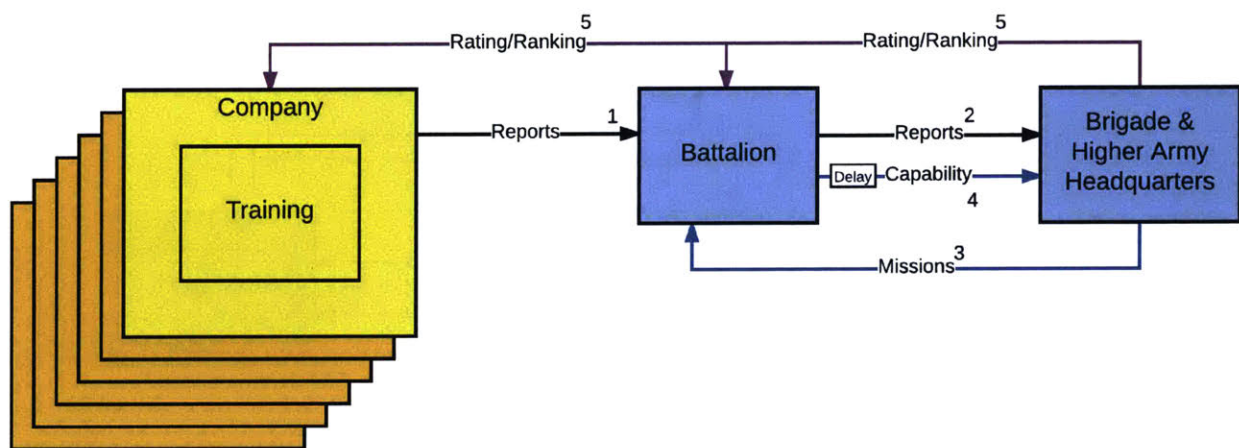


Figure 13: SVN Loop 2 - Inflated Battalion Reports

When examined jointly, the battalion and company reports play significant roles in each loop. The companies' self-incentive seems intensified by the battalion's similar situation. While it may be interesting to look at possible effects at the next higher echelon, the tactical reports tend to be so heavily influenced at the lower levels as to not need additional inflation at higher levels to meet acceptable standards. Additionally, the scale at which tactical readiness reports are delivered

offers little variability, which may add to this dilemma in a zero-defeat type organization. Any efforts to distribute the gathering of information to parcel future missions would have the effect of reducing these incentive loops. We see this with highly active commands which spend extensive organizational energy monitoring and evaluating their subordinate units through means besides the subordinate units' reports. For example, battalion commanders that spend a lot of time inspecting and watching company level training have additional information to compare the company's reports against. This reduces the possibility of receiving, or accepting, inflated reports from the subordinate company. While this situation does not mean that the battalion's reports will be more accurate, it does affect the first loop. Battalions in this situation may continue to inflate their reports for several other reasons – competition, unit perception, the battalion's future missions, ect.

While this analysis may seem simplistic or simply not possible for an organization that holds the values of integrity and honor as core to their proper function, it was found as common place. Interviews with previous company commanders frequently included an instance of this interaction. In each example, a company commander was pressured by the battalion to change an initially reported training readiness data point to a more satisfactory figure. In most cases this would be to the effect of changing of a P (practiced) to a T- (trained) on a company METL task. This fairly common phenomenon raises concern because these reports are meant to influence future military decision making. A 2015 report found that at a certain level, staff officers recognized the data subordinate units passed failed to accurately represent the current training status and often neglected it (Wong and Gerras 2015). This scenario means that Army staffs routinely, and knowing, make force projection decisions without the input of tactical readiness reports.

In summary, this analysis finds that companies and battalions maintain incentives to inflate their tactical readiness reports due to the structure of the system. Because units rely so heavily on training reports to influence the assignment of missions and these missions impact future ratings, reports often fail to capture the true level of training readiness. The battalion seems particularly impactful in this system as they sway both identified incentive loops.

## 6. Current Architecture

### 6.1. TEO Analysis

The Army's method to evaluate the tactical training of infantry units relies on a series of processes and the adherence of standards. TEOs are the bedrock of this training process. The TEO acts as the official standard for individual and collective tasks. Underneath the revised evaluation system, TEOs should be used to evaluate collective task performance and combined with manning and task context information to develop a more holistic training readiness metric. This adjustment makes a previously vague process much more defined, but relies on the correct usage of the TEOs. TEOs have not been used consistently during at least the past decade of infantry training. Leader's and evaluators commonly use them for reference, but rarely use them as intended. Trainers habitually adapt the TEO based off experience to better fit a particular training event. In fact, over seven years of practice and interviews of over ten company level trainers failed to find a single example of a unit utilizing the TEOs in whole and as prescribed.

The following analysis aims to characterize the TEOs as a method of evaluation. The collective task, Conduct Support by Fire (Platoon-Company) from Training Circular 3-21.10 – Infantry Rifle Company Collective Task Publication, will serve as a representative task for this analysis. Infantry companies commonly train this task as part of their tactical development because it serves as a component in nearly all tactical maneuvers. By evaluating this task, we can better understand what the Army is currently assessing, providing feedback for, and possibly why trainers do not use the method.

#### 6.1.1. Too long / Cumbersome

One of the most typical comments collected during interviews reference the length of the TEO format. For our reference task, the manual outlines the standards in 22 steps and 5 pages. It is worth noting that when sub-steps are teased out of the manual there are actually 75 steps that should be evaluated for this task. While this may not seem too hefty of a document, we must remember that often this task pairs with one or several other tasks to comprise a single training event. For example, Conduct an Attack in an Urban Area (Platoon-Company) (07-2- 1261) lists 24 supporting collective tasks that could be evaluated in the same exercise. Our reference task lists 10 supporting collective tasks.

<b>SUPPORTING COLLECTIVE TASKS</b>	
<b>Task Number</b>	<b>Task Title</b>
07-2-1324	Conduct Tactical Movement (Platoon-Company)
07-2-1396	Employ Obstacles (Platoon-Company)
07-2-3027	Integrate Direct Fires (Platoon-Company)
07-2-3036	Integrate Indirect Fire Support (Platoon-Company)
07-2-5009	Conduct a Rehearsal (Platoon-Company)
07-2-5027	Conduct Consolidation and Reorganization (Platoon-Company)
07-2-5063	Conduct Composite Risk Management (Platoon-Company)
07-2-6063	Maintain Operations Security (Platoon-Company)
08-2-0003	Treat Casualties
08-2-0004	Evacuate Casualties

*Figure 14: Conduct a Support By Fire - Supporting Collective Tasks*

Each of these collective tasks has an associated TEO that should be used for evaluation. Only 5 of the 10 supporting collective tasks for the reference task are in the same manual. For the tasks within this manual there are 19 additional pages of evaluations. If a trainer referenced all TEOs for the tasks a unit performed in a typical training event they would carry several dozen pages of evaluations, much of which would overlap in subtasks evaluated. In practice, trainers could eliminate the supporting collective tasks and only use the TEOs to evaluate METL tasks in this new detailed method. Even with this adjustment, trainers would consistently evaluate 2-3 tasks for a training exercise and would still require over a dozen pages for every training iteration.

While most interviewed leaders felt the TEOs had extremely helpful information, it seems the length and inability to whittle down the format to a single list of steps to evaluate a large training event makes it burdensome to utilize. They often reference the documents to help them understand how a unit should conduct a task, but stopped short of using the given format.

#### 6.1.2. Binary (Go / No-Go)

Another interesting characteristic of this method is the exclusive use of binary evaluation. Evaluators categorize tasks as complete or not in a "Go/No-Go" demarcation. This style attempts to make the evaluation of a subtask quick and objective. A unit or individual either completed the subtask as required or did not. The forms utilize no other type of scale, but do categorize and prioritize some subtasks as critical or leader tasks. Figure 15, shows that all 75 of the reference tasks evaluations were binary in nature.

### TEO Analysis - Feedback Granularity

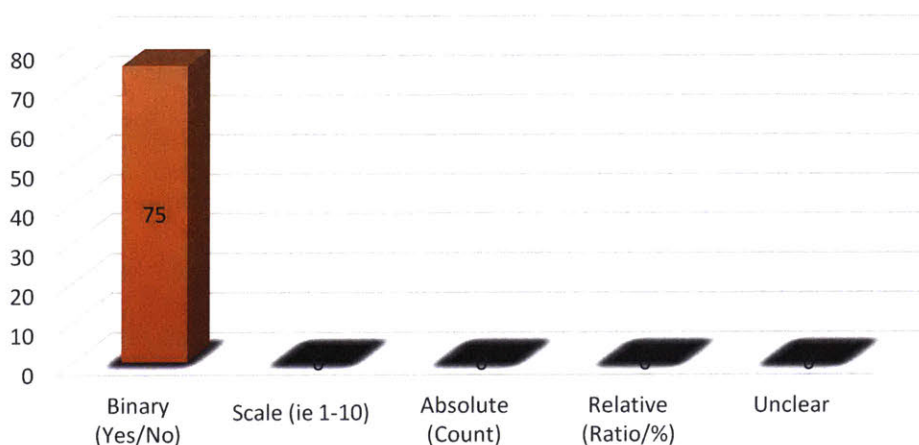


Figure 15: TEO Analysis - Feedback Granularity

While this style of evaluation helps trainers make quick marks, it limits the ability of evaluators to identify those units that perform really well and those that need severe help. The extremes and nuances of proficiency fail to appear. In response to this limited deviation, trainers either consciously or sub-consciously seemed to change the way in which they utilized the form, changing the TEO interpretation in response to available resources.

#### 6.1.3. Objectivity and Subjectivity

The goal of many evaluation methods is to provide objective metrics to help characterize the performance of a unit. The Army understands the need to have standardized and objective means to evaluate its units. TEOs try to utilize objective metrics when possible. Understandably, objective metrics are difficult to utilize due to the changing context and training environment in which most training is conducted. Surprisingly, the TEOs do offer many objective metrics with 61% of the reference task deemed objective. Figure 16 shows the objective/subjective breakdown of the reference task.



### TEO Analysis - Feedback Objectivity

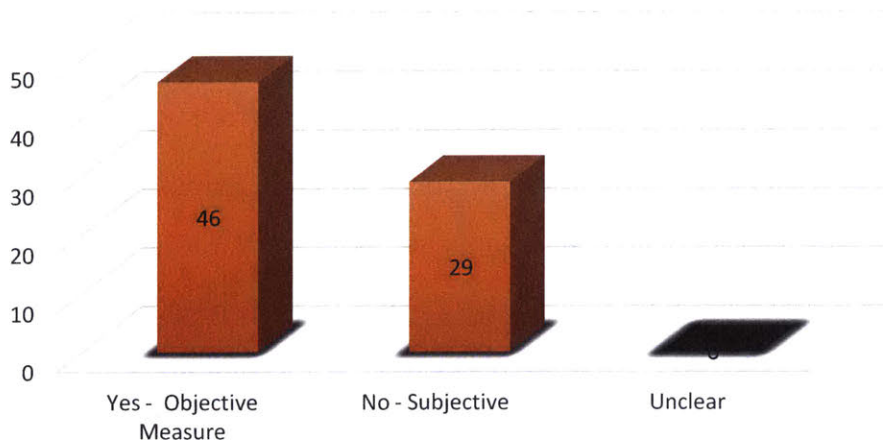


Figure 16: TEO Analysis - Feedback Objectivity

Upon further analysis, it seems that the objective metrics refer to process adherence portions of a task. For example, one subtask of step four questions if the unit “Conducts liaison with maneuver elements to integrate anti-fratricide measures.” This single statement is objective. A unit either did or did not complete this task. However, what is missing is the effectiveness of the liaison from this type of analysis. Nowhere in the TEO does an evaluator have the ability to comment or indicate the effectiveness of this action. By utilizing these types of objective measures, the TEOs reinforce process best practices but fail to seek the effectiveness of a unit’s process employment.

The remaining subjective measures allow trainers to search for effectiveness. In fact, many of the tasks deemed subjective specifically ask if a unit conducted an action “effectively.” This sort of subjective metric helps a trainer key in on the actual performance outcomes of unit rather than simply process adherence. During interviews, these types of questions seemed to be the types of things a senior enlisted trainer focused on – “Was the unit effective?” This finding indicates the Army’s recent campaign for outcomes based training has permeated the force.

Interviewees described a shifting standard that could not be captured in the traditional TEOs. A unit could be a technical “Go” for particular training event, but still need significant improvement in the eyes of the trainer. Leaders described a situation where they changed the interpretation of the TEO standards based upon the time and resources remaining to train. If time and resources existed, the unit continued training regardless of proficiency. Trainers did not waste

## A Systems Analysis of the Army's Tactical Evaluation Process

resources, but simply felt that because the training event had already been planned and resources were available the unit should take advantage of the opportunity to generate as much proficiency as possible.

### 6.1.4. Adherence Based

TEOs tended to focus on process adherence to rather than performance outcomes. When examining the reference task for these feedback types, care was taken to try and tease out those tasks that differed from simple process adherence and focused on process effectiveness. The difference can be extremely subtle, but the implications significant. Process adherence for example might state “unit conducts a rehearsal.” This feedback simply checks if a unit is following the prescribed process without regard to the effectiveness of the unit’s process. Feedback labeled as simply as process focuses more on the effectiveness a process, but not performance outcomes. An example from the reference task would be “unit leader issues clear and concise tasking, orders, and instructions to include ROE.” This is still a process rather than a performance outcome check. Interestingly enough all of the process tasks fall into the subjective category as well.

### TEO Analysis - Feedback Types

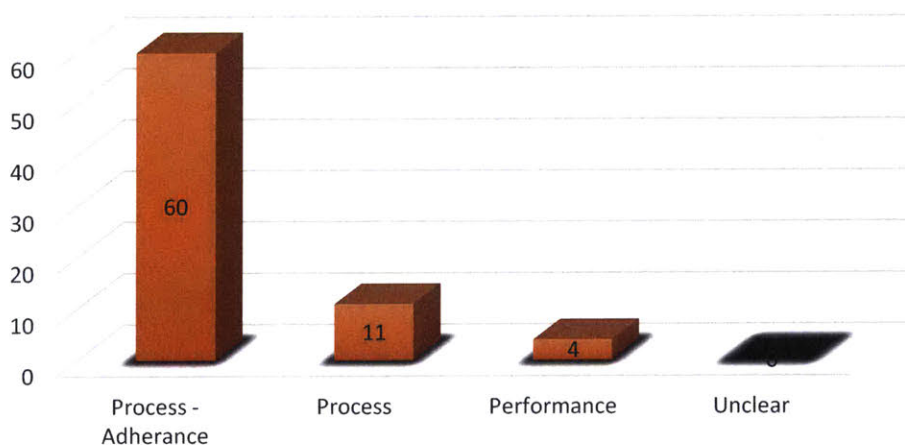


Figure 17: TEO Analysis - Feedback Type

The reference task did offer a few performance outcome feedback prompts that focused on the ability of the unit to engage the enemy. These prompts however all took subjective slants, using phrases such as “acquires, suppresses, and/or destroys identified enemy elements using the appropriate weapon systems.” Nowhere in this method does the evaluator compare objective

## A Systems Analysis of the Army's Tactical Evaluation Process

measures such as engagement accuracy, the distribution of fires, frequency of hits, or movement times.

### 6.1.5. Team/Individual Focus

Interestingly, the reference TEO focuses primarily on leader actions as opposed to team. After each check was evaluated to determine the subject of evaluation, the reference task showed that 69% of these checks focused on feedback for the leader. In effect this suggests that this task is evaluated at as an individual leader task more so than a collective team task.

### TEO Analysis - Feedback Levels

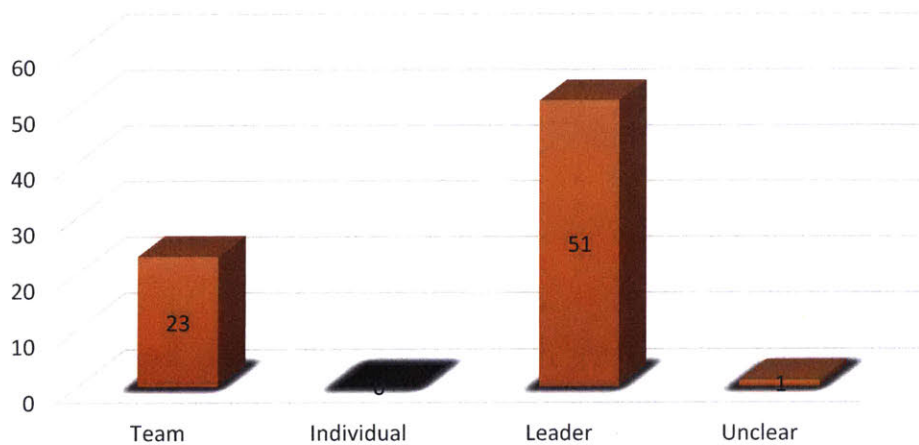


Figure 18: TEO Analysis - Feedback Levels

Upon further investigation, interviews found that the primary focus of trainers remained on collective team action. Trainers tended to focus their attention and comments on collective actions such as communication and the coordination of activity, which can be guided by leader actions. Intriguingly, this suggests that trainers focused on team processes and outcomes to evaluate individual leader performance. For example, a leader may not have to be as vocal or directive if they've trained their soldiers to perform a collective function to satisfaction. Trainers tended only to focus on leader actions when team collective actions failed to produce satisfactory results.

### 6.1.6. Feedback Categorical Focus

To further understand the focus of the reference task, each component was placed into categories according to its competency focus, stage of execution, and warfighting function. The assumption of this initial analysis was that TEOs provided a comprehensive and thorough methodology to assess tactical units. Categorical bucketing would provide a means to demonstrate that the scope and focus of the TEOs covered all areas. Competency focus broke down into five categories – Plan, Shoot, Move, Communicate, and Render-Aid. Infantry units often split their training focuses into competencies. These categories take many forms but often include a variation of the chosen categories. Stage of execution comes from a segregation in the TEO method. Each TEO breaks its subtasks into these categories. ADP 3-0 states that “a warfighting function is a group of tasks and systems (people, organizations, information, and processes) united by a common purpose that commanders use to accomplish missions” (“ADP 3-0: Operations” 2016). The Army commonly uses these warfighting functions to help with planning and operational employments. These three categorical types provide additional context to the focus of the reference TEO.

Analysis within the competency focuses found that nearly 71% of feedback revolved around planning. This type of feedback offers itself easily to assessment, can be objective (process-adherence), and tends to be conducted by only one person (leaders). While not a surprise, the lopsidedness of the assessment does draw some concern. Additionally, trainers usually use this TEO for large scale, resource intensive, soldier intensive, and time intensive training events. Planning, and its assessment could primarily be conducted in other types of environments with less personnel and resource requirements.

### TEO Analysis - Feedback Competency Focus

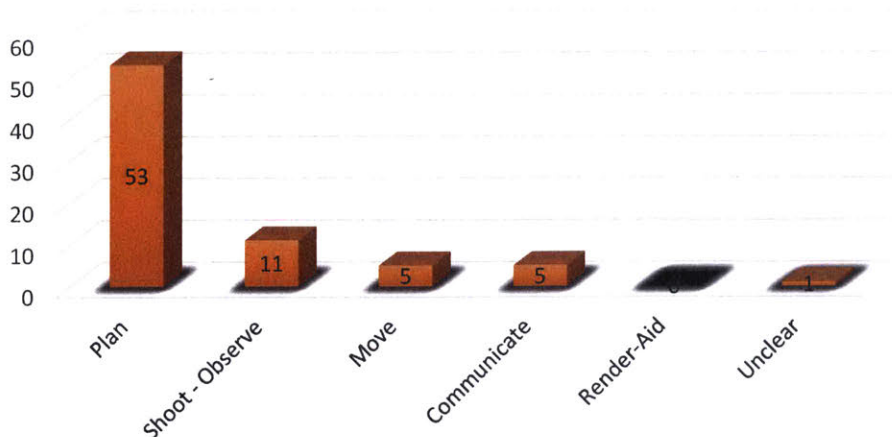


Figure 19: TEO Analysis - Feedback Competency Focus

Figure 20 shows similar results when tasks were categorized by stage of execution. The Plan and Prepare stages primarily contained tasks that fell within the Plan competency focus. Only 25% of tasks occupied the Execution Stage, another indicator of the format's unbalanced emphasis on leader planning tasks.

### TEO Analysis - Stage of Execution

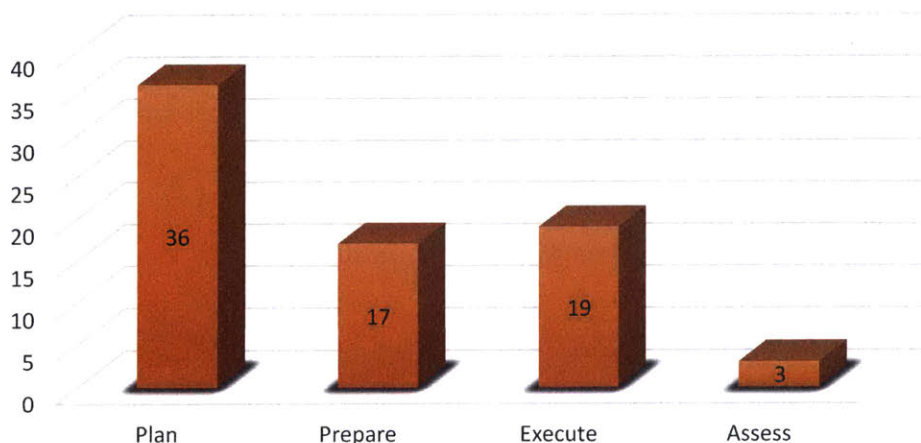


Figure 20: TEO Analysis - Feedback Stage of Execution

Characteristic of previous findings, the Command and Control Warfighting Function dominated all others. Interestingly enough each category did have representation. Feedback tended to focus

on the maneuver of elements and their protection more than the effectiveness of their engagements.

### TEO Analysis - Warfighting Functions

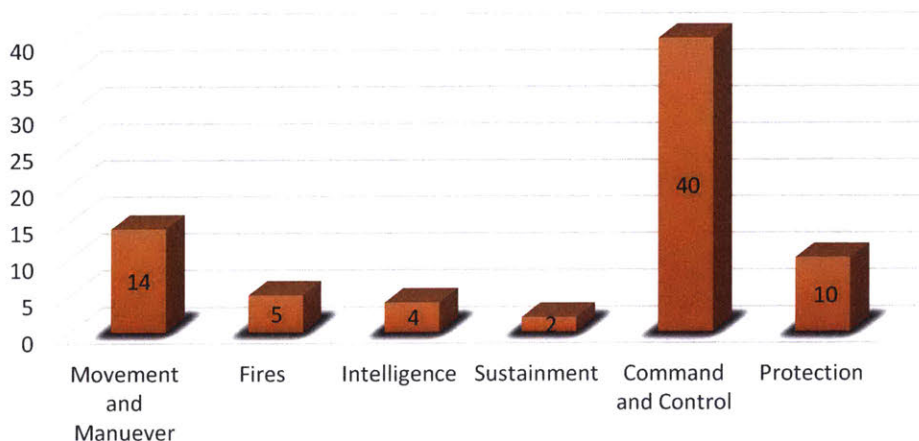


Figure 21: TEO Analysis - Warfighting Functions

In an effort to conceptualize the intended or maybe unintended focus of the TEOs, each subtask was analyzed for the proposed intent. This offered similar results of the previous categorical analysis, but did provide slight nuance in other areas. Figure 22 shows that TEOs are primarily focused on planning abilities, but also have five other focuses. Four of the focuses have nearly equivalent distributions – Control of Units, Security/Protection, Lethality, and Situational Awareness.

## TEO Analysis - What are we trying to measure?

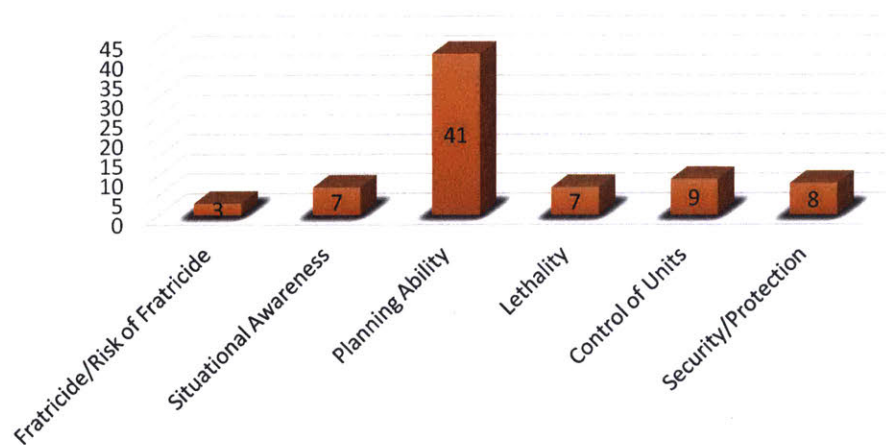


Figure 22: TEO Analysis - Measurement Intent

This final categorical breakdown more closely aligned with the comments received during interviews of infantry trainers. However, the trainers suggested that their feedback tends to be more evenly distributed than the TEOs would suggest. As previously mentioned, trainers tended to focus on team processes and performance first and then look for proximate cause if a deficiency appeared. This type of evaluation or methodology of prioritizing assessments seemed to be aimed at ensuring that units received the feedback they needed as opposed to covering all prescribed categories.

### 6.2. TEO Analysis Summary

Analysis of the TEO reference task suggests that the TEOs primarily focus on process adherence leader tasks to ensure that planning and preparation for operations occurs in accordance with best practices. TEOs do not focus on team processes and tend to avoid performance or outcome feedback. Infantry trainers utilize the TEO as a reference guide to help train units, but do not utilize the form for assessments. Trainers abandon the TEO format in favor of an adhoc and flexible system that focuses on team processes and performance monitoring. Trainers use this method to triage team performance, identifying gaps in capability and then search for explanations and remedies to the recognized deficiency. This change in methodology yields severe inconsistencies between trainers and an inability to compare like units on even the same task.

## 7. System Dynamics Modeling

### 7.1. Overview

System Dynamics is a modeling technique, based heavily on control theory and non-linear dynamics, that has proven effective in simulating complex and dynamic situations (Sterman 2000). Its principles and methods often aim to help policy makers and non-modelers better understand a situation by involving them in the process and breaking down interactions into bite size, almost intuitive, chunks of information. Key facets of its methods are casual loop diagrams and reference modes.

Casual loop diagrams are meant to help capture the variables of interest and their general interactions with each other. This style of diagramming is helpful when refining the studied system, interactions, and influences that drive the observed dynamics. Variables are linked together with simple arrows and positive or negative signs. These arrows depict a direct relationship (input or output) while the sign reflects how that relationship behaves. For example, if Variable A is linked with variable B with a positively assigned arrow, this denotes that increases in A will have an associated increase in Variable B (Figure 22). You'll notice in this diagram that creates a loop, where Variable B also effects Variable A, but in a negative fashion. This is known as a balancing loop, also commonly called negative or self-correcting loop. These loops tend to counteract change (Sterman 2000). While many loops developed during casual loop diagramming will be balancing loops, modelers will encounter some loops that amplify actions. These loops are called reinforcing or positive loops. Figure 24 shows a reinforcing loop, where Variable A increases Variable B, which in turn increases Variable A. These types of interactions can cause tremendous growth/decline and instability in systems. Peter Senge does a brilliant explanation of this phenomenon, explaining how positive word of mouth can drive

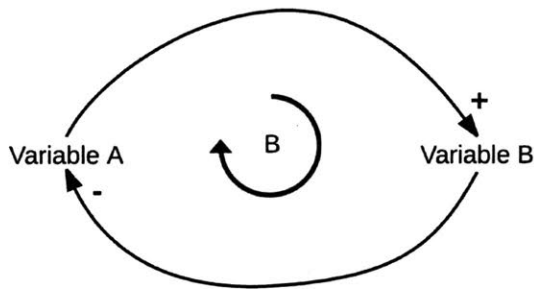


Figure 23: Balancing Loop

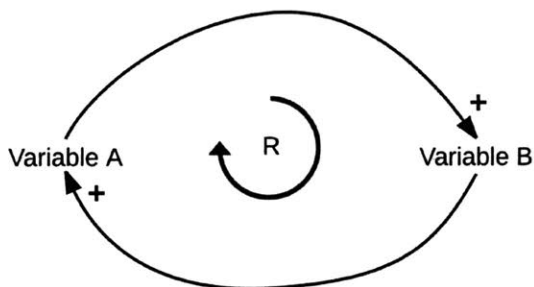


Figure 24: Reinforcing Loop

or output) while the sign reflects how that relationship behaves. For example, if Variable A is linked with variable B with a positively assigned arrow, this denotes that increases in A will have an associated increase in Variable B (Figure 22). You'll notice in this diagram that creates a loop, where Variable B also effects Variable A, but in a negative fashion. This is known as a balancing loop,

also commonly called negative or self-correcting loop. These loops tend to counteract change (Sterman 2000). While many loops developed during casual loop diagramming will be balancing loops, modelers will encounter some loops that amplify actions. These loops are called reinforcing

or positive loops. Figure 24 shows a reinforcing loop, where Variable A increases Variable B, which in turn increases Variable A. These types of interactions can cause tremendous growth/decline and instability in systems. Peter Senge does a brilliant explanation of this phenomenon, explaining how positive word of mouth can drive



sales, create more satisfied customers, and increase the positive word of mouth (Senge 2006). This example illustrates a beneficial reinforcing loop, though one can image several destructive reinforcing loops.

Reference modes are “a set of graphs and other descriptive data showing the development of the problem over time (Sterman 2000).” These graphs show the direct relationships between variables or the development of a variable across time. They are helpful because they allow us to describe interactions in piecemeal or general trends. By breaking a system apart into these reference modes, we are better able to understand the dynamics of a system, build confidence in our modeling, and identify those subcomponents of a system that drive the variables of interest. These reference modes are often derived from literature, interviews with stakeholders, historic data, or empirical experiments.

## 7.2. Model Scope

For this work, the scope of the model has been limited to allow adequate detail and rigor in analysis. The model seeks to capture the relevant dynamics involved while an infantry company conducts a 72-hour team training event. This type of training event and time scale is characteristic within the infantry for team, squad, platoon, and company live fire exercises. These live fire exercises often serve as a certifying or culminating event for that team size. The viewpoint of the model focuses on the development of capability of the team, the evaluator, and the training methodology employed. The following sections will explain the base model and reference modes that were utilized to conduct analysis. The full model and variable calculations can be found in the Annex.

## 7.3. B1: Train the Task

When a team conducts training, it is investing in capability. This capability builds and evaluators are able to perceive this building of capability and decipher a gap between the desired state and perception. Any gap spurs the increase in training intensity and continued investment in capability until the gap is closed.

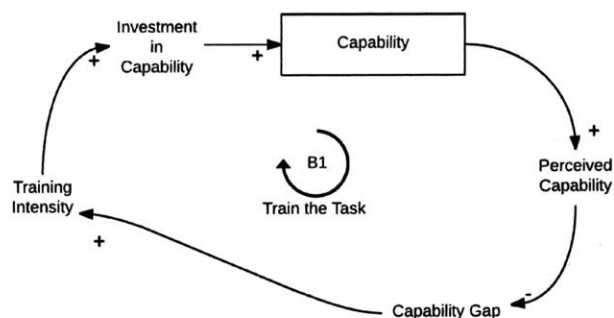


Figure 25: B1 - Train the Task

This is a classic balancing loop and the typical balancing loop that dominates a practitioner's initial thoughts on team training.

#### 7.4. B2: More Effort

As a unit conducts training and builds capability they face several obstacles to their continued development. One of these obstacles is that effort is not linear. As is characteristically found in sports and academics, the amount of effort needed to move from novice to intermediate levels of proficiency is drastically smaller than the level of effort needed to move from intermediate to master levels of proficiency. Chi outlines this phenomenon in her explanation of experts, their development, and the difference between novices. She notes that there are even instances where no amount of effort can transform the intermediate performer to an expert (Chi 2006).

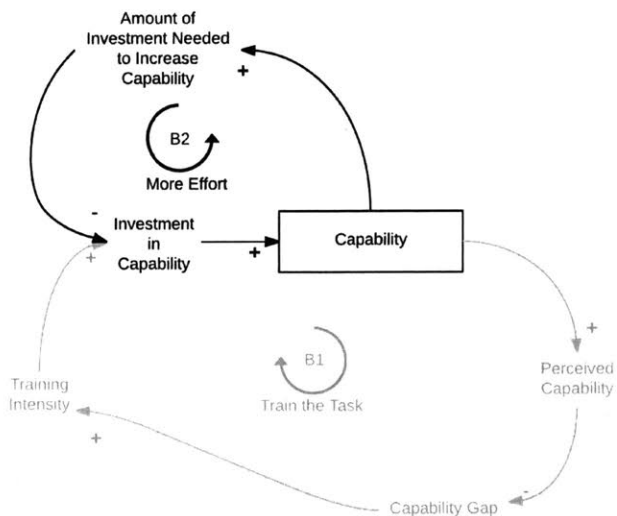


Figure 26: B2 - More Effort

#### 7.5. B3: Learn to Train

During the course of training, the evaluator gains additional experience which increases their capacity. This in turn increases their ability to perceive capability, decreases the capability gap, and reduces the training intensity. This interaction forms the bases of the Learn to Train balancing loop. While the relative experience gained during a training exercise is small when compared to the

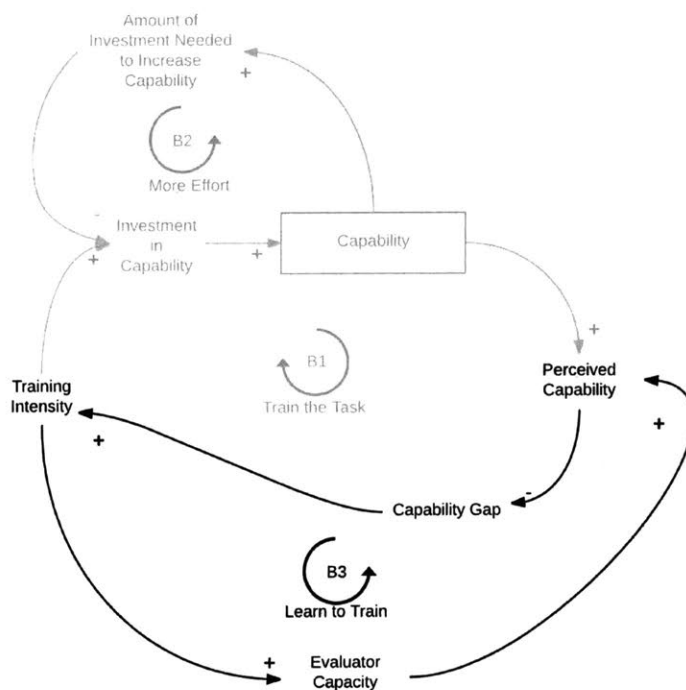


Figure 27: B3 - Learn to Train

## A Systems Analysis of the Army's Tactical Evaluation Process

entire experience of a career infantryman, it does matter and have noticeable impacts on the training exercise.

### 7.6. R1: Evaluator Fatigue

An often-encountered side effect of training is fatigue (Figure 28). The traditional view point looks at this from a team's ability to develop capability, but there is another element of the system that is susceptible to fatigue – the evaluator. This has real implications within the infantry training environment as teams often get breaks between interactions, but the evaluator often remains unchanged. As training intensity increases this increases evaluator fatigue, which in turn reduces the evaluators capacity, reduces the perceived capability, increases the capability gap, and ultimately increases the training intensity. This is the only reinforcing loop within the system.

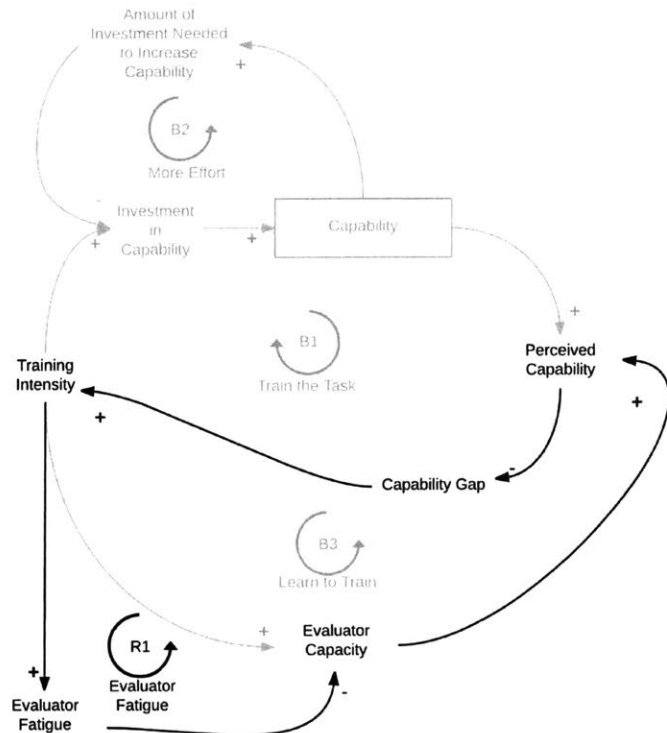


Figure 28: R1 - Evaluator Fatigue

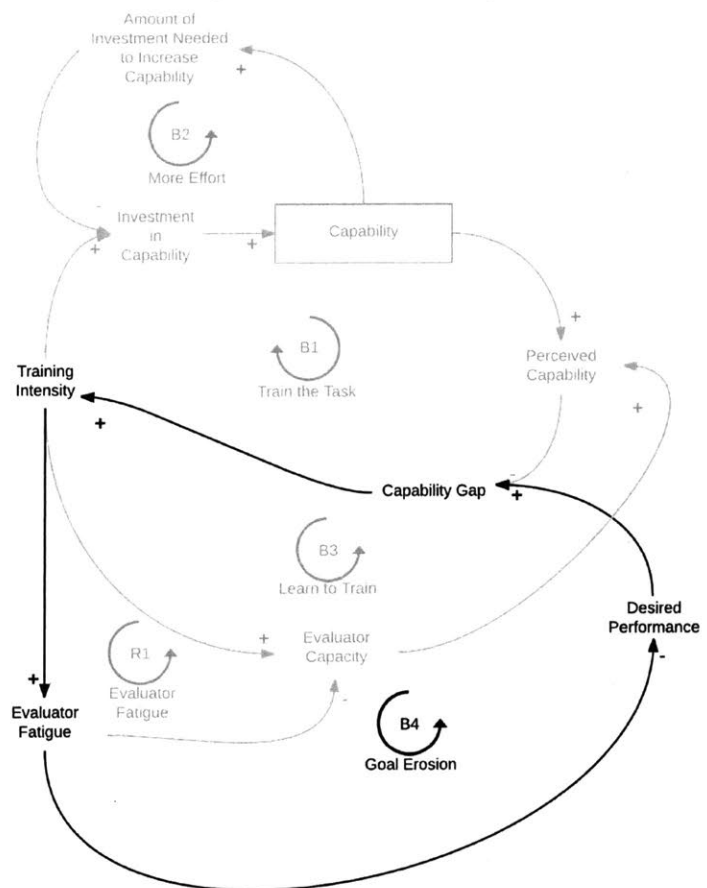


Figure 29: B4 - Goal Erosion

intensity, and balances their fatigue (Figure 29). This is a classic phenomenon that has been documented in several industries. While we tend to think that the Army might be more resilient to these types of standard erosion tendencies, practitioners seem to agree that this reduction in acceptance level is pervasive in most training events. Evaluators often start with very high or elaborate standards and expectations, with a constant recalibration during the training event. In the end, the standard becomes what is safe, and capable for all units to achieve during the given time.

### 7.8. Reference Modes of Interest

The following reference modes are the most important and interesting variable to variable interactions from the model. These are important to characterize because their help create the more dynamic and surprising results from our loop interactions. These interactions often define the strength and timing of a loops dominance.

#### 7.8.1. Effect of Evaluator Capability on Feedback Quality



Figure 30: Effect of Evaluator Capability on Feedback Quality

There is a certain amount of evaluator capability that is solely focused on ensuring a safe training event. This level of capability yields no productive feedback to the training unit. Once the evaluator has surpassed this threshold, they are

able to devote attention to evaluating and providing training feedback to the unit. Feedback at this level and beyond has additional benefit to a training units development of capability. The relationship remains linear because there was no relevant literature found that mapped the effect of this relationship within the tactical context. Chi does provide examples and explanation as for the general relationship we expect to see. Evaluators with more experience should develop more expertise and therefore be able to provide more valuable feedback (Chi 2006).

### 7.8.2. Experience of Evaluators and Capability of Evaluators

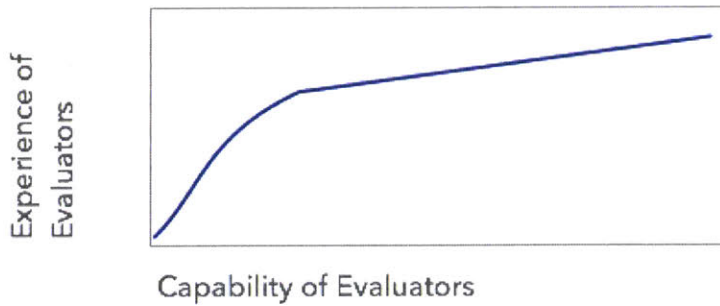


Figure 31: Effect of Evaluator Experience on Capability of Evaluators

The more experience an evaluator gains while evaluating a particular task, the better they become at evaluating this task. In this model we expect only a slight improvement in the evaluator’s capability due to the time of

exposure and the amount of deviation within the exercise. However we do see diminished returns as experience, in solitude, often fails to produce true experts. Chi describes this development of experts and the need for a subject to not only gain experience, but have exposure to the right experiences and feedback to develop the highest levels of capability (Chi 2006). Evaluators make their gains early on and then fall in a “groove.” This becomes exemplified in a more traditional learning curve shaped reference mode.

### 7.8.3. Effect of Fatigue of Evaluators on Evaluator Capability

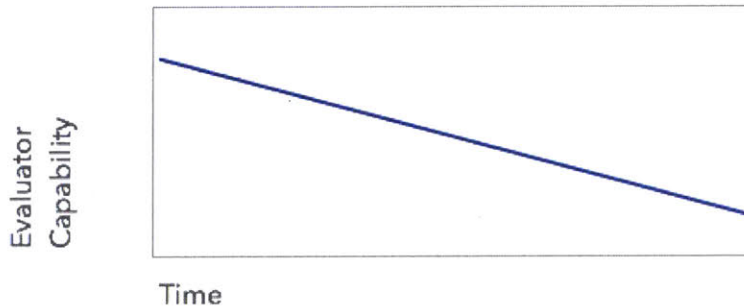


Figure 32: Effect of Time on Evaluator Capability

A group out of the Sandia National Lab explored several models that were developed to show the effects of fatigue on performance. Studies have shown a general cap in this timeframe (0-100hrs) of around 20% efficiency. Based off their

analysis it was determined that Neurobehavior Model (Jewett & Kronauer) was best as

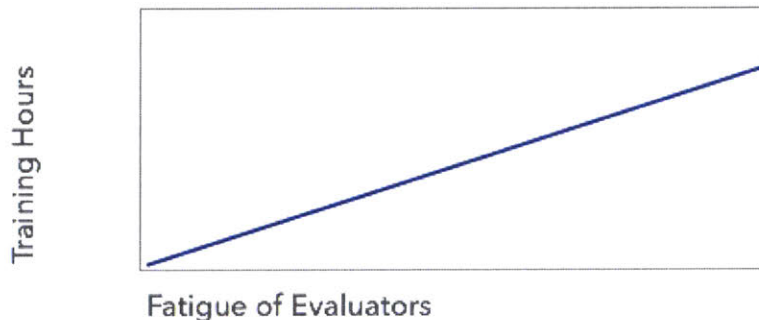


Figure 33: Effect of Training Hours on Evaluator Fatigue

approximating tasks similar to this simulation (Lawton, Miller, and Campbell 2005). The final reference mode is a simple linear decrease with performance falling to nearly 20% by 72 hours of continuous effort (Figure 32). As

the evaluator continues training in this simulation, they accumulate fatigue from continued work and low or no rest cycles (Figure 33). Previous analysis suggest this simple model is best at approximating fatigue and performance capability within this simulation's time domain (Lawton, Miller, and Campbell 2005).

#### 7.8.4. Perception Gap on Quality of Feedback

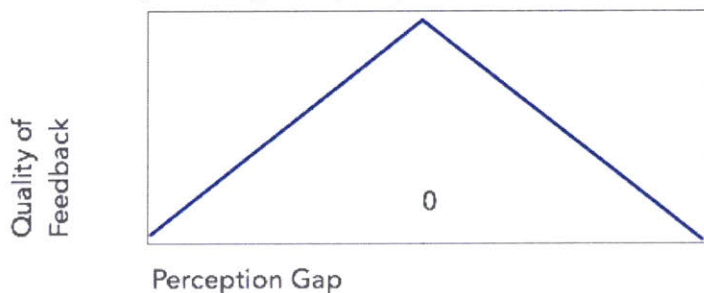


Figure 34: Effect of Quality of Feedback on Perception Gap

The closer an evaluator understands the true capability of a team the better the quality of the feedback they can present. The more noise or divergent their perception of the team's capability, the less effect their feedback will have on team

performance. For example, when a baseball coach fully understands the deficiencies of a player's swing, they are able to provide highly effective cues and guidance to that player to improve their swing. If the coach does not understand the deficiencies, the perception gap is high. This has a direct effect on the ability of the coach to provide useful feedback to a player. This relationship is captured in this simple reference mode where the maximum possible quality of feedback appears when the perception gap is zero. Any movement, positive or negative, in perception gap has a negative effect on the quality of feedback.

#### 7.8.5. Capability on Perception Gap

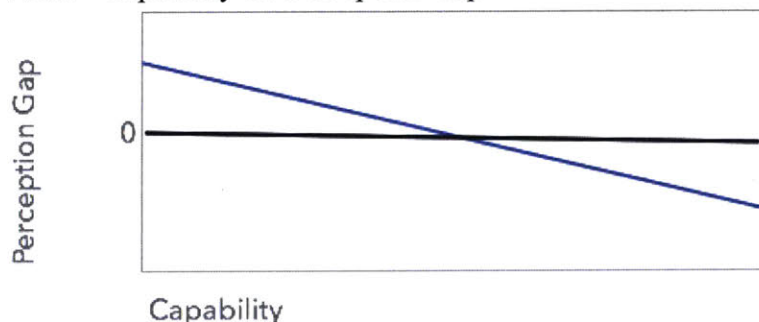


Figure 35: Effect of Capability on Perception Gap

Evaluators have more difficulty understanding the fringes of performance. When a team performs extremely poor it is difficult to tell what is holding the team back. It may also be difficult to understand their performance

because of the high interdependency of tasks. If one teammate performs especially bad, the team's poor performance may be attributed to this actor. On the other end, performances of higher caliber

tend to be hard to distinguish between. At this level, the nuisances of performance are too small for most evaluators to distinguish. This leads to the tendency for evaluators to skew their performances towards the average. They over estimate underperformers and under estimate those truly spectacular teams.

### 7.8.6. Work and Schedule Pressure

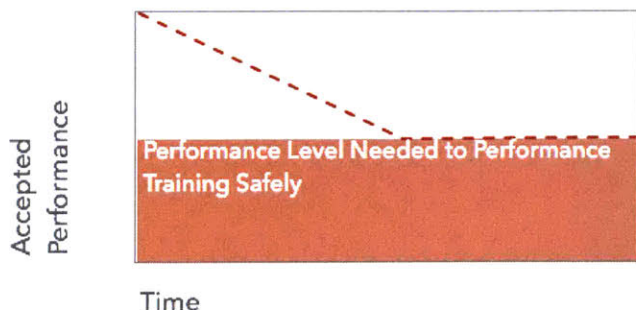


Figure 36: Effect of Time on Accepted Performance

Previous work in the service industry has demonstrated the effects of work pressure on time and quality standards of work (Sterman 2000; Oliva 2001). In line with this analysis, practitioners expressed a general trend to adjust standards based off of time available and general fatigue. The less time

available to train and the higher the fatigue, the lower the accepted performance. It is worth noting that there is a level at which no evaluator will accept performance because it is unsafe to conduct the training. This relationship is exemplified in Figure 36, with the floor level of performance set at 50.

### 7.8.7. Effect of Augmentation on Evaluator Capacity

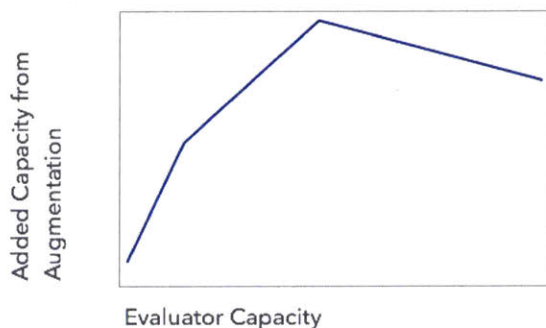


Figure 37: Effect of Augmentation on Evaluator Capacity

Perception augmentation (adding data, sensors, and objective measures) does not affect all capabilities of evaluators the same. Novices cannot leverage information in the same manner as experts. Experts will utilize the additional information in a more efficient manner, discarding the bits of information that provide no added benefit. Additionally, experts may not need to see the data to be able to provide

excellent feedback - so it won't improve their performance as much as someone in the middle (Chi 2006; Gilhooly et al. 1997). This produces a unique reference mode where evaluators with average capacity gain the most from augmentation while novice evaluators gain very little and experts actually gain less than average users.

### 7.9. Modeling Moral Intent

Within the model, Moral Intent is the propensity of a unit to inflate a report of capability from their perceived capability. Moral Intensity is controlled by many things within the tactical training context. The placement of the evaluator (self-evaluation, 3<sup>rd</sup> Party – multiple levels), reporting standardization, objective metrics and sampling through augmentation, opportunity for retraining, and organizational expectations all have serious impacts on a unit's moral intensity. Because of the influence, but different placement within the system, a battalion's moral intensity has been modeled separately from the company's. The company has the first cut at a report, providing the initial report for the battalion. This initial report has already been subject to the moral intensity of the company leadership. It is then subject to additional vetting and approval from the battalion. Practitioners described several instances where this interaction took place. A company would provide slides for a pending meeting to brigade and receive comments/suggestions from battalion. If a company had reported a capability that might be an outlier or paint the unit in a bad light, battalion highly suggested that the report be changed. In some instances, this was not a request. This structuring of moral intensity allows us to better understand these dynamics and its effect on a company's final capability report.

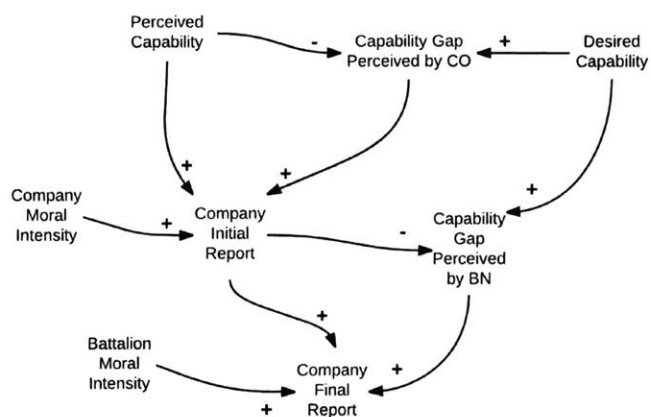


Figure 38: Modeling Moral Intent

to brigade and receive comments/suggestions from battalion. If a company had reported a capability that might be an outlier or paint the unit in a bad light, battalion highly suggested that the report be changed. In some instances, this was not a request. This structuring of moral intensity allows us to better understand these dynamics and its effect on a company's final capability report.

### 7.10. Sensitivity Analysis

A sensitivity analysis is a common technique used to both validate parameter bounds and test interaction assumptions and gains. To help provide a better understanding of the model limits and usable alternatives, each variable was interrogated across a spectrum of values while holding all other variables constant. This was done to test model response both within the model development stage and again for a sensitivity analysis.

During the course of sensitivity analysis, Feedback Format and Augmentation demonstrated interesting interactions. Feedback Format displayed highly sensitive features, but in a linear response. High levels of Feedback Format developed higher levels of capability. This



response ended up mitigated by the More Effort balancing loop. At high levels of Feedback Format, the rate of capability development capped and was no longer effected by increases in Feedback Format. While this intuitively made sense, there is a limit to training and development growth regardless of feedback, the sensitivity of the feature suggested we should treat it with only mild adjustments to capture realistic changes.

Augmentation exhibited similar capped features. At the high ends of Evaluator Initial Capacity, Augmentation effects had no significant effect. Augmentation had the greatest effect on average evaluators, with diminishing returns as evaluators increased capacity. These two variables also displayed a similar cap to the rate of capability development as Feedback Format. This further demonstrated the inherent limits to capability development regardless of condition. No matter how many sensors, how experienced the evaluators, or comprehensive the format – teams can only develop so fast. Salas mentions this limits to growth phenomenon with regards to the limits of enhanced feedback, simply that teams exhibited a limit to develop rate regardless of how comprehensive the post exercise feedback (Salas, Nichols, and Driskell 2007; Salas et al. 2012).

Do to the sensitivity of certain features and inherent balancing mechanisms within the model, the tradespace was restricted in terms of parameters tested and levels utilized. This helps ensure that alternatives outside of the evaluated space are not considered as viable. For example, a Feedback Level of 5 (1 being nominal) would simply restrict the model to a capped capability development output and present unrealistic growth. Similarly, Augmentation levels over .2 produce unrealistic effects with today's current capabilities and restrictions.

## **8. Alternative Tactical Training Architectures**

### **8.1. Architecture Generation and Reasoning**

Once the model was established and calibrated, it can be helpful in evaluating different alternatives. The alternatives came from first evaluating the possible levers available for tactical evaluation. From research and interviews, several hooks had been inserted into the model to help capture these different features/levers. A SWOT (Strengths, Weaknesses, Opportunities, and Traps) assessment on the current architecture greatly helped focus the development of features and their levels.

The current system has several strengths. Evaluators and the Army understand the current process and are comfortable implementing at least their version of it. Generational learning has ensured that the training the Army conducts produces effective, adaptable, and highly capable teams. The current environment also allows commanders significant flexibility in both developing a training program and evaluating their units. The system has historically treated the Army well.

The system also exhibits several weaknesses. Report and feedback quality, consistency, and quantity vary greatly between units and evaluators. Inflation, or the tendency of units to report high levels of capability, have made capability reports less useful or completely useless for higher level unit's planning or assessment. Maybe even more concerning is the lack of a consistent baseline with which to compare units. There is currently no realistic method to assess a unit in their current configuration and say they are better, or worse, than a unit from Vietnam – despite the amazing advances in technology. The idea that the Army's tactical units have continued to improve is simply a subjective assessment or hunch. Even at the most basic level these units have difficulty assessing changes. For example, recently the infantry changed their primary weapon system, M4s, to have heavier barrels and exchange three round burst functionalities for fully-automatic. This change came with little or no recommendation for how to utilize this different functionality. There were serious debates about how to use this new function within a tactical unit. Could it replace the Squad Automatic Weapon? When does fully-automatic make sense? Are higher ammo consumption rates sustainable for expeditionary maneuvers? Because units had no assessment method to test their tactics with this new feature, most units simply suggested no one use fully-automatic. Human evaluators have limitations. Interviews suggest evaluators operate at full utilization and prioritize safety over assessment/feedback. It is also worth reiterating the effects

of fatigue. It was not uncommon to find examples of extreme evaluator fatigue during a training event – limited sleep, high iteration count, lots of walking, etc. It was also noted that evaluators often have a difficult time providing comparative feedback. A unit could complete one iteration, receive feedback, conduct retraining, and upon their subsequent iteration the evaluator had a difficult time comparing their performances to see if the feedback and retraining was effective. In general, this complete reliance on humans produces extreme variance in the consistency and quality of feedback.

Opportunities within the tactical training environment abound. The range facilities currently rarely provide any assistance to evaluators through data collection. The ability to collect this data for utilization proves a significant opportunity that sports and aviation teams have recently leveraged. GPS watches, heart rate monitors, smart phones, and the internet of things are continually improving and becoming a normal part of a soldier's everyday life. Soldiers are more prepared now to integrate data and technology into the training environment than ever before. Additionally, there are already different methods of evaluation emerging from the Army. The National Training Center has codified a version of competency modeling, commonly called the "Big 10." Competency modeling is a flexible evaluation method that has significant academic and practical roots (Campion et al. 2011). Interviews found that several units had developed similar methods for use in their training.

Attempting change to this system poses threats as well. The most poignant concern is that inserting technology or new methods will yield lower capability due to the additional effort needed in learning and developing the change. This is commonly known as the capability trap in system dynamics (Repenning and Sterman 2016). Poor outcomes or performance at the onset of a system change can lead to scrapping adoption prior to realization of the true capability of the new system. Additionally, some system architectures can restructure the established value flows and upset the organizational power balance. People typically fear change and will attempt to maintain the status quo unless they can be assured of improvement or sustainment of their political power.

### 8.2. Levers and the Morphological Matrix

The SWOT analysis of the established system as well as previous research led to the development and identification of four variables or levers that could be utilized to adjust the system

## A Systems Analysis of the Army's Tactical Evaluation Process

and develop different system architectures. The four variables are Feedback Format, Initial Evaluator Experience, Evaluator Organizational Placement, and Perception Augmentation.

Feedback Format represents the effectiveness of a particular format in terms of helping a unit improve their functioning and develop more capability. Currently, the format is highly variable between units and more contingent on environment context and generational learning than doctrine. Within the model, this baseline value is 1. Forcing the usage of the TEOs would actually drop the effectiveness of feedback. The TEO has been abandoned by practitioners for a reason and forcing the usage would distract from what generational learning has deemed most effective. Forcing its usage however does have immediate effects on the moral intensity of units. The standardization of a reporting format should lower a unit's propensity to inflate reports because they can reasonably assume that all units are using the same standards of evaluation. The other option would codify a new feedback format that more closely mimics the outcome of the Army's generational learning. This would increase the format's effectiveness, through a combination of best practices, and reduce moral intensity through standardization as previously discussed. The development of this format would be burdensome, but provide increased effectiveness of feedback and be more easily implemented than the abandoned format.

Initial Evaluator Experience represents the general level of expertise that an evaluator has at the onset of a training exercise. It was assumed that battalion commanders and command sergeants major would be considered experts (twenty years' experience). Their amount of experience in a particular task was estimated in training hours (72 hours of experience a year) and then used as reference. The baseline experience level for a company was six years, or 6/20 of the reference expert level of experience. Adding evaluator training would increase the experience level a marginal amount. In this scenario, one additional year of experience was assumed. If the infantry implemented a master trainer type program, it was assumed to add two years' experience to the evaluator. These values would be expected to fluctuate depending on the quality and duration of study and are intended to be representative in this work.

Evaluator Organizational Placement represents the effect of proximity on moral intensity (Jones 1991; Jones and Ryan 1998). Evaluators that report on their own unit will have more propensity to inflate a report than someone that is echelons above/away from the training unit. The effect is modeled linearly with a cap at two levels removed. As such, a company level evaluator would have no impact on the baseline moral intensity of a company report, a battalion level

## A Systems Analysis of the Army's Tactical Evaluation Process

evaluator would slightly reduce the moral intensity of a company report, and a brigade or higher level evaluator would have a significant reduction in moral intensity. It is worth noting that implementing this higher-level evaluator faces significant organizational obstacles and may not be feasible do to manning availability and culture.

Perception Augmentation is the objective sampling of a training event through sensors. This variable represents the non-destructive and automatic collection of objective metrics during a training event. Its construction within the model acts as an additive to evaluator capacity. The construction and utilization of this system would have to be thoroughly thought out and investigated to ensure that it added and did not distract from training evaluation. Simply implementing sensors or throwing data at evaluators without training or review could not be assumed to have positive effects. The levels of minimal and moderate augmentation represent the implementation of a vetted and studied system that would either be automatically intuitive to evaluators or need minimal training. The idea of this architectural choice is that sensors could offload the over utilized human evaluator so that they could focus on more complex training interactions or use the data to help compare iteration runs.

By combining each of these architectural choices we can develop a morphological matrix (see Table 1). This matrix represents choices that can be independently made with each variable to develop unique and substantively different system architectures.

Lever	Option 1	Option 2	Option 3
<b>Feedback Format</b> (Low is .8, Baseline is 1, High is 1.2)	Force Usage of TEOs (.8 Feedback Format Quality, -.1 CO and BN Moral Intensity for Standardized Format)	Allow Evaluators to choose Feedback Format (1)	Study and adjust Feedback Format to mimic generational learning (1.2 for Feedback Effectiveness, -.1 CO and BN Moral Intensity for Standardized Format)
<b>Initial Evaluator Experience</b> (Measured in Hours Training Experience with 432 as the current baseline for Company level leaders)	Keep current training programs and focus (432 Hours of Experience)	Build Evaluator Training Program (Increase experience to 504 Hours)	Use Master Trainers - BN/BDE/DIV Level (Increase Experience to 576 Hours)
<b>Evaluator Organizational Placement</b> (0 is Company Level, .1 is BN, .2 is BDE or Higher)	Keep CO Level Evaluators (0)	Develop BN Level Evaluators (.1)	Develop BDE/DIV Level Evaluators (.2)
<b>Perception Augmentation</b> (0 - .2)	No Augmentation (0)	Minimal Augmentation (.1)	Moderate Augmentation (.2)

Table 1: Tactical Evaluation Morphological Matrix

### 8.3. Sampling Possible Tactical Training Architectures

There are two possible architectures that must to analyzed for reference, the baseline and OBJ-T. The baseline represents the status of the tactical evaluation training system prior to implantation of OBJ-T while OBJ-T represents the recently implemented changes of the system. Because of the qualitative nature of this model, all possibilities should be analyzed in reference to these alternatives.

The morphological matrix can be represented as a 4 x 3 matrix. This means that there are  $3^4$  or 81 system architectures possible in this configuration. Unfortunately, simulating all these possibilities would not only be a time consuming and cumbersome, but would be difficult to meaningfully represent. To expedite analysis, a subset of 14 alternatives was selected utilizing the Latin Hypercube methodology. This method essential selects an evenly distributed random sample of architectures to allow for quick analysis before performing more detailed analysis on those architectures that prove promising. These 14 alternates were added to the baseline and OBJ-T alternatives for a total of 16 simulations for initial analysis.

## 9. Tradespace Analysis

### 9.1. Initial Simulations

The sixteen initial conditions were constructed within the model and their associated data captured for each simulation run. The data representing each simulation run was parsed and compiled so that all simulations could be evaluated in a comparative analysis. The following graphs represent the primary graphical tools used to identify trends, interactions, and dominance within the tradespace.

The Capability vs Inflation graph (Figure 39) plots a simulation's final developed capability against its report's inflation. The size of the circle represents the evaluator's organizational placement. The bigger the circle, the further removed the evaluator is from the training unit. For example sim12's evaluator is within the training unit, so it has the smallest sized dot. Sim7 has an evaluator at least two level's removed, so it has the largest sized dot. In general we see horizontal patterning. The small dots (internal evaluators) align generally on the right side of the tradespace and therefore tend to have the highest inflation. Interestingly enough there is no

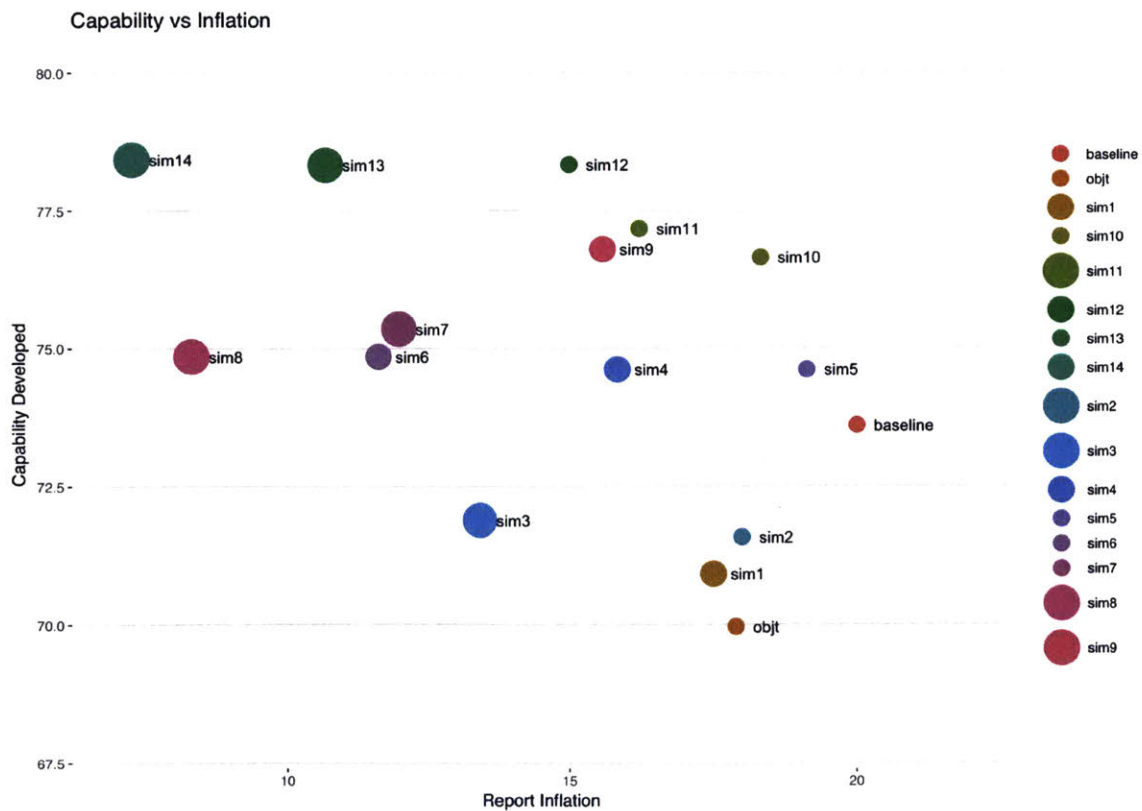


Figure 39: Capability vs Inflation - Tradespace 1

## A Systems Analysis of the Army's Tactical Evaluation Process

vertical patterning in this tradespace, which suggests that evaluator placement has an impact on inflation, but not as much if any bearing on capability development. This aligns with our mental model of the tactical training environment. Evaluator placement shouldn't affect the development of capability, but would have impacts on reporting.

To look at the impacts of evaluator experience, capability developed was plotted against total training hours, with color representing evaluator experience and circle size corresponding to inflation (Figure 40). The most interesting point about this tradespace is the strong negative correlation between capability and training hours. In general, the most efficient evaluation systems develop their capability quickly. The longer a unit spends training, the lower the levels of developed capability. There also seems to be a mild, but noticeable, effect from initial evaluator experience. The trend shows that systems with higher levels of evaluator experience tend to develop the most capability with the most efficiency.

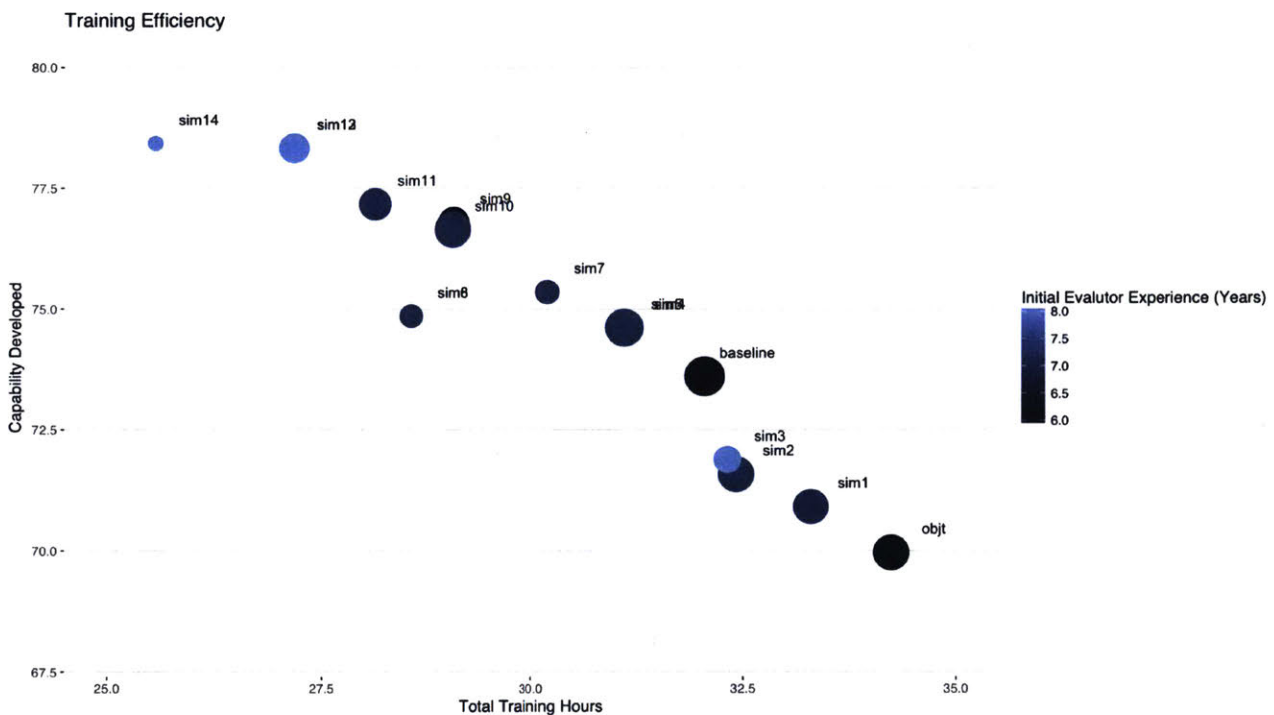


Figure 40: Training Efficiency - Tradespace 2

To help further understand the impacts of augmentation, reported capability was plotted against developed capability with color representing augmentation level and size representing evaluator organizational placement (Figure 41). In this representation, the solid line depicts perfect reporting. Circles closest to this line have less inflation that circles that are farther away. As such, we see that augmentation seems to have a dampening effect on report inflation. Sim5, sim8, and



## A Systems Analysis of the Army's Tactical Evaluation Process

sim14 all have the highest levels of augmentation and are closest to the perfect reporting line. We also notice that evaluator organizational placement has a similar effect in report inflation. Those simulations with internal evaluators (objt, baseline, sim2, sim5, sim10, sim11, and sim12) all report the highest levels of inflation when compared to the other simulations.

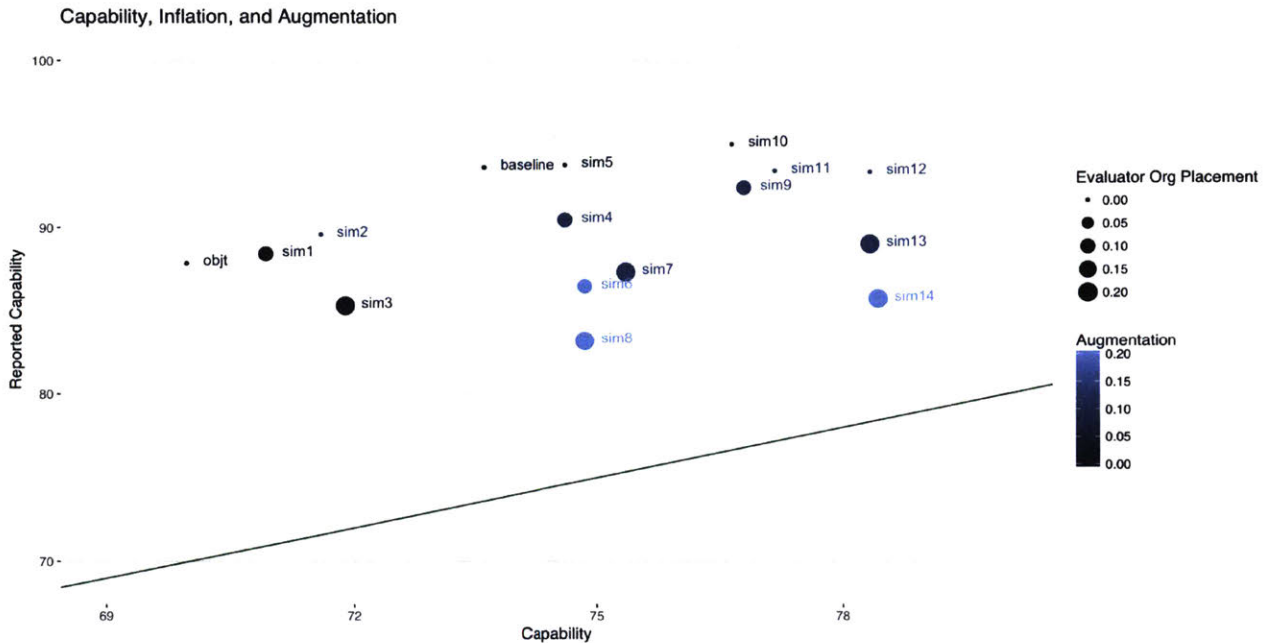


Figure 41: Capability, Inflation, and Augmentation - Tradespace 3

The final comparative analysis emphasizes feedback format (Figure 42). In this representation reported capability is mapped against capability, with size associated with evaluator organizational placement, shape referencing level of augmentation, and color indicating level of feedback format. In this graph, a significant pattern emerges. Feedback format, essentially the effectiveness of feedback on the development of capability, has a positive correlation on both reported capability and capability. This means that the feedback format has a direct interaction on the development of capability while the standardization of the report plays into the amount of inflation. The effect of augmentation clearly expresses itself in this depiction as well, with increasing levels of augmentation bringing simulation points closer to the “perfect” reporting line.

## A Systems Analysis of the Army's Tactical Evaluation Process

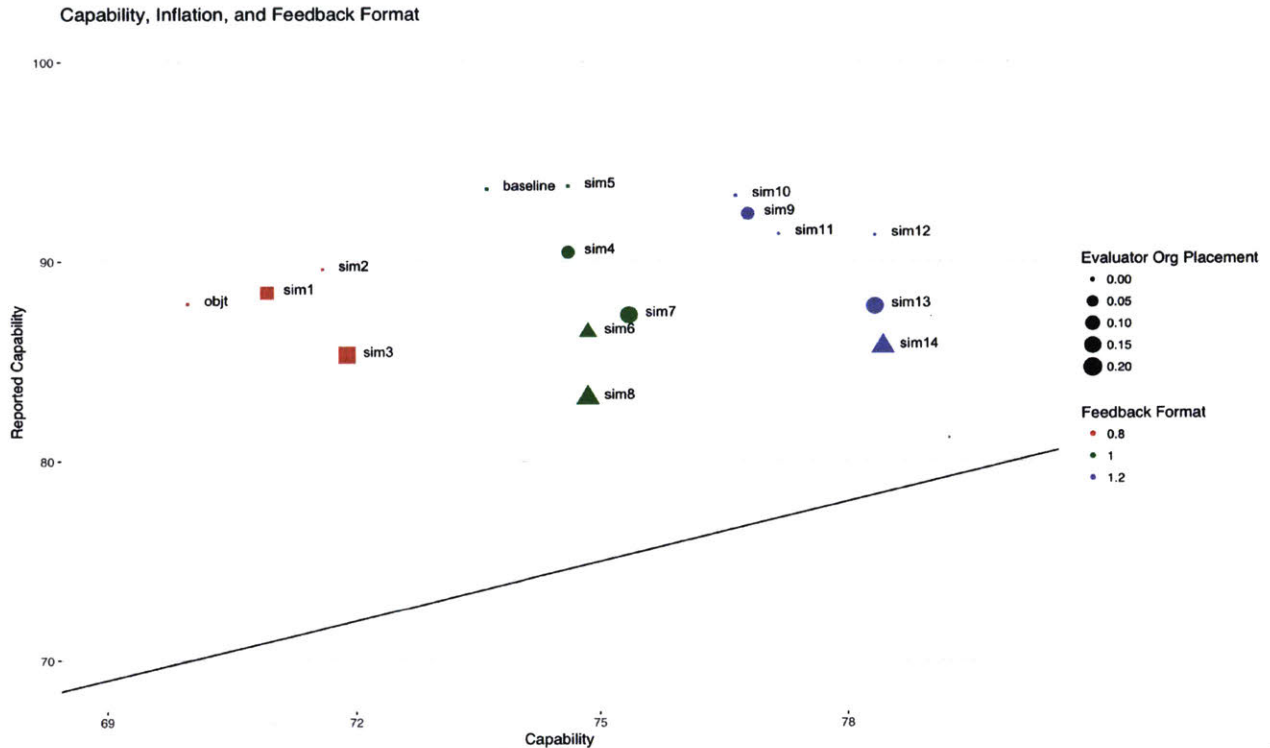


Figure 42: Capability, Inflation, and Feedback Format - Tradespace 4. While all levers showed significant impact on the system, Feedback Format proved the single greatest lever interrogated in the tactical evaluation system. The standardization of reporting and the increased development of capability have a combined effect at reducing inflation.

The initial analysis allows us to draw the following conclusions about our tactical evaluation architectures:

- Augmentation decreases report inflation, with marginal positive impact to capability development
- Evaluator Organizational Placement decreases report inflation – with no impact on capability development
- Initial Evaluator Capability has slight positive effects on capability development but is mitigated by the bounds expressed in this model
- Feedback Format has moderate positive effects on capability development and decreases report inflation

### 9.2. Refinement

Based upon the initial analysis from the sub-setted simulations, another set of simulations was selected to investigate and generate opportunities. The basis of selection was, knowing the

impacts of each of the levers, to identify a portfolio of organizationally acceptable alternatives for the Army to consider. Some of the alternatives face significant organizational obstacles and would not be seriously considered as an initial solution. For example, implementing the highest level of evaluator organization placement (two or high levels removed from training organization) would drastically reduce report inflation without impacts to capability development. However, this option faces significant organizational challenges due to culture and soldier availability. From this organizational understanding the following alternatives emerged.

#### 9.2.1. Basic Alternatives

The following systems were selected as the most basic options that the Army could select from the available options with the aim of minimizing cost and organizational obstacles.

Baseline – (Revert to Generational Learning). This alternative has no standardized feedback format, augmentation, or external evaluation component. It relies on the generational learning to adapt doctrine into the most effective form possible. This is the system prior to OBJ-T.

OBJ – T – (Standardize Format Usage). This configuration seeks to improve reporting by implementing a standardized reporting matrix improving the previously highly subjective interpretations. This reduces report inflation, but slightly reduces the capability developed in a system. The reduction is associated with the adoption of the less effective/abandoned TEO feedback format. It does not utilize augmentation or external evaluators.

Sim15 – (Improved Feedback Format). This alternative would study the gains made from the generational learning of feedback and codify it into a system of standardized reports. This would improve the quality of feedback, generating more capability, and its standardized use would reduce report inflation.

#### 9.2.2. Enhanced Alternatives

The following systems have been classified as enhanced because they involve more aggressive development for implementation, but also provide more value. These alternatives require more than one adjustment from the Baseline or OBJ-T alternatives.

Sim16 – (Baseline Feedback, Evaluator Training, Minimal Augmentation). This option attempts to maximize the benefits of generational learning on feedback format and a small train

the trainer program. These choices will increase capability development while the augmentation will amplify the evaluator effects and mitigate report inflation. Because of the strong influence of feedback format, this is the only option that keeps the generational learning. All remaining options utilize the improved feedback which seeks to capture the lessons learned from the generational learning and codify them into competency modeling formats for standardization.

Sim17 – (Improved Feedback, Minimal Augmentation). In addition to improved feedback, this alternative utilizes minimal augmentation to help evaluators and lessen report inflation.

Sim10 –(Improved Feedback, Evaluator Training). This option provides basic evaluator training, boosting the initial experience level and adding to the gains of improved feedback.

Sim11– (Improved Feedback, Evaluator Training, Minimal Augmentation). Sim11 builds on Sim12 by adding minimal augmentation. This has the effect of marginally improving the evaluator's capability and reducing report inflation.

Sim18 – (Improved Feedback, Master Trainer (Internal)). This alternative explores the impacts of providing master level training to evaluators in conjunction with improved feedback format.

Sim12 – (Improved Feedback, Master Trainer (Internal), Minimal Augmentation). Sim12 adds minimal augmentation to Sim18's parameters, reducing report inflation and increasing the evaluators capability.

### 9.2.3. Long Term Alternatives

Some of the alternatives have attributes that would take a significant amount of time and study to successfully implement. These have been categorized as long term alternatives because they offer increased benefit, but with a higher organizational cost and longer lead time. Key to these alternatives is a more integrated and capable augmentation system. This type of system would seek to capture metrics of process and performance with non-destructive means, evaluate the data, and provide evaluators and trainees insights into their iteration – ideally with recommendations for remediation. This next step in augmentation would require significant research, tuning, and organizational investment. It would however provide objective insights into a unit's capability.

Sim19 – (Improved Feedback, Evaluator Training, Moderate Augmentation). This alternative uses the gains from improved feedback format and evaluator training seen in Sim18, but adds moderate augmentation.

Sim20 – (Improved Feedback, Mater Trainer (Internal), Moderate Augmentation). This option contains similar parameters as Sim12, but increases augmentation to a moderate level.

### 9.2.4. Unit Training Sampling

External evaluators are a huge organizational hurdle. Having them at every training event is simply not a realistic option. However; because of the standard progression of the infantry training methodology, there is an opportunity to conduct sampling at a specific team even with external evaluators. In these alternatives, external evaluators could be utilized only for a specific event type of a unit's training progression. For example, the squad live fire exercise is a culminating event that happens towards the end of a units collective training. This would be a fantastic option to utilize sampling to calibrate previous (and future) reporting from the unit. It would also be possible to lock certain aspects of the training event to allow for more transparent and usable comparisons. Holding training variables constant (scenario, start point, target array, resources ...) would allow for external evaluators to help provide accurate comparisons between units that execute that training event – even across broad time domains. It would be similar to how the Army conducts individual and gunnery marksmanship training (set target array, distances, ammunition ...). Figure 43 demonstrates the relative effect of holding certain environmental conditions constant and its effects on evaluation. Those training events that have less task and context variability have the opportunity to be more objectively evaluated.

### Task Variability and Assessment Objectivity

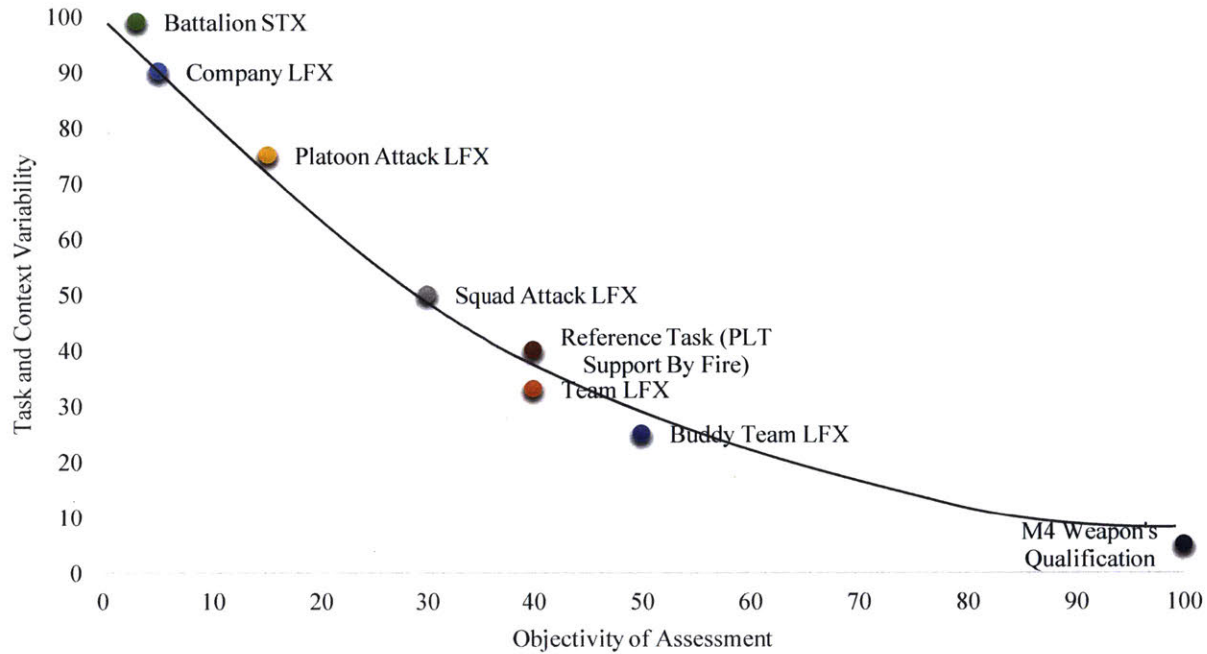


Figure 43: Task Variability and Assessment Objectivity.

When context and task variability reduces, we expect to see our ability to objectively assess capability to increase. This relationship is convex as increases in organizational focus increase in size non-linearly. Organization size and the amount and variability of tasks a unit conducts increases significantly at each level. With this type of sampling system we could start to see changes in unit capability and establish baselines with which we could compare units as equipment, manning, and training techniques progress. With this idea of progression sampling, the following alternatives were selected to examine relative effectiveness:

Sim21 – (Improved Feedback, Master Trainer, Evaluator Organizational Placement (2+ Levels, No Augmentation)). This alternative utilizes parameters that simulate the effect of having a master trained external evaluator (at the brigade or higher level) evaluate a company with an improved feedback format.

Sim13 – (Improved Feedback, Master Trainer, Evaluator Organizational Placement (2+ Levels, Minimal Augmentation)). Sim13 is similar to Sim21, but adds minimal augmentation, boosting evaluator capability and reducing inflation through increased objectivity.

## A Systems Analysis of the Army's Tactical Evaluation Process

Sim14 – (Improved Feedback, Master Trainer, Evaluator Organizational Placement (2+ Levels, Moderate Augmentation)). Sim14 increases the level of augmentation to a moderate level from Sim13.

### 9.2.5. Refined Option Analysis

The refined options were evaluated using the model in the same fashion as the initial alternatives. The tradespace graphs were adjusted slightly to help identify the different category from which each alternative originated.

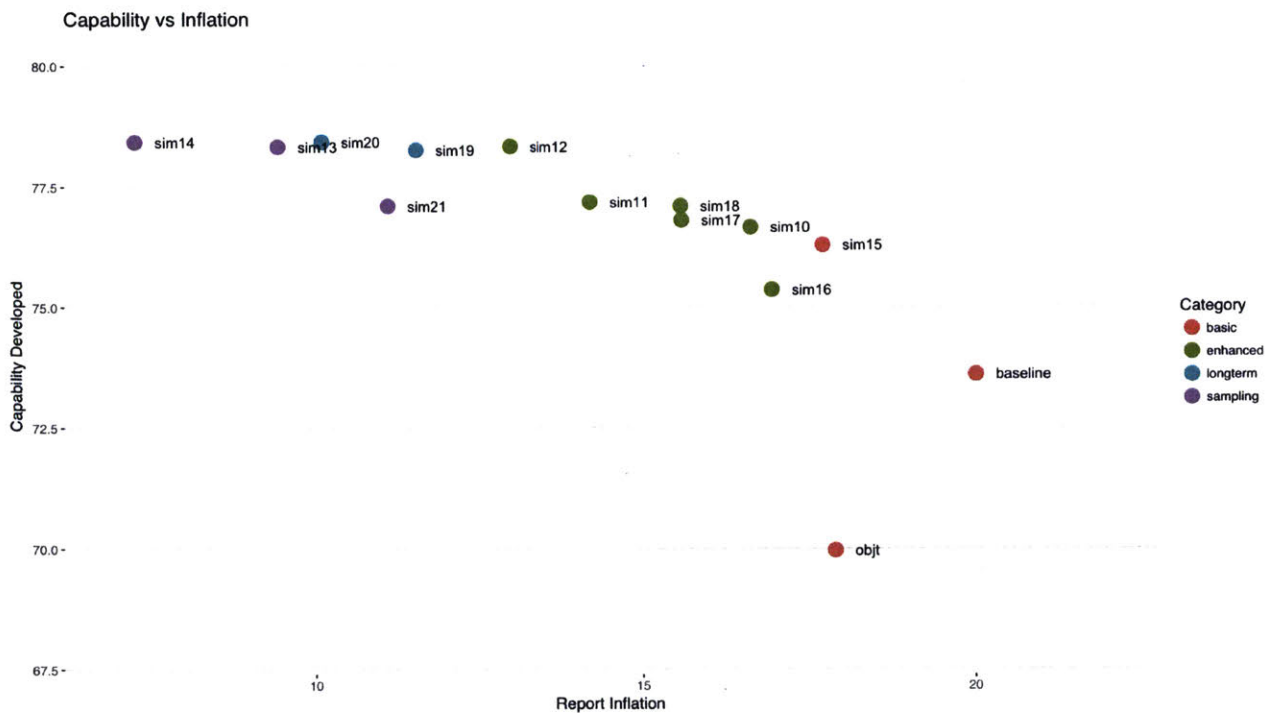


Figure 44: Capability vs Inflation - Refined Options

We see that report inflation seems to be differentiated by simulation category, with basic having the most inflation and sampling and long term alternatives having the least. There is some overlap between sampling and long term (Sim20 and Sim21), but the trend holds as original analysis suggested. Developed Capability is relatively flat between all alternatives except OBJ-T and Baseline. This is not surprising as alternatives were selected to both improve capability development and reduce report inflation. There is however less variance between the selected alternatives, in regards to capability development, than was originally hypothesized.

## A Systems Analysis of the Army's Tactical Evaluation Process

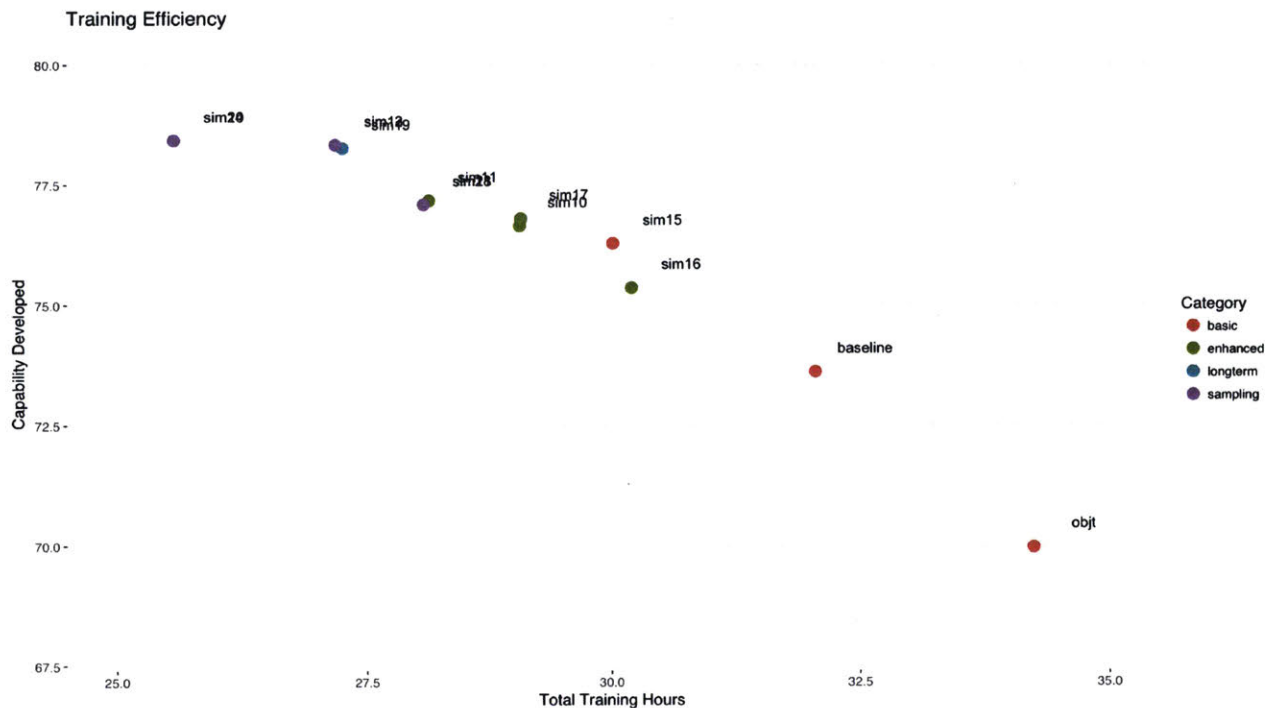


Figure 45: Training Efficiency - Refined Options

The Training Efficiency trend holds between the refined and initial simulations. The effect of evaluator fatigue seems to overpower any significant gains in longer duration training events. Alternatives that gain the highest levels of capability do so in the most efficient fashion as opposed through greatly training intensity (longer hours).

Within the refined alternatives, augmentation displays more of a dampening effect on report inflation than an increased in capability development. Figure 46 highlights this fact by coloring the alternatives by their associated level of augmentation. Alternatives with higher levels of augmentation trend closer to the “perfect reporting” line.



# A Systems Analysis of the Army's Tactical Evaluation Process

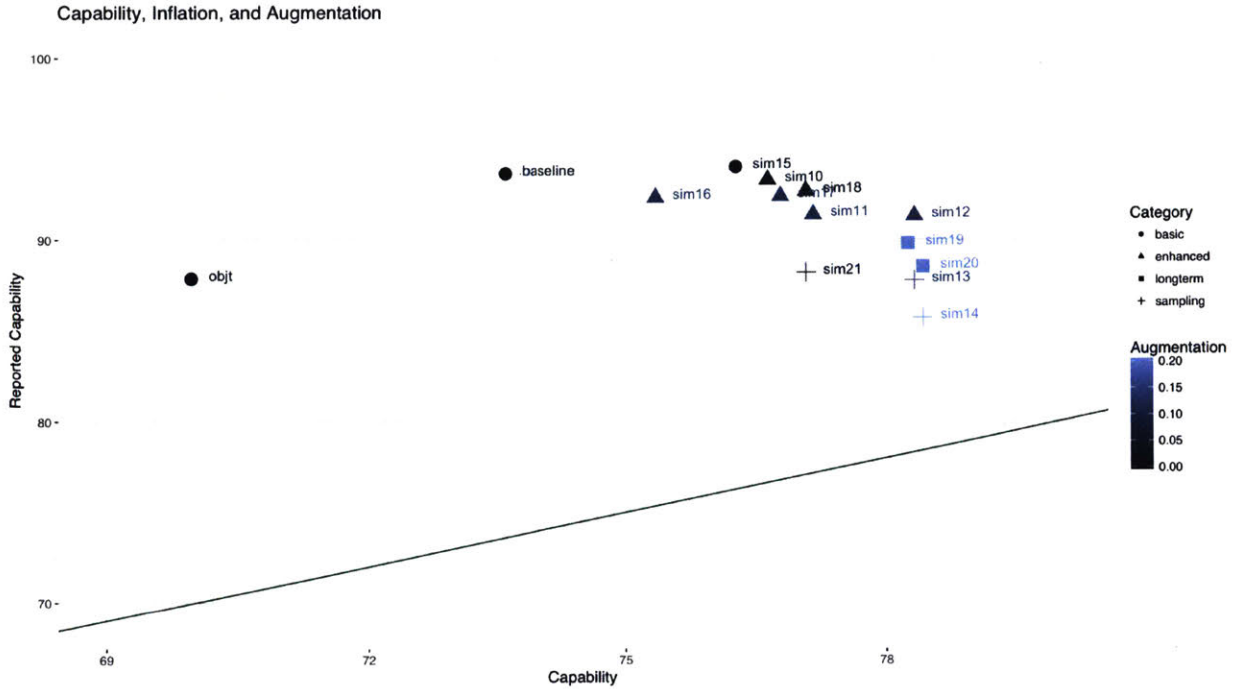


Figure 46: Capability, Inflation, and Augmentation - Refined Options

Interestingly enough the basic alternative category holds the greatest spread of capability developed. In Figure 47, we see that sim15, of the basic category, clusters near the enhanced alternatives. The change in feedback format, and the effects of standardization, dramatically change the system outcomes. This refined analysis again demonstrates the significant effects of this parameter and suggests that investment in improving and standardizing feedback format provide significant gains.

# A Systems Analysis of the Army's Tactical Evaluation Process

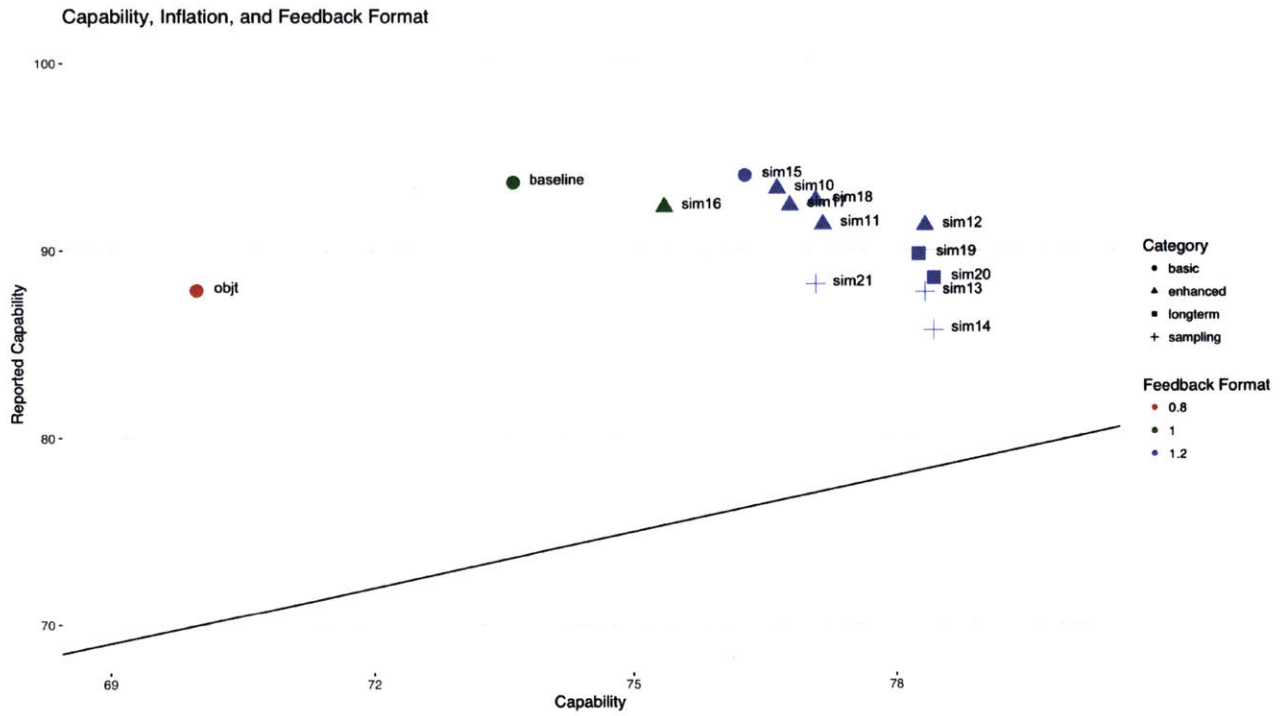


Figure 47: Capability, Inflation, and Feedback Format - Refined Options

## 10. Limitations and Recommendations

### 10.1. Limitations

The system dynamics model uses assumptions from several areas outside of its original field of study. While attempts were made to identify relevant literature and empirical evidence for interactions and gains within the model, the tactical domain has not been studied with the same rigor as many other fields. Using assumptions from other domains limits the ability of the model to deliver truly quantitative results. Unfortunately, data collection on tactical training within the Army fails to capture details or demonstrate enough consistency to use for calibration. The dynamics and response of the model match certain expectations and previous experiences of the author and other consulted practitioners but cannot be used for finite quantitative analysis. At best, this is a qualitative design tool that should be used for investigation of alternatives and to guide future work.

Some of the features and alternatives contain vague descriptions that do not lend themselves to immediate implementation. For example, augmentation does not specifically indicate what would be measured, how it would be measured, or how that information would be utilized. Assumptions during alternative generation rely on the belief that these alternatives would be studied and implemented in a positive and consistent manner.

Additionally this analysis restricts itself to the development of capability during the execution of a representative 72 hours exercise. It may be helpful to further expand the analysis to include the impacts of different alternatives on a larger timescale (12-18 months). Expanding the time domain would help understand additional interactions and possible effects of alternatives on the more comprehensive task of preparing a unit for war from individual to culminating events. This type of analysis might also demonstrate additional or stronger artifacts of the tactical development process. For example, the flexibility of a unit's schedule and resources allows them to address training deficits that were identified in training event. Increases in flexibility might reduce the tendency of units to inflate reports because they can still address these gaps in capability before culminating events.

### 10.2. Recommendations

#### 10.2.1. Key Takeaways

The biggest takeaway from this endeavor is the need for the infantry, and perhaps the Army, to begin to seriously study our trade. The inability to find comparable literature or empirical evidence within in the domain considerably dampens any possible insights or conclusions. We do not sufficiently understand how process, performance, and feedback intertwine within the tactical environment. Our usage of best practice and lesson's learned has allowed us to develop an extremely capable fighting force. However, this type of learning often relies on failures in combat. Often the lessons are inculcated to a degree where even the established practitioners have difficulty applying new or novel equipment or techniques (i.e. automatic fire on M4s). We must reframe the idea of tactical training to include the idea of experimentation. Key to this idea is thinking of training as a platform for data driven military study. Training must serve as a living lab so that ideas and the evolution of tactics can proliferate the force at much higher speeds (Pentland 2015). We must not simply learn drills, but learn how to most efficiently win in combat.

The TEOs were abandoned for a reason. While almost all evaluators said that they used them for reference, the overarching consensus was that they were not effective as is. We must seek to fully understand the gains from our generational learning and codify them into a better format. While we may not completely understand why they work, years of trial and error and yielded a more effective method for feedback. NTC utilizes competency modeling to augment their reporting and the literature seems to suggest this is a more optimal solution for tactical training (Campion et al. 2011; Michael T. Brannick et al. 1997). Analysis shows this to have a significant impact across all simulations and should be leveraged as such.

Augmentation provides a method to collect truly objective data for evaluation and future study, but integration will take significant effort to make useful for tactical evaluations. Because we currently have an immature understanding of the linkages between feedback type, level, and focus, any data collected on these features will need to be studied for its effect on future capability. This data will also need to be integrated into the tactical evaluation system and provide for easy interpretation from evaluators and trainees. These tasks are not trivial, as most sports teams have demonstrated in their search for insightful analytics. However, to ignore this possibility due to its difficulty will hurt future study and the development of its capability. The Army cannot hope that

## A Systems Analysis of the Army's Tactical Evaluation Process

other domains (sports, aviation, medicine...) will develop a method to integrate analytics and take that as an off the shelf solution. The Army must build its augmentation system itself. There are easy first steps that could possibly provide amazing improvements. Currently, training events do not capture target feedback with any fidelity. This is an example of an easy first step to augmenting the training environment for both evaluation and study's sake.

### 10.2.2. Vision for the Future: Now – Ten Years

Results from the analysis provide a few insights into the possible paths forward for the Army's tactical evaluation system. Because the costs and relative time requirements for some alternatives varies, it is helpful to think about these items in terms of the previously prescribed categories (basic, enhanced, long term, and sampling). Regardless of alternative, the value of an improved feedback format cannot be overlooked. This is the easiest of all implementations and provides perhaps the smallest cultural hurdle. For this reason the first step in this tactical training path must be the study and standardized adoption of the feedback developed through generational learning (see sim15 in Figure 48).

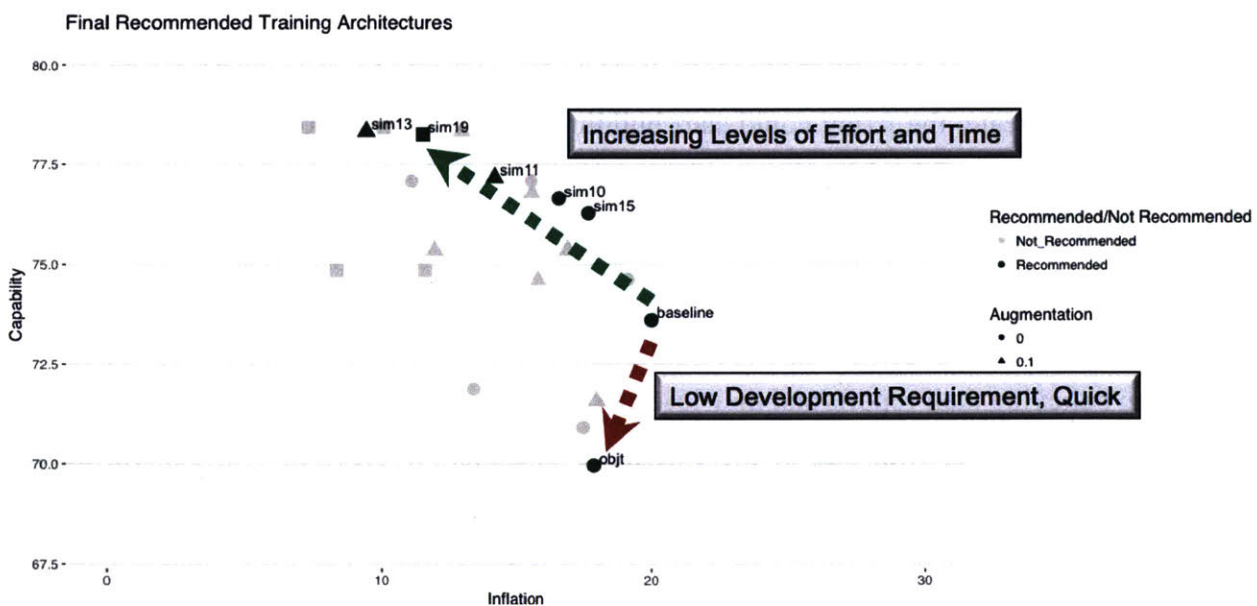


Figure 48: Recommend Tradespace. There are two distinct approaches to reducing inflation within the Army's Tactical Evaluation System. Inflation can be mitigated by standardizing reporting and forcing usage. This approach has extremely low development requirements, but unintentionally reduces the teams developed capability. The other approach standardizes reports, but invests in improving the feedback format to leverage generational learning and possibly integrate evaluator training and sensor augmentation.

As the next step, we look to the enhanced alternatives. In an effort to increase standardization, it would be important to develop and include a portion of evaluator training that

addresses the new format and the reasons for its usage. This will not only increase overall usage through exposure, but provide a mechanism to ensure that all evaluators fully understand how to implement this new doctrinal tool effectively. This training could take place at home station or during any of the pre-established mid-career educational courses. Because of the inundated curriculum, it is recommended that this training be basic (sim10 and sim11). While increased training through a master training program (sim12) does increase the overall capability developed, this increase is marginal and comes with more significant organizational costs. In addition to evaluator training, it is recommended that minimal augmentation be developed for use as an evaluation tool. Minimal augmentation not only reduces report inflation, but it lays the basis for continued study and empirical comparison of training units. While not the explicit goal of this work, the necessity of the continued study of the tactical domain cannot be understated. For these reasons, sim11 should be the midterm goal for Army tactical evaluation.

Unit performance sampling is the final recommendation of this analysis. This is a slightly different suggestion than the previous, because it is a temporary evaluation method or condition that leverages the effect of having external evaluators (see sims 13, 14, and 21). The Army has traditionally used external evaluators to provide feedback and help train units, but often the evaluators do not provide performance reports (CTCs). By selecting specific events for an external evaluator to adjudicate, the Army could start building like comparisons between units in a more objective manner. The advantage of this system is that it could be implemented quickly and have broader effects outside of the targeted event. For example, a unit may be less prone to inflate their capability prior to an externally evaluated event because of the possibility of dissonance between expectations and performance. If a unit consistently rates itself high, but fails a training event, people stop believing their reports and they lose credibility. After a few cycles, units would be able to calibrate their ratings with that of the external evaluators. As previously discussed, holding environment and task variability as constant as possible would amplify the Army's ability to conduct comparisons.

### 10.2.3. Vision for the Future: The Next 50 Years

As our understanding of the intricacies of teamwork, performance, and process develop within the tactical domain, evaluations will become extremely detailed and prescriptive in remediation as well as capability aggregation. For example, a platoon conducts a training event and at the end each soldier and team receives a series of tasks to work on and metrics for comparison. These outputs develop remedial training guidance or prescription, with which could be used to predict the platoons new capability level after this individual and sub-unit remediation – all without having to conduct the platoon exercise again. This type of system is demonstrated in sim19 with evaluator training, improved feedback formats, and moderate levels of augmentation. As the Army's understanding of the tactical domain increases, training progressions and tradeoffs will be easier to predict, evaluate, and optimize.

Additionally, the study of the tactical domain, combined with sensors, and analytics could yield predictive models that could be operationalized in a combat scenario. Often mission allocation and planning rely on assumptions and subjective assessments of unit status. What if a unit's performance and status could be objectively measured during combat operations to drive decision making? It is not unrealistic to think that, with dedicated effort, the Army could have systems that allow leaders to see unit fatigue and performance statistics in real time and be able to predict future performance on pre-defined tasks. Planners could augment their mission allocations, resource commitments, and risk tolerances according to a more accurate picture of unit readiness.

While this work does not significantly outline this very real future, it does layout the near and mid-term importance and system gains that can be achieved through increased diligent study of military teams. Without this study and the continued development of tactical sensors, the more futuristic ideas and benefits will remain out of reach. It is my sincere hope that this work and the possibility of improving our future operational capabilities helps shape conversation and nudges our military leaders to consider new opportunities within the tactical domain.

## 11. Bibliography

- "ADP 1: The Army." 2012. HEADQUARTERS DEPARTMENT OF THE ARMY.
- "ADP 3-0: Operations." 2016. HEADQUARTERS DEPARTMENT OF THE ARMY.  
[https://www.army.mil/e2/rv5\\_downloads/info/references/ADP\\_3-0\\_UL0\\_Oct\\_2011\\_APD.pdf](https://www.army.mil/e2/rv5_downloads/info/references/ADP_3-0_UL0_Oct_2011_APD.pdf).
- Bommer, William H., Jonathan L. Johnson, Gregory A. Rich, Philip M. Podsakoff, and Scott B. Mackenzie. 1995. "On the Interchangeability of Objective and Subjective Measures of Employee Performance: A Meta-Analysis." *Personnel Psychology* 48 (3): 587–605. doi:10.1111/j.1744-6570.1995.tb01772.x.
- Brannick, Michael T., Eduardo Salas, Pegasus Professor and University Trustee Chair Department of Psychology Eduardo Salas PhD, and Carolyn W. Prince. 1997. *Team Performance Assessment and Measurement: Theory, Methods, and Applications*. Psychology Press.
- Campion, Michael A., Alexis A. Fink, Brian J. Rugeberg, Linda Carr, Geneva M. Phillips, and Ronald B. Odman. 2011. "Doing Competencies Well: Best Practices in Competency Modeling." *Personnel Psychology* 64 (1): 225–62. doi:10.1111/j.1744-6570.2010.01207.x.
- Cannon-Bowers, Janis A., and Eduardo Salas. 1997. "A Framework for Developing Team Performance Measures in Training." In *Team Performance Assessment and Measurement: Theory, Methods, and Applications*, edited by M. T. Brannick, E. Salas, and C. Prince, 45–62. Series in Applied Psychology. Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.
- Chi, Michelene. 2006. "Two Approaches to the Study of Experts' Characteristics." In *The Cambridge Handbook of Expertise and Expert Performance*, 21–30.  
<http://learnlab.org/uploads/mypslc/publications/chi%20two%20approaches%20chapter%202006.pdf>.
- Craft, Jana. 2013. "A Review of the Empirical Ethical Decision-Making Literature: 2004-2011." *Journal of Business Ethics* 117 (2): 221–59. doi:10.1007/s10551-012-1518-9.
- Dickinson, Terry L., and Robert M. McIntyre. 1997. "A Conceptual Framework for Teamwork Measurement." In *Team Performance Assessment and Measurement: Theory, Methods, and Applications*, edited by M. T. Brannick, E. Salas, and C. Prince, 19–43. Series in Applied Psychology. Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.
- Feng, Wen. 2013. "Strategic Management for Large Engineering Projects : The Stakeholder Value Network Approach." Thesis, Massachusetts Institute of Technology.  
<http://dspace.mit.edu/handle/1721.1/80983>.
- Feng, Wen, Edward F. Crawley, Olivier L. de Weck, Rene Keller, and Bob Robinson. 2010. "Dependency Structure Matrix Modelling for Stakeholder Value Networks." *MIT Web Domain*. <http://dspace.mit.edu/handle/1721.1/81156>.
- "FM 3-21.10: The Infantry Rifle Company." 2006. HEADQUARTERS DEPARTMENT OF THE ARMY.
- "FM 3-21.20: The Infantry Battalion." 2006. HEADQUARTERS DEPARTMENT OF THE ARMY.
- "FM 3-96: Brigade Combat Team." 2015. HEADQUARTERS DEPARTMENT OF THE ARMY.
- "FM 7-0: Train to Win in a Complex World." 2016. HEADQUARTERS DEPARTMENT OF THE ARMY.

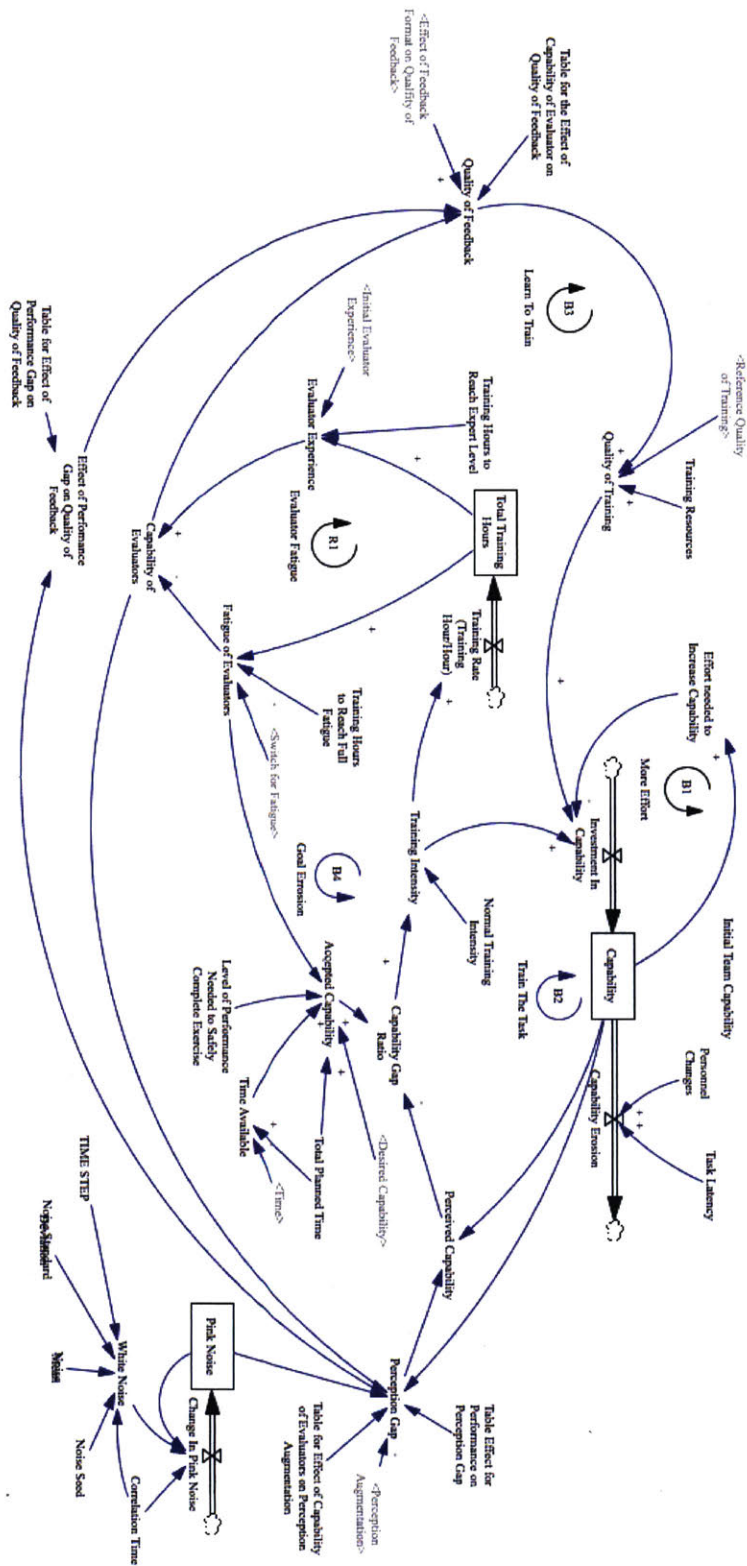


- Gilhooly, K. J., P. McGeorge, J. Hunter, J. M. Rawles, I. K. Kirby, C. Green, and V. Wynn. 1997. "Biomedical Knowledge in Diagnostic Thinking: The Case of Electrocardiogram (ECG) Interpretation." *European Journal of Cognitive Psychology* 9 (2): 199–223. doi:10.1080/713752555.
- Jones, Thomas M. 1991. "Ethical Decision Making by Individuals in Organizations: An Issue-Contingent Model." *The Academy of Management Review* 16 (2): 366–95. doi:10.2307/258867.
- Jones, Thomas M., and Lori Verstegen Ryan. 1998. "The Effect of Organizational Forces on Individual Morality: Judgment, Moral Approbation, and Behavior." *Business Ethics Quarterly*.  
<http://libproxy.mit.edu/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=cookie,sso,ip,uid&db=edsjsr&AN=edsjsr.10.2307.3857430&site=eds-live>.
- Klein, Cameron, Deborah DiazGranados, Eduardo Salas, Huy Le, C. Shawn Burke, Rebecca Lyons, and Gerald F. Goodwin. 2009. "Does Team Building Work?" *Small Group Research* 40 (2): 181–222. doi:10.1177/1046496408328821.
- Kleingeld, Ad, Heleen van Mierlo, and Lidia Arends. 2011. "The Effect of Goal Setting on Group Performance: A Meta-Analysis." *Journal of Applied Psychology* 96 (6): 1289–1304. doi:<http://dx.doi.org/10.1037/a0024315>.
- Lawton, Craig, Dwight Miller, and James Campbell. 2005. "Human Performance Modeling for System of Systems Analysts: Soldier Fatigue." SAND2005-6569. Sandia National Laboratory. <http://prod.sandia.gov/techlib/access-control.cgi/2005/056569.pdf>.
- Martin, Luc J., Albert V. Carron, and Shauna M. Burke. 2009. "Team Building Interventions in Sport: A Meta-Analysis." *Sport & Exercise Psychology Review* 5 (2): 3–18.  
<http://libproxy.mit.edu/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=cookie,sso,ip,uid&db=a9h&AN=43803527&site=eds-live&scope=site>.
- McEwan, Desmond, Geralyn R. Ruissen, Mark A. Eys, Bruno D. Zumbo, and Mark R. Beauchamp. 2017. "The Effectiveness of Teamwork Training on Teamwork Behaviors and Team Performance: A Systematic Review and Meta-Analysis of Controlled Interventions." *PLOS ONE* 12 (1): e0169604. doi:10.1371/journal.pone.0169604.
- Nahrgang, Jennifer D., D. Scott DeRue, John R. Hollenbeck, Matthias Spitzmuller, Dustin K. Jundt, and Daniel R. Ilgen. 2013. "Goal Setting in Teams: The Impact of Learning and Performance Goals on Process and Performance." *Organizational Behavior and Human Decision Processes* 122 (1): 12–21. doi:10.1016/j.obhdp.2013.03.008.
- "No Time, Literally, for All Requirements | Association of the United States Army." 2017. Accessed January 19. <https://www.ausa.org/articles/no-time-literally-all-requirements>.
- O'Dea, A, P O'Connor, and I Keogh. 2014. "A Meta-Analysis of the Effectiveness of Crew Resource Management Training in Acute Care Domains." *POSTGRADUATE MEDICAL JOURNAL* 90 (1070): 699-U47.  
<http://libproxy.mit.edu/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=cookie,sso,ip,uid&db=edswsc&AN=000345916800007&site=eds-live&scope=site>.
- Oliva, Rogelio. 2001. "Tradeoffs in Responses to Work Pressure in the Service Industry." *California Management Review* 43 (4): 26–43.  
<http://libproxy.mit.edu/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=cookie,sso,ip,uid&db=bth&AN=5244711&site=eds-live>.

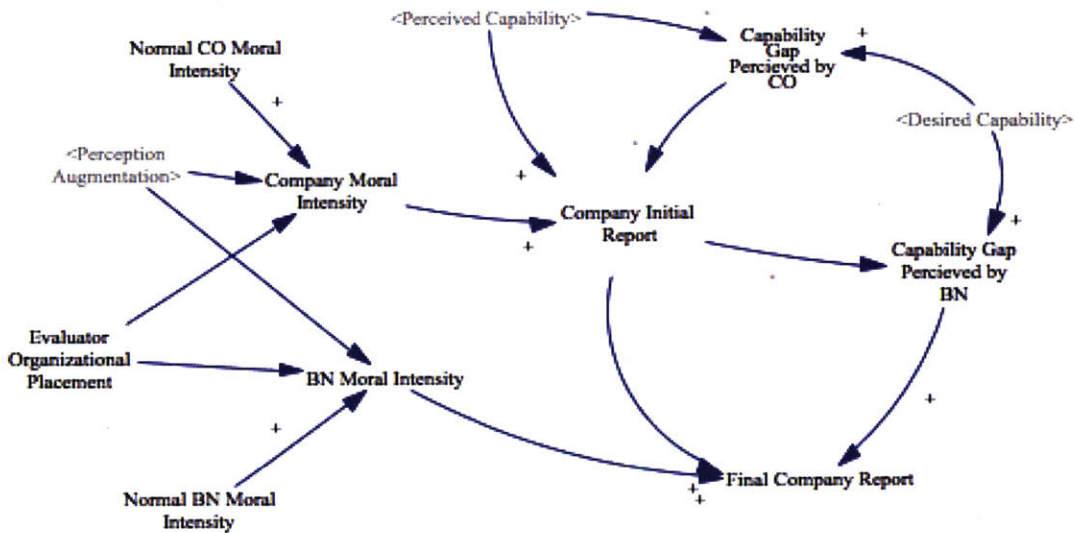
- Pentland, Alex. 2015. *Social Physics: How Social Networks Can Make Us Smarter*. Reissue edition. Penguin Books.
- Repenning, Nelson, and John Sterman. 2016. "Nobody Ever Gets Credit for Fixing Problems That Never Happened: Creating and Sustaining Process Improvement." Accessed September 14. [http://web.mit.edu/nelsonr/www/Repenning=Sterman\\_CMRSu01\\_.pdf](http://web.mit.edu/nelsonr/www/Repenning=Sterman_CMRSu01_.pdf).
- Salas, Eduardo, Deborah DiazGranados, Cameron Klein, C Shawn Burke, Kevin C Stagl, Gerald F Goodwin, and Stanley M Halpin. 2008. "Does Team Training Improve Team Performance? A Meta-Analysis." *Human Factors* 50 (6): 903–33. <http://libproxy.mit.edu/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=cookie,sso,ip,uid&db=mdc&AN=19292013&site=eds-live&scope=site>.
- Salas, Eduardo, Diana R. Nichols, and James E. Driskell. 2007. "Testing Three Team Training Strategies in Intact Teams: A Meta-Analysis." *Small Group Research* 38 (4): 471–88. doi:10.1177/1046496407304332.
- Salas, Eduardo, Drew Rozell, Brian Mullen, and James E. Driskell. 1999. "The Effect of Team Building on Performance: An Integration." *Small Group Research* 30 (3): 309–29. doi:10.1177/104649649903000303.
- Salas, Eduardo, Scott I. Tannenbaum, Kurt Kraiger, and Kimberly A. Smith-Jentsch. 2012. "The Science of Training and Development in Organizations: What Matters in Practice." *Psychological Science in the Public Interest: A Journal of the American Psychological Society* 13 (2): 74–101. doi:10.1177/1529100612436661.
- Senge, Peter M. 2006. *The Fifth Discipline: The Art & Practice of The Learning Organization*. Revised & Updated edition. New York: Doubleday.
- Sterman, John D. 2000. *Business Dynamics: Systems Thinking and Modeling for a Complex World with CD-ROM*. HAR/CDR edition. Boston u.a.: McGraw-Hill Education.
- Taylor, Frederick Winslow. 1911. *Principles of Scientific Management*. New York, London: Harper & Brothers.
- "United States Army: Organizational Information." 2017. *Organization | The United States Army*. Accessed January 9. <https://www.army.mil/info/organization/>.
- Wong, Leonard, and Stephen J. Gerras. 2015. "Lying to Ourselves: Dishonesty in the Army Profession." ARMY WAR COLLEGE CARLISLE BARRACKS PA STRATEGIC STUDIES INSTITUTE, ARMY WAR COLLEGE CARLISLE BARRACKS PA STRATEGIC STUDIES INSTITUTE. <http://www.dtic.mil/docs/citations/ADA615274>.

12. Annexes

12.1. System Dynamics Base Model



12.2. System Dynamics Moral Intensity



12.3. System Dynamics Model Code

Accepted Capability=  
 $\text{MAX}(\text{Level of Performance Needed to Safely Complete Exercise}, \text{MIN}(\text{Desired Capability} * (1 - \text{Fatigue of Evaluators}), (\text{Desired Capability} + \text{Level of Performance Needed to Safely Complete Exercise}) * (\text{Time Available} / \text{Total Planned Time})))$   
 Units: Capability [0,100]  
 The level of performance that will give a unit a "Go" or "Trained" designation. The level at which we stop investing in developing capability.

Capability= INTEG (  
 Investment In Capability-Capability Erosion,  
 Initial Team Capability)  
 Units: Capability [0,100]  
 The level of a unit's capability in a particular task or mission.

Capability Erosion=  
 Task Latency\*Personnel Changes  
 Units: Capability/Hour

Capability Gap Ratio=  
 $\text{MAX}(0, (1 - \text{MIN}(1, \text{Perceived Capability} / \text{Accepted Capability})))$

## A Systems Analysis of the Army's Tactical Evaluation Process

Units: Dmnl [0,1]

The Performance Gap is the ratio between the Perceived Performance and the Accepted Performance.

Capability of Evaluators=

$\text{MAX}(0, \text{Evaluator Experience} - \text{Fatigue of Evaluators})$

Units: Dmnl [0,1]

The capability of evaluators to capture performance.

Change In Pink Noise=

$(\text{White Noise} - \text{Pink Noise}) / \text{Correlation Time}$

Units: \*\*undefined\*\*

Correlation Time=

$\frac{2}{}$

Units: \*\*undefined\*\*

Desired Capability=

100

Units: Capability [50,100,10]

The level of performance that is initially desired at the beginning of a training event.

Effect of Feedback Format on Quality of Feedback=

1

Units: Dmnl [0.8,1.2,0.2]

This is the effect of your feedback format on the quality of your feedback. Better formats/methods produce better feedback while poor or poorly understood formats/methods will be discarded and not utilized.

Effect of Performance Gap on Quality of Feedback=

Table for Effect of Performance Gap on Quality of Feedback(Perception Gap)

Units: Dmnl

Effort needed to Increase Capability=

$\text{Capability} / 115$

Units: Capability/Training Hours

This is the amount of effort needed to increase capability. At low levels the effort needed to increase capability is extremely low, but it is almost impossible to reach the maximal amount of capability possible.

Evaluator Experience=

$(\text{Initial Evaluator Experience} + \text{Total Training Hours}) / \text{Training Hours to Reach}$

Expert Level

## A Systems Analysis of the Army's Tactical Evaluation Process

)^1

Units: Dmnl

Level of experience that evaluators have evaluating mission/task. Can have an initial value which represents the previously acquired experience before the training event.

Fatigue of Evaluators=

(Total Training Hours/Training Hours to Reach Full Fatigue)\*Switch for Fatigue

Units: Dmnl [0,1]

The relative fatigue of the evaluators due to evaluating training. This is a by-product of walking lanes, wearing kit, and all the associated mental and physical stresses that are experienced by holding this position.

Initial Evaluator Experience=

432

Units: Training Hours [0,1440,72]

The level of experience that the evaluators start the training event with in Training Hours. We assume normal level is a company commander who probably has 6 years of experience.  $72 * 6 = 432$ .

Initial Team Capability=

25

Units: Capability [0,75,25]

This is the initial amount of capability a team has the task to be trained.

Investment In Capability=

MAX(0,Quality of Training\*Training Intensity\*(1-Effort needed to Increase Capability

))

Units: Capability/Hour

The rate at which you are adding capability.

Level of Performance Needed to Safely Complete Exercise=

50

Units: Capability

Noise Mean=

0

Units: \*\*undefined\*\* [0,3,1]

Noise Seed=

3

Units: \*\*undefined\*\*

## A Systems Analysis of the Army's Tactical Evaluation Process

Noise Standard Deviation=

3

Units: \*\*undefined\*\* [0,6,1]

Normal Training Intensity=

0.75

Units: Training Hours/Hour

Perceived Capability=

$\text{MIN}(\text{SMOOTH}(\text{Capability} + \text{Perception Gap}, 3, 25), 100)$

Units: Capability [0,100]

The proportion of the actual performance that is understood by the evaluators. What is "seen."

Perception Augmentation=

0

Units: Dmnl [0,0.5,0.1]

The level of augmentation or feedback for perception. The baseline is 0, as there is no additional equipment or help for the evaluators to perceived or determine performance. A 1 would be perfect ability to capture performance.

Perception Gap=

$\text{Table Effect for Performance on Perception Gap}(\text{Capability}) * \text{Pink Noise} * \text{MAX}(0, (1 - (\text{Capability of Evaluators} + \text{Perception Augmentation}))) * (\text{MAX}(0, 1 - (\text{Capability of Evaluators} + \text{Perception Augmentation} * \text{Table for Effect of Capability of Evaluators on Perception Augmentation}(\text{Capability of Evaluators}))))$

Units: Capability

Difference between what is perceived and what is actual performance. This table produces the dampening effect of perception on the tails of performance.

Personnel Changes=

0

Units: Dmnl

Pink Noise= INTEG (

Change In Pink Noise,  
Noise Mean)

Units: \*\*undefined\*\*

Quality of Feedback=

$\text{MAX}(0, (\text{Table for the Effect of Capability of Evaluator on Quality of Feedback}(\text{Capability of Evaluators}) + \text{Effect of Feedback Format on Quality of Feedback}))$

## A Systems Analysis of the Army's Tactical Evaluation Process

) \*Effect of Performance Gap on Quality of Feedback)

Units: Dmnl

Quality of Training =

$(\text{Training Resources} + \text{Quality of Feedback}) * \text{Reference Quality of Training}$

Units: Capability/Training Hours

The Quality of Training is dependant on the level of resourcing and the quality of feedback. It is the amount of capability you gain for every training hour.

Reference Quality of Training =

0.7

Units: Capability/Training Hours [0.5,1.5,0.1]

The normal amount of capability gained per hour of training under nominal conditions.

Switch for Fatigue =

1

Units: \*\*undefined\*\* [0,1,1]

This turns on and off the effects of fatigue on evaluator capability. 0 is off. 1 is on.

Table Effect for Performance on Perception Gap(

$[(0,-20)-(100,20)],(0,10),(50,0),(100,10)$ )

Units: \*\*undefined\*\*

Table for Effect of Capability of Evaluators on Perception Augmentation(

$[(0,0)-(1,1)],(0,0),(0.1,0.5),(0.5,1),(1,0.75)$ )

Units: Dmnl

Perception augmentation (adding data, sensors, and objective measures) does not effect all capabilities of evaluators the same. Novices cannot leverage information in the same manner as experts. Additionally, experts may not need to see the data to be able to provide excellent feedback - so it won't improve their performance as much as someone in the middle.

Table for Effect of Performance Gap on Quality of Feedback(

$[(-10,0)-(10,10)],(-10,0),(0,5),(10,0)$ )

Units: Dmnl

Table for the Effect of Capability of Evaluator on Quality of Feedback(

$[(0,0)-(1,1)],(0,0),(0.25,0),(0.6375,0.5),(1,1)$ )

Units: \*\*undefined\*\*

Below the .25 level, evaluators are only focused on safety and not feedback. Once an evaluator gains capability over .25 they are able to direct some of their attention to providing quality



## A Systems Analysis of the Army's Tactical Evaluation Process

feedback to training units.

Task Latency=  
0

Units: Capability/Hour

The rate at which a unit loses proficiency due to time spent not training.

Time Available=  
Total Planned Time-Time

Units: Hour

Amount of hours left to train.

TIME STEP = 1

Units: Hour [0,?]

The time step for the simulation.

Total Planned Time=  
100

Units: Hour

The amount of time that is originally planned for training. You cannot train past this time because of resource constraints.

Total Training Hours= INTEG (  
"Training Rate (Training Hour/Hour)",  
0)

Units: Training Hours

Total amount of training hours spent during a training event.

Training Hours to Reach Expert Level=  
1440

Units: Training Hours

If we assume that a CSM or Battalion Commander is an expert training on a particular event, and they conduct this type of event annually, then expertise is the length of the event times about 20. In this case  $72 * 20 = 1440$ .

Training Hours to Reach Full Fatigue=  
24\*4

Units: Training Hours

Training Intensity=

IF THEN ELSE( Capability Gap Ratio > 0 , MIN(1,Normal Training Intensity+Capability Gap Ratio), 0 )

Units: Training Hours/Hour [0,1]

## A Systems Analysis of the Army's Tactical Evaluation Process

Training Intensity is the ratio of time you spend training in hours. For example, a value of .75 means you spend 45 min of every hour training.

"Training Rate (Training Hour/Hour)"=  
Training Intensity

Units: Training Hours/Hour

Training Resources=  
0.25

Units: Dmnl

The level at which a training event is resourced effects the amount of learning rate. More realistic training lends itself to higher rates of learning. The model has been adjusted so that .25 is representative of a high quality live fire training exercise.

White Noise=

Noise Mean+Noise Standard Deviation\*(Correlation Time/TIME  
STEP)\*(RANDOM UNIFORM  
( -1, 1 , Noise Seed ))  
Units: \*\*undefined\*\*